Event Numeration in Situation Model Construction:

Two *e* or Not to *e*

Fate O’Gara

A thesis submitted for the degree of

Doctor of Philosophy in Psycholinguistics

Department of Language and Linguistics

University of Essex

January 2024

Abstract

Comprehenders routinely construct abstracted representations of the situations described by language, integrate these representations with existing knowledge, and store these in memory for later access. How negation and distributivity interact with the underlying events of verbal predicates that aggregate into these message-level representations (known as situation models) is not well understood. Addressing this question is the principal focus of this thesis.

Across three related studies, I examine (i) the effect of verbal negation on access to conceptual information inside a negated event via a probe word task, including whether negation of elided and explicit clauses influences access to representations differently; (ii) the effect of ambiguously distributive constructions on event representations via a self-paced reading task, measuring responses to critical words consistent or inconsistent with collective and distributive readings, including whether manipulations of the timespan in which the event occurs affects these representations; and (iii) via two EEG experiments that measure prediction error and reanalysis effects, the role of negation and distributivity on a well-attested event processing phenomenon, wherein comprehenders possess *a priori* assumptions that events will be ordered singularly and contiguously.

This thesis demonstrates that verbal negation suppresses event representations from forming part of the developing mental model and violates comprehenders' expectations of event contiguity, with the effect being different in explicit clauses and elided ones. Ambiguously distributive and overtly distributive constructions lead comprehenders to reanalysing their expectations of singular-event contiguity, while both collective and distributive representations remain accessible following ambiguous texts; a preference for distributive readings is observed where the timespan in which the event can occur is extended. Taken together, these results have consequences for our theories of language comprehension: comprehenders monitor and mentally represent the number of events within their linguistic input, such that how events are enumerated influences the construction of a comprehenders' message-level situation model.

This work was supported by the Economic and Social Research Council's South-East Network for Social Sciences (SeNSS) studentship award.

Supervisors: Dr Kyle Jerro, Dr Laurel Lawyer

Contents

Chapter 1

On Events

and their Mental Representations

1.1 – Introduction

With language we tell stories: stories about current affairs, stories about our histories, stories about our futures, stories about our everyday experiences, even stories about fantastical worlds that do not exist. At the heart of storytelling are events. Events are the building blocks of the complex mental representations comprehenders develop during the processing of linguistic texts, whether spoken, signed, or codified in writing, known as situation models (Zwaan, 2016).

One can represent language in a variety of ways: the orthographic symbols in the word *dog*; the diagrammatic branching of a syntactic tree; the dots and dashes of Morse code. Each of these, and many more, represent some level of linguistic information, but none of these are the words or sentences that we utter or, more importantly, what we hold in our minds. The notion of mental representations captures both the intuition and observation that language users mentally encode certain aspects of words, phrases, and clauses as some abstracted (non-tangible) representation in memory, available for later access (Aitchison, 2012; Craik, 1943; Johnson-Laird, 2001; 2004). These mental representations are observed across multiple conceptual levels, ranging from the phonetic to the propositional, and beyond (Zwaan, 2016).

The high-level mental representation of the story described by some linguistic input is what is known as the situation model (Johnson-Laird, 1983; Kintsch, 1988). These situation models are argued to arise not only from single sentences, but also across inter-related strings of linguistic input, such as the text within a novel (McNerney et al., 2011). Explaining how these story-level mental representations are constructed (as well as encoded within and retrieved from memory) is central to theories of language comprehension and processing at levels beyond individual words (McNamara & Magliano, 2009). The representation of stories is intrinsically tied to the representation of the events – the things that happen – that comprise them, such that event representations are widely considered to be the fundamental unit of situation model construction (Ferstl et al., 2005; Huff et al., 2014; Kuperberg, 2016; Kuperberg & Jaeger, 2016; Zacks, 2020; Zwaan, 1996; Zwaan, 2016; Zwaan & Radvansky, 1998). However, to date, this body of literature has overwhelmingly focussed on single event representations – or sequences of one event leading into another event within a narrative – with little deliberation afforded to what happens to situation model construction when there are no events to mentally represent, or there are ambiguously many events to choose from. The following research project, therefore, aims to expand our knowledge of how the mental representations of events described within multi-sentence narratives arise and interact to establish event-complex representations that are more than the sum of their parts, and how these differ when no event has occurred or when ambiguously many events have occurred. Specifically, via two common structures in English – negation of verb phrases (which typically encode events) and plurality denoting expressions (that give rise to collective and distributive interpretations of events) – this thesis by paper examines on-line responses to short narratives in which the number of events that comprise the situation is manipulated, resulting in hypothesised representational differences between stories with no events, those with single events, and those with ambiguously many events. The centrality of number, or quantity, in the representation of events modified by negation and plurality denoting expressions in this thesis, gives rise for the need of an umbrella term to refer to the process or mechanism by which comprehenders determine how many events compose into their larger mental model. I will, therefore, use the term *event numeration* to capture the relatedness of the investigations that comprise this study.

With events being fundamental to this research, an important question is ‘what is an event?’. Cognitively construed, events are an emergent conceptual phenomenon that result from changes in the state or location (both spatial and temporal) of entities in the world, or within some described world (Altmann & Ekves, 2019). Interestingly, events conceived while perceiving a visual scene, like a movie, share the same representational properties as events conceived during the processing of stories we tell with language (Knott & Takac, 2020). There is no unique construction dedicated to the production of event meanings in language; verb semantics plays a crucial role, though, such as via the distinctions in the lexical aspect of accomplishments, activities, and achievements (Vendler, 1967), but adverbials and the modifiers and arguments of verbs also seem to introduce and range over event phenomena, such that some linguists have even proposed event arguments in the logical form of predicates to capture how event meaning manifests (Davidson, 1967; Parsons, 1990). To examine how event numeration – the determining of the number of events – contributes to story-level representations (i.e., situation models), I manipulate events expressed by Vendlerian V+NP accomplishment predicates, such as *ate an apple* or *bought a book*, wherein the event described is telic (has an endpoint), durative (spans time) and dynamic (in that there is a change of state in the entity undergoing the action event). Across three inter-related studies that form this thesis, I modify accomplishment predicates like those above using negation and plurality denoting expressions to shape whether the underlying semantics conveys a single event, no event, or multiple events, while obtaining some psycholinguistic processing measure of readers' access to critical conceptual representations consistent or inconsistent with one, less than one, or more than one event.

To illustrate, by negating an accomplishment construction, as in *didn't buy a book*, the situation described becomes one in which a book-buying event, arguably, did not occur. Probe word response times and event related potentials (ERPs) – see chapters 2 and 4, respectively – are obtained from comprehenders reading narratives in which such a construction is embedded. The results are then be compared to responses in an otherwise identical narrative with a non-negated accomplishment. By including a plurality denoting expression prior to the accomplishment verb phrase, as in *two students bought a book*, the situation described becomes ambiguous as to whether a single book-buying event or multiple book-buying events took place. Self-paced reading times to critical words and ERPs – see chapters 3 and 4, respectively – are obtained from comprehenders reading narratives with these ambiguous constructions. These findings are then compared to an otherwise identical narrative with a non-ambiguous control. Additional manipulations of both negated and plurality denoting accomplishments are also included to test a number of sub-questions, such as: whether event representations are differentially affected in explicit clauses and elided clauses, whether tense and aspectual features that redefine the temporal space in which events can occur impact these representations, and whether increasing gradations of negation or plurality give rise to an increased probability of constructing non-singular event representations.

Although negation and the phenomenon of distributivity each have a sizeable body of literature in both semantic theory and psychology, the processing of negation with respect to mental representations of the situation described has primarily developed from research on the negation of nouns and adjectives – with a dearth of research on the negation of verbal predicates that are frequently central to event phenomena – while the processing of plurality denoting expressions in ambiguously distributive constructions has received little theoretical connection to situation model construction altogether. Moreover, how event numeration interacts with well attested event-processing phenomena like the iconicity assumption – the tendency for comprehenders to implicitly assume that the order of events in a narrative progress from one (single event) to the next in a chronological and contiguous fashion, such that violations of this order result in increased processing costs (Duran et al., 2007; Ohtsuka & Brewer, 1992; Mandler, 1986; Zwaan, 1996) – is both not understood and ripe for investigation.

This research, therefore, explores an understudied space within the domain of mental representations, and promises to enrich our understanding of how comprehenders construct complex situation models when single events occur, when ambiguously many events occur, and when the event described did not occur, as well as their respective impact on known event phenomena like the iconicity assumption. Some existing hypotheses about situation model construction are intentionally left to the side throughout this thesis to promote simplicity, namely the role of embodied knowledge and mental simulations (discussed in section 1.3), although all findings and insights observed in this thesis should later be integrated with these factors in any full theoretical account of story-level representations.

Thus, chapter 1 of this thesis introduces evidence for the psychological reality of situation models, various strands of research on the nature of events, and the role of factors other than events that are well-known to contribute to story-level representations but whose full integration lies outside the scope this project. Chapter 2 introduces existing research on negation processing and then sets out the methods and results of a probe word experiment designed to investigate the role of verbal negation on event representations. Chapter 3 examines prior literature on ambiguously distributive constructions, then discusses the methods and results of a self-paced reading experiment exploring the role of these constructions on the representation of events. Chapter 4 reviews previous work on the iconicity assumption, a robustly observed event processing phenomenon, and then reports on the methods and results of two electroencephalography (EEG) experiments which look at how negation and distributivity interact with comprehenders expectations about the order of events within a narrative. Finally, chapter 5 synthesises the conclusions from the central three studies and place them within the context of the larger material on situation model construction; limitations of the present research and recommendations for future research are discussed. Throughout this thesis I use the terms situation model, mental model, and discourse/story-level/message-level representation interchangeably. There may be arguments for delineating these terms more precisely, but the distinctions one might draw between these terms are not relevant for this project.

1.2 – Representing Situations

1.2.1 – Evidence for the Situation Model

Findings from a wide body of psycholinguistic experiments motivate the core assumption of this thesis that comprehenders mentally model the situation described by narrative texts. For example, consider the following sentences, adapted from Bransford et al. (1972).

1) Three turtles rested on a floating log and a fish swam beneath them.
2) Three turtles rested on a floating log and a fish swam beneath it.

The situation described by sentences (1) and (2) are the same; despite a switch in anaphoric pronoun at the end of each sentence, what *happened* in these mini-stories is essentially identical – the fish swam beneath where the turtles were, which happens to be on top of a floating log. In fact, after presenting participants with a sentence like (1) Bransford et al. found that participants were unable to appropriately determine if they had encountered (1) or a similar example like (2) when presented with these options as probes in a task just three minutes after the original stimulus. This is not the case, however, with (3) and (4).

3) Three turtles rested beside a floating log and a fish swam beneath them.
4) Three turtles rested beside a floating log and a fish swam beneath it.

The only linguistic difference between (3) and (4) compared with (1) and (2) is the preposition *beside* instead of *on* modifying the verb *rested*. Nevertheless, these second set of sentences have a very different meaning to each other, unlike (1) and (2) which had essentially the same meaning. Note: both (1) and (3) contain the same anaphoric pronoun, *them*, as do (2) and (4), with *it*. The meaning of (3) is that the fish swam beneath the turtles who were next to a log, but the meaning of (4) is that the fish swam beneath the log which was next to the turtles. When presented with a sentence like (3), participants from Bransford et al.'s study were much more accurate at being able to identify the sentence they encountered when presented with sentences (3) or (4) as probes. In other words, Bransford et al. showed that comprehenders do not retain exact linguistic representations in memory: they were unable to distinguish a *them*-sentence such as (1) from an *it*-sentence such as (2) where the situation described the same thing, but this was not the case for sentences (3) and (4), where the situations described were different. It is possible (Bransford et al. point out) that this may simply be a consequence of some difference in the semantics of the prepositions *on* and *beside*; although, if this were the case, it would not explain similar patterning with a subsequent experiment the researchers conducted in this study (see sentences (5) through (8) below).

5) The box is to the right of the tree.
6) The box is to the left of the tree.
7) The tree is to the left of the box.
8) The tree is to the right of the box.

In this subsequent experiment, Bransford et al. showed participants sentences like (5), and found that, in a similar subsequent probe task three minutes later, it was the sentences which preserved the situation described that were most likely to be identified as having been previously presented in the experiment (i.e. examples (5) and (7), with the former being slightly preferred and the other two options being largely dispreferred). These results reinforce the initial experiment's findings that an exact linguistic representation of the sentence in long-term memory cannot be the right picture, and instead comprehenders produce some sort of mental abstraction of what happened in the text and retrieve this abstraction from long-term memory storage rather than the literal word-for-word text when required or prompted to do so.

Bransford et al.'s research was originally pitched in a discussion over interpretive versus constructive approaches to sentence memory, essentially does the reader rely on structural or semantic descriptions in the input and concluded that the research supported the latter. The framing of these findings has shifted in the five decades since this research was conducted, however. Subsequent researchers looking at sentence and discourse processing have re-interpreted the findings of Bransford et al. as evidence that a conceptual representation is constructed incrementally during the unfolding of an utterance, accumulating in a context-sensitive abstraction of what the text was about, and it is this abstraction that is ultimately stored in memory for later retrieval (see Ferstl et al., 2005; Huff et al., 2014; Johnson-Laird, 1983; Kintsch, 1988; Kuperberg, 2016; Kuperberg & Jaeger, 2016; Lin & Matsumi, 2022; Marslen-Wilson, 1975; Nieuwland & van Berkum, 2006; Raudszus et al., 2019; Zwaan, 1996). This context-sensitive abstraction is what is known as a situation model.

Research since Bransford et al.'s (1972) study has provided linguists with deeper insights into the nature of situation models. It has become apparent that comprehenders track and integrate various sources of conceptual and linguistic information, as the findings of the studies discussed in the remainder of this section illustrate.

Zwaan et al. (1998) review evidence that comprehenders' abstract representations of texts and utterances must be comprised of at least five dimensions or indices – causality, temporality, and spatial location, alongside the objects and protagonists involved in the situation, and their motivations or intentions. While comprehenders needed to be prompted to monitor spatial continuity in more complex situations, such as by being offered a detailed map of the setting, comprehenders showed reliable increases in reading times to discontinuities along the other four dimensions or indices. Similar observations were made in related research by Zwaan et al. (1995a). These findings were also supported in independent work by Albrecht and O'Brien (1993) who show that comprehenders are aware of and track inconsistencies about the characters and their motivations described earlier in a narrative and their subsequent actions in order to maintain global coherence.

Breaks in causal continuity (i.e., a lack of direct link between the current sentence and relevant prior story information) also results in increased reading times compared to sequences which explicitly frame these story connections, as comprehenders must instead infer the causal links themselves, resulting in more processing costs (Zwaan et al., 1995b). Indeed, take this passage from Samet and Schank (1984), shown in (9).

9) In a little Danish town, two fishmongers exchanged blows. Anders, by far the stronger, had a cousin in prison. When he was first convicted, Anders was living in Italy. Anders has a wife who lost her bathing cap. Her car is at this moment double-parked.

This narrative is unusual and rather difficult to parse because each sentence introduces a new protagonist or switches the location or time in focus, including failing to adhere to a clear chronology; moreover, it lacks clear and explicit connections between sentences to establish causal links. In other words, the narrative is odd because it does not describe a unique situation, but several small and seemingly unrelated ones, and thus readers will have a difficult time representing this in memory. A single situation model will be difficult to construct here.

Interestingly, time appears to be an especially important component of the situation models comprehenders construct. For example, Dowty (1986) recognised that readers appear to be inclined toward narratives that unfold in a chronological order; the author argued that for a given sentence in narrative discourse, comprehenders expect its temporal index to immediately follow the temporal index of the previous sentence, unless the given sentence contains adverbial information to the contrary. Dowty formalised this observation in his Temporal Discourse Interpretation Principle (TDIP), which he assumes is part of one's mental grammar (of English, at least). In more experimental work, Ohtsuka and Brewer (1992) observed that a mismatch between the temporal expectation and reality of a narrative's chronological structure actually led to decrements in accuracy on a subsequent comprehension task. Moreover, Mandler (1986) demonstrated how comprehenders were sensitive to a mismatch in the order of mention and order of occurrence, in line with Dowty's (1986) and Ohtsuka and Brewer's (1992) observations, but found this was only when the underlying events were not explicitly connected in a causal relation, leading them to conclude that a mental model of the temporal order of causally connected events might pre-exist, but this representation must be constructed by comprehenders for non-causally related events.

A fascinating conclusion drawn from data in studies like this is that comprehenders appear to hold *a priori* beliefs about the order of events described in discourse, and this can be modelled as part of their discourse representation. This belief is referred to as the iconicity assumption.

To illustrate, take examples (10) and (11), from Zwaan (1996).

10) The president opened the door, looked around, and coughed.

11) The president coughed, looked around, and opened the door.

In the above sentences, the natural interpretation is one in which the activities, achievements and accomplishments unfold in a temporal order that matches their order of mention. Bar-Lev and Palacas (1980) posit a Semantic Command Constraint (SCC) in the grammar (of English, at least) that ensures that, where sentences (or clauses) are conjoined, the second conjunct is not prior to the first, chronologically or causally. Between the SCC and Dowty's (1986) TDIP, mentioned above, it is clearer why the constructions in (10) and (11) lead to the interpretations they do. However, some psychologists have gone a step further, arguing for a non-linguistic origin to these interpretative rules. It has been proposed that, without cues to the contrary, the default assumption comprehenders hold is that the order in which events are reported in a narrative corresponds to the situations' chronological order, and that comprehenders make this assumption because it maps onto their own experience of chronological time in day-to-day life, hence being iconic (Zwaan, 1996; Duran et al., 2007). The TDIP and SCC are likely linguistic analogues of this temporal assumption about event ordering.

The body of research discussed here supports the notion that situation models are multifaceted, that comprehenders are sensitive to manipulations of key indices within the models they construct, and that time is an especially important component of mental representations of situations described. Furthermore, as Zwaan and Radvansky (1998) note, comprehenders appear to frequently make use of grammatical cues, such as aspectual information, temporal adverbials, word order and tense – and are sensitive to manipulations of these – in order to identify referents and the intentional and causal links between them as part of the process of building a coherent representation of the total linguistic input. It appears reasonably safe to conclude, then, that situation models are probably a real psychological phenomenon that are not only essential to how we process language but are heavily influenced by one's linguistic input. An important question that follows is how these models are constructed.

1.2.2 – From Texts to Events

In this section, I review two theoretical accounts for how comprehenders take some linguistic input and translate it into an abstracted representation of the situation described: the Construction-Integration model and the Event-Indexing model. Both models offer fruitful explanations of how situation models are constructed, although the former lacks the insight that most researchers now agree upon, which is that events are the fundamental unit of situation models construction, while the latter lacks some of the predictive precision that relates raw linguistic input to abstracted mental representations. Some combination of these models is likely to be true, and thus both models are referenced throughout this thesis where applicable. To motivate the importance of events in our theories of story-level representations, the discussion of these models also include an interlude on the nature of events themselves.

1.2.2.1 – The Construction-Integration Model

In the decade or so following the observations by Bransford et al. (1972) and the scenarios of fish swimming beneath turtles and logs, early attempts were made – such as Johnson-Laird's (1983) classic monograph on the science of language, inference, and consciousness – to theoretically motivate the concept of mental models as a fundamental property of the human cognitive system. Around the same time, the first substantive efforts to develop an explanation of how comprehenders transform the signs and arbitrary symbols of some text on a page into an abstracted representations of the situation described by that text were being proposed. Van Dijk and Kintsch's (1983) analysis of the micro- and macro-structures of a text, and Kintsch's (1988) subsequent expansion of these ideas into the Construction-Integration (CI) model, is perhaps the best-known and most long-lasting of these proposals.

The CI model was developed as a computational and connectionist counterproposal to existing top-down, expectation-based accounts of discourse comprehension at the time, where the processing of sentences was assumed to be a matter of the comprehender predicting each successive linguistic item in the unfolding input from those already parsed based on the comprehender's syntactic knowledge; for example, see Winograd (1983). While prediction as a theory of processing was revitalised in subsequent decades (see DeLong et al., 2005; Federmeier & Kutas, 1999; Kuperberg & Jaeger, 2016; Szewczyk & Schriefers, 2013; van Berkum, 2008; *inter alia*), the CI model proposes that discourse comprehension is largely data-driven (or bottom-up) and rule-based, with little in the way of conscious processes, proceeding along two ordered steps: knowledge construction and knowledge integration. During the construction phase, it was proposed that the linguistic representations of words are extracted from the text on a page and, in combination with the syntactic rules of the language, are formulated into a microstructure of individual propositions composed of predicates and their arguments; where argument overlap occurs between any two or more microstructural propositions these are then combined into an interrelated, macrostructural propositional network (Wharton & Kintsch, 1991). The microstructure and macrostructure of a text might, in some cases, be identical, as in the case of single-sentence discourses (van Dijk & Kintsch, 1983). At this point, a crude mental model of the discourse is represented via the nodal connections of the propositional network but is not yet interpretable; the rules used to generate this macrostructure are presumed weak enough to allow for the right interpretation to be among those generated, but other interpretations will also be generated that are irrelevant or inappropriate (Kintsch, 1988). To resolve the problem, the propositional network is then integrated with existing experience and knowledge – itself assumed to be an associative network – in a process of spreading activation which strengthens the contextually appropriate elements and inhibits undesirable ones to fine-tune the model for evaluation (Kintsch, 1988; Wharton & Kintsch, 1991). This fine-tuning results in a final text representation, or what has since come to be known as the situation model.

Evidence for this CI model can be seen in Kintsch et al. (1990) and Ratcliff and McKoon (1989) who independently observed that, in sentence-matching experiments, surface similarity is strongest during retrieval tasks, while paraphrases, inferences and unrelated distractors show diminishing activation in a graded fashion; furthermore, the former study found that expert readers utilised their knowledge more successfully, but required more processing time, while novice readers (who lacked situational understanding) made quicker judgements based on the surface characteristics of the experimental sentences.

While the CI model continues to receive some currency in the discourse processing literature, and notions of a dual system which makes use of constructing bottom-up information and integrating the composite product with existing levels of representation remains a central, if sometimes understated, feature in many accounts of situation modelling, the CI model is less widely reported than it once was, although the merits of its insights are still valuable and theoretically instructive, especially when attempting to explain how discourse representations arise from raw linguistic input. As Zwaan (2016) notes, however, it has largely been accepted that event representations, and not propositions, are the basic unit of the situation model, suggesting that the CI model is not the complete picture. The following subsection reviews what researchers mean by events, beginning with a formal analysis of events, followed by how events manifest cognitively, and ending with a discussion of the relationship between events and representational objects.

1.2.2.2 – An Interlude on the Nature of Events

1.2.2.2.1 – Formalising Events

Perfetti and Britt (1995) argue that propositional representations are insufficient for understanding texts and there is evidence that comprehenders instead attend to verb semantics (Zwaan et al., 1995a). Propositions have truth values – they are either true or false – and the situations in which a proposition is true is said to be its truth conditions (Ball & Rabern, 2018); events, on the other hand, do not have truth values, and are instead the sort of thing a sentence may be a true or false description of (Lasersohn, 1998). In this sense, events and situations are alike. But what exactly do linguists and psychologists mean by an event, and why is it that their representations, rather than propositions, are now considered the basic unit of mental models?

Continuous sensorimotor processes of perception are naturally segmented into discrete units: objects in the world and 'the things that happen', the latter being the broadest conception of events (Knott & Takac, 2020). Stated more precisely, events are discrete segments of time at a given location conceived by an observer to have a beginning and an end (Zacks & Tversky, 2001). These are both non-linguistic notions of events, such as the events of an unfolding movie scene or the evolution of actions and processes in the environment around us. Interestingly, linguists describe events in a similar fashion. According to Weiser (2008), events have three properties: they have causes and effects, they exist within a temporal and spatial location, and they happen. In Maienborn's (2011) estimation, events are spatiotemporal entities in the world that can vary in the way they are realised, but she notes that the search for the ontological criteria for identifying events remains ongoing. From the perspective of lexical aspect (or *aksionsart*), non-stative eventualities such as activities (*she danced*), achievements (*they reached the mountaintop*), accomplishments (*he ate the banana*) and semelfactives (*she knocked*) possess the known ontological criteria for conveying an event (Comrie, 1976; Vendler, 1967). Each of these eventuality types either share the temporal dimensions of durativity and telicity with another type, or they differ: accomplishments are telic (have an end point) and durative (they persist), while semelfactives are atelic and non-durative; activities are atelic and durative, and achievements are telic and non-durative. Time appears to be an essential property of events and is encoded grammatically through various tense and aspectual realisations, as well as lexically through temporal adverbials.

There is no single type of construction, at any level of the grammar, that is uniquely dedicated to the encoding of events; in fact, events might emerge from the use of individual lexical items (like the noun *destruction*) or more complex morphosyntactic units, like a Vendlerian accomplishment (such as *ate an apple*), or even from entire discourses (Bohnemeyer et al., 2011). Maienborn (2011) highlights additional diagnostics for determining where an event might be found, at least on a linguistic level: events, she argues, combine with locative and temporal modifiers (*on the table*; *after class*), with manner adverbials and comitatives (*in a hurry*; *with their friends*), and can serve as infinitival complements of perception verbs (*kiss Sam* in *Alex saw Mary kiss Sam)*. Since the work of Davidson (1967) – and certainly since Parsons (1990) – events have been considered a necessary component in the logical form of action sentences and are typically represented with an *e* event argument of predicates in the logical form. The distribution of events is such that it is not only verbs that introduce these

Davidsonian event arguments, but also prepositions, adverbs, adjectives, and nouns (Maienborn, 2011). In (12), an *e* event argument is an argument of the verb, *kiss*, expressed in the event-semantics of (13); that event argument, however, is shared across the locations, manners, times, and individuals involved in the entire proposition. Details relating to matters of tense and definiteness have been excluded from the analysis in (13) for simplicity.

12) At midnight, Alison kissed Barbara in the kitchen.

13) ∃*e* [KISS(*e*) & AGENT(*e*, a) & THEME(*e*, b) & IN(*e*, k) & AT(*e*, m)]

Example (13)[1] can be read as there is some event *e*, such that it is a kissing event, and Alison is the agent of that event, and Barbara is the theme of that event, and that event was in the kitchen, and that event was at midnight. Under this event-semantics view, all (dynamic) verbs are predicates ranging over events, and the verb's surface-level syntactic arguments are introduced by way of thematic roles, such as agent and theme, expressing a binary relationship between events and their participants (Maienborn, 2011).

What are the motivations for positing an event argument in the semantics of action sentences, though? Aside from its power in capturing the intuition that events, or "things that happen", are central to the meaning of non-stative verbs, the following empirical motivations inform the insights of Davidson (1967) as well as those who have since adopted the event-semantic approach to meaning analysis. Logical entailment is an important semantic relationship between sentences; if some sentence A entails some sentence B, then if A is true, B must be

[1] Formal approaches to the semantics of sentences typically adopt a meta-language using logical notation to avoid or minimise circularity in capturing meaning within a language. Various symbols, such as '∃' and '&' are often used, with the former corresponding to an existential quantifier binding some logical constant or variable (in this case the event *e*) and expressing existential commitment to that thing, and the latter corresponding roughly to the English conjunction 'and'. Predicates, such as KISS, IN and AT, as well as thematic roles such as AGENT and THEME appear in small caps followed (in brackets) by their argument or arguments, conventionally represented as a single character. In event-based semantics, all predicates also take the event (*e*) itself as an argument. Everything inside the square brackets is treated as being in scope of the existential quantifier. Event-based formalisms such as the one in (13) are used throughout this thesis to support various parts of the discussion and analysis.

true. For example, the sentence *There is a fluffy, brown cat* entails both the sentences *There is a fluffy cat* and *There is a brown cat*, and they all entail the sentence *There is a cat*. Logically, the initial sentence can be expressed as $\exists x$ [CAT(x) & FLUFFY(x) & BROWN(x)], meaning there is some individual x in the world such that x is a cat, x is fluffy, and x is brown, while the latter is represented as $\exists x$ [CAT(x)]. If the former is true, the latter must be true. Equally, sentence (12) above entails sentences (14) and (15), while (14), (15) and (12) each entail (16).

14) Alison kissed Barbara in the kitchen.

$\exists e$ [KISS(e) & AGENT(e, a) & THEME(e, b) & IN(e, k)]

15) At midnight, Alison kissed Barbara.

$\exists e$ [KISS(e) & AGENT(e, a) & THEME(e, b) & AT(e, m)]

16) Alison kissed Barbara.

$\exists e$ [KISS(e) & AGENT(e, a) & THEME(e, b)]

Interestingly, the conjunction of *I have a fluffy cat* and *I have a brown cat* does not entail the original sentence *I have a fluffy, brown cat*, as the x entities introduced in the semantics of the former two sentences could be referring to different cats, whereas there is just a single cat that is fluffy and brown in the latter; similarly, the conjunction of (14) and (15) does not entail (12), as the kissing event could have been a different one in (14) and (15), such that Alison kissed Barbara in the kitchen on Tuesday while she also kissed Barbara at midnight on Saturday, but the kissing event described in (12) is a single event that is in the kitchen and at midnight. The similarity in the entailment relations between the entities x and events e described by these two sets of scenarios calls for a similar underlying analysis. This was Davidson's (1967) seminal insight in positing an event argument introduced by predicates.

That events are arguments of the individuals, locations, and times expressed in a sentence at a semantic level establishes a noteworthy parallel to the observations from Zwaan et al. (1998), mentioned earlier, that comprehenders' abstracted representations of the situations described by texts are comprised of, among other properties, the temporal and spatial locations pertinent to the underlying events, as well as the objects and protagonists involved. Research in these areas developed independently, but similar conclusions about the nature of events were reached by linguists attempting to formally model what speakers know about the meaning of sentences and by psychologists attempting to understand the nature of the mental representations comprehenders construct while reading or hearing such sentences. Before continuing to describe an alternative to the Construction-Integration model as an explanation for situation model construction that incorporates events rather than propositions as the basic unit, events from the perspective of cognitive processing need to be discussed.

1.2.2.2.2 – Events in Cognition

Events play a crucial role in cognitive processing. There is a parallel between event-semantic formulations and event representational structures during processing, including events conceived outside of language. Knott and Takac (2020) go so far as to claim that the event representations activated in working memory during the sensorimotor experience of events (such as during the perceptual processing of a visual scene) are fundamentally the same kind of event representations that are studied by linguists formalising the semantics of action sentences, such that sentences reporting events are interpreted as traces of sensorimotor routines through which perceptual events are experienced. Pettijohn et al. (2016) argue that whether individuals experience a live event through perceptual means, or represent an event encoded linguistically in some narrative, they parse that stream of information into manageable units in a process of event segmentation. These segments are then stored in memory (Sargent et al., 2013). This capacity to segment events appears to develop early in life, with children showing a strong bias towards using spatio-temporal individuation strategies to discriminate objects and events both linguistically and non-linguistically (Shipley & Shepperson, 1990; Wagner & Carey, 2003).

In events both linguistic and otherwise, people continuously develop predictions about what will happen next in a situation, guided by an event model (Radvansky & Zacks, 2011; Zacks, 2020). An event model is a representation of what is currently happening (focussing more on the event representations themselves rather than their integration into larger situation models).Where a prediction error occurs (i.e., where something the individual does not anticipate happens) this results in the updating of that event model (see Altmann & Mirković, 2009; Bohnemeyer et al., 2007; Magliano & Zacks; 2011; Radvanksy & Zacks, 2014; Zacks et al., 2007; *inter alia*). Prediction errors tend to occur where features of some action or activity change, such as a shift in the entities or the temporal and spatial locations depicted in a visual or narrative scene, or a shift in the causal and motivation relationships that exist between entities within that scene; it is this feature change that is unexpected and that results in the prediction error, requiring the event model to be updated (Zacks, 2020). Consequently, an increase in prediction error is the catalyst for updating the current event model, leading to the perception of an event boundary and consequently the segmenting of events (Altmann & Ekves, 2019).

Recently, the Event Horizon model (Radvansky, 2012; Radvansky & Zacks, 2017), which draws upon discoveries about event segmentation processes, has been proposed in order to explain how comprehenders' event models, for both events in studies of visual perception as well as in studies of events encoded by the grammar, interact with memory and account for how some event boundaries appear to impair memory (Radvansky & Copeland, 2010; Radvansky et al., 2011), while some appear to improve it (Pettijohn & Radvansky, 2016a; Swallow et al., 2009). A fuller treatment of the Event Horizon model and non-linguistic event segmentation processes lies beyond the scope of this thesis; however, event segmentation and individuation research has been central in the recognition of events in language processing and, by extension, in the literature on mental representations of the situations described by texts.

1.2.2.2.3 – Events and Objects

A final important point relating to the topic of events is that they are closely related to mental representations of entities or objects[2]. Events share a great deal of conceptual real-estate with objects. Objects are said to occupy a spatial location and exclude other objects from that location, are able to move through that location and others, as well as endure through time in a manner in which they are wholly present at every point at which they exist; events, on the other hand, are understood to more easily tolerate co-location (such as sitting on a chair and reading while sat on that chair)[3], and take up time by having different parts of that event be at those different times, such as the incremental theme of mowing a lawn, which changes state and grows closer to completion as time passes (Dowty, 1991; Dretske, 1967; Quinton, 1979).

While events and objects are evidently different in how they manifest in time and space, the similarities in how events and objects can be described suggests that they may be ontologically related concepts. Goodman (2012) certainly thought so, arguing that objects and events are things of the same kind that simply differ in their state. Looking at the visual recognition and representation of events, Nevatia et al. (2004) similarly describes events as changes of state in an object. The semantics of certain verbs, such as accomplishments, are generally considered

[2] The notion of a (representational) object used here is broader than that used in reference to, say, syntactic objects, and instead refers to any kind of discourse entity, be it animate or otherwise, that forms part of the situation described. I use the terms entity and object interchangeably, in this respect.

[3] Lemmon (1967) makes the radical proposal that if two events occupy the same portion of space and time they are, in effect, identical. Maienborn (2011) notes that this would unfortunately result in examples like our chair sitting and reading activities mentioned in the parenthetical above being treated as the same event, which is probably not a claim we want to make. Davidson (1969) alternatively proposes that two events are identical if they have the same causes and effects, although a fully developed account of event causation is still under development (see Eckardt (2000) for an attempt at addressing this gap).

to consist of multiple subevents[4], with one or more objects in the event undergoing change between an initial state of affairs and an end state of affairs, such as the lawn in *mowed the lawn* (Rappaport Hovav & Levin, 1998). Cooper (1986) treats the semantics of sentences and discourses as the description certain kinds of histories (or courses of events) and not just the truth or falsity of these constructions in a situation; similarly, Altmann and Ekves (2019) argue that events, cognitively speaking, are ensembles of intersecting object histories in which one or more objects have changed state.[5] This conception of events as object histories, i.e., changes in state of objects across time and space, is one that is adopted in this thesis and is thus a notion that is implicit in any subsequent reference to events in this work.

In sum, events, like situations, are the sorts of things propositions are said to be a true or false description of. Events are objects having undergone or are in the process of undergoing a state of change, with spatiotemporal characteristics, such as telicity and durativity, that distinguish them from the typical conception of a 'static' object. When encoded linguistically, events are not introduced by any specific construction, and this insight has been formally modelled as an event argument over which predicates, including prepositions and thematic roles, range. This

---

[4] The accomplishment subevents referred to here are something like a more primitive action event and the change of state it brings about, formalised as [[x ACT] CAUSE [BECOME [y STATE]]] in which the first subevent [x ACT] is a representation associated with an activity and the second subevent is [BECOME [y STATE]], which is associated with the representation of an achievement, connected by a causal semantic primitive. Note, this conceptualisation of subevents is different from but connected to that which is articulated in Bohnemeyer et al. (2007) and Bohnemeyer et al. (2011) through their discussion of the Macro-Event Property (MEP) and the observation that events are mereological (part-whole) structures, in which parts of and combinations of events are themselves conceptualised as events. A mereological property of events is examined more closely in chapter 3 in the discussion of the collective and distributive interpretations of ambiguous events.

[5] Changes in state here is used in a broader sense in the psycholinguistic literature than change-of-state conceptions in some verb-semantic scholarship, whereby a verb like *clean* inherently produces a change of state in its syntactic object while a verb like *wipe*, indicating contact via movement, does not (Levin & Rappaport Hovav, 1992). A change of state in the event representation literature can refer to any shift in a discourse entity's 'static' properties, such that both *clean* and *wipe* induce a change in the history of the agent of the action (itself a representational object), thus reflecting the dynamic nature of events. It is my view that these two conceptions of changes in state have not been properly teased apart, especially in event representations resulting from manipulations of verb semantics. It is also worth noting that this conception of events as changes in state is not a settled matter. Cleland (1991), for example, proposes that events are individuated by changes but argues for an analysis in which it is not changes in the state of objects but changes in their determinable properties (such as their location, colour, and temperature) that result in the individuation of events. To my knowledge, these claims have not been tested.

event-semantic formalism shares parallels with experimental observations of how comprehenders process event information in real-time. Furthermore, there appears to be a significant overlap in the conception of non-linguistic events and those encoded in the grammar of a language, in that event segmentation and prediction updating mechanisms lie at the heart of both visual and linguistic processing. Both formal and cognitive enterprises converge on the idea that events are a basic unit of discourse representations. Theories of situation model construction therefore ought to incorporate events in their ontology.

With these considerations about events in mind, I will revisit the central topic of this section – how mental models are constructed – and compare the Construction-Integration (CI) model with an alternative which places events, rather than propositions, squarely at the centre of analysis: the Event-Indexing (EI) model.

1.2.2.3 – The Event-Indexing Model

Earlier in this section, I introduced the CI model and noted that it considered propositions rather than events as the fundamental unit of situation model construction. While the CI model has offers rich explanations of various data in the field (see Britton & Gülgöz, 1991; McNamara et al., 1996; O'Reilly & McNamara, 2007), a theoretical model which incorporates events captures several essential properties of discourse representations that the CI model does not. One such model that has gained traction in the comprehension literature over the years is the Event-Indexing (EI) model.

McNamara and Magliano (2009), in a review of comprehension research at the time, remark that a core assumption of the EI model is that the human cognitive-linguistic system is attuned to perceive events (i.e., changes in the states of objects). Unlike the CI model which focussed on propositional networks and argument overlap, the EI model emphasises the role of verb semantics in conveying information about the underlying event information, in parity with findings from work in semantic theory and event segmentation research, and proposes that situational coherence is developed across multiple dimensions of continuity which are indexed for temporal, causal, spatial and motivational properties of objects and their histories relative to existing events introduced by predicates in the preceding input (Zwaan et al., 1995a). The claims made by proponents of the EI model are, therefore, consistent with insights in the event-semantics literature, in which predicates range over event arguments, and findings that demonstrate the importance of events in cognitive processing.

The notions of construction and integration remain central to the EI model, which claims that situation model construction proceeds along three phases: a *current model* that is constructed now, at Time t_n, where a comprehender interprets the events of a particular clause or sentence (Clause c_n); an *integrated model* that is the existing global representation of all connected models through Times t_1 to t_{n-1}, produced by a comprehender while reading Clauses c_1 through c_{n-1}; and a *complete model*, which is the sum of both *current* and *integrated models*, that is stored in memory after all textual input has been processed (Zwaan et al., 1995a; Zwaan & Radvansky, 1998). The process of incorporating the *current model* with the *integrated model* is that of updating[6].

As discussed in section 1.2.1, situation models are multi-dimensional, consisting of casual, intentional, spatial, temporal, and object information (also see Johnson-Laird, 1983; Zwaan et al., 1998), and readers are sensitive to manipulations of these dimensions. Comprehenders

[6] The complete model is updated when new information is received in the input, including after extended passages of time, such as when the comprehender reflects on their interpretation of the events described and perhaps infers new connections between information in the models (Zwaan & Radvansky, 1998). A particularly rich account of situation model updating was proposed and developed by Kuperberg (2016), Kuperberg and Jaeger (2016), and Xiang and Kuperberg (2015), in which mental representations exist across multiple, increasingly complex connectionist layers, including an event-sequencing and event-structural layer, where information is propagated from one layer to another via probabilistic (Bayesian) hypotheses about upcoming input. The finer details of this account are beyond the scope of this thesis (and to this author's knowledge the model has yet to gain significant ground in the wider literature) but given its emphasis on 'event layers' and the evident centrality of events to mental models, it may present an interesting avenue for research into event processing in the future.

construct their *current model* along these five dimensions – or indices (Zwaan & Radvansky, 1998).

17) Peter took the elevator to the fifth floor. He went to talk to his professor about his grade.

For a text like (17), adapted from Zwaan and Radvansky (1998), upon reading the second sentence the comprehender will attempt to integrate an unfolding representation with an extant one. The reader will recognise the pronoun as a cue to search previous input or context for a suitable referent, and lacking context will identify Peter as a suitable candidate sharing presumed features such as [male]; they will also likely establish a representational token for another object, the professor, inferring that “he”, Peter, is likely her student. The goal of the event described in the second sentence will also be attributed to Peter – that is, Peter’s goal in riding the elevator is updated to be understood as a means of going to speak with his professor. Peter’s series of actions, including the elevator ride, are inferred to be a result of the grade he received (possibly a bad one), hence him wanting to talk with his teacher about it. Without explicit references to changes in spatial location, the reader will infer that no change has taken place, and that the location of the professor is on (or assumed to be on) the fifth floor; note, that while interpreting the first sentence, readers will establish a representational token of the elevator object, recognising that it and – by extension – Peter have transitioned from one spatial state (not the fifth floor) to another (the fifth floor), developing the histories of the objects described. As with space, with no explicit reference to shifts in time the reader will expect the ‘going’ event described in the second sentence to be at a time immediately after Peter’s implicit alighting of the elevator, although this does not have to be the case, as (18) illustrates.

18) Peter took the elevator to the fifth floor. An hour later, he went to talk to his professor about his grade.

While (18) is interpretable, there is an unnaturalness about it; the temporal index of the second sentence does not match the default expectation that events unfold from one event to the next in a chronological and contiguous fashion – the iconicity assumption introduced in an earlier section. As mentioned previously, violations of this assumption about the temporal order of events leads to increased reading times and decrements in comprehension accuracy (Mandler, 1986; Ohtsuka & Brewer, 1992; Zwaan et al., 1995b; Zwaan, 1996). Behavioural data like these provide supporting evidence for an account of mental modelling in which events are the fundamental unit of discourse representations.

1.2.2.4 – Comparing Models of Comprehension

Comprehenders, according to the EI model, track temporal information about the unfolding events, and are surprised when their expectations of temporal ordering and contiguity is violated; the CI model, with a focus on propositions and argument overlap, is unable to equally explain observations like this. Additional evidence for the EI model can be seen in Rinck and Weber (2003), who found that changes in the time, location and objects involved in a narrative situation resulted in increased reading times; similar observations were made by Ehrlich and Johnson-Laird (1982) who provided participants with continuous narrative descriptions like (19) and discontinuous narrative descriptions like (20), finding that participants were faster at reading the former than the latter.

19) The knife is in front of the pot. The pot is behind the dish. The dish is left of the glass.
20) The knife is in front of the pot. The glass is behind the dish. The pot is left of the glass.

It is worth noting that the findings from Ehrlich and Johnson-Laird are consistent with both the EI and CI models, as argument overlap between propositions in the experimental stimuli is equally capable of explaining this data, but importantly, research across a range of experimental studies and various analyses of sentence and verb meaning, have emphasised the essentiality of events in the representations comprehenders form while processing (narrative) language, suggesting that the EI model is perhaps a better explanations of the data than the CI model[7].

One advantage of the CI model, however, is that it relies on a relatively small number of assumptions about what contributes towards comprehension processes, with an emphasis on bottom-up mechanisms, rules of syntax, and argument overlap. The CI model operates on the assumption that propositions are constructed and form – through macrostructural propositional networks – the basic unit of situation models. While a significant consequence of its simplicity is that it can fail to explicitly capture the centrality of event information during processing and the mechanisms that guide comprehenders toward situation modelling (McNamara & Magliano, 2009), the CI model does offer a tighter account of bottom-up processes in translating surface-level forms of text into qualitatively rich mental representations of the situation described. Conversely, while the EI model is more explicit about the role of events and the kind of information comprehenders attend to at higher levels of representation, it is not particularly clear how readers and interlocutors extract the underlying events encoded by the meaning of verbs and their arguments.

The role of knowledge is also important in both models, but they differ with respect to how they treat knowledge. Knowledge is implicit in the EI model in the top-down inferential processes that update the *current model* with the *integrated model*, such as in the world knowledge inherent in the implication that the professor in example (17) was inside a building, probably a university, ensuring the construction of a coherent mental model of the total situation described in which Peter took an elevator up inside the building that the professor probably worked in. Words like "professor" establish an entire *frame* in which a potentially

[7] Due to its increased attention on event representations and their compilation into situation models, the EI model is only a model of narrative text comprehension – as a general model of language comprehension, and as an explanation of expository or argumentative texts, the EI model does not fare as well (McNamara & Magliano, 2009). Arguably, the CI model, with its focus on situation modelling, shares a similar problem. It is in these areas of discourse processing that alternative models of comprehension might be preferred.

infinite number of concepts are generated, based on previous experience and knowledge of the word, such as *working in a university*, *conducting research*, *marking assignments*, etc. (Aitchison, 2012; Fillmore, 2006). The CI model is less detailed about what it is that comprehenders know or attend to but is more explicit about how that knowledge strengthens or inhibits various interpretations by treating that knowledge as part of a connectionist process of spreading activation from conceptual node to conceptual node. The role of knowledge, both encyclopaedic and embodied, in situation model construction is explored in more detail in the next section.

Overall, both models are valuable tools in explaining how situation models are constructed, but approach the problem from different vantage points, engendering certain advantages and disadvantages. As alluded to, a scientifically complete model with respect to mental representations is likely to be a blend of (at least) both these accounts. As an exploration of such a putative 'super-model' lies outside the scope of the present thesis, I draw upon insights of both the CI and EI models, where applicable, to account for the data in my experiments and situate the results in the larger theoretical conversation.

1.3 – Other Elements of Representations

Knowledge of the world, including associations with how we perceive the world and our experiential interactions with it, all appear to be important mechanisms of mental model construction. Broadly speaking, a formula for the necessary components of situation models would be something like [event representations + encyclopaedic knowledge + embodied knowledge].[8] While this thesis focusses on the former, with particular consideration given to

[8] Event representations consist of the changes in the temporal, causal, motivational, and spatial states of objects and their interactions with one other, derived from some (linguistic or non-linguistic) input. Context, both linguistic and non-linguistic, could arguably be treated as discrete elements within this formula; however, the role of context will be subsumed under the label of encyclopaedic/world knowledge for the purposes of this thesis.

the number of events that are mentally represented, both encyclopaedic knowledge and embodied knowledge contribute important elements in any complete theory of narrative comprehension processes. To maintain simplicity while examining the effects of negation and plurality denoting expressions on situation model construction that comprise the main thrust of this thesis, the knowledge-based factors of this formula have not been manipulated or measured in the experiments that form the central three chapters of this project. However, any claims and findings of this thesis should be considered in conjunction with insights from research in these areas of knowledge contribution. The present section reviews research pertinent to the role of both encyclopaedic and embodied knowledge. The ideas presented here will be revisited in chapter 5, where some suggestions are made for future research that integrates the key conclusions of the central three chapters with those about the role of knowledge introduced here.

1.3.1 – The Role of Encyclopaedic Knowledge

To begin, consider the situation described in (21), adapted from Bransford and Johnson (1972).

21) The procedure is actually quite simple. First you arrange things into different groups depending on their makeup. Of course, one pile may be sufficient, depending on how much there is to do. If you have to go somewhere else due to lack of facilities, that is the next step; otherwise you are pretty well set. It is better to do too few things at once than too many. Remember, mistakes can be expensive. At first, the whole procedure will seem quite complicated. Soon, however, it will become just another fact of life.

Without context, the details of this passage can be difficult to process, interpret and recall, as observed across several experiments by Bransford and Johnson. However, when given knowledge to aid comprehension, such as an appropriate context picture – something like ‘a washing machine’ for the situation in (21) – readers recall more information from the text than those who receive insufficient knowledge. These findings were replicated by Long and Spooner (2010), who, surprisingly, also observed that supporting knowledge of the text enhanced participants' memory for even the context in which they read the text, such as the colour of a border around said text, or even the colour of the experimenter's shirt.

Comprehenders' existing knowledge of the world is also important and aids in comprehension. Schneider and Körkel (1989) demonstrated how domain expertise in a topic was more important than the verbal skills participants demonstrated in study of recall ability. Assuming that degree of recall is a function of what is abstracted away and encoded by the reader during sentence and discourse processing, their findings illustrate the impact of knowledge on mental modelling.

More recent research from Nieuwland and van Berkum (2006) found that, given no context, a sentence like (23) would result in larger N400 effect than (22).

22) The peanut was salted.

23) The peanut was in love.

The N400 is a negative-going wave observed in electroencephalographic (EEG) research, typically appearing 300-500ms post-stimulus, with a graded modulation based on the degree of congruency the critical stimulus has with its preceding context (Payne et al., 2015) – the lower the congruency the larger the N400. A larger N400 was observed to (23) as it clearly violates animacy rules germane to real-world knowledge: peanuts do not enjoy emotions like love, but they are, however, frequently salted. That said, when presented with a preceding context in which peanuts are established to have animacy, and are fully anthropomorphised, EEG findings to (23) show that the N400 effect disappears, and that processing is easier for comprehenders than the sentence in (22).

In summary, whether pre-existing or provided, knowledge interacts with how comprehenders process and ultimately model representations of the situation described by some text.

1.3.2 – The Role of Embodied Knowledge

In addition to encyclopaedic knowledge, embodied knowledge appears to be important too. One of the most significant contributions to discovering the underlying mechanisms behind situation model construction has been the observation that abstract representations modelled by comprehenders during language processing are unlikely to be entirely arbitrary and amodal (Glenberg & Kaschak, 2003; Harnad, 1990). Barsalou (1999) recognised that a purely conceptual system of cognition does not capture the richness of experience in mental modelling. Perceptual experience matters.

Barsalou (1999) makes the case that sensory-motor areas of the brain capture bottom-up patterns of activation during perceptual experience, such as touching something sharp, and then represents these experiences as perceptual symbols in memory, wherein the internal structure of these symbols is modal and analogically related to the perceptual states that produced them. These perceptual symbols are then reactivated, in a top-down fashion, during a relevant stimulus, such as encountering the word *sharp*, which subsequently elicits one's experiences of sharpness. Barsalou argues that qualities of objects in the world, such as BLUE, SHARP or HOT, are extracted from the perceptual experiences of the individual and stored in memory in a symbolic representation of the phenomena previously observed. The kinds of representations formed in this way are not limited to just sensory experience, but matters of proprioception (jumping, singing, dancing) and introspection (happiness, hunger, love), although the latter are much less understood (Evans, 2009). Thus, unique representations and their associations, based on prior experiences, are reactivated for each person who encounters words like *love* or *dance,* or even phrases like *a purring cat*. These representations are likely to be somewhat phenomenologically different for everyone due to unique individual experiences.

Neurolinguistic evidence for the kinds of perceptual representations described above are observed in the literature. Words involving actions related to different body parts, such as *kick* and *punch*, appear to activate motor and pre-motor cortices (de Lafuente & Romo, 2004). Via excitatory transcranial magnetic stimulation (TMS), with magnetic pulses below motor threshold, Pulvermüller et al. (2005) stimulated hand and leg motor areas of the left hemisphere of right-handed speakers of English while they read common arm- and leg-related words, finding that processing of these words during a lexical decision task are affected differentially based on the motor area being stimulated: arm words are responded to faster than leg words when arm area TMS takes place, and vice versa. In other research, words related to physical objects that were inherently graspable, and had a specific function when grasped, elicited greater activation in fronto-parietal sensorimotor regions of the brain compared to objects which were still graspable but lacked any specific or typical function while held (Rueschemeyer et al., 2010). Kurby & Zacks (2013) demonstrate, via two functional magnetic resonance imaging (fMRI) studies, that reading high-imagery clauses activates brain areas associated with motor and perceptual representations, with this effect being stronger when such clauses can be easily integrated into a coherent situation model compared to a more discontinuous passage of discourse.

From a more cognitive perspective, Glenberg and Kaschak (2002) showed how task response times are influenced by perceptual representations. As part of a sensibility-judgment task, participants were instructed to indicate that some stimulus sentence was 'sensible' by moving part of their body in one direction and to indicate some stimulus was 'not sensible' as movement in the opposite direction, such as moving an arm away from the body or toward the body, respectively. The sentences consisted of situations like "close the drawer" and "open the drawer". Interestingly, participants had more difficulty in making sensibility judgements where the movement implied by the stimulus sentence, such as moving away from the body for closing a drawer and toward the body for opening a drawer, was opposite to the direction of movement required to indicate the correct answer. Responses were facilitated when the direction of movement was the same. As a consequence of findings like this, most psycholinguists working in discourse comprehension research now claim that people appear to *simulate* themselves within the situations described by texts, drawing upon embodied knowledge in the construction of these mental simulations (for example, see Barsalou, 1999; Bergen & Wheeler, 2010; Glenberg, 1997; Kaup et al., 2007; Mak & Willems, 2019; Speed & Majid, 2020; Zwaan & Pecher, 2012).

It is not only ourselves that we simulate, however – other objects in the situation are simulated too. For example, Stanfield and Zwaan (2001) presented participants with a sentence like (24); shortly after, participants were presented with a picture whose orientation matched or mismatched the situation implied by the text.

24) The pencil is in the cup.

In one version, the picture was vertical (as a pencil would typically be if inside a cup, such as for holding), while the other was orientated horizontally (which is much less likely to reflect a typical situation of a pencil being inside a cup, unless that cup were knocked over, for example). Participants were faster at responding to the match condition, which was interpreted by the researchers as evidence that during textual comprehension readers generate a modal representation which aligns with prior knowledge about what the text is describing (in this case a vertically orientated pencil), and they reactivate and use this representation when responding to the task. In essence, they simulate their typical experience of a pencil in a cup.

25) There was no eagle in the sky.

Similar findings are seen in Kaup et al. (2007), who presented participants with sentences like (25), and found that participants, in a subsequent sentence-picture-matching task, responded faster to an image of an eagle with outstretched wings (consistent with the form eagles typically take when in the sky) than an image of an eagle with folded wings and in a resting state (inconsistent with the form eagles typically take when in the sky), even though the sentence contained the negation marker ‘no’. Although the prototypicality of these images was not controlled for, findings like this have again been interpreted as evidence that comprehenders construct an experiential representation of the situation described – birds fly with outstretched wings, not closed ones – and it is this that is re-activated when the picture-matching task becomes relevant.

Inspired by examples like the above studies on perceptual and modal representations, the recent trajectory of research on situation modelling has moved away somewhat from its raw linguistic influences, which appeared to be in focus during the 1970s through to the early 2000s, to the matter of mental simulations and embodied cognition (see Horchak & Garrido, 2020; Jachmann et al., 2019; Kang et al., 2020; Kok & Cienki, 2017; Koning et al., 2017; Lin & Matsumi, 2022; Seger et al., 2020; Venhuizen et al., 2016; Xu & Liu, 2022; Zhang et al., 2021; *inter alia*). Indeed, as Miller and Johnson-Laird (1976) almost prophetically muse, “any sharp division between perception and conception seems questionable”.

Embodied knowledge (or, more precisely, simulations) are, therefore, a deeply important consideration when theorising on the abstracted representations comprehenders create and store in memory. The meanings of words, and related conceptual properties of events, such as time, causality and location, among others, clearly cannot give a full accounting of how we construct mental representations of the sentences we encounter: world knowledge, perception, personal experience, and even, it seems, emotions matter (see van Berkum (2019) for a recent discussion on the interplay between language comprehension and emotion, and Gernsbacher et al. (1992) for evidence that comprehenders mentally represent the emotional states of others).

1.3.3 – A Situation Without Knowledge

While encyclopaedic and embodied knowledge are demonstrably essential for a full accounting of situation model construction, it is a primary contention of the present thesis that the linguistic contributions of these mental models, particularly as they relate to event representations, are still woefully understudied. Levin (2019) comments that many accounts of mental modelling have thus far minimised or ignored the role grammatical information may play in the construction of situation models. Over a decade ago, Bergen and Wheeler (2010) similarly noted that we know very little about what components of an utterance drive the representations comprehenders produce in the pursuit of a coherent mental model of some text. I would agree that this is still true today.

Naturally, we cannot and should not neglect the breadth of testable claims about the linguistic influences on modelling situations and the event representations that comprise them. We still possess very little empirical evidence about how grammatical features like tense and telicity, conjunction and negation, plurality and serial verb constructions scaffold or otherwise contribute to discourse representations. This list is not exhaustive; and while telicity is perhaps the best studied of these (see Becker et al., 2013; Malaia et al., 2012; Malaia, 2014; Pinango et al., 1999; Wagner & Carey, 2003; *inter alia*), and grammatical constructions not found in European languages (particularly English) are the least studied (Ünal et al., 2021), the current thesis attempts to address some of that understudied research-ground by looking at negation and plurality denoting expressions and their impact on how comprehenders determine the number of events, if any, that take place within a described situation.

This question regarding the number of events represented is of particular significance to our theories of comprehension, as the existing literature, whether discussing the representation of events or their impact on mental simulations, has overwhelmingly focussed on single events (or states), or sequences of single events in some narrative; however, both negation (in constructions like *didn't buy a book*) and plurality denoting expressions (such as *two students* in *two students bought a book*) are widely used by speakers (of English) and, in the context of eventualities like accomplishments, frequently give rise to event meanings where either an event did not occur or ambiguously many events occurred. Understanding better the role these linguistic structures have on event representational processes remains an important and outstanding task in the science of language comprehension.

To maintain simplicity in the following investigation, I will, to the best of my ability, intentionally put to one side the ways in which negation and plurality denoting expressions might interact with existing pragmatic, contextual or intertextual knowledge (although it is impossible to exclude entirely). Moreover, I will intentionally ignore the role of – and the non-linguistic complications introduced by – embodied cognition in situation model construction, which draw upon concepts like experience and sensorimotor perception. Instead, I will attempt to focus solely on how negation and plural denoting expressions, as linguistic constructions, contribute to the developing event representations elicited by short narratives in an experimental context. This project will therefore act as something of a test to see if focussing on the role of understudied grammatical structures to the exclusion of the role of important factors like embodiment and simulations (and, where possible, confounding inferences like world knowledge) can still be productive in advancing the science. I will revisit the role of knowledge explored in this section in chapter 5, where I discuss potential further research that might be pursued which integrates findings from the experiments conducted in the central three chapters with work on world knowledge, embodiment, and the notion of simulations.

The rest of this thesis is, therefore, devoted to addressing how verbal negation and plurality denoting expressions in short English narratives affect event representations, with knowledge and simulations temporarily put aside. The central theme of this research, termed *event numeration*, is thus how many events – one, more than one, or none – form part of the formula [event representations + encyclopaedic knowledge + embodied knowledge] when event-introducing accomplishments (that contain the properties of telicity, durativity and dynamicity associated with events) are modified by negation or a plurality denoting expression. An important measure of the effect these linguistic constructions have on mental representations will be to test their impact on the well-attested expectation comprehenders hold that events should unfold in singularly ordered contiguous fashion, known as the iconicity assumption. Addressing these matters will offer some response to the relative dearth of research on the grammatical contributions to event representations, while revealing more about how events – particularly when non-singular – contribute to the development of situation model construction, improving our understanding of language comprehension processes. In the following section, I provide a brief outline of the remaining chapters of this work.

1.4 – The Structure of This Thesis

This work is produced as a thesis by paper. Chapter 1 has set the scene for the various theoretical and methodological concepts relevant to this project. The central three chapters (2, 3 and 4) are written in the style of academic articles, with individual literature reviews, methods, results and discussion sections. Chapter 5 bookends this thesis by summarising and reflecting on the findings and conclusions of the preceding chapters.

More specifically, chapter 2 examines how the inclusion of negated verb phrases within critical sentences of narrative discourse affects event representation by measuring the effect of negation on accessibility to conceptual information inside the negated verb phrase via a subsequent probe recognition task. Critical sentences consist of two clauses – one full and one elided – which are manipulated across four conditions: explicit clause negated, elided clause negated, both negated, and neither negated. Findings are compared to existing research on the negation of nouns and adjectives on information accessibility and the function of negation during language processing (see Giora et al., 2004; Kaup, 2001; Kaup et al., 2006; Kaup et al., 2007; Lüdtke et al., 2008; MacDonald & Just, 1989; *inter alia*), and what this means for how events contribute to the situation models comprehenders produce when in scope of negation. See *Appendix A* for the materials used in this study.

Chapter 3 examines how the inclusion of numeral modifiers and conjoined noun phrases, collectively called plurality denoting expressions (Champollion, 2019), within critical sentences of narrative discourse affect event representations by capturing reading times to consistent and inconsistent continuations with either collective (single event) or distributive (multiple event) interpretations of the situation described. Critical sentences consist of four conditions: non-plurality denoting, plurality denoting via numeral modifiers, plurality denoting via conjunction, and plurality denoting via both constructions. Narratives that introduce accomplishment events in the past perfect and in the present perfect are compared to determine whether the situationally relevant timespan that differs between these two aspectual tenses biases comprehenders to a distributive interpretation over a collective one. Findings are compared to existing literature on distributivity processing (e.g., Frazier et al., 1999; Maldonado et al., 2017). See *Appendix B* for the materials used in this study.

Chapter 4 examines, via event-related potential (ERP) research, the effect of both verbal negation and plurality denoting expressions on the well-attested observation in event representation research that comprehenders expect events to be ordered singularly and contiguously (Event X → Event Y → Event Z) – the iconicity assumption (Dowty, 1986; Zwaan, 1996). If comprehenders' conception of events is fundamental to this assumption, then it is predicted that negated event information in the middle of three ordered events comprising a short narrative should lead to event discontiguity and be treated by the comprehender as a violation of their underlying assumption about event ordering being contiguous (Event X → Some Time Passes BUT No Event Occurs → Event Z). This violation is predicted to be observable in the ERP waveform in negated event conditions compared to a non-negated control time-locked to the introduction of the third event (see Kamide, 2008; Kuperberg, 2016; O'Gara & Lawyer, 2019). Similarly, if multiple events are mentally represented following a plurality denoting expression in the critical second sentence of the narrative (Event X → Event Y AND Event Y' → Event Z), differential ERP patterns are expected when compared to an unambiguously non-distributive control. Findings from these two experiments are reviewed in light of existing research on the iconicity assumption and the role of event information in the construction of a coherent mental model of the situation described by some text. See *Appendices C and D* for the materials used in this study.

While these core chapters deal with the on-line processing of negation and plurality denotating expressions, and their effect on *a priori* expectations of event ordering, Chapter 5synthesises these findings into a larger discussion of event numeration – how comprehenders determine the number of events (0, 1 or >1) during processing – drawing connections to theoretical accounts like the CI and EI models, and recommending future work that builds a bridge to research on knowledge and simulations in this area of enumerating events. Chapter 5 also reviews the limitations of the present thesis, reflecting on certain methodological challenges regarding data collection during the covid-19 global pandemic. Additional areas of research are also proposed that shine a spotlight on other understudied linguistic constructions (including in languages other than English) on the development of abstracted representations of events and their composition into coherent situation models in memory.

Chapter 2

The Function of Negation in Event Representation

2.1 – Background

Negation, in on-line processing experiments, has proven to be a useful empirical tool in uncovering the mechanisms behind the construction of coherent mental representations of the situations described in texts. This chapter examines the effects of negation on event-introducing verbs, which remains an understudied area of mental representation research. Inspired by a psycholinguistic, probe-word experiment by MacDonald and Just (1989), the present study measures response times to probes following short narratives containing negated elided clauses, negated syntactically explicit clauses, or both, and compares these to probe responses times in non-negated conditions. The results here suggest that the negation of event-introducing verbs has different processing consequences from the negation of objects within the situation described. Consequently, this work has theoretical implications for how comprehenders access conceptual information and build representations following negated contexts.

2.1.1 – The Negation of Nouns

In MacDonald and Just (1989), participants were asked whether a probe word (a noun) had appeared in a previously presented sentence. Where that probe word corresponded with a negated noun, participants were slower to indicate that it had occurred in the preceding passage than for a non-negated noun. For example, in (26), participants were slower to identify that the word *bread* had appeared than they were when *cookies* was presented as the probe.

26) Mary baked some cookies but no bread.

27) Mary baked no cookies but some bread.

It was assumed this observation reflects the function of negation in sentence and discourse processing. The negative particle (*no*, in this case) is assumed to lead to reduced accessibility of the negated concept (*bread*), reflected in response times to the recall task (Giora, 2007; Kaup & Dudschig, 2020; Kaup & Zwaan, 2003; Kaup et al., 2006). Importantly, this effect was also observed for *cookies* when negated instead of *bread*, in examples like (27), showing there is nothing intrinsic to the concept of *bread* which is more difficult for comprehenders to access; similarly, the order of the nouns in the critical sentence had no bearing on the differential response times, indicating that negation is the contributing factor in this task (MacDonald & Just, 1989). Negation, therefore, might be operating as a cue or instruction to suppress the negated concept, making it less accessible than an equivalent non-negated concept (Gernsbacher & Jescheniak, 1995; Giora et al., 2004; Giora et al., 2007; Lea & Mulligan, 2002; Lüdtke et al., 2008).

The suppression of information hypothesis in this formulation has problems, though, as many of the above authors note. Consider (28), in which the subject *animals* is negated with the negative marker *no*.

28) No animals were harmed in the making of this programme.

It is not the case that the concept of *animals* is suppressed here; in fact, (28) presupposes that animals were involved, and the health and welfare of any and all animals which were present in the programme is understood to be the salient matter. Similarly, consider example (29) from Kaup (2001), in which the negated concept, *photographs*, is again not suppressed.

29) Elizabeth burned the letters but not the photographs.

The situation described here is one that includes both *letters* and *photographs,* such that both are salient entities, but it happens that while the letters were burned the photographs were not, a qualitatively different circumstance than that captured in (26), where *bread* was actually not part of the situation described. Example (26) does not describe a situation where *bread* is salient; *bread* is simply not part of that situation. The verb *bake* in (26) is, frequently, a verb of creation (see Piñón, 2008), denoting the coming into being of the referent of its internal argument (e.g. *bread*). Under negation, acts of creation are denied (Partee et al., 2011).

Consequently, this denial of creation results in a situation in which *bread* is not present. Kaup (2001) shows negation does not affect destruction verbs in the same way, explaining (29) above. The suppression of negated information, it seems, is not obligatory, and may depend on the linguistic structures it occurs within (Giora et al., 2004). In a position which is still the consensus today, Kaup (2001) postulates that the accessibility or inaccessibility of particular concepts, whether in scope of negation or not, is a function of their representation in the situation model – the mentalised abstraction of the holistic amalgam of the events, concepts and contexts encoded within and inferred from some text by the comprehender (Johnson-Laird, 1983; Kintsch, 1988; Radvansky & Zacks, 2011; Xiang and Kuperberg, 2015; Zwaan, 2016). The degree of accessibility of these concepts, however, may be modified by the inclusion of negation, as the examples from MacDonald and Just (1989), above, show.

2.1.2 – The Negation of Verbs

Much of the work on negation has been conducted by looking at the phenomenon across nouns (as discussed in the previous section) and adjectives (see Farshchi et al., 2021; Kaup et al., 2006; Lüdtke et al., 2008; Orenes, 2021); only recently in studies of action embodiment and inhibition (and their neural correlates) has the negation of verbs, where negation has scope over the event itself, been the predominant focus (see Aravena et al., 2012; Beltrán et al., 2019; de Vega et al., 2016; Montalti et al., 2023; Tettamani et al., 2008). Little has been written about the cognitive effects of negating verbs on event representations and situation model construction. Events are widely considered to be the basic units of mental models and the representations comprehenders construct about the situations described in texts (Altmann & Ekves, 2019; Bohnemeyer & Pederson, 2010; Elman & McRae, 2019; Knott & Takac, 2020; Kuperberg, 2016; Kuperberg and Jaeger, 2016; McNamara & Magliano, 2009; Metusalem et al., 2012; Pettijohn & Radvansky, 2016a; Pettijohn et al., 2016; Radvanksy & Zacks, 2014; Radvansky & Zacks, 2017; Zacks, 2020; Zwaan, 2016; Zwaan & Radvansky, 1998; *inter alia*).

Therefore, understanding how negation interacts with event phenomena should afford us deeper insights into the creation of situation models, in much the same way as the processing of negation in noun and adjective phrases[9] has revealed important cognitive and representational mechanisms in play.

The negation of verbs, sometimes referred to as sentential negation, demonstrates some interesting scopal properties in relation to the underlying events introduced by those verbs. Champollion (2011), using event semantic ontology, shows how the sentential negation of (30) can be captured by the logical form of (31), but not by (32), which is trivially true for all events in which John did not laugh, and is not a natural interpretation of the sentence in (30)[10].

30) John didn't laugh.

31) $\neg\exists e$ [LAUGH(e) & AGENT(e, j)]

'There is no event in which John laughed'

32) * $\exists e$ $\neg$[LAUGH(e) & AGENT(e, j)]

'There is an event in which John does not laugh'

In (31), the laughing event did not happen – there is no situation of which it is a part; in fact, the proposition it expresses is only true in a world in which a laughing event by John is not so. There may be pragmatic inferences or implicatures that could be induced based on context, such as John not laughing at a joke because he was furrowing his brow in confusion, and this

---

[9] Research on the negation of adjectives has uncovered a two-step processing route during the interpretation of binary negated contexts (Kaup et al, 2006; Orenes et al; 2014), whereby two states of affairs are initially represented for a sentence like *the door is not open*: the negated state of affairs (*the door is open*) and the actual state of affairs described (*the door is closed*), with the comprehender shifting attention after some time (around 1300ms) to the actual state of affairs, as evidenced in naming task response times and visual world eye-tracking tasks. These findings demonstrate the important role of negation in mental representation research.

[10] This observation has interesting parallels with when negation is treated as a quantifier over entities rather than events, suggesting that similar underlying principles may be in effect; thus, insights from nominal negation may be applicable to sentential negation. The sentence *No boy laughed* has the semantic form $\neg\exists x$[BOY(x) & $\exists e$[LAUGH(e) & AGENT(e, x)]], and is read as 'there is no laughing event that is done by a boy', but not *$\exists e$[$\neg\exists x$[BOY(x) & LAUGH(e) & AGENT(e, x)]], paraphrased as 'there is an event that is not a laughing event by a boy'. The latter is a trivially true proposition and one that does not reflect the natural reading of this sentence (Champollion, 2011).

inference might be mentally represented as an event that comprises the larger narrative situation, but such implicatures rely on social or discourse contexts which may not be available to a comprehender, especially if presented in isolation, and so this does not ultimately undermine Champollion's analysis. Nonetheless, additional complications arise with sentential negation when considering examples like the following.

33) Paul did not stop at the red light.

34) John did not arrive; it surprised Mary.

Weiser (2008) discusses a number of scenarios in which the $\neg\exists e$ analysis of (31) might not apply, including examples like (33), where through our world knowledge of road networks and traffic systems we actually imagine Paul engaged in the event of driving dangerously through the intersection under a red light – an event thus generated by a negated action sentence – and (34), in which the pronoun refers back to the 'negative event' of *John not arriving* in the preceding sentence, although Weiser prefers to treat this antecedent as a 'state' or 'fact' rather than a 'negative event'. Indeed, Bernard and Champollion (2018) note that it is not even conceptually or formally clear what it is for an event to be 'negative'. I set aside these complications for the purposes of this study, focussing instead on unambiguously non-events, with limited pragmatic or contextual implications, expressible in the form $\neg\exists e$. For a more detailed discussion of the pragmatic influences on negation processing see Kaup and Dudschig (2020), Nieuwland and Kuperberg (2008), and Tian & Breheny (2019).

If there exists a psychological reality to the semantics of examples like (31), the negation of action verbs should lead to the suppression of that event forming part of the larger situation described in discourse contexts.[11] For example, if *John read a book and Mary didn't bake a*

[11] The motivation for my claim here stems from formal analyses, like (31), wherein the existence of the event is formally denied, and thus any predicates which that event is an argument of are not represented as part of the larger situation described. In situation model terms, this should lead to a lack of representation in memory of the action and arguments of the event-denoting verb that is negated.

cake, the only event that comprises the total situation described is that of the reading event; semantically, and presumably psychologically, there is no event in which Mary baked a cake that could contribute to the developing situation model. While the negated event is not assumed to contribute to an unfolding mental model, a surface representation of the clause should still be processed and potentially retained (for a short time) by comprehenders. This differs from negated noun phrase contexts, such as those in MacDonald and Just (1989). While expressions like 'baked no bread' lead to the object being suppressed (and lacking existential commitment in the formal logic), the baking event itself still happens, which the sentence in (26), *Mary baked some cookies but no bread*, exemplifies. Slower reaction times to a subsequent probe word task were observed in these negated contexts as comprehenders had to 'search' through a message-level representation that did not contain the probed object inside of it. Negated events, on the other hand, such as 'didn't bake bread' lead to the baking event itself, including all the predicates it is part of, lacking existential commitment and therefore both the action and any arguments are expected to not be represented in memory at the message level. In other words, there would be nothing at this level of representation to 'search' through during a subsequent probe word task. However, a surface-level representation of the text should still be accessible for a short time.

Surface representations are sometimes explicit in models of comprehension (e.g. Kintsch, 1988), although they are often left unstated, as they are assumed to merely act as cues to generate abstracted mental representations of the events and entities described within the discourse (McNamara & Magliano, 2009). Message-level representations seem to influence lexical access of target words in priming studies (Morris, 1994; Traxler et al, 2000), but surface linguistic representations that consist of orthographic and phonetic information also prime repetitions of previously encountered words (Cloitrew & Bever, 1988; Ledoux et al., 2006; Tanenhaus et al., 1985); indeed, merely mentioning a target in a pre-task context facilitates processing, even after a delay (Lüdtke et al., 2008). The strength of surface representations can vary, with highly-stylised literary texts or highly-attentive reading approaches resulting in better recall and recognition of surface-level features in a subsequent task (Zwaan, 1991). Memory for surface representations is typically weaker, and decays faster, than semantic or propositional 'gists' of sentences (Anderson, 1974; Begg & Wickelgren, 1974; Kintsch et al., 1990; Zimny, 1987), but, unlike event representations, surface representations consisting of the exact orthographic and phonetic traces of words in their grammatical structure should survive negation. For that reason, surface representations are likely to play a more significant role in tasks involving retrieval of information from clauses that negate entire events than in clauses that do not, as comprehenders will not have access to the more salient situational representations (or 'gists') to draw upon.

Experimental evidence of the effect of verbal negation on situation model construction is notably lacking in the literature and, as a result, is an empirical gap in need of addressing. The present study therefore examines whether sentential negation suppresses an underlying event from forming part of a larger discourse representation, leading to only surface-level representations being accessible, as measured through behavioural responses to an on-line task.

In the following experiment, participants are presented with a probe recognition task, similar in design to that of MacDonald and Just (1989) discussed above, where one set of the probes corresponds with the nominal object comprising the negated verb phrase (e.g. *book* in (*didn't*) *read a book*). Accordingly, it is predicted that there will be differential response times to the probe recognition task based on whether the critical clause is negated or not. However, unlike MacDonald and Just, where negation ranged over objects and the concepts associated with them but left some larger event (such as a *baking* event) intact, the current experiment is designed to supress the event itself from forming part of the developing situation model. As a result, while the negation of nominal concepts led to decreases in reaction time compared to non-negated conditions in previous literature, the present study should leave only the surface representations of the words active in negated conditions with nothing of the critical event to be encoded in an abstracted mental model of the situation described. A negated condition, in which only surface representations are utilised, is predicted to *decrease* response times in a probe task compared to a non-negated control, in which both event and surface representational information is active (where the former is stronger and more salient than the latter). While this hypothesis is somewhat speculative, with no prior research being done in this area, the direction of this predicted effect is assumed as comprehenders are expected to draw upon only surface linguistic representations in negated event conditions without having to retrieve and 'search' through abstracted mental models of the aggregate situation for the object that corresponds with the probe, as would be the case in the control condition. The central underlying assumption here is that 'searching' through message-level representations is more process intensive than accessing surface-level traces of the linguistic text, and this ought to be reflected behaviourally in reaction times to the probe word task. To my knowledge, prior to this study, this assumption remains untested, and so the results of this experiment promise to provide insights into the relative processing costs of accessing message- and surface-level representations during retrieval tasks.

Both syntactically explicit clauses, with the full verb phrase predicate spelled out, and syntactically elided clauses, with only pro-form 'do' (realised as sequences like *did so too* and *didn't either*), are included in this study to compare access to their respective representations. Response times to probe words that have no surface representation in the preceding narrative, such as words not found in the narrative, are expected to be recognised and responded to slower than probes that are repeated, but these probes should show no differences between negated and non-negated narrative conditions, as they do not form part of the critical event being suppressed.

2.2 – Experiment 1: Methods

2.2.1 – Participants

A total of 62 participants were recruited via the Amazon Mechanical Turk (MTurk) platform to take part in this experiment. All participants were over 18-years old, native English speakers, and were offered a small monetary reward for their participation. Due to the prominence of MTurk in the United States, all participants were from and were, at the time of the study, currently living in the US. Due to a programming error, additional biodata information such as age and gender was not able to be matched with responses for this experiment. Ethical approval was obtained for this study from the University of Essex Ethics Committee.

2.2.2 – Materials

For this study, a total of 80 two-sentence narratives were constructed in a 4x2 design. Each narrative was told in first-person by some protagonist and a friend of theirs. The first sentence of each narrative introduced a critical event in the form of a V+NP accomplishment verb (Vendler, 1957; 1967), such as *reading a book*, *writing a novel*, *climbing a mountain*, and *building a house*, and so on. An indefinite article was chosen to avoid presupposing given information in the narrative, and to avoid the implication that the objects of these verbs were somehow already relevant (Giora et al., 2004; Tottie, 1994).

This initial, critical sentence took the form of two clauses, as illustrated in Table 1. The initial clause of this sentence contained an explicit mention of the critical event and was either in the affirmative (conditions A & B) or was negated (conditions C & D); the second clause was always an elided clause that implied the same conceptual accomplishment event as the main clause and was also either in the affirmative (A & C) or negated (B & D). The second sentence of each narrative was a filler sentence. These filler sentences introduced some other, non-critical, event (also in the form of V + NP accomplishments) that the protagonists in the story took part in. After the presentation of each two-sentence narrative, participants were presented with a probe word[12]. The probe word displayed was from one of two categories:

i. Critical Probe: a noun that appeared within the critical event (accomplishment verb phrase) of the first sentence (e.g. *book* in *read a book*);

ii. Non-Narrative Probe: a semantically unrelated noun which did not appear anywhere in the preceding narrative (or in any narrative in the study).

Table 1: Examples of two-sentence narrative structures for all four conditions.

Condition	Critical Sentence	Filler Sentence	Probe	Probe Type
A	Noah wrote a poem last Tuesday, and I did too.	This morning I tidied my bedroom and later we stole a purse.	poem	Critical
A	Elijah wrote a poem last Tuesday, and I did too.	This morning I pranked my brother and later we sang a song.	tiger	Non-Narrative
B	Liam wrote a poem last Tuesday, but I didn't.	This morning Liam shaved his legs and later we drank some lemonade.	poem	Critical
B	Lucas wrote a poem last Tuesday, but I didn't.	This morning Lucas revised for an exam and later we watched a movie.	tiger	Non-Narrative
C	Ava didn't read a book last week, but I did.	This morning I walked in the country and later we bought a present.	book	Critical
C	Amelia didn't read a book last week, but I did.	This morning I drew some manga and later we explored a castle.	shelf	Non-Narrative
D	Isabella didn't read a book last week, and I didn't either.	This morning Isabella listened to some music and later we drove to the store.	book	Critical
D	Evelyn didn't read a book last week, and I didn't either.	This morning Evelyn cuddled a teddy bear and later we played a game.	shelf	Non-Narrative

[12] The experiment was designed such that both of the two probe word types had an equal chance of being assigned to the end of the four narrative conditions, resulting in 10 of the 20 trials of condition A being accompanied by a critical probe word and 10 being followed by a non-narrative probe; likewise, 10 trials from condition B were accompanied by both a critical probe and a non-narrative probe, and so on, such that each probe type appeared with each condition 10 times.

Finally, to reduce the possibility that participants were simply memorising certain key words from the text, each trial also included a short 3-letter memory task. At the beginning of each trial, participants were presented with a 3-letter string from the English alphabet. No letter appeared more than once in any 3-letter string, and no sequence included or resembled an English word or had the phonotactics of English that would make it relatively easy to recall as a single chunk or syllable. Before each probe word task, a single letter of the English alphabet appeared on screen. In 50% of trials, this letter appeared in the 3-letter sequence; in 50% of trials, it did not. Participants were instructed to respond whether this letter had appeared in the initial sequence for this trial. (See *Appendix A* for a full list of experimental stimuli.)

2.2.3 – Procedure

This experiment was conducted on the online experimental software package PsyToolkit (Stoet, 2010, 2017), and distributed via Amazon Mechanical Turk's requester system, a crowdsourcing marketplace for surveys and experiments. Data was collected online using this method due to Covid-19 complications limiting safe access to face-to-face data collection.

All stimuli were presented in Arial 20 typeface, in white font, on a black background. The 3-letter sequence, which formed part of the memory allocation distractor task, was presented at the start of each trial for 1500ms. Following this, each sentence of the narrative was presented to participants, in full, for a total of 4000ms each. Only one sentence was presented at a time. After the second sentence, the single letter recall element of this task was presented until either the participants confirmed or rejected this as appearing in the previous sequence by selecting the appropriate option on their keyboard, or ten seconds had elapsed – likewise with the subsequent probe word task. At the end of each trial participants could progress to the next trial at their own pace by pressing the space bar or take a break if they wished. Participants were given visual feedback for both the letter-sequence task and the probe word task in the form of

‘Correct’ or ‘Wrong response, or too slow’ messages on screen. Trial order was pseudorandomised across four blocks, with all participants seeing all trials.[13]

2.3 – Experiment 1: Results

2.3.1 – Analysis

While Amazon’s Mechanical Turk platform has been found to replicate common psychological response-time measures in Stroop, Flanker and Simon tasks (Crump et al., 2013), as well as other psycholinguistic measures such as filler-gap dependency processing (Enochson & Culbertson, 2015), the culture and patterns of behaviour among some workers on the MTurk platform introduces additional complications. Ford (2017) notes an alarming trend among MTurk users of what they and Smith et al. (2016) call ‘speeders’: respondents who essentially speed through surveys without reading questions carefully and only applying the minimum effort required in terms of mental engagement to be paid as soon as possible. To account for this, as well as increase the likelihood that the data under analysis came from participants performing in good faith, participants with lower than 50% accuracy in the post-narrative probe word task were immediately excluded from the final analysis.

Response times from the remaining participants were then trimmed, with unrealistic response times (below 100ms) and particularly long response times (over 3000ms) removed. In line with MacDonald and Just (1989), individual participant means were then calculated, and trials with response times more than 2.5 standard deviations away from each participant’s mean were rejected. Finally, to further reduce the chance of including participants who did not perform

[13] As this study used a repeated-measures design, each (narrative) item selected for the experimental trials could have an influence on the results, such as a familiarity effect from participants seeing the same item across different conditions. Standardly, a Latin Square counterbalancing system of lists is used to create a within-item design to control for this potential problem. Due to experimenter oversight, this practice was not implemented in this study. However, steps were taken to maximise variation of character names and other elements in each narrative item, such that no two-sentence narrative was actually identical, even outside of the critical manipulations of negation, minimising any potential confound from this oversight.

the task in good faith, or ‘speeding’, any participant with fewer than 50% of total trials remaining were excluded from the final analysis. One additional participant was excluded due to performing very poorly in one of the conditions, with no accurate responses or times within the thresholds established. The number of participants included in the final analysis was 36.

Remaining response times were then log-transformed and modelled using a linear mixed effects regression model using the *lmer* function (lme4 package) in RStudio. Participant and Trial were included as random effects. Log-transformed Reaction Times were set as the dependent variable, with Condition and Probe Type as independent variables. Word Length and log-transformed word Frequency were also included as factors within the model, with the full model taking the form of lmer(logRT ~ Condition*ProbeType + LogFreq + Length + (1 | PPT) + (1 | Trial). Categorical fixed factors in the model (Condition and ProbeType) were numerised using the deviation contrast scheme prior to running the model. Planned contrasts were then conducted (emmeans package) directly comparing the three negated conditions to the control condition and then each other, with p-values adjusted using Bonferroni’s correction. Finally, to further explore the shape of the data, a Bayesian analysis was conducted (using the rstanarm and bayestestR packages) with 10 chains and 5000 iterations (1000 warmup), and the credible interval set at 0.95; prior distributions were set as weakly informative defaults in the model. Length and frequency were again included in the model, with participant and trial as random effects. Probability of direction effects, credible intervals (CI), Bayes factors (BF), and percentage in region of practical equivalence (ROPE) values are reported. Only correct trials were included in these reaction time analyses.

Finally, a subsequent linear mixed effects model was also run using accuracy of probe responses as the dependent variable instead of reaction times. All fixed and random factors were coded the same as in the reaction time model. Only participants included in the reaction time model were included in the accuracy model to prevent undesirable confounds introduced by probable ‘speeders’.

2.3.2 – Findings

When presented with critical probes, participants responded faster to negated contexts than non-negated contexts, with the strongest facilitation effect observed where both the main and elided clause of the preceding critical sentence were negated (i.e. condition D), with the slowest mean response time to narratives in which both clauses in the critical sentence were in the affirmative (condition A). When presented with non-narrative probes, responses times across conditions were roughly equivalent. Mean response times are illustrated in Fig. 1 and reported with standard deviations and interquartile ranges in Table 2 below.[14]

Testing for significance, the linear mixed effects regression model showed an interaction effect in which the negation of both clauses (D) resulted in faster reading times for narrative probes than non-narrative probes (β = -.047, SE = .015, t = -3.08, p < .01). No other interaction effects were observed: responses to probes following condition C (negative-affirmative) narratives were not significantly faster when that probe appeared in the narrative compared to when it did not (β = -.0138, SE = .015, t = -0.93, p = NS); likewise, responses to probes following condition B (affirmative-negative) were not significantly faster for narrative probes (β = .0235, SE = .015, t = 1.57, p = NS). A significant main effect was observed for condition C (negative-affirmative), where probes following that condition were responded to faster than the control (β = .01, SE = .015, t = -2.42, p < .05). No other factor was significant in the model: condition B (β = -.005, SE = .015, t = -0.39, p = NS), condition D (β = -.01, SE = .015, t = -0.67, p = NS),

[14] Mean word length across all conditions was 5.6 characters per word. For conditions A and B, mean word length was 5.75 characters per word; for conditions C and D, mean word length was 5.45 characters per word. Mean frequency across all conditions was 222.85 instances per million (SUBLEX UK). For conditions A and B, mean frequency was 140.86 instances per million; for conditions C and D, mean frequency was 304.83 instances per million. This imbalance was unintended. However, as no condition showed statistically significant results by frequency or word length, this oversight is unlikely to have influenced the results.

LogFreq (β = -.002, SE = .005, t = -0.44, p = NS), Length (β = -.007, SE = .006, t = -1.29, p = NS), ProbeType (β = -.0006, SE = .01, t = -0.06, p = NS).

Simple effect comparisons between the control condition and the three negated conditions for narrative probes reveal that condition D was significantly faster than the control (β = -.148, SE = .04, t = -3.991, p < .001), as was condition C (β = -.136, SE = .04, t = -3.796, p < .001). Although condition B showed a direction of effect of being faster than the control, this difference was not significant (β = -.066, SE = .04, t = -1.835, p = NS). No difference was observed between conditions C and B (β = -.069, SE = .04, t = -1.886, p = NS), D and B (β = -.081, SE = .04, t = -2.141, p = NS), or D and C (β = -.012, SE = .04, t = -0.308, p = NS).

Simple effect comparisons for non-narrative probes revealed no significant results: B-A (β = -.036, SE = .03, t = -1.154, p = NS); C-A (β = -.045, SE = .03, t = -1.413, p = NS); D-A (β = .021, SE = .03, t = 0.648, p = NS); C-B (β = -.008, SE = .03, t = -0.264, p = NS); D-B (β = .058, SE = .03, t = 1.744, p = NS); D-C (β = .066, SE = .03, t = 2.015, p = NS). Reaction time results for each condition for both critical and non-narrative probes are represented in Fig. 2.

Table 2: ***Summary of Mean Response Times to experimental conditions by probe type.***

Condition	Probe Type	Response Time (ms)		
		Mean	SD	IQR
A	Critical	860.8	379.2	536.5
B	Critical	794.1	355.9	502.6
C	Critical	794.6	353.2	422.5
D	Critical	766.0	371.0	489.3
A	Non-Narrative	830.1	398.1	463.5
B	Non-Narrative	794.9	369.9	378.0
C	Non-Narrative	809.1	402.2	497.5
D	Non-Narrative	816.7	374.0	416.3

Fig. 1: ***Mean Response Times to experimental conditions by probe type (blue = critical probe; red = non-narrative probe).***

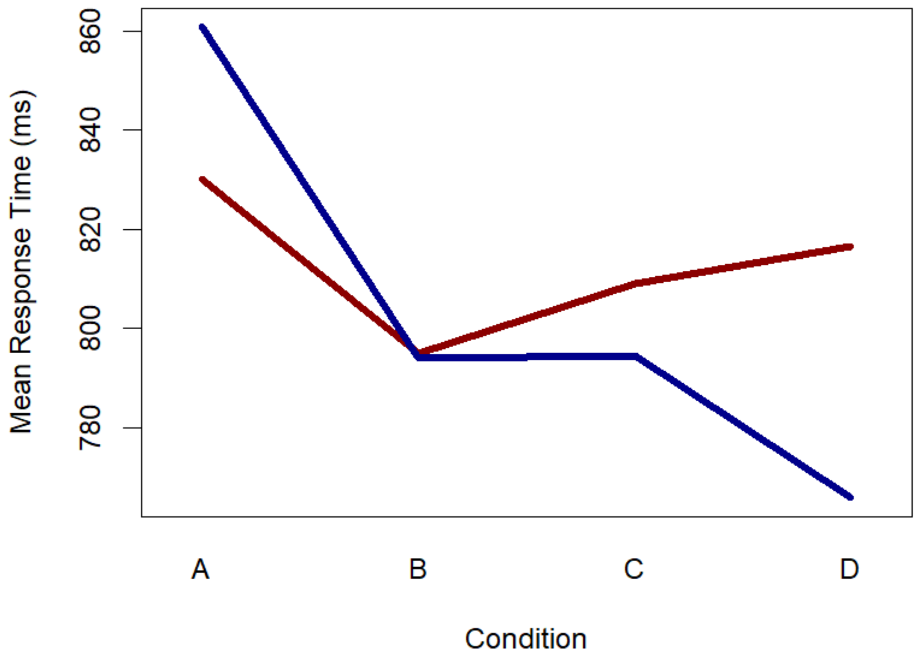
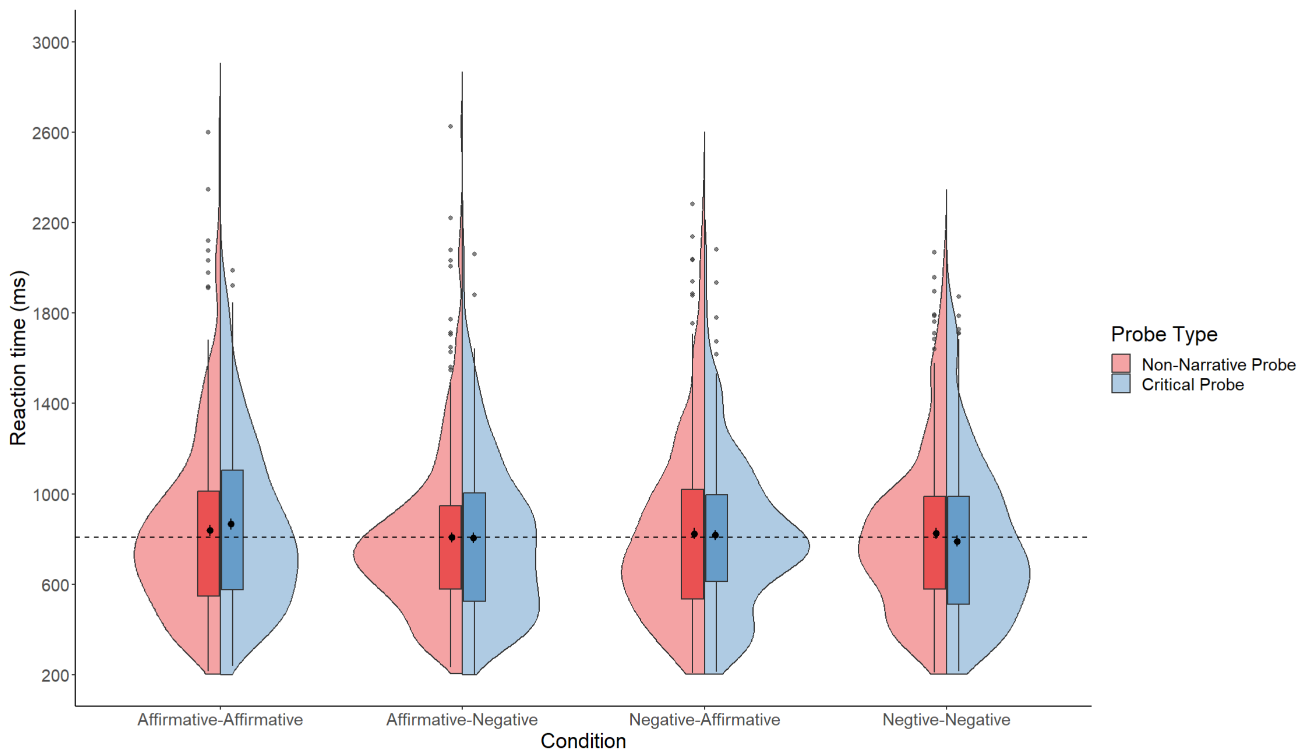


Fig. 2: ***Split violin and boxplots of experimental conditions by probe type.***



To better estimate and describe the shape of this data, a Bayesian analysis was conducted. For critical probes, posterior distributions show that the effect of the affirmative-negative condition (B) compared to affirmative-affirmative condition (A) has a probability of 96.94% of being negative / faster (Median = -0.07, 95% CI [-0.14, 0.00], BF = 0.069) with 31.83% falling in the region of practical equivalence (ROPE), meaning we are unlikely to be able to reject the null hypothesis. However, both the negative-affirmative condition (C) and the negative-negative condition (D) showed a probability of over 99% of being negative /faster than condition A (Median = -0.14, 95% CI [-0.21, -0.07], BF = 11.22), <1% in ROPE; Median = -0.15, 95% CI [-0.22, -0.07], BF = 29.93), <1% in ROPE) suggesting we are likely to be able to reject the null hypothesis for these comparisons.

These effects were not observed for non-narrative probes – neither for condition B (87.74% probability of being negative / faster (Median = -0.04, 95% CI [-0.10, 0.03], BF = 0.021) with 67.87% falling in ROPE), condition C (91.89% probability of being negative / faster (Median = -0.04, 95% CI [-0.11, 0.02], BF = 0.029) with 58.41% falling in ROPE), nor condition D (72.59% probability of being positive / slower (Median = 0.02, 95% CI [-0.05, 0.09], BF = 0.013) with 83.41% falling in ROPE) – suggesting the null hypothesis should not be rejected for responses to non-narrative probes.

Accuracy of probe responses was also modelled from the data via a linear mixed effects regression model. Mean accuracy for critical probes was 69%, while mean accuracy for non-narrative probes was 77%. (Note: only accurate trials were included in the preceding reaction time analysis.) Accuracy was quite consistent across narrative conditions A, B and C for both critical (74%, 72%, 70%) and non-narrative probes (81%, 81%, 80%), although accuracy was higher for non-narrative probes. Interestingly, accuracy was lower for all probes following narratives in condition D – for critical probes, 62%, and for non-narrative probes, 67%. Statistical analysis show that there was no significant difference in accuracy for any of the three non-control conditions when the probe was critical or non-narrative, and no interactions of condition and probe type on accuracy (Condition D : Narrative Probe – (β = .013, SE = .014, t = 0.932, p = NS); Condition C : Narrative Probe – (β = -.010, SE = .014, t = -0.722, p = NS); Condition B : Narrative Probe - (β = -.009, SE = .015, t = -0.632, p = NS). While the main effect of Probe Type approached significance (β = -.038, SE = .019, t = -1.98, p = .052), suggesting that comprehenders trended in the direction of being more likely to respond accurately to probes not found in the preceding narratives than to probes that were found in those narratives, no other factor in the model reached or approached significance: Condition B (β = .006, SE = .015, t = 0.429, p = NS); Condition C (β = .021, SE = .014, t = 1.505, p = NS); Condition D (β = -.016, SE = .014, t = -1.124, p = NS); LogFreq (β = .003, SE = .005, t = 0.691, p = NS); Length (β = -.001, SE = .006, t = -0.179, p = NS).

2.4 – Discussion

In the MacDonald and Just (1989) study that provided the inspiration for the present research, participants were presented with a probe word following a short sentence that either negated or did not negate a noun phrase and the conceptual object it represented. For example, in (26) above, repeated below, responses to the probe word *bread* were slower than to the probe *cookies*, and vice versa for (27), where in the former *bread* was negated in the critical sentence but not *cookies*, while in the latter *cookies* was negated but not *bread*.

26) Mary baked some cookies but no bread.

27) Mary baked no cookies but some bread.

Implications about the function of negation on the psycholinguistic processing of objects introduced by and negated within texts developed out of MacDonald and Just's findings, leading Kaup (2001) to claim that the accessibility or inaccessibility of concepts – whether in scope of negation or not – is a function of their representation in the reader's developing situation model. However, negation within the MacDonald and Just (1989) paper, as well as in most of the subsequent work on negation processing and representational access, has centred on noun phrases – or, in some cases, adjectives (see Kaup et al. (2006) and Orenes et al. (2014), and their work on sentences like *the door is not open* compared to *the door is closed*). A noticeable gap in the field was an exploration of the function of negation on accessibility to concepts when verb phrases, which typically encode the event itself as well as the objects within them, are negated. Given the centrality of events in the construction of situation models (Altmann & Ekves, 2019; Elman & McRae, 2019; Knott & Takac, 2020; Kuperberg, 2016; McNamara & Magliano, 2009; Metusalem et al., 2012; Pettijohn et al., 2016; Radvansky & Zacks, 2017; Zacks, 2020; Zwaan, 2016; *inter alia*), the present research aimed to address this glaring omission in the literature, adopting several of the design elements of MacDonald and Just's (1989) probe word task alongside critical sentences with negated and non-negated verb phrases. Negation in this study ranged over VPs in either full and syntactically explicit clauses, conjoined elided clauses, or both, and were compared to a non-negated control.

The results of this study show that for critical probes (i.e. nouns that appeared within the first sentence), mean response times were 860.8ms when no negation was included in the preceding narrative. Importantly, these response times were *slower* than response times to any of the three negated conditions. Only the conditions in which the explicit clause was negated (C and D) reached statistical significance, reflecting in interaction terms and main effects within the main model and within subsequent planned comparisons between conditions. Condition B, where only the elided clause was negated, failed to reach the threshold of significance, both as a main effect in the model and in subsequent comparisons of simple effects. A Bayesian analysis of the data suggested that while confidence would be high in rejecting the null hypothesis when comparing conditions C and D to the non-negated control of condition A, this was not the case for condition B, with 31.83% of the credible interval falling in the region of practical equivalence (or null region), and a Bayes factor of less than 0.1. It is noteworthy, that while the mean response times for conditions B and C were similar (as seen in Table 1 and Fig. 1), the interquartile range, and therefore the spread of the data, differed between these conditions, with condition C showing a much tighter range than B or A, illustrated in the boxplots of Fig. 2. Consequently, the conditions in which negation ranged over full clauses reached statistical significance, whereas negation within only elided clauses did not seem to significantly affect accessibility to probes, such that the observed effect seems to be predicated on the negation of the explicit clause which contains the explicit textual reference to the concept being probed. Interestingly. no significant difference was observed between the three negated conditions, suggesting a lack of a gradation effects with increased negation, although the mean response times for each condition do hint at a possible trend in that direction. Further work emphasising the role of negation in elided constructions and the effects of multiple instances of negation on accessibility to probed nominal concepts is encouraged.

Non-narrative probes (i.e. nouns which were semantically unrelated to critical probes and did not appear anywhere in the preceding narratives) were also included at the end of 50% of the trials. Their inclusion acted as baseline to measure whether the effect observed in the critical probe task was due to the manipulation of negation within the preceding narrative on subsequent accessibility to the object that appeared within that event. As the objects introduced by the non-narrative probes did not appear in any negated or non-negated sentence, it was expected that there should be no significant differences in response times to these probes in any of the four experimental conditions. The results of this experiment indeed show no significant difference between any of the four conditions to non-narrative probes, supporting the conclusion that the effects observed to the critical probes were a direct result of the manipulations of VP negation. Word frequency effects, typically observed in psycholinguistic processing tasks (see Brysbaert et al. (2018) for a recent review) failed to reach significance in the models for both critical and non-narrative probes..Although only accurate trials were considered in the reaction time analysis, an analysis of the accuracy of responses to probe words across the four conditions was also conducted, showing no interaction between condition and probe type, nor any main effect of condition, word frequency or word length. Nonetheless, probe type as a main effect did approach significance, suggesting that comprehenders were more likely to respond accurately when that probe was a non-narrative probe than when it had appeared in the preceding text (critical probe). Moreover, while not statistically significant, accuracy to probes following condition D – where both elided and explicit clause negation was in effect – was lower than to the other three conditions for both critical and non-narrative probes. These findings suggest that processing multiple instances of negation is possibly more resource taxing, resulting in lower accuracy to the experimental task than the other conditions;

furthermore, comprehenders seem to be more adept at accurately gauging that a non-narrative word did not appear in the text than they are at determining that a critical probe did appear in the narrative, regardless of condition. Unfortunately, the original paper by MacDonald and Just (1989) did not analyse their accuracy results, so no comparison is able to be made. As statistical significance was ultimately not reached on any measure in this model, no additional consideration of the accuracy results will be made here. Nevertheless, as interesting implications for probe word tasks (especially involving negation) are potentially observed, further work in this area is recommended.

Returning to the response time measures central to this study, then, these findings demonstrate that negation of accomplishment verb phrases within explicit clauses decreases response times to nominal probe words that appeared within the critical event (e.g. *book* in (*didn't*) *read a book*). However, negation of elided clauses, which only implicitly reference the accomplishment event and the object within that event (e.g. *and I did so too* or *and I didn't* (*either*)) do not appear to significantly reduce response times to critical probes, although there is trend in that direction that warrants further investigation in future research. These results of nominal accessibility following verbal negation in the data above reveal striking differences when compared with MacDonald and Just's (1989) study and the subsequent body of research examining the function of negation on accessibility to objects when negation ranges over nouns. Crucially, while a repeated finding in the literature is that the negation of nouns (e.g. *no bread*) non-obligatorily supresses the concept introduced by that noun from forming part of the comprehender's unfolding situation model and subsequently reduces accessibility to that concept, the present study discovered something quite different. When the event itself, including any entities germane to that event, is under the scope of negation and supressed from forming part of the developing mental model, accessibility to relevant nominals (but not nominals that did not appear within the narrative) seems to be increased.

However, it is possible that what is happening in cases of verbal negation is not that object concepts actually have increased accessibility in the situation model, but that the reduced response times here reflect that there are actually two parallel processes of information access available during subsequent retrieval of representational content, such as in probe word tasks, and that in negated verbal contexts, particularly when the context is syntactically explicit, comprehenders rely on one of these methods more, as the other is not available to them.

The first of these processes would relate to message-level access, and would be associated with Kaup's (2001) assertion that concept accessibility is a function of its representation within the situation model; the second, would relate to surface-level representations, which consist of orthographic and phonetic information, and which facilitate processing, even after a delay (Cloitrew & Bever, 1988; Ledoux et al., 2006; Lüdtke et al., 2008; Tanenhaus et al., 1985). While memory for surface representations is typically considered to be weaker, and decays faster, than message-level representations (Anderson, 1974; Begg & Wickelgren, 1974; Zimny, 1987), surface representations are strong early in the time-course of retrieval (Ratcliff & McKoon, 1989) and have a long history of being considered as crucial elements of language comprehension models (Kintsch, 1988; McNamara & Magliano, 2009; Zwaan & Radvansky, 1998). While in negated nominal contexts both surface- and message-level representations would be active, in negated verbal contexts the event itself does not 'happen' and would therefore be suppressed from forming part of the developing model. As a result, comprehenders of negated verbal contexts – at least in the case of unambiguously non-events expressible in the logical form $\neg\exists e$, as discussed in example (31) above – would not have a message-level representation of the event to access to be able to retrieve conceptual information; they would have only the surface-level representations available to them.

If this interpretation is correct, for condition A, where the critical sentence contained no verbal negation, both a message-level and surface-level representation of the critical noun would be available, with the message-level representation being the more salient of the two for competent readers (Kintsch, 2018; Kintsch et al., 1990). Contrastingly, for a negated condition such as D, where both full and elided clauses in the critical sentence were negated, no message-level representation of that action event and pertinent object (e.g. *reading* and *book* in *didn't read a book*) would be constructed, leaving only the surface-level representations available. Interestingly, when only one clause is negated, and the negated clause is explicit as in condition C, this facilitation/suppression effect is observed, but when the negated clause is elided as in condition B, the difference in response times fails to reach significance, perhaps because there is no explicit orthographic form of the critical noun in the elided clause, meaning the expression in which that word is found is non-negated, just like the control condition. However, there does seem to be a trend in the direction of a facilitated response in the negated elided condition demanding further investigation.

An important takeaway is that the findings of this study are consistent with Kintsch et al. (1990) and Kintsch (2018), who argue that sentences are represented in memory at several levels of representation in a connectionist-style network of spreading activation: a surface-level, a level which reflects the semantic content of the text, and the situation model (i.e., the abstracted representation of what the text is about). The present chapter makes no meaningful claims about this second level of text-based semantics, which is presumed flexible enough to allow for multiple competing representations of the same structure, as representations (and thus any concepts being probed) are not committed to at this level (Kintsch et al., 1990). The network rejects inappropriate interpretations that have been constructed when integrating them with top-down inferences at the level that corresponds with the situation model (*idem*.). Therefore, the message-level and surface-level routes suggested as explanation of this chapter's findings have parallels – and are consistent – with the surface-level and situation model level representational pathways of language comprehension proposed by Kintsch and others.

Notably, the present findings have an additional theoretical implication: that responses to conditions C and D were significantly faster than A suggests that relying on surface-level representations is potentially more efficient than accessing a comparatively content-rich mental model of the situation described. This intriguing hypothesis could – and should – be tested in subsequent studies on the negation and suppression of events. Surface representations do decay rapidly, however, and so the effect observed here may be limited to tasks performed only a short time after the experimental manipulation, a factor which ought to be more precisely measured in subsequent work, too.

As a final point, in the present experiment, as mentioned, probe words were presented to participants very shortly after they read a preceding narrative, but in every-day linguistic interactions people often need to access information from more than 10 or 20 seconds ago. While this may mean that the observation regarding surface-level and message-level processing efficiencies is limited to a small bandwidth of the human linguistic experience, a captivating conclusion emerges: it is advantageous that humans are predisposed to developing these less processing-efficient message-level situation models of the linguistic input we encounter, as long-term language comprehension would be difficult, if not impossible, without them.

Looking to the future, then, more research must be conducted on the role of negation in verb phrases. Prior to this study, perhaps the best account of the intersection between negation processing and situation modelling was provided by Kaup (2001), discussed above, but much of the evidence in support of these claims has come from experiments on the negation of nouns and adjectives: event-introducing (or -suppressing), verbal negation remains unfortunately understudied. The results of this chapter suggest that the picture is more complicated than has traditionally been assumed. One area of this research that is less clear and demands future investigation is the precise role of negation in elided contexts; there appears to be a trend toward facilitated response times compared to a non-negated control, but these failed to reach statistical significance in the experiment presented here, perhaps as the concept being probed was textually explicit in another event which did form part of the comprehenders' developing situation model. Consequently, the effect of syntactically explicit full clause on the construction of event-based representations appears to be somewhat stronger than elided clauses: negating the former has a more substantial impact on comprehenders' reliance on surface representations than the latter. Additional research should also look to explore negated verbal contexts which do not suppress the underlying event or its message-level representation, as in examples like (33): *Paul did not stop at the red light*. In such constructions, world knowledge relating to traffic and driving would integrate with the semantic information encoded in the words and phrases and be interpretable as something like *Paul drove through the red light*. These pragmatic constructions present wider challenges for a full account of the function of negation in event representation, and thus ought to be part of the effort to push the boundaries of the science.

2.5 – Conclusion

Negation, in on-line processing experiments, has historically been a useful tool in uncovering the mechanisms behind the construction of situation models. This chapter examined the effects of negation on event-introducing verbs in both syntactically explicit clauses and elided clauses by presenting participants with a probe word which either appeared in the critical narrative or did not appear in that narrative. Response times to these probes were obtained, and results show that response times to critical probes were faster in negated contexts, particularly where the negated clause was syntactically explicit, but no such effect was observed for non-narrative probes. These findings challenge existing accounts of the function of negation in subsequent accessibility to conceptual information by suggesting that comprehenders utilise (at least) two parallel routes to retrieve information – a message-level route and a surface-level route – and that in cases where the event itself is prevented from composing into the unfolding situational model as a result of verbal negation, comprehenders rely on fading surface-level representations which provide a more economical route from a processing perspective. This chapter adds to the existing body of research on the construction of situation models and the real-time processing of negation.

Chapter 3

Representing Ambiguously Distributive Events

in the Perfect Timespan

3.1 – Background

This chapter examines how ambiguity interacts with event representations by testing the effect plurality denoting expressions have on the processing and interpretation of event-introducing verbs. Building on research that explores how comprehenders process and determine the meaning of ambiguous constructions where an event either ranges over a collective set of entities or summarises over individual references to each entity separately, the present study asked participants to read short two-sentence narratives while measuring their reading times to critical nouns that were either consistent or inconsistent with these possible interpretations. Results of this study show that a preference for a collective interpretation is ultimately reached, but – contrary to previous findings – representations of both interpretations appear to remain active beyond the sentence that introduced the ambiguity. This is not the case, however, for all types of construction, as those that introduce events with an extended timespan for multiple events to potentially occur are more likely to lead comprehenders to favour a distributive reading where the event ranges over and is performed by each entity separately. Consequences for these findings on event representation and situation model construction are discussed.

3.1.1 – Ambiguity and Distributivity

Language is rife with ambiguity, from phonetic ambiguity (Ingvalson et al., 2011; Rogers & Davis, 2017) and morphological ambiguity (Järvikivi et al., 2009; Xiang et al., 2011) to lexical (Duffy et al., 1988; Khanna & Boland, 2010; Meylan et al., 2021) and structural ambiguity (Frazier & Rayner, 1982; Frisch et al., 2002; Spivey-Knowlton et al. 1993). However, in language processing and representation research, it has been argued that there are levels of representation above the lexical or syntactical; one of the most important of such levels for research on comprehension processes is considered to be the situation model, a holistic amalgam of event and object representations abstracted from the linguistic input that is integrated with existing encyclopaedic knowledge to form a mental model of what the text is about (Altmann & Ekves, 2019; Bohnemeyer & Pederson, 2010; Elman & McRae, 2019; Knott & Takac, 2020; Kuperberg, 2016; Kuperberg and Jaeger, 2016; McNamara & Magliano, 2009; Metusalem et al., 2012; Pettijohn & Radvansky, 2016a; Pettijohn et al., 2016; Radvanksy & Zacks, 2014; Radvansky & Zacks, 2017; Zacks, 2020; Zwaan, 2016; Zwaan & Radvansky, 1998; *inter alia*).

Not much has been written about ambiguity in language that affects situation model construction or the event representations that compose them. One notable area that has received some focus on ambiguous events is that of distributivity, a mereological property of plurality denoting constructions in language that relates predicates to their (multiple) arguments, although the focus here has been more on psycholinguistic processing implications (Dotlačil & Brasoveanu, 2021; Frazier et al., 1999; Maldonado et al., 2019) and the nature of a putative semantic operator[15] ranging over predicates and their arguments (Champollion, 2019; Lasersohn, 2013; Link, 1987; Roberts, 1987), with less being said about the role of distributivity on mental model construction.

Distributivity is observed in cases where a predicate, such as an event-introducing verb, is combined with a plurality denoting expression – for example, quantified noun phrases like *two friends*, or the conjunction of multiple entities like *Jack and Jill*. That event-introducing predicate is then understood to hold of each member of that expression (Champollion, 2019).[16] For example, a distributive reading of (35) would be one in which the protagonist Rebecca

[15] Attempts to formally model the semantics of distributivity has led to the postulation of a distributivity operator (D operator) that applies to a set of events, typically denoted by a verb phrase, and returns another event predicate that applies to any entity whose atomic parts each satisfy that event (Champollion, 2019). The intuitive meaning and function of this D operator is equivalent to the English adverb 'each', such that when a D operator is applied to a predicate in an expression like *two boys bought a book*, it returns a proposition which can be paraphrased as 'two boys each bought a book'. Whether such a putative distributivity operator is psychologically real lies outside the scope of the present chapter; but see Dotlačil & Brasoveanu (2021) for discussion.

[16] Stative verbs also interact with distributive constructions, but as this chapter is primarily concerned with an analysis of non-stative events, all subsequent references to predicates will assume them to be event-introducing in the form of Vendlerian (1957; 1967) activities, achievements, or accomplishments.

gave two cakes, one to each friend (i.e., there were two instances of cake-giving); a similar reading can be obtained for (36), with conjunction.

35) Rebecca gave a cake to two friends.

36) Rebecca gave a cake to Jack and Jill.

Collective, or non-distributive readings, are also obtainable from sentences with plurality denoting expressions (Lasersohn, 1998); (35) and (36) above can both be interpreted as single (collective) cake-giving events, with only a single cake being given in each. Moreover, the plurality denoting expression can be either the object or the subject of a sentence (or both); plurality denoting subjects also give rise to ambiguous event interpretations, as in (37).[17]

37) Two friends gave a cake to Rebecca.

The number of cakes that are represented within a comprehender's situation model during (and after) reading sentences like (37) depends upon their interpretation of the number of events that occurred; that is to say, if one could determine how many cakes in (37) form part of a comprehender's mental model of the situation described, by extension one could determine whether a single event was mentally represented, or multiple events were represented. Given that reading times to plural or singular nouns presented in some task sentence is assumed to reflect consistency or inconsistency with a preceding context, such that inconsistent contexts lead to slower reading times (Jegerski, 2014), adopting a self-paced reading method to a task sentence that follows constructions like (37) above ought to aid in measuring the number of event representations relevant to the developing situation model during comprehension of a text (see section 3.2 below for further discussion and implementation of this method).

[17] Some researchers have argued that it might not be strictly ambiguity, but underspecification, that leads to differences in interpretation following distributive constructions (see Champollion (2020) for a recent state of the art). It is beyond the scope of this study to tease apart these two competing explanations, although see Dotlačil and Brasoveanu (2021) who find evidence in support of the ambiguity account.

3.1.2 – Covert and Overt Distributivity

Distributivity can be both overt and covert (Champollion, 2016a; 2016b). Examples like (37) are covertly distributive, that is they are ambiguous between collective and distributive interpretations. The inclusion of adverbials, such as *together* and *each*, can disambiguate these constructions, leading to unmistakable collective or distributive readings – these constructions are therefore considered overt (Champollion, 2019). In work examining disambiguating adverbs, Frazier et al. (1999) discovered that, in general, comprehenders seem to prefer collective interpretations over distributive ones. By measuring participant eye movements during a reading task with sentences possessing distributive properties (i.e., conjoined noun phrases) that were then disambiguated by an adverb at either the beginning or end of the verbal predicate, as in (38) below, Frazier at al. found evidence that comprehenders appear to be more likely to commit themselves to a collective reading sometime during the processing of the predicate (and before arriving at said disambiguating adverb), as late-disambiguated distributive constructions were slightly more taxing to process for readers than the early-disambiguated alternative (where the adverb appeared before the predicate rather than after).

38) Sam and Maria (*each/together*) carried one suitcase (*each/together*) at the airport.

Importantly, this increased processing cost was only observed to the adverb *each* that forces a distributive interpretation, with no difference observed when comparing the early and late constructions with the adverb *together*, that forces a collective one. This extra processing demand for sentences like *Sam and Maria carried one suitcase each at the airport* was treated by the researchers as evidence that comprehenders default to a non-distributive interpretation in cases where the plurality denoting expression is not immediately disambiguated.

Curiously, comprehenders are not always able to articulate their reasons for this preference to the collective (Clifton & Frazier, 2012), although Harris et al. (2013) propose that collective interpretations may be preferred simply because they are less cognitively costly than distributive ones (with only one event mentally represented instead of two). These apparent biases are averaged effects across participant samples, though; individual variation in collective or distributive preference, strength of preference, and cause of preference, cognitive or otherwise, has, to my knowledge, not been thoroughly explored.

Of course, not every sentence with plurality denoting expressions has overt disambiguating adverbials like those in Frazier et al. (1999); distributivity can be and often is covert and undisambiguated (Champollion, 2016b; 2016c) – as seen with examples (35-37) above. Without disambiguating adverbials, these covert expressions are fundamentally ambiguous between the two readings. Crucially, arriving at a distributive or collective interpretation following these covert constructions is a process of ambiguity resolution. While there seems to be an eventual preference for the collective, it remains an open question as to whether both collective and distributive representations are active during comprehension, or if one (such as the distributive interpretation) is quickly suppressed in favour of the other.

Studies on lexical and syntactic ambiguity have demonstrated that ambiguity processing likely proceeds in parallel rather than in a serial fashion (Gibson & Pearlmutter, 2000; Levy, 2008; Logačev & Vasishth, 2013; McClelland et al., 1989; McRae et al., 1998; Snell & Grainger, 2019; van Gompel et al., 2000) but this avenue of inquiry has not been properly extended to the processing and representation of ambiguous event-introducing predicates with plurality denoting arguments. The primary aim of this chapter is therefore an attempt to build on existing distributivity research to address this knowledge gap. To achieve this, Imeasure reading times to critical nouns consistent or inconsistent with collective and distributive interpretations following ambiguous (covert) sentences to determine how many objects (and thus events) are represented in the comprehender's developing mental model. This research differs from Frazier et al. (1999) in that disambiguating adverbials, like *together* and *each*, are not included. A non-distributive condition, without any plurality denoting expression, is used as a control.

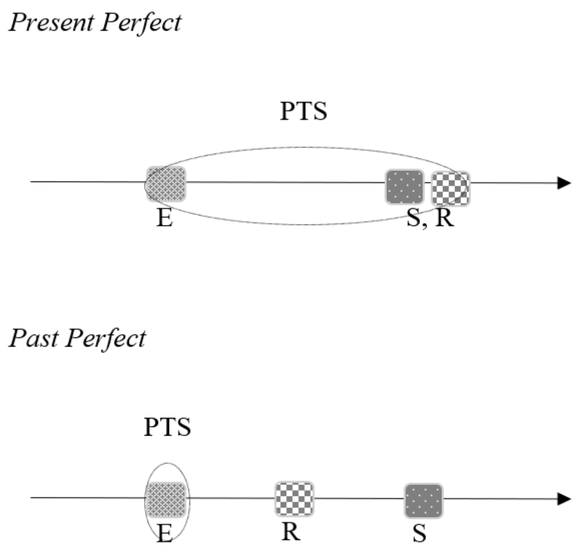
3.1.3 – Temporal Ambiguity and Distributivity

The secondary aim of this chapter is to explore another understudied question regarding ambiguity in these distributive constructions: that of temporal ambiguity. A representational complication with ambiguously distributive constructions (and even overtly distributive constructions) is whether the distributive interpretation (that of multiple cake-giving events) is treated as occurring at the same time or at different times, with constructions like (37) above being vague enough to allow for either reading. In effect, this ambiguity allows for a total of three possible readings of sentences like (37): a single, collective cake-giving event of a single cake that occurs at a single point in time; two, distributive cake-giving events (and accordingly two cakes) that occur at the same time; or two, distributive cake-giving events that occur at different times. The role of temporal ambiguity in the representation of distributive constructions is profoundly understudied, a fact made more egregious by the importance of temporality on events and in the development of coherent situations models. This chapter, therefore, additionally explores whether the interpretation (and thus mental representation) of ambiguously distributive constructions is affected, and can indeed be forced, by manipulations of grammatical aspect which alter the timespan in which the event or events can occur.

Much has been written about the effects of aspect on event representation (see Becker et al., 2013; Bergen & Wheeler, 2010; Ji & Papafragou, 2022; Levin, 2019; Madden & Zwaan, 2003; Malalia, 2014; O'Bryan et al., 2013; Pinango et al., 1999; Romagno et al., 2012; Wagner and Carey, 2003; *inter alia*), although the majority of this literature has attended to the telic-atelic distinction and the real-time processing of temporality, or the segmentation of events in bounded and unbounded contexts. While telicity might offer some productive insights in an exploration of the role of temporal ambiguity in distributive constructions, this chapterinstead looks at a lesser-examined aspectual property of language – the perfect timespan – which, due to introducing a variable-length time interval where events can happen, is hypothesised to have a substantive impact on the representation of plurality denoting expressions in which a distributive reading of two (or more) events can occur at different times.

Events introduced by the present perfect (in English) are subordinate to the time of the utterance, where the event is understood to occur within a timespan that extends into the past but also contains and is relevant to the present moment; for events introduced by the past perfect, those events occur within a timespan that includes that event somewhere in the past but this does not extend to incorporate the time of the utterance (Spejewski, 1998).

Fig. 3: Reichenbachian analysis of tense with respect to the perfect timespan, adapted from Spejewski (1998).



Adopting a Reichenbachian (1947) analysis of tense, the timespan represented in present and past perfect constructions are as illustrated in Fig. 3, where the arrow indicates the progression of time from past to future, S is the time of the utterance or speech, E is the time of the reported event or state of affairs, and R is a reference point, identified from context, that is used in calculating the temporal location of E with respect to some other time (e.g. 'Chris had left New

York… *by some other (unspecified) point in time*'). The perfect timespan (PTS) in which events can occur for past and present are the circled areas in Fig. 3.

With minimal syntactic differences between past and present perfect predicates, but substantive differences in terms of their conceptual timespans, the PTS becomes an excellent diagnostic tool for testing claims about temporal ambiguity in distributive constructions. As a result of this extended timespan, it is predicted that comprehenders will be more likely to adopt a distributive interpretation of a temporally ambiguous sentence where the predicate is in the present perfect than an otherwise identical construction in the past perfect[18].

To that end, by comparing reading times to critical nouns consistent or inconsistent with distributive and collective interpretations of ambiguous constructions that appear in the perfect aspect, the following experiment aims to address two important research questions relevant to the mental representation of events in the contexts of ambiguous sentences:

> **RQ1:** Although it is known that a collective interpretation is eventually favoured, when reading ambiguously distributive (covert) constructions, do comprehenders commit early on to a collective interpretation of the events described in a sentence, or do both distributive and collective representations remain active?
>
> **RQ2:** Given the temporal ambiguity discussed, does the extended timespan of the present perfect aspect increase the likelihood that comprehenders adopt a distributive interpretation over a collective one when compared to past perfect constructions with a relatively reduced timespan?

[18] Due to the atelic and durative nature of Vendlerian (1957; 1967) activities, the continuous construction would offer a similar, although not identical, extended timespan; however, this would be observable in both the past and present progressive aspect – wherein the timespan for the past would encompass time intervals around an event before S (the time of utterance) but would not include S, while the present would encompass time intervals around an overlapping S and event – and so this pair do not offer the same experimentally suitable comparison that the past and present perfect constructions offer, where one has an extended timespan but the other does not. Nonetheless, the progressive might be an interesting avenue for future research on distributivity representations.

3.2 – Experiment 2: Methods

3.2.1 – Participants

A total of 57 participants were recruited via the Amazon Mechanical Turk (MTurk) platform to take part in this experiment. All participants were native English speakers who were more than 18-years old. Due to the prominence of MTurk in the United States, all participants were from the US and were living in the US at the time of the study. Participants were a mix of genders, with biodata information related to age, level of education, and amount of reading per week (coded as literacy) obtained in a pre-task questionnaire. Monetary compensation was provided upon completion of the experiment. Ethical approval was obtained for this study from the University of Essex Ethics Committee.

3.2.2 – Materials

To address the two primary research questions, 64 two-sentence narratives were presented to participants in a self-paced reading task; 32 of these narratives were in the past perfect and were designed to address the initial research question relating to whether both collective and distributive interpretations are active during processing, while the other 32 narratives were in the present perfect, selected to investigate the second research question and the impact of extended aspectual timespans on temporal ambiguity and distributivity preferences.

In the set of past perfect narratives, the first sentence acted as a context sentence which introduced an event in the past perfect, such as *had bought a book*, in the form of a V+NP accomplishment verb (Vendler, 1957; 1967). This context sentence had four variations: in one, the event was conducted by a singular subject (e.g. *Sally*), coded as ‘condition A’; in another, a plurality denoting subject introduced by the numeral *two* (e.g. *two chefs*) preceded the event, coded as ‘condition B’; in the third, a conjoined subject (e.g. *Amy and Sally*) preceded the event, and was coded as ‘condition C’; and in the final type, the predicate followed the conjunction of two plurality denoting subjects (e.g. *two students and* two *chefs*), and was coded as 'condition D', as illustrated in Table 3 below. Including variations of ambiguously distributive constructions, one with a numeral, one with conjunction, and the other with both, allowed for the testing of whether these different types of plurality denoting expression differentially affected collective or distributive readings, including whether multiple instances of covert distributivity biases comprehenders toward a distributive reading. Adopting a four-way paradigm was largely exploratory and was done in alignment with the design of the experimental conditions of Chapter 2 to test for differences and gradations of effect.

Table 3: List of example stimuli, coded for condition, with critical words underlined; plural and non-plural critical words were equally distributed across conditions.

Timespan	Condition	Context Sentence	SPR Task Sentence	Plural Type
Past Perfect	A	Billy had bought a book.	He gave the <u>book</u> to Ruby last month.	Non-Plural
Past Perfect	B	Two teachers had bought a book.	They gave the <u>books</u> to Gwen last Sunday.	Plural
Past Perfect	C	Charles and Billy had bought a book.	Charles gave the <u>books</u> to Anne last Thursday.	Plural
Past Perfect	D	Two writers and two teachers had bought a book.	The writers gave the <u>book</u> to Lisa last night.	Non-Plural
Present Perfect	A	Sally has baked a cake.	She gave the <u>cakes</u> to Noah on Monday.	Plural
Present Perfect	B	Two chefs have baked a cake.	They gave the <u>cake</u> to Evan on Wednesday.	Non-Plural
Present Perfect	C	Amy and Sally have baked a cake.	Amy gave the <u>cake</u> to Owen on Saturday.	Non-Plural
Present Perfect	D	Two students and two chefs have baked a cake.	The students gave the <u>cakes</u> to Andy on Tuesday.	Plural

The second sentence of each narrative was a task sentence, where participants were asked to read one word at a time in a self-paced reading (SPR) manner. The task sentence described the object of the preceding sentence being given away to a third party (e.g., *Amy gave the cake to Owen on Saturday*). However, the critical noun in this task sentence varied, appearing in the singular form (i.e., *the cake*) in 50% of trials and in the plural (i.e., *the cakes*) in the remaining trials. Where the singular was included, this presupposed that only a single cake was baked (or book was bought, etc.) in the preceding event; where the plural was found, multiple cakes and thus multiple baking events were implied. The plurality of this critical noun would either be consistent or inconsistent with a comprehender's mental representation of the number of events introduced in the context sentence; a collective interpretation would be consistent with a singular noun, while a distributive interpretation would be consistent with a plural noun. Reading times to the critical noun were obtained[19], with longer reading times assumed to reflect inconsistency – and shorter reading times reflecting consistency – with the comprehender's representation of the number of events (Jegerski, 2014). Reading time measures were used to address RQ1 by examining whether the experimental manipulations in the context sentences – different types and degrees of ambiguously distributive constructions (conditions B, C, D) – affected reading times to these critical nouns. Condition A, in which there was only a singular subject in the context sentence, acted as an experimental control against which the other conditions were compared, as A was predicted to strongly favour continuations with non-plural critical nouns in the task sentence due to being unambiguously non-distributive.

To address RQ2, reading times to critical pluralised continuations following the other 32 narratives in the present perfect, consisting of accomplishments like *has*/*have baked a cake*, were analysed. The grammatical structure of the context sentence in this set was otherwise identical to the past perfect narratives, including the four-way distinction between conditions (again, coded as A, B, C and D, as illustrated in Table 3), as was the subsequent SPR task sentence and the plural / non-plural alternation of the critical word. (See *Appendix B* for a full list of experimental stimuli.)

[19] Reading times to the spillover region of the subsequent prepositional phrase were also obtained but are not discussed further in this chapter due to failing to reach significance on any measure. A sentence-final temporal adverbial such as *on Saturday* was included in the task sentence to avoid sentence-wrap up effects on the critical noun or spillover region.

3.2.3 – Procedure

This experiment was conducted using the online experimental software package PsyToolkit (Stoet, 2010, 2017), and distributed via Amazon Mechanical Turk's requester system, a crowdsourcing marketplace for surveys and experiments. Data was collected online using this method due to Covid-19 restrictions limiting safe access to face-to-face data collection.

Stimuli were presented in Arial 20 typeface, in white font, on a black background. The context sentence was presented to participants in full for a maximum of 30 seconds, allowing participants to read it in their own time and move onto the second sentence when they were ready by pressing the space bar on their keyboard. The task sentence was presented to participants one word at a time in a self-paced reading fashion, with a maximum duration of 3 seconds each. However, participants were instructed to press space on their keyboard to move onto the next word at their own pace. A centralised fixation point, indicated by a '+' symbol, was displayed between the context and task sentences for 1000ms to focus attention to where the individual words of the narrative's second sentence would appear.

At the end of each trial, participants were asked a naturalness rating question in the form of *How much sense did the story you just read make?* Responses were given using a rating scale from 1-5, with 1 coded as 'not much sense' and 5 as 'complete sense'. Participants were required to use their mouse cursor to click on their preferred option in a deliberate effort to increase participant engagement with the task and reduce the impact of 'speeders' among the participant pool (see Analysis section below for discussion). At the end of each trial participants could progress to the next trial at their own pace by again pressing the space bar or take a break if they desired. Trial order was pseudorandomised across four blocks, with all participants seeing all trials.[20]

---

[20] As this study used a repeated-measures design, each (narrative) item selected for the experimental trials could have an influence on the results, such as a familiarity effect from participants seeing the same item across different

3.3 – Experiment 2: Results

3.3.1 – Analysis

Although MTurk has been used to successfully replicate common psychological measures, such as response times to Stroop and Flanker tasks, as well as measures like filler-gap dependencies (see Crump et al., 2013; Culbertson, 2015), researchers such as Ford (2017) and Smith et al. (2016) have noted that a number of respondents, branded 'speeders', have been observed to rapidly progress through surveys and experiments on this platform in order to be paid as quickly as possible. With these observations in mind, and with an aim to reduce the impact of potential 'speeders', participants in the present study were excluded from the final analysis if more than 50% of their trials showed speeded responses in the self-paced reading measures (or exceeded the 3 second duration of each item in the task sentence). A liberal 100ms was chosen as a lower cut-off point for speeded responses.[21] Due to the high number of potential 'speeders' on the MTurk platform, the original participant sample size was reduced from 57 to 26, with reading times for the remaining participants systematically trimmed, rejecting individual participant responses more than 2.5 standard deviations away from that individual's mean. Reading times to critical words were residualised by word length using the following formula: (reading time / word length) x mean word length. These residualised reading times were then log transformed to normalise for statistical analysis. Data was analysed

conditions. Standardly, a Latin Square counterbalancing system of lists is used to create a within-item design to control for this potential problem. Due to experimenter oversight, this practice was not implemented in this study. However, steps were taken to maximise variation of character names and other elements in each narrative item, such that no two-sentence narrative was actually identical, even outside of the critical distributivity manipulations, minimising any potential confound from this oversight.

[21] Valid reading times of less than 100ms are generally not possible (Luce, 1986). A higher threshold of 200ms was considered, but as individual participants' reading times were subsequently trimmed to exclude values more than 2.5 standard deviations away from their individual means, the lower absolute cut-off was chosen. It is worth noting that other researchers have also adopted a similar 3000ms upper cut-off for native speakers reading times (Havik et al., 2009; Roberts & Felser, 2011). Also see Jegerski (2014) for a review of SPR cut-offs.

using R Studio (R version 4.3.1) and modelled using linear mixed effects regression with the *lmer* function (lme4 package).

To address the first of this study's research questions looking at whether both collective and distributive interpretations are active during on-line processing, the first model looked at reading times to critical words for past perfect constructions. Log-transformed residualised reading times were established as the dependent variable, with an interaction term for Condition and Plural Type as the predictor. Critical word frequency, as well as participant age, gender, education, and literacy were also included as factors in the model. Participant and Trial were included as random effects, with the full model taking the form of lmer(LogRT ~ Condition*Plural_Type + Age + Sex + Education + Literacy + (1 | PPT) + (1 | Trial). Categorical fixed factors were numerised using the deviation coding scheme prior to running the model. Planned contrasts were then conducted (emmeans package), comparing reading times of each of the distributive conditions to the non-distributive controls in both plural and non-plural contexts, as well as to their counterpart conditions in the other plurality (e.g., condition B with a plural continuation against B with a non-plural continuation). P-values for multiple comparisons were adjusted using Bonferroni's correction.

To address the second research question regarding potential commitment differences to distributive contexts in extended time spans, the linear mixed effect model above and ordinal regression model (discussed below) were repeated for the present perfect constructions.

Finally, to analyse differences in naturalness ratings reported by participants at the end of each trial, an ordinal logical regression model (*polr* function) was implemented; odds ratios and *p*-values for rating differences between plural and non-plural continuations for each of the four distributivity conditions are reported, for both past perfect and present perfect narratives.

3.3.2 – Findings

3.3.2.1 – Research Question 1

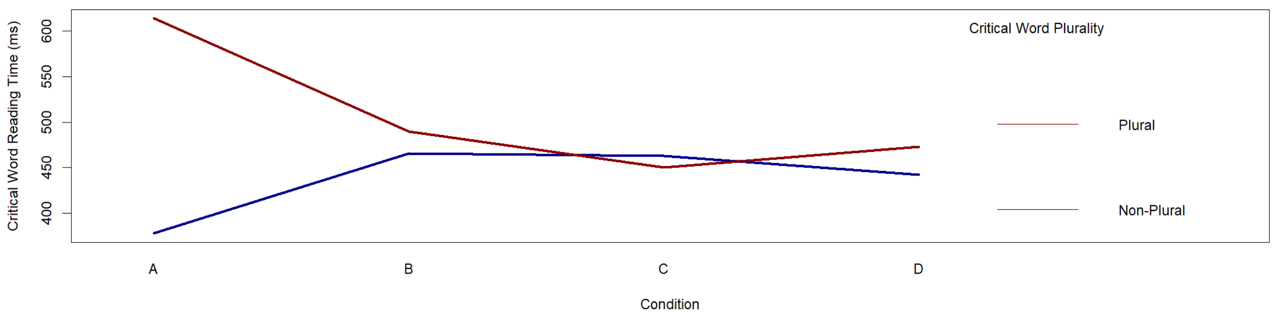
In past perfect narrative contexts, the linear mixed effects regression model with Condition and Plural Type as interaction terms showed a marginally significant effect of plurality, with reading times to plural words read slower than non-plural words (β = 0.1, SE = 0.05, t = 1.77, p = .09). No interaction between plural type or any of the distributive conditions reached significance – Condition B : Plural (β = -0.04, SE = 0.09, t = -0.44, p = NS); Condition C : Plural (β = -0.08, SE = 0.09, t = -0.88, p = NS); Condition D : Plural (β = -0.02, SE = 0.09, t = -0.23, p = NS). Main effects of condition also failed to reach significance – Condition B (β = -0.0002, SE = 0.09, t = -0.003, p = NS); Condition C (β = -0.02, SE = 0.09, t = -0.203, p = NS); Condition D (β = -0.01, SE = 0.09, t = -0.128, p = NS). Age (β = 0.007, SE = 0.008, t = 0.839, p = NS);, gender (β = -0.22, SE = 0.21, t = -1.042, p = NS);, education (β = -0.08, SE = 0.10, t = -0.836, p = NS); and literacy (β = 0.08, SE = 0.09, t = 0.906, p = NS) all failed to reach significance in the model. Fig. 4 shows some interesting differences across conditions, though. Running simple effect comparisons between conditions revealed more about the marginal effect of plurality reported above. Comparing the reading times to plural and non-plural continuations following a non-distributive narrative (condition A) revealed a statistically significant difference in responses (β = 0.48, SE = .22, t = 2.231, $p < .05$). However, this was not the case across any other condition. There was no difference between readings times to plural and non-plural continuations following narratives in condition B (β = 0.108, SE = .22, t = 0.499, p = NS), condition C (β = 0.026, SE = .22, t = 0.120, p = NS), or condition D (β = 0.148, SE = .22, t = 0.686, p = NS), as seen in Fig. 4. Interestingly, no significant difference was observed between reading times to plural continuations in the control condition compared to any of the distributive conditions (B-A (β = -0.218, SE = .22, t = -1.009, p = NS), C-A (β = -0.277, SE = .22, t = -1.286, p = NS), D-A (β = -0.209, SE = .22, t = -0.972, p = NS)); likewise, statistical significance was not reached when comparing reading times to non-plural continuations in the control condition to any of the distributive conditions (B-A ($\beta = 0.155$, SE $= .22$, $t = 0.720$, $p =$ NS), C-A ($\beta = 0.177$, SE $= .22$, $t = 0.823$, $p =$ NS), D-A ($\beta = 0.123$, SE $= .22$, $t = 0.573$, $p =$ NS)). Mean reading times and standard deviations are reported in Table 4 and a fuller representation of the data via a split-violin plot can be observed in Fig. 5.

Participants were also asked to rate the naturalness of the narratives they had just read. For past perfect narratives, participants strongly favoured continuations with non-plural critical words in all conditions, rating these as higher than continuations with plural critical words. An ordinal logistic regression model showed that for condition A, the control, continuations with non-plural critical words were 6.25 times more likely to be rated higher than continuations with plural words ($p < .001$). For condition B, continuations with non-plural critical words were 2.22 times more likely to be rated higher than continuations with plural words ($p < .05$); for condition C, continuations with non-plural critical words were 4.23 times more likely to be rated higher than continuations with plural words ($p < .001$); and for condition D, continuations with non-plural critical words were 2.56 times more likely to be rated higher than continuations with plural words ($p < .05$); These findings are represented in Fig. 6.

Table 4: Mean (residualised) reading times to critical words in task sentence across condition and plural type for past perfect and present perfect timespans.

Timespan	Condition	Plural Type	Reading Time (ms)	
			Mean	Standard Deviation
Past Perfect	A	Plural	614.1	363.9
Past Perfect	B	Plural	489.8	352.2
Past Perfect	C	Plural	450.5	309.2
Past Perfect	D	Plural	472.8	308.4
Past Perfect	A	Non-Plural	377.9	267.8
Past Perfect	B	Non-Plural	465.4	318.4
Past Perfect	C	Non-Plural	462.8	306.3
Past Perfect	D	Non-Plural	442.4	338.3
Present Perfect	A	Plural	406.1	304.1
Present Perfect	B	Plural	387.7	291.6
Present Perfect	C	Plural	417.1	343.7
Present Perfect	D	Plural	434.3	387.5
Present Perfect	A	Non-Plural	620.1	421.9
Present Perfect	B	Non-Plural	633.2	347.8
Present Perfect	C	Non-Plural	650.0	374.9
Present Perfect	D	Non-Plural	709.5	527.3

Fig. 4: Plotted mean (residualised) reading times to critical words across conditions and plural type for past perfect (top) and present perfect (bottom) timespans.



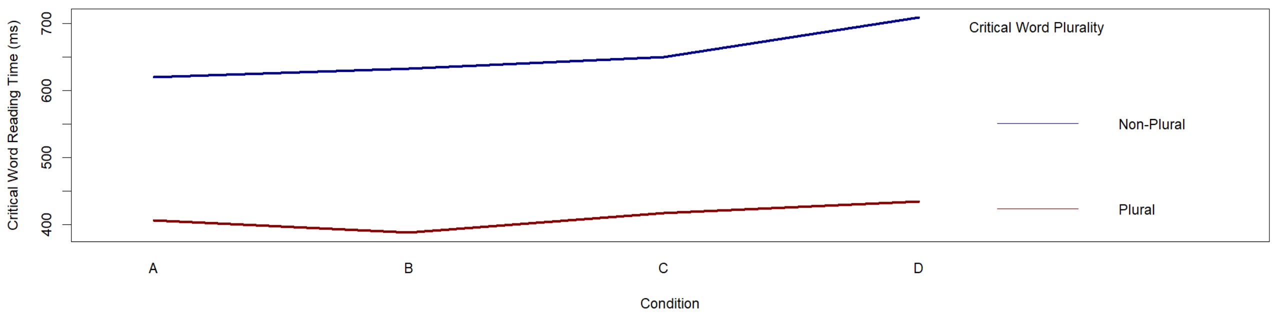


Fig. 5: Split violin plots of residualised reading times to critical words across conditions by plural type in past perfect narratives – global mean indicated by horizontal line.

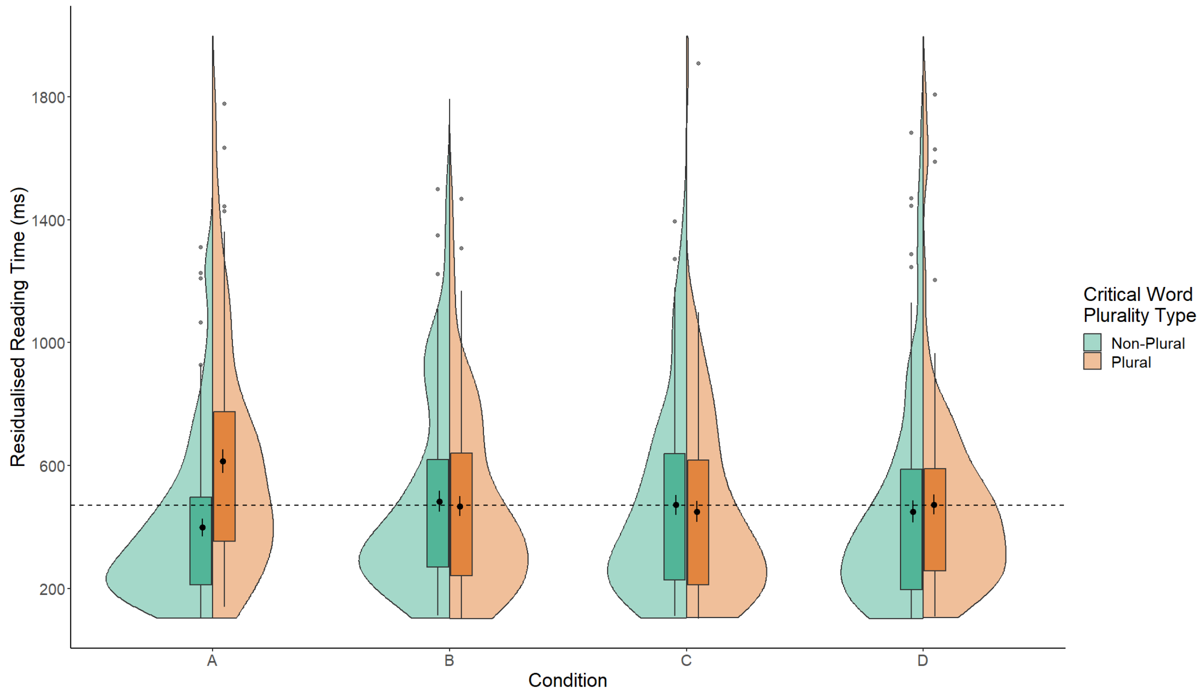


Fig. 6: Boxplots of explicit naturalness ratings (from 1-5) across conditions by plural type in past perfect narratives.

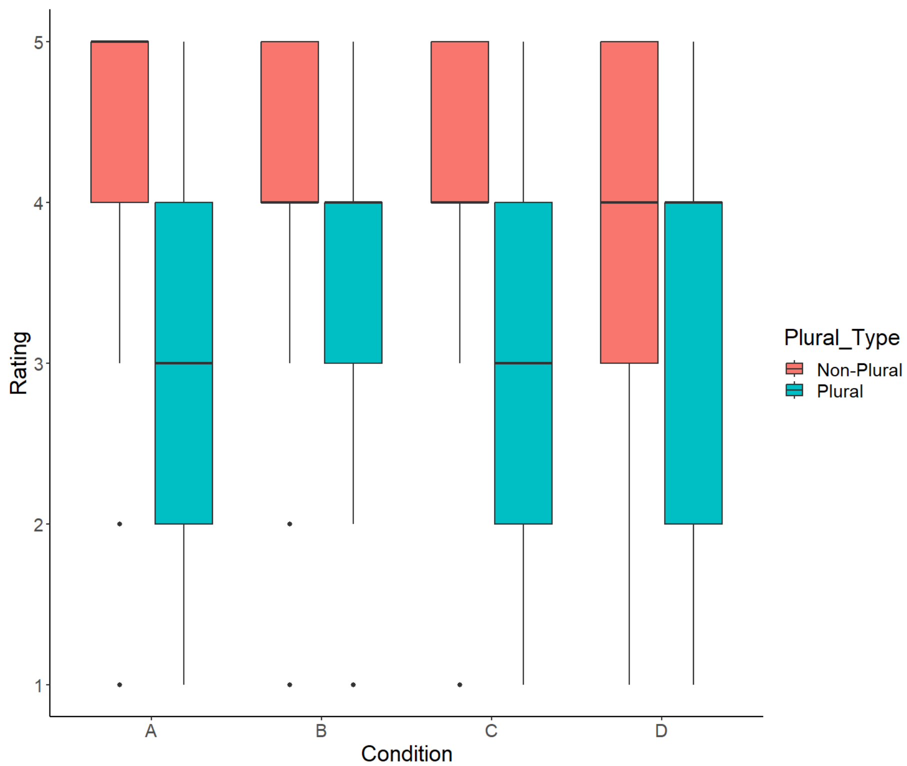


Fig. 7: Split violin plots of residualised reading times to critical words across conditions by plural type in present perfect narratives. – global mean indicated by horizontal line.

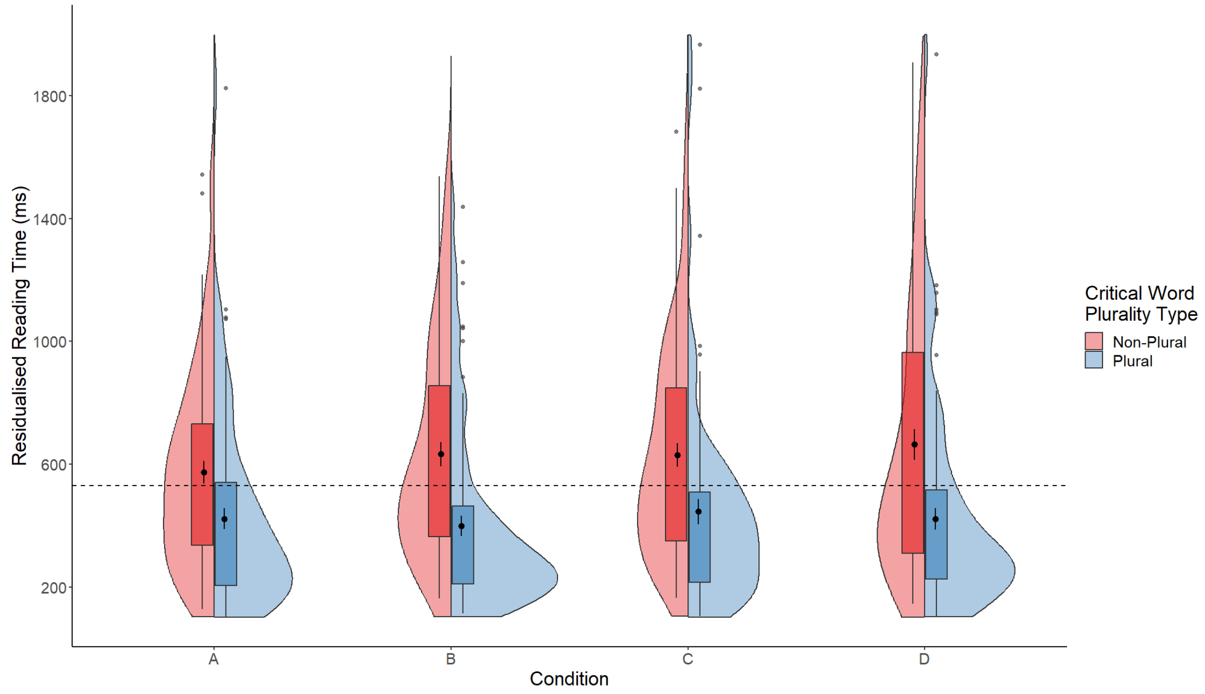
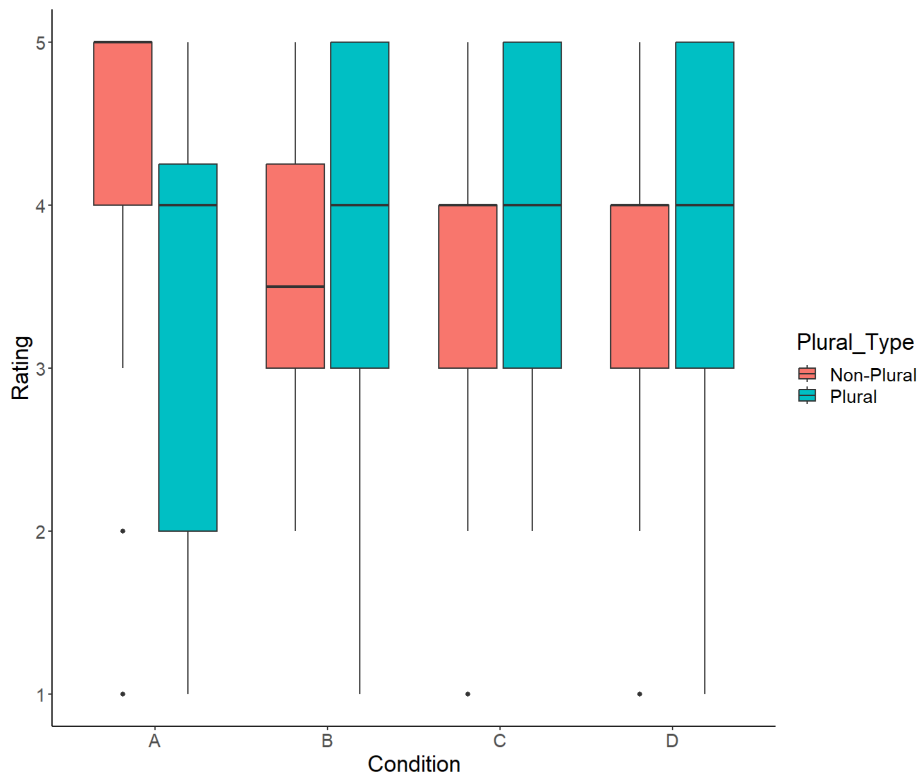


Fig. 8: Boxplots of explicit naturalness ratings (from 1-5) across conditions by plural type in present perfect narratives.



3.3.2.2 – Research Question 2

Moving onto addressing RQ2, and the set of present perfect narratives, participants in all three distributive conditions – as well as, surprisingly, the non-distributive control – exhibited faster reading times to critical words in the plural than the non-plural (mean reading times and standard deviations for each condition are again reported in Table 4). The linear mixed effects regression model with Condition and Plural Type as interaction terms showed a significant main effect of plurality, with reading times to plural continuations being read faster than to non-plural continuations (β = -0.252, SE = 0.05, t = -5.527, p < .001), as illustrated in Fig. 4 above. No interaction terms involving plurality were observed, however – Condition B : Plural (β = -0.006, SE = 0.079, t = -0.074, p = NS); Condition C : Plural (β = -0.017, SE = 0.079, t = -0.211, p = NS); Condition D : Plural (β = 0.005, SE = 0.079, t = 0.058, p = NS). There was also no main effect of condition – Condition B (β = -0.015, SE = 0.079, t = -0.186, p = NS); Condition C (β = -0.011, SE = 0.079, t = -0.144, p = NS); Condition D (β = 0.041, SE = 0.079, t = 0.526, p = NS). Effects of age (β = 0.006, SE = 0.009, t = 0.672, p = NS), gender (β = -0.12, SE = 0.24, t = -0.501, p = NS), education (β = -0.018, SE = 0.118, t = -0.152, p = NS) and literacy (β = 0.122, SE = 0.098, t = 1.248, p = NS) were all non-significant in the model.

As predicted, simple effect comparisons show that reading times to the plural were significantly faster for condition B than its counterpart preceding a non-plural continuation (β = -0.516, SE = .18, t = -2.828, p < .05); reading times to the plural were significantly faster for condition C than its counterpart (β = -0.538, SE = .18, t = -2.936, p < .05); the same is true for condition D (β = -0.495, SE = .18, t = -2.726, p < .05). Oddly, faster responses to plural continuations were also observed in the non-distributive control condition, which was marginally significant (β = -0.468, SE = .18, t = -2.564, p = .07). Given that condition A is patently non-distributive, the observation that plural continuations are read faster than non-plural continuations was entirely unexpected. Looking at individual participant means for condition A to better interpret this non-hypothesised result, it was observed that 20 of the participants responded faster to plural continuations – ranging from 39ms faster to an astonishing 883ms faster. Even if the participant who had the extreme 883ms difference was excluded from the model, plural continuations following condition A were still faster than to non-plural continuations and retained marginal significance, although to a lesser degree (β = -0.447, SE = .18, t = -2.455, p = .09). No statistical difference was observed between any of the distributive conditions and the control when looking solely at plural continuations (B-A (β = -0.026, SE = .18, t = -0.142, p = NS); C-A (β = -0.038, SE = .18, t = -0.207, p = NS); D-A (β = 0.037, SE = .18, t = 0.203, p = NS)), or between any of the distributive conditions and the control when considering only non-plural continuations (B-A (β = 0.039, SE = .18, t = 0.214, p = NS); C-A (β = 0.052, SE = .18, t = 0.285, p = NS); D-A (β = 0.086, SE = .18, t = 0.470, p = NS)). The distribution of reading times observed among participants to present perfect narratives are illustrated in Fig. 7 above, which shows much more variability in reading times to non-plural continuations than to plural ones.

As with research question 1, participants were asked to rate the naturalness of the narratives they had just read. For condition A, the control, participants were 3.55 times more likely to rate continuations with non-plural words higher than continuations with plural critical words (p < .001), which matched the rating pattern observed for the past perfect narratives in the previous section, but interestingly did not align with the unexpected reading time results observed here for the present perfect narratives, where plural continuations were read faster. Participants were 1.62 times more likely to rate the naturalness of narrative continuations with a plural critical word following sentences in condition B higher than similar continuations with a non-plural critical word, but this did not reach significance (p = 0.23). Likewise, while continuations following condition C were 1.17 times more likely to be rated higher if that continuation was plural rather than non-plural, this again did not reach significance (p = 0.71). Finally, continuations following condition D were 1.68 times more likely to be rated higher if that continuation was plural rather than non-plural, but again failed to reach significance (p = 0.20). These findings are illustrated in Fig. 8.

3.4 – Discussion

3.4.1 – Research Question 1

The initial research question of this study asked whether comprehenders commit early to a collective interpretation of the events described in an ambiguously distributive sentence, or if both distributive and collective representations remain active during processing and the construction of mental representations of the situation described. If the former were true, there should be on-line evidence that comprehenders treat ambiguously distributive texts as if only a single event occurred, while, if the latter were true, comprehenders should not only treat ambiguously distributive constructions differently from an unambiguously non-distributive control in that they should show no preference for either continuations that are consistent with a single event or continuations that are consistent with multiple events during an on-line processing task. Considering that previous insights in the literature suggested comprehenders ultimately default to a collective interpretation (Frazier et al., 1999), it was predicted that comprehenders would show an eventual preference for a collective reading during a post-trial off-line judgement task. The reading time and rating results from the past perfect narratives in the above experiment go some way to supporting these predictions, suggesting that while comprehenders do ultimately favour collective interpretations, in line with Frazier et al. (1999), they likely maintain both collective and distributive representations during on-line processing of texts.

Mean reading times to plural and non-plural continuations across the three distributive conditions were almost identical, with no statistically significant difference observed between them. This was not so for the non-distributive (control) narratives, where reading times to plural continuations were slower than to non-plural continuations (adjusted for length of word). This statistically significant difference is in line with the expectation that a non-distributive control condition should be consistent with a collective interpretation, and that an inconsistent plurality-continuation would result in additional processing costs and would be reflected in increased reading times to that word. That such an effect was only observed to the non-distributive narratives but not the three distributive narrative conditions suggests that comprehenders treat distributivity contexts differently from non-distributive ones; that is to say, comprehenders did not treat distributive contexts as being consistent with collective- or distributive-consistent continuations – both were acceptable. The type of plurality denoting expression included (numeral modifier, conjunction, or both) does not seem to affect this lack of immediate commitment to the collective. The exciting takeaway here is that ambiguously distributive constructions seem to be compatible with both collective and distributive continuations, such that during processing both representations are likely active for some time, unlike non-distributive controls, where only a collective interpretation is acceptable. However, more data needs to be collected in this area, as statistically significant differences between the distributive and control conditions were not observed in (p-value adjusted) direct comparisons. The commitment to a collective interpretation does seem to be reached eventually, however, as the results from the off-line naturalness rating task demonstrate. Ratings with non-plural continuations were significantly higher than plural continuations across all four conditions, indicating an ultimate preference for a single event interpretation, no matter the condition. These rating data demonstrate that comprehenders ultimately commit to a collective interpretation when explicitly prompted to make a judgment, in line with previous research.

As with non-event-based psycholinguistic research which established that ambiguity processing seems to proceed in parallel rather than in a serial fashion (Gibson & Pearlmutter, 2000; McClelland et al., 1989; McRae et al., 1998; Levy, 2008; Logačev & Vasishth, 2013;

Snell & Grainger, 2019; van Gompel et al., 2000) the present study demonstrates both representations of singular events and representations of multiple events are likely to be active and processed in parallel during the reading of ambiguously distributive constructions. These observations have consequences for our understanding of how comprehenders compose and integrate events with their encyclopaedic knowledge to construct mental representations of the situation described. If representations of both single and multiple events are active during the reading of event-ambiguous texts, this complicates our picture of situation model construction and demands further investigation into how comprehenders deal with representations of these multiple parallel events, including how each representation contributes to the larger mental model (or models) being constructed.

Some of our existing theories of comprehension, such as the connectionist-inspired Construction-Integration (CI) model (Kintsch, 1988; van Dijk & Kintsch, 1983; Wharton & Kintsch, 1991), propose that representations of the situation described are built bottom-up from rule-based, subconscious processes, wherein comprehenders extract linguistic representations of predicates and their arguments via the language's morphosyntax, and formulate these into a macrostructural propositional network which generates multiple interpretable meanings, with one of these eventually being committed to. The present finding that both collective and distributive interpretations are active in parallel might be explained by both representations being active at this macrostructural phase of situation model construction, after which one of the interpretations is chosen (i.e., when explicit prompted to make a judgement).

This explanation faces a critical problem, though. The mechanism by which the propositional representations are constructed in the CI model is based on argument overlap; this is not consistent with how distributive and collective meanings are thought to arise. Instead, it is understood that the event underlying the verbal semantics ranges over its arguments or atomic elements within those arguments, such that a collective reading is generated by the former while a distributive reading comes about via the latter. For an event-introducing English verb like ‘buy’, this can formally be expressed by the following formula, adapted from Champollion (2016b), and read as for all buying events *e*, that event consists of one or more buying events, *e*’, whose thematic roles (θ) are mereological atoms; that is to say, if John and Mary bought a book, the (collective) book buying event consists of (at least two) subevents which map onto different agents, one being John, the other, Mary.

Event Semantics of the Distributivity of a ‘Buying’ Event

$\forall e[\textsc{buy}(e) \rightarrow e \in \lambda e'[\textsc{buy}(e') \;\&\; \text{Atom}(\theta(e'))]]$

This crucial mereological relationship between collective events and distributive events is not captured by the CI model, or the properties of argument overlap. A model of comprehension which priorities events as part of its ontology, while also maintaining some semblance of the construction-integration process, is the Event-Indexing (EI) model (Pettijohn & Radvansky, 2016b; Zwaan, 2016; Zwaan & Radvansky, 1998; Zwaan et al., 1998). This model of event representation considers how events interact with individuals, locations, times and causations, including how comprehenders track these dimensions to update their representations in real-time (Rinck & Weber, 2003); this model maps well onto event semantic formulations, like the one above, that consider events to be arguments of not only verbs, but all predicates, including those that convey time, location and the entities involved (Maienborn, 2011). However, the EI model’s focus on top-down integration processes and representational updating means it often lacks the predictive power (of the CI model) or the logical apparatus (of event semantics) that is needed to explain how bottom-up linguistic input is translated into a mental representational format. As a result, it, too, is probably not sufficient in its current form to explain the observation that both collective and distributive readings remain accessible during language processing.

Consequently, these two influential accounts of situation model construction (CI and EI) do not seem to be explanatorily powerful enough alone to capture the present data. While the EI model is probably the (psychological) approach most consistent with event semantic formalisations of ambiguously distributive constructions, and the CI model is arguably the most precise in capturing the representational space these pluralistic interpretations might occupy, neither is sufficient in their current design to fully capture the results of this study. Expressed in another way, there are clearly aspects of event representational phenomena derived from certain constructions, like plurality denoting expressions, that are not yet adequately modelled by our cognitive theories.

3.4.2 – Research Question 2

The second research question asked whether the pattern observed for past perfect constructions would hold for present perfect constructions, where the timespan in which an event or set of events can occur (and are relevant) was extended. Given the temporal ambiguity inherent in distributive constructions, it was predicted that, for present perfect narratives, comprehenders would be more likely to accept an interpretation consistent with multiple events having occurred and would thus read plural critical words in a continuation faster than non-plural critical words. However, this was predicted to apply only in the distributive conditions, where temporal ambiguity was inherent, and not in the unambiguously single-event control condition, which was expected instead to pattern like the past perfect contrast between plural and non-plural reading times, with the mismatching plural read slower than the non-plural. A preference for continuations with a plural word in distributive narratives was therefore expected in the subsequent naturalness rating task as well, whereas the reverse was expected for non-distributive (control) conditions. The results of this study generally support these predictions, but there were some unexpected findings, chiefly relating to differences between the reading times and off-line ratings in the non-distributive condition.

As predicted, for the unambiguous control, naturalness ratings favoured continuations with a non-plural critical word, consistent with a single event interpretation. This matched the pattern observed in the set of past perfect narratives, discussed above. However, for the three distributive conditions, the naturalness ratings showed no significant difference between plural and non-plural continuations. This ran counter to the findings in the past perfect set of narratives, wherein non-plural continuations were favoured across all four conditions. These results also did not match the prediction that comprehenders would favour plural continuations in the explicit rating task following temporally ambiguous events in an extended timespan. Instead, comprehenders show no preference for either continuation in a subsequent rating task. While for past perfect narratives, comprehenders ultimately commit to a collective interpretation of the event(s) described, no matter the experimental condition, the failure to reach statistical significance when comparing ratings to distributive conditions with plural and non-plural continuations in present perfect narratives suggests that manipulations of the perfect timespan simply reduces the likelihood of comprehenders committing to that 'default' collective interpretation, rather than forcing a distributive one.

The picture painted here becomes more interesting when looking at the reading time results in present perfect narratives. Reading times to plural critical words following sentences in the three distributive conditions (B, C and D) were, as expected, faster than to non-plural words, suggesting that participants were more likely to accept continuations consistent with multiple events than with continuations that implied that just a single event had taken place. There was no substantive difference in reading times between any of the three distributive conditions to plural words, or between any of the three distributive conditions to non-plural words, suggesting – as with the past perfect set of narratives – that the type or quantity of plurality denoting expression has no impact on readers' interpretations.

As a result, while the off-line rating data did not show a clear commitment to a distributive (or collective) reading in these three conditions, the reading time data supports the conclusion that comprehenders produce stronger distributive representations than collective ones during on-line processing, treating non-plural continuations as mismatching. The combined reading time and rating data related to these distributive constructions provides evidence that the manipulation of the perfect timespan does, indeed, affect distributivity preferences and thus the number of events that are mentally represented during the on-line comprehension of a text, consistent with experimental predictions, but these preferences do not survive explicit prompting to make a judgement, wherein neither plural nor non-plural continuations are favoured. Whether the bias to default to a collective interpretation when prompted is still relevant to present perfect narratives that otherwise favour distributive readings in real-time, resulting in the null effect, is unknown.

The intriguing notion that falls out of these differences between the reading time data and rating data in the present perfect is that explicit prompts via the naturalness ratings do not necessarily match on-line processing results, and that what is immediately mentally represented during reading may not be the same as eventual commitments to an interpretation. When taken together with insights from the past perfect narratives in the previous section, there are good reasons to assume that event representations are somewhat fluid, allowing for strong distributive preferences during the unfolding of a sentence, with collective representations potentially being active in parallel; representations are then committed to by explicit prompting, wherein there seems to be a tendency toward favouring collective interpretations or (in the case of present perfect constructions) both interpretations equally. These findings emphasise the importance of continued work in the field on how linguistic structures and features, particularly those that are understudied like distributivity, affect event representations, including how these are processed, and their representations accessed, in real-time. Indeed, further research that explores the difference between off-line commitments and on-line representations of events in temporally ambiguous contexts remains vital.

These innovative results are not without their complications, though. A rather unexpected finding observed to the control condition warrants some discussion. Contrary to what was predicted – and the inverse of the respective off-line ratings – reading times to the control, which lacked any plurality denoting expression, were also faster to plural critical words than non-plural words, and this difference was marginally significant. Although an extended timespan is extant during all present perfect narratives, condition A was not temporally ambiguous; there was only one event, at one time, that took place. Why this reading time finding was observed is not immediately clear. Most of the limited available psycholinguistic research on distributivity has found that in controls or baseline trials comprehenders typically show a dispreference for distributive readings (see Frazier et al. 1999; Syrett & Musolino, 2013) – a finding which also matches general intuitions about these constructions. Notably, this finding persisted even after individual participant means were reviewed and extreme values excluded from the model. Maldonado et al. (2017) similarly and unexpectedly observed a preference for distributive interpretations in their non-distributive control trials as well, speculating that perhaps their methodology, choice of stimuli, or novel experimental design[22] could have influenced their results. What causes this bias toward distributive readings in controls both here and in Maldonado et al. (2017) is unclear, including whether they are related or not, but given that the aspectual contrast in the present study has not, to my knowledge, been applied to distributive constructions in any psycholinguistic experiment before, it is possible that some other factor in distributivity processing which is not fully understood is at play. An

[22] Across three sentence-picture matching experiments, Maldonado et al. (2017) used relations between abstracted shapes, such as 'Two [shape 1] are connected to three [shape 2]', to test priming effects among pairs of distributive constructions, extending their findings to more natural constructions like 'two boys have three balloons'. Baseline trials included the unambiguous 'A [shape 1] is connected to two [shape 2]', where only a single set of connections from shape 1 is drawn (roughly, although not exactly, analogous to the single event in condition A above).

attempt to replicate these findings is encouraged, with a particular view to exploring the circumstances in which participants, both individually and as a group, might favour distributive readings in non-distributive contexts.

Another matter worth noting in the analysis of the present perfect results is that unlike past perfect narratives where the event of the initial sentence is understood to be temporally ‘before’ the past time event of the subsequent critical sentence, this is not necessarily the case for present perfect narratives. The latter has an initial clause wherein the underlying event is understood to be relevant to the present time / time of utterance, which may extend to a temporal interval ‘after’ the past time event of the critical sentence. Consequently, the implemented present perfect narratives might have a more complex temporal relationship in their interpretation than those in the past perfect. While there is no good theoretical reason to assume this would give rise to the pattern of effects observed in the data for the present perfect condition – particularly the unexpected (marginally significant) observation relating to the control condition – a thorough analysis of this data must at least be mindful of this potential added complexity in the chronology of narratives that contain the present perfect. Future research on this topic must do the same. Looking ahead, a possible solution to the temporal complexities inherent in this paradigm would be to use future time in the critical second sentence instead, which would establish a chronological relationship between event 1 and event 2 with no potentially confounding temporal overlap, further improving the validity of the design.

Although the mystery of the reading times to the control in present perfect narratives is still outstanding, the general findings of this chapter support the existing science that suggests there is an eventual preference for a collective interpretation, while also contributing two unique findings: first, both collective and distributive interpretations appear to be active (or at least accessible) during on-line processing, with commitment to the collective coming later; second, temporally ambiguous distributive events are less likely to be interpreted in the collective when in the expanded timespan of the present perfect aspect, with, instead, a strong preference for the distributive demonstrated in the reading time data. Consequently, this research emphasises the importance of continued investigation into distributivity and its interaction with different kinds of linguistic structures, while also highlighting just how little is known about what is involved in the real-time construction of ambiguous event representations. Future research should explore further the relationship between on-line and off-line measures of distributivity, while seeking to explain how ambiguous events contribute to comprehenders' mental models.

3.5 – Conclusion

This chapter built on research that explored how comprehenders process and determine the meaning of ambiguous constructions where an event either ranges over a collective set of entities or summaries over individual references to each entity separately. Results of the study demonstrated that following past perfect constructions, a preference for a collective interpretation is ultimately reached in both ambiguous and non-ambiguous contexts, matching previous observations in the literature; however, a particularly novel finding relating to these constructions is that both collective and distributive interpretations remain active during the real-time comprehension of a text, with readers only committing to a collective interpretation when explicitly prompted. For present perfect constructions – where the timespan in which events can occur is extended – comprehenders are seen to have strong a preference for distributive interpretations during the on-line processing of ambiguous events, but this preference does not survive when explicitly prompted to make a naturalness judgement, where instead either a collective or no clear preference is reached. These findings promote new insights into how ambiguously distributive events are processed, while motivating future efforts to examine in more detail the relationship between on-line and off-line measures of distributivity and the role of event-permitting aspectual timespans on the developing situation models of comprehenders. Establishing the precise role of plurality denoting expressions on mental representations remains an ongoing challenge for our theories of language processing.

Chapter 4

Negation and Distributivity in Assumptions of Temporal Iconicity

4.1 – Background

During language processing, comprehenders hold *a priori* expectations about the order of events in a narrative, known as the iconicity assumption. Violations of these expectations result in increased processing costs and, similar to other prediction error effects in the wider literature, are reflected in ERP components time-locked to critical continuations consistent or inconsistent with these expectations. The effect of various grammatical constructions which are hypothesised to influence the mental representations of narrative events – such verbal negation which encodes something did not happen, and plurality denoting expressions that give rise to ambiguously distributive interpretations – is not well evidenced in the literature. Two EEG experiments comprising this study reveal that negation suppresses representations of events from forming part of a comprehenders' developing mental model, thus impacting expectations about the order of unfolding events, while ambiguously distributive constructions, in which multiple competing representations are active (or accessible), result in integration difficulties with comprehenders' expectations about single event to single event contiguity within a narrative.

4.1.1 – ERPs and Prediction Violations

During the processing of linguistic input, comprehenders maintain expectations about upcoming stimuli. High-probability continuations reduce on-line processing costs compared to lower-probability continuations, with these effects being observable in event-related potential components time-locked to the onset of these continuations (Kamide, 2008; Loerts et al., 2013). Initially discovered by Kutas and Hillyard (1980), who presented participants sentences like (39), marked differences in event related potentials are observed depending on whether the critical continuation is semantically consistent with the preceding context or not.

39) He buttered the bread with…

For example, a low-likelihood continuation like *socks* elicits a more negative-going event-related potential (ERP) between 300ms-500ms post-stimulus than a high-likelihood continuation like *jam*. This negative-going neuro-electrical signal, averaged across both participants and trials, is known as the N400, and its increased amplitude in the semantically inconsistent condition is assumed to reflect the additional computational resources required in processing the unexpected critical word, typically being strongest at centroparietal sites of the scalp (Kamide, 2008).

N400 effects are observed widely in the literature on prediction.[23] In Polish, where gender directly encodes animacy on both nouns and adjectives, prediction-inconsistent gender-marked adjectives elicit larger N400 effects than prediction-consistent adjectives (Szewczyk & Schriefers, 2013), while in Dutch, following narratives translated as (40) below, noun phrase continuations which contained a gender-marked adjective mismatching the expected noun's gender[24] show larger N400 effects than gender-matching adjectives, suggesting that readers use their knowledge of the wider discourse rapidly enough to anticipate specific upcoming words as the sentence is unfolding (van Berkum et al., 2005).

40) The burglar had no trouble locating the secret family safe. Of course, it was situated behind a…

Similarly, examining the indefinite article in English, which is sensitive to whether the onset of the subsequent word begins with a consonant or vowel, DeLong et al. (2005) observed that N400 effects to indefinite articles differ in a graded fashion as a function of contextual

---

[23] Prediction effects are observed in measures other than ERPs, too. For example, in a visual-world eye-tracking paradigm looking at Japanese, a verb-final language, Kamide et al. (2003) found that when the first argument *weitoresu* ('waitress') was marked with the nominative case (*-ga*) and the second argument *kyaku* ('customer') was marked with the dative case (*-ni*), comprehenders were more likely to look at a picture of a hamburger prior to the presentation of the verb compared to when *kyaku* was marked with the accusative case (*-o*).

[24] Van Berkum et al. (2005) found that 83% of people in a prior cloze procedure suggested the word *schilderij* ('painting') as the best continuation of this sentence. *Schilderij* has neuter gender with a 'zero' suffix.

constraint, such that the continuation *a kite* in (41) is more constrained (and more expected) than *an aeroplane*, and thus encountering *an* at the end of this sequence results in a larger N400 effect to the article compared to *a*.

41) The day was breezy. The boy went outside to fly…

Where the input – whether a morphological, lexical, or sentential continuation – does not match comprehenders expectations about upcoming stimuli, a prediction error occurs (Fitz & Chang, 2019). This prediction error is reflected in the N400 component widely reported in the literature, which appears in a graded modulation based on the degree to which the critical stimulus being measured is semantically consistent with the preceding context (Kuperberg, 2016; Payne et al., 2015).

Where integration of the new material is particularly difficult, however, the prediction error is sometimes followed by an attempt (typically below the level of consciousness) to reanalyse the developing mental representation of the described situation in light of the unexpected linguistic input (Kuperberg & Jaeger, 2016; Kuperberg et al., 2020). The potential reanalysis of the sentence or discourse and its underlying message-level representation is reflected via another ERP component, the P600, which is a late positive-going waveform between 500ms-900ms post-stimulus, frequently associated with syntactic parsing errors and ambiguity resolution (Bornkessel et al., 2004; Friederici et al., 2002; Frisch et al., 2002; Hagoort et al., 1993; Kaan & Swaab, 2003; Kaan et al., 2000), but has also been observed to reflect a rapid reanalysis effort to make sense of some unanticipated linguistic input and better reflect the broader statistical structure of the environment (Kuperberg & Jaeger, 2016). Thus, the P600 indexes general integration difficulty of some continuation with its context and the comprehender's attempt to resolve this, where bottom-up information – be it syntactic, semantic, morphological, etc. – is evaluated with respect to top-down expectations (Brouwer et al., 2012;

Brouwer et al., 2017; Delogu et al., 2019; DeLong et al., 2011; Otten & van Berkum, 2008; Regel et al., 2014; Tanner et al., 2017). Anterior site P600s appear to be correlated with disconfirmed prediction in congruent texts, while posterior P600s, maximally in parietal and centroparietal regions, are a result of reanalysis in less congruent texts (see van Petten and Luka (2012) for discussion). Disentangling the function of these topographically different components remains a matter of ongoing debate, however. Interestingly, where a P600 is observed following semantically anomalous or prediction-inconsistent constructions, an N400 is not always present (Hoeks et al., 2004; Kolk et al., 2003; Kuperberg et al., 2003; van Herten et al., 2005; Wang et al., 2009), suggesting these components reflect two distinct processes that only sometimes overlap.[25]

Both the N400 and P600 components are highly instructive neural responses that researchers of language can, and often do, consider when attempting to model the predictive mechanisms comprehenders rely on during linguistic processing. The present chapter, therefore, builds upon existing ERP research to test a relatively understudied expectation comprehenders hold about the order of events described within a linguistic text known as the iconicity assumption and measures the cognitive costs when this message-level expectation clashes with the number of events implied by the preceding context.

4.1.2 – The Iconicity Assumption and Prediction Violations

Events are central to comprehension. Events can be described as the changes in the temporal, causal, motivational, and locational states of objects and entities that comprise a described situation, including the relationships that hold between those objects (Altmann & Ekves, 2019; Zwaan, 2016). Typically, events are linguistically encoded by verbs and their arguments

[25] A semantic P600 without a corresponding N400 has been referred to as the 'semantic illusion' (Brouwer et al., 2017), while semantic P600s following an N400 are usually called post-N400-positivities or PNPs (van Petten & Luka, 2012).

(Rappaport Hovav & Levin, 1998; Vendler, 1957; Zwaan et al., 1995a), but it is worth noting that research into event processing routinely supports the notion that there is substantial overlap between linguistically-derived event representations (such as in a sequence like *ate an apple*) and those derived from perceptual experience (such as visually observing a person eating food served to them by a waiter), such that the fundamental mechanisms of discriminating one event from another or segmenting some event into smaller sub-events is quite possibly the same whether the input is via language or sensorimotor stimuli (see Altmann & Ekves, 2019; Bailey et al., 2017; Folli & Harley, 2006; Kurby & Zacks, 2012; Metusalem et al., 2012; Ozyurek et al., 2015; Sakarias & Flecken, 2019; Zacks & Tversky, 2001; Zacks et al., 2001a; Zacks et al., 2001b; Zacks et al., 2009; *inter alia*). As a result, insights into sensorimotor experience often translate well when analysing event representations in language.

One such example is the iconicity assumption. Duran et al. (2007) describe how our real-life experience of the chronological and continuous passage of time is mapped onto comprehenders' expectations about the way events described within some narrative text will unfold (hence the term iconic). Comprehenders appear to prefer narrative continuations that introduce new events which are temporally contiguous with the event introduced in the preceding sentence, such that continuations, unless elaborative in nature, are expected to follow a strict chronology of some Event X followed by some Event Y followed by some Event Z, with minimal gaps in time between these described events (Briner et al., 2012; Dery & Koenig, 2015; de Vega et al., 2004; Dowty, 1986; He et al., 2015; Kaiser, 2019; Magliano & Schleich, 2000; Nuthmann & van der Meer, 2005; Therriault et al., 2006; van der Meer et al., 2002; Xu & Kwok, 2019; Yevseyev, 2012; Zarcone & Demberg, 2021; Zwaan, 2016; Zwaan & Radvansky, 1998; Zwaan et al., 1998). Deviations from this expected ordering result in increased processing costs (Duran et al., 2007). For example, Zwaan (1996) demonstrated that even the chronological distance between the events encoded in two consecutive clauses is sensitive to comprehenders' expectations about how they will unfold. By looking at examples like (42) and (43), Zwaan illustrated how non-contiguous adverbials such as *an hour later* in (43) demand additional processing resources compared to contiguous adverbials like *at that moment* in (42), manifesting in increased reading times.

42) The professor started analysing the data. At that moment, her phone rang.

43) The professor started analysing the data. An hour later, her phone rang.

Mismatches between the narrated and expected order of events, whether these include temporal adverbials or not, lead to elevated processing costs (Mander, 1986), decrements in accuracy (Ohtsuka & Brewer, 1992), and increased discourse-level computations reflected in marked negativity in ERPs as early as 300ms post-stimulus to non-contiguous, iconicity-inconsistent continuations (Münte et al., 1998). Other work provides evidence that violations of iconicity actually pattern like the prediction errors discussed above in the ERP waveform, producing graded N400 effects based on the degree of iconicity violation (O'Gara & Lawyer, 2019).[26] Given that the iconicity assumption and comprehenders' default expectations about event ordering are essentially *a priori* predictions, and mismatches of the anticipated ordering result in additional processing costs, it is no surprise that prediction error-like effects should be observed in the ERP waveform.

Surprisingly, research into the iconicity assumption has broadly examined only simple situations in which a single event leads into another event, which may match or mismatch in contiguity or chronology, as in (42) and (43) above; to my knowledge, no study has attempted to examine what happens to comprehenders' expectations about event ordering when a more

[26] Non-linguistic event chronology is also detectable in other ERP components; Baetens et al. (2014) show that mismatches between the content of some visually presented scene (i.e., line drawings illustrating some social script event sequence, like a proposal followed by a wedding) and its expected chronology result in an observable left anterior negativity (LAN) typically associated with violations of syntactic expectations of things like word order. Elsewhere, Holle and Gunter (2007) show how non-linguistic iconicity-mismatches result in predictable N400 effects; for example, where an iconic gesture mismatched the linguistic description provided alongside it.

complex sequence of events is presented that include otherwise fairly common linguistic features which give rise to varied event representations – those of verbal negation and plurality denoting expressions. The negation of an event (e.g., *John didn't buy a novel*) results in that event not happening, and, if an event ranges over a plurality denoting expression (e.g., *John and Mary bought a novel*), it is ambiguous as to whether a single event or multiple events took place.

In other words, in some sequence of contiguous events, say event X → Y → Z, if event Y is negated, and therefore did not happen, the contiguity of the narrative is broken – there is a temporal 'break' between events X and Z – and thus temporal iconicity should be violated. Similar to other contiguity violations within the literature on the iconicity assumption, integration difficulties via a prediction error-like effect should occur. If measured using ERPs, when compared to a non-negated control sequence, it is hypothesised that a centroparietal N400 component, reflecting the prediction error (Kuperberg, 2016), will be observed to a critical continuation that establishes event Z as being contiguous with the (negated) event Y. It is possible that a subsequent P600 reflecting reanalysis processes may also be present following negated contexts, but as not all N400 effects occur with a post-N400-positivity this latter postulation is merely exploratory.

The matter of plurality denoting expressions within temporally iconic narratives is predicted to function differently. In some sequence of contiguous events, X → Y → Z, if event Y ranges over a plurality of individuals, such that it is ambiguous as to whether there was a single collective event or a distributive one in which multiple discrete 'event Ys' took place, it is not expected that a prediction error-like effect will occur, typically reflected in an N400, as a sequence of contiguous events remains intact; instead, it is hypothesised that comprehenders will hold the expectation that the sequence of described events proceeds iconically from a single event Y to a single event Z, and that, when presented with an ambiguously distributive event Y, they should experience prediction-based integration difficulties with their developing message-level representation as a result of the ambiguously multiple Y event. Resultantly, when compared to a disambiguated, non-distributive control, a P600 component, likely with an anterior topography reflecting prediction-based integration difficulties within congruent texts (van Petten & Luka, 2012), should be observed to critical continuations that establish event Z as being contiguous with the (ambiguously numerous) event Y(s).

Accordingly, this chapter consists of two electroencephalography (EEG) experiments, each examining short narratives with one of the two linguistic features mentioned – verbal negation and plurality denoting expressions – and measures averaged ERPs to contiguous narrative continuations of the manipulated events.

For negation, in Experiment 3, consideration is given to the role of both explicit and elided clauses on negated events, and their effect on iconicity violations. Given a relative dearth of research on the processing of verbal negation (see chapter 2 for further discussion), it is unknown whether negation in syntactically explicit clauses (e.g., *John didn't buy a novel*) and negation in elided clauses (e.g., *Mary didn't either*) differentially affects comprehenders' default expectations about event ordering within the narrative, although there is some suggestion from response time data to a probe word task that events encoded in explicit and elided clauses are, on some level, differentially represented in memory (again, see chapter 2 for discussion).

By adopting such an approach, the sequence of contiguous events, X → Y → Z, will thus have event Y corresponding to a sentence that consists of a conjoined explicit and elided clause (e.g., *While there, John bought a novel and Mary did too*). In one experimental condition, both clauses are negated, and as such no Y event(s) corresponding to that sentence will have a coherent representation. Consequently, an N400 to a critical continuation in the narrative that implies contiguity with the negated event is expected. Comparisons are made to a condition in which only the explicit clause is negated, and one in which only the elided clause is negated, to test whether iconicity assumption violations are observed in all three. While previous probe word research has shown differential effects between elided and explicit clauses (see chapter 2), access to probes in negated contexts and the impact of those negated contexts on expectations of event contiguity are not guaranteed to be the same. As a result, while an N400 reflecting iconicity violations is expected to the 'both' condition, hypotheses are less motivated and more exploratory regarding potential prediction error effects in the other negated conditions. A condition where neither clause is negated, and thus no iconicity violation is expected, is implemented as a control.

Finally, for plurality denoting expressions, in Experiment 4, comparisons between an ambiguously distributive event and both overt distributive and overt collective events (i.e., events that are modified by the disambiguating adverbs *together* and *separately*, respectively) are also made. Comprehenders do seem to eventually prefer a collective interpretation of ambiguously distributive constructions over a distributive one, as shown in Frazier et al. (1999), but there are some compelling reasons to think that both interpretations are active for some time during online processing (see chapter 3 for discussion), such that comparisons with both overt distributive and overt collective constructions here is warranted.

4.2 – Experiment 3: Methods

4.2.1 – Participants

A total of 20 participants were recruited to take part in this experiment. All participants were members of the University of Essex community and were currently working towards either an undergraduate or postgraduate degree. All participants were native English speakers who were more than 18-years old. Participants were a mix of genders, with ages ranging from 18-49. Monetary compensation was provided upon completion of the experiment. Ethical approval was obtained for this study from the University of Essex Ethics Committee.

4.2.2 – Materials

Participants were presented with 64 three-sentence narratives in a reading task; the first two sentences were presented in full, with the third sentence being presented in a serial visual presentation (SVP) format. Each three-sentence narrative told a short story about the characters John and Mary, alongside a friend of theirs, who was introduced in the final sentence and whose name changed across each trial. The situation described in each narrative formed a temporally iconic contiguous series of events, with event X being introduced in sentence one, event Y in sentence 2, and event Z in sentence 3; this contiguity was imposed using key temporal adverbials in each sentence (i.e., *at the same time*, *while there*, and *a few moments/minutes later*, respectively).[27] The second sentence of each narrative included an explicit clause conjoined with an elided clause, such as *While there, John bought a novel and Mary did too*.

[27] While, technically, sentence two of each narrative consisted of two clauses, and sentence three consisted of a conjoined VP with a single, shared subject, both have their elements connected via conjunction, which, in English, results in a temporal interpretation where (absent of any iconicity violating temporal adverbials or tense/aspect manipulations that shift the temporal frame in which the events described are understood to have taken place) the second conjunct is considered concurrent with – or, at least, contiguous with – the first (see Bar-Lev & Palacas (1980) and their Semantic Command Constraint (SCC)). For simplicity, sentences 2 and 3 are considered to encode Events Y and Z, respectively, but a slightly more precise analysis could be one in which there is some event Ya concurrent or contiguous with some event Yb, followed by some event Za concurrent or contiguous with some event Zb. To avoid unnecessarily complications in the analysis, a simple X → Y → Z model is adopted.

This sentence appeared in one of four conditions that manipulated the type and degree of negation in that sentence, matching the four-way design of the conditions in chapter 2. In condition A (affirmative-affirmative), both explicit and elided clauses were not negated, and acted as the control. In condition B (affirmative-negative), the elided clause was negated, as in *While there*, *John bought a novel, but Mary didn't*. In condition C (negative-affirmative), the explicit clause was negated, as in *While there, John didn't buy a novel, but Mary did*. Finally, in condition D (negative-negative), both clauses were negated. Examples of the full three-sentence narratives can be seen in Table 5. EEG waveforms were measured to the word *later* in the contiguity-defining temporal adverbial *a few moments later* in the final sentence; this measurement follows the manipulation of the second narrative sentence to examine the effect of negated event contexts on the comprehenders' iconicity assumption.

Averaged ERPs to a narrative final word which consisted of either a predictable, semantically good continuation in 50% of trials (e.g., greeted Mary with a *hug*), or non-predictable, semantically odd continuation in 50% of trials (e.g., greeted Mary with a *spatula*) were also obtained. A comprehension question about some aspect of the narrative (such as, how did John enter the shop, who their friend was, how they were greeted by their friend, or who bought what) was presented at the end of each trial. Accuracy of answers to the comprehension questions and ERPs to the narrative-final word were used to ensure participants were both paying attention to the text and – given the novelty of the present study's design around the manipulation of negation and its effect on the iconicity assumption – to ensure semantically induced ERP components typical of unexpected continuations (like the N400) were being accurately measured during this experiment. (See *Appendix C* for a full list of experimental stimuli.)

Table 5: List of example stimuli, coded for condition, with ERP time-locked words underlined; good and bad continuations were equally distributed across conditions.

Condition	Sentence 1 (Event X)	Sentence 2 (Event Y)	Sentence 3 (Event Z)	Continuation
A	John hopped gleefully into his local music shop at the same time as Mary.	While there, John bought a record and Mary did too.	Beth arrived a few moments later and greeted John with a hug.	Good
A	John stepped confidently into his local book shop at the same time as Mary.	While there, John bought a novel and Mary did too.	Sally arrived a few moments later and greeted John with a spatula.	Bad
B	John traipsed absently into his local music shop at the same time as Mary.	While there, John bought a record but Mary didn't.	Becky arrived a few moments later and greeted John with a smile.	Good
B	John trotted excitedly into his local book shop at the same time as Mary.	While there, John bought a novel but Mary didn't.	Joan arrived a few moments later and greeted John with a trampoline.	Bad
C	John charged quickly into his local music shop at the same time as Mary.	While there, John didn't buy a record but Mary did.	Emma arrived a few moments later and greeted John with a hug.	Good
C	John skipped happily into his local book shop at the same time as Mary.	While there, John didn't buy a novel but Mary did.	Jess arrived a few moments later and greeted John with a spatula.	Bad
D	John paced slowly into his local music shop at the same time as Mary.	While there, John didn't buy a record and Mary didn't either.	Carol arrived a few moments later and greeted John with a smile.	Good
D	John walked nervously into his local book shop at the same time as Mary.	While there, John didn't buy a novel and Mary didn't either.	Gina arrived a few moments later and greeted John with a trampoline.	Bad

4.2.3 – Procedure

The 64 three-sentence narratives were presented to participants while they sat 100cm away from a 25” Dell monitor in a dark room. Sentences 1 and 2 were presented to participants in full and remained on screen until the participant pressed ‘space’ on the keyboard in front of them to proceed. Sentence 3 was presented in a serial visual presentation (SVP) format for a random integer between 500ms and 700ms (SVP format typically ranges from 500ms-1000ms (Kliegl et al., 2012), but, given the length of each trial, a shorter duration was adopted for this experiment). Materials were designed on the OSF program OpenSesame 3.0 (Mathôt & March, 2022; Mathôt et al., 2012). Trials were pseudorandomised and split into four blocks of 16 trials, with a short break offered after each block, with all participants seeing all trials[28] A practice block of 4 trials, which did not replicate any sentences from the experimental stimuli, was also included at the start. After each narrative, participants were asked to respond to a comprehension question about the narrative by pressing either the ‘a’ or ‘l’ key on their keyboard. These keys were coded to mean ‘yes’ or ‘no’ prior to the start of each experiment and were counterbalanced across participants.

While participants read these sentences, EEG was recorded using BioSemi (ActiView version 7.07, September 16, 2016) from 64 sites in line with the international 10-20 system. Two additional electrodes were positioned over the left and right mastoids for contralateral referencing of all EEG electrodes. A further three external electrodes were situated around the right eye to detect blinking artefacts for later removal. ‘Ground’ electrodes consisted of the

[28] As this study used a repeated-measures design, each (narrative) item selected for the experimental trials could have an influence on the results, such as a familiarity effect from participants seeing the same item across different conditions. Standardly, a Latin Square counterbalancing system of lists is used to create a within-item design to control for this potential problem. Due to experimenter oversight, this practice was not implemented in this study. However, steps were taken to maximise variation of character names and other elements in each narrative item, such that no three-sentence narrative was actually identical, even outside of the critical manipulations of negation, minimising any potential confound from this oversight. Similar considerations were made in the design of Experiment 4, discussed in section 4.4 below.

Driven Right Leg (DRL) passive electrode and the Common Mode Sense (CMS) active electrode. EEG responses were recorded to the critical word *later* in the contiguity-defining temporal adverbial introduced in sentence 3, as well as to the final word of sentence 3. EEG triggers for each of the four conditions at the critical word *later*, as well as for both good and bad continuations at the final word, were coded in the PySerial Python library for serial port communications on OpenSesame.

4.3 – Experiment 3: Results

4.3.1 – Analysis

Of the 20 participants recruited, two showed significant electrode connection problems that could not be resolved during breaks leading to data collection being abandoned early. Furthermore, a 75% threshold for accuracy across responses to comprehension questions was used to exclude participants who were not attending to the task, although no participant fell below this level, with all achieving at least 85% accuracy.

Raw EEG data obtained from the remaining participants was analysed using the EEGLAB toolbox v2023.0 (Delorme, 2023; Delorme & Makeig, 2004) and ERPLAB toolbox v9.20 (Lopez-Calderon & Luck, 2014) within MathWorks MATLAB (R2023a). In line with existing literature (Acunzo et al., 2012; Luck, 2005a), to eliminate slowly varying responses in the dataset, the sampling rate was reduced to an analysable 256Hz and each participant's data was filtered with an IIR Butterworth high-pass of 0.1Hz and low-pass of 30Hz; DC bias was also removed. Following Independent Component Analysis (ICA), blinking artefacts were eliminated using a combination of the ICLabel plugin and visual identification of blink components in the output. Data was referenced to the external mastoid electrodes and bins were assigned based on the EEG triggers set around the sentence-three critical adverb *later* in each condition, as well as the narrative-final word in each trial. Bin-based epochs were then established with a timeframe of -200ms to 1000ms around the onset of the stimulus related to each trigger. Artefact rejection of the epoched data using a moving peak-to-peak voltage threshold of 120Hz within a 50ms window step was conducted for each participant. Two further participants were removed from the final analysis (final $n = 16$) due to a large proportion of trials rejected (>25%).

As the previously discussed literature demonstrated (see Kamide, 2008; van Petten & Luka, 2012), hypothesised N400 components following conditions which violate semantic expectations (such as temporal iconicity) and potential post-N400-positivities reflecting reanalysis and integration difficulties of less congruent texts in the form of semantic P600s are likely to be maximal in centroparietal regions. Consequently, the following electrode locations for statistical analysis were used: CPz, CP1, CP2, CP3, CP4, Pz, P1, P2, P3, P4, Cz, C1, C2, C3, C4. The N400 analysis window was treated as 300ms-500ms post-stimulus (see Šoškić et al., 2022), with the P600 window treated as 500ms-900ms post-stimulus (Kaan, 2023). Latency effects of the N400 were also considered across the conditions, and an extended N400 window of 300ms-600ms was also analysed (see Holcomb (1993) for examples of non-age-related N400 latency windows) to accommodate for observed latencies in the negative-going components. For the narrative-final continuation, both a 300ms-500ms N400 window and 500ms-900ms P600 window were analysed at the same electrode sites. Averaged ERP data across trials and participants were plotted alongside difference waves[29], and individual participant ERPs were collated within the above time windows and statistically analysed on RStudio (R version 4.3.1) using the *lmer* function (lme4 package), with mean amplitude as the dependent variable and condition as the predictor; age and gender were included as fixed

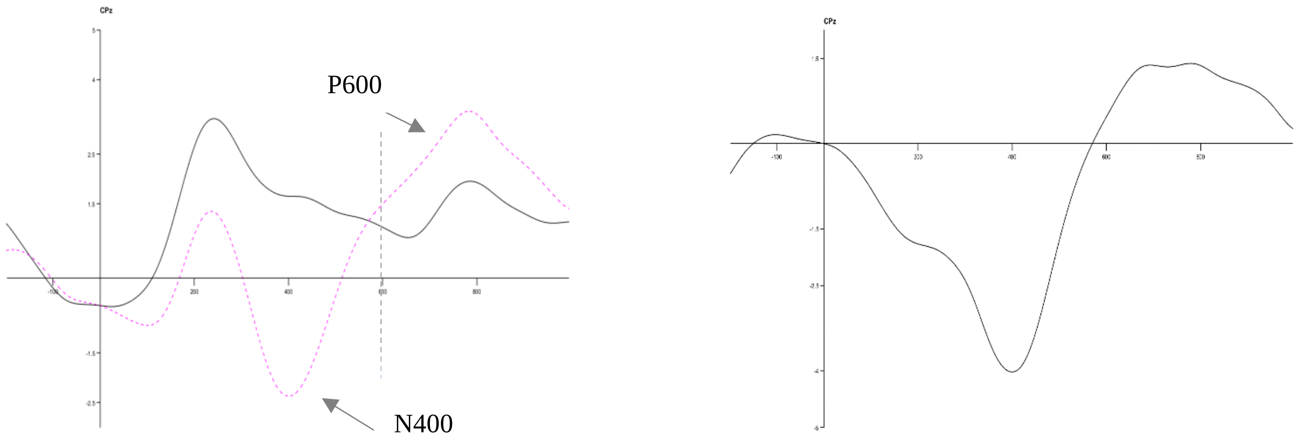
[29] Difference waves show voltage difference between two conditions at each time point across ERP waveforms. Each condition's ERP waveform is a summed average of the underlying source waveforms in the raw data. With difference waves, underlying source waveforms that are similar are eliminated while those with differences are subtracted from each other and then plotted to show averaged voltage difference between experimental conditions. Difference waves presented in this chapter were generated using the in-built function on ERPLAB toolbox.

effects within the models, while participant was included as a random effect (lmer(Amplitude ~ Condition + Age + Sex + (1 | PPT). Electrode was not included as a random effect in the final analysis due to issues of singular fit. As the fixed factor of Condition was categorical, this factor was numerised using a deviation coding scheme prior to running the model. Planned simple effect contrasts were run from this model using the *emmeans* and contrast functions (emmeans package), comparing each of the three experimental conditions to the control. P-values were adjusted using Bonferroni's correction.

4.3.2 – Findings

Semantically 'bad' narrative-final continuations which lacked predictability showed a statistically significant difference in the 300ms-500ms post-stimulus waveform characteristic of the N400 component when compared to predictable, semantically 'good' continuations (β = -3.2034, t = -9.261, p < .001), demonstrating that semantic effects, such as predictability, were observable using in the task adopted, thus increasing confidence that ERPs observed to the critical manipulations of negation on the iconicity assumption were likely to be valid. Amplitudinal differences between bad and good narrative-final continuations in the 500ms-900ms P600 time window were marginally significant (β = .4652, t = 1.811, p = .07), such that a post-N400-positivity reflecting reanalysis was potentially in evidence (see general discussion below). These findings are illustrated in Fig. 9.

Fig. 9: Averaged ERPs at CPz node to bad narrative-final continuations compared to good narrative-final continuations (left), with difference wave (right). Mean critical word offset represented by dashed vline (601ms).

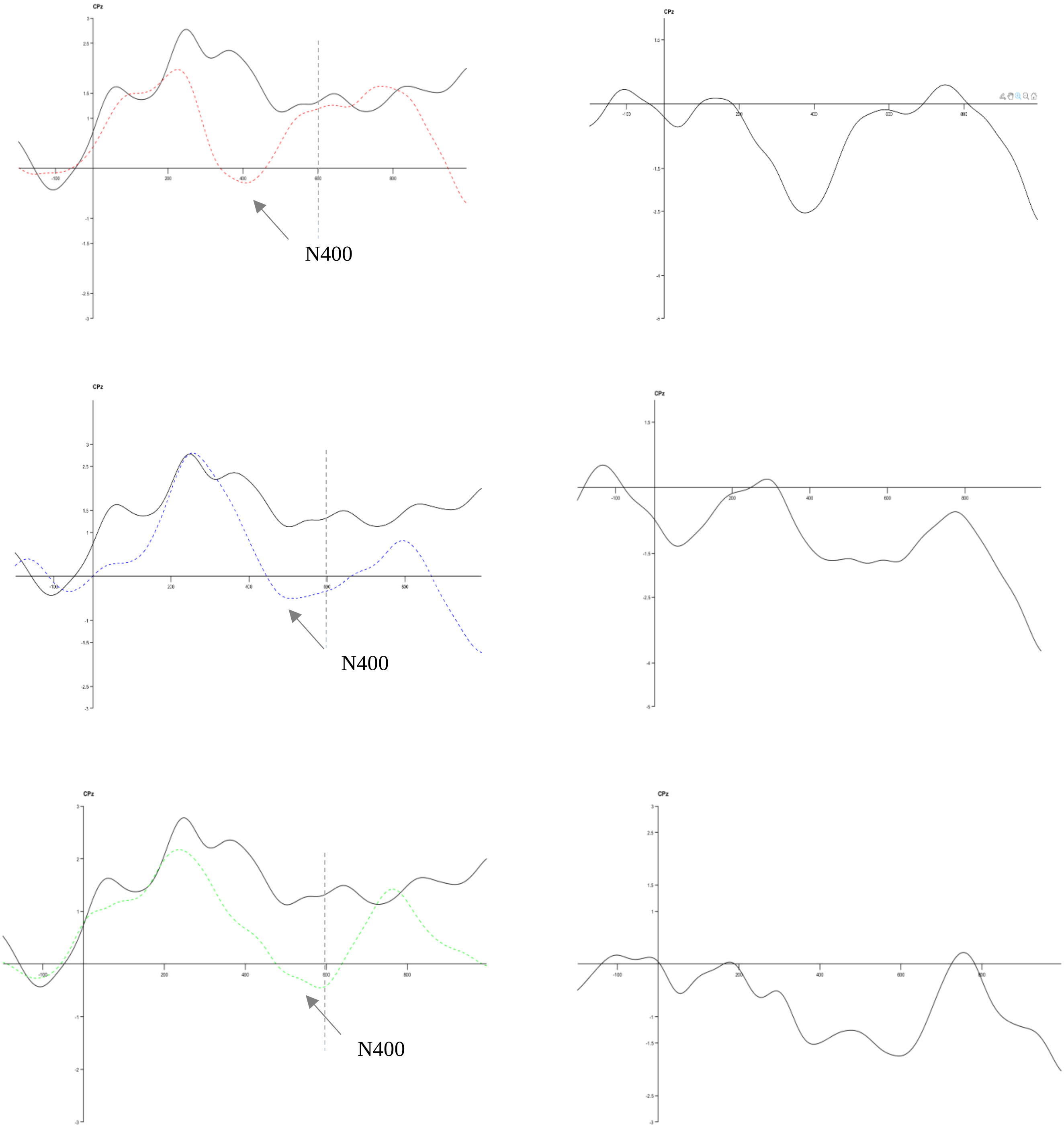


Turning to the role of event negation on comprehenders' expectations about temporal iconicity, when considering the 300ms-500ms time window for the N400 typical of prediction errors, the deviation coded linear mixed effects model showed that condition B was significantly different from the grand mean (β = -1.092, SE = 0.27, t = -4.121, p < .001). However, conditions C and D were not significantly different from that mean (β = -0.006, SE = 0.27, t = -0.022, p = NS; β = 0.137, SE = 0.27, t = 0.518, p = NS). Neither age (β = -0.016, SE = 0.1, t = -0.162, p = *NS*) nor sex (β = 2.257, SE = 2.02, t = 1.118, p = NS) was significant in the model. Planned contrasts comparing each experimental condition with the control show that the average peak amplitude of condition B was significantly different (β = -2.053, SE = 0.43, t = -4.743, p < .001). However, condition C was only marginally significant (β = -0.967, SE = 0.43, t = -2.223, p = .077). and condition D failed to reach significance (β = -0.824, SE = 0.43, t = -1.903, p = *.172*). These findings are illustrated in Fig. 10.

Considering the averaged waves in Fig. 10, it appears to be the case, at least for the CPz node, that the negative-going waves that defined the pattern observed for conditions C and D, where the explicit clause was negated, possess a peak morphology that differs somewhat from condition B, where only the elided clause was negated. This was confirmed by examining peak times across the three negated conditions within the defined centroparietal electrode sites. While condition B had an average peak of 448ms post-stimulus, peaks in condition C were produced significantly later than to condition B ($\beta = 40.174$, $t = 4.815$, $p < .001$); similarly, peaks in condition D were produced significantly later than to condition B ($\beta = 30.278$, $t = 3.629$, $p < .001$); on the other hand, there was no difference in peak times between conditions C and D ($\beta = 9.9$, $t = 1.186$, p = NS). Therefore, considering the possibility that the hypothesised prediction error-like effects in conditions C and D, with explicit clause negation, had a different latency (to be discussed below), an extended time window of 300ms-600ms was used in an additional analysis of these data. These findings show that while condition C and D are still not significantly different from the grand mean ($\beta = -0.135$, SE = 0.26, $t = -0.522$, p = NS; $\beta = -0.135$, SE = 0.26, $t = -0.522$, p = NS), planned contrasts between conditions C and A ($\beta = -1.01$, SE = 0.42, $t = -2.408$, $p < .05$) and D and A ($\beta = -1.05$, SE = 0.42, $t = -2.499$, $p < .05$) do reveal significant effects of experimental manipulation. Condition B remains significant in the main model ($\beta = -0.571$, SE = 0.26, $t = -2.215$, $p < .05$), and in simple effect comparison with condition A ($\beta = -1.55$, SE = 0.42, $t = -3.444$, $p < .05$). Neither age ($\beta = -0.009$, SE = 0.09, $t = -0.104$, p = NS) nor sex ($\beta = 1.755$, SE = 1.90, $t = 0.925$, p = NS) were significant in this extended window model.

Finally, a 500ms-900ms time window, characteristic of a putative P600, was analysed across the four conditions to determine if any reanalysis effect was caused by the experimental manipulations. However, none of the three negated conditions resulted in significant reanalysis effects. There was a pronounced negativity observable in each of the three negated conditions after around 800ms, as illustrated in Fig. 10, but as the bin-based epoch for analysis was defined as -200ms-1000ms around the stimulus, this pattern largely fell outside the scope of the study.

Fig. 10: Averaged ERPs at CPz node to condition B compared to condition A (top), condition C compared to condition A (middle), and condition D compared to condition A (bottom), with difference waves included on the right. Mean critical word offsets represented by dashed vline (601ms).



4.3.3 – Interim Discussion

The iconicity assumption that comprehenders hold regarding the contiguous ordering of events in a narrative is, arguably, an *a priori* prediction. Violations of expectations about upcoming linguistic input, whether at a morphological, lexical, or syntactic level, or indeed at a contextually determined discourse level, show robust negative-going ERP effects roughly 300ms-500ms after the presentation of the offending or anomalous stimulus (DeLong et al., 2005; Fitz & Chang, 2019; Kuperberg, 2016; Nieuwland & van Berkum, 2006; Szewczyk & Schriefers, 2013; van Berkum et al., 2005). This effect, known as the N400, was hypothesised to be observed to violations of the contiguity of some sequence of events (Event X → Event Y → Event Z) in the experimental narratives presented in the current study. Suppression of the intermediate event (Event Y) of the narrative, via verbal negation, was hypothesised to result in a violation of temporal iconicity, particularly in condition D, as the sequence of contiguous events would be broken, and an N400 effect would be observed reflecting integration difficulties and a possible prediction error-like effect associated with that violation. Comparisons between verbal negation inside an explicit clause and verbal negation inside an elided clause were also examined, as was the effect of negation on the iconicity assumption when both types of clauses were negated.

When compared to a non-negated control condition, an N400 effect was observed to all three negated conditions, providing evidence that the experimental manipulations were responsible for violations of comprehenders' expectations about the contiguous ordering of events in the narratives, and demonstrating that verbal negation suppresses the event encoded by that verb and its arguments from forming part of the larger developing mental representation of the situation described, whether in elided or explicit clauses, or both.

Curiously, the peak morphology of the N400 components observed across the conditions differed in their latency. While in the condition where the elided clause alone was negated (condition B) the averaged N400 peaked around 450ms post-stimulus, the N400 in the other two negated conditions – in which the explicit clause was negated – was characterised by a peak around 480ms-490ms post-stimulus, with no statistical difference observed between the condition in which only the explicit clause was negated (condition C) and the condition in which both clauses were negated (condition D). Therefore, it can be said that no gradation of effects with increased negation in the critical sentence were observed.

Typically, N400 effects are observed within the 300ms-500ms post-stimulus time window, and, outside of age-effects, have been argued to be quite robust across experiments in falling within this period (Kutas & Iragui, 1998; Kutas & Federmeier, 2009; Kutas & Federmeier, 2011). However, some researchers, such as Holcomb (1993), Berger and Coch (2010), Zhao et al. (2017) and Lehtonen et al. (2012), have observed N400 latency effects to manipulations like distorted linguistic input, 'internet' or 'text' language, and second language processing. Others, looking at semantic processing and repetition, have suggested that N400 latency may be less stable than originally assumed (see Renoult et al., 2012). The statistically significant latency effect observed in the present study to conditions C and D, while still falling within the 300ms-500ms time window, prompted an additional analysis which adopted a slightly extended window of 300ms-600ms, in line with the analysis in Holcomb (1993), to better capture the difference in the overall shape of the negative-going components in these conditions. This subsequent analysis returned improved statistical significance for the two conditions in which the explicit clause was negated.

These latency observations allude to the possibility that, while both the negation of elided clauses and the negation of explicit clauses violate comprehenders' expectations about temporal iconicity, they may do so differentially. Caution should be taken, however, to not overinterpret these results, as this latency effect was not hypothesised, nor theoretically motivated. While the apparent latency of the N400 in conditions C and D compared to B encouraged an extended time window for analysis, treating that same latency as evidence of some underlying representational difference might be a stretch. Moreover, ERP peaks and components like the N400 are not the same thing, and differences in peak latency do not necessarily correspond to differences in the underlying component (Luck, 2005b). Nevertheless, these observations are intriguing and promotes further inquiry. Admittedly, some possible explanations for the observed latency might simply be an artefact of limited statistical power, or even the design of the study in which sentence 2 of the narrative consists of both an explicit and elided clause. Perhaps repeating the study while changing the second sentence to contain only a single explicit clause or a single elided clause would help address this potential confound. Replication and expansion of this research is, therefore, desirable with an eye to examine the factors that could contribute to this significant latency effect.

Finally, while some N400 components to prediction errors in the literature, including, arguably, the narrative-final continuation in this study (see general discussion), show post-N400-positivities in the 500ms-900ms post-stimulus time window, no P600 component, reflecting reanalysis of the prior linguistic input, was observed to the critical word *later* in any of the negated conditions in this experiment. Collectively, the N400 and lack of P600 suggest that while verbal negation suppresses events from contributing to the unfolding situation model, resulting in a violation of the comprehenders' iconicity assumption of contiguous event ordering, no reparsing or reappraisal of the preceding narrative took place, at least within the hypothesised time window to the critical word *later*. Interestingly, a pronounced negative-going wave with an onset roughly 800ms was observed in this time window to each of the three negated conditions when compared to the non-negated control. However, this potential component largely fell outside the established epoch of analysis (-200ms to 1000ms), as well as the wider theoretical scope of the study, and so limited commentary will be made here (although see general discussion for a potential, if speculative, interpretation of this wave in light of other findings from this chapter). This observation, nonetheless, inspires further consideration of late negative deflections in cases like this in future EEG work on the iconicity assumption and its interaction with negation and the suppression of events and their representations.

The findings of Experiment 3 reveal that integration difficulties associated with violations of the iconicity assumption pattern like prediction errors in the ERP waveform. Given that this assumption is likely an *a priori* expectation that comprehenders hold about the chronological and contiguous ordering of events in a narrative, there are good theoretical reasons for thinking that predictions about the likelihood of certain continuations are being made and then either confirmed or disconfirmed during the time-course of reading. However, caution should be applied in making too strong a claim here. While the effects are clearly similar to prediction errors in the wider literature, there is one important aspect of prediction-based research that is commonly used but not replicated in this study – that of cloze probability. Cloze probability is typically an off-line measure of the likelihood of a certain lexical or grammatical continuation (Loerts et al. 2013). While this is incorporated into the pre-task requirements of many prediction-based studies, often with another independent sample of participants, it is difficult to see how such a method could be similarly applied to events and their ordering, although tense- and aspect-related information strongly associated with the encoding of events and their temporality might be sensitive to this kind of task (see Lopukhina et al. (2021) for a discussion on tense and cloze probability). To improve confidence in the conclusion that iconicity assumption violations are, fundamentally, prediction errors, further investigation in this area, using something like cloze probability, is encouraged.

In sum, verbal negation – whether in elided or explicit clauses, or both – results in the suppression of event information and the violation of comprehenders' expectations about the temporal contiguity of the unfolding events in a narrative. These integration difficulties via prediction error-like effects are reflected in N400 components to all three negated conditions. A gradation of violating effects resulting from both clauses being negated is not observed, however, as there is no difference between the condition in which only the explicit clause is negated and the condition where both clauses are negated. No reanalysis effects were observed, characterised by a post-N400-positivity, although (as P600s do not always accompany N400 prediction error effects) no strong hypotheses were made about their inclusion in the data anyway. Some noteworthy areas for further research are identified, namely an ERP latency difference to violations of comprehenders' iconicity assumption between negated explicit clauses and negated elided clauses, and a pronounced negative-going wave with an onset around 800ms post-stimulus that appeared in the negated conditions only.

4.4 - Experiment 4: Methods

4.4.1 – Participants

A total of 19 participants were recruited to take part in this experiment examining the effect of distributivity on assumptions of temporal iconicity. All participants were again members of the University of Essex community on either undergraduate or postgraduate degree programmes. All participants were native English speakers who were more than 18-years old and were a mix of genders and ages. Monetary compensation was provided upon completion of the experiment. Ethical approval was obtained for this study from the University of Essex Ethics Committee.

4.4.2 – Materials

Participants were presented with 48 three-sentence narratives in a reading task; the first two sentences were presented in full, with the third sentence presented in an SVP format. Likewise, the situation described in each narrative formed a temporally iconic contiguous series of events, with event X being introduced in sentence one, event Y in sentence 2, and event Z in sentence 3, with iconicity imposed via the use of temporal adverbials. The second sentence of each narrative described a buying event which was modified by a plurality denoting expression – the use of the conjoined noun phrase subject, *John and Mary* – giving rise to an ambiguity regarding the individuals over which the event ranged (i.e., a single event or multiple events). This second sentence appeared in one of three conditions: in condition A, the sentence included a disambiguating adverb which forced a collective interpretation of the event (*While there, John and Mary bough a book together*); in condition B, the sentence included an adverb which forced a distributive interpretation of the event (*While there, John and Mary bought a book separately*); in condition C, the sentence did not include a disambiguating adverb, and was therefore ambiguous between the two interpretations. Example narratives within this paradigm are illustrated in Table 6.

Table 6: List of example stimuli, coded for condition, with ERP time-locked words underlined; good and bad continuations were equally distributed across conditions.

Condition	Sentence 1 (Event X)	Sentence 2 (Event Y)	Sentence 3 (Event Z)	Continuation
A	John walked nervously into his local book shop at the same time as Mary.	While there, John and Mary bought a novel together.	Alice arrived a few moments later and greeted Mary with a hug.	Good
A	John paced slowly into his local music shop at the same time as Mary.	While there, John and Mary bought a record together.	Diana arrived a few moments later and greeted Mary with a spatula.	Bad
B	John skipped happily into his local book shop at the same time as Mary.	While there, John and Mary bought a novel separately.	Emily arrived a few moments later and greeted Mary with a smile.	Good
B	John charged quickly into his local music shop at the same time as Mary.	While there, John and Mary bought a record separately.	Paula arrived a few moments later and greeted Mary with a trampoline.	Bad
C	John trotted excitedly into his local book shop at the same time as Mary.	While there, John and Mary bought a novel.	Abby arrived a few moments later and greeted Mary with a hug.	Good
C	John traipsed absently into his local music shop at the same time as Mary.	While there, John and Mary bought a record.	Tina arrived a few moments later and greeted Mary with a spatula.	Bad

As with Experiment 3, EEG waveforms were recorded to the word *later* in the contiguity-defining temporal adverbial *a few moments later* in the final sentence to test the effect of ambiguously distributive contexts on the iconicity assumption. As prior research has shown that comprehenders eventually default to a collective interpretation of the events described in ambiguously distributive contexts, the overt collective reading of condition A acted as the control in this study. Condition B, with an overt distributive reading, was also included as the findings of chapter 3 of this thesis demonstrated that comprehenders appear to activate both collective and distributive representations during the processing of ambiguously distributive events, and so comparisons between conditions C and B (and A and B) were warranted.

Finally, as with Experiment 3, averaged ERPs time-locked to the narrative final word which consisted of a predictable, semantically good continuation or non-predictable, semantically bad continuation of sentence 3 were obtained to ensure participants were attentive to the task and ERP components to prediction effects were being accurately measured. A comprehension question about the narrative at the end of each trial was also included, with the overall accuracy of responses captured for each participant. (See *Appendix D* for a full list of experimental stimuli.)

4.4.3 – Procedure

The 48 three-sentence narratives were presented to participants and data was obtained for each in the same manner as Experiment 3, with the only difference being that the current experiment structure consisted of three randomised blocks of 16 trials, rather than four, due to only three conditions in this experiment compared to the four in the former (again, with all participants seeing all trials). See section 4.2.3 for details.

4.5 – Experiment 4: Results

4.5.1 – Analysis

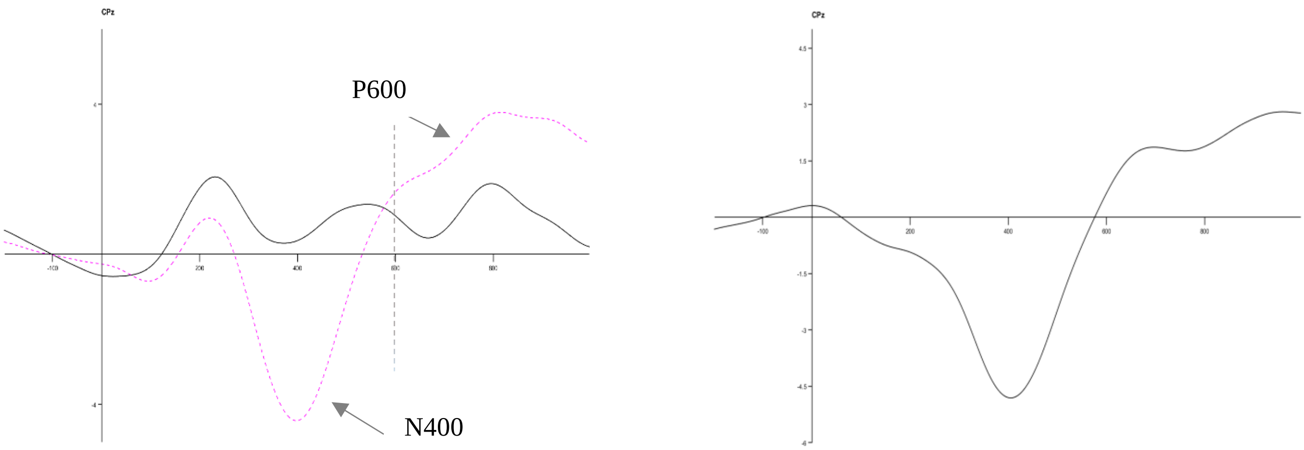
Of the 19 participants recruited, two showed significant electrode connection problems that could not be resolved during breaks leading to data collection being abandoned early. No participants were excluded due to falling below the 75% accuracy threshold in the comprehension questions.

Preparation of the remaining raw EEG data for Experiment 4 utilised the EEGLAB and ERPLAB toolboxes in an identical fashion to Experiment 3 (see section 4.3.1 for further details). Artefact rejection led to one additional participant being removed from further analysis (final n = 16) due to a large percentage of trials being eliminated (>25%). As previously discussed, disconfirmed prediction in congruent texts – which the experimental narratives are – has been associated with anterior P600 effects (see van Petten & Luka, 2012); therefore, the hypothesised P600 component following the ambiguously distributive condition which is not consistent with expectations of temporal iconicity of some single event Y followed by some event Z is expected to have a frontal electrode distribution. The following locations were used for statistical analysis: Fz, F1, F2, F3, F4, FCz, AFz, AF3, AF4, FC1, FC2, FC3, FC4. The P600 analysis window was treated as 500ms-900ms post-stimulus (see Kaan, 2023). For the narrative-final continuation, both a 300ms-500ms N400 window and 500ms-900ms P600 window were analysed at centroparietal scalp sites in line with Experiment 3 (see section 4.3.1). Averaged ERP data across trials and participants were again plotted alongside difference waves, and individual participant ERPs within the above time window were collated and analysed on RStudio using the *lmer* function, with mean amplitude by condition, where condition was numerised using deviation coding prior to running the model. As with Experiment 3, age and gender were also included as fixed effects within the model, while participant was included as a random effect (lmer(Amplitude ~ Condition + Age + Sex + (1 | PPT). The *emmeans* and *contrast* functions were likewise used for planned contrasts across the conditions, with Bonferroni correction used for p-value adjustment with multiple comparisons.

4.5.2 – Findings

Like the first experiment, ERPs to narrative-final continuations which were either predictable (good) or non-predictable (bad) were measured within the 300ms-500ms time window at defined centroparietal electrode cites characteristic of the N400 component typical of prediction errors, demonstrating that participants were indeed attending to the narrative and making predictions about upcoming stimuli. Bad continuations showed a statistically significant negative-going waveform when compared to good continuations (β = -4.1114, t = -12.30, $p < .001$); furthermore, a significant effect of reanalysis, characterised by a positive-going wave in the 500ms-900ms time window, was also observed (β = 1.6502, t = 4.514, $p < .001$). The averaged waveforms are illustrated in Fig. 11.

Fig. 11: Averaged ERPs at CPz node to bad narrative-final continuations compared to good narrative-final continuations (left), with difference wave (right). Mean critical word offset represented by dashed vline (601ms).

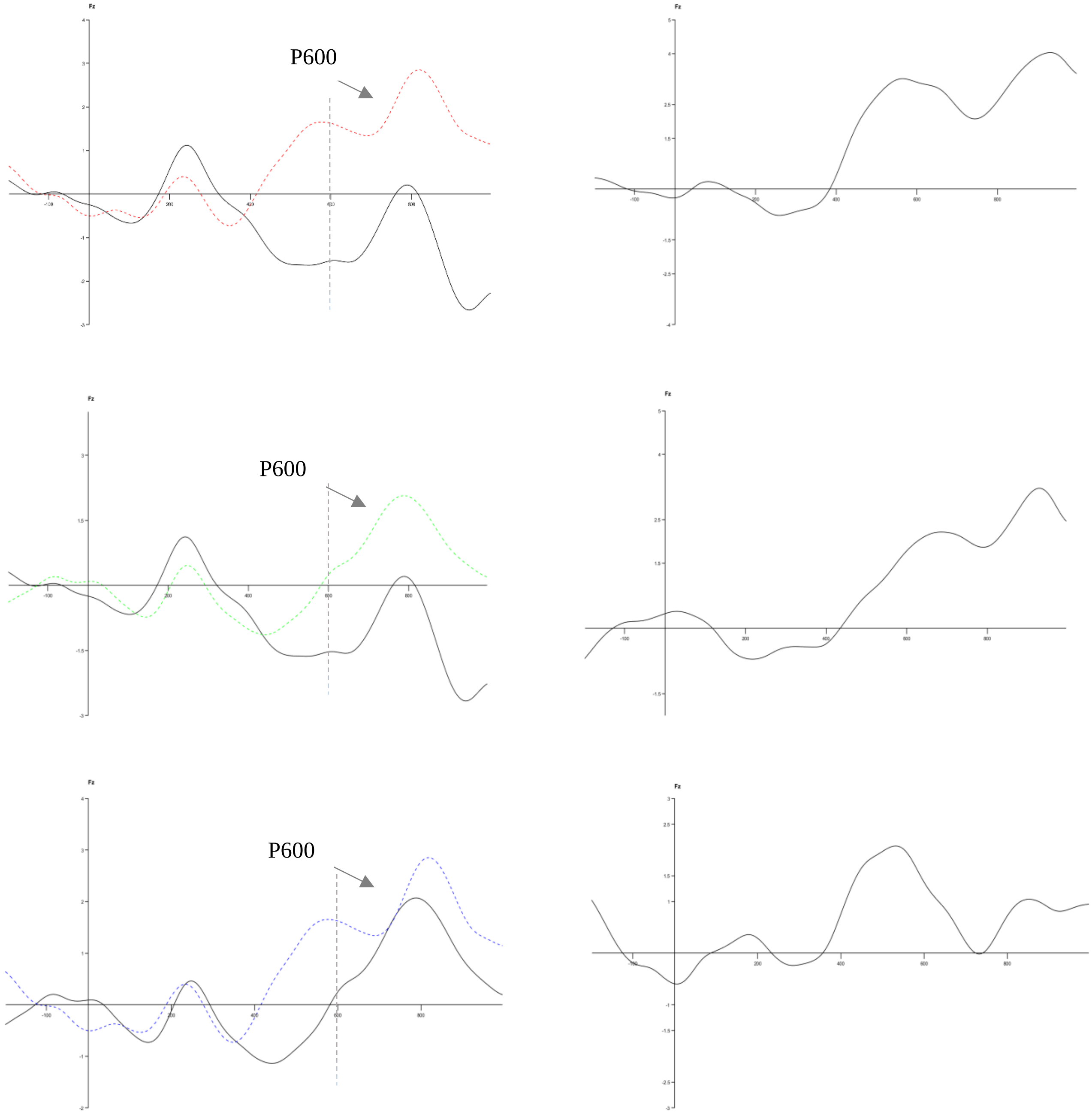


Unlike Experiment 3 and the narrative-final continuations, a prediction error ERP effect – the N400 – was not expected to ambiguously distributive narrative conditions at the critical word *later*, as there would be some sequence of contiguous events across all three sentences in the narrative, in all conditions. Instead, integration difficulties at the word *later* with ambiguously distributive narratives (condition C) were expected when compared to a disambiguated non-distributive control (condition A). These integration difficulties, due to prediction disconfirmation of assumed single-event contiguity, were expected to be maximal at frontal –

rather than centroparietal – scalp locations since the text remained coherent even following the experimental manipulations, which is not the case for posterior P600s.

Running a linear mixed effects regression model on the mean amplitude at defined anterior electrodes between 500ms-900ms revealed that the waveform to condition C was significantly more positive going than the mean of all levels for this factor (β = 0.886, SE = 0.22, t = 4.000, $p < .001$). Condition B, however, did not reach significance when compared to this grand mean (β = 0.081, SE = 0.22, t = 0.367, p = NS). Similarly, neither age (β = -0.058, SE = 0.07, t = -0.815, p = NS) nor gender (β = -1.742, SE = 1.46, t = -1.196, p = NS) were significant in the model. However, planned simple effect comparisons between each condition showed that the mean amplitude of both condition C (β = 1.854, SE = 0.38, t = 4.831, $p < .001$) and B (β = 1.049, SE = 0.38, t = 2.734, $p < .05$) were significantly different from the control, but C was only marginally different from B (β = 0.805, SE = 0.38, t = 2.097, p = .11). The P600 observed to condition B peaked 746ms post-stimulus, averaged across trials, participants, and defined electrodes. Comparing peak latencies of conditions C and B (via an identical model structure to the above analysis but with latency as the dependent variable) did not show a significant difference between these waveforms (β = 14.554, t = 1.497, p = NS), even though the onset of the observed component to condition C was earlier (almost 200ms earlier at the Fz node). The averaged waveforms at the Fz node are illustrated in Fig. 12, with difference waves.

Fig. 12: Averaged ERPs at Fz node to condition C compared to condition A (top), condition B compared to condition A (middle), and condition C compared to condition B (bottom), with difference waves included on the right. Mean critical word offsets represented by dashed vline (600ms).



4.5.3 – Interim Discussion

Ambiguously distributive events in an otherwise temporally iconic narrative were hypothesised to require reanalysis by comprehenders at the contiguity-defining critical word *later*, reflected in a positive-going ERP shift – when compared to an overt non-distributive control – between 500ms-900ms post-stimulus, known as the P600 (Brouwer et al., 2017; Kuperberg et al., 2020: Regel et al., 2014). This prediction was made because, in some sequence of events (Event X → Event Y → Event Z), if the intermediate event (Event Y) was ambiguous as to whether a collective reading should be adopted or a distributive reading should, it was assumed comprehenders would likely have difficulty integrating their developing representations with their *a priori* expectation that events should proceed contiguously and singularly from one event to the next in a chronological fashion. This integration difficulty following prediction disconfirmation in an otherwise congruent text was hypothesised to be maximal in anterior electrode sites in line with previous research (van Petten & Luka, 2012). Comparisons between an ambiguously distributive condition and a disambiguated collective condition were made. Inclusion of an overt distributive condition allowed for additional comparisons with the other two and demonstrated fascinating implications of its own.

Supporting the hypothesis that ambiguously distributive events would result in integration difficulties with comprehenders' iconicity expectations, a P600 to defined frontal scalp sites was observed in the ambiguously distributive condition (e.g., *John and Mary bought a novel*) to the critical adverb *later* when compared to the control condition of an overtly collective event (e.g., *John and Mary bought a novel together*). However, the overt distributive condition (e.g., *John and Mary bought a novel separately*) also demonstrated a P600 effect when compared to the collective control. At first glance, these findings suggest that it is not, as hypothesised, the ambiguity between collective and distributive readings (characteristic of the ambiguous condition C) that necessarily results in the integration difficulty with comprehenders' expectations of single event to single event contiguity within a narrative, but, perhaps, the adoption of a non-collective interpretation of events.

Considering this 'first glance' interpretation of the data, it is possible that, as comprehenders seem to eventually default to a collective interpretation of events unless otherwise prompted (see Frazier et al., 1999), a non-collective narrative (such as conditions B and C) is treated as somewhat inconsistent with their expectations, resulting in integration difficulty, which is thus reflected in the observed P600 components. However, this analysis is unlikely to be correct, as, if it were, it implies the P600 to condition C was not due to ambiguity, but due to comprehenders ultimately settling on a non-collective interpretation (i.e. a distributive one), which runs counter to previous findings. While not explicitly measured in this task, it is unlikely that comprehenders would eventually settle on a distributive reading of condition C over a collective one, given the strong bias to the collective in previous research. Although some grammatical constructions do appear to bias comprehenders to a distributive reading of ambiguous events, such as the present perfect, where the timespan in which the event can occur is extended (see chapter 3 above), this is not applicable to the present study where the simple past is adopted, similar to Frazier et al. (1999). As a result, it is doubtful the reanalysis effects observed are due to simply non-collective interpretations of conditions B and C.

Instead, a more likely explanation of the observed data is one in which no commitment was yet made to condition C, and ambiguity of different types is present in both conditions B and C when encountering the critical word *later*, resulting in similar integration difficulties and requiring reanalysis. To illustrate, for condition B, being overtly distributive, it is still technically ambiguous as to when the two events took place. The buying events by Mary and John might have occurred at the same time or at different times: i.e., John bought a novel and Mary, separately, bought a novel, and these occurred simultaneously, perhaps with the aid of different shop assistants at the same moment in time, or they might have happened at different times, maybe hours apart. This inherent temporal ambiguity is a fundamental property of distributive texts (albeit one that is woefully understudied; but see chapter 3 which analyses the effect of different timespans on ambiguously distributive texts). Meanwhile, condition C is ambiguous in a different way, between a collective interpretation and a distributive one, where both representations may be active (or at least accessible) at the critical adverb *later*, resulting in integration difficulty with comprehenders' iconicity assumption about single event to single event contiguity, reflected in the observed P600 compared to the collective control. If correct, then comprehenders have not settled on an interpretation by the introduction of the critical adverb but are attempting – and failing – to rapidly integrate multiple (narratively inconsistent) representations into their unfolding mental model of the situation described.

Intriguingly, the positive-going component to condition C had a more positive mean amplitude than condition B (which was marginally significant). While researchers must be cautious to not over-interpret ERP results (and marginal significance effects), taken together, these findings seem to show a potential gradation of effects: a distributive event is possibly more difficult to integrate within a temporally iconic narrative than a collective one, perhaps due to the inherent temporal ambiguity in such constructions, while additional integration difficulties are observed where that critical event is ambiguously distributive, indicating comprehenders' expectations about temporal iconicity are being substantially reassessed. These data strengthen the claim that both distributive and collective readings are active (or at least accessible) during the processing of ambiguously distributive narratives and that neither a collective interpretation nor a distributive one is immediately committed to; these data also indicate that the plurality of event representations accessible in C produces greater integration difficulty to the contiguity-defining adverb *later* than the events in the temporally ambiguous condition B, although further research is in this area is warranted to see if more robust effects can be demonstrated.

Overall, these findings support the hypothesis that, compared to overtly collective events, ambiguously distributive events are more difficult to integrate with expectations of single event to single event contiguity within a narrative. However, they also reveal that overtly distributive events similarly result in integration difficulties, perhaps due to the fundamental temporal ambiguity present in all distributive events. An apparent gradation of P600 effects across the non-collective conditions tantalisingly point to different degrees of integration difficulty reflecting the different types of ambiguity present. In line with findings from chapter 3, these data support the conclusion that both collective and distributive representations are active during the processing of ambiguously distributive events, while also demonstrating that these multiple representations interfere with our *a priori* expectations of event ordering.

4.6 – General Discussion

Bringing the observations of the two experiments that comprise this chapter together, three areas of interest are worth additional discussion: the marginal P600 observed to the narrative-final continuation in Experiment 3, the negative-going shift around 800ms post-stimulus in negated conditions of Experiment 3, and the superficial similarity between one of the conditions in Experiment 3 and another condition in Experiment 4 that nonetheless seem to show different effects in the 500ms-900ms ERP time window.

While not essential to the core hypotheses of this chapter regarding the effects of verbal negation and covert distributivity on the expectations of temporal iconicity in short narratives, each experiment nevertheless included a measurement of prediction error and reanalysis to an unexpected continuation at the end of each trial to act as a baseline and improve confidence that components observed in the novel methodological design regarding the iconicity assumption were likely to be valid. In Experiment 3, a statistically significant N400 component was observed following the presentation of a non-predictable ('bad') continuation when compared to a predictable (‘good’) continuation; furthermore, a marginally significant post-N400-positivity was observed. This putative P600 was treated as being in evidence, as factoring in the data from the (identical) continuations in Experiment 4 – where both an N400 and P600 reached significance – there were good reasons to believe that the trend of a P600 in Experiment 3 was indeed a real effect. These components, with a centroparietal maximal, demonstrated both prediction error-like effects (N400) and reanalysis or integration difficulties (P600) to continuations that were not aligned with comprehenders’ expectations of their linguistic input.

Second, while, in Experiment 3, an N400 effect was observed to the critical word *later* at centroparietal sites following negated contexts, demonstrating that negation of intermediate events (in both elided and explicit clauses) violates comprehenders’ assumptions about event contiguity, no subsequent reanalysis or integration difficulty, reflected in a P600, was observed. This is not terribly surprising, as not all N400 effects in the literature are accompanied by a post-N400-positivity, or vice versa (see Hoeks et al., 2004; Kolk et al., 2003; Kuperberg et al., 2003; van Herten et al., 2005; Wang et al., 2009 – but also see van Petten & Luka, 2012, for discussion). Moreover, no clear hypotheses about a potential P600 and the conditions in which it might be found were theoretically justified prior to this study. Given that no reanalysis effects were observed following the critical manipulations of negation, but a marginal P600 was observed to the non-predictable narrative-final continuation, further investigation into the possible causes of this disparity is merited, a point made more emphatic by the observation of the pronounced negative-going waves with an onset around 800ms to the three negated conditions, but not to the non-negated control. From another perspective, one could argue that all three negated conditions did show a positive deflection (see Fig. 10) followed by this pronounced negativity in the 500ms-900ms time window, characteristic of a post-N400-positivity waveform, whereas the non-negated control shows minimal voltage variation during that period. While remaining highly speculative at this time, it is possible that this is a trend in the direction of linguistic reanalysis that may have failed to reach significance due to a lack of statistical power. That the P600 in the narrative-final continuation was only marginally significant, as well, justifies further consideration of this suggestion in future work.

Finally, a P600 to anterior sites was observed in Experiment 4 to the critical word *later* following both the ambiguously distributive and overtly distributive conditions when compared to a non-distributive control. The ambiguously distributive condition in Experiment 4 (*John and Mary bought a novel*) is semantically similar to the control condition (*John bought a novel and Mary did too*) in Experiment 3. However, while the former had a pronounced positivity in the 500ms-900ms time window (see Fig. 12), the latter did not.[30] Although these two conditions were from different tasks, with different participants, are were therefore not directly measured against one another, the pattern of findings between them indicate that the coordination of two entities in a subject over which an event ranges likely affects distributivity representations (and thus temporal iconicity expectations) differently than the conjunction of two clauses, each with its own subject and verbally-encoded event. That is to say, while condition A of Experiment 3 and condition C of Experiment 4 appear superficially alike, the type of coordination in these expressions probably resulted in different representational processes during the above tasks. More research to explicitly test the validity of these claims is recommended.

[30] Although not reported here, no positivity between 500ms-900ms was observed in anterior sites for any of the conditions (including the control) of Experiment 3, either.

4.7 – Conclusion

This chapter examined the effects of verbal negation and ambiguously distributive constructions on event representations by measuring the impact of their inclusion on a well-attested event processing phenomenon known as the iconicity assumption, wherein comprehenders expect, without evidence, that events will be orders in a singular and contiguous fashion, based on their experience of the passage of time. Two EEG experiments were conducted, measuring averaged ERPs across trials and participants at electrode sites defined by existing research on prediction error and reanalysis effects. Experiment 3 showed that verbal negation, whether in elided or explicit clauses, within the middle of an otherwise contiguous narrative, leads to violations of comprehenders' expectations about event ordering, providing support for the conclusion that the representations of these events are suppressed from forming part of the larger developing mental model the situation described. Meanwhile, Experiment 4 demonstrated that both ambiguously distributive texts (where both collective and distributive readings are possible) and overtly distributive texts (which are still temporally ambiguous) result in integration difficulties when compared to a non-distributive, non-ambiguous control, with these effects appearing in a graded fashion based on the degree of ambiguity present. Taken together, these results emphasise the importance of additional consideration of understudied linguistic features – such as negation and distributivity – and their interaction with comprehenders' event representations, while highlighting the value of using the iconicity assumption as a tool to test broader claims about situation model construction.

Chapter 5

Reflections on Event Numeration

5.1 – Overview

This project began with the notion of story-level representations – abstracted mental models of the situation described by some linguistic input – and the theoretical importance of the events that comprise them. It was observed that while research into event representational processes in language comprehension has blossomed over the last few decades, with a recent turn to the fascinating discoveries that events in language and visual perception share fundamental properties and that these representations interact with our embodied knowledge and experience, there remains a relative dearth of research on various grammatical contributions in the construction of event representations; furthermore, much of the extant literature has examined event representations of a singular nature, or sequences of events in which one event leads into another (see chapter 1 for a full discussion). A considerable gap in our knowledge of event representations, and thus the story-level situation models they compose into, was thus identified: what happens to these representations when more than one or less than one event occurs, and how do these non-events or pluralistic events contribute to the construction of a comprehenders' developing mental model?

Two types of linguistic structure – negation over verb phrases (which typically encode events) and plurality denoting expressions (that give rise to ambiguously distributive events) – were recognised as being able to offer constructive insights into non-singular event representations, while also being both common in usage and understudied with respect to situation model construction. Through three related studies, testing the role of these linguistic constructions on event representations following probe words (chapter 2), narrative continuations (chapter 3), and their impact on *a priori* expectations about event ordering (chapter 4), this thesis aimed to increase our understanding of how comprehenders determine the number of events and in what way they are mentally represented. Viewed together, the role of negation and distributivity on event representations was operationalised as the umbrella term *event numeration*, and its effect on situation model construction became the central theme of this thesis.

Careful consideration was given to the scope of this work. Situation model construction has been shown to be a complex phenomenon, consisting of event representations – including the monitoring of the changes in the temporal, locational, motivational and causal relationships that hold between the entities or objects that comprise events – encyclopaedic (or world) knowledge – involving contextual, pragmatic, schematic and intertextual information about worlds, fictional or otherwise – and embodied knowledge – through which comprehenders seemingly mentally simulate the described situation, often including themselves, in a conceptual space, activating associated sensorimotor regions of the brain while doing so (see chapter 1 for review). Given this complexity, and the intention to examine event-complex phenomena via negation and distributivity, a decision was made to simplify the model of mental representations under analysis. Knowledge, both encyclopaedic and embodied, was purposefully sidelined for the duration of this project to maintain that simplicity. However, as previously remarked, all findings from the present research should be considered alongside insights from these temporarily abandoned topics, as a full accounting of situation model construction necessarily contains them. I will speak more of this during section 5.4 of this chapter, where I look to future work that builds on the novel observations of this study and incorporates the valuable perspectives offered by research on these two types of knowledge.

First, however, in section 5.2, I look back and reflect upon some methodological challenges faced throughout the completion of this work, their impact on the collection of data for chapters 2 and 3, and why – following the practices adopted – confidence in the validity and reliability of the findings is both justified and sound. In section 5.3, I then summarise those findings, situating them within the wider literature on models of comprehension, while drawing upon both the Construction-Integration (CI) and Event-Indexing (EI) models previously introduced.

As mentioned, section 5.4 revisits the significance of knowledge in situation model construction with respect to event numeration, while also turning an eye to potential research that looks directly at the role of other understudied linguistic features on event representations, both singular and complex. In the final section, 5.5, I reflect holistically on the successes and limitations of this project and bring this work to its conclusion.

5.2 – Methodological Challenges

Before reviewing the results of this multi-stage project and the way the findings inform upon each other, including their implications for our theories of language comprehension, I must first reflect on the impact of the covid-19 global pandemic on the collection of response time and reading time data for two of this thesis' chapters. While, ideally, these experiments would have taken place in an experimenter-controlled computer laboratory with identical machine set-up for each participant, due to face-to-face national restrictions in the UK and at the University of Essex during the height of the pandemic, when these data were collected, all participants included in chapters 2 and 3 were sampled from an online source: Amazon's Mechanical Turk. In this section, I discuss some of the literature regarding conducting psycholinguistic experiments online, with a particular focus on the use of Mechanical Turk (henceforth, MTurk), and argue – due to the methods adopted to ensure the robustness and reliability of the final datasets – that confidence in the results of chapters 2 and 3, despite being collected online, is entirely warranted.

5.2.1 – Collecting Data Online

This thesis was undertaken, and a fair amount of the data was collected, during the height of the covid-19 pandemic, engendering the use of online data collection for chapters 2 and 3. In a world where participants are likely to have very different hardware set-ups at home, and research attests differences in measured response times to different hardware (Anwyl-Irvine et al., 2021; Plant et al., 2003; Plant & Turner, 2009; Pronk et al., 2020), alongside the fact that there exists a persistent view that conducting web experiments might be unsuitable for research because of this technological variance (Brand & Bradley, 2012) or low-quality participant engagement (Aust et al.; 2013), the question of whether results from online data are valid and reliable became particularly important for this project.

Although no comparison was made to compare the findings of chapters 2 and 3 with an offline replication of the same study, several published projects that include this exact type of comparison are found throughout the literature, which endorse the lack of my own. Looking at a sample of this wider literature, the concerns around online data collection – based on the observation that responses obtained online frequently differ from those gathered in a researcher-controlled environment – might be slightly overestimated, as while differences are indeed observed, those differences, particularly when considering psycholinguistic measures, tend to be in matters of latency rather than differences of effect.

For example, de Leeuw and Motz (2015) directly compared reaction times to a visual search task, where participants had to identify a target letter or image in an array of distractors, using both offline and online modalities (MATLAB's Psychophysics Toolbox and JavaScript, respectively). The authors found that while response times measured through a web-browser by way of JavaScript had a mean RT around 25ms longer, both approaches were equally sensitive to manipulations of experimental conditions, with the same statistically significant effects being observed.

Likewise, Reimers and Stewart (2007) found that response times to a binary correct-incorrect decision task in an online Flash-based experiment were, on average, around 20ms slower than an identical experiment run on a single computer in the researchers' laboratory, and 30ms slower than a Linux-based version of that same experiment also conducted in the lab; crucially, all versions of the experiment revealed similar effects with respect to task. Replicating the general findings of other comparison studies, Linnman et al. (2006) found robust Stroop effects in both an online and laboratory-administered version of the same experiment, although responses in the online version were significantly slower on average than the laboratory one.

Conversely, not all research indicates that response times are slower in online versions of the same task. Kim et al. (2019) tested whether there were any differences between an E-Prime study conducted on a single computer in laboratory conditions and one conducted through PsyToolkit (Stoets, 2010; 2017) on participants' personal internet-connected computers, and found that while both versions of this experiment showed similar differences in mean RTs to male and female names paired with stereotypical occupations, average responses times were quicker in the PsyToolkit version of the experiment. Kim et al. argue that this was possibly because participants felt more natural in their home environment, with no experimenter present, and were thus less affected by any social desirability bias. The suggestion here is that online studies might be more ecologically valid than laboratory ones in some cases.

Elsewhere, Hilbig (2016) also showed that the size of an experimental manipulation might be stronger during some online tasks compared to offline ones. Hilbig ran a lexical decision task, where participants had to identify if a stimulus was a real-word or a non-word, across three conditions: in a standardised laboratory with the E-Prime software, in that same laboratory but on a web-browser running Javascript, and at home on a web-browser running Javascript. Consistent with much of the other literature, all three conditions were sensitive to the lexical decision task; however, word frequency tended to have a greater effect on RTs in the online versions of the experiment than the offline one.

Viewed together, these studies demonstrate that not only are RT effects replicable in online versions of experiments, and that differences are often just in matters of latency, but sometimes the effects are stronger, or RTs are faster, when completed online. While the initial intention was to conduct the experiments for chapters 2 and 3 of this thesis in a researcher-controlled lab at the University of Essex, the shift in approach demanded by the pandemic is likely not to be as confounding as originally feared, and the results might even be more ecologically valid as a result. Indeed, distrust of the validity and reliability of online experiments in psycholinguistic research is probably, on some level, a little misguided.

Nevertheless, care should still be taken to ensure that certain confounds applicable to online data collection, such as the pool of self-selecting participants, some of whom may not be performing in good faith, do not interfere with the collection and analysis of reliable data. One such challenge was faced in this project due to the adoption of Amazon's Mechanical Turk for the recruitment of its participants.

5.2.2 – Participants on MTurk

MTurk is an online crowdsourcing website designed for the acquisition of participant data collection in experiments, surveys, and questionnaires, demanding as little as a couple of minutes or over an hour of their time, such that its popularity grew tenfold in the decade prior to the pandemic (Walter et al., 2019) and has continued to be popular since (Albert & Smilek, 2023; Newman et al.; 2021). Its name is based on an 18th Century mechanical chess-playing automaton which had the appearance of a human-sized mannequin wearing a turban and fur-trimmed robes, owned by Hungarian nobleman Wolfgang von Kempelen (Glaeser & Strouhal, 2000). Kempelen's Turk is remembered to have defeated nearly every opponent it faced, but, in a *Wizard of Oz*-style twist, there was a human chess master concealed inside the machine all along. Today, MTurk allows researchers, known as requesters, to submit their surveys and experiments to the platform in the form of a specified number of Human Intelligence Tasks (HITs), after which self-selected participants, known as workers, elect to take part in one of the available HITs or not. Restrictions can be placed on which workers are able to apply to complete a HIT, such as age, language, or nationality, though Paolacci et al. (2010) found about 70-80% of the site's workers at that time were from the United States. Researchers can reject a completed HIT if justification is provided, such as the quality of responses not being of the required standard (e.g., too fast or performing at chance), whereupon the HIT is returned to the pool of available participants for someone else to complete. Typically, HITs will include a link to the survey or study on an external platform that uses a compatible web-programming language, such as Javascript or HTML Chapters 2 and 3 of this thesis included such links to experiments designed on PsyToolkit, itself based on Javascript. Specifying that the experiments in the current research project required native English speakers resulted in all participants recruited being residents of the US.[31] Occasionally, HITs were rejected and returned to the participant pool, but this was quite rare, as most of the data that ended up being excluded was removed during the preparation of the raw data for analysis.

Although collecting psycholinguistic data online is likely to be as reliable and valid as laboratory methods, as discussed in the previous section, and MTurk has itself been found to replicate common psychological response-time measures in Stroop, Flanker and Simon tasks (Crump et al., 2013), as well as other psycholinguistic measures such as filler-gap dependency processing (Enochson & Culbertson, 2015) and scalar implicature processing across incremental decision tasks and web-based eye-tracking tasks (Degen et al., 2021), the culture

[31] Participants who took part in both studies were monetarily compensated for their contribution, with Amazon taking an additional 10% commission on all payments. Aker et al. (2012) found that the level of payment offered to participants on MTurk can improve performance on tasks where answers are binary (correct or incorrect), encouraging participants to respond attentively; on the other hand, for more subjective judgements or tasks, including those in psycholinguistic research, there appears to be no correlation between the rate of pay and the quality of the responses (Buhrmester et al., 2011). Paolacci and Chandler (2014) suggested that participants on MTurk have high levels of intrinsic motivation to complete tasks regardless of payment, although see Moss et al. (2020) for a view of the ethics of compensation rates on MTurk, and Albert and Smilek (2023) for a discussion of attentiveness among MTurk workers. Aguinis et al. (2021) also provides an interesting meta-analysis of recent MTurk usage in data collection and provides recommendations about matters like compensation rates.

and patterns of behaviour among some workers on the MTurk platform introduces additional complications that needed to be considered during this thesis: those of ‘speeders’ and bots.

Ford (2017) reviews participant behaviour on MTurk, from the perspective of surveys in advertising, noting an alarming trend of what they and Smith et al. (2016) call ‘speeders’: respondents who essentially speed through surveys without reading questions carefully and applying only the minimum effort required in terms of mental engagement to complete the HIT and be paid as soon as possible. Yarrish et al. (2019) suggests that some ‘speeders’ might, in fact, be bots, an ironically apt observation considering MTurk’s namesake. Yarrish et al. define bots as computer programs that complete online forms at a faster rate than would be possible for humans – a keyboard macro activating keys at a faster pace than a human, especially if activated at a regular pace, could also be considered a bot. So-called ‘speeders’ or bots were, unfortunately, observed in the collection of data for chapters 2 and 3 of this thesis, such that measures both before the acquisition of data, including the design of the task, and in the preparation of the data for statistical analysis were undertaken to minimise their impact on the final results.

Early in the pandemic, a pilot of the data collection process for chapter 2 was conducted, in which 20 participants were recruited from MTurk using a primitive version of the experiment. In response to this survey, response times were overwhelmingly below 100ms, especially as the experiment progressed, indicative of ‘speeders’ or perhaps macro-like bots in the sample[32]; indeed, some participants had RTs across the board which were both unrealistic for the task and surprisingly uniform (within a millisecond or two of each other).

[32] Many psycholinguistic tasks have empirically motivated thresholds, with anything less than 100ms certainly being non-human-like behaviour (Luce, 1986).

To get around problems like this, Hauser et al. (2019) recommends the inclusion of warnings as a method to increase the quality of the data and attentiveness among participants when conducting research on platforms like MTurk; they also made other suggestions for potentially improving data quality such as including 'red-herrings' and 'catch-questions'.[33] The latter options were not adopted in the experiments of chapters 2 and 3 to avoid extending the length of the discourse-heavy tasks even further, but warnings were added to the start, including an adjustment to a task in chapter 3 which demanded more engagement than a simple button press.

The included warnings made clear to participants that responses were being recorded and analysed and that frequent unrealistic response times *may* result in the HIT being rejected. While this was rarely acted upon in the actual data collection process, when piloted with an additional 10 participants, these warnings seemed to improve the overall response times, such that based on a simple eyeballing of the raw data file, substantially fewer responses fell into the <100ms time range. These pilot data suggested that the additional warnings were having a positive effect on generating fewer 'speeders' or bots, or at least discouraging them from accepting the HIT in the first place, and thus more realistic response times were being obtained. Naturally, these warning were included in both finalised experiments for chapters 2 and 3.

A final adjustment made to address the problem of bots was an alteration to how a critical task in chapter 3 was completed. Unlike chapter 2, in which a measure of accuracy was collected that could be used to trim inattentive participants at a later point in the analysis process, chapter 3 included no such measure, due to its design already requiring a naturalness rating instead.

[33] Yarrish et al. (2019) argue that it can be difficult to differentiate bots from non-bots, and real participants from human bots (people who are paid low amounts of money by third parties to rapidly finish the task in an effort to acquire the compensation provided by the MTurk requester), although they suggest work-around techniques, such as open-ended questions or password protected sections of the study. These methods, they note, bring their own disadvantages, such as requiring extra steps for well-meaning participants and increasing time demands on both participant and researcher in completing and reviewing more open-ended materials. Furthermore, they point out that 'catch questions', like those suggested by Hauser et al. (2019), might also be problematic as they may end up 'catching' real, attentive participants who may have simply been confused by the question.

From a desire to keep the discourse-heavy experiment as short as possible, an additional task to measure accuracy was not added. For the experiment in chapter 3, participants were instead required to use mouse movements to highlight and then select their preferred option for the naturalness rating question (as opposed to continuing to press keys on their keyboard). This aspect of the design was introduced to increase participant engagement in the task and reduce the impact of macro-like bot behaviour among the participant pool.

With these measures in place, and several steps taken to prepare the raw response time data for final analysis, including the exclusion of data that may still be indicative of bot-like behaviour or mistakes in responses, such as low accuracy and fast response times compared to each individual's mean (see Analysis sections of chapters 2 and 3), the final pool of participants included in the present studies are likely to been performing in good faith, with steps taken to address the confound of potential 'speeders' and bots that can affect data collected online. In line with suggestions by Yarrish et al. (2019), I recommend frequent data monitoring of the raw output when conducting research online via platforms such as, but not limited to, MTurk, where 'speeders' and bots may be prolific. While additional burdens may be placed on the researcher's time or on the design of the study, workarounds like warnings, questions, passwords, or variations in task type, alongside careful review of the raw output files and preparation of the data for analysis, can be effective at filtering out potential bots and 'speeders' from the study. Caution should be used with blanket rejections of HITs (or the equivalent on other platforms), as, if not used sparingly and with good justification, this might lead to a Type 1 error, whereby a real, attentive participant could be inaccurately considered a bot.

Viewed together with research discussed in the previous section on validity and reliability of online data collection, there are good reasons for trusting psycholinguistic research collected online. Care, however, must be taken in the process of obtaining that data to eliminate 'speeders' or bots buried in the participant pool. Confidence in the present findings is justified, as several methods, both in the distribution of the experiments and in the preparation of the data, in line with recommended suggestions throughout the field, have been implemented with success to ensure the remaining participants were human-like, attentive and performing the tasks in good faith. While the experiments in chapters 2 and 3 were not replicated offline, chapter 4, which was conducted in an experimenter-controlled environment, reveals findings about the role of negation and distributivity on event representations that, in many respects, support the conclusions drawn from the experiments reported in chapters 2 and 3, as isreviewed below, further strengthening conviction in the validity and reliability of these results.

5.3 – Implications of the Research

To address this thesis' central question – how event numeration contributes to situation model construction – three interrelated studies were conducted. These studies, comprising chapters 2,3 and 4 of this thesis, examined the role of verbal negation (assumed to suppress an event from forming part of the larger mental model) and plurality denoting expressions (which produce ambiguously distributive event meanings) on mentalised event representations. The effects of negation and distributivity on comprehenders' *a priori* expectations of event contiguity were also examined. This section evaluates the core findings of these studies, focussing first on the composite results of negation and second on the combined results of distributivity, before finally integrating them together within the larger theoretical context of situation model construction introduced in chapter 1.

5.3.1 – Summary of Results: Negation

This thesis demonstrates that negation which ranges over a verbal predicate suppresses the event encoded by that verb and its arguments from forming part of the comprehenders' developing mental representation of the situation described. This conclusion is evidenced in both chapter 2, wherein access to critical concepts inside the event differs as a function of whether that event is negated or not, and chapter 4, in which intermediate events in an otherwise temporally iconic narrative differ as to whether they violate comprehenders' default expectation that events will be ordered in a contiguous fashion, depending on if that intermediate event was negated or not.

Negated verbal predicates suppress message-level representations, forcing comprehenders to rely instead on fading surface-level representations, which appear to be more economical in terms of processing, resulting in faster responses to critical probes in negated conditions than non-negated conditions, a finding which flips the observations in previous publications on the negation of nominals (in which the event remains intact, but the concept being probed is itself suppressed), leading to slower responses to those critical probes. Excitingly, these data suggest that the function of verbal negation on situation model construction differs from nominal negation – a novel contribution to the language comprehension literature.

Negated verbal predicates violate comprehenders iconicity assumption, reflected in prediction error effects in the averaged ERPs to critical continuations which establish that expected contiguity, highlighting not only that negation suppresses events from composing into a comprehenders' developing discourse representation (a finding replicated in both chapters 2 and 4), but that the widely-attested iconicity assumption is an *a priori* prediction that can be operationalised to test claims about event representations more broadly. Remarkably, this thesis appears to be the first study, to my knowledge, to demonstrate the potential empirical and theoretical applications of the iconicity assumption as a tool to improve our knowledge of when (or if) events are encoded in memory.

Finally, negated verbal predicates in syntactically elided and explicit clauses seem to differentially affect the representations formed – or at least how they are processed – as negation of explicit clauses to temporally iconic continuations show a significant latency effect when compared to those in elided passages; furthermore, negation in explicit clauses shows significant decreases in response times to critical probes, whereas negation in elided clauses does not reach statistical significance. These differences may be minor or may simply be an artefact of the studies' design, but, given the replication of a difference between elided and explicit clause negation across chapters 2 and 4, future inquiry is encouraged to precisely delineate the potential difference between these linguistic constructions and their effect on comprehenders' mental models.

5.3.2 – Summary of Results: Distributivity

This thesis also discovered that ambiguously distributive constructions generate both collective and distributive representations which remain active (or, at least, accessible) to comprehenders during the unfolding of a text. This conclusion finds evidence in both chapter 3, in which differential reading times to critical continuations are observed following non-distributive events but not ambiguous events, and chapter 4, where that ambiguity leads to increased integration difficulty with comprehenders' expectations of single event to single event contiguity when compared to narratives with a disambiguated, non-distributive event.

Ambiguously distributive constructions, whether introduced by a numeral-modified subject or a conjoined subject, over which an event ranges, lead to no preference in whether a plural or non-plural continuation is included in the input, whereas a strong preference for non-plural continuations is observed following non-distributive events, demonstrating that mental representations consistent with a distributive interpretation, as well as a collective one, are active during the processing of narratives with ambiguous events. A commitment to a collective interpretation is made eventually, however, such as when prompted by an explicit judgment of the naturalness of the prose. These observations matter for the field, as they both support and challenge existing psycholinguistic accounts of distributivity: that collectivity is ultimately preferred aligns with the existing literature; that both collectivity and distributivity are acceptable to readers during on-line processing is one of the unique contributions of this work to the wider conversation in event representation research.

Ambiguously distributive constructions in otherwise contiguous sequences of events also produce integration difficulties, reflected in components associated with linguistic reanalysis in the averaged ERPs to critical continuations that reinforce default expectations of event ordering. When compared to a disambiguated, non-distributive event, these reanalysis effects are observed in both covertly distributive clauses and overtly distributive clauses, both of which are ambiguous; the size of this effect appears in a graded fashion based on the type of ambiguity, with constructions that are ambiguous between collective and distributive readings resulting in (potentially) greater integration difficulties than constructions that are temporally ambiguous but unmistakeably distributive. These observations replicate the valuable finding from chapter 3 that both collective and distributive representations are accessible following ambiguously distributive events, while underscoring the utility of the iconicity assumption as a tool to test claims about event representational phenomena more broadly.

Finally, this thesis demonstrates that certain grammatical structures can influence the representational bias comprehenders hold following ambiguously distributive events. In utterances where the timespan in which an ambiguous event occurs is extended, such as those in the present perfect – where the temporal scope is stretched over some event in the past and remains relevant until the present moment – comprehenders' prefer continuations consistent with a distributive reading, wherein multiple events occurred. Although this effect fades when explicitly tasked to make a linguistic naturalness judgement, suggesting that the aforementioned eventual bias toward the collective might persist, these findings illustrate the importance of continued empirical research on the role of understudied grammatical structures on events and their mental representations.

5.3.3 – Modelling Event Numeration

Chapter 1 introduced several models of language comprehension and representation proposed by researchers across the field, two of which were given focal significance: the Construction-Integration (CI) model and the Event-Indexing (EI) model. While these are not the only models to have been proposed or gain significant traction in the literature, it is my contention that these models offer our most empirically motivated bottom-up and top-down accounts, respectively, of representational construction and update. Mention was made, also, of the Event Horizon (EH) model, which builds upon event segmentation work in visual perception, and has been formidable in recent years in explaining the relationship between events, both linguistic and otherwise, and how they are individuated and bounded. Each of these three models (CI, EI and EH) offer different vantage points through which events are understood. It is likely, then, that some combination of all three, alongside insights from experimental event semantic work, reflects the true state of affairs with respect to the mental representation of events and situation model construction. It is important to stress, however, that while this thesis was principally concerned with the role of event numeration on situational representations, it was not intended to be, nor has the scope to be, a treatise on the finer points of some putative 'super-model' that incorporates the best of the above. Instead, I use this section to review the strengths, and possible limitations, of the focal CI and EI models in capturing the key findings of chapters 2, 3 and 4, with a view to how these bottom-up and top-down approaches explain how comprehenders enumerate events. Consideration of how other models like the EH model (Radvansky, 2012), Story Grammars (Brewer & Lichtenstein, 1980), Discourse Representation Theory (Kamp & Reyle, 1993) or the Bayesian-inspired generative framework of event comprehension proposed by Kuperberg (2016), account for the reported data has been and will continue to be minimal, as they fall outside the defined scope of this project; nevertheless, future analysis of other models of comprehension with respect to this work is encouraged.

The findings on negation in the above chapters show differences in accessibility to conceptual representations based on whether the message-level representation of an event is suppressed by negation or not. The Construction-Integration model provides a potential explanation for these findings through proposing that there are three parallel representational pathways during situation model construction (Kintsch, 2018; Kintsch et al., 1990; van Dijk & Kintsch, 1983): first[34], surface-level representations that are characterised by the exact words and phrases encountered in the input; second, the semantic content of propositions extracted from the input, reflected in the global macrostructure of the text; and finally, the situation model, which is not the text itself, but the abstracted message described by that text (often embedded in domains of pre-existing, perceptual and contextualised knowledge). The first and final of these representational stages are important in explaining the present data on negation.[35]

The central observations of chapter 2 revealed that, when compared to a non-negated clause, verbal negation of accomplishments results in faster responses to critical probes, which was interpreted as evidence of the event being suppressed from forming part of the unfolding message-level representation. The implication, here, is that comprehenders would have to rely on their fading surface-level representations of the text to access the probed concept instead. The original insight these data offer is that – given that comprehenders ultimately responded quicker to probes when relying on surface-level representations only – adopting a surface-level

---

[34] While the terms first, second, and final have been used here, this is for rhetorical convenience only, as strictly speaking these proceed in parallel, with inferences at the level of the situation model often informing information at the other two, and vice versa (Kintsch et al., 1990).

[35] The 'second' of these three representational pathways would roughly correspond with the linguistic representations of logical form modelled by semanticists (Kintsch et al., 1990), such that operators that range over predicates and their arguments, like negation (¬) or distributivity (D), might be represented here if they had psychological reality. This representational stage is flexible enough to accommodate multiple competing representations, with commitment to abstracted concepts and events achieved only under the 'final' representational stage (*idem*.). Access to probed concepts is, therefore, likely to be strong in surface-level representations, at least early in the time-course of processing, where comprehenders will possess memory traces of the corresponding set of orthographic symbols (Lüdtke et al., 2008; Ratcliff & McKoon, 1989), and to message-level representations of the events and objects that compose them, for long periods following the presentation of the stimulus (Davoudi & Moghadam, 2015).

access route to representations may be more economical, at least in the short term, than searching through abstracted message-level representations for that concept.

The effect of message-level suppression was strongest when both explicit and elided clauses were negated, and thus no event representations were formed, leading to the shortest average response times. However, a comparison between response times to probes in conditions where both clauses were negated and responses to conditions where only the explicit clause was negated did not reach statistical significance, suggesting that the critical factor in the suppression of message-level representations – and the reliance on fading surface-level information – is the negation of the explicit clause in which the textual reference to the concept being probed is found. Interestingly, there did seem to be a trend in the direction of this suppression effect when only the elided clause was negated – wherein the explicit orthographic reference to the concept being probed was inside another verbal predicate which did end up forming part of the unfolding situation model – but, again, this did not reach statistical significance, indicating there is a probable difference between elided and explicit negation in terms of accessibility to their respective representational concepts.

Accordingly, the bottom-up process of abstracting text into mentalised situations, via multiple parallel representational stages inherent within the CI architecture, provides an elegant explanation of how the negation of verbs, particularly within syntactically explicit clauses, might produce the observed effect of faster responses to probe words, while still being consistent with evidence of nominal negation (where the message-level representation remains intact) showing the reverse behavioural pattern (see MacDonald & Just, 1989).

While findings on event negation in this thesis are explainable by considering representations in the first and final pathways of the CI model, findings on distributivity are equally explainable by the second and final levels of this system: the macrostructure and the situation model.

Chapter 3 revealed, in line with previous research on ambiguously distributive constructions, that comprehenders, on average, commit to a collective interpretation – as evidenced in the post-narrative naturalness ratings – but this was not the case while they read through the narrative itself. While reading, comprehenders appeared to access both collective and distributive representations of the same text regardless of the plurality of the critical continuation, with reading times suggesting that both interpretations were acceptable following ambiguously distributive events but not non-distributive ones. The macrostructural route of representation construction in the CI model specifically allows for multiple, potentially competing, interpretations of the larger situation being described (i.e., both collective and distributive versions of the text). Commitment to an interpretation is reached when the connectionist network underlying the system, following a process of spreading activation, rejects the inappropriate interpretations that have been constructed; multiple competing propositions at that macrostructural level are integrated with top-down inferences at the level of the situation model, whereupon a single interpretation is selected (Kintsch et al., 1990). That both representations are accessible to plural and non-plural continuations in ambiguously distributive texts suggests that comprehenders are tapping into this rich pool of potential representations at the macrostructural level, while the naturalness ratings (which show a preference for a collective reading) are indicative of comprehenders committing to a single interpretation when prompted, which they access from their abstracted situation model of the text. The representational system of the CI model thus offers a simple, elegant explanation for both the findings on negation and distributivity observed throughout this thesis.

The CI model as an explanatory account of the observed findings does face a small problem, though: its emphasis on propositions, predicates, and arguments as the machinery of representational structure, as opposed to events. As discussed in chapter 1, events are now widely considered to be the fundamental unit of situation models (Ferstl et al., 2005; Huff et al., 2014; Kuperberg, 2016; Kuperberg & Jaeger, 2016; Zacks, 2020; Zwaan, 1996; Zwaan, 2016; Zwaan & Radvansky, 1998; *inter alia*). In fact, negation and distributivity are formally analysable with respect to events, as in (31), repeated in (44), below, adapted from Champollion (2011), and in (45), adapted from Champollion (2016b).

44) John didn't laugh.

$\neg\exists e$ [LAUGH(e) & AGENT(e, j)]

'There is no event in which John laughed'

45) 'Buy' in… John and Mary bought a book.

$\forall e$[BUY(e) $\rightarrow e \in \lambda e'$[BUY($e'$) & Atom($\theta(e')$)]][36]

'For all buying events *e*, that event consists of one or more buying events, *e*', whose thematic roles (θ) are mereological atoms'

An event-based construction-integration model, that maintains the notion of the three discrete, but parallel, representational pathways, would provide a compelling theoretical account of the data obtained in this thesis, as well as offer valuable insights into the mechanisms underlying event numeration during situation model construction. No existing theory, to my knowledge, has precisely captured the parallel representational system of the CI model with a detailed, bottom-up account of how events encoded in the linguistic input are abstracted and combined to produce aggregate representations of the situation that builds upon our semantic knowledge of the compositional meanings of event-based expressions.

One empirically successful attempt at an event-based construction-integration model, though, is the Event-Indexing model (Zwaan, 2016; Zwaan et al., 1995a; Zwaan & Radvansky, 1998), which claims that situation model construction proceeds along three similar, although not identical, phrases to the CI model: a current event model, where comprehenders interpret the events of a particular clause in their input; an integrated model that is the global amalgam of

[36] This formulation can be generalised to $||D|| = \lambda V\, \lambda e[e \in \lambda e'[V(e') \,\&\, \text{Atom}(\theta(e'))]]$, where *V* is some verb, and D is a putative Distributivity operator in the semantics, roughly equivalent in meaning to the English adverb *each*.

all pre-existing representations relevant to that input; and a complete model, consisting of the sum of current and integrated models. Changes in discourse factors such as the time events occur, or the protagonists involved in those events, leads to the development of a new current model, and the updating of the complete and integrated ones. This remains a continuous, parallel process throughout the comprehension of a text.

The difference in phases, or pathways, between the EI and CI models reflects their difference in perspective on situation model construction. The EI model focusses on representational update via comprehenders' monitoring of changes in the dimensions (or indices) of events, relating to, but not limited to, time, location, protagonist, causation, and intention, while the CI model emphasises how the meaning of words and phrases are extracted from the input and translated into discourse-level representations. The EI model is, therefore, not overtly preoccupied with the same kinds of bottom-up mechanisms which translate literal texts to mental representations found in the CI model, and thus the EI model does not exactly satisfy the requirements of the idealised, bottom-up, explanatory model recommended above, nor explains the results of the above chapters – in terms of surface-level and ambiguous representations – as effectively as the CI model does.

However, with its emphasis on higher-level representational factors in the construction of events like temporality, causality and changes in the objects or entities under discussion, the EI model is suited to a very different kind of explanation of the data obtained in this thesis. Instead of describing how negation suppresses events, or how plurality denoting expressions create simultaneous collective and distributive representations of the same text, the EI model provides a framework for talking about what comprehenders – or at least the language processing mechanisms that drive them – are attending to in a narrative with respect to the events that comprise that narrative. Vitally, the central theme of this thesis has been that comprehenders extract semantic information from their linguistic input – including when the events encoded by that input are modified by verbal negation and plurality denoting expressions – to determine whether a single event, multiple events, or no events occurred. In other words, comprehenders count events. For all the precision of the CI model in explaining the data above, that implication does not naturally fall out of the model. It is made explicit, however, under the framework of the EI model – assuming, crucially, that one of the higher-level indices (or dimensions) relating to events that comprehenders track in the construction of new event models is number.

While five indices (or dimensions) have been suggested, these were never intended to be an exhaustive list (see Zwaan & Radvansky, 1998). The contention of this thesis is that comprehenders, when constructing mental representations of the situation described, also track the dimension of number. This seems reasonable when talking about the lack of bread or the inclusion of photographs in examples like (26) and (29), repeated as (46) and (47), as well as when determining the number of recipients to represent in (35), repeated as (48), below.

46) Mary baked some cookies but no bread.

47) Elizabeth burned the letters but not the photographs.

48) Rebecca gave a cake to two friends.

But it is not just matters of protagonist that intersect with number. As the findings of chapter 4 demonstrate, time, and expectations around events unfolding in time, intersect with number, too. Where an event did not occur at a critical point in an otherwise contiguous narrative – whether in a syntactically explicit or elided clause – this breaks the assumed chronology, leading to a prediction error effect in the averaged ERPs to a contiguity enforcing adverbial; where an event in an otherwise contiguous narrative is ambiguous, either between a collective and distributive reading, or temporally ambiguous in its distributivity, integration difficulties are observed, reflected in an ERP component typical of reanalysis. How comprehenders track number with respect to events seems to lead to different event models relating to the narrative situations being constructed, resulting in the reported violations and integrations difficulties of chapter 4. When supplemented alongside existing research on event representations, these data endorse the conclusion that during situation model construction comprehenders monitor multiple event-related factors, including time, location, protagonist, causation, and intention – as well as, importantly, number.

In summary, the two focal theories of situation model construction adopted in this thesis (the CI and EI models) each provide different perspectives on the data obtained in the course of its completion, and offer potential explanations for the observed results, from different levels of representational structure being accessed at specific points in the input, to what the comprehenders' processor might be monitoring in its determination of whether to construct a new event model, a new set of event models, or no event model at all. The explanations offered by the CI and EI frameworks admittedly remain somewhat speculative, though, as more research on negation and distributivity needs to be conducted to improve confidence in these claims. A further point of consideration is that, either due to an intentional lack of emphasis on bottom-up mechanisms or limited application of event semantic concepts in its machinery, neither of these models account for all aspects of the complex process that is situation model construction, or account for the present data in its entirety, such that a combination of these approaches was adopted to explain the results of this thesis. Other perspectives and paradigms, such as the Event Horizon model (Radvanksy, 2012; Radvansky & Zacks, 2017) or the more formally semantic Discourse Representation Theory (Kamp, 1981; Kamp & Reyle, 1993), among others, continue to generate fruitful deductions about representations of language at levels beyond the word or phrase, and, although these models lay beyond the scope of this project, they might have equally insightful things to say about event numeration, too.

5.4 – Looking Ahead

Throughout this thesis, several suggestions were made regarding possible future directions for research which builds upon the findings – and sometimes the oddities – observed in the central three chapters. These include: the marginally significant trend in the direction of a facilitation effect in negated elided contexts to probe words; the difference between on-line and off-line measures of distributivity, as well as the somewhat bizarre observation of a preference toward distributive-consistent continuations following a non-distributive context in the present perfect; and both the latency difference of the N400 to explicit and elided negation, and the pronounced negativity in negated contexts around 800ms post-stimulus that could be indicative of later stages of a P600 component which failed to reach significance in the reported sample. In what follows, additional recommendations for further research that incorporates the non-linguistic knowledge discussed in chapter 1, as well as suggestions that exclude it, are put forth.

5.4.1 – To Knowledge and Simulations

Early in this project, certain elements of representational composition were considered, reflected in the purported formula: [event representations + encyclopaedic knowledge + embodied knowledge]. While attention throughout this thesis has prioritised the first element within this formulation, adopting the perspective of the other two has awarded researchers constructive data that has pushed on the boundaries of comprehension science (see section 1.3 for discussion). Both negation and distributivity, while looked at solely through the lens of event representation in this work, could equally be analysed from the perspective of knowledge, embodied or encyclopaedic, and its impact on situation model construction. For example, as mentioned in chapter 2, further work on negation could consider the role of inferences that prevent the suppression of an event, in sentences like (33), repeated as (49), where world knowledge relating to driving and the nature of traffic lights would integrate with the semantic information encoded in the linguistic expression and be interpreted as something like *Paul drove through the red light* – a non-suppressed event. Our understanding of the representations of non-suppressed negated events is indisputably in need of further study.

49) Paul did not stop at the red light.

Additional efforts in the field could examine whether the fMRI effects in areas of the brain associated with motor and perceptual representations following high-imagery clauses, (Kurby & Zacks, 2013; discussed in section 1.3.2) persist when those clauses – and thus the events they convey – are negated. Similarly, verbal negation and thus the suppression of the event could influence behaviours that have evidenced the importance of mental simulations in language comprehension, like the directional movement of a response showing differential latencies when matching or mismatching the described situation, such as the openings and closings of drawers (see Glenberg & Kaschak, 2002; discussed in section 1.3.2). If events are suppressed – and sufficient time is provided for negation to integrate with the meaning of these expressions before the measurement, as with the present study (see Kaup et al. (2006) and Orenes (2021) for a discussion of the two-step process of negation integration) – these sensorimotor effects are predicted to disappear following negated contexts.

Distributivity could also be explored from the perspective of knowledge. While the experimental conditions in chapters 3 and 4 were intentionally constrained to be compatible with both collective and distributive readings, not every event is pragmatically compatible with an atomic or group interpretation. For example, consider (50) and (51).

50) Ten women lifted the five-tonne machine.

51) John and Mary passed their driving test.

Based on world knowledge relating to weight and human strength, as well as how driving exams typically unfold, there is a considerably lower likelihood of a distributive interpretation than a collective one in (50), with the reverse being so in (51). However, as in Nieuwland and van Burkum (2006), who showed that context – realistic or otherwise – can realign expectations of acceptable discourse, such as peanuts falling in love, context could be experimentally manipulated to test whether altering comprehenders preconceptions about these situations can constrain the distributivity interpretation that is committed to, or that is active during processing. Suggestions like these show that our current research into sentences with verbal negation and plurality denoting expressions have only begun to scratch the surface of our understanding of mental model construction.

5.4.2 – To Language

A pillar and, arguably, a strength of this thesis has been its focus on the linguistic (as opposed to embodied) contributions to event representations. By intentionally sidelining elements of situation model construction relating to perceptual experience and simulations that have been prolific and informative in mental model science over the last couple of decades, the present research acted as something of a canary-in-the-mine: can emphasising just the role of understudied grammatical constructions, like verbal negation and plurality denoting expressions, on mental representations to the exclusion of embodied cognition still be empirically and theoretically fruitful? I believe this thesis and the results it reports have demonstrated that the answer is yes. While research into the role of embodiment and mental simulations goes on and will likely continue to make significant contributions to our knowledge, with possible suggestions for further inquiry in these areas outlined in the previous section, linguists, who may have little interest in non-linguistic factors like sensorimotor experience, nonetheless possess a wellspring of tools via understudied grammatical constructions that can and should be used to advance our understanding of situation modelling. The science of mental representations can blossom, even without the inclusion of simulations. Here are six suggestions for continued efforts in this promising enterprise:

a) Related to distributivity, and unexplored in this thesis, is the concept of cumulativity. Cumulativity is understood as an event which ranges over and is distributive upon two separate plurality denoting arguments, as in *three boys saw two girls*, leading to an interpretation in which each of the three boys saw at least one of the girls, and each of the girls was seen by at least one of the boys (see Champollion, 2020). Cumulative representations have not been tackled here but present an intriguing opportunity to build upon our knowledge of ambiguity processing via plurality denoting expressions.

b) Negated events present some interesting challenges not widely examined. In a sentence like *John didn't laugh for two hours*, what was it that occurred for the duration described by the *for*-adverbial; if no event is mentally represented, what is it that the duration of time relates to? Are negated events still conceptually represented as some span of time? Is 'not laughing' itself an event – a negative event (see Bernard, 2018)? These questions and more inspire deeper analyses of representations of negated events.

c) With events understood to be changes in the states of objects (Altmann & Ekves, 2019), representational restorations of that original state following such a change, as with the restitutive uses of *again* and *back* in sentences like *Jasmine opened the door and then slammed it shut again / back shut* (see Iyer, 2022), demand to be examined with more experimental scrutiny as part of efforts to better determine the relationship between objects, changes in their states, and the events they aggregate into.

d) Not all events are introduced by verbs. Nouns such as *destruction* in the phrase *the destruction of the city* and *nap* in *a nap* are understood not to be entities but descriptions of events. A deeper examination of events introduced via non-verbal predicates, as well as potential representational differences between verbs and non-verbal events, especially during early child language acquisition and the development of mental representations – see He & Wittenberg (2019) for a recent project – is encouraged.

e) Looking at states inside relative clauses, Marx and Wittenberg (2022) examined back- and forward-shifted temporal interpretations in sentences like *the girl fed the rabbit that was near the mushroom*, finding that forward-shifted interpretations (where the subordinate was treated as being after the matrix event in a conceptualised chronology of the situation) were significantly less acceptable to comprehenders than back-shifted interpretations (where the subordinate was treated as occurring before the matrix event), suggesting that comprehenders anchor temporal expressions by relating them to another temporal interval, with contextually salient events more likely to be chosen as that anchor. An extension of this research, in which relative clauses contain dynamic events, as in *Rebecca fed the rabbit that was chased by the cat* would advance our understanding of the relationship between multiple events. Such an approach would even open the door for additional work on event numeration and the temporal anchoring of negated and distributive constructions: *Rebecca didn't feed the rabbit that was chased by the cat*; *John and Mary each fed the rabbit that was chased by the cat*.

f) Finally, much of the research into event representations and situation model construction has heavily relied on English (or western European languages), with few of the thousands of extant languages throughout the world being considered (Ünal et al., 2021), a limitation that also affects this thesis. Consideration of linguistic structures like serial verb constructions – in which two or more (potentially event-introducing) verbs combine within a single clause that share no morphological marking to indicate subordination or coordination, and which are independently affected by features like negation (Lovestrand, 2021) – are profoundly understudied in cognitive science and thus may offer substantive insights into event processing and representational phenomena that are obscured by focussing on almost-exclusively western European varieties. Similarly, research on understudied modalities, such as sign languages, which are frequently iconic and pluractional (i.e., they entail the existence of a multitude of events via a reduplication of verbal forms; see Kuhn and Aristodemo (2017) for discussion), promises to illuminate unknown aspects of event representational construction and retrieval during comprehension.

While this list is far from exhaustive, the research conducted throughout this thesis demonstrates that efforts to contribute to our knowledge of situation model construction without relying on the informative – and fashionable – role of mental simulations remains insightful and much needed. In looking ahead to future research, this chapter has made an assortment of recommendations for possible studies which examine psycholinguistically underrepresented grammatical structures that integrate with the elements of encyclopaedic and embodied knowledge in situation model construction, while simultaneously emphasising the value of continued work that omits them. Through either approach, the future of mental model research looks bright.

5.5 – Concluding Remarks

With language we tell stories. The story of this thesis is that, during language processing, comprehenders enumerate the number of events (0, 1, >1) encoded by the input, through verbal negation and plurality denoting expressions, to mentally represent the situation described. Opening with a discussion of the psychological reality of situation models, this project established that event representations are their fundamental unit, but that most of the existing literature on the subject had examined single events, or sequences of single events, with little consideration given to the complexities of negated events or ambiguously distributive events despite the prolific use of expressions that give rise to these meanings among speakers of the English language. The functions of world knowledge and mental simulations, which have offered valuable insights on the construction of discourse representations, were explored, but a constrained model of representation which excluded these factors was adopted to maintain simplicity while analysing the effects of the selected topic: event numeration. This research was framed within a comparison of two influential theories of situation modelling, known as the Construction-Integration (CI) and Event-Indexing (EI) models.

Chapters 2, 3 and 4 examined the effects of verbal negation and ambiguously distributive constructions on event representational access and comprehension processes through measures of response times to probe words, reading times to distributive-consistent or -inconsistent continuations, and averaged ERPs to critical adverbials which prescribed a sense of temporal contiguity with the preceding negated or distributive clauses. While consideration was given to the potential limitations of the online data collection process, particularly relating to MTurk users, confidence in the validity and reliability of the results were maintained through the application of careful methods which mitigated the impact of participants performing in bad faith. Confidence was further improved via the overlap of results between the in-person and online chapters, which together demonstrate that verbal negation suppresses event representations, although the effect may be different in syntactically explicit and elided clauses, and that both collective and distributive representations are active (or accessible) following ambiguously distributive texts, even though a collective interpretation is ultimately reached.

The implications of these findings were reviewed from the perspectives of the CI and EI models, and an explanation for the observed effects that drew upon representational structure at a surface-level, a macrostructural semantic level, and an integrated situation level was proposed; these findings were also generalised to high-level event-based indices, whereupon it was argued that part of situation model construction is comprehenders' attentiveness to the *number* of events within a narrative. Suggestions for future research that incorporate world knowledge and mental simulations in negation and distributivity processing, building upon the current findings, were recommended, while additional proposals for continued research that adopts a simplified representational model, wherein non-linguistic factors are sidelined, were outlined to stress the value of language-based contributions on the mental representation of events. In sum, the story of this project is one in which several original contributions to our understanding of events, both empirical and theoretical, have been motivated by looking at verbal negation and plurality denoting expressions from a processing perspective, with this thesis acting as a vanguard to future efforts to tell data-driven stories about understudied linguistic structures on situation model construction.

References

Acunzo, D. J., MacKenzie, G., & van Rossum, M. C. (2012). Systematic biases in early ERP and ERF components as a result of high-pass filtering. *Journal of neuroscience methods*, *209*(1), 212-218.

Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk research: Review and recommendations. *Journal of Management*, *47*(4), 823-837.

Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon*. (4th ed.) John Wiley & Sons.

Aker, A., El-Haj, M., Albakour, M. D., & Kruschwitz, U. (2012). *Assessing Crowdsourcing Quality through Objective Tasks.* Paper presented at the 8th International Conference on Language Resources and Evaluation, Istanbul, Turkey.

Albert, D. A., & Smilek, D. (2023). Comparing attentional disengagement between Prolific and MTurk samples. *Scientific Reports*, *13*(1), 20574.

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of experimental psychology: Learning, memory, and cognition*, *19*(5), 1061.

Altmann, G., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review*, *126*(6), 817.

Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive science*, *33*(4), 583-609.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive psychology*, *6*(4), 451-474.

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior research methods*, *53*, 1407-1425.

Aravena, P., Delevoye-Turrell, Y., Deprez, V., Cheylus, A., Paulignan, Y., Frak, V., & Nazir, T. (2012). Grip force reveals the context sensitivity of language-induced motor activity during "action words" processing: Evidence from sentential negation. *PLoS One*, *7*(12), e50287.

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior research methods*, *45*, 527-535.

Baetens, K., Van der Cruyssen, L., Vandekerckhove, M., & Van Overwalle, F. (2014). ERP correlates of script chronology violations. *Brain and cognition*, *91*, 113-122.

Bailey, H. R., Kurby, C. A., Sargent, J. Q., & Zacks, J. M. (2017). Attentional focus affects how events are segmented and updated in narrative reading. *Memory & cognition*, *45*, 940-955.

Ball, D., & Rabern, B. (2018). *The science of meaning: Essays on the metatheory of natural language semantics*. Oxford University Press.

Bar-Lev, Z., & Palacas, A. (1980). Semantic command over pragmatic priority. *Lingua*, *51*(2-3), 137-146.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and brain sciences*, *22*(4), 577-660.

Becker, R. B., Ferretti, T. R., & Madden-Lombardi, C. J. (2013). Grammatical aspect, lexical aspect, and event duration constrain the availability of events in narratives. *Cognition*, *129*(2), 212-220.

Begg, I., & Wickelgren, W. A. (1974). Retention functions for syntactic and lexical vs semantic information in sentence recognition memory. *Memory & Cognition, 2,* 353-359.

Beltrán, D., Morera, Y., García-Marco, E., & Vega, M. D. (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.

Bergen, B., & Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain and language*, *112*(3), 150-158.

Berger, N. I., & Coch, D. (2010). Do u txt? Event-related potentials to semantic anomalies in standard and texted English. *Brain and Language*, *113*(3), 135-148.

Bernard, T. (2018). Negation in event semantics with actual and nonactual events. In *ConSOLE 2018-26th Conference of the Student Organization of Linguistics in Europe*, 350-366.

Bernard, T., & Champollion, L. (2018). Negative events in compositional semantics. In *Semantics and Linguistic Theory* (Vol. 28, pp. 512-532).

Bohnemeyer, J., Enfield, N. J., Essegbey, J., Ibarretxe-Antuñano, I., Kita, S., Lüpke, F., & Ameka, F. K. (2007). Principles of event segmentation in language: The case of motion events. *Language*, 495-532.

Bohnemeyer, J., Enfield, N. J., Essegbey, J., & Kita, S. (2011). The macro-event property. *Event representation in language and cognition*, 43-67.

Bohnemeyer, J., & Pederson, E. (2010). *Event representation in language and cognition* (Vol. 11). Cambridge University Press.

Bornkessel, I. D., Fiebach, C. J., & Friederici, A. D. (2004). On the cost of syntactic ambiguity in human language comprehension: An individual differences approach. *Cognitive Brain Research*, *21*(1), 11-21.

Brand, A., & Bradley, M. T. (2012). Assessing the effects of technical variance on the statistical outcomes of web experiments measuring response times. *Social Science Computer Review*, *30*(3), 350-357.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive psychology*, *3*(2), 193-209.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of verbal learning and verbal behavior*, *11*(6), 717-726.

Brewer, W. F., & Lichtenstein, E. H. (1980). Event schemas, story schemas, and story grammars. *Center for the Study of Reading Technical Report; no. 197*.

Briner, S. W., Virtue, S., & Kurby, C. A. (2012). Processing causality in narrative events: Temporal order matters. *Discourse Processes*, *49*(1), 61-77.

Britton, B. K., & Gülgöz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of educational Psychology*, *83*(3), 329.

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

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

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45-50.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?. *Perspectives on psychological science*, *6*(1), 3-5.

Champollion, L. (2011). Quantification and negation in event semantics. *Baltic International Yearbook of Cognition, Logic and Communication*, *6*(1), 3.

Champollion, L. (2016a). Overt distributivity in algebraic event semantics. *Semantics and Pragmatics*, *9*, 16-1.

Champollion, L. (2016b). Covert distributivity in algebraic event semantics. *Semantics and Pragmatics*, *9*, 15-1.

Champollion, L. (2016c). Ten men and women got married today: Noun coordination and the intersective theory of conjunction. *Journal of Semantics*, *33*(3), 561-622.

Champollion, L. (2019). Distributivity in formal semantics. *Annual review of linguistics*, *5*, 289-308.

Champollion, L. (2020). Distributivity, collectivity and cumulativity. In L. Matthewson, C. Meier, H. Rullmann, & T. E, Zimmermann (Eds.), *The Wiley Blackwell companion to semantics*. Wiley: Hoboken.

Cleland, C. (1991). On the individuation of events. *Synthese*, 229-254.

Clifton Jr, C., & Frazier, L. (2012). Interpreting conjoined noun phrases and conjoined clauses: Collective versus distributive preferences. *Quarterly Journal of Experimental Psychology*, *65*(9), 1760-1776.

Cloitrew, M., & Bever, T. G. (1988). Linguistic anaphors, levels of representation, and discourse. *Language and cognitive processes*, *3*(4), 293-322.

Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems* (Vol. 2). Cambridge university press.

Cooper, R. (1986). Tense and discourse location in situation semantics. *Linguistics and philosophy*, 17-36.

Craik, K. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS one*, *8*(3), e57410.

Davidson, D. (1967). The logical form of action sentences. In In Nicholas Rescher (ed.), *The logic of decision and action* (pp. 81-95). Pittsburgh: University of Pittsburgh Press.

Davidson, D. (1969). The individuation of events. In *Essays in honor of Carl G. Hempel: A tribute on the occasion of his sixty-fifth birthday* (pp. 216-234). Dordrecht: Springer Netherlands.

Davoudi, M., & Moghadam, H. R. H. (2015). Critical review of the models of reading comprehension with a focus on situation models. *International Journal of Linguistics*, *7*(5), 172-187.

Degen, J., Kursat, L., & Leigh, D. D. (2021). Seeing is believing: Testing an explicit linking assumption for visual world eye-tracking in psycholinguistics. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43).

de Lafuente, V., & Romo, R. (2004). Language abilities of motor cortex. *Neuron*, *41*(2), 178-180.

de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, *48*, 1-12.

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.

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

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

Delorme, A. (2023). EEGLAB (https://github.com/sccn/eeglab), GitHub.

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.

Dery, J. E., & Koenig, J. P. (2015). A narrative-expectation-based approach to temporal update in discourse comprehension. *Discourse Processes*, *52*(7), 559-584.

de Vega, M., Morera, Y., León, I., Beltrán, D., Casado, P., & Martín-Loeches, M. (2016). Sentential negation might share neurophysiological mechanisms with action inhibition. Evidence from frontal theta rhythm. *Journal of Neuroscience*, *36*(22), 6002-6010.

de Vega, M., Robertson, D. A., Glenberg, A. M., Kaschak, M. P., & Rinck, M. (2004). On doing two things at once: Temporal constraints on actions in language comprehension. *Memory & cognition*, *32*(7), 1033-1043.

Dotlacil, J., & Brasoveanu, A. (2021). The representation and processing of distributivity and collectivity: ambiguity vs. underspecification. *Glossa*, *6*(1), 1-22.

Dowty, D. R. (1986). The effects of aspectual class on the temporal structure of discourse: semantics or pragmatics?. *Linguistics and philosophy*, 37-61.

Dowty, D. (1991). Thematic proto-roles and argument selection. *language*, *67*(3), 547-619.

Dretske, F. I. (1967). Can events move?. *Mind*, *76*(304), 479-492.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of memory and language*, *27*(4), 429-446.

Duran, N. D., McCarthy, P. M., Graesser, A. C., & McNamara, D. S. (2007). Using temporal cohesion to predict temporal coherence in narrative and expository texts. *Behavior Research Methods*, *39*, 212-223.

Eckardt, R. (2000). Causation, contexts, and event individuation. In J. Higginbotham, F. Pianesi, A. C. Varzio (Eds.), *Speaking of Events* (pp. 127-143). New York: Oxford University Press.

Ehrlich, K., & Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of verbal learning and verbal behavior*, *21*(3), 296-306.

Elman, J. L., & McRae, K. (2019). A model of event knowledge. *Psychological Review*, *126*(2), 252.

Enochson, K., & Culbertson, J. (2015). Collecting psycholinguistic response time data using Amazon Mechanical Turk. *PloS one*, *10*(3), e0116946.

Evans, V. (2009). *How words mean: Lexical concepts, cognitive models, and meaning construction*. Oxford University Press.

Farshchi, S., Andersson, A., van de Weijer, J., & Paradis, C. (2021). Processing sentences with sentential and prefixal negation: An event-related potential study. *Language, Cognition and Neuroscience*, *36*(1), 84-98.

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

Ferstl, E. C., Rinck, M., & Cramon, D. Y. V. (2005). Emotional and temporal aspects of situation model processing during text comprehension: An event-related fMRI study. *Journal of cognitive Neuroscience*, *17*(5), 724-739.

Fillmore, C. J. (2006). Frame semantics. *Cognitive linguistics: Basic readings*, *34*, 373-400.

Fitz, H., & Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, *111*, 15-52.

Folli, R., & Harley, H. (2006). What language says about the psychology of events. *Trends in cognitive sciences*, *10*(3), 91-92.

Ford, J. B. (2017). Amazon's Mechanical Turk: a comment. *Journal of Advertising*, *46*(1), 156-158.

Frazier, L., Pacht, J. M., & Rayner, K. (1999). Taking on semantic commitments, II: collective versus distributive readings. *Cognition*, *70*(1), 87-104.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology*, *14*(2), 178-210.

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of psycholinguistic research*, *31*, 45-63.

Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition*, *85*(3), B83-B92.

Gernsbacher, M. A., Goldsmith, H. H., & Robertson, R. R. (1992). Do readers mentally represent characters' emotional states?. *Cognition & Emotion*, *6*(2), 89-111.

Gernsbacher, M. A., & Jescheniak, J. D. (1995). Cataphoric devices in spoken discourse. *Cognitive psychology*, *29*(1), 24-58.

Glaeser, G., & Strouhal, E. (2000). Von Kempelen's Chess-Playing Pseudo-Automaton and Zu Racknitz's Explanation of Its Controls, in 1789, In M. Ceccarelli (Eds.) *International Symposium on History of Machines and Mechanisms – Proceedings HMM 2000* (pp. 351-360). Springer, Dordrecht.

Glenberg, A. M. (1997). Mental models, space, and embodied cognition. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 495–522). American Psychological Association.

Glenberg, A. M., & Kaschak, M. P. (2003). The body's contribution to language. *Psychology of learning and motivation*, *43*, 93-126.

Gibson, E., & Pearlmutter, N. J. (2000). Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research*, *29*, 231-240.

Giora, R. (2007). “A good Arab is not a dead Arab incitement": On the accessibility of negated concepts. *Explorations in pragmatics: Linguistic, cognitive, and intercultural aspects*, *1*, 129.

Giora, R., Balaban, N., Fein, O., & Alkabets, I. (2004). Negation as positivity in disguise. In *Figurative language comprehension* (pp. 233-258). Routledge.

Giora, R., Fein, O., Aschkenazi, K., & Alkabets-Zlozover, I. (2007). Negation in context: A functional approach to suppression. *Discourse Processes*, *43*(2), 153-172.

Goodman, N. (2012). *The Structure of Appearance* (3rd ed.). Dordretch: Reidel Publishing Company.

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.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1-3), 335-346.

Harris, J. A., Clifton, C., Jr., & Frazier, L. (2013). Processing and domain selection: Quantificational variability effects. *Language and Cognitive Processes, 28*(10), 1519-1544.

Hauser, D., Paolacci, G., & Chandler, J. (2019). Common concerns with MTurk as a participant pool: Evidence and solutions. In *Handbook of research methods in consumer psychology* (pp. 319-337). Routledge.

Havik, E., Roberts, L., Van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning*, *59*(1), 73-112.

He, A. X., & Wittenberg, E. (2020). The acquisition of event nominals and light verb constructions. *Language and Linguistics Compass*, *14*(2), e12363.

He, X., Yang, H., Zhang, W., Zhao, X., & Xie, Y. (2015). Dynamic Updating Process of Readers’ Temporal Situation Model: From Short-term Working Memory to Long-term Working Memory. *Acta Psychologica Sinica*, *47*(10), 1235.

Hilbig, B. E. (2016). Reaction time effects in lab-versus Web-based research: Experimental evidence. *Behavior Research Methods*, *48*, 1718-1724.

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

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of cognitive neuroscience*, *19*(7), 1175-1192.

Horchak, O. V., & Garrido, M. V. (2020). Is complex visual information implicated during language comprehension? The case of cast shadows. *Cognitive Science*, *44*(7), e12870.

Huff, M., Meitz, T. G., & Papenmeier, F. (2014). Changes in situation models modulate processes of event perception in audiovisual narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1377.

Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of phonetics*, *39*(4), 571-584.

Iyer, J. (2022). Back to restitutive readings again. In *Proceedings of SALT 32*, 599-618.

Jachmann, T. K., Drenhaus, H., Staudte, M., & Crocker, M. W. (2019). Influence of speakers' gaze on situated language comprehension: evidence from event-related potentials. *Brain and cognition*, *135*, 103571.

Järvikivi, J., Pyykkönen, P., & Niemi, J. (2009). Exploiting degrees of inflectional ambiguity: Stem form and the time course of morphological processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 221.

Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 20-49). New York: Routledge.

Ji, Y., & Papafragou, A. (2022). Viewers Spontaneously Represent Event Temporal Structure. In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society*. (Vol. 44, No. 44).

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (No. 6). Harvard University Press.

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in cognitive sciences*, *5*(10), 434-442.

Johnson-Laird, P. N. (2004). The history of mental models. *Psychology of reasoning: Theoretical and historical perspectives*, *8*, 179-212.

Kaan, E. (2023). Investigating Adaptation to And-Coordination in English: An ERP Study. *Languages*, *8*(3), 207.

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.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of cognitive neuroscience*, *15*(1), 98-110.

Kaiser, E. (2019). Order of mention in causal sequences: Talking about cause and effect in narratives and warning signs. *Discourse Processes*, *56*(8), 599-618.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, *2*(4), 647-670.

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

Kamp, H. (1981). A theory of truth and semantic representation. *Formal semantics-the essential readings*, 189-222.

Kamp, H. & Reyle, U. (1993). *From Discourse to Logic*. Dordrecht: Kluwer.

Kang, X., Eerland, A., Joergensen, G. H., Zwaan, R. A., & Altmann, G. T. (2020). The influence of state change on object representations in language comprehension. *Memory & cognition*, *48*, 390-399.

Kaup, B. (2001). Negation and its impact on the accessibility of text information. *Memory & cognition*, *29*(7), 960-967.

Kaup, B., & Dudschig, C. (2020). Understanding negation: Issues in the processing of negation. In V. Déprez & T. Espinal (Eds.), *The Oxford Handbook of Negation* (pp. 635–655). Oxford University Press, USA.

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.

Kaup, B., & Zwaan, R. A. (2003). Effects of negation and situational presence on the accessibility of text information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 439.

Kaup, B., Zwaan, R. & Lüdtke, J. (2013). 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). London: Lawrence Erlbaum Associates.

Khanna, M. M., & Boland, J. E. (2010). Children's use of language context in lexical ambiguity resolution. *Quarterly Journal of Experimental Psychology*, *63*(1), 160-193.

Kuhn, J., & Aristodemo, V. (2017). Pluractionality, iconicity, and scope in French Sign Language. *Semantics and Pragmatics*, *10*(6), 1-49.

Kim, J., Gabriel, U., & Gygax, P. (2019). Testing the effectiveness of the Internet-based instrument PsyToolkit: A comparison between web-based (PsyToolkit) and lab-based (E-Prime 3.0) measurements of response choice and response time in a complex psycholinguistic task. *PloS one*, *14*(9), e0221802.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological review*, *95*(2), 163.

Kintsch, W. (2018). Revisiting the construction—integration model of text comprehension and its Implications for Instruction. In D. Alvermann, N. J. Unrau, M. Sailors, & R. B Ruddell (Eds.), *Theoretical models and processes of literacy* (pp. 178-203). Routledge.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and language*, *29*(2), 133-159.

Kliegl, R., Dambacher, M., Dimigen, O., Jacobs, A. M., & Sommer, W. (2012). Eye movements and brain electric potentials during reading. *Psychological research*, *76*, 145-158.

Knott, A., & Takac, M. (2021). Roles for event representations in sensorimotor experience, memory formation, and language processing. *Topics in Cognitive Science*, *13*(1), 187-205.

Kok, K., & Cienki, A. (2017). Taking simulation semantics out of the laboratory: Towards an interactive and multimodal reappraisal of embodied language comprehension. *Language and Cognition*, *9*(1), 1-23.

Kolk, H. H., Chwilla, D. J., Van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and language*, *85*(1), 1-36.

de Koning, B. B., Wassenburg, S. I., Bos, L. T., & Van der Schoot, M. (2017). Size does matter: Implied object size is mentally simulated during language comprehension. *Discourse processes*, *54*(7), 493-503.

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, cognition and neuroscience*, *31*(5), 602-616.

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

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

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive brain research*, *17*(1), 117-129.

Kurby, C. A., & Zacks, J. M. (2012). Starting from scratch and building brick by brick in comprehension. *Memory & cognition*, *40*, 812-826.

Kurby, C. A., & Zacks, J. M. (2013). The activation of modality-specific representations during discourse processing. *Brain and language*, *126*(3), 338-349.

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

Kutas, M., & Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *108*(5), 456-471.

Kutas, M., & Federmeier, K. D. (2009). N400. *Scholarpedia*, *4*(10), 7790.

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

Lea, R. B., & Mulligan, E. J. (2002). The effect of negation on deductive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(2), 303.

Ledoux, K., Camblin, C. C., Swaab, T. Y., & Gordon, P. C. (2006). Reading words in discourse: The modulation of lexical priming effects by message-level context. *Behavioral and Cognitive Neuroscience Reviews*, *5*(3), 107-127.

Levin, B. (2019). On Dowty's "Thematic Proto-Roles and Argument Selection". In *A Reader's Guide to Classic Papers in Formal Semantics: Volume 100 of Studies in Linguistics and Philosophy* (pp. 103-119). Cham: Springer International Publishing.

Levin, B., & Hovav, M. R. (1992). The perspective from unaccusativity. *Thematic structure: Its role in grammar*, *16*, 247.

Lin, Y., & Matsumi, N. (2022). Visuospatial working memory and the construction of a spatial situation model in listening comprehension: An examination using a spatial tapping task. *Cognitive Processing*, 1-14.

Linnman, C., Carlbring, P., Åhman, Å., Andersson, H., & Andersson, G. (2006). The Stroop effect on the internet. *Computers in human behavior*, *22*(3), 448-455.

Lasersohn, P. (1998). Generalized distributivity operators. *Linguistics and philosophy*, 83-93.

Lasersohn, P. (2013). *Plurality, conjunction and events* (Vol. 55). Springer Science & Business Media.

Lehtonen, M., Hultén, A., Rodríguez-Fornells, A., Cunillera, T., Tuomainen, J., & Laine, M. (2012). Differences in word recognition between early bilinguals and monolinguals: Behavioral and ERP evidence. *Neuropsychologia, 50*(7), 1362-1371.

Lemmon, E. J. (1967). Comments on D. Davidson's "The Logical Form of Action Sentences". In N. Resher (Ed.), *The Logic of Decision and Action*. Pittsburgh, PA: University of Pittsburgh Press, 96–103.

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

Link, G. (1987). Generalized quantifiers and plurals. In P. Gärdenfors (Ed.*), Generalised Quantifiers* (pp. 151-180). Reidel Publishing Company.

Loerts, H., Stowe, L. A., & Schmid, M. S. (2013). Predictability speeds up the re-analysis process: An ERP investigation of gender agreement and cloze probability. *Journal of Neurolinguistics*, *26*(5), 561-580.

Logačev, P., & Vasishth, S. (2016). A multiple-channel model of task-dependent ambiguity resolution in sentence comprehension. *Cognitive Science*, *40*(2), 266-298.

Long, D. L., & Spooner, A. (2010). Placing a text in context. *Psychonomic bulletin & review*, *17*(2), 237-242.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience*, *8*, 213.

Lopukhina, A., Lopukhin, K., & Laurinavichyute, A. (2021). Morphosyntactic but not lexical corpus-based probabilities can substitute for cloze probabilities in reading experiments. *PloS One*, *16*(1), e0246133.

Lovestrand, J. (2021). Serial verb constructions. *Annual Review of Linguistics*, *7*, 109-130.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

Luck, S. J. (2005a). *An Introduction to the Event-Related Potential Technique*. Cambridge, Massachusetts: The MIT Press.

Luck, S. J. (2005b). Ten simple rules for designing and interpreting ERP experiments. In T. C. Handy (Ed.). *Event-related potentials: A methods handbook*. MIT press.

Lüdtke, J., Friedrich, C. K., De Filippis, M., & Kaup, B. (2008). Event-related potential correlates of negation in a sentence–picture verification paradigm. *Journal of cognitive neuroscience*, *20*(8), 1355-1370.

MacDonald, M. C., & Just, M. A. (1989). Changes in activation levels with negation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 633.

Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations?. *Memory & cognition*, *31*(5), 663-672.

Magliano, J. P., & Schleich, M. C. (2000). Verb aspect and situation models. *Discourse processes*, *29*(2), 83-112.

Magliano, J. P., & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive science*, *35*(8), 1489-1517.

Maienborn, C. (2011). Event semantics. In K. von Heusinger & P. Portner, *Semantics* (pp. 232-266). Berlin: de Gruyter.

Mak, M., & Willems, R. M. (2019). Mental simulation during literary reading: Individual differences revealed with eye-tracking. Language, Cognition and Neuroscience, 34 (4), 511–535.

Malaia, E. (2014). It still isn't over: Event boundaries in language and perception. *Language and Linguistics Compass*, *8*(3), 89-98.

Malaia, E., Ranaweera, R., Wilbur, R. B., & Talavage, T. M. (2012). Event segmentation in a visual language: Neural bases of processing American Sign Language predicates. *NeuroImage*, *59*(4), 4094-4101.

Maldonado, M., Chemla, E., & Spector, B. (2017). Priming plural ambiguities. *Journal of Memory and Language*, *95*, 89-101.

Maldonado, M., Chemla, E., & Spector, B. (2019). Revealing abstract semantic mechanisms through priming: The distributive/collective contrast. *Cognition*, *182*, 171-176.

Mandler, J. M. (1986). On the comprehension of temporal order. *Language & Cognitive Processes*, *1*, 309-320.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, *189*(4198), 226-228.

Marx, E., & Wittenberg, E. (2022). Event structure predicts temporal interpretation of English and German past-under-past relative clauses. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, No. 44).

Mathôt, S., & March, J. (2022). Conducting linguistic experiments online with OpenSesame and OSWeb. *Language Learning*, *72*(4), 1017-1048.

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

McClelland, J. L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and cognitive processes*, *4*(3-4), SI287-SI335.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction*, *14*(1), 1-43.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of learning and motivation*, *51*, 297-384.

McNerney, M. W., Goodwin, K. A., & Radvansky, G. A. (2011). A novel study: A situation model analysis of reading times. *Discourse Processes*, *48*(7).

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*(3), 283-312.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of memory and language*, *66*(4), 545-567.

Meylan, S. C., Mankewitz, J., Floyd, S., Rabagliati, H., & Srinivasan, M. (2021). Quantifying lexical ambiguity in speech to and from English-learning children. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43).

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Harvard University Press.

Montalti, M., Calbi, M., Cuccio, V., Umiltà, M. A., & Gallese, V. (2023). Is motor inhibition involved in the processing of sentential negation? An assessment via the Stop-Signal Task. *Psychological Research*, *87*(1), 339-352.

Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 92.

Moss, A. J., Rosenzweig, C., Robinson, J., Jaffe, S. N., & Litman, L. (2023). Is it ethical to use Mechanical Turk for behavioral research? Relevant data from a representative survey of MTurk participants and wages. *Behavior Research Methods*, 1-20.

Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, *395*(6697), 71-73.

Nevatia, R., Hobbs, J., & Bolles, B. (2004). An ontology for video event representation. In *2004 Conference on Computer Vision and Pattern Recognition Workshop* (pp. 119-119). IEEE.

Newman, A., Bavik, Y. L., Mount, M., & Shao, B. (2021). Data collection via online platforms: Challenges and recommendations for future research. *Applied Psychology*, *70*(3), 1380-1402.

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.

Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of cognitive neuroscience*, *18*(7), 1098-1111.

Nuthmann, A., & Van Der Meer, E. (2005). Time's arrow and pupillary response. *Psychophysiology*, *42*(3), 306-317.

O'Bryan, E., Folli, R., Harley, H., & Bever, T. G. (2013). Evidence for the use of verb telicity in sentence comprehension. *Syntax and its limits*, *48*, 80-104.

O'Gara, F. & Lawyer, L. (2019). *Are Iconicity Assumption Violations Prediction Errors?* (Unpublished master's dissertation). University of Essex, Colchester, UK.

Ohtsuka, K., & Brewer, W. F. (1992). Discourse organization in the comprehension of temporal order in narrative texts. *Discourse Processes*, *15*(3), 317-336.

O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse processes*, *43*(2), 121-152.

Orenes, I. (2021). "Looking at" Negation: Faster Processing for Symbolic Rather Than Iconic Representations. *Journal of Psycholinguistic Research*, *50*(6), 1417-1436.

Orenes, I., Beltrán, D., & Santamaría, C. (2014). How negation is understood: Evidence from the visual world paradigm. *Journal of memory and language*, *74*, 36-45.

Otten, M., & Van Berkum, J. J. (2008). Discourse-based word anticipation during language processing: Prediction or priming?. *Discourse processes*, *45*(6), 464-496.

Özyürek, A., Furman, R., & Goldin-Meadow, S. (2015). On the way to language: Event segmentation in homesign and gesture. *Journal of child language*, *42*(1), 64-94.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making*, *5*(5), 411-419.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current directions in psychological science*, *23*(3), 184-188.

Parsons, T. (1990). *Events in the semantics of English: A study in subatomic semantics.* MIT Press,

Partee, B. H., Borschev, V., Paducheva, E. V., Testelets, Y., & Yanovich, I. (2011). Russian genitive of negation alternations: The role of verb semantics. *Scando-Slavica*, *57*(2), 135-159.

Payne, B. R., Lee, C. L., & Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. *Psychophysiology*, *52*(11), 1456-1469.

Perfetti, C. A., Britt, M. A., & Georgi, M. (1995). *Text-based learning and reasoning: Studies in history*. Hillsdale, NJ: Erlbaum.

Pettijohn, K. A., & Radvansky, G. A. (2016a). Walking through doorways causes forgetting: Event structure or updating disruption?. *Quarterly Journal of Experimental Psychology*, *69*(11), 2119-2129.

Pettijohn, K. A., & Radvansky, G. A. (2016b). Narrative event boundaries, reading times, and expectation. *Memory & cognition*, *44*, 1064-1075.

Pettijohn, K. A., Thompson, A. N., Tamplin, A. K., Krawietz, S. A., & Radvansky, G. A. (2016). Event boundaries and memory improvement. *Cognition*, *148*, 136-144.

Piñango, M. M., Zurif, E., & Jackendoff, R. (1999). Real-time processing implications of enriched composition at the syntax–semantics interface. *Journal of psycholinguistic research*, *28*, 395-414.

Piñón, C. (2008). Verbs of creation. In J. Dölling, T. Heyde-Zybatow, & M. Schäfer (Eds.), *Event structures in linguistic form and interpretation* (pp. 493-521). Walter de Gruyter.

Plant, R. R., Hammond, N., & Whitehouse, T. (2003). How choice of mouse may affect response timing in psychological studies. *Behavior Research Methods, Instruments, & Computers*, *35*(2), 276-284.

Plant, R. R., & Turner, G. (2009). Millisecond precision psychological research in a world of commodity computers: New hardware, new problems?. *Behavior Research Methods*, *41*(3), 598-614.

Pronk, T., Wiers, R. W., Molenkamp, B., & Murre, J. (2020). Mental chronometry in the pocket? Timing accuracy of web applications on touchscreen and keyboard devices. *Behavior Research Methods*, *52*, 1371-1382.

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793-797.

Quinton, A. (1979). Objects and events. *Mind*, *88*(350), 197-214.

Radvansky, G. A. (2012). Across the event horizon. *Current Directions in Psychological Science*, *21*(4), 269-272.

Radvansky, G. A., & Copeland, D. E. (2010). Reading times and the detection of event shift processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 210.

Radvansky, G. A., Krawietz, S. A., & Tamplin, A. K. (2011). Walking through doorways causes forgetting: Further explorations. *Quarterly journal of experimental psychology*, *64*(8), 1632-1645.

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 608-620.

Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition*. Oxford University Press.

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current opinion in behavioral sciences*, *17*, 133-140.

Rappaport Hovav, M., & Levin, B. (1998). Building verb meanings. *The projection of arguments: Lexical and compositional factors*, 97-134.

Ratcliff, R., & McKoon, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology*, *21*(2), 139-155.

Raudszus, H., Segers, E., & Verhoeven, L. (2019). Situation model building ability uniquely predicts first and second language reading comprehension. *Journal of Neurolinguistics*, *50*, 106-119.

Regel, S., Meyer, L., & Gunter, T. C. (2014). Distinguishing neurocognitive processes reflected by P600 effects: Evidence from ERPs and neural oscillations. *PloS one*, *9*(5), e96840.

Reichenbach, H. (1947). The tenses of verbs. *Time: From concept to narrative construct: A reader*, 1-12.

Reimers, S., & Stewart, N. (2007). Adobe Flash as a medium for online experimentation: A test of reaction time measurement capabilities. *Behavior research methods*, *39*(3), 365-370.

Renoult, L., Wang, X., Calcagno, V., Prévost, M., & Debruille, J. B. (2012). From N400 to N300: Variations in the timing of semantic processing with repetition. *NeuroImage, 61*(1), 206-215.

Rinck, M., & Weber, U. (2003). Who when where: An experimental test of the event-indexing model. *Memory & Cognition*, *31*(8), 1284-1292.

Roberts, C. (1987). *Modal subordination, anaphora, and distributivity*. (Unpublished doctoral thesis). University of Massachusetts Amherst.

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, *32*(2), 299-331.

Rogers, J. C., & Davis, M. H. (2017). Inferior frontal cortex contributions to the recognition of spoken words and their constituent speech sounds. *Journal of cognitive neuroscience*, *29*(5), 919-936.

Romagno, D., Rota, G., Ricciardi, E., & Pietrini, P. (2012). Where the brain appreciates the final state of an event: The neural correlates of telicity. *Brain and language*, *123*(1), 68-74.

Rueschemeyer, S. A., Pfeiffer, C., & Bekkering, H. (2010). Body schematics: On the role of the body schema in embodied lexical-semantic representations. *Neuropsychologia*, 48(3), 774-781.

Sakarias, M., & Flecken, M. (2019). Keeping the result in sight and mind: General cognitive principles and language-specific influences in the perception and memory of resultative events. *Cognitive Science*, *43*(1), e12708.

Samet, J., & Schank, R. (1984). Coherence and connectivity. *Linguistics and philosophy*, 57-82.

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., ... & Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241-255.

Schneider, W., & Körkel, J. (1989). The knowledge base and text recall: Evidence from a short-term longitudinal study. *Contemporary educational psychology*, *14*(4), 382-393.

Seger, B. T., Hauf, J. E., & Nieding, G. (2020). Perceptual simulation of vertical object movement during comprehension of auditory and audiovisual text in children and adults. *Discourse Processes*, *57*(5-6), 460-472.

Shipley, E. F., & Shepperson, B. (1990). Countable entities: Developmental changes. *Cognition*, *34*(2), 109-136.

Smith, S. M., Roster, C. A., Golden, L. L., & Albaum, G. S. (2016). A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples. *Journal of Business Research*, *69*(8), 3139-3148.

Snell, J., & Grainger, J. (2019). Readers are parallel processors. *Trends in Cognitive Sciences*, *23*(7), 537-546.

Šoškić, A., Jovanović, V., Styles, S. J., Kappenman, E. S., & Ković, V. (2022). How to do better N400 studies: reproducibility, consistency and adherence to research standards in the existing literature. *Neuropsychology Review*, *32*(3), 577-600.

Speed, L. J., & Majid, A. (2020). Grounding language in the neglected senses of touch, taste, and smell. *Cognitive neuropsychology*, *37*(5-6), 363-392.

Spejewski, B. (1996, August). Temporal subordination and the English perfect. In *Semantics and Linguistic Theory* (Vol. 6, pp. 261-278).

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology*, *47*(2), 276-309.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological science*, *12*(2), 153-156.

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior research methods*, *42*, 1096-1104.

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24-31.

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*(2), 236-257.

Syrett, K., & Musolino, J. (2013). Collectivity, distributivity, and the interpretation of plural numerical expressions in child and adult language. *Language acquisition*, *20*(4), 259-291.

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

Tanenhaus, M. K., Carlson, G. N., & Seidenberg, M. S. (1985). Do listeners compute linguistic representations. *Natural language parsing: Psychological, computational, and theoretical perspectives*, *10*.

Tanner, D., Grey, S., & van Hell, J. G. (2017). Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. *Psychophysiology*, *54*(2), 248-259.

Tettamanti, M., Manenti, R., Della Rosa, P. A., Falini, A., Perani, D., Cappa, S. F., & Moro, A. (2013). Negation in the brain: Modulating action representations. In *The Equilibrium of Human Syntax* (pp. 345-366). Routledge.

Therriault, D. J., Rinck, M., & Zwaan, R. A. (2006). Assessing the influence of dimensional focus during situation model construction. *Memory & cognition*, *34*(1), 78-89.

Tian, Y., & Breheny, R. E. (2019). Pragmatics and negative sentence processing. In C. Cummins & N. Katsos (Eds.), *The Oxford Handbook of Experimental Semantics and Pragmatics* (pp. 195-207). Oxford: Oxford University Press.

Tottie, G., & Kastovsky, D. (1994). Any as an indefinite determiner in non-assertive clauses: Evidence from Present-day and Early Modern English. *Studies in early modern English*, 413-27.

Traxler, M. J., Foss, D. J., Seely, R. E., Kaup, B., & Morris, R. K. (2000). Priming in sentence processing: Intralexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research*, *29*, 581-595.

Ünal, E., Ji, Y., & Papafragou, A. (2021). From event representation to linguistic meaning. *Topics in Cognitive Science*, *13*(1), 224-242.

Van Berkum, J. J. (2008). Understanding sentences in context: What brain waves can tell us. *Current directions in psychological science*, *17*(6), 376-380.

Van Berkum, J. J. (2019). Language comprehension and emotion: where are the interfaces, and who cares?. In G. I. de Zubicaray & N. O. Schiller (Eds.), *The Oxford Handbook of Neurolinguistics* (pp. 736-768). New York: Oxford University Press.

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

Van der Meer, E., Beyer, R., Heinze, B., & Badel, I. (2002). Temporal order relations in language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 770.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. London: Academic Press.

Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In *Reading as a perceptual process* (pp. 621-648). North-Holland.

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

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

Vendler, Z. (1957). Verbs and times. *The philosophical review*, *66*(2), 143-160.

Vendler, Z. (1967). Facts and events. *Linguistics in philosophy*, 122-146.

Venhuizen, N. J., Crocker, M. W., & Brouwer, H. (2019). Expectation-based comprehension: Modeling the interaction of world knowledge and linguistic experience. *Discourse Processes*, *56*(3), 229-255.

Wagner, L., & Carey, S. (2003). Individuation of objects and events: A developmental study. *Cognition*, *90*(2), 163-191.

Walter, S. L., Seibert, S. E., Goering, D., & O'Boyle, E. H. (2019). A tale of two sample sources: Do results from online panel data and conventional data converge?. *Journal of Business and Psychology*, *34*, 425-452.

Wang, L., Hagoort, P., & Yang, Y. (2009). Semantic illusion depends on information structure: ERP evidence. *Brain research*, *1282*, 50-56.

Weiser, S. (2008). The Negation of Action Sentences. *Proceedings of ConSOLE XV*, *357*, 364.

Wharton, C., & Kintsch, W. (1991). An overview of construction-integration model: a theory of comprehension as a foundation for a new cognitive architecture. *ACM Sigart Bulletin*, *2*(4), 169-173.

Winograd, T. (1983). *Language as a cognitive process: Volume 1: Syntax*. United States.

Xiang, M., Harizanov, B., Polinsky, M., & Kravtchenko, E. (2011). Processing morphological ambiguity: An experimental investigation of Russian numerical phrases. *Lingua*, *121*(3), 548-560.

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, cognition and neuroscience*, *30*(6), 648-672.

Xu, X., & Kwok, S. C. (2019). Temporal-order iconicity bias in narrative event understanding and memory. *Memory*, *27*(8), 1079-1090.

Xu, Z., & Liu, D. (2022). Perceptual simulation in language comprehension and Chinese character reading among third-grade Hong Kong children. *Educational Psychology*, *42*(5), 587-606.

Yarrish, C., Groshon, L., Mitchell, J., Appelbaum, A., Klock, S., Winternitz, T., & Friedman-Wheeler, D. G. (2019). Finding the signal in the noise: Minimizing responses from bots and inattentive humans in online research. *The Behavior Therapist*, *42*(7), 235-242.

Yevseyev, V. (2012). Non-iconic chronology in Russian narrative texts. Its types and motivations. *Zeitschrift für Slawistik*, *57*(1), 46-70.

Zacks, J. M. (2020). *Ten Lectures on the Representation of Events in Language, Perception, Memory, and Action Control.* Boston: Brill.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L. & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature neuroscience*, *4*(6), 651-655.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, *138*(2), 307.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological bulletin*, *133*(2), 273.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*(1), 3-21.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of experimental psychology: General*, *130*(1), 29.

Zarcone, A., & Demberg, V. (2021). Interaction of Script Knowledge and Temporal Discourse Cues in a Visual World Study. *Discourse Processes*, *58*(9), 804-819.

Zhao, Q., Ke, W., Tong, B., Zhou, Z., & Zhou, Z. (2017). Creative processing of internet language: Novel N400 and LPC. *Acta Psychologica Sinica*, *49*(2), 143.

Zhang, Q., Yin, Y., & Lin, H. (2021). Are Situation Models Implemented in Modality-Specific Simulation?. *International Journal of Social Science and Education Research*, *4*(4), 36-41.

Zimny, S. T. (1987). Recognition memory for sentences from discourse. (Unpublished doctoral thesis). University of Colorado, Boulder.

Zwaan, R. A. (1991). Some parameters of literary and news comprehension: Effects of discourse-type perspective on reading rate and surface structure representation. *Poetics*, *20*(2), 139-156.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, memory, and cognition*, *22*(5), 1196.

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic bulletin & review*, *23*, 1028-1034.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995a). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological science*, *6*(5), 292-297.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995b). Dimensions of situation model construction in narrative comprehension. *Journal of experimental psychology: Learning, memory, and cognition*, *21*(2), 386.

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PloS one*, *7*(12), e51382.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, *123*(2), 162.

Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific studies of reading*, *2*(3), 199-220.

Appendix A – Chapter 2 Experiment 1: Materials

Condition	Critical Sentence	Filler Sentence	Probe	Type
A	Noah wrote a poem last Tuesday, and I did too.	This morning I tidied my bedroom and later we stole a purse.	poem	Critical
A	Elijah wrote a poem last Tuesday, and I did too.	This morning I pranked my brother and later we sang a song.	tiger	Non
A	John learned a dance last Wednesday, and I did too.	This morning I drew some manga and later we stole a purse.	dance	Critical
A	Jackson learned a dance last Wednesday, and I did too.	This morning I pranked my brother and later we sang a song.	oven	Non
A	Evan sailed an ocean last month, and I did too.	This morning I listened to some music and later we stole a purse.	ocean	Critical
A	Jack sailed an ocean last month, and I did too.	This morning I drew some manga and later we sang a song.	bush	Non
A	Will cooked a turkey last March, and I did too.	This morning I bought a present and later we stole a purse.	turkey	Critical
A	Joshua cooked a turkey last March, and I did too.	This morning I tidied my bedroom and later we drank some lemonade.	banjo	Non
A	Nick touched an artefact last Thursday, and I did too.	This morning I revised for an exam and later we stole a purse.	artefact	Critical
A	Nathan touched an artefact last Thursday, and I did too.	This morning I walked in the country and later we sang a song.	flotsam	Non

A	Paul washed a plate last week, and I did too.	This morning I tidied my bedroom and later we stole a purse.	plate	Critical
A	Adrian washed a plate last week, and I did too.	This morning I played a game and later we sang a song.	truck	Non
A	Dean hunted a deer last Sunday, and I did too.	This morning I cuddled a teddy bear and later we stole a purse.	deer	Critical
A	Austin hunted a deer last Sunday, and I did too.	This morning I played a game and later we sang a song.	shirt	Non
A	Sean fought an officer last month, and I did too.	This morning I listened to some music and later we drank some lemonade.	officer	Critical
A	Chase fought an officer last month, and I did too.	This morning I drew some manga and later we sang a song.	planet	Non
A	Tony kicked a football last night, and I did too.	This morning I bought a present and later we stole a purse.	football	Critical
A	Louis kicked a football last night, and I did too.	This morning I tidied my bedroom and later we drove to the store.	tower	Non
A	Karl kissed a girl last Monday, and I did too.	This morning I revised for an exam and later we stole a purse.	girl	Critical
A	Alan kissed a girl last Monday, and I did too.	This morning I walked in the country and later we sang a song.	ostrich	Non
B	Liam wrote a poem last Tuesday, but I didn't.	This morning Liam shaved his legs and later we drank some lemonade.	poem	Critical
B	Lucas wrote a poem last Tuesday, but I didn't.	This morning Lucas revised for an exam and later we watched a movie.	tiger	Non

B	Jacob learned a dance last Wednesday, but I didn't.	This morning Jacob shaved his legs and later we listened to some music.	dance	Critical
B	Sebastian learned a dance last Wednesday, but I didn't.	This morning Sebastian revised for an exam and later we watched a movie.	oven	Non
B	Joseph sailed an ocean last month, but I didn't.	This morning Joseph stole a purse and later we drew some manga.	ocean	Critical
B	Luke sailed an ocean last month, but I didn't.	This morning Luke drove to the store and later we watched a movie.	bush	Non
B	Julian cooked a turkey last March, but I didn't.	This morning Julian stole a purse and later we sang a song.	turkey	Critical
B	Chris cooked a turkey last March, but I didn't.	This morning Chris sang a song and later we watched a movie.	banjo	Non
B	Patrick touched an artefact last Thursday, but I didn't.	This morning Patrick tidied his bedroom and later we played a game.	artefact	Critical
B	Thomas touched an artefact last Thursday, but I didn't.	This morning Thomas drank some lemonade and later we watched a movie.	flotsam	Non
B	Christian washed a plate last week, but I didn't.	This morning Christian shaved his legs and later we watched a movie.	plate	Critical
B	Nolan washed a plate last week, but I didn't.	This morning Nolan explored a castle and later we drank some lemonade.	truck	Non

B	Dominic hunted a deer last Sunday, but I didn't.	This morning Dominic shaved our legs and later we explored a castle.	deer	Critical
B	Miles hunted a deer last Sunday, but I didn't.	This morning Miles drew some manga and later we watched a movie.	shirt	Non
B	Wesley fought an officer last month, but I didn't.	This morning Wesley tidied his bedroom and later we stole a purse.	officer	Critical
B	Harrison fought an officer last month, but I didn't.	This morning Harrison drove to the store and later we watched a movie.	planet	Non
B	Tristan kicked a football last night, but I didn't.	This morning Tristan stole a purse and later we explored a castle.	football	Critical
B	George kicked a football last night, but I didn't.	This morning George sang a song and later we watched a movie.	tower	Non
B	Jasper kissed a girl last Monday, but I didn't.	This morning Jasper tidied my bedroom and later we drank some lemonade.	girl	Critical
B	Milo kissed a girl last Monday, but I didn't.	This morning Milo drank some lemonade and later we watched a movie.	ostrich	Non
C	Ava didn't read a book last week, but I did.	This morning I walked in the country and later we bought a present.	book	Critical
C	Amelia didn't read a book last week, but I did.	This morning I drew some manga and later we explored a castle.	shelf	Non

C	Emily didn't craft a model last Friday, but I did.	This morning I listened to some music and later we bought a present.	model	Critical
C	Madison didn't craft a model last Friday, but I did.	This morning I bought a present and later we explored a castle.	fridge	Non
C	Chloe didn’t climb a mountain last year, but I did.	This morning I pranked my brother and later we drove to the store.	mountain	Critical
C	Zoey didn’t climb a mountain last year, but I did.	This morning I played a game and later we explored a castle.	flower	Non
C	Eleanor didn't build a house last Monday, but I did.	This morning I listened to some music and later we bought a present.	house	Critical
C	Violet didn't build a house last Monday, but I did.	This morning I walked in the country and later we explored a castle.	shark	Non
C	Bella didn't eat an apple last Sunday, but I did.	This morning I shaved my legs and later we bought a present.	apple	Critical
C	Samantha didn't eat an apple last Sunday, but I did.	This morning I drew some manga and later we explored a castle.	cage	Non
C	Sarah didn't pet a dog last night, but I did.	This morning I walked in the country and later we bought a present.	dog	Critical
C	Ivy didn't pet a dog last night, but I did.	This morning I drew some manga and later we explored a castle.	rocket	Non
C	Alexa didn't host a ceremony last Saturday, but I did.	This morning I listened to some music and later we bought a present.	ceremony	Critical

C	Peyton didn't host a ceremony last Saturday, but I did.	This morning I explored a castle and later we drank some lemonade.	tissue	Non
C	Aubree didn't cross a lake last Tuesday, but I did.	This morning I pranked my brother and later we bought a present.	lake	Critical
C	Kylie didn't cross a lake last Tuesday, but I did.	This morning I walked in the country and later we explored a castle.	horse	Non
C	Bailey didn’t slap a sister last year, but I did.	This morning I listened to some music and later we bought a present.	sister	Critical
C	Eden didn’t slap a sister last year, but I did.	This morning I drew some manga and later we drank some lemonade.	kettle	Non
C	Sydney didn't attend a theatre last week, but I did.	This morning I shaved my legs and later we bought a present.	theatre	Critical
C	Josie didn't attend a theatre last week, but I did.	This morning I drew some manga and later we explored a castle.	chest	Non
D	Isabella didn't read a book last week, and I didn't either.	This morning Isabella listened to some music and later we drove to the store.	book	Critical
D	Evelyn didn't read a book last week, and I didn't either.	This morning Evelyn cuddled a teddy bear and later we played a game.	shelf	Non
D	Camila didn't craft a model last Friday, and I didn't either.	This morning Camila sang a song and later we shaved our legs.	model	Critical

D	Luna didn't craft a model last Friday, and I didn't either.	This morning Luna tidied her bedroom and later we played a game.	fridge	Non
D	Natalie didn’t climb a mountain last year, and I didn't either.	This morning Natalie watched a movie and later we bought a present.	mountain	Critical
D	Nora didn’t climb a mountain last year, and I didn't either.	This morning Nora sang a song and later we played a game.	flower	Non
D	Hannah didn't build a house last Monday, and I didn't either.	This morning Hannah explored a castle and later we played a game.	house	Critical
D	Aurora didn't build a house last Monday, and I didn't either.	This morning Aurora shaved her legs and later we drove to the store.	shark	Non
D	Claire didn't eat an apple last Sunday, and I didn't either.	This morning Claire bought a present and later we tidied my bedroom.	apple	Critical
D	Willow didn't eat an apple last Sunday, and I didn't either.	This morning Willow sang a song and later we drank some lemonade.	cage	Non
D	Allison didn't pet a dog last night, and I didn't either.	This morning Allison listened to some music and later we drew some manga.	dog	Critical
D	Piper didn't pet a dog last night, and I didn't either.	This morning Piper cuddled a teddy bear and later we played a game.	rocket	Non
D	Julia didn't host a ceremony last Saturday, and I didn't either.	This morning Julia sang a song and later we walked in the country.	ceremony	Critical

D	Clara didn't host a ceremony last Saturday, and I didn't either.	This morning Clara drove to the store and later we played a game.	tissue	Non
D	Maria didn't cross a lake last Tuesday, and I didn't either.	This morning Maria watched a movie and later we cuddled a teddy bear.	lake	Critical
D	Ashley didn't cross a lake last Tuesday, and I didn't either.	This morning Ashley revised for an exam and later we played a game.	horse	Non
D	Jasmine didn’t slap a sister last year, and I didn't either.	This morning Jasmine explored a castle and later we sang a song.	sister	Critical
D	Esther didn’t slap a sister last year, and I didn't either.	This morning Esther revised for an exam and later we played a game.	kettle	Non
D	Daisy didn't attend a theatre last week, and I didn't either.	This morning Daisy bought a present and later we drove to the store.	theatre	Critical
D	Alana didn't attend a theatre last week, and I didn't either.	This morning Alana sang a song and later we played a game.	chest	Non

Appendix B – Chapter 3 Experiment 2: Materials

Condition	Critical Sentence	Task Sentence	Plural Type	Timespan
A	James had flown a plane.	He gave the planes to Emma last week.	Plural	Past Perfect
A	Alice has chopped a log.	She gave the log to Liam on Monday.	Non-Plural	Pres Perfect
A	Billy had bought a book.	He gave the book to Ruby last month.	Non-Plural	Past Perfect
A	Sally has baked a cake.	She gave the cakes to Noah on Monday.	Plural	Pres Perfect
A	Lucy had washed a car.	She gave the cars to Dean last night.	Plural	Past Perfect
A	Paul has built a robot.	He gave the robot to Mary on Tuesday.	Non-Plural	Pres Perfect
A	Carol had watched a movie.	She gave the movie to Ryan last night.	Non-Plural	Past Perfect
A	Michael has repaired a computer.	He gave the computers to Gina on Sunday.	Plural	Pres Perfect
A	Ryan has played a game.	He gave the game to Luna on Wednesday.	Non-Plural	Pres Perfect
A	Tina had written a song.	She gave the songs to Juan last night.	Plural	Past Perfect
A	Jack has constructed a toilet.	He gave the toilets to Kara on Tuesday.	Plural	Pres Perfect
A	Beth had designed a statue.	She gave the statue to Neil last month.	Non-Plural	Past Perfect
A	Ian has broken a piano.	He gave the piano to Zoey on Thursday.	Non-Plural	Pres Perfect
A	Diana had uncovered a relic.	She gave the relics to Luka last weekend.	Plural	Past Perfect

A	Josh has spilt a milkshake.	He gave the milkshakes to Faye on Thursday.	Plural	Pres Perfect
A	Ioana had painted a watercolour.	She gave the watercolour to Milo last night.	Non-Plural	Past Perfect
B	Two pilots had flown a plane.	They gave the planes to Abby last night.	Plural	Past Perfect
B	Two lumberjacks have chopped a log.	They gave the log to Jack on Saturday.	Non-Plural	Pres Perfect
B	Two teachers had bought a book.	They gave the book to Gwen last Thursday.	Non-Plural	Past Perfect
B	Two chefs have baked a cake.	They gave the cakes to Evan on Wednesday.	Plural	Pres Perfect
B	Two mechanics had washed a car.	They gave the cars to Gary last Friday.	Plural	Past Perfect
B	Two scientists have built a robot.	They gave the robot to Tara on Monday.	Non-Plural	Pres Perfect
B	Two critics had watched a movie.	They gave the movie to Paul last Friday.	Non-Plural	Past Perfect
B	Two technicians have repaired a computer.	They gave the computers to Fran on Saturday.	Plural	Pres Perfect
B	Two children have played a game.	They gave the game to Lola on Sunday.	Non-Plural	Pres Perfect
B	Two singers had written a song.	They gave the songs to Omar last week.	Plural	Past Perfect
B	Two artisans have constructed a toilet.	They gave the toilets to Alma on Sunday.	Plural	Pres Perfect
B	Two sculptors had designed a statue.	They gave the statue to Hank last week.	Non-Plural	Past Perfect
B	Two weightlifters have broken a piano.	They gave the piano to Rose on Saturday.	Non-Plural	Pres Perfect

B	Two archaeologists had uncovered a relic.	They gave the relics to Kobe last Monday.	Plural	Past Perfect
B	Two diners have spilt a milkshake.	They gave the milkshakes to Kira on Wednesday.	Plural	Pres Perfect
B	Two artists had painted a watercolour.	They gave the watercolour to Chad last Sunday.	Non-Plural	Past Perfect
C	John and James had flown a plane.	John gave the planes to Lily last Tuesday.	Plural	Past Perfect
C	Mary and Alice have chopped a log.	Mary gave the log to Luke on Friday.	Non-Plural	Pres Perfect
C	Charles and Billy had bought a book.	Charles gave the book to Anne last Sunday.	Non-Plural	Past Perfect
C	Amy and Sally have baked a cake.	Amy gave the cakes to Owen on Saturday.	Plural	Pres Perfect
C	Abby and Lucy had washed a car.	Abby gave the cars to Finn last Thursday.	Plural	Past Perfect
C	Alan and Paul have built a robot.	Alan gave the robot to Jade on Saturday.	Non-Plural	Pres Perfect
C	Sophie and Carol had watched a movie.	Sophie gave the movie to Joel last weekend.	Non-Plural	Past Perfect
C	Steve and Michael have repaired a computer.	Michael gave the computers to Rosa on Wednesday.	Plural	Pres Perfect
C	Ryan and Oliver have played a game.	Oliver gave the game to Anya on Friday.	Non-Plural	Pres Perfect
C	Tina and Katie had written a song.	Katie gave the songs to Lyle last weekend.	Plural	Past Perfect

C	Jack and Thomas have constructed a toilet.	Thomas gave the toilets to Kate on Saturday.	Plural	Pres Perfect
C	Beth and Emily had designed a statue.	Emily gave the statue to Mack last night.	Non-Plural	Past Perfect
C	Ian and Will have broken a piano.	Will gave the piano to Lucy on Monday.	Non-Plural	Pres Perfect
C	Diana and Rachel had uncovered a relic.	Rachel gave the relics to Nash last night.	Plural	Past Perfect
C	Josh and Andrew have spilt a milkshake.	Andrew gave the milkshakes to Jana on Sunday.	Plural	Pres Perfect
C	Ioana and Rosie had painted a watercolour.	Rosie gave the watercolour to Ivan last week.	Non-Plural	Past Perfect
D	Two astronauts and two pilots had flown a plane.	The astronauts gave the planes to Nora last Friday.	Plural	Past Perfect
D	Two hunters and two lumberjacks have chopped a log.	The hunters gave the log to Adam on Wednesday.	Non-Plural	Pres Perfect
D	Two writers and two teachers had bought a book.	The writers gave the book to Lisa last night.	Non-Plural	Past Perfect
D	Two students and two chefs have baked a cake.	The students gave the cakes to Andy on Tuesday.	Plural	Pres Perfect
D	Two racers and two mechanics had washed a car.	The racers gave the cars to Hugh last month.	Plural	Past Perfect
D	Two engineers and two scientists have built a robot.	The engineers gave the robot to Vera on Sunday.	Non-Plural	Pres Perfect
D	Two journalists and two critics had watched a movie.	The journalists gave the movie to Karl last month.	Non-Plural	Past Perfect

D	Two programmers and two technicians have repaired a computer.	The programmers gave the computers to Elsa on Tuesday.	Plural	Pres Perfect
D	Two children and two teenagers have played a game.	The teenagers gave the game to Ruth on Monday.	Non-Plural	Pres Perfect
D	Two singers and two dancers had written a song.	The dancers gave the songs to Seth last Tuesday.	Plural	Past Perfect
D	Two artisans and two inventors have constructed a toilet.	The inventors gave the toilets to Lara on Wednesday.	Plural	Pres Perfect
D	Two sculptors and two masons had designed a statue.	The masons gave the statue to Eric last Sunday.	Non-Plural	Past Perfect
D	Two weightlifters and two musicians have broken a piano.	The musicians gave the piano to Nina on Friday.	Non-Plural	Pres Perfect
D	Two archaeologists and two historians had uncovered a relic.	The historians gave the relics to Cory last Tuesday.	Plural	Past Perfect
D	Two diners and two waiters have spilt a milkshake.	The waiters gave the milkshakes to Eden on Friday.	Plural	Pres Perfect
D	Two artists and two illustrators had painted a watercolour.	The illustrators gave the watercolour to Aden last Saturday.	Non-Plural	Past Perfect

Appendix C – Chapter 4 Experiment 3: Materials

Condition	Sentence 1 (Event X)	Sentence 2 (Event Y)	Sentence 3 (Event Z)	Continuation
A	John hopped gleefully into his local music shop at the same time as Mary.	While there, John bought a record and Mary did too.	Beth arrived a few moments later and greeted John with a hug.	Good
A	John hopped gleefully into his local game shop at the same time as Mary.	While there, John bought a puzzle and Mary did too.	Hannah arrived a few moments later and greeted Mary with a kiss.	Good
A	John hopped gleefully into his local sports shop at the same time as Mary.	While there, John bought a ball and Mary did too.	Zoe arrived a few moments later and greeted John with a wave.	Good
A	John hopped gleefully into his local card shop at the same time as Mary.	While there, John bought a present and Mary did too.	Nicola arrived a few moments later and greeted Mary with a smile.	Good
A	John hopped gleefully into his local electronics store at the same time as Mary.	While there, John bought a computer and Mary did too.	Carmen arrived a few moments later and greeted John with a hug.	Good
A	John hopped gleefully into his local art store at the same time as Mary.	While there, John bought a paintbrush and Mary did too.	Nadine arrived a few moments later and greeted Mary with a kiss.	Good
A	John hopped gleefully into his local corner store at the same time as Mary.	While there, John bought a lager and Mary did too.	Monique arrived a few moments later and greeted John with a smile.	Good
A	John hopped gleefully into his local car store at the same time as Mary.	While there, John bought a convertible and Mary did too.	Eva arrived a few moments later and greeted Mary with a wave.	Good
A	John stepped confidently into his local book shop at the same time as Mary.	While there, John bought a novel and Mary did too.	Sally arrived a few moments later and greeted John with a spatula.	Bad

A	John stepped confidently into his local food shop at the same time as Mary.	While there, John bought a sandwich and Mary did too.	Madison arrived a few moments later and greeted Mary with a cloud.	Bad
A	John stepped confidently into his local weapon shop at the same time as Mary.	While there, John bought a gun and Mary did too.	Delilah arrived a few moments later and greeted John with a stalactite.	Bad
A	John stepped confidently into his local gift shop at the same time as Mary.	While there, John bought a souvenir and Mary did too.	Kimberly arrived a few moments later and greeted Mary with a furnace.	Bad
A	John stepped confidently into his local toy store at the same time as Mary.	While there, John bought a figurine and Mary did too.	Jacinta arrived a few moments later and greeted John with a grapefruit.	Bad
A	John stepped confidently into his local furniture store at the same time as Mary.	While there, John bought a sofa and Mary did too.	Fleur arrived a few moments later and greeted Mary with a fence.	Bad
A	John stepped confidently into his local DIY store at the same time as Mary.	While there, John bought a drill and Mary did too.	Anette arrived a few moments later and greeted John with a deodorant.	Bad
A	John stepped confidently into his local curry store at the same time as Mary.	While there, John bought a vindaloo and Mary did too.	Reem arrived a few moments later and greeted Mary with a flamethrower.	Bad
B	John traipsed absently into his local music shop at the same time as Mary.	While there, John bought a record but Mary didn't.	Becky arrived a few moments later and greeted John with a smile.	Good
B	John traipsed absently into his local game shop at the same time as Mary.	While there, John bought a puzzle but Mary didn't.	Viola arrived a few moments later and greeted Mary with a wave.	Good

B	John traipsed absently into his local sports shop at the same time as Mary.	While there, John bought a ball but Mary didn't.	Willow arrived a few moments later and greeted John with a kiss.	Good
B	John traipsed absently into his local card shop at the same time as Mary.	While there, John bought a present but Mary didn't.	Jasmine arrived a few moments later and greeted Mary with a hug.	Good
B	John traipsed absently into his local electronics store at the same time as Mary.	While there, John bought a computer but Mary didn't.	Eliana arrived a few moments later and greeted John with a wave.	Good
B	John traipsed absently into his local art store at the same time as Mary.	While there, John bought a paintbrush but Mary didn't.	Ursula arrived a few moments later and greeted Mary with a smile.	Good
B	John traipsed absently into his local corner store at the same time as Mary.	While there, John bought a lager but Mary didn't.	Sara arrived a few moments later and greeted John with a kiss.	Good
B	John traipsed absently into his local car store at the same time as Mary.	While there, John bought a convertible but Mary didn't.	Olga arrived a few moments later and greeted Mary with a hug.	Good
B	John trotted excitedly into his local book shop at the same time as Mary.	While there, John bought a novel but Mary didn't.	Joan arrived a few moments later and greeted John with a trampoline.	Bad
B	John trotted excitedly into his local food shop at the same time as Mary.	While there, John bought a sandwich but Mary didn't.	Scarlett arrived a few moments later and greeted Mary with a rainbow.	Bad
B	John trotted excitedly into his local weapon shop at the same time as Mary.	While there, John bought a gun but Mary didn't.	Ruby arrived a few moments later and greeted John with a kettle.	Bad

B	John trotted excitedly into his local gift shop at the same time as Mary.	While there, John bought a souvenir but Mary didn't.	Gemma arrived a few moments later and greeted Mary with a typewriter.	Bad
B	John trotted excitedly into his local toy store at the same time as Mary.	While there, John bought a figurine but Mary didn't.	Teresa arrived a few moments later and greeted John with a duvet.	Bad
B	John trotted excitedly into his local furniture store at the same time as Mary.	While there, John bought a sofa but Mary didn't.	Agnes arrived a few moments later and greeted Mary with a volcano.	Bad
B	John trotted excitedly into his local DIY store at the same time as Mary.	While there, John bought a drill but Mary didn't.	Erika arrived a few moments later and greeted John with a cupboard.	Bad
B	John trotted excitedly into his local curry store at the same time as Mary.	While there, John bought a vindaloo but Mary didn't.	Tara arrived a few moments later and greeted Mary with a broom.	Bad
C	John charged quickly into his local music shop at the same time as Mary.	While there, John didn't buy a record but Mary did.	Emma arrived a few moments later and greeted John with a hug.	Good
C	John charged quickly into his local game shop at the same time as Mary.	While there, John didn't buy a puzzle but Mary did.	Rosie arrived a few moments later and greeted Mary with a kiss.	Good
C	John charged quickly into his local sports shop at the same time as Mary.	While there, John didn't buy a ball but Mary did.	Lucy arrived a few moments later and greeted John with a wave.	Good
C	John charged quickly into his local card shop at the same time as Mary.	While there, John didn't buy a present but Mary did.	Daisy arrived a few moments later and greeted Mary with a smile.	Good

C	John charged quickly into his local electronics store at the same time as Mary.	While there, John didn't buy a computer but Mary did.	Renata arrived a few moments later and greeted John with a hug.	Good
C	John charged quickly into his local art store at the same time as Mary.	While there, John didn't buy a paintbrush but Mary did.	Alina arrived a few moments later and greeted Mary with a kiss.	Good
C	John charged quickly into his local corner store at the same time as Mary.	While there, John didn't buy a lager but Mary did.	Sylvie arrived a few moments later and greeted John with a smile.	Good
C	John charged quickly into his local car store at the same time as Mary.	While there, John didn't buy a convertible but Mary did.	Anisa arrived a few moments later and greeted Mary with a wave.	Good
C	John skipped happily into his local book shop at the same time as Mary.	While there, John didn't buy a novel but Mary did.	Jess arrived a few moments later and greeted John with a spatula.	Bad
C	John skipped happily into his local food shop at the same time as Mary.	While there, John didn't buy a sandwich but Mary did.	Chloe arrived a few moments later and greeted Mary with a cloud.	Bad
C	John skipped happily into his local weapon shop at the same time as Mary.	While there, John didn't buy a gun but Mary did.	Sadie arrived a few moments later and greeted John with a stalactite.	Bad
C	John skipped happily into his local gift shop at the same time as Mary.	While there, John didn't buy a souvenir but Mary did.	Ruth arrived a few moments later and greeted Mary with a furnace.	Bad
C	John skipped happily into his local toy store at the same time as Mary.	While there, John didn't buy a figurine but Mary did.	Mathilda arrived a few moments later and greeted John with a grapefruit.	Bad

C	John skipped happily into his local furniture store at the same time as Mary.	While there, John didn't buy a sofa but Mary did.	Aubrey arrived a few moments later and greeted Mary with a fence.	Bad
C	John skipped happily into his local DIY store at the same time as Mary.	While there, John didn't buy a drill but Mary did.	Yvonne arrived a few moments later and greeted John with a deodorant.	Bad
C	John skipped happily into his local curry store at the same time as Mary.	While there, John didn't buy a vindaloo but Mary did.	Jiang arrived a few moments later and greeted Mary with a flamethrower.	Bad
D	John paced slowly into his local music shop at the same time as Mary.	While there, John didn't buy a record and Mary didn't either.	Carol arrived a few moments later and greeted John with a smile.	Good
D	John paced slowly into his local game shop at the same time as Mary.	While there, John didn't buy a puzzle and Mary didn't either.	Sofia arrived a few moments later and greeted Mary with a wave.	Good
D	John paced slowly into his local sports shop at the same time as Mary.	While there, John didn't buy a ball and Mary didn't either.	Brooklyn arrived a few moments later and greeted John with a kiss.	Good
D	John paced slowly into his local card shop at the same time as Mary.	While there, John didn't buy a present and Mary didn't either.	Ashley arrived a few moments later and greeted Mary with a hug.	Good
D	John paced slowly into his local electronics store at the same time as Mary.	While there, John didn't buy a computer and Mary didn't either.	Alba arrived a few moments later and greeted John with a wave.	Good
D	John paced slowly into his local art store at the same time as Mary.	While there, John didn't buy a paintbrush and Mary didn't either.	Ioana arrived a few moments later and greeted Mary with a smile.	Good
D	John paced slowly into his local corner store at the same time as Mary.	While there, John didn't buy a lager and Mary didn't either.	Yvette arrived a few moments later and greeted John with a kiss.	Good

D	John paced slowly into his local car store at the same time as Mary.	While there, John didn't buy a convertible and Mary didn't either.	Samara arrived a few moments later and greeted Mary with a hug.	Good
D	John walked nervously into his local book shop at the same time as Mary.	While there, John didn't buy a novel and Mary didn't either.	Gina arrived a few moments later and greeted John with a trampoline.	Bad
D	John walked nervously into his local food shop at the same time as Mary.	While there, John didn't buy a sandwich and Mary didn't either.	Grace arrived a few moments later and greeted Mary with a rainbow.	Bad
D	John walked nervously into his local weapon shop at the same time as Mary.	While there, John didn't buy a gun and Mary didn't either.	Natalie arrived a few moments later and greeted John with a kettle.	Bad
D	John walked nervously into his local gift shop at the same time as Mary.	While there, John didn't buy a souvenir and Mary didn't either.	Sydney arrived a few moments later and greeted Mary with a typewriter.	Bad
D	John walked nervously into his local toy store at the same time as Mary.	While there, John didn't buy a figurine and Mary didn't either.	Adele arrived a few moments later and greeted John with a duvet.	Bad
D	John walked nervously into his local furniture store at the same time as Mary.	While there, John didn't buy a sofa and Mary didn't either.	Colette arrived a few moments later and greeted Mary with a volcano.	Bad
D	John walked nervously into his local DIY store at the same time as Mary.	While there, John didn't buy a drill and Mary didn't either.	Kara arrived a few moments later and greeted John with a cupboard.	Bad
D	John walked nervously into his local curry store at the same time as Mary.	While there, John didn't buy a vindaloo and Mary didn't either.	Mei arrived a few moments later and greeted Mary with a broom.	Bad

Appendix D – Chapter 4 Experiment 4: Materials

Condition	Sentence 1 (Event X)	Sentence 2 (Event Y)	Sentence 3 (Event Z)	Continuation
A	John walked nervously into his local book shop at the same time as Mary.	While there, John and Mary bought a novel together.	Alice arrived a few moments later and greeted Mary with a hug.	Good
A	John paced slowly into his local music shop at the same time as Mary.	While there, John and Mary bought a record together.	Diana arrived a few moments later and greeted Mary with a spatula.	Bad
A	John walked nervously into his local food shop at the same time as Mary.	While there, John and Mary bought a sandwich together.	Holly arrived a few moments later and greeted John with a kiss.	Good
A	John paced slowly into his local game shop at the same time as Mary.	While there, John and Mary bought a puzzle together.	Mia arrived a few moments later and greeted John with a cloud.	Bad
A	John walked nervously into his local weapon shop at the same time as Mary.	While there, John and Mary bought a gun together.	Nora arrived a few moments later and greeted Mary with a wave.	Good
A	John paced slowly into his local sports shop at the same time as Mary.	While there, John and Mary bought a ball together.	Ivy arrived a few moments later and greeted Mary with a stalactite.	Bad
A	John walked nervously into his local gift shop at the same time as Mary.	While there, John and Mary bought a souvenir together.	Rachel arrived a few moments later and greeted John with a smile.	Good
A	John paced slowly into his local card shop at the same time as Mary.	While there, John and Mary bought a present together.	Isabel arrived a few moments later and greeted John with a furnace.	Bad
A	John walked nervously into his local toy store at the same time as Mary.	While there, John and Mary bought a figurine together.	Francesca arrived a few moments later and greeted Mary with a hug.	Good

A	John paced slowly into his local electronics store at the same time as Mary.	While there, John and Mary bought a computer together.	Ines arrived a few moments later and greeted Mary with a grapefruit.	Bad
A	John walked nervously into his local furniture store at the same time as Mary.	While there, John and Mary bought a sofa together.	Bertha arrived a few moments later and greeted John with a kiss.	Good
A	John paced slowly into his local art store at the same time as Mary.	While there, John and Mary bought a paintbrush together.	Betina arrived a few moments later and greeted John with a fence.	Bad
A	John walked nervously into his local DIY store at the same time as Mary.	While there, John and Mary bought a drill together.	Dominique arrived a few moments later and greeted Mary with a smile.	Good
A	John paced slowly into his local corner store at the same time as Mary.	While there, John and Mary bought a lager together.	Joelene arrived a few moments later and greeted Mary with a deodorant.	Bad
A	John walked nervously into his local curry store at the same time as Mary.	While there, John and Mary bought a vindaloo together.	Mona arrived a few moments later and greeted John with a wave.	Good
A	John paced slowly into his local car store at the same time as Mary.	While there, John and Mary bought a convertible together.	Tatiana arrived a few moments later and greeted John with a flamethrower.	Bad
B	John skipped happily into his local book shop at the same time as Mary.	While there, John and Mary bought a novel separately.	Emily arrived a few moments later and greeted Mary with a smile.	Good
B	John charged quickly into his local music shop at the same time as Mary.	While there, John and Mary bought a record separately.	Paula arrived a few moments later and greeted Mary with a trampoline.	Bad

B	John skipped happily into his local food shop at the same time as Mary.	While there, John and Mary bought a sandwich separately.	Fiona arrived a few moments later and greeted John with a wave.	Good
B	John charged quickly into his local game shop at the same time as Mary.	While there, John and Mary bought a puzzle separately.	Evelyn arrived a few moments later and greeted John with a rainbow.	Bad
B	John skipped happily into his local weapon shop at the same time as Mary.	While there, John and Mary bought a gun separately.	Hazel arrived a few moments later and greeted Mary with a kiss.	Good
B	John charged quickly into his local sports shop at the same time as Mary.	While there, John and Mary bought a ball separately.	Bella arrived a few moments later and greeted Mary with a kettle.	Bad
B	John skipped happily into his local gift shop at the same time as Mary.	While there, John and Mary bought a souvenir separately.	Jade arrived a few moments later and greeted John with a hug.	Good
B	John charged quickly into his local card shop at the same time as Mary.	While there, John and Mary bought a present separately.	Taylor arrived a few moments later and greeted John with a typewriter.	Bad
B	John skipped happily into his local toy store at the same time as Mary.	While there, John and Mary bought a figurine separately.	Saskia arrived a few moments later and greeted Mary with a wave.	Good
B	John charged quickly into his local electronics store at the same time as Mary.	While there, John and Mary bought a computer separately.	Esmerelda arrived a few moments later and greeted Mary with a duvet.	Bad
B	John skipped happily into his local furniture store at the same time as Mary.	While there, John and Mary bought a sofa separately.	Claudia arrived a few moments later and greeted John with a smile.	Good
B	John charged quickly into his local art store at the same time as Mary.	While there, John and Mary bought a paintbrush separately.	Ivana arrived a few moments later and greeted John with a volcano.	Bad

B	John skipped happily into his local DIY store at the same time as Mary.	While there, John and Mary bought a drill separately.	Georgette arrived a few moments later and greeted Mary with a kiss.	Good
B	John charged quickly into his local corner store at the same time as Mary.	While there, John and Mary bought a lager separately.	Chantal arrived a few moments later and greeted Mary with a cupboard.	Bad
B	John skipped happily into his local curry store at the same time as Mary.	While there, John and Mary bought a vindaloo separately.	Rita arrived a few moments later and greeted John with a hug.	Good
B	John charged quickly into his local car store at the same time as Mary.	While there, John and Mary bought a convertible separately.	Aisha arrived a few moments later and greeted John with a broom.	Bad
C	John trotted excitedly into his local book shop at the same time as Mary.	While there, John and Mary bought a novel.	Abby arrived a few moments later and greeted Mary with a hug.	Good
C	John traipsed absently into his local music shop at the same time as Mary.	While there, John and Mary bought a record.	Tina arrived a few moments later and greeted Mary with a spatula.	Bad
C	John trotted excitedly into his local food shop at the same time as Mary.	While there, John and Mary bought a sandwich.	Esther arrived a few moments later and greeted John with a kiss.	Good
C	John traipsed absently into his local game shop at the same time as Mary.	While there, John and Mary bought a puzzle.	Luna arrived a few moments later and greeted John with a cloud.	Bad
C	John trotted excitedly into his local weapon shop at the same time as Mary.	While there, John and Mary bought a gun.	Riley arrived a few moments later and greeted Mary with a wave.	Good

C	John traipsed absently into his local sports shop at the same time as Mary.	While there, John and Mary bought a ball.	Madelyn arrived a few moments later and greeted Mary with a stalactite.	Bad
C	John trotted excitedly into his local gift shop at the same time as Mary.	While there, John and Mary bought a souvenir.	Clara arrived a few moments later and greeted John with a smile.	Good
C	John traipsed absently into his local card shop at the same time as Mary.	While there, John and Mary bought a present.	Molly arrived a few moments later and greeted John with a furnace.	Bad
C	John trotted excitedly into his local toy store at the same time as Mary.	While there, John and Mary bought a figurine.	Paloma arrived a few moments later and greeted Mary with a hug.	Good
C	John traipsed absently into his local electronics store at the same time as Mary.	While there, John and Mary bought a computer.	Ramona arrived a few moments later and greeted Mary with a grapefruit.	Bad
C	John trotted excitedly into his local furniture store at the same time as Mary.	While there, John and Mary bought a sofa.	Edith arrived a few moments later and greeted John with a kiss.	Good
C	John traipsed absently into his local art store at the same time as Mary.	While there, John and Mary bought a paintbrush.	Eloise arrived a few moments later and greeted John with a fence.	Bad
C	John trotted excitedly into his local DIY store at the same time as Mary.	While there, John and Mary bought a drill.	Juliet arrived a few moments later and greeted Mary with a smile.	Good
C	John traipsed absently into his local corner store at the same time as Mary.	While there, John and Mary bought a lager.	Hilde arrived a few moments later and greeted Mary with a deodorant.	Bad

C	John trotted excitedly into his local curry store at the same time as Mary.	While there, John and Mary bought a vindaloo.	Anastasia arrived a few moments later and greeted John with a wave.	Good
C	John traipsed absently into his local car store at the same time as Mary.	While there, John and Mary bought a convertible.	Khadija arrived a few moments later and greeted John with a flamethrower.	Bad