

INFERRING TEMPORAL ORDER IN LANGUAGE COMPREHENSION

Elena Marx

submitted to

Central European University

Department of Cognitive Science

*In partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Cognitive Science*

Primary supervisor: Eva Wittenberg

Secondary supervisor: Ágnes Kovács

Vienna, 2025

Copyright Notice

Inferring temporal order in language comprehension © 2025 by Elena Marx is licensed under [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International \(CC BY-NC-SA 4.0\)](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Declaration of Authorship

I, the undersigned, Elena Marx, candidate for the PhD degree in Cognitive Science declare herewith that the present thesis titled “Inferring temporal order in language comprehension” is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography.

I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person’s or institution’s copyright.

I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

The present thesis includes work that has been published in the following papers:

Marx, E., & Wittenberg, E. (2025). Temporal construal in sentence comprehension depends on linguistically encoded event structure. *Cognition*, 254, 105975.

<https://doi.org/10.1016/j.cognition.2024.105975>

Marx, E., Iwan, O., & Wittenberg, E. (2025). Dynamicity predicts inferred temporal order in complex sentences: Evidence from English, German, and Polish. *Cognitive Science*, 49, e70038. <https://doi.org/10.1111/cogs.70038>

Marx, E., & Wittenberg, E. (2024). Eventuality type predicts temporal order inferences in discourse comprehension. *Glossa Psycholinguistics*, 3, 1–24.

<https://doi.org/10.5070/G60116579>

Marx, E., Jardón, N., & Wittenberg, E. (2025). The state-before-event inference emerges across tenses. *Open Mind*, (9), 726–745. https://doi.org/10.1162/opmi_a_00207

.....
Elena Marx

Declaration of GenAI

I used [Google Cloud Text-to-Speech API](#) for the creation of the auditory stimuli in Chapter 3. The tool was used to provide pre-recorded sentence materials, with neutral intonation across conditions.

I used [ChatGPT](#) and [deepL](#) to proofread my own work before submitting. The tool was used to point out grammatical errors and stylistic impasses.

Abstract

How do people order events in their mind when listening to a story? Language often provides explicit hints as to whether something has happened, will happen or happened before or after something else. However, in many cases, the temporal order between linguistically described situations is underspecified and needs to be inferred. This thesis proposes that such inferences are grounded in more general cognitive principles that take event dynamicity – namely, whether situations are inherently static or change over time – as a starting point. Specifically, I propose that comprehenders infer more salient events to happen against the backdrop of less salient states.

Four lines of experiments serve to test this proposal experimentally. The first study provides first empirical evidence that temporal interpretation is systematically driven by encoded event dynamicity. Participants reliably matched complex situation descriptions to visual sequences in which described states preceded described events. The second study replicates this pattern of results in three languages in an act-out task, demonstrating that this interpretative preference arises in the absence of visual scaffolding. The third study extends the linguistic context under investigation from complex sentence descriptions to discourse. The fourth study delves deeper into the underlying mechanisms of this temporal inference by using different tenses, thereby testing competing explanations based on linguistic encoding and more general cognitive principles.

Overall, this thesis sheds light on the broader question of how event cognition and language interact to form mental representations of temporal relations. Its findings support models of event comprehension that treat event structure, more specifically the presence or absence of dynamic change, as a central organizing principle for understanding temporal structure in language.

Acknowledgments

Eva – for supporting me on every step of the way, not only with immediate advice and encouragement whenever I needed it, but also with your hands-on approach to solving problems. I have learned so many things from you, but most importantly, to face new challenges not with fear or hesitation, but with curiosity and excitement.

Ágnes Kovács – for your valuable feedback and discussions, which served as inspiration for the developmental questions I address at the end of this thesis.

Jesse Snedeker – for generously hosting me during my research stay, and for pushing me to critically evaluate my assumptions and to always consider alternative perspectives.

Natalia Jardón – for your ideas, our discussions, and the collaboration throughout the years.

Attila Balla – for fixing all the things that so often exasperated me.

Martín – for listening, and much more.

All LCL members and the Department of Cognitive Science – for the stimulating input, and for making it such a pleasure to work and discuss in an environment that encourages rigorous scientific thinking.

Réka Finta and Andrea Jenei – for your help, patience, and support, for all of which I owe you a special thank you.

Table of Contents

1. Introduction	1
1.1. Understanding events in language and cognition.....	3
1.1.1. Events and states from a linguistic perspective.....	3
1.1.2. Events from an event cognitive perspective	9
1.2. Connecting the two perspectives	15
1.3. The structure of this thesis	17
2. Mapping visual temporal order onto relative clause constructions	19
2.1. Introduction	19
2.1.1. Event structure in language and perception.....	22
2.1.2. The case of relative clauses.....	25
2.1.3. The current studies: Predictions and outline.....	32
2.2. Methods	35
2.2.1. General materials and procedure.....	35
2.2.2. General analysis.....	37
2.2.3. Experiment 1a	38
2.2.4. Experiment 1b: Terminated states	42
2.2.5. Experiment 1c: Non-restrictive relative clauses	45
2.2.6. Experiment 2a	49
2.2.7. Experiment 2b: Non-restrictive relative clauses	52
2.2.8. Experiment 2c: Linear order	56
2.2.9. Experiment 3	61
2.2.10. Between-language comparison.....	64
2.3. General Discussion.....	68
2.4. Conclusion.....	73
2.5. Supplementary materials	74
2.5.1. Participants	74
2.5.2. Materials and procedure	74
2.5.3. Statistical analysis.....	75
2.5.4. Results	76
2.5.5. Discussion.....	76

3.	Inferring temporal order in English, German and Polish relative clause constructions ...	78
3.1.	Introduction	78
3.1.1.	Encoding temporal relations in language.....	79
3.1.2.	Temporal interpretations in relative clauses	81
3.1.3.	Dynamicity and temporal construal: Previous research	82
3.1.4.	The current studies	86
3.2.	Methods	90
3.2.1.	General materials and procedure.....	90
3.2.2.	General analysis.....	92
3.2.3.	Experiment 1: English.....	93
3.2.4.	Experiment 2: German.....	96
3.2.5.	Experiment 3: Polish.....	101
3.2.6.	Between-language comparison.....	105
3.3.	General discussion.....	108
3.4.	Conclusion.....	112
4.	Beyond relative clauses – Temporal order inference in discourse comprehension.....	113
4.1.	Introduction	113
4.2.	Experiment 1	119
4.2.1.	Methods.....	119
4.2.2.	Results	121
4.2.3.	Discussion of Experiment 1	124
4.3.	Experiment 2	125
4.3.1.	Methods.....	125
4.3.2.	Results	128
4.3.3.	Discussion of Experiment 2	132
4.4.	General Discussion.....	132
5.	Temporal order inference across tenses – Evidence for a Figure-Ground bias.....	136
5.1.	Introduction	136
5.1.1.	Event order through tense-mediated temporal anchoring.....	139
5.1.2.	Future as a modal, not a tense.....	143
5.1.3.	Alternative accounts to derive temporal order from language	145
5.1.4.	The current studies and predictions	150

5.2.	Experiment 1: Past tense	151
5.2.1.	Participants	151
5.2.2.	Materials.....	151
5.2.3.	Procedure	152
5.2.4.	Statistical analysis.....	153
5.2.5.	Results	154
5.2.6.	Discussion of Experiment 1	155
5.3.	Experiment 2: Future tense	156
5.3.1.	Participants	156
5.3.2.	Materials, procedure and analysis	156
5.3.3.	Results	157
5.3.4.	Discussion of Experiment 2	159
5.4.	General Discussion.....	160
5.5.	Conclusion.....	163
6.	General discussion of the thesis and future directions	164
6.1.	Mapping visual temporal order onto relative clause constructions.....	166
6.2.	Inferring temporal order from relative clause constructions across languages	169
6.3.	Temporal order inference in discourse comprehension.....	172
6.4.	Temporal order inferences across tenses	174
6.5.	Future research	177
6.5.1.	Temporal order inference in development	178
6.5.2.	Temporal order inferences across languages.....	181
6.5.3.	Temporal order inference across domains	185
6.5.4.	Linking temporal order inference and causal reasoning	188
7.	Conclusion.....	192
8.	References	193

1. Introduction

Events are everywhere in human experience: whether as short as a sneeze or as long as doing a PhD, we experience events as discrete meaningful units that structure our everyday lives. Crucially, events never occur in isolation but connect in time with other events. These temporal relations are important for how we make sense of the world: For instance, if you first see someone fall off their bike and then notice a car braking, you perhaps conclude that the driver stopped to check on the cyclist. If, however, the braking maneuver happened first, your interpretation is likely to be a different one: namely, that the car's abrupt stop had scared the cyclist into falling off their bike. The interpretation of this situation thus depends crucially on the temporal order between its subparts.

When events and the temporal relations between them are not directly observable, we often learn about them through language. Yet, temporal order in language is not always explicitly encoded: Although relations can be directly expressed, for instance, through temporal adverbs (1), or more subtly through tense contrasts (2), other linguistic constructions leave the temporal order between their subparts underspecified: In relative clause constructions as in (3), it is not clear which event precedes the other. In discourse structures, too, temporal order is often left implicit (4).

- (1) *The car braked **before/after** the cyclist fell.*
- (2) *The car **had braked** for the cyclist who **fell**.*
- (3) *The car braked for the cyclist who fell.*
- (4) *The car braked. The cyclist fell.*

Temporal underspecification rarely leads to communication failure where speakers need to pause in order to clarify the temporal structure of their descriptions. Instead, comprehenders simply infer how events and situations are ordered in time. These inferences can be guided by

causal reasoning or contextual knowledge, but when such cues are absent or causal relations are equally plausible, as in the example above, the question becomes: Which cues do people rely on to understand how and when situations happen in time?

In this thesis, I investigate this question in depth, focusing on the role of event dynamicity. Specifically, I ask whether people use the dynamic properties of a described situation – that is, whether descriptions denote a dynamic event or a static state – to construe the order between situations in time. Across four series of psycholinguistic experiments, I find a systematic effect of event dynamicity on temporal order inference: Stative situations tend to be understood as happening before eventive ones. This effect appears to be independent of task differences, language-specific properties of how event structure is encoded, syntactic constructions and tense meaning.

More broadly, this thesis uses temporal order inference as a window into how people understand and mentally represent situations and events in terms of their dynamic properties. Event dynamicity, understood as the presence or absence of change over time, has been acknowledged in linguistic theory as part of lexical aspectual classification, but it has rarely been examined in isolation or subjected to systematic experimental investigation. In event cognition, by contrast, event dynamicity has been a core concept, constituting how the mind builds event representations from perception. Nevertheless, the lack of dynamic change as its static counterpart has been widely overlooked in this line of research. The goal of this thesis is therefore to connect these two separate, but highly relevant literatures: linguistic research on events and their internal structure (reviewed in Section 1.1.1.) and event cognition research on how events are construed from perception (reviewed in Section 1.1.2.). In both review sections, I focus on how each field has thought about event dynamicity: While my review of the linguistic literature examines the distinction between stative and eventive predicates, or how the presence and absence of change has been analyzed and classified within the linguistic system, my review

of the event cognitive literature focuses on how dynamic change underlies mechanisms of event perception. Together, the two perspectives provide the conceptual basis for this thesis and motivate its central research question, outlined in Section 1.1.3. The empirical chapters of the thesis are then presented in Section 1.2.

1.1. Understanding events in language and cognition

1.1.1. Events and states from a linguistic perspective

Events and the way in which language expresses different event types have long been at the heart of theorizing about the world. In this tradition, events have typically been treated as real-world occurrences and the linguistic behavior of their corresponding descriptions – that is, verbs – as indicators of ontological differences amongst them. Aristotle, for instance, distinguished between MOVEMENTS (κίνησις) which are directed towards an end, and ACTUALITIES (ἐνέργεια) which are always realized at their beginning (Aristotle, *Metaphysics*, Book IX, Θ6, 1048b), arguing that grammatical differences can capture this ontological contrast in the world.¹ Twentieth-century accounts revisited this idea, explicitly incorporating the contrast between dynamically unfolding and static situations: In opposition to mind-body-dualism, Ryle (1949) argued the real dualism lies not between the mental and the physical but between two situational types: DISPOSITIONS and OCCURRENCES. Dispositions, whether mental states (e.g., *to know*) or physical states (e.g., *to be feathered*), are neither temporally bounded (i.e., cannot combine with temporal *for*-phrases) nor internally dynamic (i.e., cannot combine with adverbs like *gradually*). Unlike occurrences (e.g., *chase*), they would therefore not constitute real instances

¹ The so-called 'Imperfective Paradox' (cf. Dowty, 1977; Kenny, 1963; Bach, 1986; Tylor 1977) states that movement verbs in progressive forms do not entail that they reached the endpoint, even though they are defined by them (1), but activity verbs without an inherent endpoint do (2).

- (1) *Teresa is walking home* ≠ *Teresa has walked home*.
- (2) *Teresa is playing the maracas* ≠ *Teresa has played the maracas*.

in the world. Kenny (1963), in turn, distinguished between STATES, PERFORMANCES, and ACTIVITIES, using tense diagnostics to capture the underlying difference between human emotion, volition, and action.

Linguists soon became interested in verb types and their grammatical properties, referring to them as lexical aspect distinctions. The goal was not merely ontological, but to classify linguistic representations of events and predict grammatical differences in the linguistic system. This marked the point at which the linguistic study of events began to focus on their internal structural properties. One of the first accounts in this respect was Vendler's categorization of STATE, ACTIVITY, ACCOMPLISHMENT and ACHIEVEMENT verbs, reflecting different "time schemata" (Vendler, 1967, p. 21): State verbs (12) and activity verbs (13) both denote periods of time; but only states can be divided into indivisible moments, to which they apply equally. Accomplishment verbs (14) denote a culmination point that naturally terminates their duration. Achievement verbs (15) denote such culmination points as definite, indivisible moments without duration.

(5) STATES: *know, possess, desire, want, like, dislike, love, hate, rule*

(6) ACTIVITIES: *run, walk, swim, push, pull, smoke*

(7) ACCOMPLISHMENTS: *write a book, build a house, grow up, recover*

(8) ACHIEVEMENTS: *win a race, draw a circle, spot, recognize, die*

Subsequent accounts have recast Vendler's descriptive time schemata as binary-valued dimensions of event structure (see Table 1, following Kearns, 2000; Olsen, 1994; Van Valin & LaPolla, 1997; Smith, 1999; see also Filip, 2019), making it possible to analyze event structure along distinct, individual dimensions.² This shift towards a dimensional analysis has formally

² Throughout this thesis (unless otherwise noted), I will use the names of these dimensions – dynamicity, telicity and durativity – as shortcuts for both values of a respective dimension: For instance, by referring to the 'event dynamic properties' of a predicate, I refer to the dimension as a whole, thus entailing both the presence and absence of dynamic change in the linguistic encoding.

introduced event dynamicity as a core concept in lexical aspect, yet it always remained somewhat of an outlier amongst the other two dimensions.

Table 1 Feature-based accounts of event-structural dimension, encoded in verbs and predicates

	Dynamicity	Telicity	Durativity
Activity	+	-	+
Achievement	+	+	-
Accomplishment	+	+	+
State	-	-	+

One reason for this imbalance may lie in the ontological vs. linguistic debate that has long engaged theoretical research: In other words, do these event-structural dimensions reflect the properties of the underlying event in the world or do they belong to their linguistic encoding (Bach, 1986; Bennett & Partee, 1972; Carlson, 1981; Dowty, 1972, 1979; Filip, 1999, 2019; Gill, 1993; Krifka, 1992; Levin, 1993; Levin & Rappaport Hovav, 2005; Michaelis, 2011; Parsons, 1990; Piñón, 1995, 1997; Rothstein, 2004; Taylor, 1977; van Hout, 2000; Verkuyl, 1972; Hoekstra, 1992)? In this regard, telicity and durativity have often been attributed to linguistic encoding, given that features easily change across syntactic contexts: For instance, atelic activity verbs like *to eat*, when combined with a singular object (i.e., *eat an apple*), become telic accomplishments that culminate once the apple is eaten. Similarly, punctual achievement verbs such as *to knock* can be aspectually coerced into durative activities when combined with temporal adverbs (e.g. *Sara knocked on the door for two hours*).

Event dynamicity, by contrast, appears to be less penetrable by such manipulations of linguistic context, suggesting that it is more closely tied to the underlying situations themselves rather than the linguistic representations (Kamp, 1980; but see Rothstein, 2004). This stability distinguishes dynamicity from other event-structural dimensions. While some accounts have

argued that eventive verbs become more stative under progressive aspect, perfect tense or in habitual readings (Herweg, 1991; Katz, 2003; Michaelis, 2011; Parsons, 1990; Zucchi, 1998; but see Bertinetto, 1994), the underlying dynamic event remains accessible and part of the linguistic representation. This parallels with the effects of grammatical aspect on telicity, where aspectual markers can modulate the realization of endpoints – that is, the boundedness of a situation – whether they are inherently encoded culmination points or not (Depraetere, 1995; but see de Swart & Verkuyl, 1999).

Another line of research has focused on more fine-grained distinctions within lexical aspect (Dowty, 1972, 1979; Husband, 2010; Jackendoff, 1990, 1991; Pustejovsky, 1991; Rappaport Hovav & Levin, 1998; Rothstein, 2004; Verkuyl, 1972): So-called decompositional accounts have searched for semantic or conceptual primitives whose compositional structure can explain the semantic and grammatical behavior of verb classes, as proposed in (1-4).³

How the decompositional tradition has thought about – and singled out – the category of non-dynamic states is particularly revealing: On the one hand, states have been treated as basic semantic primitives from which other, more complex events can be derived (Dowty, 1979; Jackendoff, 1990, 1991; Pustejovsky, 1991; Rothstein, 2004, 2013; Smith, 1999). At the same time, it has been argued that stative predicates – although unified in their lack of including change⁴ – might require further internal decomposition: Inspired by the stage-level/individual-level distinction (Carlson, 1977; Kratzer, 1995; Milsark, 1974), Dowty (1979) distinguished three stative subclasses, differing in terms of their durational properties: OBJECT-LEVEL STATIVES (e.g., *to like*, *to know*) express permanent states that hold for all moments of an extended interval; INTERVAL STATIVES (e.g. *to sit*, *to lie*, *to stand*) describe stable positional states

³ Such primitives are usually treated as part of lexical meaning, but some accounts have located them at a syntactic level, arguing that event structure is encoded via functional projections (Ramchand, 2008; see also Borer, 2003, 2005; Rosen, 1999).

⁴ But see Rothstein (2004, 2013) who ties change more closely to the dimension of telicity.

that extend over intervals rather than moments; and MOMENTARY STAGE-PREDICATES (e.g., *to be hungry*) denote transient states that can be true for single moments.

Non-decompositional accounts have reached similar conclusions but reframed stative subclasses in terms of other event-structural dimensions: Carlson (1981) and Bach (1986) introduced the class of DYNAMIC STATES, corresponding to Dowty's interval predicates, while Maienborn (2007) grouped interval and momentary-stage predicates as bounded DAVIDSONIAN STATES, in contrast to more persistent, unbounded KIMIAN STATES. Olsen (1994) added STAGE-LEVEL STATES to her classification, characterizing them by a [+telic] feature, in line with expectations that stage-level states have natural endpoints. Crucially, these accounts have broadened Vendler's traditional conceptions of states which has been limited to permanent object-level predicates.⁵

In sum, formal linguistic research has meticulously analyzed semantic and grammatical differences among verbal predicates. When stative predicates have been examined explicitly, they have often either been treated as a unified class, or analyzed through the lens of boundedness and duration – two event-structural concepts that are conceptually closer to telicity and durativity than to event dynamicity. Dynamicity itself has remained a somewhat implicit background assumption, rather than at the core of systematic investigation on stative predicates. More importantly, however, as a formal discipline, this line of work has rarely asked whether the proposed predicate distinctions also affect event construal – in other words, whether they systematically give rise to distinct mental representations of situations during language comprehension. This, so the impression, has been mostly taken as self-evident.

⁵ This broadening has also challenged classical grammatical diagnostics of stativity, which focus on unboundedness and volitional control (Lakoff, 1966): Unlike traditional states (e.g., *to know*), positional states (e.g., *to sit*) are compatible with the progressive (e.g., *Emil was sitting in the kitchen* vs. *#Emil was knowing French*) and momentaneous temporal adverbials (*Emil sat in the kitchen at 3 p.m.* vs. *Emil knew French at 3 p.m.*). They also permit imperatives (*Sit!* vs. *Know French!*) which suggests that they involve some degree of agency or control to initiate or terminate the state (i.e., sitting down and getting up, cf. Hoekstra, 1992).

The question of how different predicates affect mental representations has been addressed by psycholinguistic research, but here, too, the emphasis has largely been on dimensions other than event dynamicity: Experimental work has shown that telicity and boundedness guide temporal inferences on event duration, often in interaction with event knowledge about the typical duration of described situations (Flecken & Gerwien, 2013; but see Milczarski, 2021 for cross-linguistic differences). Furthermore, both affect the representation and availability of event participants during on-line processing (Becker, Ferretti, & Madden-Lombardi, 2013; Malaia, Wilbur, & Weber-Fox, 2012; Misersky, Slivac, Hagoort, & Flecken, 2021); and drive inferences about event completion (Madden & Zwaan, 2003; Minor, Mitrofanova, Guajardo, Vos, & Ramchand, 2023; Minor, Mitrofanova, & Ramchand, 2022), further interacting with durativity (Foppolo, Bosch, Greco, Carminati, & Panzeri, 2021) or event knowledge about the typical duration (Magliano & Schleich, 2000). For durativity, psycholinguistic studies have likewise found systematic effects on duration inferences (Bott, 2010; Bott & Gattnar, 2015), with interpretations being modulated by syntactic framing (Wittenberg & Levy, 2017). Durativity furthermore supports event individuation inferences, with punctual predicates generally being understood as discrete, countable units (Barner, Wagner, & Snedeker, 2008; see also Jardón, Marx, & Wittenberg, 2025), as well as inferences of event repetition, where punctual events are often coerced to mean several instances of the same event (Paczynski, Jackendoff, & Kuperberg, 2014; Pickering, McElree, Frisson, Chen, & Traxler, 2006; Piñango, Winnick, Ullah, & Zurif, 2006; Todorova, Straub, Badecker, & Org, 2000).

A significantly smaller body of research has examined stative predicates more directly: Gennari and Poeppel (2002, 2003) found that stative predicates are processed significantly faster than eventive predicates, attributing this processing advantage to the reduced semantic complexity of stative predicates (cf. Coll-Florit & Gennari, 2011, who argue instead for a durational explanation). Dery and Koenig (2015) further demonstrated that not all stative descriptions

pattern alike: In a story-continuation task, temporary states prompted readers to move narrative time, whereas more permanent states led them to hold narrative time static.

Despite such findings, event dynamicity itself has received comparatively little attention in psycholinguistic research. A likely reason is its perceptual, rather than linguistic, grounding: Whether or not a situation undergoes dynamic change is part of general event knowledge, and cannot easily be modulated by syntactic context or grammatical marking. That is, there is no grammatical operation that can turn an inherently dynamic eventive predicate like *to run* into a genuinely static one or vice versa. The only way to achieve such a shift is to reframe the situation conceptually, using a different, stative predicate, thus drawing attention to a different aspect of the broader context (i.e., the location or properties of the runner rather than the act of running itself). By contrast, telicity and durativity are far more sensitive to linguistic factors: As outlined above, endpoints can easily be added or removed, and durativity can be stretched or compressed, simply by modulating argument structure and aspectual marking, and without changing the entire predicate. These properties make both dimensions not only tractable for experimental manipulation, but also more directly aligned with classic psycholinguistic questions about how the linguistic system shapes mental representations. Event dynamicity, on the other hand, is a dimension that lies truly at the interface between perception and language. Precisely because of this, it offers a test case for assessing how conceptual distinctions, rooted in perceptual experience, are mapped onto linguistic meaning and affect linguistic event construal.

1.1.2. Events from an event cognitive perspective

A research field in which event dynamicity, however, takes center stage is event cognition: From this perspective, events are not treated as ontological types or in terms of their linguistic representations, but as mental representations which the mind carves up from the continuous

stream of experience: As such, events have often been defined as segments of space and time which an observer conceives as discrete units (Zacks & Tversky, 2001, but see Yates, Sherman, & Yousif, 2023).

Crucially, these event representations are not merely by-products of perception, but form the basic input of episodic memory (Kurby & Zacks, 2008; Sargent et al., 2013; Sonne, Kingo, & Krøjgaard, 2017; Swallow, Zacks, & Abrams, 2009), guide attentional and predictive processes (Coull & Nobre, 1998; Nobre & Van Ede, 2018; Pace et al., 2020), and support action planning (Bailey, Kurby, Giovannetti, & Zacks, 2013; Hommel, Müsseler, Aschersleben, & Prinz, 2001; Lashley, 1951) as well as social comprehension (Kuperberg, 2021). In that sense, events serve as cognitive organizing units that enable efficient processing and functioning within an ever-changing environment.

Given their centrality across many cognitive domains, a fundamental question has been how event representations are construed in the first place. Here, the concept of event dynamicity – more specifically, the notion of dynamic change – lies at the core of most proposed mechanisms (cf. Yates et al., 2023): One influential line of research has argued that event representations result from the predictability structure of the perceptual stream (Kurby & Zacks, 2008; Radvansky & Zacks, 2014; Rouhani, Norman, Niv, & Bornstein, 2020; Zacks, Speer, Swallow, Braver, & Reynolds, 2007; Zacks & Tversky, 2001; see also Elman, 1990). On this view, people continuously generate predictions about how their environment will unfold, based on the incoming perceptual input and so-called event models,⁶ which represent the current state of the world in working memory. Whenever prediction error spikes, that is, when the incoming input has not been correctly predicted, the currently active event model is updated and replaced by a

⁶ In the literature, the term ‘event model’ is often used as an umbrella term, encompassing both ‘situation models’, which are construed during language comprehension (e.g., Zwaan, Magliano, & Graesser, 1995; Zwaan & Radvansky, 1998) and ‘experience models’ which are derived from perceptual experience (Radvansky & Zacks, 2011, 2014). Throughout this thesis, I mostly use the term ‘situation model’ in this tradition to emphasize that the focus lies on linguistic, rather than perceptual, construal.

new one that better fits the sensory input and yields higher prediction accuracy. This updating process of event models demarcates the boundaries of one event to the next in perceptual experience. Critically, according to this view, prediction error is always triggered by unexpected or unpredictable change. Or as Zacks and colleagues (2007, p. 277) put it: “[e]vent segmentation depends on change. When the world is static, prediction is easy.”

Others have stressed that event representations may also arise from anticipating dynamic change through event knowledge (Baldwin & Kosie, 2021; Shin & DuBrow, 2021). This view builds on the observation that event boundaries are oftentimes highly predictable, for instance, through routines, plans or schedules. At the same time, environments can be very noisy so that prediction error alone may not always be a helpful cue for event construal. Inference-based accounts have therefore proposed that generalizations over event types, formed through accumulated experience, allow the cognitive system to update event models based on learned probability distributions. Rather than merely reacting towards surprising changes in their environment, people proactively predict them, as evidenced by heightened attention before changes actually occur (Hard, Recchia, & Tversky, 2011; Monroy, Gerson, & Hunnius, 2018). Through such a mechanism, the structure and endpoint of an ongoing event, as well as what is likely to follow, can be inferred by drawing on knowledge of the latent causal relations within and between events in the world. Similar event-knowledge-based mechanisms have been proposed for language processing and learning, where predictions about upcoming input is mediated by event knowledge or statistical regularities in prior experience (Elman, 2009; McRae & Matsuki, 2009; Saffran et al., 1996, 1999).

However, prediction error and event knowledge are not the only ways to construct event representations. Since prior event knowledge is not always available (e.g., abstract, non-causal events, cf. Diamond & Levine, 2020; or entirely novel situations, cf. Levine, Buchsbaum, Hirsh-Pasek, & Golinkoff, 2019), and prediction error is often confounded with contextual

change, it can be difficult to determine which factor drives event construal. Another account has therefore suggested that event representations might arise through the cognitive system's sensitivity to dynamic change itself, rather than interpreting it in order to register event boundaries: Supporting this view, Siefke et al. (2019) disentangled prediction error from contextual change in a source memory task, by introducing change that was frequent and thus highly predictable, compared to stable contexts that were rare, and thus the less predictable exceptions to the rule. Even when change was fully predictable, it nevertheless drove memory performance, suggesting that the cognitive system may treat changes in the environment as structuring cues, when forming event-like mental representations (see also Clewett, Gasser, & Davachi, 2020; and Schapiro et al., 2013).

None of the underlying mechanisms are mutually exclusive but illustrate different routes by which event representations emerge. What all accounts, however, have in common is that event representations and their boundaries are tightly linked to change. In this respect, change can take different forms: One source of change can be physical, that is, through low-level motion features in the perceptual stream that manifest as shifts in direction and velocity. Such motion features have been shown to play a critical role in parsing events (Hard et al., 2011; Speer, Swallow, & Zacks, 2003), yet mainly when people identify fine-grained sub-events, nested in larger event structures, rather than more coarse-grained units of perceptual experience (Baldwin & Kosie, 2021; Zacks, 2004; Zacks & Tversky, 2001; see also Hespous, Grossman, & Saylor, 2010; Hespous, Saylor, & Grossman, 2009). Similarly, changes in location or – in the case of movies or written narratives – changes in time can cue event boundaries (Magliano, Miller, & Zwaan, 2001; Radvansky & Copeland, 2006; Radvansky, Tamplin, & Krawietz, 2010; but see Gomez et al., 2025).

Change can also affect the state of an object or event participant. Building on this observation, Altmann and Ekves (2019) have proposed that event representations can be conceptualized as

intersecting object histories – that is, as records of all objects involved in an event as well as their respective changes of state, tied together through their cooccurrence over space and time. On this view, event representations consist of stored object state-changes, whereas other components of an event (e.g., causal structure, event roles etc.) emerge as higher-level abstractions from these underlying primitives. Evidence for this proposal comes from language comprehension where people seem to simultaneously activate different object states when processing event descriptions (Hindy, Altmann, Kalenik, & Thompson-Schill, 2012; Kang, Eerland, Joergensen, Zwaan, & Altmann, 2020; Misersky et al., 2021). A similar idea can be found in Gibson’s (1975) approach, where events correspond to changes in the layout of surfaces – ranging from minimal changes such as repositioning an object, to more substantial changes such as the reshaping or the sudden appearance/disappearance of an object from the visual field.

Change can also relate to more top-down or inferred information, such as shifts in goal structure. Indeed, experimental work suggests that event boundaries are more strongly associated with the completion of goals than with surface-level changes such as shifts in location and event participants (Gomez et al., 2025; see also Zacks, Tversky, & Iyer, 2001), movement dynamics (Zacks, 2004) or object states (Mathis & Papafragou, 2022). Moreover, goal completions appear to be more salient segmentation points than onsets (Candan Şimşek, Aydın, & Huff, 2025; Levine, Hirsh-Pasek, Pace, & Golinkoff, 2017). Event boundaries can even arise from the observer’s own goal structure (Wang & Egner, 2022) and more broadly from first-person changes such as shifts in affect or motivation (Wang, Adcock, & Egner, 2024, see also Magliano, Radvansky, Forsythe, & Copeland, 2014). Whether a goal has been achieved can also be linked to the broader concept of boundedness, namely whether an event has reached its natural endpoint or culmination. Boundedness has been shown to lead to abstract event

categorization (Ji & Papafragou, 2020, 2022) as well as different patterns of attentional allocation both during verbal and non-verbal encoding tasks (Sakarias & Flecken, 2019).

Overall, event dynamicity – understood as physical or conceptual change – is manifold, often coinciding in many of these dimensions (Huff, Meitz, & Papenmeier, 2014; Pettijohn & Radvansky, 2016; Zacks, Speer, & Reynolds, 2009; Zacks, Speer, Swallow, & Maley, 2010; see also Zwaan & Radvansky, 1998). More importantly, it is central to theories of event cognition. The way in which it is conceptualized, however, is strikingly different from its treatment in the linguistic literature on event structure. Rather than being treated as a binary feature, that is, as the mere presence or absence of dynamic change, event cognitive approaches tend to view it as a relative property: In a continuously changing environment, not every fluctuation is construed as an event, but only certain types or magnitudes of change reach a threshold necessary to be perceived as event boundaries, whether those are triggered through prediction error, anticipatory processing, or a general sensitivity to meaningful change.⁷ This conceptualization of change also ties into how events themselves are conceived: From an event cognitive perspective, events are primarily those that are “inherently dynamic in nature” (Kuperberg, 2021, p. 258, see also Altmann & Ekves, 2019), whereas non-changing states are more broadly described as situations.⁸ The only notable exception, to the best of my knowledge, is Gibson’s (1975) account of event perception, who places both the detection of change and of non-change or permanence at the heart of an event perceptual system.

⁷ See Yates et al. (2023) for a similar point as well as Pettijohn and Radvansky (2016), Radvansky and Copeland (2010), Rinck and Weber (2003) and Zwaan et al. (1995, 1998) for empirical evidence.

⁸ This exclusion of stativity is also implicit in theories of visual change detection: Rensink (2000), for instance, argues that change is detected whenever attention is directed to disruptions of coherence that is, moments when visual scenes becomes unstable. In this framework, stability is equivalent to coherence, not to the absence of change. Static aspects of a scene are not coherently bound in the first place, and therefore fall outside the focus of attention. See also Royden, Wolfe and Klempen (2001) who demonstrate that visual search of dynamic items in a static context is easier than of static items within a field of moving distractors.

1.2. Connecting the two perspectives

The central goal of this thesis is to bring together the two literatures I have reviewed in the previous sections by focusing on what both, in different ways, take as a central concept: dynamic change. Specifically, I take the linguistic distinction between non-dynamic states and dynamic events – a binary contrast originating from theories on lexical aspect – as a conceptual starting point and highlight it as a critical, yet underexplored, factor in event construal, following the event cognitive tradition.

The thesis's research question is to what extent event dynamicity guides how people understand and mentally represent situations during language comprehension, tapping into a more general cognitive mechanism that structures perceptual experience through its sensitivity to dynamic change. I focus on how comprehenders infer temporal order, when interpreting linguistic descriptions of more complex situations, in which temporal relations are not explicitly encoded. Two aspects of my research question are worth emphasizing: First, this thesis focuses on language comprehension and the construal of temporal relations through linguistic input, rather than through perception. Language offers a unique perspective on event construal as it requires comprehenders to build mental representations without direct perceptual access to the unfolding events themselves. Linguistic descriptions can flexibly direct attention to different parts of a situation (e.g., *Tomas sliced tomatoes*), including the situation as a whole (e.g., *Tomas prepared dinner*), or the connections between them (e.g., *Tomas sliced tomatoes before he chopped some herbs*), but can also leave many aspects of a situation underdetermined, such that much of the situational structure must be actively inferred. Previous work has shown that such linguistic inferences are driven by general event knowledge, such as scripts and schemas (Schank & Abelson, 1977), providing scaffolds not only for unmentioned aspects of a situation but also about their temporal structure (Nieuwland, 2015; Nuthmann & Van Der Meer, 2005; Van Der Meer, Beyer, Heinze, & Badel, 2002; Zarcone & Demberg, 2021). Moreover, listeners not only

represent the events mentioned in a description but also their background settings (Horchak & Garrido, 2022) and interrelations (Claus & Kelter, 2006; Townsend, 1983), even when these relations remain implicit or go unnoticed (Rinck, Gámez, Díaz, & De Vega, 2003; Rinck, Hähnel, & Becker, 2001; Therriault & Raney, 2007). This makes linguistic construal an especially insightful subject to study how events and their temporal relations are mentally represented and how they emerge from an interplay of both cognitive and linguistic factors.

Second, the focus on temporal order within complex situations is not an arbitrary choice, but follows directly from both literatures. In event cognitive research, a key question concerns how events, derived from the perceptual stream, are structured, in particular, how their temporal links and hierarchical relations are established (Radvansky & Zacks, 2014; Zacks, 2004; Zacks et al., 2001). Accordingly, this thesis builds on the idea that change is important not only for segmenting the perceptual input into discrete, meaningful units, but for providing “a constant and ever-present source of temporal information” (Stojić & Nadasdy, 2024, p. 2). In linguistic research, on the other hand, formal accounts have long sought to capture how individual descriptions are temporally and conceptually connected within larger narratives, also recognizing the effect of event classes on temporal interpretation (Hinrichs, 1986; Hopper, 1979; Kamp & Reyle, 1993; Lascarides & Asher, 1991, 1993; Reinhart, 1984): A central observation from these accounts has been that stative and eventive descriptions have different consequences for narrative time. While eventive predicates tend to advance a narrative in time, thus creating interpretations of temporal progression, stative predicates serve to elaborate on the broader situation and tend to overlap in time with described events. Such formal accounts, however, lack a cognitive explanation for why states and events should give rise to different temporal relations in discourse. This thesis therefore approaches this question from a more cognitive perspective, exploring how temporal relations are systematically driven by event

dynamicity – which may, in turn, also underlie the linguistic patterns formalized by such frameworks of discourse comprehension.

1.3. The structure of this thesis

The first empirical chapter (Chapter 2) provides direct evidence for the hypothesis that event dynamicity systematically guides the way in which temporal order is construed in language comprehension: A line of sentence-video-matching studies, conducted in German and English, demonstrates a robust preference to interpret descriptions of static states as preceding descriptions of dynamic events (published as Marx & Wittenberg, 2025). These findings not only establish a first empirical basis for the research question in this thesis, but also test predictions from linguistic accounts regarding temporal interpretations in the specific linguistic context I investigated: relative clause constructions.

In the second empirical chapter (Chapter 3), I examine whether this temporal order preference also arises from language alone, without visual input explicitly distinguishing between alternatives to interpret temporally underspecified sentences. Using an act-out task, the line of studies in this chapter replicates the results from Chapter 2 as an inference in three typologically different languages (published as Marx, Iwan, & Wittenberg, 2025). Chapter 3 therefore allows to further assess whether the effects of event dynamicity rely on language-specific encoding of event structure, by addressing questions of cross-linguistic generalizability.

Chapter 4 tests the scope of the effect of event dynamicity across linguistic contexts and contrasts it with another well-established temporal order inference: the iconicity principle, according to which the order of mention is assumed to reflect the order of situations in the world. The experiments reported in this chapter (published as Marx & Wittenberg, 2024) collect explicit temporal order judgments, investigating whether a state-before-event inference holds in minimal discourse and whether it interacts with an inference of iconic temporal order. The

findings in Chapter 4 show that both temporal order inferences are at play, but that event-structural properties outweigh the impact of iconicity, not only when one situation is described as stative and the other one as eventive (i.e., differences in event dynamic properties), but also when two stative descriptions differ in their durational properties. In this way, Chapter 4 also highlights the interplay between different dimensions of event structure.

In Chapter 5, I address the underlying mechanism that gives rise to the inference that states unfold before events. The working hypothesis, developed in Chapter 2, puts forth a linguistically grounded explanation where temporal and aspectual encoding play a central role in how two situations are temporally linked. By examining temporal inferences in a different tense context (i.e., the simple future), I contrast predictions from this linguistic account with a broader cognitive explanation: a Figure-Ground distinction in the event domain. Two temporal order judgment experiments (published as Marx, Jardón, & Wittenberg, 2025) provide evidence for the latter account, while also offering new insights into the semantics of the future tense.

Chapter 6 sums up the main findings of the thesis and situates them in the bigger research question of how event cognition and language interact in the context of event construal. It also outlines directions for future research, to test the Figure-Ground account in more depth and to address open questions.

2. Mapping visual temporal order onto relative clause constructions

2.1. Introduction

Events are basic constructs of human cognition: We experience holidays, sneezes, wars, and concerts as discrete units. One way the constant flow of everyday life is organized is through temporal sequence: Each event is preceded by events that have happened before, and it is followed by events that are about to happen. In this paper, we ask how temporal sequence is derived from temporally underspecified linguistic descriptions, and how such descriptions map onto visually perceived event sequences. In language, what determines whether an event is construed as preceding or following another event? We find that one important factor answering this question is whether something is described as a static or dynamic situation: States precede events.⁹

Our starting point is an observation based on data from perception: Namely, that dynamic events often unfold against unchanging states. When you enter your kitchen in the morning to drink a cup of coffee, you are likely to pay more attention to the brewing event where coffee dynamically drips into your cup than to the fact that your coffee machine statically stands on the kitchen counter.

This distinction between states and events is also fundamental in natural language. That is, verbs describing states, like *to stand*, differ from verbs describing events, like *to brew*, in their grammatical properties: *The coffee machine stood on the kitchen counter for years* is a natural

⁹ There is considerable variation in the terminology surrounding event structure (cf. Casati & Varzi, 2008). From here on, we use the following terms to minimize confusion between linguistic and cognitive notions: When we talk about non-linguistic events *or* states, we call them 'situations'. Situations consist of non-dynamic 'states' and change-of-state 'events'. We leave the technical terms 'event cognition', 'event perception' and 'event structure' unchanged and understand them to include both events and states. The linguistic encodings of situations are called 'linguistic event type'. They are expressed linguistically as 'stative descriptions' and 'eventive descriptions', respectively. For the purposes of the current studies, we do not further distinguish between Aktionsarten (Vendler, 1967). Instead, we follow Olsen (1994) who defines described events in language as states with added dynamicity.

sentence, but *The coffee brewed for years* needs some semantic gymnastics, for instance, coming up with an enriched interpretation of a repeated, habitual action, as opposed to a yearslong brewing event.

Here, we argue that this perceptual and linguistic distinction between stative and eventive situations could serve to guide inferences about temporal sequence within complex situations. Specifically, we ask: Do people use their knowledge of the internal structure of things that happen over time to infer how they are temporally ordered?

Temporal sequence is crucial in several cognitive processes, from perception to planning: People sequence the unfolding perceptual input into consecutive situations based on the degree of change in the environment (Huff et al., 2014; Newton, 1973; Zacks et al., 2009, 2001). Temporal sequence has also been linked to memory, with amnesic patients showing impaired recall of order of situations (Downes, Mayes, MacDonald, & Hunkin, 2002; Gillis, Quinn, Phillips, & Hampstead, 2013; Shimamura, Janowsky, & Squire, 1990). Furthermore, memory retrieval is improved when the retrieved information is linked to temporal boundaries (Kurby & Zacks, 2011; Radvansky, 2012; Sargent et al., 2013; Sonne, Kingo, & Krøjgaard, 2017; Swallow, Zacks, & Abrams, 2009; Zacks, Speer, Vettel, & Jacoby, 2006, see also Tulving, 1984), and the temporal structure of situations guides attentional focus (Nobre & Van Ede, 2018; Olson & Chun, 2001), facilitating both action learning (O'Reilly et al., 2008) and action planning (Bailey, Kurby, Giovannetti, & Zacks, 2013; see also Hommel, Müsseler, Aschersleben, & Prinz, 2001; Lashley, 1951).

In linguistic descriptions, information about temporal aspects is inherent. For instance, the examples in Figure 1 illustrate how different linguistic expressions map onto specific temporal slices of a situation (see Figure 1A, which shows a short video clip as a sequence of still frames). Depending on the description you choose, different parts of the scene will make up a reader's mental model of the situation (Gleitman & Papafragou, 2012; Slobin, 1996; Talmy, 1988):

Reading *The rabbit hopped over to the girl* maps onto the situation in (1B), where the rabbit is first somewhere else and then moves towards the girl. The description *The girl fed the rabbit with the carrot*, on the other hand, only maps onto the parts of the situation in (1C).

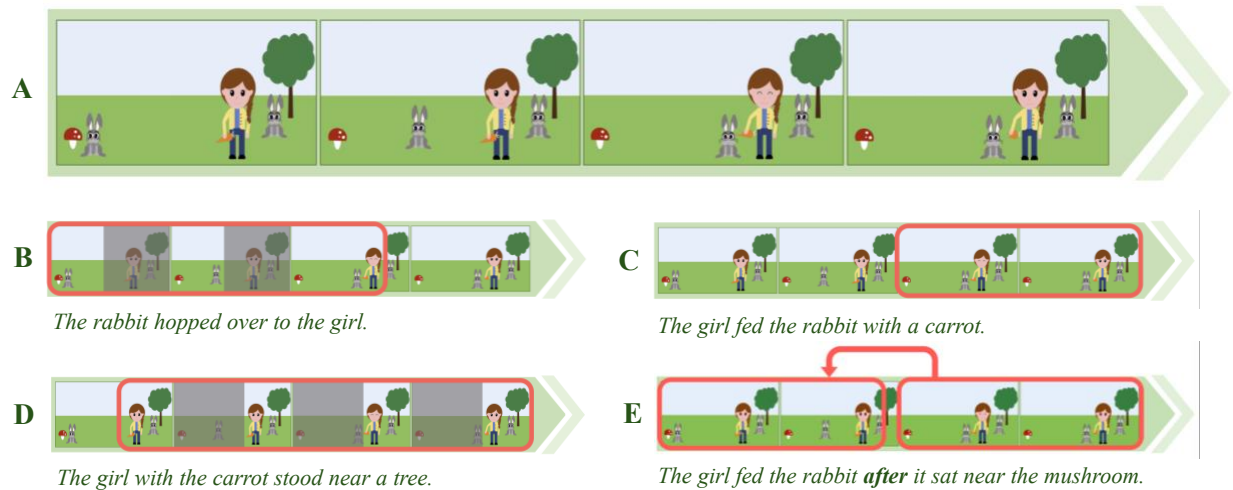


Figure 1 A. Illustration of an animation of a complex situation. B-E: Linguistic descriptions relating to the animation, and the possible mental situation models derived from them.

However, not everything in Figure 1A is a dynamic event: Some parts of this situation are static. While we intuitively may not pay much attention to static situations (Gibson, 1975b), we can nevertheless talk about them (Bach, 1986; Dowty, 1979; Katz, 2003; Michaelis, 2011). Therefore, the *The girl with the carrot stood near a tree* can be used to refer only to the non-dynamic situational parts of the scene (Figure 1D), which perceptually are less salient, as we know that the focus of visual attention is typically on dynamic features of scenes (Han, Sun, Hu, Han, & Shao, 2014).

With regard to temporal sequence within a complex situation, like in Figure 1A, language also provides explicit means to relay the order of situations, such as the conjunction *after* in Figure 1E; but unlike in Figure 1E, order does not always have to be explicitly mentioned.¹⁰ For

¹⁰ There is great diversity in how temporal relations are mapped onto different linguistic devices across languages, from tense to temporal adverbs to connectives (Bybee & Dahl, 1989; Comrie, 1985; Haspelmath et al., 2005; Núñez & Cooperrider, 2013), but note that they are neither universal (i.e., tenseless languages such as Mandarin, Dyirbal, Yucatec; Bohmeyer, 2009; Comrie, 1985; Tonhauser, 2015), nor is their use obligatory (e.g., for Mandarin temporal adverbs see Lin, 2003, 2006).

instance, in a past-under-past relative clause construction such as *The girl fed the rabbit that was near the mushroom*, tense morphology (i.e., *fed*, *was*) establishes the temporal relation of each situation to the time of utterance (i.e., both happened in the past), but the way in which these situations are ordered relative to each other is not straightforward: That is to say, the description may or may not lead to a construal corresponding to Figure 1E. To understand the temporal order of situations in past-under-past relative clause constructions, comprehenders need to rely on other cues to construct their situation models. Here, we argue that one of those cues could be the distinction between stative and eventive descriptions. More to the point, the difference between linguistic event types, and the resulting differences in the individual event construals, could guide temporal construal of complex events in past-under-past relative clauses.

Crucially, there is evidence that temporal inferences are calculated automatically during language comprehension. For instance, readers track the temporal aspects of a narrative (Claus & Kelter, 2006; Townsend, 1983), even when temporal information is implicit (Therriault & Raney, 2007) or not actively noticed (Rinck, Hähnel, & Becker, 2001, but see Rinck, Gámez, Díaz, & De Vega, 2003). That is, even when, on the face of it, establishing temporal structure is not strictly needed, there is evidence that temporal inferences are generated as a by-product of language comprehension.

In what follows, we focus on insights into event structure from linguistics and event perception (2.1.1), and review theories on temporal interpretations of relative clauses (2.1.2), leading to predictions for the current studies (2.1.3).

2.1.1. Event structure in language and perception

From a cognitive perspective, the analysis of event structure has typically started from the observation that, in order to make sense of the world, our minds parse the ongoing stream of

experience into successive (and hierarchical) situation models (Radvansky & Zacks, 2011). The way in which these situation models are parsed is determined by the perception of salient changes. Put differently, there seems to be a fundamental perceptual difference between states and events (i.e., changes-of-state): Events are more salient than states because they attract more of people's attentional resources and create the structural boundaries within the perceptual input (Clewett et al., 2020; Huff et al., 2014; Kurby & Zacks, 2008; Ongchoco, Yates, & Scholl, 2023; Swallow et al., 2011; Zacks et al., 2007). In line with this, these boundaries at the peak of dynamicity can be seen as closing off a history of past states, as has been argued for scenes involving objects undergoing change (Altmann & Ekves, 2019).

In linguistics, on the other hand, the analysis of event structure has often been couched in terms of verbal predicates (Koenig, 2015; Maienborn, 2011). This has led to a rich catalogue of verb types differing with regard to properties such as durativity or telicity reflected in linguistic behavior across grammatical contexts (Bach, 1986; Bertinetto, 1994; Borer, 1991; Comrie, 1976, 1985; Depraetere, 1995; Dowty, 1979; Katz, 2003; Kratzer, 1995; Levin, 1993; Michaelis, 2011; Parsons, 1990; Vendler, 1967; Verkuyl, 1972). Importantly, and in line with event perception research, linguistic analyses have argued for a distinction between eventive and stative predicates – a distinction that is not only based on linguistic tests but can also be found in psycholinguistic data: Previous studies have shown that stative predicates are processed faster than eventive predicates (Gennari & Poeppel, 2003) and that, within a narrative, reading times increase as a function of the amount of changes (Pettijohn & Radvansky, 2016; Radvansky & Copeland, 2010; Rinck & Weber, 2003; Zwaan, Magliano, & Graesser, 1995; Zwaan, Radvansky, Hilliard, & Curiel, 1998).

Thus, arising independently from both perceptual and linguistic data, there is evidence that events are more salient than states. We established this difference also in a language production task as a norming study for this paper: When English and German native speakers (total N=84)

linguistically described the video materials we used for the current study (i.e., short clips as represented in Figure 1A), they used about nine times more eventive than stative predicates in both languages, with more than half of the participants omitting stative predicates from their descriptions altogether (Figure 2, see Section 2.5 for details on the experimental procedure and results). While we cannot say whether this is because participants paid more attention to dynamic state-changes than to static states, as suggested by event cognition research, they did at the very least find events more worthy of mention.

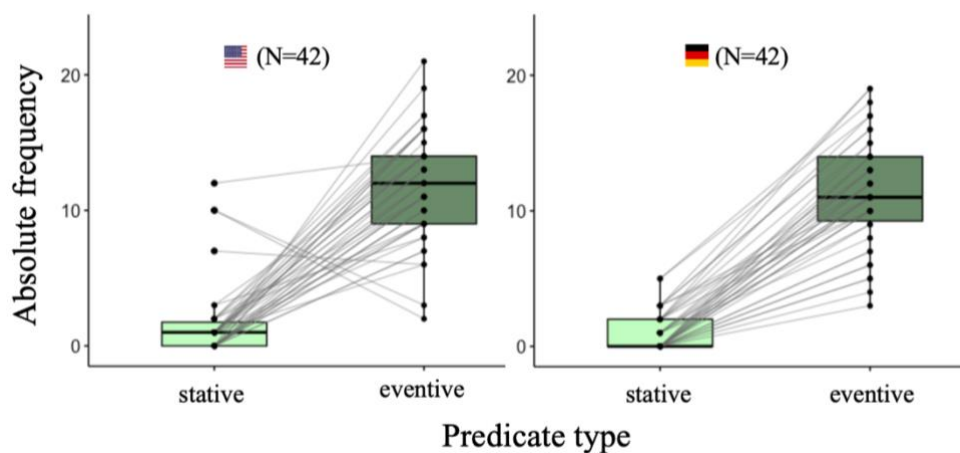


Figure 2 Absolute frequencies of stative and eventive predicates in a language production task, describing the critical video materials of the study series. Dots and lines represent individual participants.

Overall, our hypothesis is that the linguistic distinction between states and events, which parallels a similar difference in perception, reliably predicts inferences of temporal order within complex situations, such as in Figure 1A. In particular, we ask whether people derive temporal orders under conditions of uncertainty based on the linguistic state-event distinction.

In what follows, we review current theories of how temporal inferences are drawn during sentence comprehension, and contrast these with an account rooted in event structure.

2.1.2. The case of relative clauses

To understand how people make inferences about temporal order, we focus on past-under-past relative clause constructions such as (1) where a simple past relative clause, introduced by a relative pronoun *that*, modifies a noun phrase in the matrix clause, also in the simple past.

- (1) *The girl **fed**^{PAST} the rabbit that **was**^{PAST} near the mushroom.*

The rationale for this choice is twofold: First, relative clauses are frequent and well-understood devices to link two propositions to one another in a relationship that is semantically neutral: Unlike conjunctions (*because, after, if*), the relative pronoun (*that/who* in English) simply links a noun (i.e. *rabbit* in ex. 1) to a proposition that includes that noun as an argument. That is, no specific causal, temporal, or conditional inferences between the main clause and the relative clause are triggered by the relative pronoun (e.g., Thomson & Martinet, 1985).

Second, past-under-past relative clauses are excellent test cases for three linguistic approaches that make different predictions on how temporal inferences are drawn during sentence comprehension in general: syntax-driven accounts (cf. Abusch, 1997; Enç, 1986, 1987; Ogihara, 1995, 1996; Ogihara & Sharvit, 2012; Stowell, 2007; Von Stechow & Grønn, 2013a), discourse pragmatic accounts (cf. Chierchia & McConnell-Ginet, 1990; Depraetere, 1996; Potts, 2005), and accounts rooted in non-linguistic event cognition (Carroll & von Stutterheim, 2010; Klein, 1994, 2000, 2008a, 2009; von Stutterheim, Carroll, & Klein, 2003).

In English past-under-past relative clauses such as (1), there are at least three logically possible ways in which the two situations described in past tense can be temporally related: Any model would argue that the sentence in (1) can express that the girl fed the rabbit while it was near the mushroom. Such a simultaneous interpretation (depicted in Figure 3A) is possible as long as the only requirement of the past tense morpheme is met: Both the event and the state must have occurred before the utterance time.

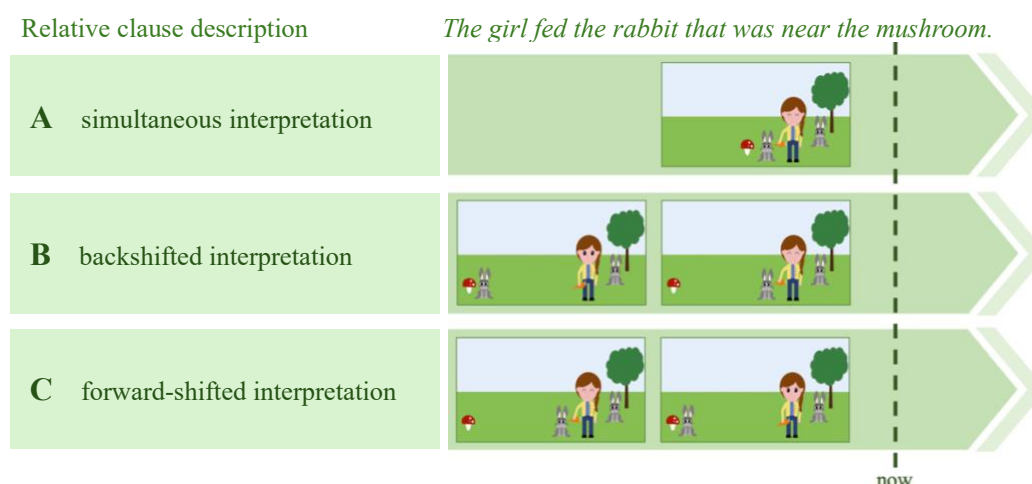


Figure 3 Logical possibilities of ordering a past relative clause description (A) simultaneous, (B) backshifted, or (C) forward-shifted relative to a past main clause description

There are, however, two additional logical possibilities of temporal order for past-under-past relative clauses: In principle, the state described in the relative clause (being near the mushroom) could precede the event described in the main clause, so that the rabbit was near the mushroom before the feeding event (i.e., a so-called back-shifted interpretation of the relative clause, see Figure 3B). Alternatively, the state described by the relative clause could also follow the event described by the main clause. In such an interpretation, the girl would have fed the rabbit before it was near the mushroom (i.e., a so-called forward-shifted interpretation of the relative clause, see Figure 3C).

Both interpretations fulfil the past tense requirement to locate a situation before the utterance time, and become explicit when temporal information is added through tense or temporal adverbials (see 2, for an unambiguous back-shifted interpretation, and 3, for an unambiguous forward-shifted interpretation).

(2) *The girl **fed** the rabbit that **had been** near the mushroom.*

(3) *The girl **fed** the rabbit that **was** near the mushroom **afterwards**.*

The question, however, is whether such shifted interpretations are in fact a reflection of how people build their temporal models of a described situation, as in (1). This question has been answered differently in different approaches, which we will now review in turn.

2.1.2.1. Syntactic structure licenses temporal interpretations

From a syntactic perspective, it has been argued that people derive temporal relations between situation descriptions from the structural properties of a sentence (e.g., Abusch, 1997; Enç, 1986, 1987; Ogihara, 1995; Ogihara, 1996; Stowell, 2007). This argument is based on the observation that understanding the syntactic structure of a sentence often determines how events, situations, properties, and states relate to each other (Rohde, Levy, & Kehler, 2011): For instance, in a sentence such as *The girl fed the rabbit sitting near the mushroom*, a comprehender must decide whether sitting near the mushroom is attached to the rabbit (i.e., *The girl fed the sitting rabbit*), the girl (i.e., *The sitting girl fed the rabbit*), or modifies the entire feeding event (i.e., *The sitting girl fed the sitting rabbit*). Syntax-driven accounts have therefore claimed that the syntactic status of different clause types corresponds to the underlying semantics of a sentence and determines how an embedded tense is related to a main clause tense (Abusch, 1997; Enç, 1986, 1987; Ogihara, 1995; Ogihara, 1996; Stowell, 2007).

The argument is as follows: Complement clauses are verbal arguments (Figure 4, left). As such, they should always be evaluated relative to the time encoded by the main clause (cf. Abusch, 1997; Enç, 1987; Kusumoto, 1999; Ogihara, 1996; Ogihara & Sharvit, 2012; Stowell, 2007, for a general discussion of tense interpretation in embedded contexts). In contrast, relative clauses are syntactic adjuncts of antecedent head nouns, which they modify (Figure 4, right). A side

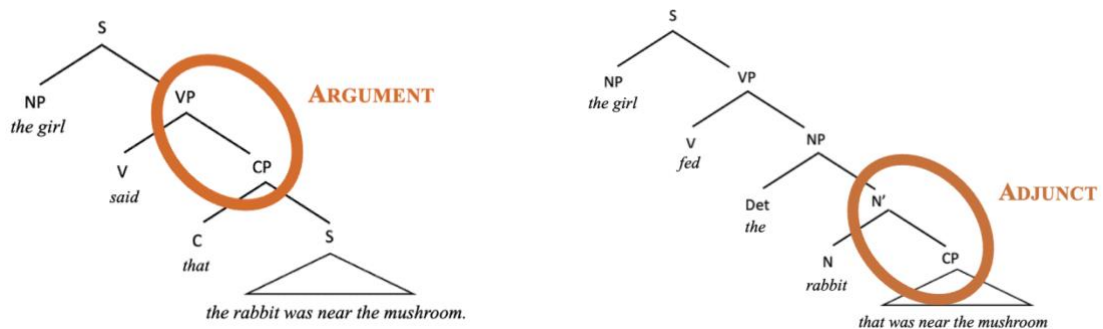


Figure 4 Syntactic structure of past-under-past complement clauses (left) and past-under-past relative clauses (right)

effect of this structural property is that tenses in relative clauses should always be interpreted independently of tense in the main clause.¹¹

Thus, from a syntax-driven perspective, all logical possibilities of temporal order in Figure 3 should be feasible interpretations of past-under-past relative clause descriptions (for a similar analysis of temporal adjunct clauses, see Arregui & Kusumoto, 1998; Kubota, Lee, Smirnova, & Tonhauser, 2009; Von Stechow & Grønn, 2013b, but see Ogihara, 1996).

2.1.2.2. Temporal interpretations are driven by discursive pragmatic function

Syntactic structure may determine temporal interpretations in yet another way: Rather than the syntactic dependencies between tenses, people might take into account the discursive function of relative clauses when modelling the temporal relations of a complex situation. From such a perspective, temporal interpretations are determined at the interface of syntax and pragmatics.

In pragmatics, there is a broad distinction between background and foreground information where the former “embellish[es] and elaborate[s] upon the foreground”, and the latter “charts the progress of a narrative through time” (Ehrlich, 1987, p. 373). Relative clauses serve either to restrict the space of potential referents (i.e., restrictive relative clauses) or to add more information about the antecedent head noun to the discourse (i.e., non-restrictive relative clauses, see Stockwell, Schacter, & Partee, 1973; Fox & Thompson, 1990; Lambrecht, 1988; Quirk, Greenbaum, Leech, & Svartvik, 1985; Thomson & Martinet, 1985; Levinson, 1983; Cinque, 2020).

In both cases, the content of relative clauses should not contribute to the progression of the storyline, but provide background information against which a narrative unfolds. In particular, it has been argued that non-restrictive relative clauses fall within the domain of backgrounded

¹¹ Analyses differ as to the locus of these properties: One proposal argues that adjunct clauses are not governed by the main verb phrase (Enç, 1987); another, that they cannot be arguments to intensional predicates (Abusch, 1997; von Stechow & Grønn, 2013a), but they can move out of the scope of the main verb to a higher structural position (Ogihara, 1996; Stowell, 2007).

information (cf. Chierchia & McConnell-Ginet, 1990; Depraetere, 1996; Potts, 2005), whereas main clauses are generally identified with the discursive foreground.

While background in discourse is not necessarily and exclusively temporal, it has been pointed out that there is a connection between pragmatic background and temporality (Hopper, 1979; Talmy, 1975; Reinhart, 1984; Wårvik, 2004; see also Vega, Rinck, Díaz, & León, 2007): In terms of temporal interpretations, then, a focus on the background function of relative clauses would predict that a situation described by a relative clause should be restricted to taking place either at the same time as, or before, a situation described by a main clause (corresponding to Figures 3a and 3b).¹²

2.1.2.3. Event structure determines temporal construal

Alternatively, people's temporal interpretations may not be driven by the syntactic properties, nor the discourse-pragmatic function of past-under-past relative clauses, but by the internal structure of the described situations themselves. In contrast to syntactic and discourse-pragmatic theories, an event-structural perspective suggests that complex temporal structures are established through different linguistic means (Carroll & Stutterheim, 2010; Klein, 1994, 2000, 2008, 2009; von Stutterheim, Carroll, & Klein, 2003). Crucially, these temporal structures can be parametrized as follows:

First, each temporal expression (e.g. *The girl fed the carrot to the rabbit*) maps to a situation time which always refers to a non-finite component of the utterance (e.g., feeding). That is, the

¹² Note that there is an alternative version of the discourse-pragmatic account, where embedded situation descriptions may not enter people's temporal models of a situation at all (cf. Hopper, 1979, Gunkel, 2007; Hofherr, 2014, but also acquisition research, showing that young children frequently ignore situations encoded by relative clauses or evaluate them according to their pragmatic function, de Villiers, Tager Flusberg, Hakuta, & Cohen, 1979; Hamburger & Crain, 1982; but see Eisenberg, 2002). Since such a perspective yields the same predictions as structure-driven accounts – that is, all logical temporal orders of two past descriptions should be acceptable for past-under-past relative clauses – we do not pursue these alternative accounts in the present paper. In addition, as we argued above, there is evidence that temporal inferences are generated even in the absence of explicit temporal information.

situation time includes knowledge about the typical duration and structure of an event (e.g. that feeding a carrot to a rabbit changes the state of the carrot over time).

Second, temporal expressions evoke another temporal parameter, the so-called topic time (cf. Klein, 1994, 2000, 2008, 2009), which refers to the finite component of the utterance that includes, for instance, tense and aspect. Most relevant for the current studies, simple past indicates that the topic time lies within the situation time, but also includes its rightmost boundaries (i.e., culmination point in terms of achievements and accomplishments, a post-state in terms of state and activity eventualities, see Vendler, 1967). Therefore, topic time relates the situation time to the utterance time. For instance, *The girl fed the rabbit* happened before utterance time, and the action is concluded.

Third, and most importantly for our purpose, a temporal anchor links the topic time to other temporal intervals.¹³ In the simplest cases, a temporal anchor is congruent with the utterance time: When hearing *The girl fed the rabbit*, comprehenders can evaluate only the sentence relative to the utterance time. In more complex descriptions, however, the context or previous discourse may provide other salient times as temporal anchors. For instance, both the tense and the temporal adverb in a sentence such as *The girl fed the rabbit yesterday at 5:30 pm* relate the topic time to a temporal anchor in the past (i.e., *yesterday at 5:30 pm*). Importantly, because simple past always selects a subset of the situation time including its end boundary (i.e., culmination point/post-state), this also constrains the starting boundary of a situation, that is, when it started relative to its temporal anchor: For *The girl fed the rabbit yesterday at 5:30 pm*, the sentence is true when the girl started feeding the rabbit before the time of the temporal anchor, as long as she was still feeding at 5:30 pm. At the same time, the sentence cannot be true when the feeding situation begins significantly after this time (e.g., yesterday at 5:45 pm).

¹³ See also other formal accounts of tense (Hinrichs, 1986; Partee, 1973; Reichenbach, 1947) who argue for the existence of a time of reference, however with conceptual overlaps with Klein's topic time.

In that sense, the way in which the temporal order between described situations is construed should be constrained by the way in which people link them to temporal anchors.

For a past-under-past relative clause like in (1), there are two ways of identifying the temporal anchor of the sentence: The temporal anchor could either be placed at the time of the situation described by the main clause (i.e., the girl feeding the rabbit), or at the time of the event described by the relative clause (i.e., the rabbit being near the mushroom). Whereas, in the first case, the rabbit must have been near the mushroom before the girl fed it (as in Figure 3b), in the second case, the girl must have fed the rabbit before it was near the mushroom (as in Figure 3c). But how do people decide which of the two temporal intervals should temporally anchor the other one?

We argue that, at this point, the conceptual distinction between events and states comes into play: If we accept that events are cognitively more salient than states, then we should find that descriptions encoding more salient events (i.e., the girl feeding the rabbit) are identified as the temporal anchor for descriptions of less salient states (i.e., the rabbit being near the mushroom). An event-structural account would thus predict that in past-under-past relative clause constructions, stative descriptions may begin before eventive descriptions in time, but not after it, due to the constraints of the simple past on the leftmost boundary; and this should be the case regardless of their syntactic or discourse-pragmatic properties in the sentence. Crucially, such a perspective is not only consistent with the fact that semantic distinctions, rather than structural properties, seem to determine how well people recall complex temporal sentences (Clark & Clark, 1968), but also with the fact that different expectations about the temporal location of described situations are grounded in event-structural properties (Gennari, 2004; cf. Hinrichs, 1986 on the temporal status of eventive and stative predicates).

2.1.3. The current studies: Predictions and outline

Overall, the three theoretical accounts reviewed here make different predictions about the temporal interpretation of past-under-past relative clauses: A crucial difference – between syntax-driven accounts, on the one hand, and discourse-pragmatic and event-structural accounts, on the other – is whether temporal relations are construed between tense in the relative clause and tense in the main clause.

Syntax-driven accounts (Figure 5A) would predict that both forward-shifted and back-shifted interpretations should, in principle, be available for the sentences in (4): Given that the embedded tense of the relative clause should be interpreted as depending only on the utterance time, none of the different encodings of linguistic event type should privilege any particular temporal order.

- (4) a. *The girl **fed** the rabbit that **was** near the mushroom.*
 b. *The rabbit that the girl **fed** **was** near the mushroom.*
 c. *The girl **stood** near the rabbit that **hopped** towards/away from the mushroom.*
 d. *The girl **stood** near the rabbit that **was** near the mushroom.*
 e. *The girl **fed** the rabbit that **hopped** towards/away from the mushroom.*

From a discourse pragmatic perspective, people's temporal interpretations of the data in (4) should yield the following pattern (Figure 5B): If people use the discourse function of relative clauses to establish the background of a described complex situation, their interpretations should be driven by the linguistic form: That is to say, for any sentence in (4), relative clause descriptions should be ordered first.

By contrast, an event-structural account (Figure 5C) would predict clear interpretative preferences as a function of event-structural differences: Since cognitively more salient eventive descriptions should be identified as temporal anchors for less salient stative descriptions, the situation described by the relative clause should occur before the situation

described by the main clause in (4a), but after it in (4b-c). When there are no event-structural differences, as in (4d-e), these preferences should disappear.

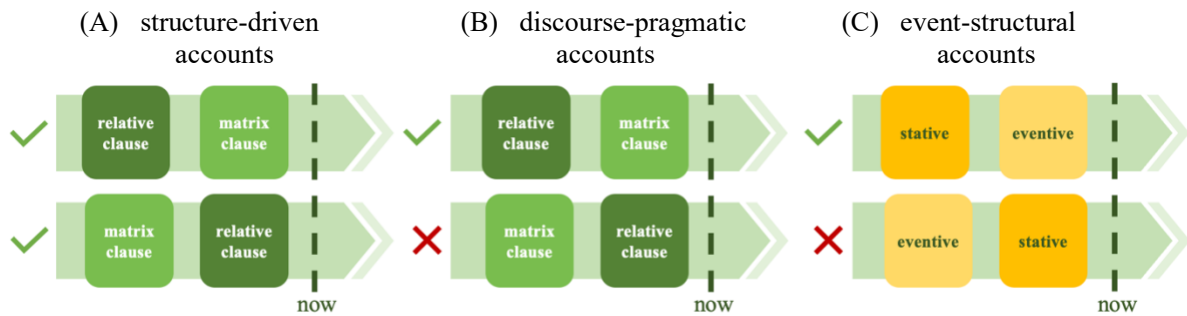


Figure 5 Predicted ratings for temporal sequences from structure-driven accounts (A), discourse pragmatic accounts (B) and event-structural accounts (C) for past-under-past relative clause constructions

To test these different predictions, we conducted seven pre-registered experiments with English past-under-past relative clauses, building on previous results (Marx & Wittenberg, 2022). Each experiment was replicated in German to further investigate the effects of language-specific properties on temporal interpretations: In other complex sentences such as complement clauses (e.g., *The girl said that the rabbit was near the mushroom*), an embedded stative predicate in past tense can be interpreted simultaneously with the time of the matrix clause, in addition to signaling anteriority – a phenomenon also known as Sequence of Tense (SoT). While English is clearly a SoT language (cf. Abusch, 1997; Declerck, 1988; Ogiwara, 1995; Stowell, 2007), the case is less clear for German (Boskovic, 2012, but see Giorgi & Pianesi, 2000; von Stechow, 2005). Furthermore, unlike in English, any potential temporal ambiguities can be resolved via German subjunctive marking (Eckardt, 2021; Helbig & Buscha, 2001). Therefore, we tested our predictions in German as well, for additional crosslinguistic evidence from a non-strictly SoT language.

The logic of our studies was inspired by the experimental design of Dermidache and Lungu (2008), who investigated children's temporal construals of complement and relative clauses in French: 5-to-7-year-olds saw one of two sequences of situations, followed by a linguistic

prompt. The children were then asked whether the prompt was an adequate description of the previous sequence. The objective of the study was to investigate which temporal orderings of situations were in principle acceptable to children, rather than simply prompting their dominant interpretations.

In our studies, we closely followed the experimental design of Dermidache and Lungu (2008). We used short video clips, followed by linguistic descriptions of the videos. We asked participants to judge whether they accepted the sentence as matching the video. However, in order to pit the predictions of the syntax-driven, discourse-pragmatic, and event-driven accounts against each other, we reduced the linguistic context to relative clauses and the past tense: Adult participants watched animated short clips like represented in Figure 1A, and then judged whether a corresponding past-under-past relative clause description matched the video. We tested the hypothesis that linguistic event type (eventive vs. stative descriptions) impacts temporal construal: Experiment 1 contrasts the predictions of syntax-driven accounts against predictions from both other accounts. The main clause always described an event, and the relative clause always described a state (e.g., 4a). We used while-clauses as critical contrasts, which, due to the semantics of this temporal connective, should elicit low ratings for both sequences of situations. Two follow-up experiments controlled for potential confounds of the visual materials (Experiment 1b) and the effects of reference resolution (Experiment 1c).

Experiment 2 pits the predictions of the event-structural perspective against the predictions of discourse-pragmatic accounts: We inverted the mapping between syntactic encoding and linguistic event type, so that main clauses always encoded stative descriptions and relative clauses always encoded eventive descriptions (e.g., 4b, Experiment 2a). As in Experiment 1b, we examined the influence of reference resolution (Experiment 2b) and further eliminated differences in information structure (e.g., 4c, Experiment 2c).

Experiment 3 examined how past-under-past relative clauses are interpreted in the absence of event-structural differences in order to explicitly contrast the predictions of syntax-driven accounts with those of discourse-pragmatic accounts: Here, we compared relative clause constructions that contained either two eventive or two stative descriptions (e.g., 4d-e).

2.2. Methods

2.2.1. General materials and procedure

In all experiments, participants first watched a short animation showing two events with two characters involved (e.g., a girl feeding a rabbit, the rabbit hopping towards/away from the girl), embedded in two static positions of one of the characters (e.g., the rabbit being near the mushroom, the girl standing near the rabbit). In total, we created eight video pairs as shown in Figure 6, which lasted between 9 and 12 seconds and included a distractor location (e.g., a barn) and a distractor character (e.g., a second rabbit). These video pairs were used in all experiments except Experiments 1b, 1c, and 2b, where we edited either the video length or the presence of the distractor character to control for alternative explanations (see Materials for each experiment). Video pairs differed in terms of the temporal sequence: One event happened once after one state (e.g., ‘feeding’ frames after ‘being’ frames, Figure 6A) and once before it (e.g., ‘feeding’ frames before ‘being’ frames, Figure 6B); the other event happened once before the



Figure 6 Temporal arrangement of the displayed stative situations (i.e., ‘being’, ‘standing’ frames) and the eventive situations (i.e., ‘feeding’, ‘hopping’ frames) in the video materials.

other state ('hopping' frames before 'standing' frames, Figure 6A) and once after it ('hopping' frames after 'standing' frames, Figure 6B).¹⁴

After each clip, followed by a 1s blank screen, participants read a corresponding linguistic description in the center of the screen and rated how well the sentence matched the video on a 5-point rating scale. We constructed eight critical sentence pairs per experiment, using simple past for the English experiments and preterite for the German versions (the historical and structural equivalent of the English simple past, Helbig & Buscha, 2001; Klein, 2000, Comrie, 1995). The specific linguistic materials for each experiment are outlined individually in the following chapters, but they all contained a main clause and a relative clause describing the events and states shown in the video.

Throughout the experiments, videos were classified according to their respective linguistic description, that is, in terms of the temporal ordering between main and relative clause: Back-shifted videos always showed the relative clause situation occurring before the main clause situation, whereas in forward-shifted videos, the relative clause scene always happened after the main clause scene.

Participants completed 30 trials per experiment: Each session began with two practice trials to familiarize participants with the experimental procedure. After practice, participants performed eight critical trials, sixteen filler trials and four sanity checks in a randomized order. Filler trials consisted of sentences that either linked two eventive descriptions using coordinating conjunctions (e.g., *The dog barked and the bird flew away*) or specified the location of a single eventive description using prepositional phrases (e.g., *The ghost scared the girl in front of the houses*). Ten filler trials were constructed in such a way that the linguistic descriptions were incompatible with the preceding video, in order to elicit negative judgements about some of the

¹⁴ All materials can be found under <https://osf.io/6ae5m/>.

linguistic stimuli. In the remaining six filler trials, the sentence descriptions matched the corresponding videos.

While the filler trials remained the same across experiments, the sentences in sanity check trials were modified to match the linguistic manipulations of each experiment (see the respective sections for further details). In line with critical trials, the filler and sanity check trials consistently used the simple past for the English and the preterite for the German sentence materials. Sanity check animations showed two situations occurring simultaneously – a temporal sequence that is predicted by any theoretical account of embedded tense. Sanity checks should therefore elicit high ratings across trials.

To counterbalance between subjects, participants were randomly assigned to one of four lists using a Latin-square design (sentence type, video sequence, test item). All experiments were programmed in PsychoPy and conducted online at Pavlovio.org.

2.2.2. General analysis

Only complete data sets were included in the analyses. We fitted linear mixed effects regression models (i.e., *lmer* from the *lme4* package, Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2014). Significance of the full model was assessed using likelihood ratio tests between a full model and reduced models that individually excluded fixed predictors and interactions, but included the same random effect structure (Barr, Levy, Scheepers, & Tily, 2013). In all experiments, sentence type and video sequence were contrast-coded as fixed effects, and participants and items were included as random intercepts to control for potential variance due to these two factors. We used a minimal intercept-only structure to ensure model convergence.

In a second step of the analysis, planned pairwise comparisons were conducted to assess the interactions of the regression models. Preregistrations for all conducted experiments and analyses conducted are available at <https://osf.io/6ae5m/> and <https://osf.io/8hwmj>.

2.2.3. Experiment 1a

2.2.3.1. Participants

We recruited 61 English speakers from Amazon Mechanical Turk via the CloudResearch platform (Litman, Robinson, & Abberbock, 2017) and 51 German participants via the clickworker.com platform, as well as from the researcher's personal environment. Overall, the experiment lasted approximately 20 minutes and participants were only included in the analysis if all trials were completed and the sanity check criterion was met (i.e., less than 75% of the unambiguously true sanity check trials were rated high). Based on these criteria, we excluded 11 English and 3 German participants, thus, leaving 50 English and 48 German data sets for analysis.

This and all following experiments were performed in compliance with institutional guidelines and have been approved by the Psychological Research Ethics Board of (PREBO) of the Central European University (reference number 2023/02). Informed consent was obtained.

2.2.3.2. Materials

In Experiment 1a, critical sentence pairs consisted of two types of clauses: First, a main clause describing an event was combined with a relative clause specifying the state of the patient character (5). Second, we included *while*-clause constructions as control sentences (6), for which no shifted interpretations should be available – i.e., none of the temporal sequences in Figure 3 – due to the semantics of these temporal connectives. Parallel to the relative clause

constructions, *while*-clause constructions described the event in the main clause and the state in the subordinate clause (6).

- (5) MC_{EVENTIVE} – RC_{STATIVE}

[*The girl fed the rabbit*]_{MC} [*that was near the mushroom*]_{RC}

- (6) MC_{EVENTIVE} – WHILE_{STATIVE}

[*The girl fed the rabbit*]_{MC} [*while it was near the mushroom*]_{WHILE}

The German critical sentences were close translations of the English material, encoding an eventive description in the main and a state in the relative clause (7) or *während*-clause (9).

- (7) MC_{EVENTIVE} – RC_{STATIVE}

[*Das Mädchen fütterte den Hasen,*]_{MC} [*der neben dem Pilz saß*]_{RC}

- (8) MC_{EVENTIVE} – WÄHREND_{STATIVE}

[*Das Mädchen fütterte den Hasen,*]_{MC} [*während er neben dem Pilz saß*]_{WÄHREND}

In both English and the German materials, sanity checks mirrored the critical sentence pairs: Two relative clause trials, similar to (5) and (7), were contrasted with two trials consisting of *while*- or *während*-clause constructions, similar to (6) and (8).

We used the animated videos described in Section 2.2.1: Back-shifted sequences were videos as in Figure 6A, so that the state encoded by the relative clause (i.e., being near the mushroom) referred to the first video frames, whereas forward-shifted sequences were videos as in Figure 6B, where the relative clause state referred to the last frames of the video.

2.2.3.3. Results

English and German ratings are shown in Figure 7: Overall, participants rated video sequences in combination with relative clause descriptions higher than while-clause descriptions, with a clear preference for back-shifted sequences. For the English data, the statistical analysis

revealed a main effect of sentence type: For both video sequences, while-clause descriptions were rated significantly lower ($\text{mean}_{\text{WH}}=1.84$, $\text{SD}=1.26$) than relative clause constructions ($\text{mean}_{\text{RC}}=3.39$, $\text{SD}=1.57$, $Df=1$, $\chi^2=129.50$, $p<0.001$), indicating that the task was easy to understand and participants responded as predicted to mismatching video-sentence pairs.

Furthermore, participants rated forward-shifted sequences ($\text{mean}_{\text{FORW}}=2.26$, $\text{SD}=1.49$) as significantly lower than back-shifted sequences ($\text{mean}_{\text{BACK}}=2.96$, $\text{SD}=1.68$, main effect of video sequence: $Df=1$, $\chi^2=29.37$, $p<0.001$). Importantly, there was a significant interaction between sentence type and video sequence ($Df=1$, $\chi^2=17.56$, $p<0.001$) which was driven by the relative clauses according to planned pairwise comparisons: *While*-clauses were rated low regardless of the temporal sequence in the videos ($\text{SD}_{\text{WH-BACK}}=1.35$, $\text{SD}_{\text{WH-FORW}}=1.17$; $\beta=-0.08$, $t=-0.90$, $p=0.37$), whereas relative clauses were rated as significantly higher when they described back-shifted ($\text{SD}_{\text{RC-BACK}}=1.30$) compared to a forward-shifted sequences ($\text{SD}_{\text{RC-FORW}}=1.59$; $\beta=-0.06$, $t=-5.9$, $p<0.001$).

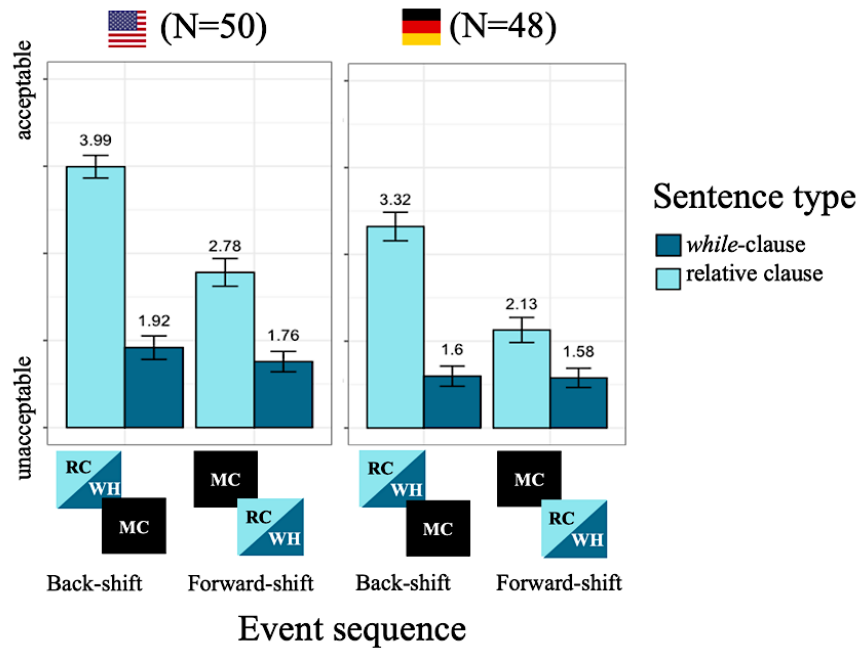


Figure 7 Mean ratings for English clause types describing shifted event sequences in Exp. 1a. The y-axis corresponds to the 5-point rating scale, as participants saw it in each trial, the x-axis shows video sequences classified according to the temporal ordering between main and relative clause. Error bars represent Standard Errors.

The German data replicated this pattern of results: We found main effects of sentence type ($Df=1$, $\chi^2=75.88$, $p<0.001$) and video sequence ($Df=1$, $\chi^2=23.10$, $p<0.001$) as well as a significant interaction ($Df=1$, $\chi^2=22.65$, $p<0.001$), resulting from the ratings of relative clauses: Whereas *während*-clauses were rated low in all video sequences ($SD_{WH-BACK}=1.14$, $SD_{WH-FORW}=1.09$; $\beta=-0.01$, $t=-0.12$, $p=0.90$), relative clauses were rated significantly higher in combination with back-shifts ($SD_{RC-BACK}=1.61$) than with forward-shifts ($SD_{RC-FORW}=1.42$; $\beta=-0.60$, $t=-5.44$, $p<0.001$).

2.2.3.4. Discussion of Experiment 1a

The results of Experiment 1a suggest that when a main clause describes an event and a relative clause describes a state, the embedded situation tends to be interpreted relative to the main clause event: Both English and German participants judged video sequences significantly lower when the relative clause situation occurred after the main clause situation (i.e., forward-shifted sequences) than when the relative clause situation occurred before the main clause situation (i.e., back-shifted sequences).

Overall, this pattern of results suggests that the temporal interpretation of past-under-past relative clauses is not exclusively determined by the syntactic properties of a sentence: While from a syntax-driven perspective, main and relative clause are structurally independent and should license both temporal sequences as possible interpretations, this was not the case according to our data. Rather in line with an event-structural and a discourse-pragmatic perspective, participants preferred video sequences in which a salient eventive description in the main clause anchored a backgrounded stative description in the relative clause.

However, there is another reasonable explanation for our findings: In the forward-shifted condition, the video clips generally ended after the patient of the described event had moved to the location described by the stative predicate (e.g., the rabbit hopping to the mushroom). Thus,

rather than reflecting people's sensitivity to the internal structure of situations, ratings for the forward-shifts may have been low because participants did not perceive the states in the videos as fully in the past: That is, participants may have detected a mismatch between the state in the video and the simple past or preterite in the linguistic description, forcing a terminated interpretation of the predicate at the time of speech. To control for this alternative, we ran a modified version of the experiment in which the states in the visual materials were unambiguously in the past.

2.2.4. Experiment 1b: Terminated states

2.2.4.1. Participants

50 native English speakers and 56 native German speakers were recruited via the Testable Minds platform and clickworker.com, respectively. We excluded 2 English and 4 German participants due to incomplete data sets, as well as 4 German participants due to failing the sanity check criterion. In total, we submitted 48 English and 48 German complete data sets to analysis.

2.2.4.2. Materials

The linguistic materials were taken from Experiment 1a, but critical video material for the forward-shifted sequences (i.e., Figure 6B) was edited: After moving to the location of the described state, the patient character (e.g., the rabbit) lingered at the location (e.g., the mushroom) for two or three seconds, before leaving the scene on one side of the screen. Apart from these changes, the visual materials (i.e., Figure 6A) were the same as in Section 2.2.1; back- and forward-shifted sequences were coded as in Experiment 1a.

2.2.4.3. Results

As in Experiment 1a, video sequences – especially back-shifted sequences – were rated higher in combination with relative clauses, and equally low in combination with *while*-clauses (see Fig. 8). We found a main effect of sentence type ($Df=1$, $\chi^2=93.55$, $p<0.001$), a main effect of video sequence ($Df=1$, $\chi^2=16.90$, $p<0.001$) and a significant interaction in the English data ($Df=1$, $\chi^2=8.97$, $p=0.003$). Planned pairwise comparisons revealed that the interaction was driven by the fact that relative clauses were rated lower in combination with forward-shifted ($SD_{RC-FORW}=1.57$) than with back-shifted sequences ($SD_{RC-BACK}=1.22$; $\beta=-0.49$, $t=-4.86$, $p<0.001$). In contrast, *while*-clauses were judged low in all video sequence conditions ($SD_{WH-BACK}=1.55$, $SD_{WH-FORW}=1.48$; $\beta=-0.07$, $t=-0.62$, $p=0.54$).

Similarly, the German data yielded main effects of sentence type ($Df=1$, $\chi^2=85.10$, $p<0.001$) and video sequence ($Df=1$, $\chi^2=6.80$, $p=0.009$) and a significant interaction ($Df=1$, $\chi^2=18.06$, $p<0.001$), which resulted from differences between the relative clause conditions (SD_{RC-

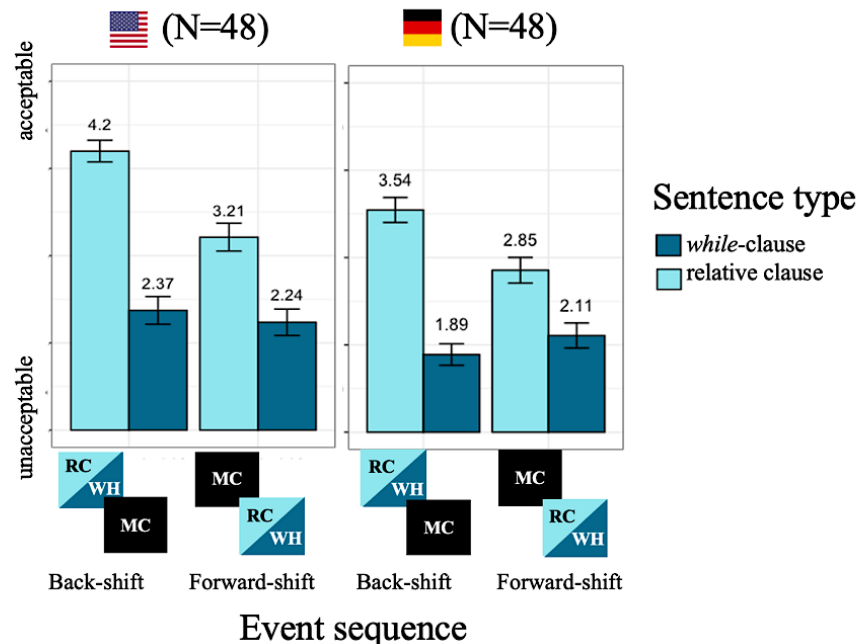


Figure 8 Mean ratings for English clause types describing shifted event sequences in Exp. 1b. The y-axis corresponds to the 5-point rating scale, as participants saw it in each trial, the x-axis shows video sequences classified according to the temporal ordering between main and relative clause. Error bars represent Standard Errors.

BACK=1.40, SD_{RC-FORW}=1.43; β =-0.34, t =-3.37, p =0.001), rather than the *während*-clause conditions (SD_{WH-BACK}=1.21, SD_{WH-FORW}=1.40; β =0.10, t =1.15, p =0.25).

2.2.4.4. Discussion of Experiment 1b

Experiment 1b replicated the pattern of results from Experiment 1a: Participants rated past-under-past relative clause constructions lower for forward-shifted sequences than for back-shifted sequences. Given that the states in the video materials were unambiguously terminated (i.e., the rabbit hopping away from the mushroom), the overall lower ratings for forward-shifts cannot be attributed to a potential mismatch between the semantics of the past tense in the linguistic descriptions and the visually perceived situation.

However, our visual materials allow for another alternative explanation: Following the experimental design of Dermidache and Lungu (2008), we included a distractor character in each video clip, in the form of a second potential referent of the patient character (e.g., a second rabbit at a distractor location, see Figure 6). Therefore, the videos may have forced a contrastive alternative interpretation for the object NP which could have influenced temporal judgements in two different ways: On the one hand, the competing NP referent may have shifted people's attention to the main clause situation in order to identify the correct character in the videos. If so, the description of the relative clause would need to be true at the time of the main clause description, rendering forward-shifted sequences worse, only in the context of reference resolution. On the other hand, the contrast alternatives may have distracted participants from the temporal structure of the sentence altogether: If so, people should have focused on whether the relative clause description identified the correct character, with the temporal order between the two situations being less critical to their judgments. This would predict lower judgments for forward-shifted sequences in the absence of a second distractor character in the videos.

To investigate these potential influences of reference resolution, we conducted another modified version of Experiment 1a, excluding the visual contrast alternatives from the visual materials.

2.2.5. Experiment 1c: Non-restrictive relative clauses

2.2.5.1. Participants

54 native English speakers were recruited via testable.com. We excluded 2 participants and a further 4 participants due to incomplete data sets and our sanity check criterion respectively. For the German data, we collected 53 data sets via clickworker.com. 5 participants were excluded due to their performance in the sanity check trials so that, in total, we submitted 48 English and 48 German data sets to statistical analysis.

2.2.5.2. Materials

We used the same linguistic materials and coding of video sequence as in the previous experiments (see Section 2.2.1): For back-shifts, the relative clause state (i.e., being near the mushroom) was shown in the first frames of the animation (Figure 6A), and for forward shifts, it was shown at the end of the video (Figure 6B).

The animations were edited to remove the second potential referent for the object NP in the main clause from all critical videos. Since the overall pattern of results was the same for Experiments 1a and 1b, and to keep variability in video length to a minimum, we edited the original videos from Experiment 1a: That is, in the forward-shifted sequences, the videos always ended after the patient character (e.g., the rabbit) had reached the state location (e.g., the mushroom).

2.2.5.3. Results

Participants' ratings are shown in Figure 9. In line with the previous experiments, forward-shifts were judged lower for relative clause descriptions than back-shifts, but overall, both video sequences were rated higher in combination with relative clauses than with *while*-clauses. In the English data, there was a main effect of sentence type ($Df=1$, $\chi^2=64.66$, $p<0.001$), a main effect of video sequence ($Df=1$, $\chi^2=22.16$, $p<0.001$) and a significant interaction ($Df=1$, $\chi^2=15.52$, $p<0.001$). This pattern was replicated by the German data (main effect of sentence type: $Df=1$, $\chi^2=33.15$, $p<0.001$; video sequence: $Df=1$, $\chi^2=4.85$, $p=0.028$; interaction: $Df=1$, $\chi^2=9.54$, $p=0.002$).

For both the English and German data, pairwise comparisons revealed that the interaction was driven by people's judgements in the relative clause conditions (English: $SD_{RC-BACK}=1.47$, $SD_{RC-FORW}=1.48$; $\beta=-0.53$, $t=-4.94$, $p<0.001$; German: $SD_{RC-BACK}=1.53$, $SD_{RC-FORW}=1.45$; $\beta=-0.31$, $t=-2.84$, $p=0.005$) rather than from differences between the *while*-clause ($SD_{WH-BACK}=1.47$, $SD_{WH-FORW}=1.48$; $\beta=0.01$, $t=0.01$, $p=0.999$).

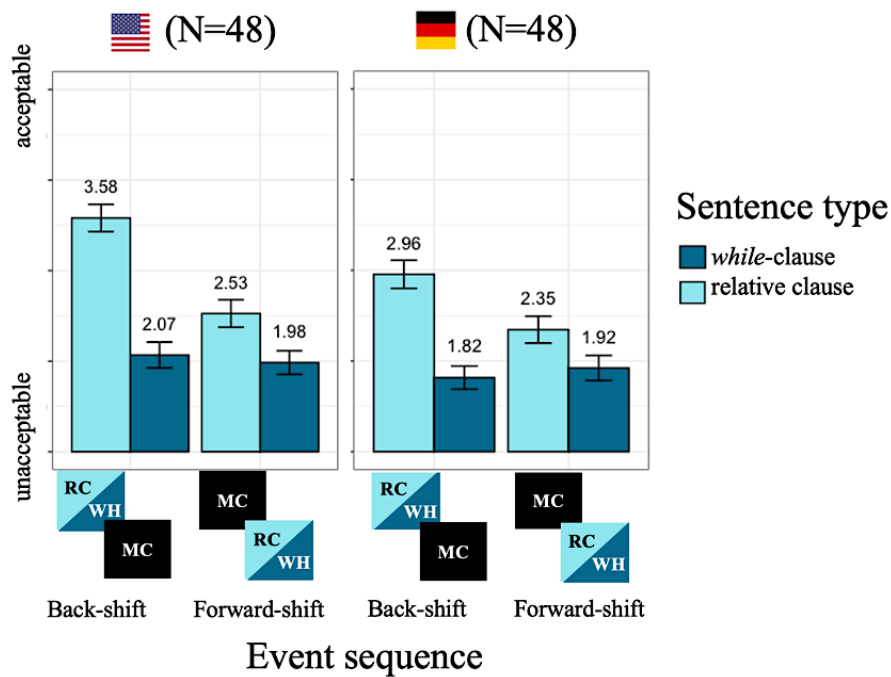


Figure 9 Mean ratings for English clause types describing shifted event sequences in Exp. 1c. The y-axis corresponds to the 5-point rating scale, as participants saw it in each trial, the x-axis shows video sequences classified according to the temporal ordering between main and relative clause. Error bars represent Standard Errors.

$_{\text{BACK}}=1.40$, $SD_{\text{WH-FORW}}=1.27$; $\beta=-0.04$, $t=-0.43$, $p=0.67$) or the *während*-clause conditions ($SD_{\text{WH-BACK}}=1.25$, $SD_{\text{WH-FORW}}=1.35$; $\beta=0.05$, $t=0.57$, $p=0.57$).

We conducted exploratory analyses between Experiments 1a and 1c, to investigate potential quantitative differences as a function of reference resolution: The data sets from the two experiments were pooled, and the number of possible object NP referents was added to the regression model as a contrast-coded fixed effect (i.e., two referents vs. one referent). Apart from the previous pattern of results (i.e., the two main effects and the interaction between sentence type and video sequence), we found no main effect of the number of possible referents in English ($Df=1$, $\chi^2=0.22$, $p=0.63$) and in German ($Df=1$, $\chi^2=0.57$, $p=0.44$). In English, however, the differences between *while*- and relative clauses were smaller in Experiment 1c ($mean_{\text{RC}}=3.05$, $SD_{\text{RC}}=1.57$ vs. $mean_{\text{WH}}=2.03$, $SD_{\text{WH}}=1.33$) than in Experiment 1a ($mean_{\text{RC}}=3.39$, $SD_{\text{RC}}=1.57$ vs. $mean_{\text{WH}}=1.84$, $SD=1.26$), whereas, in German, judgments differed significantly less between video sequences in Experiment 1c ($mean_{\text{BACK}}=2.39$, $SD_{\text{BACK}}=1.50$ vs. $mean_{\text{FORW}}=2.13$, $SD_{\text{FORW}}=1.41$) than in Experiment 1a ($mean_{\text{BACK}}=2.46$, $SD_{\text{BACK}}=1.64$ vs. $mean_{\text{FORW}}=1.85$, $SD_{\text{FORW}}=1.29$). Statistically, this was reflected by a significant interaction between sentence type and number of possible referents in both languages (English: $Df=1$, $\chi^2=9.25$, $p=0.002$, German: $Df=1$, $\chi^2=4.13$, $p=0.04$) and a significant interaction between video sequence and number of possible referents in German ($Df=1$, $\chi^2=4.58$, $p=0.03$).

2.2.5.4. Discussion of Experiment 1c

In Experiment 1c, we replicated the pattern of results from Experiments 1a and 1b: Despite the absence of the distractor character, participants rated relative clause descriptions significantly lower when they followed a forward-shifted sequence than when they followed a backward-shifted sequence. Experiment 1c demonstrates that the removal of a contrast alternative from

the referential space did not lead to qualitative differences in people's temporal interpretations for past-under-past relative clauses, although differences were less pronounced when only one potential NP referent was available.

Overall, the results of Experiments 1a-c support the claim that event-structural properties of a sentence, rather than its syntactic encoding, determine the way in which temporal relations are construed for past-under-past relative clauses. However, these results provide only indirect evidence for an event-structural account, leaving open an alternative explanation at the interface of syntax and pragmatics: Forward-shifted sequences may have conflicted with a pragmatic interpretation strategy based on the kind of information relative clauses usually convey in discourse. On such a view, relative clauses generally convey backgrounded and main clauses foregrounded information (Chierchia & McConnell-Ginet, 1990; Depraetere, 1996; Potts, 2005). In our study, participants may have been sensitive to the different discourse functions of main and relative clauses: Video sequences in which relative clause situations preceded main clause situations could have been rated higher because embedded situations, as backgrounds, establish the context relative to which main situations are interpreted. Such an interpretive strategy – a relative-clause-first strategy – would predict a similar pattern of results to the event-structural account for Experiments 1a-c.

In order to distinguish between a discourse-pragmatic explanation and event structure as the primary determinant of temporal interpretations of relative clauses, we conducted a second set of experiments with an inverted mapping between linguistic event type and syntax. If the event-structural properties of a sentence establish temporal order, we would predict forward-shifted sequences – i.e., video sequences that show relative clause descriptions after the main clause descriptions – to be ameliorated in sentences that encode the event in the relative clause and the state in the main clause: Here, a less salient stative description should be anchored by a more salient eventive description, regardless of its structural encoding in a complex sentence. If, on

the other hand, relative clauses are inherently backgrounded, regardless of the linguistic event type of the predicate, back-shifted sequences should elicit higher ratings than forward-shifted sequences, even when the mapping between event structure and syntactic encoding is inverse.

2.2.6. Experiment 2a

2.2.6.1. Participants

We recruited 49 native English speakers and 54 native German speakers online (testable.org, clickworker.com) and excluded 1 English and 6 German participants due to their underperformance in the sanity check trials. In total, we submitted 48 English and 48 German datasets to analysis.

2.2.6.2. Materials

In Experiment 2a, linguistic materials were altered, systematically manipulating the syntactic encoding of the two linguistic event types (i.e., stative vs. eventive descriptions): In stative relative clauses, the main clause described an event and the relative clause described a state (9). In eventive relative clauses, a state was encoded by the main clause and an event by the relative clause (10). In both conditions, the semantic content of predicates remained constant such that sentence pairs always denoted the same event (e.g., feeding) and the same location of the patient character in the videos (e.g., being near the mushroom).

(9) MC_{EVENTIVE} – RC_{STATIVE}

[The girl fed the rabbit]_{MC} [that was near the mushroom]_{RC}

(10) MC_{STATIVE} – RC_{EVENTIVE}

[The rabbit] [that the girl fed]_{RC} [was near the mushroom]_{MC}

German linguistic stimuli closely matched the English sentence pairs, contrasting stative relative clauses (11) with eventive relative clauses (12).

(11) MC_{EVENTIVE} – RC_{STATIVE}

[*Das Mädchen fütterte den Hasen,*]_{MC} [*der neben dem Pilz saß*]_{RC}

(12) MC_{STATIVE} – RC_{EVENTIVE}

[*Der Hase,*] [*den das Mädchen fütterte,*]_{RC} [*saß neben dem Pilz*]_{MC}

In sanity check trials, two sentences encoded the state in the relative clause, like (9) and (11), and two sentences described the state in the main clause, like (10) and (12).

Visual materials were the same as in Experiment 1a (see Section 2.2.1), but the coding of video sequences differed: For stative relative clauses, backshifted videos showed the relative clause state (i.e., being near the mushroom) before the main clause event (i.e., feeding, Figure 6A) and forward-shifted videos showed it after (Figure 6B). For eventive relative clauses, back-shifted sequences related to videos where the relative clause description (i.e., feeding girl) was shown before the main clause state (i.e., standing, Figure 6B), and forward-shifted sequences related to videos where the relative clause event was shown after the main clause event (Figure 6A).

2.2.6.3. Results

Ratings from both languages are shown in Figure 10: Participants rated back-shifted sequences higher than forward-shifted sequences when the relative clause encoded a state. By contrast, forward-shifts were rated higher than back-shifts in combination with eventive relative clauses. For English, we found no main effect of sentence type ($Df=1$, $\chi^2=0.11$, $p=0.74$) or video sequence ($Df=1$, $\chi^2=0.11$, $p=0.74$), however, the interaction between the two predictor variables was significant ($Df=1$, $\chi^2=122.19$, $p<0.001$). For German, there was no main effect of sentence type ($Df=1$, $\chi^2=0.05$, $p=0.82$), but a main effect of video sequence ($Df=1$, $\chi^2=5.25$, $p=0.02$) as well as a significant interaction ($Df=1$, $\chi^2=53.18$, $p<0.001$).

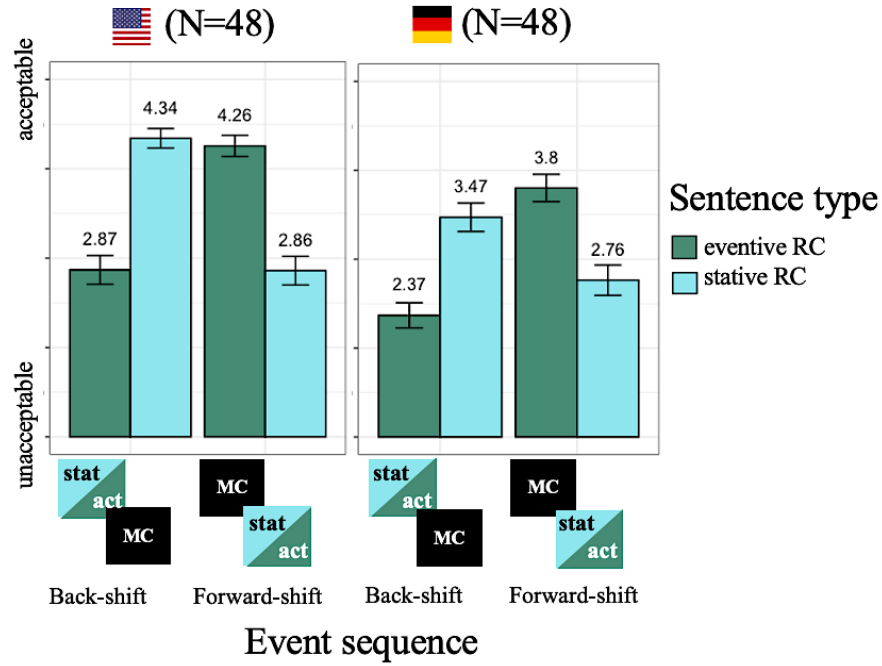


Figure 10 Mean ratings for English clause types describing shifted event sequences in Exp. 2a. The y-axis corresponds to the 5-point rating scale, as participants saw it in each trial, the x-axis shows video sequences classified according to the temporal ordering between main and relative clause. Error bars represent Standard Errors.

Planned pairwise comparisons showed that the interaction resulted from different video sequences in both the eventive relative clause (English: $\beta=0.69$, $t=6.96$, $p<0.001$; German: $\beta=0.72$, $t=6.85$, $p<0.001$) and the stative relative clause condition (English: $\beta=-0.74$, $t=-7.62$, $p<0.001$; German: $\beta=-0.35$, $t=-3.03$, $p=0.003$): While participants rated back-shifted sequences higher for stative relative clauses (English: $SD_{STATE-BACK}=1.08$, $SD_{STATE-FORW}=1.57$; German: $SD_{STATE-BACK}=1.57$, $SD_{STATE-FORW}=1.67$), the opposite was the case for eventive relative clauses: Forward-shifts were judged significantly higher than back-shifts (English: $SD_{EVENT-BACK}=1.57$, $SD_{EVENT-FORW}=1.16$; German: $SD_{EVENT-BACK}=1.39$, $SD_{EVENT-FORW}=1.51$).

2.2.6.4. Discussion of Experiment 2a

Experiment 2a asked whether forward-shifts were rated lower in the previous experiments because (a) they deviated from people's temporal intuitions based on the event-structural properties of a sentence (i.e., cognitively less salient stative descriptions are generally anchored

by more salient eventive descriptions) or (b) they conflicted with pragmatic expectations as to what function relative and main clauses have in discourse (i.e., generally encoding backgrounded vs. foregrounded information).

We found that reversing the mapping between linguistic event type and syntactic encoding also reversed the pattern of results from Experiments 1a-c: When the main clause described an event and the relative clause a state, the embedded state was generally interpreted prior to the main clause. However, when the state was described by the main clause and the event described by the relative clause, main clauses were more likely to be ordered prior to relative clauses.

Overall, Experiment 2a demonstrates that people do not assume that relative clauses provide the context of main clause situations. On the contrary, judgments were highly sensitive to event-structural differences such that video sequences in which described states happened first were preferred, regardless of their structural encoding in main or subordinate clause. Experiment 2a thus provides evidence that event structure, rather than pragmatics, impacts people's temporal interpretations of complex sentences.

A follow-up experiment aimed to replicate this pattern of results, while controlling for the effects of reference resolution along the lines of Experiment 1c (i.e., eliminating the distractor character from the critical video clips).

2.2.7. Experiment 2b: Non-restrictive relative clauses

2.2.7.1. Participants

49 native English and 49 native German speakers were recruited online via Testable Minds and clickworker.com. After excluding one participant in each language due to the sanity check criterion, we submitted 48 English and 48 German datasets in total to statistical analysis.

2.2.7.2. Materials

Linguistic materials were the same as in Experiment 2a. In line with Experiment 1c, we removed the second potential object NP in all critical video materials. Back- and forward-shifted sequences were coded as in Experiment 2a.

2.2.7.3. Results

The pattern of results replicated our findings in Experiment 2a (see Fig. 11): Compared to forward-shifts, back-shifts were rated higher for stative relative clauses, but lower for eventive relative clauses. In the English data, we did not find a main effect of sentence type ($Df=1$, $\chi^2=0.02$, $p=0.90$), nor a main effect of video sequence ($Df=1$, $\chi^2=0.62$, $p=0.43$), but the interaction turned out significant ($Df=1$, $\chi^2=63.83$, $p<0.001$). Similarly, in the German data, there was no main effect of sentence type ($Df=1$, $\chi^2=0.68$, $p=0.41$), but a main effect of video sequence ($Df=1$, $\chi^2=5.09$, $p=0.02$) and a significant interaction between the two independent variables ($Df=1$, $\chi^2=69.33$, $p<0.001$).

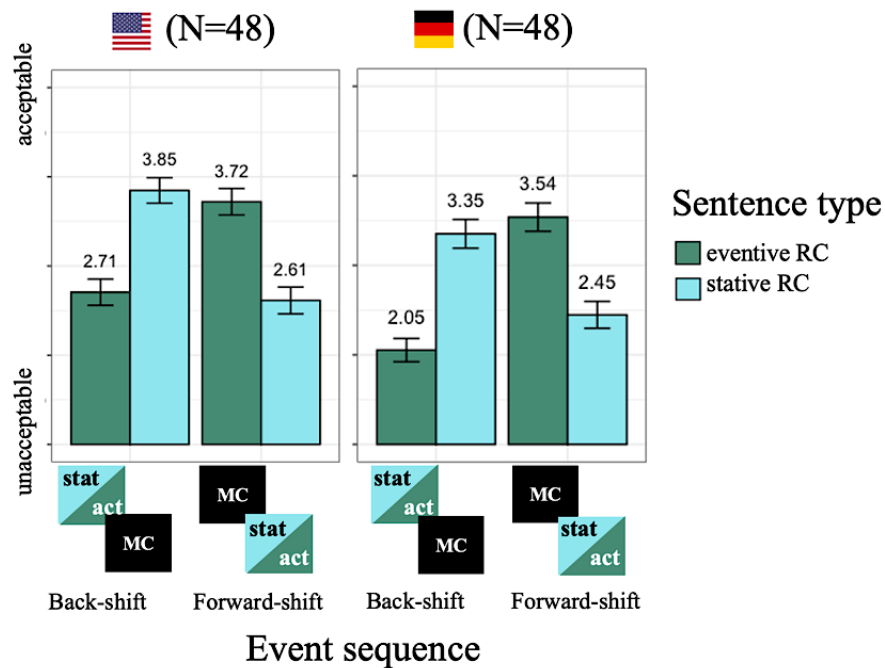


Figure 11: Mean ratings for English clause types describing shifted event sequences in Exp. 2b. The y-axis corresponds to the 5-point rating scale, as participants saw it in each trial, the x-axis shows video sequences classified according to the temporal ordering between main and relative clause. Error bars represent Standard Errors.

Planned pairwise comparisons showed that participants rated back-shifted sequences significantly higher with stative relative clauses ($SD_{\text{STATE-BACK}}=1.40$, $SD_{\text{STATE-FORW}}=1.47$; $\beta=-0.62$; $t=-5.94$, $p<0.001$), while preferring forward-shifted sequences with eventive relative clauses ($SD_{\text{EVENT-BACK}}=1.44$, $SD_{\text{EVENT-FORW}}=1.45$; $\beta=0.51$, $t=4.86$, $p<0.001$). This was also the case for German stative ($SD_{\text{STATE-BACK}}=1.57$, $SD_{\text{STATE-FORW}}=1.47$; $\beta=-0.45$; $t=-4.13$, $p<0.001$) and eventive relative clauses ($SD_{\text{EVENT-BACK}}=1.28$, $SD_{\text{EVENT-FORW}}=1.55$; $\beta=0.74$, $t=7.25$, $p<0.001$).

2.2.7.4. Discussion of Experiment 2b

The findings of Experiments 2a and 2b suggest that event structure determines how people construe temporal relations in past-under-past relative clause descriptions: Participants generally rated sentences higher when the described state happened first in a video sequence, followed by a more salient eventive description. This was true no matter which of the two linguistic event types was encoded by relative or main clause – a pattern of results neither predicted by standard syntax-driven accounts of embedded tense, nor by a discourse pragmatic strategy where relative clauses are interpreted as the default background of a complex situation. There are, however, two caveats regarding our results: For one thing, the effects could have been solely driven by the visual materials. In both Experiments 2a and 2b, back-shifted sequences in the context of stative relative clauses (Fig. 6A) functioned as forward-shifted sequences in the eventive relative clause condition. Likewise, forward-shifted sequences for stative relative clauses (Fig. 6B) were always used as the forward-shifted sequences for eventive relative clauses. Therefore, participants could have been prone to rate the same video materials either high or low – independent of the linguistic descriptions we provided. While we did not find such a preference in the while-control conditions in Experiment 1a-c, these concerns cannot be entirely remedied: While-clauses could have patterned differently because

they never express sequential order so that their strong semantic restrictions may have overridden people's preferences for one particular type of video.

Yet, an even stronger objection is that not only the linguistic event type of main and relative clause was switched in Experiment 2a and 2b: In keeping the semantic content between critical sentences largely constant, we also changed the attachment of the relative clause from the object (9) to the subject of the main clause (10). Furthermore, we varied the syntactic constituent from which the complementizer was extracted within the relative clause: *That* could either replace the subject (9) or object in the relative clause (10).

Crucially, both structural properties of relative clauses may bear potential confounds: Relativizing the sentential subject also puts the relative clause into topical position. Forward-shifts may thus have been ameliorated because information was packaged differently across critical sentences (i.e., topic vs. focus). Additionally, forward-shifted sequences matched the linear order of the sentence which may have triggered another interpretative strategy, well established in discourse pragmatics: Situations that are ordered first in speech are assumed to be ordered first in time (e.g., Diessel, 2008; Horn, 2022; Levelt, 1989; Tai, 1983).

Secondly, it has been shown that most object relative clauses are harder to process than most subject relative clauses (Staub, 2010; Can Mutlu, Canbeyli, & Saybaşı, 2023; but see Chen, Ning, Bi, & Dunlap, 2008; Real & Christiansen, 2007) as well as to acquire (Brandt, Diessel, & Tomasello, 2008; Brown, 1971; Costa, Lobo, & Silva, 2011; Diessel & Tomasello, 2005; Friedmann & Costa, 2010; McKee, Snedeker, & McDaniel, 1998). In other words, contrasting object- and subject-extracted relative clauses may place the eventive description in the relative clause into linguistic focus for reasons other than the internal structure of the complex situation. Our results may therefore reflect saliency differences that are grounded in factors other than the event structure of a sentence.

In Experiment 2c, we controlled for such confounding factors, by keeping both the information structure as well as the syntactic properties of the modified noun constant across critical sentences: Relative clauses were always subject-extracted and only relativized the object NP of the main clause.

2.2.8. Experiment 2c: Linear order

2.2.8.1. Participants

We recruited 54 native English speakers (testable.org), though 6 participants had to be excluded because they failed to reach the sanity check criterion. 53 German speakers participated in Experiment 2c, with 3 of them being excluded due to their performance in the sanity check trials and 2 because of incomplete data sets. 48 English and 48 German datasets were submitted to the statistical analysis.

2.2.8.2. Materials

To keep the syntactic properties of the relative clause constructions consistent, critical sentence pairs were modified such that main and relative clauses systematically referred to different aspects of the video scenes (see Figure 6): While stative relative clauses were the same as in Experiment 2a-b, such that the event (i.e., feeding) was encoded by the main clause and the target location by the relative clause (i.e., being near the mushroom, see 13), in eventive relative clauses in Experiment 2c, the change of location was described in the relative clause (i.e., hopping to or away from the mushroom) and the stative spatial relation between the characters was described in the main clause (i.e., standing near each other). Importantly, we used sentences such as (14a) in combination with forward-shifted sequences and sentences such as (14b) in combination with back-shifted sequences to ensure that descriptions matched the videos.

(13) MC_{EVENTIVE} – RC_{STATIVE}

[*The girl fed the rabbit*]_{MC} [*that was near the mushroom*]_{RC}

(14) MC_{STATIVE} – RC_{EVENTIVE}

a. [*The girl stood near the rabbit*]_{MC} [*that hopped to the mushroom*]_{RC}

b. [*The girl stood near the rabbit*]_{MC} [*that hopped away from the mushroom*]_{RC}

Sanity check trials used the same descriptions as in Experiment 2a. As in the previous experiments, German sentence pairs (15-16) were constructed in line with the English materials (13-14).

(15) MC_{EVENTIVE} – RC_{STATIVE}

[*Das Mädchen fütterte den Hasen,*]_{MC} [*der neben dem Pilz saß*]_{RC}

(16) MC_{STATIVE} – RC_{EVENTIVE}

a. [*Das Mädchen war neben dem Hasen,*]_{MC} [*das zu dem Pilz hoppelte*]_{RC}

b. [*Das Mädchen war neben dem Hasen,*]_{MC} [*das vom Pilz weghoppelte*]_{RC}

The same video materials were used as in Experiment 2a, but unlike in Experiment 2a, back- and forward-shifts related to the same videos across critical sentences: Back-shifted videos showed the relative clause descriptions (i.e., being near the mushroom, hopping away from the mushroom) before the main clause (i.e., feeding, standing), as in Figure 6A; and forward-shifted videos showed the relative clause descriptions after the main clause descriptions (Figure 6B).

2.2.8.3. Results

In Experiment 2c, participants still rated back-shifted sequences higher than forward-shifted sequences for stative relative clauses, and vice versa for eventive relative clauses. However, we found slight differences compared Experiments 2a and 2b (see Figure 12): Forward-shifts were generally rated higher in the eventive relative clause condition than back-shifts in the stative

relative clause conditions. This was also reflected by the statistical analysis: While there was still no main effect of video sequence ($Df=1$, $\chi^2=0.36$, $p=0.55$), we found a main effect of sentence type ($Df=1$, $\chi^2=4.59$, $p=0.03$), on top of the significant interaction ($Df=1$, $\chi^2=74.29$, $p<0.001$) in the English data, demonstrating that eventive relative clause constructions were generally rated lower than stative relative clause constructions ($\text{mean}_{\text{RC-EVENT}}=3.08$, $\text{SD}_{\text{RC-EVENT}}=1.51$ $\text{mean}_{\text{RC-STATE}}=3.51$, $\text{SD}=1.59$). The pattern of results was replicated by the German data (main effect of sentence type: $Df=1$, $\chi^2=10.11$, $p=0.001$; no main effect of video sequence: $Df=1$, $\chi^2=0.37$, $p=0.54$; significant interaction: $Df=1$, $\chi^2= 32.45$, $p<0.001$).

In both languages, planned pairwise comparisons revealed that the interaction was driven by differences between the two video sequences for both stative and eventive relative clauses: When combined with stative relative clause descriptions, back-shifted sequences were rated significantly higher (English: $\text{SD}_{\text{STATE-BACK}}=1.38$; German: $\text{SD}_{\text{STATE-BACK}}=1.55$) than forward-shifted sequences (English: $\text{SD}_{\text{STATE-FORW}}=1.59$; $\beta=-0.57$, $t=-5.35$, $p<0.001$, German: $\text{SD}_{\text{STATE-FORW}}=1.47$; $\beta=-0.37$, $t=-3.42$, $p<0.001$). In combination with eventive relative clauses, on the

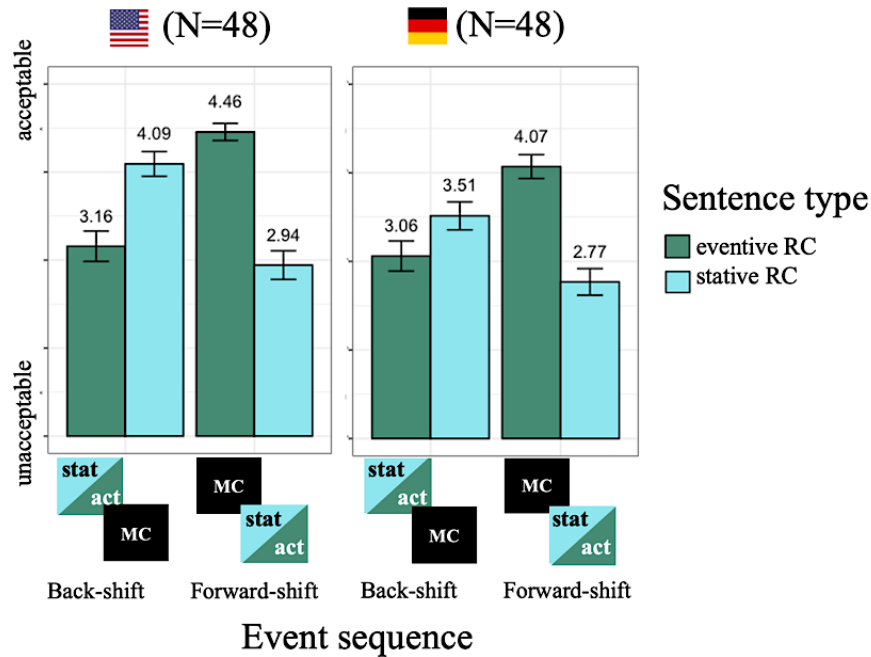


Figure 12 Mean ratings for English clause types describing shifted event sequences in Exp. 2c. The y-axis corresponds to the 5-point rating scale, as participants saw it in each trial, the x-axis shows video sequences classified according to the temporal ordering between main and relative clause. Error bars represent Standard Errors.

other hand, back-shifts were judged significantly lower (English: $SD_{\text{EVENT-BACK}}=1.68$, German: $SD_{\text{EVENT-BACK}}=1.66$) than forward-shifts (English: $SD_{\text{EVENT-FORW}}=0.95$; $\beta=0.65$, $t=6.57$, $p<0.001$; German: $SD_{\text{EVENT-FORW}}=1.32$; $\beta=0.50$, $t=4.65$, $p<0.001$).

To investigate the main effect of sentence type more thoroughly, we also ran non-preregistered pairwise comparisons between the forward-shifted eventive and the back-shifted stative conditions (i.e., the low bars in Fig. 7) as well as between the back-shifted eventive and the forward-shifted stative conditions (i.e., the high bars in Fig. 7). Results showed that, both in the English as well as the German data, there was only a significant difference between the high rated conditions (English: $\beta=0.36$, $t=2.12$, $p=0.04$, German: $\beta=0.56$, $t=2.68$, $p=0.008$), not the low rated conditions (English: $\beta=0.21$, $t=0.90$, $p=0.37$, German: $\beta=0.29$, $t=1.28$, $p=0.199$).

2.2.8.4. Discussion of Experiment 2c

Experiment 2c was set up to (a) show that participants did not have particular preferences towards the video materials and (b) keep the structural properties – rather than semantic content – of the critical sentence pairs stable. In doing so, we controlled for potential structural and information structural influences in our previous findings. Overall, we found the same overarching tendency as in previous experiments: Judgments were predicted by event-structural differences, such that relative clause constructions were rated higher when the corresponding videos first showed a situation relating to the state description instead of the event description. The results from Experiment 2c thus replicate Experiment 2a-b, demonstrating that temporal order was derived independently from the state's encoding in the subordinate or the main clause. In addition, Experiment 2c verified that people's responses in Experiment 2a-b (i.e., ordering stative descriptions first) did not result from particularities of our visual materials: Different videos were rated high or low as a function of linguistic description.

Unlike in Experiment 2a and 2b, however, we found that encoding an event in the relative clause yielded higher ratings than encoding a state in the relative clause. This effect was driven by the fact that forward-shifted eventive relative clause constructions were rated higher than back-shifted stative relative clause constructions. How can this difference be explained? One major difference between the two conditions becomes evident when matching the linguistic descriptions to the video materials: In cases where the relative clause encoded a state, the back-shifted videos did not align with the linear order of the sentence, such that people first saw the situation relating to the state description but encountered it second during sentence processing. The opposite was the case for forward-shifted videos that combined with eventive relative clause descriptions: Here, the state description came first both in the linguistic (i.e., in the stative main clauses) as well as in the visual stimuli, showing the spatial relation between the characters. In that sense, linear order may have promoted people's preference to arrange a stative description prior to a more salient eventive description. At the same time, our results also reveal that linear order did not intervene with event-structural effects: Forward-shifted stative relative clauses, where the order between linguistic and visual materials matched, were not rated significantly higher than back-shifted eventive relative clauses which ran counter to the temporal order in the videos.

While our results so far suggest that, for past-under-past relative clauses, event-structural properties rather than alternative discourse pragmatic strategies determine the way in which described situations are ordered in time, it is an empirically open question how temporal order is derived in the absence of event-structural differences.

In Experiment 3, we therefore used sentences in which relative and main clause descriptions were equally salient, either containing two eventive or two stative descriptions. If, in such cases, people revert to discourse pragmatic principles, there should either be an effect of linear order (i.e., rating video sequences higher that show main clause situations first), or a general relative-

clause-first strategy should emerge (i.e., ordering relative clause situations as temporal backgrounds of main clause situations).

Note that syntax-driven accounts and event-structural accounts would predict a similar pattern of results: Both backward-shifted and forward-shifted sequences should be rated equally, either because main and embedded tenses are interpreted independently from each other or because both situation descriptions can serve as the temporal anchor of the complex sentence.

However, event structure might still affect people's temporal interpretations: It has been noted that stative descriptions tend to receive an overlapping interpretation (Dowty, 1986; Gennari, 2004; Partee, 1984). Given that our visual materials depicted clear sequences of situations without temporal overlap, double-stative sentences should therefore be harder to match to either a back-shifted or a forward-shifted sequence. We would predict that, overall, sentences containing two stative descriptions should be rated lower than double-eventive descriptions.

2.2.9. Experiment 3

2.2.9.1. Participants

38 native English speakers and 48 native German speakers were recruited via online platforms (i.e., Testable Minds, clickworker.com). One English and 3 German participants were excluded because of the sanity check criterion, resulting in 37 English and 45 German complete data sets for analysis.

2.2.9.2. Materials

We constructed new critical sentence pairs that either contained two stative or two eventive descriptions: In double-stative relative clause constructions, main clauses specified the stative spatial relation between two characters (i.e., standing near each other), while relative clauses encoded the target location of the patient character (i.e., being near the mushroom, see 17). In

double-eventive relative clauses, the event (i.e., feeding) was described by the main clause and the change of location was encoded by the relative clause (i.e., hopping to or away from the mushroom). In line with Experiment 2c, we matched the directionality of the change of location descriptions to the forward-shifted (18a) and back-shifted sequences in the video (18b).

(17) MC_{STATIVE} – RC_{STATIVE}

[*The girl stood near the rabbit*]_{MC} [*that was near the mushroom*]_{RC}

(18) MC_{EVENTIVE} – RC_{EVENTIVE}

a. [*The girl fed the rabbit*]_{MC} [*that hopped to the mushroom*]_{RC}

b. [*The girl fed the rabbit*]_{MC} [*that hopped away from the mushroom*]_{RC}

German sentence pairs mirrored the English sentence pairs (17-18), either describing a state (19) or a change of location in the relative clause (20a-b).

(19) MC_{STATIVE} – RC_{STATIVE}

[*Das Mädchen stand neben dem Hasen,*]_{MC} [*der neben dem Pilz saß*]_{RC}

(20) MC_{EVENTIVE} – RC_{EVENTIVE}

a. [*Das Mädchen fütterte den Hasen,*]_{MC} [*der zu dem Pilz hoppelte*]_{RC}

b. [*Das Mädchen fütterte den Hasen,*]_{MC} [*der vom Pilz weghoppelte*]_{RC}

Using the same video materials as in the previous experiments (see Section 2.2.1), we coded video sequences as follows: Videos that showed relative clause descriptions (i.e., being near the mushroom, hopping away from the mushroom) before the main clause (i.e., standing , feeding) were coded as back-shifted sequences (see Figure 6A); and videos that showed the relative clause descriptions after the main clause description were coded as forward-shifted sequences (see Figure 6B).

2.2.9.3. Results

In Experiment 3, participants rated video sequences similarly high in combination with double-eventive sentences, and similarly low in combination with double-stative sentences (see Fig. 13). We found a main effect of sentence type ($Df=1$, $\chi^2=131.40$, $p<0.001$), but no main effect of video sequence ($Df=1$, $\chi^2=0.16$, $p=0.69$) or interaction between the two independent variables in English ($Df=1$, $\chi^2=0.14$, $p=0.70$): Thus, sentences encoding two events were rated significantly higher than sentences encoding two states (mean_{EVENT}=4.37, SD=1.06, mean_{STATE}=2.78, SD=1.49), independent of the video sequence to which they were matched. This pattern of results was confirmed by planned pairwise comparisons, yielding no difference between back- and forward-shifted sequences for both double-stative (SD_{STATE-BACK}=1.52, SD_{STATE-FORW}=1.47; $\beta=0.002$, $t=-0.02$, $p=0.99$) and double-eventive relative clauses (SD_{EVENT-BACK}=1.15, SD_{EVENT-FORW}=0.96; $\beta=0.05$, $t=0.65$, $p=0.52$).

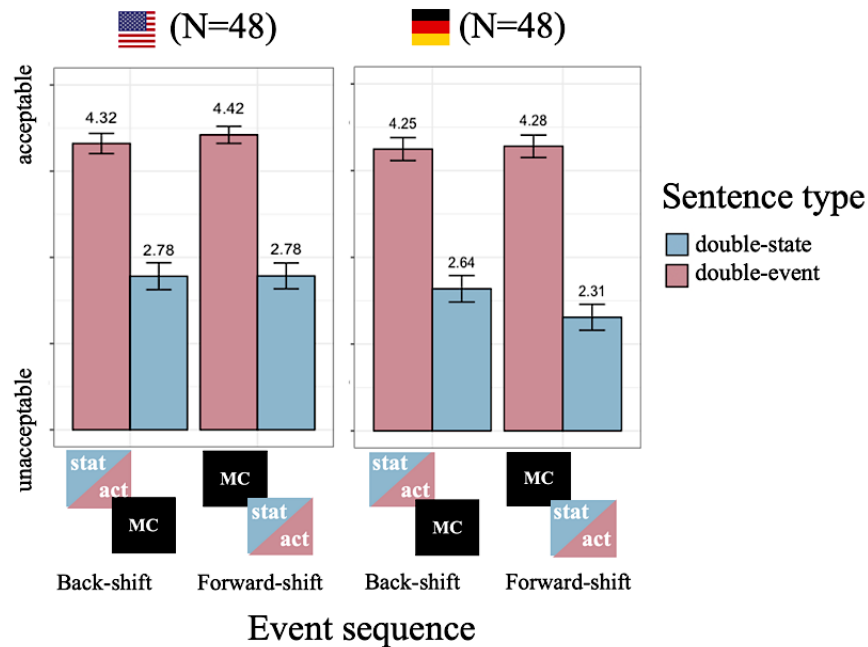


Figure 13 Mean ratings for English clause types describing shifted event sequences in Exp. 3. The y-axis corresponds to the 5-point rating scale, as participants saw it in each trial, the x-axis shows video sequences classified according to the temporal ordering between main and relative clause. Error bars represent Standard Errors.

2.2.9.4. Discussion of Experiment 3

Experiment 3 asked how people would interpret past-under-past relative clauses in the absence of event-structural differences. The question was: What guides temporal interpretations of complex situations when the internal structure of its subparts does not suffice to establish sequential order?

We found that, for relative clause descriptions, ratings did not differ across video sequences when linguistic event type was kept uniform. Unlike in previous experiments, relative clause situations could either happen before or after main clause situations such that neither linear order predicted preferences regarding people's temporal interpretations (i.e., privileging video sequences where the main clause happened first), nor did a relative-clause-first strategy (i.e., where back-shifted sequences should be always judged better than forward-shifted sequences). However, even when predicates were of the same event type, we found an effect of linguistically encoded event structure: Sentences containing two stative descriptions were generally rated lower than sentences that related to two eventive descriptions. People were thus sensitive towards the fact that states, but not events, tend to extend in time and often overlap with other situations.

Overall, Experiment 3 demonstrates that uniform linguistic event type encoding eradicates preferences for temporal order, lending additional credence to the claim that the internal structure of situations plays a crucial role in how temporal structure is mapped onto complex linguistic descriptions.

2.2.10. Between-language comparison

As argued above, one of the aims of this paper was to have additional cross-linguistic validation of the data of English, a SoT language, from German, in which the status of Sequence of Tense is less clear. As we show in Table 1, in all seven experiments, speakers of these two languages

showed the same preferences. Nevertheless, there seemed to be minor differences in German, compared to English, which we analyze below.

2.2.10.1. Statistical analysis

Due to the overall lower numerical ratings in German compared to English in most experiments (see Fig. 14), we conducted a non-preregistered analysis of language differences. We pooled the data for each experiment (i.e., 1abc, 2abc, 3) and added language (i.e., German vs. English) as contrast-coded fixed effect to the linear regression model. Furthermore, we controlled for variation between experiment versions (i.e., a, b, c) in Experiment 1 and 2 by including experiment version as a random intercept.

Table 1 Outline of all experiments including the manipulations and the preferred video sequence in both English (E) and German (G).

Exp.	Manipulation	Accepted sequence	Example sentences
1	a. MC _{EVENTIVE} – RC _{STATIC}	E: backshift G: backshift	[The girl fed the rabbit] _{MC} [that was near the mushroom] _{RC}
	MC _{EVENTIVE} – WHILE _{STATIC}	E: neither acceptable G: neither acceptable	[The girl fed the rabbit] _{MC} [while it was near the mushroom] _{RC}
	b. control: videos with clearly terminated states	replication of Experiment 1a	
	c. control: videos without ambiguous NP reference	replication of Experiment 1a	
2	a. MC _{EVENTIVE} – RC _{STATIC}	E: backshift G: backshift	[The girl fed the rabbit] _{MC} [that was near the mushroom] _{RC}
	MC _{STATIC} – RC _{EVENTIVE}	E: forward-shift G: forward-shift	[The rabbit] [that the girl fed] _{RC} [was near the mushroom] _{MC}
	b. control: without ambiguous NP reference	replication of Experiment 2a	
	c. control: sentences with the same information-structural and syntactic properties	E: backshift G: backshift E: forward-shift G: forward-shift	[The girl fed the rabbit] _{MC} [that was near the mushroom] _{RC} [The girl stood near the rabbit] _{MC} [that hopped to/away from the mushroom] _{RC}
3	MC _{STATIC} – RC _{STATIC}	E: neither acceptable G: neither acceptable	[The girl stood near the rabbit] _{MC} [that was near the mushroom] _{RC}
	MC _{EVENTIVE} – RC _{EVENTIVE}	E: both acceptable G: both acceptable	[The girl fed the rabbit] _{MC} [that hopped to/away from the mushroom] _{RC}

2.2.10.2. Results

In the first two experiments, German ratings were significantly lower overall (Exp. 1: $M_{\text{GERMAN}}=2.34$, $SD_{\text{GERMAN}}=1.50$; Exp. 2: $M_{\text{GERMAN}}=3.10$, $SD_{\text{GERMAN}}=1.61$) than English ratings (Exp. 1: $M_{\text{ENGLISH}}=2.72$, $SD_{\text{ENGLISH}}=1.62$; Exp. 2: $M_{\text{ENGLISH}}=3.49$, $SD_{\text{ENGLISH}}=1.56$). This was reflected by a main effect of language in Experiment 1 ($Df=1$, $\chi^2=19.21$, $p<0.001$) and Experiment 2 ($Df=1$, $\chi^2=17.34$, $p<0.001$). In Experiment 3, this language difference also emerged numerically ($M_{\text{GERMAN}}=2.37$, $SD_{\text{GERMAN}}=1.64$ vs. $M_{\text{ENGLISH}}=3.58$, $SD_{\text{ENGLISH}}=1.52$), but did not reach significance as a main effect ($Df=1$, $\chi^2=2.58$, $p=0.11$) or as a significant interaction (all interactions: $Dfs=1$, $\chi^2s<1.33$, $ps>0.25$).

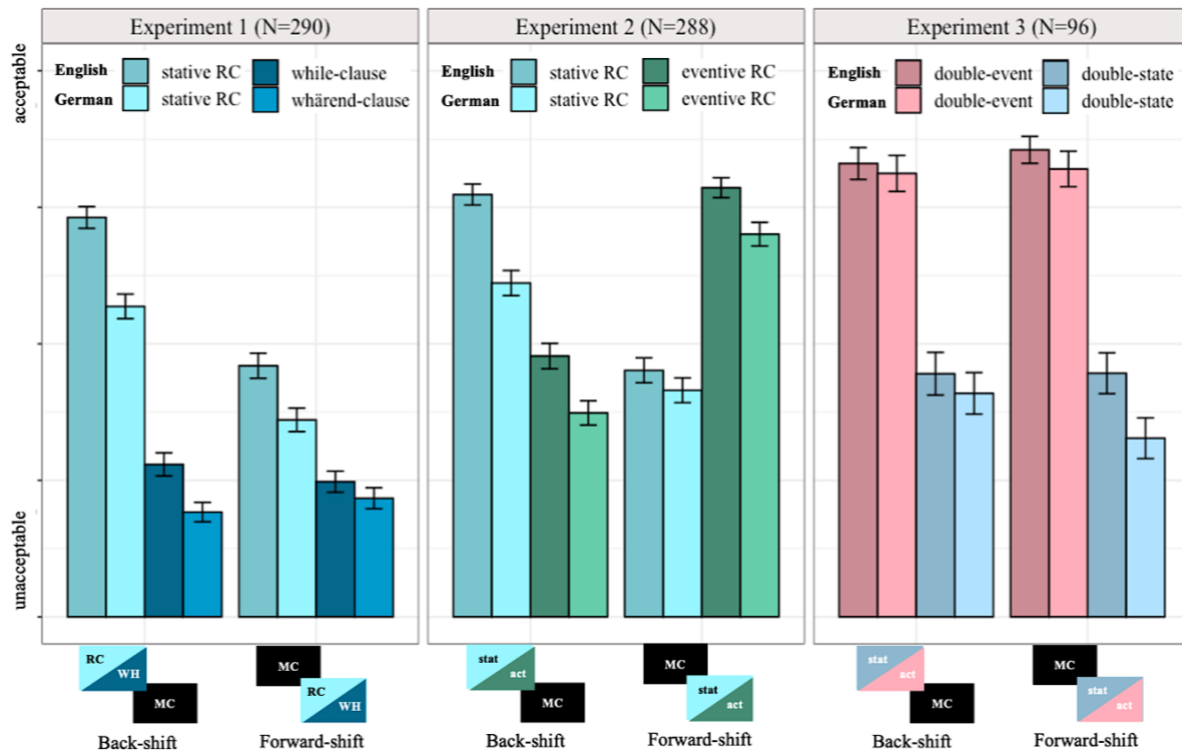


Figure 14 Mean ratings for English and German across experiments. The y-axis corresponds to the 5-point rating scale, as participants saw it in each trial, the x-axis shows video sequences classified according to the temporal ordering between main and relative clause. Error bars represent Standard Errors. Dark-shaded colors indicate English, light-shaded colors indicate German ratings.

2.2.10.3. Discussion of between-language comparisons

Overall, we found similar patterns of results in both languages, demonstrating that linguistic encoding of event structure is a strong cue to temporal order across different languages.

However, we also found cross-linguistic differences: German speakers provided generally lower ratings than English speakers, as was statistically confirmed for two out of three experiment series. What are the possible explanations for this result?

The first one is rooted in differences between the German and English tense systems. We used German preterite, which has been argued to be historically and structurally similar to the simple past form in English (Helbig & Buscha, 2001; Klein, 2000; Comrie, 1995). However, it has been observed that the preterite has started to give way to perfect tense (i.e., a phenomenon commonly known as *Präteritumschwund*, ‘disappearance of the preterite’), except in written German. The lower overall rates in German may therefore reflect an ongoing historical process of language change.

Second, and more importantly, the German tense-aspect system does not have a morphological progressive like the English gerund (Barrie & Spreng, 2009; Helbig & Buscha, 2001; Krause, 2002; Miller, DeLuca, Berndt, Iverson, & Rothman, 2017). When German speakers want to express that a situation is ongoing, they need to resort to relatively infrequent lexical or structural alternatives (see Flecken, 2011; von Stutterheim, Carroll, & Klein, 2009) that are often associated with regional vernaculars (van Pottelberge, 2004). In contrast to English, German simple tenses can therefore have a progressive interpretation, as shown in (22, see also Miller et al., 2017).

- | | |
|---|-----------|
| (22) <i>Ich schreibe/schrieb (gerade) einen neuen Absatz.</i> | (German) |
| <i>*I write/wrote (right now) a new paragraph.</i> | (English) |

But how does this use of the German preterite explain why, in our data, speakers were less willing to accept both back- and forward-shifted sequences for past-under-past relative clauses? According to event-structural accounts, progressive markers locate the topic time (i.e., the time about which the assertion is made) entirely within the situation time, whereas in simple past forms, the topic time also includes the culmination or the post-state of a situation description.

In the case of German preterite, the overlap between a progressive and a bounded interpretation may have led participants to have weaker expectations that the two described situations described must be strictly sequential. Instead, they may have preferred temporal sequences with overlapping situations – a sequence that we did not include in our visual stimuli to clearly distinguish between back- and forward-shifted interpretations. In that sense, the cross-linguistic differences we found might reflect an instance of how event conceptualization can be influenced by the way different languages slice up events in the world (cf. Gleitman & Papafragou, 2012; Slobin, 1996; Talmy, 1988).

2.3. General Discussion

In this paper, we have argued that, when people interpret temporal order in linguistic descriptions of complex events, they use as their basis the event-structural properties of each described situation: Non-dynamic states precede dynamic events.

This argument is based on a fundamental distinction between states and events both in event cognition and language: In event cognition, the perceptual units of our everyday experience are parsed based on salient, dynamic events (i.e., changes-of-state) which usually take place against a backdrop of less salient, static states. In language, stative and eventive predicates also differ, both in their grammatical behavior and how they are processed. However, the question whether the linguistic state-event distinction also determines temporal inferences in language comprehension, building on parallel differences between static and dynamic situations in perception, has so far not been investigated.

We used past-under-past relative clause constructions as an empirical test case. This allowed us to systematically manipulate properties of main and relative clauses. In addition, this approach opened new avenues to test our predictions against existing accounts of temporal interpretation in sentence comprehension. In line with our hypothesis, people prefer situations described by

stative descriptions to happen before situations described by eventive descriptions, regardless of their syntactic encoding or discourse-pragmatic functions. In the following sections, we discuss and review the results of our seven experiments, each conducted in English and German, in more detail.

All our experiments used short video clips followed by linguistic descriptions of the videos. Participants judged whether the situation seen matched the situation model construed based on the sentence they read. We consistently found that people derive temporal order depending on whether each situation is described as dynamic or static: In Experiment 1a, there were higher matching scores when the video first showed the stative situation described by a relative clause, followed by the eventive situation described by a main clause than when the eventive situation occurred before the state. This suggests that participants preferred linguistically encoded states to happen before linguistically encoded events. We replicated this preference twice, once in a follow-up experiment that controlled for potential confounds in the visual materials (Experiment 1b) and once in a follow-up experiment that further investigated how reference resolution might affect people's temporal construals (Experiment 1c).

Experiment 2a ruled out the possibility that temporal order preferences arise from the discourse-pragmatic function of relative clauses: Whereas, on such an account, participants should always rate higher those video sequences that order the situation described by the relative clause first (i.e., backshifted sequences), we found that this was not the case when the mapping between linguistic event type and syntactic encoding was reversed: Whenever a main clause introduced a stative description and the relative clauses introduced an eventive description, forward-shifted sequences in which the main clause description preceded the relative clause description were rated significantly higher than back-shifted sequences. Again, we replicated this preference for ordering stative before eventive descriptions in two follow-up experiments, showing that neither reference resolution (Experiment 2b), nor information structural factors (Experiment

2c) attenuated people's overall judgements. However, Experiment 2c showed that linear order can enhance the effects of linguistically encoded event structure: In cases where the order of situations matched between the videos and linguistic descriptions, people rated video sequences significantly higher than in mismatched cases, but only when the linear order was also consistent with the preference for ordering stative descriptions first in time.

In Experiment 3, we investigated which alternative interpretation strategy would come into play when linguistic event type does not serve as a cue to temporal order. The question was whether, in the absence of event-structural differences, relative clauses would be ordered first on the basis of their backgroundedness in the discourse, as discourse-pragmatic accounts claim. Our results show that, when event-structural properties were held constant across relative and main clauses, people had no clear preference for temporal order. While this pattern is at odds with existing analyses of the discursive function of relative clauses, it again lends support to event-structural accounts: Sentences containing two stative descriptions were generally rated lower than sentences containing two eventive descriptions. Thus, people were sensitive to event-structural properties, making use of their knowledge that states, unlike events, are not bound in time and tend to overlap.

In sum, across experiments, we found that video sequences were rated higher when a stative description preceded an eventive description, regardless of their syntactic encoding in the relative clause or the main clause. This effect follows directly from the first principles of event-structural approaches (Carroll & von Stutterheim, 2010; Klein, 1994, 2000, 2008a, 2009; von Stutterheim et al., 2003): In order to understand the temporal structure of a complex situation description, comprehenders must monitor a set of temporal parameters for each temporal expression that occurs. One of these parameters is a so-called temporal anchor that relates a described situation to another temporal interval. In principle, the temporal anchor can be the utterance time, but according to an event-structural perspective, previously mentioned or

contextually salient situations are more likely to be identified as anchors in more complex descriptions. This assumption is also in line with independent evidence, showing that people consistently link event descriptions through temporal, spatial and causal continuities (Kehler, 1994; Radvansky, Zwaan, Federico, & Franklin, 1998; Zwaan et al., 1995).

Importantly, the event-structural perspective cannot only explain why temporal order arises in past-under-past relative clause interpretations, but also how preferences for particular temporal sequences are formed: Temporal anchors are generally associated with the most salient time spans in context, so that people should use salience differences between described situations to construe their temporal models of complex situations. In everyday perception, events are more successful than states in attracting people's attention (cf. Clewett et al., 2020; Kurby & Zacks, 2008; Zacks et al., 2007) – a perceptual experience that we would expect to form part of people's general knowledge about situations. Thus, when hearing a linguistic description of a complex situation, people can use this knowledge to identify the more salient eventive description as the temporal anchor for the less salient stative descriptions. Put more simply, in language, events are temporal anchors for states.

In the present study, we tested event-structural predictions against syntax-driven and discourse-pragmatic accounts using stative descriptions of the agents' locations. However, the linguistic category of states is more diverse than what is covered in our stimuli: one can be hungry, happy, dead, or pregnant, or have a balcony, a concert ticket, a pimple, or long shiny hair. These mental, physical, or possessive stative descriptions, whether they are individual-level or stage-level predicates, may behave differently in temporal construal. Although this should be further explored, first empirical evidence points to the opposite. In a previous study (Marx & Wittenberg, 2024), we found that stative descriptions are reliably ordered before eventive descriptions, generalizing the findings presented here not only across heterogeneous stative predicates, but also across linguistic structures: When participants read consecutive sentences

that, like the relative clause constructions in the current study, left the order of described situations underspecified, they judged stative predicates to have happened first, regardless of linear order.

Furthermore, a model of linguistic event comprehension is flexible enough to allow for factors other than saliency to determine people's temporal interpretations. Consider, for example, a key motivation for syntax-driven analyses of relative clauses – a sentence such as in (21) – where the embedded stative description (i.e., being in my linguistics class) undoubtedly occurred after the main clause eventive description (i.e., raising).¹⁵

(21) *Mary raised a girl who was in my linguistics class.*

Here, identifying a temporal anchor could be either driven by causal knowledge (i.e., one can only take a university course if one has been raised) or by mere durational differences (i.e., raising someone usually takes longer than being in a linguistics course), rendering the eventive description in the main clause more stative and the stative description in the relative clause more eventive. In any case, an event-structural account provides a neat explanation for how, in the absence of explicit information about the temporal sequence within complex events, people do not use syntactic structure or discourse-pragmatic function to construct order in time, but instead draw on non-linguistic knowledge such as the internal structure of situations.

Our results are also not compatible with syntax-based accounts of temporal interpretation (cf. Abusch, 1997; Enç, 1986, 1987; Ogihara, 1995, 1996; Ogihara & Sharvit, 2012; Stowell, 2007; Von Stechow & Grønn, 2013a). According to these accounts, temporal relations are determined by underlying structural dependencies between subordinative and main clauses. Since relative clauses are syntactically independent of the main verb, the embedded tense should only be interpreted relative to the time of the utterance. In our design, each of the two temporal

¹⁵ The authors thank Roger Levy for this example.

sequences should therefore have been licensed. Clearly, this prediction is not supported by our data.¹⁶

One reason why syntax-driven accounts might not predict the temporal interpretation of past-under-past relative clauses correctly is that, for the most part, they are concerned with other types of dependent clauses (but cf. Arregui & Kusumoto, 1998; Ogiwara, 2015; Von Stechow & Grønn, 2013a, 2013b; Kubota et al., 2009). Relative clauses and other types of subordinations (e.g. temporal adjuncts), by contrast, have often been treated as independently interpreted counterexamples or only gained significance when embedded in complement clauses (e.g., Sue believed that she would marry a man who loved her, cf. Abusch, 1997, p. 17), in the scope of a future auxiliary (e.g., John will meet a man who lost his money, cf. Ogiwara, 1996, p. 161, but also Von Stechow & Grønn, 2013a; Anand & Hacquard, 2005), or in cross-linguistic analyses, tracing the ambiguities in complement clauses to underlying syntactic differences (e.g., Demirdache & Lungu, 2008; Ogiwara, 1996, 2015; von Stechow & Grønn, 2013a). In this regard, the syntactic literature still lacks a dedicated account that develops a principled model with its main focus on the temporal interpretation of past-under-past relative clauses.

2.4. Conclusion

In summary, the present set of studies empirically showed that temporal interpretation in linguistic descriptions of complex events is based on the event-structural properties of each described situation. In seven experiments in two languages, people rated video sequences lower when they violated the temporal pattern of ‘states before events’ - a result that would not be predicted by standard syntax-driven accounts of temporal interpretations in complex sentences,

¹⁶ Note that our data also refutes alternative discourse-pragmatic claims according to which relative clauses do not enter people's temporal situation models, since they only predicate or constrain discourse referents (cf. Hopper, 1979; Gunkel, 2007; Hofherr, 2014). Like the syntax-driven accounts, such analyses of relative clauses fail to explain people's preferences for particular temporal orderings, and more generally, how temporal order systematically maps onto linguistic form.

or by discourse-pragmatic accounts based on the function of relative clauses. However, our findings are consistent with a model of event comprehension that starts from the internal structure and salience of a described complex situation to explain how two different cognitive capacities - event cognition and language - must work together to form a mental representation of time.

2.5. Supplementary materials

Event-structural explanations rely heavily on independent evidence from event perception research, showing that stable states bind fewer attentional resources than cognitively more salient events (Clewett et al., 2020; Kurby & Zacks, 2008; Zacks et al., 2007). In an additional control experiment, we therefore examined whether such an asymmetry between situation types also holds true for our visual materials: We tested whether perceptual saliency differences precipitate in a language production task, predicting that, if states are perceptually less salient than events, people should use significantly fewer stative than eventive predicates in their spontaneous descriptions of the videos.

2.5.1. Participants

42 native English speakers and 42 native German speakers were recruited via online platforms (i.e., Testable Minds platform, clickworker.com).

2.5.2. Materials and procedure

The same video materials were used as in Experiments 1-3. Each participant performed two practice trials and 8 critical trials: Participants watched one of the animations and were asked to give a short, written description of the video they saw. In practice trials, participants were presented with a possible description after each video to (a) indicate the length of descriptions

expected at test and (b) make available both eventive and stative predicates. Practice trial sentences were bi-clausal *and*-constructions, containing once two stative predicates (i.e., *The cat sat in its basket and the mouse sat in front of the mouse hole*) and once two eventive predicates (i.e., *The girl waved and the cat jumped off the table*). The experiment took approximately 10 minutes, was programmed in PsychoPy and ran online on Pavlov.org. Participants were randomly assigned to one of two lists using Latin-square, to counterbalance between video sequences per item (i.e., back-shifted vs. forward-shifted sequences). Critical trials were presented in a randomized order.

2.5.3. Statistical analysis

We coded participants' responses using natural language processing methods in R (i.e., the spaCy package, Honnibal & Montani, 2017): Predicates were extracted from the verbal descriptions, lemmatized and manually coded according to their linguistic event type by the two authors (i.e., stative vs. eventive). Part-of-speech information ensured that only root verbs were included in the analysis.

We used a paired t-test to compare proportions of stative and eventive predicates produced in video descriptions. Inconclusive instances of eventive and stative predicates (e.g., *allow*, *fear*, 0.02% of all data points per language) were excluded from the statistical analysis. Before collapsing across participants, we furthermore converted each participant's raw counts of predicates into proportions of the overall produced predicates. Using proportions rather than actual quantities allowed us to level out individual differences in how exhaustive written descriptions were across participants. The analysis was preregistered on <https://osf.io/32njq>.

2.5.4. Results

In absolute frequencies, English native speakers used more eventive predicates ($M_{\text{EVENT}}=11.81$, $SD_{\text{EVENT}}=4.08$) than stative predicates when they spontaneously described the critical video clips ($M_{\text{STATE}}=1.67$, $SD_{\text{STATE}}=2.84$, see Fig. 2). This was also the case after transforming raw scores into proportions: On average, eventive predicates made up 88.4% of all predicates individual participants produced ($SD_{\text{EVENT}}=0.19$), whereas stative predicates amounted to 11.6% of predicates ($SD_{\text{STATE}}=0.19$), produced by only 26 out of 42 participants. Comparing proportions between eventive and stative predicates, this difference was statistically significant, $t(41)=12.92$, $p<0.0001$, CI [0.65, 0.89].

Similarly, in the German production task, eventive predicates were more frequent than stative predicates, both in terms of absolute frequencies ($M_{\text{EVENT}}=21.60$, $SD_{\text{EVENT}}=9.23$; $M_{\text{STATE}}=1.98$, $SD_{\text{STATE}}=2.74$) as well as in terms of proportions per participant ($M_{\text{EVENT}}=0.93$, $SD_{\text{EVENT}}=0.08$; $M_{\text{STATE}}=0.06$, $SD_{\text{STATE}}=0.08$). Crucially, 22 of the participants never produced stative predicates in their descriptions. As in the English data, the difference between eventive and stative predicate proportions was significant, $t(41)=36.52$, $p<0.0001$, CI [0.82, 0.92].

2.5.5. Discussion

In this experiment, we examined whether a widespread assumption in event perception research would precipitate in a language production task, according to which stative situations engage less attentional resources than events (Clewett et al., 2020; Kurby & Zacks, 2008; Zacks et al., 2007). When participants shortly summarized the video materials used in our study, we found that they used proportionally more eventive predicates than stative predicates, with some lacking stative predicates in their descriptions altogether. Overall, our results confirm that, during visual perception, events seem to be cognitively more salient than stable states – a circumstance that lends support to theoretical accounts that take such event-structural

differences as the starting point for how situations are ordered in time during language comprehension.

3. Inferring temporal order in English, German and Polish relative clause constructions

3.1. Introduction

When people hear a story, they build a mental representation of what happened – a so-called situation model (Barsalou, 1999; Glenberg, Meyer, & Lindem, 1987; Johnson-Laird, 1983; van Dijk & Kintsch, 1983; Zwaan, 1996; Zwaan et al., 1995). One important dimension of such models is when something happened (Radvansky et al., 1998; Rinck et al., 2001; Speer & Zacks, 2005; Theriault & Raney, 2007; Zwaan, 1996): For instance, *The vet sat in the truck* reports a situation in the past, while *The vet will hop onto the truck* tells about a situation in the future. These statements also differ in terms of other temporal properties: Sitting usually takes longer than hopping, and it describes a static, temporally homogeneous situation, rather than a dynamic, punctual change-of-state.

More complex situations, however, consist of multiple situations with different temporal properties; and while understanding how these situations are temporally related is crucial (Briner, Virtue, & Kurby, 2011; Dery & Bittner, 2016; Gouldthorp, Katsipis, & Mueller, 2018; Mandler, 1986), it is not always straightforward: A description such as *The vet sat in the truck onto which the bison hopped* could either mean that the hopping event happened before the vet sat on the truck, or that the vet sat on the truck first and then the bison hopped on it.

In this article, we investigate how people understand the temporal order of complex situations like the one just described. We focus on dynamicity of situations, like the difference between hopping and sitting, as one crucial but under-researched component in how people construe complex temporal relations from linguistic descriptions. In the discussion on eventuality distinctions, dynamicity is the third component in conjunction with telicity/boundedness and duration/durativity, which we do not consider in the present paper. The notion of dynamicity

relates to the presence or absence of change within situations in the world, and is the basis for the linguistic distinction between events and states (Dowty, 1986; Kearns, 2000; Olsen, 1994).

As such, it is not purely a linguistic concept, but rather one arising out of event perception.

Based on previous work, both from linguistics and event cognition research, we expect that dynamicity differences strongly guide the construal of situation models (Dowty, 1977; Gennari & Poeppel, 2003; Katz, 2003; Michaelis, 2011; Parsons, 2002; Pettijohn & Radvansky, 2016; Radvansky & Copeland, 2010; Rinck & Weber, 2003; Zwaan et al., 1995; Zwaan, Radvansky, Hilliard, & Curiel, 1998), and ask: When linguistic descriptions lack explicit information about the temporal relations between individual situations, how do people determine which is ordered before the other?

In three experiments, we obtain a reconstruction of people's situation models, using an act-out task in which participants recreate complex sentences with toys. Our results show that event dynamicity, i.e., whether a situation is described as a static state or as a dynamic event, is a reliable cue to temporal order: Across three languages (English, German, and Polish), we find that stative descriptions are consistently ordered before eventive descriptions. This holds true despite potential differences in temporal interpretation and linguistic encoding of dynamicity between languages.

3.1.1. Encoding temporal relations in language

In language, there is “a wide variety of strategies for describing the structure of time” (Radvansky & Zacks, 2014, p. 72): Temporal information can either be encoded grammatically through morphological tense or aspect (e.g., *The vet sat/will sit/is sitting on the truck*), or lexically via time adverbs (e.g., *yesterday, today, tomorrow*) or temporal connectives (e.g., *before, after, while*). In the case of complex situation descriptions, the temporal order between individual events can therefore be part of the linguistic meaning itself (1), when a connective

or a relational adverb directly signals how two events are temporally linked, or the temporal relation must be inferred from the absolute temporal status of each event (2) (Evers-Vermeul, Hoek, & Scholman, 2017; Kehler, 1994).

- (1) a. *The vet sat in the truck **before** the bison hopped on it.*
- b. *The vet sat in the truck onto which the bison hopped **earlier**.*
- (2) a. *The vet sat in the truck **today** onto which the bison hopped **yesterday**.*
- b. *The vet **sat** in the truck onto which the bison **had hopped**.*

However, the order in which events within complex situations occur need not be specified: Extreme examples are tenseless languages (e.g., Mandarin, Dyirbal, Yucatec, Kalaallisut, cf. Bittner, 2005; Bohnemeyer, 2009; Comrie, 1985; Tonhauser, 2015), where temporal status can be marked lexically, but this marking is not mandatory (Lin, 2003; cf. Mucha, 2012; Smith, 2008). Even in languages with obligatory tense marking, a linguistic description may not provide sufficient information to establish the order between events: This occurs, for example, in discourse such as (3), in which two past tense sentences follow each other (Hinrichs, 1986; Hopper, 1979; Horn, 2022b; Marx & Wittenberg, 2024; Schmerling, 1975). Temporal relations are also underspecified in past-under-past relative clauses such as (4), where a main clause and a relative clause each describe a situation in past tense, both involving the relativized object (i.e., *truck*).

- (3) *The vet sat in the truck. The bison hopped onto it.*
- (4) *The vet sat in the truck onto which the bison hopped.*

In such cases, it is up to the comprehender to infer the temporal order between the two described subevents. The crucial question, then, is: What guides these temporal order inferences? Here, we focus on relative clause constructions, whose possible temporal interpretations we further explore in the following.

3.1.2. Temporal interpretations in relative clauses

Relative clauses are an excellent test case for testing temporal order inferences because, like other subordinate clauses, they link two propositions together in one utterance. However, unlike other subordinate clauses, relative clauses do not systematically trigger causal, temporal, or conditional inferences between the main clause and the relative clause situation.¹⁷ Formal linguistic accounts have explained this lack of systematic inferences with the structural status of relative clauses (e.g., Abusch, 1997; Diessel, 2005; Enç, 1987; Kusumoto, 1999; Ogihara, 1995; Stowell, 2007; Von Stechow, 1995; von Stechow & Grønn, 2013b, 2013a): Since relative clauses are adjuncts and thus optional constituents in the syntactic structure, comprehenders should in principle be able to interpret the tensed predicate in the relative clause temporally independently of the tensed predicate in the main clause. For (4), as long as both the sitting and the hopping took place in the past, any logical possibility of temporal order should be a computable temporal interpretation: The bison could have hopped onto the truck either before, simultaneously with, or after the vet sat on it.

At the same time, further formal analyses of relative clauses have pointed out that speakers and hearers adhere to the discourse pragmatic function of the linguistic structure (Cinque, 2020; Fox & Thompson, 1990; Lambrecht, 1988; Levinson, 1983; Quirk et al., 1985; Stockwell et al., 1973; Thomson & Martinet, 1985): Given that relative clauses either restrict the space of potential referents (i.e., restrictive relative clauses) or simply predicate the antecedent head noun (i.e., non-restrictive relative clauses), they often only provide background information against which a main clause event unfolds; and although backgroundedness is not necessarily temporal, there seems to be a connection between pragmatic background and temporality (Hopper, 1979; Wårvik, 2004). Consequently, under such a view, comprehenders should tend

¹⁷ Relative clauses may imply causal, or at least, coherence relations in discourse processing (Hoek, Rohde, Evers-Vermeul, & Sanders, 2021; Rohde et al., 2011) but previous studies used implicit causality predicates which have been shown to elicit causal biases independently in other linguistic constructions (Hartshorne & Snedeker, 2013; Solstadder & Bott, 2022).

to locate a relative clause description before a main clause description: In (4), the bison should have hopped onto the truck before the vet sat on it.

In general, however, temporal construal is also affected by other factors such as iconicity, whereby the order of mention maps iconically onto the order of events in the world. This is often the case in linguistic discourse (Fleischman, 1990; Hopper, 1979; Horn, 2022b; Kaiser, 2019; Kamp & Rohrer, 1983; Lascarides & Asher, 1991, 1993; Ohtsuka & Brewer, 1992; Schmerling, 1975; Tai, 1983; Xu & Kwok, 2019), but also in complex sentences: Non-linear temporal adjunct clauses pose a challenge of comprehension for both children (Blything & Cain, 2016; Blything, Davies, & Cain, 2015; Clark, 1971; de Ruiter, Theakston, Brandt, & Lieven, 2018; Trosborg, 1982), and patients with language impairments (Abrahamsen & Gilbert, 1990; Natsopoulos et al., 1991; Sasanuma & Kamio, 1976). They also cause processing costs for adults (Clark & Clark, 1968; Münte, Schiltz, & Kutas, 1998; Ohtsuka & Brewer, 1992; Van Der Meer, Beyer, Heinze, & Badel, 2002; Xu & Kwok, 2019, but see Jou & Harris, 1990). If iconicity restricts the temporal interpretation in sentences like (3) or (4), then comprehenders should arrange the reported situations according to the order of mention: The bison should have hopped onto the truck only after the vet had sat on it.

In sum, the temporal relation between main and relative clause is subject to several concurring, and perhaps competing, linguistic factors. In addition, we argue that one important factor in construing temporal order is the dynamicity of the described situations.

3.1.3. Dynamicity and temporal construal: Previous research

In building situation models, language comprehension and event cognition interact (Berman & Slobin, 1994; Bunger, Papafragou, & Trueswell, 2013; Bunger, Skordos, Trueswell, & Papafragou, 2016; Carroll & von Stutterheim, 2010; Coll-Florit & Gennari, 2011; Gerwien & von Stutterheim, 2018; Gleitman & Papafragou, 2012; Madden & Zwaan, 2003; Magliano &

Schleich, 2000; Misersky, Slivac, Hagoort, & Flecken, 2021; Radvansky et al., 1998; Sakarias & Flecken, 2019; Santin, van Hout, & Flecken, 2021; Strickland et al., 2015; Ünal, Ji, & Papafragou, 2021; Wittenberg & Levy, 2017; Zwaan & Madden, 2005; Zwaan et al., 1995). When it comes to the temporal dimension of situations, the majority of studies has focused on how people derive the temporal properties of simple situations from factors such as boundedness or duration (Barner et al., 2008; Becker et al., 2013; Bott, 2010; Bott & Gattnar, 2015; Coll-Florit & Gennari, 2011; Flecken & Gerwien, 2013; Foppolo et al., 2021; Magliano & Schleich, 2000; Milczarski, 2021; Minor et al., 2022; Paczynski et al., 2014; Piñango et al., 2006; Radvansky et al., 1998; Sakarias & Flecken, 2019; Strickland et al., 2015; Wittenberg & Levy, 2017; Zwaan & Madden, 2005). However, as outlined above, relating two or more situations temporally is crucial for constructing accurate situation models of complex situations. In previous work (Marx & Wittenberg, 2022, 2025), we therefore studied people's interpretations of past-under-past relative clauses like (4), to test predictions from accounts that closely link temporal construal to dynamicity (Carroll & von Stutterheim, 2010; Klein, 1994, 2009; von Stutterheim et al., 2003): According to such theories, each tensed verb encodes a specific time, a topic time, at which a situation takes place and which is relevant, according to the speaker. Topic times are then linked to a reference point, a so-called temporal anchor, and the relationship between the two is determined by tense meaning: For simple past, the anchored topic time should begin before the temporal anchor, but never after it.

In the simplest cases of situation descriptions (e.g., *The vet sat on the truck*), temporal anchors relate to the utterance time. In more complex situation descriptions, however, they are associated with the most salient element mentioned in the description; therefore, they may be chosen on the basis of what comprehenders know about the structure of the encoded events. But how is saliency determined? One suggestion comes from perception and memory research, which has demonstrated that dynamic changes-of-state (i.e., events) are cognitively more

salient than stable states (Clewett et al., 2020; Kurby & Zacks, 2008; Zacks et al., 2007), with a corresponding distinction in the linguistic domain (Dowty, 1977; Gennari & Poeppel, 2003; Katz, 2003; Levin & Hovav, 1994; Michaelis, 2011; Olsen, 1994; Parsons, 2002; Pettijohn & Radvansky, 2016; Radvansky & Copeland, 2010; Rinck & Weber, 2003; Zwaan et al., 1995, 1998): There is a long tradition in linguistics of distinguishing between different verb and predicate types on the basis of their syntactic and semantic behavior, and one central – while sometimes overlooked – feature, along with telicity and durativity, is the presence or absence of dynamic change (see Table 1, following Olsen, 1994). In language, events – whether or not they have duration or an inherent endpoint – are states with added dynamicity.

Table 1 Linguistic features to distinguish between different types of verbs or predicates, following Olsen (1994)

	Dynamicity	Telicity	Durativity
Activity	+	-	+
Achievement	+	+	-
Accomplishment	+	+	+
State	-	-	+

Event-structural accounts of language comprehension take this distinction between dynamic and static situations in language and cognition as a starting point to predict that eventive descriptions, corresponding to more salient changes in the world, should serve as temporal anchors, and that stative descriptions, corresponding to less salient states, should serve as anchored situations.

Our previous studies confirmed the event-structural predictions in English and German tasks (Marx & Wittenberg, 2022, 2025, see also Marx & Wittenberg, 2024, as an extension to discourse comprehension): For past-under-past relative clauses, stative descriptions (i.e., *sit*) were generally preferred to happen before eventive descriptions (i.e., *hop*), regardless of

whether they were encoded in the main or the relative clause and regardless of whether the eventive predicate encoded an inherent endpoint or was unbound.¹⁸ That is, whether situations were linguistically framed as a state or as an event systematically predicted people's temporal interpretations of sentences. Durativity, as the third dimension of linguistic event structure, only affected interpretations in the absence of dynamic change (Marx & Wittenberg, 2024): longer states were ordered before shorter states.

Crucially, previous studies were mainly concerned with collecting baseline data to investigate the role of non-linguistic event cognition in linguistic temporal construal: In Marx and Wittenberg (2022, 2025), participants watched one of two video clips, in which the temporal order between two situations was manipulated, and rated on a 1-5 Likert scale whether a relative clause construction correctly described the visual scene. However, temporal construal in language comprehension may not be adequately captured by explicit acceptability judgments (for a general criticism of acceptability judgments, see Kepser & Reis, 2005; Schütze & Sprouse, 2013): One concern is that participants were presented with possible temporal interpretations in the videos, without the need to actively construe a mental representation of the described scene. Furthermore, the videos only showed two clearly sequential situations, without any potential overlap. In a strict sense, our previous results may therefore only speak to how people match linguistic descriptions to visual temporal sequences in the world, not to how they model the temporal aspects of a complex situation from language.

A second and related concern is the interpretation of gradient levels of acceptability (Schütze & Sprouse, 2013, cf. Bresnan, 2008; Keller, 2000; Barsalou, 1987): In other words, what does it mean to rate a sentence with “3”, in between “1” (unacceptable) and “5” (acceptable)? In our case, the gradient results may have reflected the strength of participants' temporal inferences, based on their event knowledge, but they may also have resulted from alternative factors:

¹⁸ The distinction between telic and atelic predicates was part of an exploratory analysis as part of the pre-registration at <https://osf.io/6ae5m/>.

Armstrong, Gleitman, and Gleitman (1983), for example, found that even clearly categorical concepts such as “even number” elicited gradient judgments to different category exemplars, suggesting that participants reinterpreted the task as judging how representative or typical particular numbers are for a concept. In our previous studies, acceptability judgments may have also been driven by the frequency or prototypicality of particular temporal interpretations, or by the context of the visual materials we used, reflecting participants’ metalinguistic knowledge or post-hoc explanations due the task design.

Finally, in the previous studies, we focused on English and German - two languages that are somewhat different in terms of temporal interpretations in other sentence constructions, but similar in the way dynamicity is linguistically encoded: whether a described situation is stative or dynamic is part of the lexical meaning of the predicate. It is therefore an open question to what extent grammatical marking of dynamicity results in a similar pattern of temporal construal.

3.1.4. The current studies

To address these questions, and to test the effects of dynamicity on temporal construal more directly, the current study uses an act-out task, adopted from studies like Thothathiri & Snedeker (2008) or Wiese et al. (2020), to tap into people’s mental construal of temporal order in language comprehension. The rationale is that temporal construal should be directly reflected in the way people act out a complex situation description: That is, those parts of the situation that are understood to have happened first in time should also be enacted first in the task. By using an act-out task, we thus did not ask participants to explicitly reason about temporal relations, while still being able to obtain the temporal dimension of their situation model. It should be noted that act-out tasks have been used in developmental studies to investigate temporal

comprehension of complex sentences (Amidon & Carey, 1972; Bauer & Mandler, 1992; Clark, 1971; Gorrell, Crain, & Fodor, 1989; Winskel, 2004).

As laid out above, formal accounts would predict that comprehenders are able to interpret the embedded description temporally independent of the main clause description, so no particular temporal order should stand out in any of the sentences in (5): Participants should perform at chance level (i.e., enact an equal number of relative clause and main clause descriptions first), assuming that they do not take other factors into account.

- (5)
- a. *The vet **sat**_{STATIVE} in the truck onto which the bison **hopped**_{EVENTIVE}.*
 - b. *The vet **climbed**_{EVENTIVE} into the truck on which the bison **stood**_{STATIVE}.*
 - c. *The vet **climbed**_{EVENTIVE} into the truck onto which the bison **hopped**_{EVENTIVE}.*
 - d. *The vet **sat**_{STATIVE} in the truck on which the bison **stood**_{STATIVE}.*

If, however, people's temporal construal of past-under-past relative clauses is systematically determined by the function of relative clauses in discourse, this should lead to relative clauses being enacted first. Another factor could be iconicity, leading to enactments that primarily follow the linear order of the sentence. Crucially for both pragmatic strategies, we would expect no differences with respect to the different dynamicity manipulations in (5): People should follow the same strategy independently of situation dynamics.

Alternatively, event dynamicity may determine the way in which event order is derived from temporally underspecified descriptions, as suggested by previous data on acceptability judgments. In this case, participants should identify more salient eventive descriptions as the temporal anchor of less salient stative descriptions in past tense sentences and construe event sequences accordingly: Stative descriptions should be enacted first, before eventive descriptions, regardless of whether they are encoded by the relative clause (5a) or the main clause (5b). In cases where dynamicity differences do not serve as a cue to temporal order (5c-

d), we would expect participants either to perform at chance level or to use one of the pragmatic strategies mentioned above (e.g., relative clauses first, linear order).

We conducted three pre-registered experiments to test these predictions, one in English (Experiment 1), one in German (Experiment 2), and one in Polish (Experiment 3): Participants heard past-under-past relative clause descriptions as in (5), which were crossed in dynamicity (stative vs. eventive description) of the main and the relative clauses. After each description, participants enacted the complex situation with toys.

The three languages were chosen to investigate the role of dynamicity in temporal construal across languages that are (a) different with respect to temporal interpretations and (b) different in the way in which dynamicity is encoded (see Table 2 for a summary).

Temporal interpretations differ in English, German, and Polish with regard to the so-called Sequence of Tense (SoT) phenomenon, which emerges for temporal interpretations in embedded contexts: In complement clause constructions (e.g., *The vet said that the bison stood on the truck*), an embedded past tense can have non-past meaning, in addition to signaling anteriority, such that it may be interpreted simultaneously with the time denoted by the matrix clause. English is a clear case of a SoT language (Abusch, 1997; Declerck, 1988; Ogihara, 1995; Stowell, 2007) and German has likewise been classified as such (Boskovic, 2012) but, to the best of our knowledge, a systematic formal and empirical analysis is lacking (see Giorgi & Pianesi, 2000; von Stechow, 2005 for alternative assessments). In contrast to English, SoT ambiguities in German complement clauses can be resolved by optional subjunctive marking (Eckardt, 2021; Helbig & Buscha, 2001). Polish, on the other hand, is a non-SoT language and expresses simultaneity between an embedded tense and the main clause only by embedding present tense (Arregui & Kusumoto, 1998; Bittner, 2014; Sharvit, 2014; but see Mucha, Renans, & Romoli, 2022). It should be noted that such SoT ambiguities only arise with embedded stative

predicates, so that speakers of SoT languages may be more aware of dynamicity as a cue to temporal interpretation.

In terms of dynamicity, both English and German encode differences between eventive and stative situations as part of the lexical meaning of the predicate, but English can also mark the ongoingness of an event grammatically with the progressive aspect (Parsons, 1989). Polish, on the other hand, has rich tense and aspect morphology, but only optional lexical distinction between states and events (cf. Kozłowska-Raś, 1987 for marking within aspectual classes); information about dynamicity is mainly derived from obligatory grammatical aspect marking (Czochralski, 1972; Comrie, 1976; Vater, 1995; Młynarczyk, 2004; Bogdan & Sullivan, 2009; Klimek-Jankowska & Błaszczak, 2020).¹⁹

Table 2 Overview of the languages tested in each experiment, including their language-specific properties of interest and the example sentences for each condition (1 – eventive/stative, 2 – stative/eventive, 3 – eventive/eventive, 4 – stative/stative).

Language	SoT	Dynamicity	Example sentences
English 	yes	lexical	1. The vet climbed into the truck on which the bison stood . 2. The vet sat in the truck onto which the bison hopped . 3. The vet climbed into the truck onto which the bison hopped . 4. The vet sat in the truck on which the bison stood .
German 	?	lexical	1. Das Vögelchen setzte sich auf den Drachen, auf dem der Junge saß . 2. Das Vögelchen saß auf dem Drachen, auf den sich der Junge setzte . 3. Das Vögelchen setzte sich auf den Drachen, auf den sich der Junge setzte . 4. Das Vögelchen saß auf dem Drachen, auf dem der Junge saß .

¹⁹ As Table 2 shows, we have not included any SoT languages in which dynamicity is grammatically encoded because, to the best of our knowledge, there is no language that has SoT properties but no inherent lexical aspect. For example, in Romance languages such as French, a SoT language, the indication of dynamicity is linked both to the lexical meaning of predicate and to the tense-aspect system (imparfait vs. passé simple/composé). Here we focus on the basic question of what role dynamicity plays in temporal construal, leaving for future research the question of how lexical encoding of dynamicity interacts with tense and aspect marking in triggering temporal inferences.

Polish 	no	grammatical	1. <i>Pływaczka usiadła obok skrzyni, na której siedziała kura.</i> 2. <i>Pływaczka siedziała obok skrzyni, na którą usiadła kura.</i> 3. <i>Pływaczka usiadła obok skrzyni, na którą usiadła kura.</i> 4. <i>Pływaczka siedziała obok skrzyni, na której siedziała kura.</i>
---	----	-------------	--

In sum, these language differences allow us to test two potential generalizations: The first is whether the effects of dynamicity extend beyond English and German, which are frequently ambiguous in terms of temporal construal of embedded stative descriptions (generalization 1). The second is whether the linguistic level at which dynamicity differences are marked mediates the effects of dynamicity seen in previous studies (generalization 2).

3.2. Methods

3.2.1. General materials and procedure

Participants were tested in a within-subjects design: They were seated at a table with a set of four toys (mostly Playmobil©) placed in a basket in front of them (see Figure 1). Three of the toys served as critical items (e.g., vet, bison, truck) and a fourth served as a distractor item (e.g., plant). At least two of the critical toys were animate (e.g., a human agent or an animal), the third was, in some cases, and the fourth in all cases, an object artifact (e.g., plant, truck).

On each trial, participants listened to a sentence (e.g., *The vet sat in the truck onto which the bison hopped*). After the sentence ended, they squeezed a rubber ducky with their non-dominant hand, to emit a squeak, and then they enacted the complex situation as described. Each enactment required manipulating the two critical animate items (e.g., vet, bison) relative to the third location item (e.g., truck). The rubber ducky served to prevent any enactment before sentence end. In addition, it forced participants to enact the sentences with one hand only. This, in turn, made the temporal dimension of their mental construals more prominent, since it forced

a sequential act out of either the relative clause or the main clause. We reasoned that, if the temporal order was either irrelevant or flexible according to participants' mental models, we should find chance-level performance across trials.

Sentences were pre-recorded using the Google Cloud Text-to-Speech API to keep intonation as neutral as possible across conditions. The linguistic materials for each experiment are described individually in the following subsections.

We used nine different sets of toys corresponding to the experimental blocks: To familiarize participants with the toy objects, they matched the toys with their corresponding name cards at the beginning of each block (see Figure 1). First, a practice block with three practice trials illustrated the task and the experimental procedure. The practice block was followed by eight experimental blocks, each consisting of 12 trials: one critical, one attention check (i.e., describing a temporally unambiguous sequence of events) and two filler trials, as well as eight trials for an unrelated study (published in Marx, Iwan, & Wittenberg, 2023).

The experimental blocks, as well as the trials within each block, were presented in a pseudorandomized order. For counterbalancing, participants were randomly assigned to one of



Figure 1 Example of a basket (upper right) and a corresponding object set.

four lists using a Latin Square design. All experiments were programmed and live-coded in PsychoPy; sessions were recorded on video.

3.2.2. General analysis

We performed two types of preregistered statistical analysis (cf. analysis plan at <https://osf.io/qgnfr/>): First, we fitted a linear mixed-effects logistic regression model (i.e., *glmer* from the lme4 package, Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2014). Model significance was assessed with likelihood ratio tests, comparing the full model to reduced models that individually excluded fixed predictors and interactions, but included the same random effects structure (Barr et al., 2013). In both experiments, the two fixed contrast-coded predictor variables were dynamicity in the main clause (stative vs. eventive) and the combination of dynamicity in the sentence (mixed: stative-eventive/eventive-stative vs. same: eventive-eventive/stative-stative). Participants and items were included as random intercepts to control for variance possibly induced by these two factors. We used random intercepts, but no random slopes, to avoid overfitting of the regression model (i.e., minimal intercept-only structure). Our dependent measure was which of the two clauses (main vs. relative clause) was

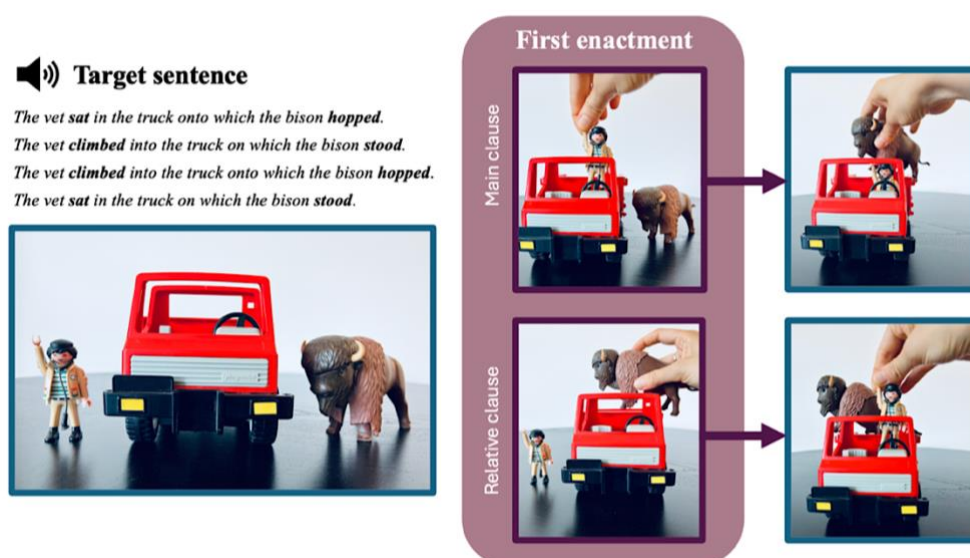


Figure 2 Staged example trial of a target sentence, either enacting the main clause first (up) or the relative clause first (down).

enacted first (see Figure 2). Second, a series of planned pairwise comparisons was conducted to assess the interaction of the regression model.

3.2.3. Experiment 1: English

3.2.3.1. Participants

49 native English speakers were recruited via the Central European University recruitment portal. We included participants in the data analysis only if all trials were completed and the criterion of at least 5 out of 8 attention control trials was met. Based on this criterion, one participant was excluded from the study, resulting in 48 complete data sets as preregistered.²⁰ As we only included correctly and unambiguously performed trials in the statistical analysis (e.g., involving only toys mentioned), 8.3% of all critical trials were excluded.

3.2.3.2. Linguistic materials

For each toy set, we constructed four critical sentences, all of which were active, intransitive past-under-past relative clause constructions which described two agents relative to a location. Descriptions were either eventive (e.g., *sit down, hop, climb*) or stative (e.g., *sit, was, stand*) and were crossed between main and relative clauses.²¹ In two of the critical sentences, dynamicity therefore differed between main and relative clause (eventive-stative 6a, stative-eventive 6b) and, in the other two, it was uniform between clauses (eventive-eventive 6c, stative-stative 6d). All linguistic material consistently used the simple past.

²⁰ We preregistered between 36 and 48 participants per experiment, based on previous act-out studies (Thothathiri & Snedeker, 2008; Wiese et al., 2020) with a maximum sample size of 35. Based on the effect size of the interaction between event type combination and event type in the main clause, the observed power was at 100.0% for all three experiments (CIs: [92.89, 100.0]; effect sizes via likelihood ratio > 1.0).

²¹ Since previous results have found no difference between bounded/unbounded eventive descriptions (Marx & Wittenberg, 2025) and an effect of duration only in the absence of dynamic differences (Marx & Wittenberg, 2024), we have left aside questions about the interaction with other event-structural dimensions such as telicity and durativity and focused only on dynamicity in linguistic encoding.

- (6) a. [*The vet **climbed**_{EVENTIVE} into the truck*]_{MC} [*on which the bison **stood**_{STATIVE}*]_{RC}
 b. [*The vet **sat**_{STATIVE} in the truck*]_{MC} [*onto which the bison **hopped**_{EVENTIVE}*]_{RC}
 c. [*The vet **climbed**_{EVENTIVE} into the truck*]_{MC} [*onto which the bison **hopped**_{EVENTIVE}*]_{RC}
 d. [*The vet **sat**_{STATIVE} in the truck*]_{MC} [*on which the bison **stood**_{STATIVE}*]_{RC}

As attention control, we included relative clause constructions that contained temporal adverbs to disambiguate between different temporal orders (e.g., *previously*, *first/then*), as well as *before-/after-*clause constructions, which forced temporal sequence due to the temporal connective (4 trials per construction type). All remaining trials consisted of simple sentences. Linguistic materials can be found at <https://osf.io/qgnfr/>.

3.2.3.3. Results

Participants enacted 93.2% of the attention control trials correctly, demonstrating both that the task was easy to perform, and interest did not decline over the course of the trials.

Participants' first enactments are shown in Figure 3: Overall, when sentences encoded both an eventive and a stative situation (pink plots), participants were more likely to enact stative situations first, regardless of whether it was in the relative clause or main clause. When clauses encoded situations of the same dynamicity (green plots), main clauses tended to be enacted first for sentences that described two eventive situations, but performance was at chance for sentences that described two stative situations.

For statistical analysis, we coded relative clause first enactments as 1, and main clause first enactments as 0. We found a main effect of dynamicity in the main clause ($Df=1$, $\chi^2=55.74$, $p<0.001$): When main clauses contained a stative description, they were more likely to be enacted first ($\text{mean}_{\text{MCstative}}=0.35$, $\text{SD}=0.48$) compared to when they contained an eventive description ($\text{mean}_{\text{MCeventive}}=0.65$, $\text{SD}=0.48$).

Furthermore, there was a main effect of dynamicity combination ($Df=1$, $\chi^2=9.50$, $p=0.002$):

Participants were more likely to enact main clauses first when clauses encoded the same dynamicity ($\text{mean}_{\text{same}}=0.45$, $SD=0.50$) than when they encoded both stative and eventive situations ($\text{mean}_{\text{mixed}}=0.55$, $SD=0.50$).

Crucially, we found a significant interaction ($Df=1$, $\chi^2=124.84$, $p<0.001$), which according to pairwise comparisons, was driven by differences both between the two mixed dynamicity conditions ($\beta=0.40$, $t=17.92$, $p<0.001$) and between the two conditions that encoded the same dynamicity ($\beta=-0.09$, $t=-2.54$, $p=0.01$): When dynamicity differed between main and relative clause, participants were more likely to enact the relative clause first when the main clause was eventive ($\text{mean}_{\text{mixed_MCeventive}}=0.95$, $SD=0.21$, first line in Figure 3), but they preferred to enact the main clause first when it was stative ($\text{mean}_{\text{mixed_MCstative}}=0.15$, $SD=0.36$, second line in Figure 3). In cases where dynamicity matched between clauses, participants were more likely to enact the main clause first when the sentence reported two eventive situations

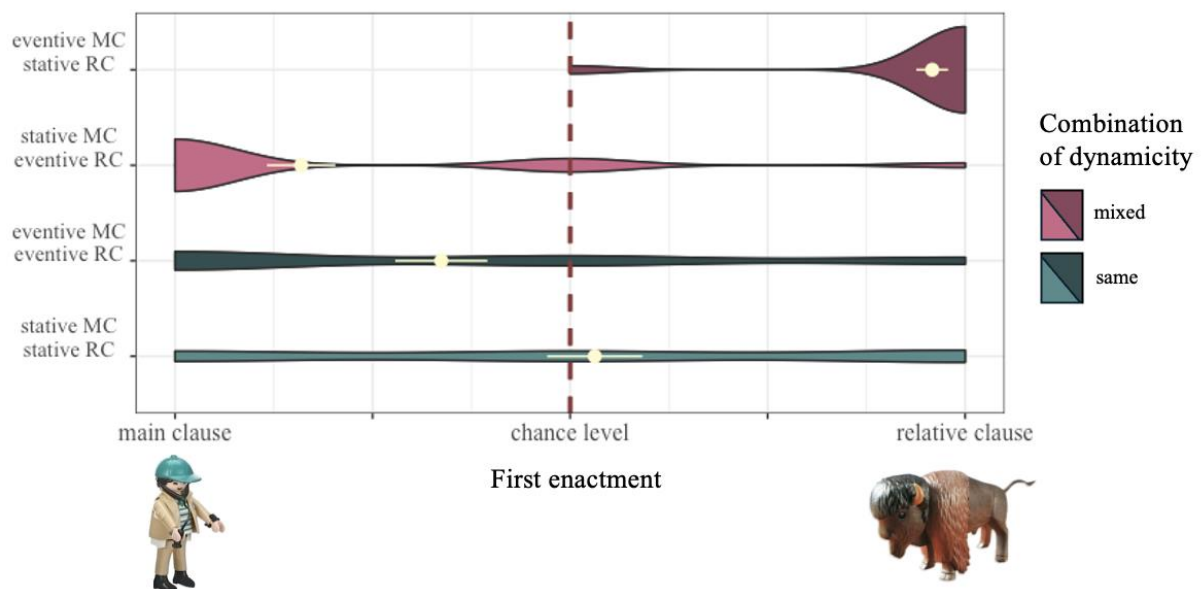


Figure 3 Mean first enactments of main or relative clause in Experiment 1 with native English participants. Dashed line depicts chance level performance. Figure illustrates the examples in (6), in which the vet appeared as subject of the main clause; first enactments of the vet figure therefore correspond to first enactment of the main clause. The bison, in contrast, appeared in the relative clause; more first enactment of the bison therefore corresponds to more first enactment of the relative clause. The width of the violin plot corresponds to means by participants; dots represent overall means, and error bars the standard error of the mean.

($\text{mean}_{\text{same_MCeventive}}=0.35$, $\text{SD}=0.48$, third line in Figure 3) than when it described two stative situations ($\text{mean}_{\text{same_MCstative}}=0.54$, $\text{SD}=0.50$, last line in Figure 3).

3.2.3.4. Discussion Experiment 1

In Experiment 1, we asked how people model the temporal relations of past-under-past relative clauses and whether dynamicity of the described situations would be used to guide inferences. When systematically manipulating dynamicity in both the main and relative clause, we found that people have a clear tendency to place the stative description first in time - at least when the other described situation is an event.

However, differences in dynamicity were also reflected between conditions in which dynamicity was the same across clauses: Participants had a weak preference to follow the iconic order of mention, i.e., the temporal order in which they learned about the situations in the sentence, when both described situations were events. For stative-stative descriptions, this preference disappeared completely, as reflected in participants' chance-level performance. This pattern of results is in line with the observation in linguistics, that events, not states, move a narrative in time (Dry, 1982; Hopper, 1979; Partee, 1984).

3.2.4. Experiment 2: German

In Experiment 2, we tested the effects of dynamicity encoding on people's temporal construal in German – a language that, like English, is typologically classified as a SoT language (but see Giorgi & Pianesi, 2000; von Stechow, 2005), but with less robust temporal ambiguity for embedded stative predicates: additional grammatical marking of mood can remedy SoT ambiguities (Eckardt, 2021; Helbig & Buscha, 2001).

Importantly for our purposes, however, German shares one important feature with English: dynamicity is part of the lexical predicate meaning. We reasoned that if lexically encoded

dynamicity is a reliable cue to temporal order across languages, not only a resolution strategy unique to English, we should obtain a similar pattern of results as in Experiment 1 with German speaking participants.

3.2.4.1. Participants

We recruited 48 native German speakers via the Central European University recruitment portal. As in Experiment 1, data sets were only submitted to data analysis if they were complete and met the criterion of performing 5 out of 8 attention control trials correctly. No participant was excluded from the analysis based on this criterion, but we excluded 3.4% of the critical trials due to incorrect performance (e.g., enactments with toys that were not mentioned).

3.2.4.2. Linguistic materials

As in Experiment 1, we used active, intransitive past-under-past relative clause constructions, describing either the movement or the posture of one of the toy agents towards or at a location. Main and relative clauses were systematically manipulated with respect to their encoded dynamicity (eventive vs. stative), resulting in two critical sentences that encoded different dynamicity types (7a, b) and two critical sentences that encoded uniform dynamicity across clauses (7c, d).

- (7) a. [*Das Vögelchen **setzte**_{EVENTIVE} sich auf den Drachen*]_{MC}
 The bird sat down REFL on the dragon
 [*auf dem der Junge **saß***_{STATIVE}]_{RC}
 on which the boy sat
- b. [*Das Vögelchen **saß***_{STATIVE} *auf dem Drachen*]_{MC}
 The bird sat on the dragon

- [*auf* *den* *sich* *der* *Junge* ***setzte***_{EVENTIVE}]_{RC}
 on which _{REFL} the boy sat
- c. [*Das* *Vögelchen* ***setzte***_{EVENTIVE} *sich* *auf* *den* *Drachen*]_{MC}
 The bird sat down _{REFL} on the dragon
- [*auf* *den* *sich* *der* *Junge* ***setzte***_{EVENTIVE}]_{RC}
 on which _{REFL} the boy sat
- d. [*Das* *Vögelchen* ***saß***_{STATIVE} *auf* *dem* *Drachen*]_{MC}
 The bird sat on the dragon
- [*auf* *dem* *der* *Junge* ***saß***_{STATIVE}]_{RC}
 on which the boy sat

We used the German preterite as the historical and structural equivalent tense of English simple past (Helbig & Buscha, 2001; Klein, 2000, Comrie, 1995). Furthermore, to keep semantic contrasts to a minimum, we used a German-specific predicative contrast: stative posture predicates (i.e., *stehen* ‘stand’, *sitzen* ‘sit’, *liegen* ‘lie’) were contrasted with their transitive counterparts (i.e., *stellen* ‘put’, *setzen* ‘place’, *legen* ‘lay down’ see De Knop, 2016; Fagan, 1991; Hermann, 2019).

Filler trials were simple sentences, and attention checks used an equal number of temporally unambiguous relative clause constructions (i.e., temporal adverbs *vorher schon* ‘previously’, *zuerst/dann* ‘first/then’) and *bevor-/nachdem*-clause constructions (i.e., 4 trials per construction type; linguistic materials can be accessed via <https://osf.io/qgnfr/>).

3.2.4.3. Results

In 93.5% of cases, the attention control was performed correctly. Participants thus understood the task and remained attentive throughout the experimental session.

Like in Experiment 1, participants were more likely to enact stative descriptions first when the other clause encoded an eventive situation (Figure 4, pink plots). This was the case regardless of whether the stative situation was encoded by the relative clause or the main clause. Unlike in Experiment 1, however, relative clauses were generally more likely to be enacted first, when the main and relative clause did not differ in their encoded dynamicity (green plots).

For statistical analysis, we again coded relative clause first enactments as 1, and main clause first enactments as 0. The statistical analysis revealed a main effect of dynamicity in the main clause ($Df=1$, $\chi^2=52.15$, $p<0.001$), but no main effect of dynamicity combination ($Df=1$, $\chi^2=0.13$, $p=0.72$): Participants were more likely to first enact stative main clauses ($\text{mean}_{\text{MCstative}}=0.45$, $\text{SD}=0.50$) than eventive main clauses ($\text{mean}_{\text{MCeventive}}=0.77$, $\text{SD}=0.42$), but on average, performances were similar between conditions where the event types differed ($\text{mean}_{\text{mixed}}=0.56$, $\text{SD}=0.50$) and conditions where they were the same across clauses ($\text{mean}_{\text{same}}=0.65$, $\text{SD}=0.48$).

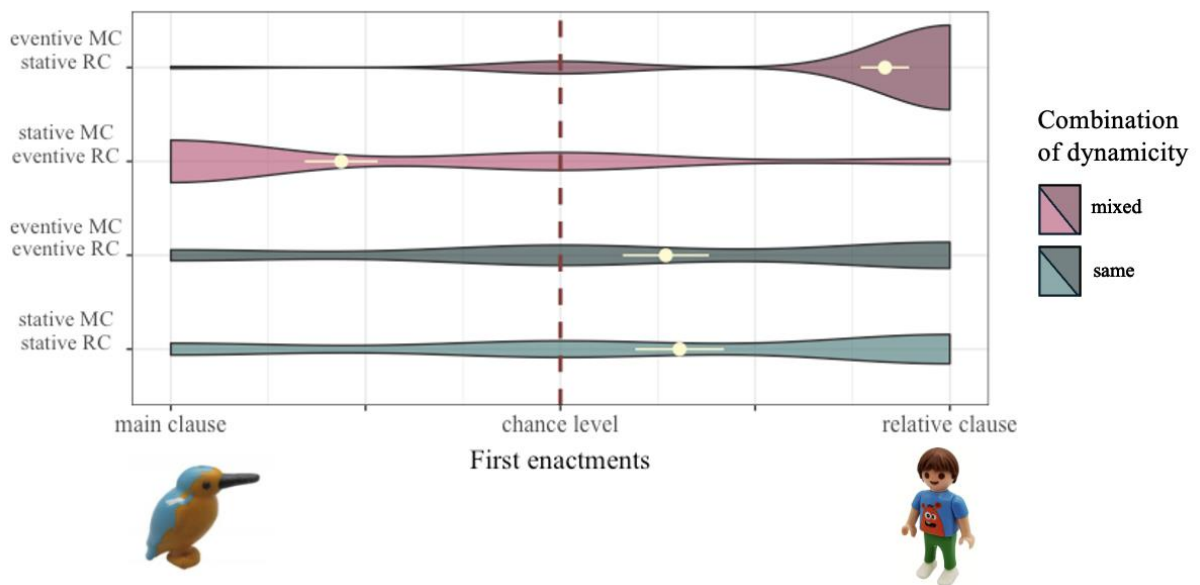


Figure 4 Mean first enactments of main or relative clause in Experiment 2 with native German participants. Dashed line depicts chance level performance. Figure illustrates the examples in (7), in which the bird appeared as subject of the main clause; first enactments of the bird figure therefore correspond to first enactment of the main clause. The boy, in contrast, appeared in the relative clause; more first enactment of the boy therefore corresponds to more first enactment of the relative clause. The width of the violin plot corresponds to means by participants; dots represent overall means, and error bars the standard error of the mean.

Crucially, the interaction between the two predictor variables was significant ($Df=1, \chi^2=61.10, p<0.001$), resulting from differences between the mixed dynamicity conditions ($\beta=0.34, t=12.94, p<0.001$): Here, relative clauses were enacted first when main clauses encoded eventive situations ($\text{mean}_{\text{mixedMCeventive}}=0.91, SD=0.28$, first line in Figure 4), but not when they encoded stative situations ($\text{mean}_{\text{mixedMCstative}}=0.23, SD=0.42$, second line in Figure 4). In cases where dynamicity was matched between clauses, participants generally preferred to enact the relative clause first ($\text{mean}_{\text{sameMCeventive}}=0.64, SD=0.48$, third line in Figure 4, vs. $\text{mean}_{\text{sameMCstative}}=0.67, SD=0.47; \beta=-0.02, t=-0.50, p=0.62$, last line in Figure 4).

3.2.4.4. Discussion Experiment 2

In Experiment 2, we aimed to understand whether effects were due to any property specific to temporal interpretations in English. We replicated the main finding of Experiment 1 in German: When one clause described a state and the other an event, participants reliably ordered the state before the event, regardless of whether it was encoded in the main or relative clause. In other words, dynamicity differences predicted the temporal order in participants' enactments. However, there was also a difference between the pattern of results in Experiment 1 and 2: When dynamicity was uniform across clauses, German participants were less sensitive towards dynamicity differences between conditions and showed an overall preference to enact relative clauses first. We will return to this cross-linguistic difference in Section 3.2.6.

In both Experiments 1 and 2, we showed that the lexical distinction between dynamic events and non-dynamic states reliably predicts temporal construal in past-under-past relative clauses. Other languages, however, use different means to express the same distinction: Slavic languages are a well-known case of marking aspect morphologically on the verb, encoding not only the speaker's perspective on a situation, but also how it is temporally and dynamically

constituted. We therefore conducted Experiment 3 in Polish – a non-SoT language for which no temporal ambiguity is reported for other embedded contexts (Mucha et al., 2023).

3.2.5. Experiment 3: Polish

Experiment 3 investigated whether the effects of dynamicity on people's temporal construals would extend to languages that (a) are non-SoT languages and thus show no temporal ambiguity in other embedded contexts, and (b) mark the distinction between events and states grammatically rather than lexically: In Polish, the perfective aspect most commonly refers to complete events, but especially in the past tense, it refers to specific events in terms of momentary occurrences, without reference to their internal temporal structure. The perfective aspect thus refers to dynamic activities that involve an instantaneous change in the event situation (Bogdan & Sullivan, 2009; Comrie, 1976; Czochralski, 1972; Młynarczyk, 2004; Vater, 1995). In this sense, situations described by perfective aspect coincide with lexically encoded events.

Polish imperfective aspect, on the other hand, indicates lack of completion or change within a described event situation and is often associated with habitual or durative states or processes as well as narrative background (Czochralski, 1972; Comrie, 1976; Vater, 1995; Młynarczyk, 2004; Bogdan & Sullivan, 2009). Thus, the imperfective aspect seems to have similar dynamic properties as lexically encoded states.

We predicted that regardless of the linguistic level at which a language marks the conceptual distinction between events and states, dynamicity should drive people's understanding of temporal sequence in language comprehension: In Experiment 3, as in the previous experiments, we expected stative situations, conveyed by imperfective aspect, to be enacted first.

3.2.5.1. Participants

As preregistered, a total of 36 native Polish speakers were recruited through CEU's recruitment portal, local language tandem cafés, and online language exchange forums. All participants met the attention control criterion and performed at least 5 out of 8 control trials correctly. Less than 1% of trials were excluded due to incorrect performance.

3.2.5.2. Linguistic materials

Critical sentences matched the materials used in Experiment 1 and 2: past-under-past relative clause constructions, describing either a dynamic movement or a static position of two toy agents, with the encoded dynamicity crossed between main and relative clause: Two critical sentences encoded with different event types in the main and relative clause (8a, d) and two critical sentences encoded the same (8b, c).

- | | | | | |
|-----|----|---|-----------------------------|--------------------------------|
| (8) | a. | [<i>Pływaczka</i> <i>usiadła</i> _{EVENTIVE} | <i>obok</i> | <i>skrzyni</i>] _{MC} |
| | | swimmer sat PRF | next.to | chest |
| | | [<i>na której</i> <i>siedziała</i> _{STATIVE} | <i>kura</i>] _{RC} | |
| | | on which sat IPFV | chicken | |
| | b. | [<i>Pływaczka</i> <i>siedziała</i> _{STATIVE} | <i>obok</i> | <i>skrzyni</i>] _{MC} |
| | | swimmer sat IPFV | next.to | chest |
| | | [<i>na którą</i> <i>usiadła</i> _{EVENTIVE} | <i>kura</i>] _{RC} | |
| | | on which sat PRF | chicken | |
| | c. | [<i>Pływaczka</i> <i>usiadła</i> _{EVENTIVE} | <i>obok</i> | <i>skrzyni</i>] _{MC} |
| | | swimmer sat PRF | next.to | chest |
| | | [<i>na którą</i> <i>usiadła</i> _{EVENTIVE} | <i>kura</i>] _{RC} | |
| | | on which sat PRF | chicken | |

d. [<i>Pływaczka</i> <i>siedziała</i> _{STATIVE}	<i>obok</i>	<i>skrzyni</i>] _{MC}
swimmer sat IPFV	next.to	chest
[<i>na której</i> <i>siedziała</i> _{STATIVE}	<i>kura</i>] _{RC}	
on which sat IPFV	chicken	

We employed the contrast between perfective and imperfective past to systematically mark different dynamic properties: Perfective forms such as *usiadła* (‘sat’) corresponded to lexically encoded events, and imperfective forms such as *siedziała* (‘was sitting’) corresponded to lexically encoded states.

Attention checks matched the relative clause constructions in the critical trials but established an unambiguous temporal event order of events, using temporal adverbs (see linguistic materials at <https://osf.io/qgnfr/>).

3.2.5.3. Results

Performance accuracy was at 97% in all attention control trials. As in the previous two experiments, participants were more likely to enact stative situations (conveyed by imperfective aspect) first, rather than eventive situations (conveyed by perfective aspect; Figure 5, pink plots). When there were no differences in aspectual marking (green plots), performance was at chance for sentences that encoded two eventive (i.e., perfective) situations, but relative clauses tended to be enacted first for sentences that encoded two stative (i.e., imperfective) situations. Again, for statistical analysis, we coded relative clause first enactments as 1, and main clause first enactments as 0. The main effect of dynamicity in the main clause ($Df=1$, $\chi^2=15.16$, $p<0.001$), but not of dynamicity combination ($Df=1$, $\chi^2=0.69$, $p=0.41$) was significant: Eventive main clauses were less likely to be enacted first (mean_{MCeventive}=0.65, SD=0.48) compared to stative main clauses (mean_{MCstative}=0.46, SD=0.50).

Like in English and German, we found a significant interaction ($Df=1$, $\chi^2=50.09$, $p<0.001$). Pairwise comparisons revealed that the interaction was driven by differences both between the two mixed dynamicity conditions ($\beta=0.29$, $t=8.34$, $p<0.001$) and between the two conditions that encoded the same dynamicity ($\beta=-0.09$, $t=-2.23$, $p=0.03$): With different event types in main and relative clause, participants enact more often the relative clause first when the main clause encoded an event ($\text{mean}_{\text{mixed_MCeventive}}=0.82$, $SD=0.39$, first line in Figure 19) than a state ($\text{mean}_{\text{mixed_MCstative}}=0.24$, $SD=0.43$, second line in Figure 5). With the same dynamicity across clauses, participants were at chance level for double-eventive sentences ($\text{mean}_{\text{same_MCeventive}}=0.49$, $SD=0.50$, third line in Figure 5) but they were more likely to enact the relative clause first for double-stative sentences ($\text{mean}_{\text{same_MCstative}}=0.68$, $SD=0.47$, last line in Figure 5).

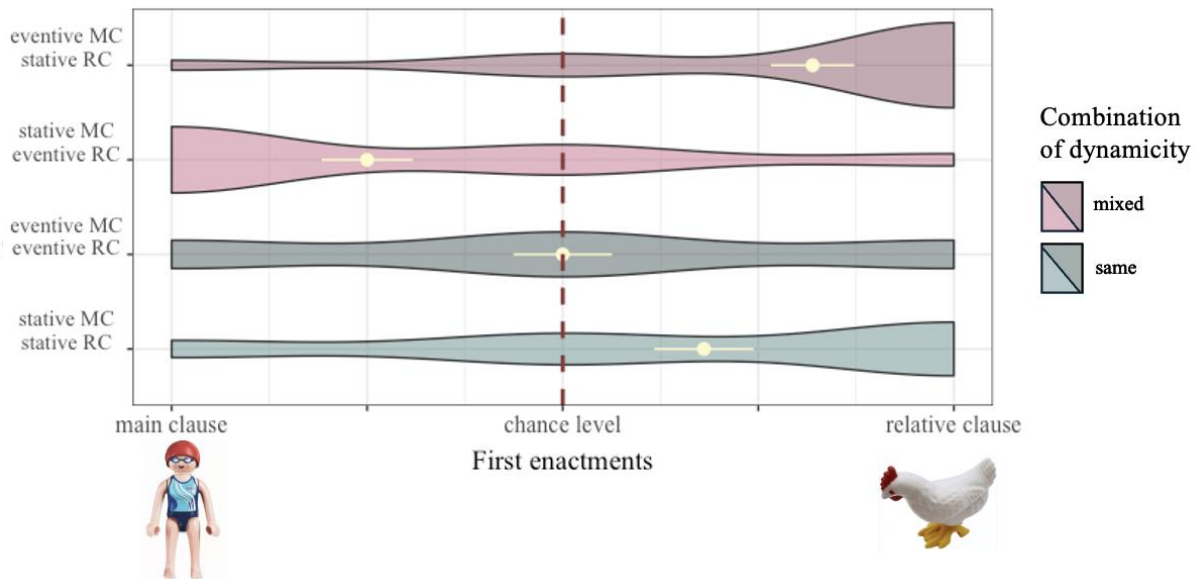


Figure 5 Mean first enactments of main or relative clause in Experiment 3 with native Polish participants. Dashed line depicts chance level performance. Figure illustrates the examples in (8), in which the swimmer appeared as subject of the main clause; first enactments of the swimmer figure therefore correspond to first enactment of the main clause. The chicken, in contrast, appeared in the relative clause; more first enactment of the chicken therefore corresponds to more first enactment of the relative clause. The width of the violin plot corresponds to means by participants; dots represent overall means, and error bars the standard error of the mean.

3.2.5.4. Discussion Experiment 3

Experiment 3 investigated whether structural differences in events also serve as a cue to the temporal order of events in non-SoT languages, and when differences are encoded by grammatical means. Overall, we replicated the pattern of results from English and German in Polish – a language without temporally ambiguous state descriptions in other embedded contexts (Mucha et al., 2023), and where aspectual marking expresses not only viewpoint differences but also the dynamic structure of events (Bogdan & Sullivan, 2009; Comrie, 1976; Czocharlski, 1972; Młynarczyk, 2004; Vater, 1995): For sentences where both a stative and an eventive situation was described, using imperfective and perfective marking respectively, participants enacted the stative description before the eventive description. The syntactic encoding of the state in the main or relative clause was irrelevant.

Like in Experiment 1, but unlike in Experiment 2, we also found a significant effect of dynamicity between conditions that encoded the same event type, though the nature of this effect differed: Performances were at chance-level when participants learned about two eventive situations, but, similar to the overall preference of German speakers in both conditions, relative clauses tended to be enacted first when sentences described two states. These differences will be discussed in the following section.

3.2.6. Between-language comparison

3.2.6.1. Statistical analysis

To investigate how the mapping of event onto temporal structure varies across languages, we conducted an exploratory analysis: We pooled the data from Experiment 1 (English), 2 (German) and 3 (Polish), and added language to the linear regression model, using a Helmert coding scheme: English and German were coded as one Helmert pair, and contrasted with Polish as the second Helmert pair.

3.2.6.2. Results

Across conditions, Polish native speakers did not enact sentences systematically different from English and German native speakers ($\text{mean}_{\text{POLISH}}=0.55$ vs. $\text{mean}_{\text{ENGLISH\&GERMAN}}=0.56$); and English native speakers did not perform differently from German native speakers ($\text{mean}_{\text{ENGLISH}}=0.50$ vs. $\text{mean}_{\text{GERMAN}}=0.61$; no main effect of language: $Df=2$, $\chi^2=2.85$, $p=0.24$).

Figure 6 shows the overall pattern of results as a combination of Figures 3-5.

However, we found a significant interaction between language and dynamicity in the main clause ($Df=2$, $\chi^2=7.45$, $p=0.02$) which, according to pairwise comparisons, was driven by differences between conditions that encoded an event in the main clause: Here, German participants were more likely to enact relative clauses first ($\text{mean}_{\text{GERMAN_eventiveMC}}=0.77$), as compared to both English participants ($\text{mean}_{\text{ENGLISH_eventiveMC}}=0.65$; $\beta=-0.58$, $z=-2.48$, $p=0.01$) and Polish participants ($\text{mean}_{\text{POLISH_eventiveMC}}=0.65$; $\beta=-0.58$, $z=-2.33$, $p=0.02$).

Language furthermore interacted with dynamicity combination ($Df=2$, $\chi^2=8.83$, $p=0.01$), such that participants performed similarly in the mixed dynamicity conditions across languages, but language differences emerged in the conditions encoding the same event type: English participants showed a weak tendency to enact main clauses first ($\text{mean}_{\text{ENGLISH_same}}=0.45$), whereas relative clauses tended to be enacted first by German ($\text{mean}_{\text{GERMAN_same}}=0.65$; $\beta=0.84$, $z=3.92$, $p<0.001$) and Polish participants ($\text{mean}_{\text{POLISH_same}}=0.58$, $\beta=0.55$, $z=2.44$, $p=0.01$).

Finally, the three-way interaction was significant ($Df=2$, $\chi^2=9.38$, $p=0.009$), showing that language differences were more pronounced in the conditions encoding the same event type, and more so when the two clauses encoded eventive situations: Polish participants performed at chance ($\text{mean}_{\text{POLISH_eventive-eventive}}=0.49$), whereas German participants tended to enact the relative clause first ($\text{mean}_{\text{GERMAN_eventive-eventive}}=0.64$) and English participants tended to enact the main clause first ($\text{mean}_{\text{ENGLISH_eventive-eventive}}=0.35$).

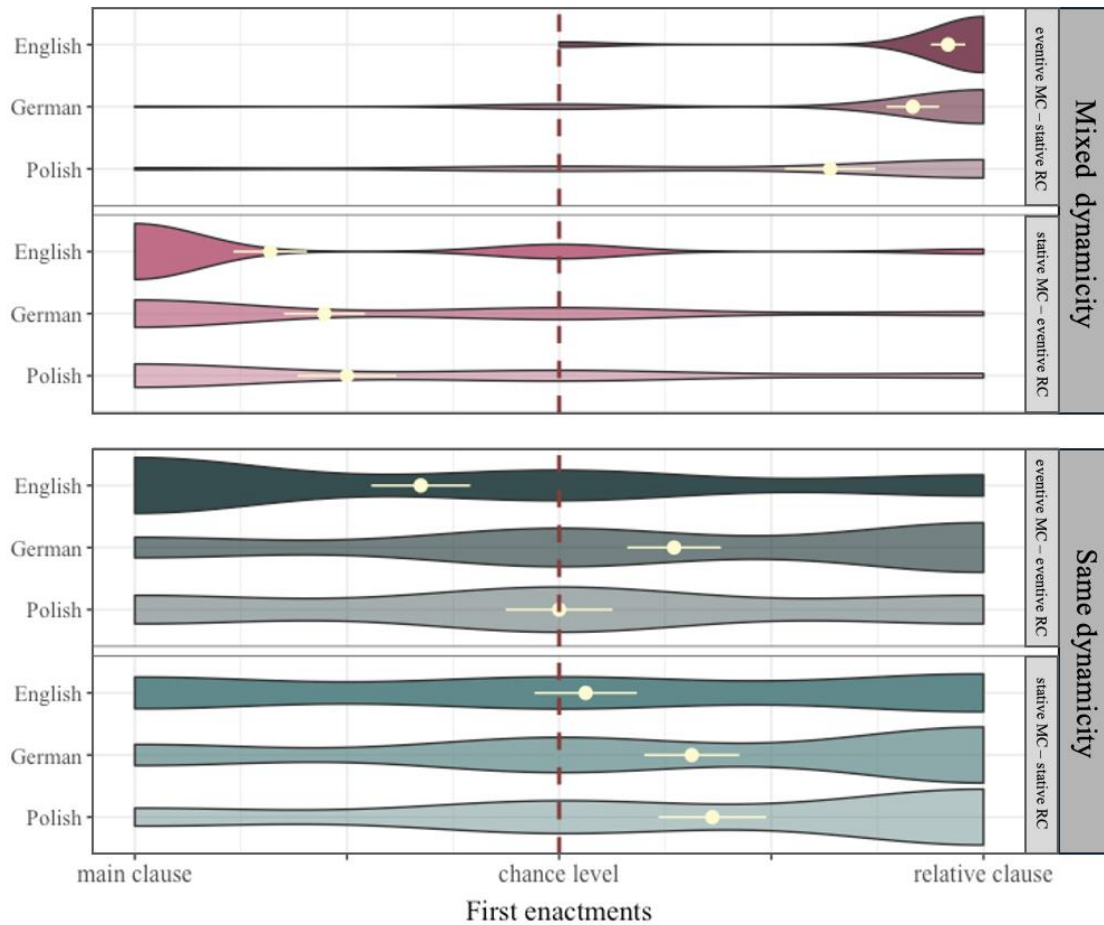


Figure 6 Mean first enactments of main or relative clause across languages (combined Figures 3-5). Dashed lines depict chance level performance, widths of violin plots correspond to means by participant, and error bars denote standard errors.

3.2.6.3. Discussion of between-language comparison

We compared three different languages to assess two generalizations: First, we asked whether linguistic encoding of dynamicity is a strong cue to temporal order across different language families, not only in SoT-languages like English and German, where dynamicity is tied to temporal ambiguity in other embedded contexts. Second, we wanted to determine whether the linguistic level at which languages encode event dynamic properties – i.e., lexicon or grammar – would interact with temporal construal derived from these properties.

The answer to both questions is clear: The between-language comparison revealed that dynamicity serves as a reliable cue to temporal order in both SoT-languages (English, German) and non-SoT languages (Polish), at least when two situations differ in dynamicity.

Any differences between languages pertained to the conditions encoding the same dynamic properties, which speaks more to the question of how participants construed temporal order in the absence of, rather than on the basis of, dynamicity differences. Here, English and German speakers used different pragmatic strategies: English participants took the iconic order of mention to reflect the temporal order of the events, but not the states in the world, whereas German overall adhered to the discourse pragmatic function of relative clauses, assuming that they generally encode situational background. Polish speakers patterned with German speakers, but more so when enacting double stative sentences.

Since our focus in this paper was on the effect of dynamicity differences, we had no clear predictions about which strategy speakers of different languages should follow in the absence of such differences. However, two points are noticeable: First, the same-dynamicity conditions in all three languages elicited comparably noisy data (green plots). Language-specific frequency, markedness, as well as discourse-pragmatic functions of relative clauses may play a role here. We suggest further research to understand why speakers rely on different cues across English, German, and Polish.

Second, and more importantly, when dynamicity was indeed mixed (pink plots), the effects were clear and consistent across languages, confirming previous results in different tasks (Marx & Wittenberg, 2022, 2025) and linguistic structures (Marx & Wittenberg, 2024). That is, state-first temporal construal seems to be a robust and reliable strategy that is employed cross-linguistically under conditions of temporal uncertainty.

3.3. General discussion

In this article, we investigated whether people would infer the temporal structure of a complex situation – that is, when a situation occurred relative to another situation – from the encoded dynamic properties of its parts. We asked: In cases where language does not provide sufficient

information to establish temporal order, is one described situation temporally ordered relative to another depending on its linguistic encoding as a state or an event? Furthermore, do dynamicity differences guide temporal construal across languages independently of Sequence of Tense, and independently of whether dynamicity is marked in a language's grammar or lexicon?

To answer these questions, we conducted three experiments in English, German and Polish – three languages that differ typologically and in terms of their tense-aspect systems – while using a linguistic context that describes two situations without any systematic causal, temporal, or conditional link: past-under-past relative clauses.

In previous work on relative clause constructions, we showed that people tend to prefer stative situations to happen before eventive situations, regardless of their syntactic encoding in either the main or relative clause, and regardless of discourse pragmatic factors, such as linearity or a background-foreground distinction between clauses (Marx & Wittenberg, 2022, 2025). However, one limitation of previous research is that participants explicitly judged how well sentences matched visual scenes. This leaves open the possibility that the observed preference was induced by factors other than the underlying effects of dynamicity on temporal construal, but related to task design (i.e., post-hoc explanations, metalinguistic knowledge, see e.g., Kepser & Reis, 2005; Schütze & Sprouse, 2013). In the current series of studies, we therefore assessed temporal construal through language, and independent of visual representations: Participants enacted relative clause descriptions, using toys as props, according to their own understanding of the sentences.

In all three languages, we replicated the previous pattern of results. When one of the clauses encoded an event and the other one a state, dynamicity predicted participants' performance: Stative situations were reliably ordered before eventive situations. This was the case regardless

of language-specific SoT properties or the linguistic encoding of dynamicity. In what follows, we discuss the implications of this cross-linguistically stable effect.

First and foremost, our findings provide further support for accounts that place event structure at the core of understanding temporal structure: In complex situation descriptions, comprehenders constantly monitor how individual subevents are linked (Kehler, 1994; Radvansky et al., 1998; Zwaan et al., 1995). In terms of temporal relations, this requires them to temporally anchor each described situation by another time or situation (Carroll & von Stutterheim, 2010; Klein, 1994, 2009; von Stutterheim et al., 2003). Although comprehenders may use utterance time as a temporal anchor, event-structural accounts predict that they are more likely to choose another, more salient situation in more complex descriptions. What our findings show is that one factor that seems to be a robust determinant of this choice is dynamicity: Dynamic events are cognitively more salient than stable states (Clewett et al., 2020; Kurby & Zacks, 2008; Zacks et al., 2007), therefore, descriptions encoding dynamic events tend to serve as temporal anchors for descriptions of non-dynamic states. Participants' enactments were in line with these predictions.

Admittedly, our interpretation of these data rests on the level of the events themselves. An alternative explanation may be that the referents involved in the acted-out situations were rendered more salient by virtue of participating in a dynamic event rather than a static situation, giving rise to temporal order not based on the event itself, but an increase of referents' saliency.²² On the basis of our data, we cannot decide whether the saliency introduced by dynamicity was related only to event participants, or to entire situations, or to both. However, we would assume that, either way, these saliency differences result in temporal construal along the lines event-structural accounts propose. By contrast, if our data reflected only saliency differences (especially for event referents) without temporal construal, our intuitive prediction

²² We thank an anonymous reviewer for this suggestion.

would be that the most salient referent – that is, the dynamic event participant – should have been enacted first, in line with previous work on agent saliency (Dowty, 1991; Goldin-Meadow, So, Özyürek, & Mylander, 2008; Hafri, Trueswell, & Strickland, 2018; Mayrhofer & Waldmann, 2014; Verfaillie & Daems, 1996) and not second, as it was the case in our data.

Second, our results show that temporal inferences, driven by dynamicity, arise independently of how differences are encoded in a given language: Participants ordered stative descriptions before eventive descriptions when structural differences between events were rooted in the lexicon, as in English (Experiment 1) and German (Experiment 2), but also when the difference between states and events was grammatically marked, as in Polish (Experiment 3). While this should be confirmed by large-scale cross-linguistic research, our findings are a first attempt to test the generalizability of the effects of dynamicity on temporal construal.

Thirdly, our findings target an important gap in the discussion on lexical and grammatical aspect (cf. Bertinetto & Delfitto, 2000; Binnick, 1991; Carlson, 1981; Comrie, 1976; de Swart, 2012; Depraetere, 1995; Dowty, 1977, 1979; Filip, 2019; Garey, 1957; Moens, 1987; Moens & Steedman, 1988; Mourelatos, 1978; Parsons, 1989; Sasse, 2002; Smith, 1986). While many studies emphasize the importance of telicity (or boundedness) and duration (or durativity) to determine aspectual class (e.g., Bertinetto & Delfitto, 2000; Comrie, 1976; Depraetere, 1995; Dowty, 1979; Filip, 2019; Lyons, 1977; Mourelatos, 1978; Pustejovsky, 1991; Verkuyl, 1972), to drive temporal inferences (Becker et al., 2013; Flecken & Gerwien, 2013; Milczarski, 2021; Wittenberg & Levy, 2017), and to understand inferences of event completion, repetition, and individuation (Barner et al., 2008; Bott, 2010; Bott & Gattnar, 2015; Foppolo et al., 2021; Minor et al., 2022; Paczynski et al., 2014; Piñango et al., 2006), dynamicity is a third, much more overlooked component of aspect in research on temporal construal (Olsen, 1994).

Overall, our results indicate that manipulating dynamicity leads to consistent temporal inferences across three languages. One may speculate that the reason why dynamicity has been

less interesting for linguistic research is that it is perceptually grounded: Unlike telicity, which is an inherently linguistic feature, susceptible to grammatical and lexical devices, it is part of event knowledge whether a situation changes or not. In perception, dynamicity is immediately evident; one might even speculate that it is less penetrable by linguistic encoding. We leave this question to future research at the interface of language and event cognition.

3.4. Conclusion

The present series of experiments demonstrates that, without specific information about temporal order, comprehenders resolve the lack of specification in complex linguistic descriptions by using the event dynamic properties of a sentence as a cue. In three act-out experiments, conducted in English, German and Polish, participants consistently preferred arranging non-dynamic states prior to dynamic events. Our results thus confirm findings from previous acceptability judgment studies, by probing the mental construal of temporal sequences. Overall, our studies provide further insights into the underlying mechanisms of event comprehension, suggesting that, across languages, event knowledge is at the heart of understanding temporal relations in language.

4. Beyond relative clauses – Temporal order inference in discourse comprehension

4.1. Introduction

Understanding temporal order is a crucial human capacity (Downes et al., 2002; Kausler et al., 1988; Sacks, 1985). When we read a story, some of the most important information to keep track of is what happens when, and how readers do this in discourse comprehension has been a vivid question in research for decades (e.g., Graesser, Singer, & Trabasso, 1994; Jakobson, 1965; Kehler, 1994, 2002; Oversteegen, 2005; Tai, 1983). Here, we focus on the temporal inferences people draw to mentally order situations expressed in successive sentences.²³ Our starting point is a classic example:

- (1) *Mary got pregnant. She got married to John.*
- (2) *Mary got married to John. She got pregnant.*

Examples (1) and (2) can be found in various versions throughout the literature on discourse comprehension over the past 70 years or so (e.g., Fleischman, 1990; Horn, 2022; Kamp & Rohrer, 1983; Lascarides & Asher, 1991, 1993; Levelt, 1989; Schmerling, 1975; Strawson, 1952; Webber, 1988). They have been used to illustrate that the order of successive sentences in a narrative is interpreted as iconically mapping to the temporal order of the expressed situations: The inferred order in (1) is that Mary was pregnant before she married John; example (2) may also trigger a causal inference, namely, that perhaps Mary and John got married because Mary was already pregnant, presumably – another inference – by John. Conversely, the

²³ There is considerable variation in the notions employed around the linguistic expression of event structure. We will use the following terminology, to minimize confusion between linguistic notions and cognitive counterparts: We refer to the order of events imagined as they happened, as ‘inferred order’, but to the order of successive sentences as ‘reported order’. When talking about non-linguistic events or states, we call them ‘situations’, and their linguistic encodings, following Bach (1986), are called ‘eventualities’. Situations comprise both non-dynamic ‘states’ and change-of state ‘events’; which are expressed linguistically as ‘stative descriptions’ or ‘eventive descriptions’. For the purpose of this paper, we do not distinguish between achievements and accomplishments.

inference triggered by (2) is that Mary and John first got married, and then, Mary became pregnant. In any case, the example pair in (1) and (2) has been presented to generations of linguistics students as evidence that inferred orders are triggered by reported order in linguistic discourse. However, this claim is incomplete on two accounts: First, there is a lack of empirical or experimental data to confirm these intuitions (Gibson & Fedorenko, 2010); and second, while the literature clearly recognizes effects of eventuality type (Hinrichs, 1986; Hopper, 1979), i.e., whether a situation is relayed using an eventive or a stative description, it is lacking a truly methodical investigation of how linguistic eventuality type systematically affects these inferences.

Most work on discourse inference has been concerned with causal inference (Briner et al., 2011; Oversteegen, 2005; Wolfe et al., 2005). Causal inference is, of course, tightly linked to temporal inference, since causes precede effects (cf. the Causal Law, Lascarides & Asher, 1993). However, not all temporal sequences are also causally related, or form contingency relations. For instance, in (2), Mary's marriage may or may not have causally contributed to the subsequent pregnancy; and even any potential causal inference in (1) is, at least in our intuition, weaker than the temporal order inference. For the moment, we therefore leave aside the question of causality and focus only on the temporal inference.

Temporal inference in discourse is made easier when two successive sentences are marked through temporal adverbs or conjunctions, or with different tense morphology. In (1') and (2'), for instance, the temporal order inferences are countercorrelated with the reported orders, because in (1'), a temporal conjunction leads to the temporal reordering of the described situations, whereas in (2'), the first sentence is marked with present tense and the second with past tense:

(1') *Mary got pregnant **after** she got married to John.*

(2') *Mary **gets** PRESENT married to John. She **got** PAST pregnant.*

There are two broad classes of theoretical accounts explaining how people infer temporal order of events in discourse through tense: one in which tense serves as an anaphora that guides the relative temporal ordering of situations (Carroll & von Stutterheim, 2010; Hinrichs, 1986; Klein, 1994, 2009; Oversteegen, 2005; Partee, 1973; von Stutterheim et al., 2003; Webber, 1988), and another in which temporal order inferences are byproducts of establishing overall coherence in the discourse, using tense as a cue (Kehler, 1994, 2002, 2006; Lascarides & Asher, 1993). Based on tense alone, both families of accounts would predict that in (1) and (2), the inferred order follows the reported order, since there is no tense differential.

While both of these accounts discuss examples of tense differentials affecting discourse interpretation, a factor that is not manipulated in (1) and (2), only one of them additionally makes clear predictions according to eventuality type, i.e. whether a situation is encoded by a stative description, or by an eventive description. Specifically, Hinrichs (1986) elaborates that stative descriptions following eventive descriptions (examples (5) and (6) below) can be interpreted as temporally overlapping with events (see also Dowty, 1986; Partee, 1984). But why should this be? Hinrichs (1986) simply refers to Vendler's (1967) Aktionsarten as the descriptive categories determining whether successive descriptions trigger sequential or overlapping inferences.

That said, a possible explanation is given by accounts that closely link non-linguistic event cognition to the temporal interpretation of linguistic descriptions, by proposing a temporal anchor relative to which other descriptions are interpreted (Carroll & von Stutterheim, 2010; Klein, 1994, 2000, 2009; von Stutterheim et al., 2003; but see also Kameyama et al., 1993; Kamp & Reyle, 1993). Specifically, it has been argued that temporal expressions, such as tenses, denote complex temporal structures that can be formalized along several interrelated temporal intervals: With each temporal expression, speakers make an assertion about a specific time – a topic time – which for the simple past, selects an interval of the time when a situation

took place, including its rightmost boundaries (i.e., a culmination point for achievements and accomplishments, or a post-state for states and activities). This topic time is then matched to another temporal interval: a temporal anchor. Such an anchor can be, among other possibilities, determined by knowledge about event structure; the crucial property is that the temporal anchor should be the most salient element in a discourse. Research on event perception has shown that events are cognitively more salient than stable states (Clewett et al., 2020; Kurby & Zacks, 2008; Zacks et al., 2007); in linguistic discourse, this should mean that eventive predicates tend to function as anchoring situations, and states tend to function as anchored situations. We find a related idea in the distinction between foreground and background, with implications for narrative progression (cf. Hopper, 1979; Wårvik, 2004).

Crucially, given that for simple past, a topic time always includes the rightmost boundary of the situation time (i.e., culmination point/post-state), this also imposes a constraint on its leftmost boundary: That is, the anchored situation cannot start after the time of the temporal anchor. Therefore, we should find that (a), if a situation is linguistically described as eventive, it should serve as a temporal anchor for a stative description. Furthermore, in two sentences in simple past, we should find that (b), in the absence of a simultaneous choice, people's temporal intuitions should order stative descriptions before eventive descriptions.

We conducted two preregistered experiments in order to systematically address these predictions, using a forced-choice task that probes inferences of temporal order. Experiment 1 manipulated linguistic eventuality in the classical examples (1) and (2) in a single-trial design. In Experiment 2, we extended Experiment 1 in two ways: First, we used a broader range of situation descriptions to assess to what extent our findings would generalize across verbs and events. Second, using multiple items allowed us to test people's temporal inferences in a within-subject design.

For both experiments, we argue that the widely reported intuitions for sentences like (1) and (2), according to which the inferred temporal order matches the reported linear order of the two sentences, only arise because both Mary's marriage and her pregnancy are linguistically conveyed by eventive descriptions (1'' and 2''):

(1'') *Mary **got pregnant** EVENTIVE. She **got married to** EVENTIVE John.*

(2'') *Mary **got married to** EVENTIVE John. She **got pregnant** EVENTIVE.*

The same correlation between reported and inferred order should also hold when the first sentence contains a stative description, such as in (3) and (4). This prediction follows from assuming an iconic mapping between sentence order and temporal order, but more importantly for our purposes, it arises independently from a general cognitive tendency to order states first in time.

(3) *Mary **was pregnant** STATIVE. She **got married to** EVENTIVE John.*

(4) *Mary **was married to** STATIVE John. She **got pregnant** EVENTIVE.*

In contrast, only from an event-structural perspective, when a stative description follows an eventive description, such as in (5) and (6), should the inferred order be opposite to the reported order:

(5) *Mary **got pregnant** EVENTIVE. She **was married to** STATIVE John.*

(6) *Mary **got married to** EVENTIVE John. She **was pregnant** STATIVE.*

In (5), the state of being pregnant should be inferred to have occurred before the wedding event; and in (6), the state of being married should be inferred to have occurred before the event of becoming pregnant. That is, (5) and (6) are predicted to reverse the temporal inference classically reported in the literature (e.g., Fleischman, 1990; Horn, 2022; Kamp & Rohrer, 1983; Lascarides & Asher, 1991, 1993; Levelt, 1989; Schmerling, 1975; Strawson, 1952; Webber, 1988).

Finally, one open question is how two stative descriptions following each other are ordered relative to each other, as in (7) and (8):

(7) *Mary **was pregnant** STATIVE. She **was married to** STATIVE John.*

(8) *Mary **was married to** STATIVE John. She **was pregnant** STATIVE.*

While none of the theoretical approaches mentioned above clearly predicts a temporal order between two states, the assumption that stative descriptions elicit overlapping inferences (Dowty, 1986; Hinrichs, 1986; Partee, 1984) might prevent people from sequentially ordering the two situations in (7) and (8) in time (see also a parallel coherence relation in Kehler, 1994, 2006, without reference to eventuality). In such a case, the two possible temporal orders should be at chance level when participants are asked to choose one over the other.

Table 1 Outline of predicted inferred temporal order based on eventuality type order in Experiment 1 and 2. Diagonal lines indicate similar predictions following from reported order and event structure.

Eventuality order	Predicted inferred order	Exp.	Example
eventive - eventive	reported order	1	<i>got pregnant – got married</i>
		2	<i>boarded a train – went on holiday</i>
eventive - eventive	reported order	1	<i>got married – got pregnant</i>
		2	<i>went on holiday – boarded a train</i>
stative - eventive	reported order = state before event	1	<i>was pregnant – got married</i>
		2	<i>was on a train – went on holiday</i>
stative - eventive	reported order = state before event	1	<i>was married – got pregnant</i>
		2	<i>was on holiday – boarded a train</i>
eventive - stative	opposite order → state before event	1	<i>got pregnant – was married</i>
		2	<i>boarded a train – was on holiday</i>
eventive - stative	opposite order → state before event	1	<i>got married – was pregnant</i>
		2	<i>went on holiday – was on a train</i>
stative - stative	opposite order → long before short state	1	<i>was pregnant – was married</i>
		2	<i>was on a train – was on holiday</i>
stative – stative	reported order = long before short state	1	<i>was married – was pregnant</i>
		2	<i>was on holiday – was on a train</i>

However, there is broad consensus in the literature that a crucial dimension of states is their duration (see Vendler, 1967; Carlson, 1977, Kratzer, 1995, for durational differences between stage/individual-level predicates, also Dery & Koenig, 2015), apart from their lack of change or dynamicity. Thus, if duration influences the way in which people conceptualize states, we would predict that differences in the duration typically associated with two stative predicates should lead to different temporal construals: That is, given that being pregnant typically lasts a shorter time than being married, pregnancies should be less stative, and thus more salient, than marriages. For (7), we would therefore predict a reversal of reported order, as in (5) and (6), whereas the double-stative sequence in (8) should follow the reported order of the two clauses as in (3) and (4). Predictions for both experiments are outlined in Table 1.

4.2. Experiment 1

4.2.1. Methods

4.2.1.1. Participants

We recruited 800 native English speakers via Prolific.co, to collect 100 observations per condition.²⁴ The experiment was programmed in PsychoPy and conducted via Pavlovia.org; it lasted approximately 2 minutes. 42 additional participants were excluded, because they terminated the experiment prematurely, or due to data recording problems.

4.2.1.2. Materials and procedure

Critical stimuli consisted of two main clauses describing two situations (i.e., pregnancy and marriage) in simple past. Situations were crossed in reported order (i.e., pregnancy first: (1), (3), (5), (7) vs. marriage first: (2), (4), (6), (8)); we also crossed eventuality type in the first

²⁴ Sample size was based on the work of Morgan et al. (2020), who used a single-trial design to collect 50 data points per condition (cf. Exp. 4), but we doubled the number of observations per cell, due to unknown effect sizes.

clause (i.e., eventive: (1-2), (5-6) vs. stative: (3-4), (7-8)) and the second clause (i.e., eventive: (1-4) vs. stative: (5-8)).

We used a single-trial, between-subjects design (see e.g., Morgan et al., 2020 Experiments 1 and 4; von der Malsburg et al., 2020 particularly the belief-estimation tasks). This choice was motivated by task adaptation effects which have been shown to emerge in a variety of experimental designs, both due to the repetition of critical trials (e.g., Balota et al., 2018; Demberg & Sayeed, 2016; Fine et al., 2013; Hammerly et al., 2019; Ness & Meltzer-Asscher, 2021; Pregla et al., 2021), and the inclusion of filler trials (e.g., Cowart, 1997; Keller, 2000; Schütze & Sprouse, 2013), commonly employed in within-subject designs (Arehalli & Wittenberg, 2021; Laurinavichyute & von der Malsburg, 2024, for discussion).

Participants were randomly assigned to one of the eight conditions, and then asked to indicate which of the two situations had occurred first by choosing between two buttons (i.e., pregnancy vs. marriage; Figure 1).

4.2.1.3. Statistical analysis

We conducted two types of statistical analyses: First, a logistic regression model was fitted in the R statistical environment (i.e., *glm* from the stats package, R Core Team, 2014) with reported order (i.e., marriage-first vs. pregnancy-first), the linguistic eventuality type in the first

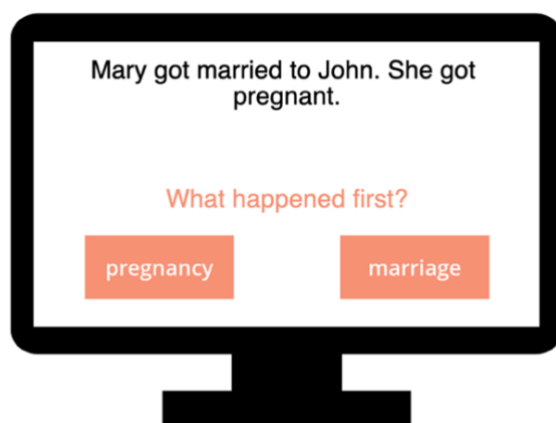


Figure 1 Example of the experimental screen participants saw in the forced-choice task, here in the eventive-eventive combination with marriage first. The side of the buttons was randomized.

sentence (i.e., stative vs. eventive), and the combination of eventuality types across sentences (i.e., same: eventive-eventive/stative-stative vs. mixed: eventive-stative/stative-eventive) as sum-contrast coded predictor variables. The dependent variable was whether people chose the situation that was reported first as having happened first. To assess the significance of the full model, we performed likelihood-ratio tests between the full model and the reduced models, excluding one of the predictors or interactions individually. Model comparison was performed with each of the three predictor variables and the four possible interactions. Second, we ran a series of planned pairwise comparisons to assess the interactions of the regression models. Preregistrations for all analyses are available at <https://osf.io/fnmq4>.

4.2.2. Results

Participants' mean response choices are shown in Figure 2 (planned pairwise comparisons between conditions are outlined in Figure 3): When the sentences described two events (dark green bars), participants' choices matched the reported order, such that the first reported situation was chosen to occur first. This dovetails with the temporal inferences for (1) and (2) as described in the literature, in which the reported order of situations iconically determines how they are ordered in time – at least when both situations are linguistically encoded as events. Accordingly, the statistical analysis revealed a significant difference in the choices participants made depending on the eventuality type of both clauses (see Figure 3.I, $\beta = -0.52$, $t = -13.77$, $p < 0.001$): That is, first reported events were chosen significantly more often when both clauses were eventive descriptions ($\text{mean}_{\text{event-event}} = 0.95$, $\text{SD} = 0.22$) than when they were stative descriptions ($\text{mean}_{\text{state-state}} = 0.43$, $\text{SD} = 0.50$).

When the two reported situations did not match in eventuality type (teal and olive bars), participants chose the first reported situation only when it was a stative description. Thus, for (3) and (4), where the first clause described a state, and the second an event, participants

followed the reported order of the two clauses, ordering the situation in the first clause first in time, but for (5) and (6), where the first clause was eventive and the second stative, the reported order was reversed: Here, participants were more likely to choose the second reported situation to have happened first.

Accordingly, the statistical analysis revealed a significant difference between the two mixed-eventuality conditions where events were reported first and the two conditions where states were reported first (see Figure 3.II, $\beta=0.71$, $t=20.55$, $p<0.001$). As predicted, participants were significantly more likely to choose first described situations when they were stative (mean_{stative_first}=0.93, SD=0.25) than when they were eventive (mean_{eventive_first}=0.22, SD=0.42). Similarly, when comparing between conditions where an eventive pregnancy was mentioned first (see Figure 3.III, $\beta=-0.75$, $t=-16.56$, $p<0.001$) as well as between conditions where an eventive marriage was mentioned first (see Figure 3.IV, $\beta=-0.69$, $t=-14.38$, $p<0.001$), the

“What happened first?”

Mean response choices and standard errors (N=800)

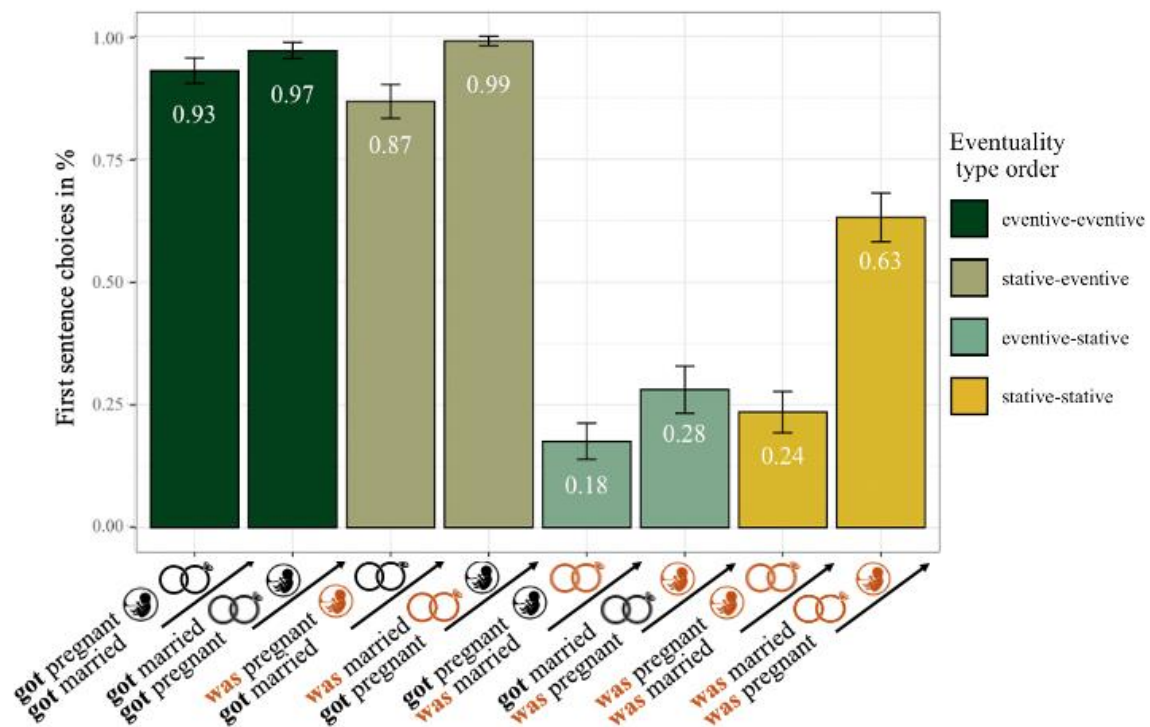


Figure 2 Mean proportions of first sentence choices across eight conditions, error bars represent standard errors. The x axis shows the reported order of the two situations, orange icons indicate stative eventuality type.

situation described first was more likely to be chosen when followed by another eventive description (eventive pregnancy first, ex. (1): $\text{mean}_{\text{eventive_pregnancy}}=0.93$, $\text{SD}=0.26$; eventive marriage first, ex. (2): $\text{mean}_{\text{eventive_marriage}}=0.97$, $\text{SD}=0.17$) than by a stative description (eventive pregnancy first, ex. (5): $\text{mean}_{\text{eventive_pregnancy}}=0.18$, $\text{SD}=0.38$; eventive marriage first, ex. (6): $\text{mean}_{\text{eventive_marriage}}=0.28$, $\text{SD}=0.45$).

Neither the main effects of eventuality in the first clause ($Df=1$, $\chi^2=3.22$, $p=0.07$) and eventuality combination ($Df=1$, $\chi^2=0.95$, $p=0.33$) reached significance, nor did the two-way interaction between reported order and eventuality combination ($Df=1$, $\chi^2=0.31$, $p=0.58$) or the three-way interaction that also included eventuality in the first clause ($Df=1$, $\chi^2=1.17$, $p=0.28$). However, we found a significant interaction between eventuality combination and eventuality in the first clause ($Df=1$, $\chi^2=363.47$, $p<0.001$): That is, the first reported eventive situation was chosen to happen first, but only if the second reported situation was also eventive. When the first clause described an event and the second a state, participants chose the state to happen first. Conversely, when the first clause described a state, it was only more likely to happen first when the second clause encoded an event.

For double-stative descriptions (ochre bars), i.e., (7) and (8), the pattern was split: When marriage was reported first, participants followed the clauses' reported order ($\text{mean}_{\text{stative_marriage}}=0.63$, $\text{SD}=0.48$), but when pregnancy was reported first, they tended to choose the second mentioned situation ($\text{mean}_{\text{stative_pregnancy}}=0.24$, $\text{SD}=0.43$). This was also reflected in the statistical analysis: There was (a) a significant difference of reported order between the double-stative conditions (see Figure 3.V, $\beta=-0.40$, $t=-6.10$, $p<0.001$), (b) a main

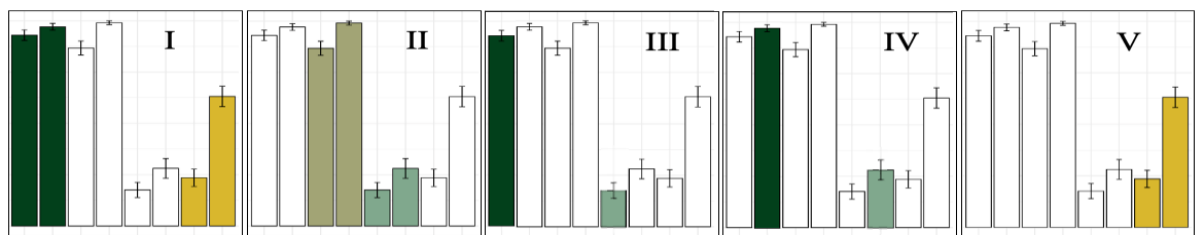


Figure 3 Outline of planned pairwise comparisons, colors of bars reflect compared conditions.

effect of reported order ($Df=1, \chi^2=29.30, p<0.001$) such that people were generally more likely to order marriage first when it was reported first ($\text{mean}_{\text{marriage_first}}=0.74, \text{SD}=0.44$) than when it was reported second ($\text{mean}_{\text{pregnancy_first}}=0.54, \text{SD}=0.50$), and (c) a significant interaction between reported order and eventuality in the first clause ($Df=1, \chi^2=5.57, p=0.018$): Overall, participants chose the first reported situation more often when it was stative and encoded marriage, rather than pregnancy; however, for first reported eventive situations, the difference between linear orders was less pronounced.

4.2.3. Discussion of Experiment 1

In Experiment 1, we asked about the mechanisms of temporal order inference in discourse, taking the classic example of Mary's pregnancy and her marriage to John as a starting point. By systematically manipulating the reported order as well as the eventuality type of both sentences, we found that linguistic eventuality was a strong predictor of people's temporal interpretations: As predicted, participants were more likely to order states before events, regardless of whether they read about the state in the first or the second sentence. In contrast, reported order, which has been argued to be a strong cue to temporal order in discourse comprehension, only guided people's temporal inferences in double-eventive sequences.

Regardless of the reported order, in double-stative descriptions participants were more likely to order marriage before pregnancy. Although this pattern of results is consistent with our assumption that people conceptualize a described state as more or less stative on the basis of its duration, and thus choose the longer state to have happened first, this pattern of results could also be due to a causal asymmetry between the two situations according to people's world knowledge: That is, people may expect marriages to happen first because of cultural conventions.

In Experiment 2, these two possibilities were examined by testing whether differences in the expected duration of states would also affect people's temporal inferences about other situation descriptions.

4.3. Experiment 2

Experiment 2 broadened our empirical base from just a single classic item (i.e., Mary's pregnancy vs. marriage) to a range of items tested in a within-participants design. In addition, we followed up on the question whether other dimensions of event structure influence the way in which people conceptualize states: that is, duration. If longer states are conceived as more stative, and thus, more backgrounded, than shorter states, we should find a systematic asymmetry between double-stative sentences that encode different durations across items.

4.3.1. Methods

4.3.1.1. Participants

Due to a technical error, we overshot our goal of 100 participants and ended up with 130 full datasets, recruited via Prolific.co. We excluded 23 additional participants who did not meet the sanity check criterion (i.e., at least 50% of the filler trial responses had to be answered correctly). The experiment took approximately 10 minutes. Only complete data sets were submitted to statistical analysis.

4.3.1.2. Materials and procedure

As in Experiment 1, the critical stimuli consisted of two sentences that were crossed with respect to their reported order (i.e., predicate 1-2 vs. predicate 2-1), the eventuality type in the first sentence (i.e., eventive vs. stative), and the eventuality type in the second sentence (i.e.,

eventive vs. stative). All sentences consistently used the simple past tense (for a full list of stimuli, see <https://osf.io/5tz6r>).

The two sentences in each item pair were created such that they described two situations that could have happened in either order. For instance, in examples (9-16), Laura's boarding the train and her going on holiday do not necessarily entail a causal order or one-directed contingency: Laura could board a train to go to her holiday destination, or she could make a train trip after starting her holiday.

- (9) *Laura **boarded**_{EVENTIVE} a train. She **went**_{EVENTIVE} on holiday.*
- (10) *Laura **went**_{EVENTIVE} on holiday. She **boarded**_{EVENTIVE} a train.*
- (11) *Laura **boarded**_{EVENTIVE} a train. She **was**_{STATIVE} on holiday.*
- (12) *Laura **went**_{EVENTIVE} on holiday. She **was**_{STATIVE} on a train.*
- (13) *Laura **was**_{STATIVE} on a train. She **went**_{EVENTIVE} on holiday.*
- (14) *Laura **was**_{STATIVE} on holiday. She **boarded**_{EVENTIVE} a train.*
- (15) *Laura **was**_{STATIVE} on holiday. She **was**_{STATIVE} on a train.*
- (16) *Laura **was**_{STATIVE} on a train. She **was**_{STATIVE} on holiday.*

While in Experiment 1, eventiveness and stativeness was based on a minimal difference of the finite verb (eventive: *to get* + adjective vs. stative: *to be* + adjective), in Experiment 2 we based the alternations on the stative and eventive meaning of the entire predicate. Therefore, there was less control over the surface features of the predicate, such as length, but we deliberately were able to broaden the range of predicates (in the above examples, copula constructions with a prepositional phrase, *to board*, *to go*).

As in Experiment 1, participants were asked to select one of two buttons to indicate which of the two sentence situations had occurred first. However, in this experiment, participants encountered 29 trials in a within-subject design: After a single practice trial, participants completed eight critical trials, one in each condition. In addition, participants also made

temporal judgments on 20 fillers, which likewise consisted of two sentences, but used either grammatical (i.e., tense/aspect) or lexical markers (i.e., adverbs and connectives) to establish an unambiguous temporal order between the two reported situations (e.g., *Julius stared out of his window. Then, he went to play with his little brother*). In half of the filler trials, the temporal order did not match the sentences' reported order, in order to induce a reversal of the reported order in some of the trials. Filler trials functioned as a sanity check.

To counterbalance between subjects, participants were randomly assigned to one of eight lists using a Latin Square design (reported order, eventuality type combination, situation item). All experiments were programmed in PsychoPy and conducted online at Pavlovia.org.

4.3.1.3. Statistical analysis

Similar to Experiment 1, we performed two types of preregistered statistical analyses (available at <https://osf.io/5tz6r>): First, we built a logistic regression model with the reported order of situations (i.e., predicate 1-2 vs. predicate 2-1), the eventuality type in the first sentence (i.e., stative vs. eventive), and the combination of eventuality types across sentences (i.e., same: eventive-eventive/stative-stative vs. mixed: eventive-stative/stative-eventive) as sum-contrast coded predictor variables in the R statistical environment (i.e., *glmer* from the stats package, R Core Team, 2014). The within-subjects design allowed us, in contrast to Experiment 1, to include participants and items as random intercepts.

As before, we used as the dependent variable whether participants chose the first sentence to have happened first in time. We performed likelihood ratio tests to compare the full model with reduced models, which excluded one of the predictors or one of the interactions at a time. Additionally, we performed a series of planned pairwise comparisons to further assess the nature of the effects and interactions of the regression models.

For analysis, predicate 1 was coded as the predicate we roughly estimated to be shorter in duration in the stative form (*being on a train* in the examples above) compared to predicate 2 (*being on holiday* in the examples above).

4.3.2. Results

Participants' mean response choices are shown in Figure 4: Overall, we replicated the pattern of results from Experiment 1, extending it to a broader range of situation descriptions as well as to a within-subjects design. All of the effects we found previously were present again, with some additional significant interactions and main effects.

As in Experiment 1, and as predicted, participants were highly likely to follow an iconic interpretation between linear order and temporal order, that is, to choose the first described event to have happened first, but only if both sentences contained eventive descriptions (dark green bars).

However, this tendency was strongly influenced by eventuality type: First reported situations were chosen significantly more often for double-eventive descriptions ($\text{mean}_{\text{event-event}}=0.83$, $\text{SD}=0.38$) than for double-stative descriptions ($\text{mean}_{\text{state-state}}=0.45$, $\text{SD}=0.50$), which appeared as a significant difference between people's response choices depending on the eventuality type of both clauses (see Figure 5.I, $\beta=0.37$, $t=9.60$, $p<0.001$).

Relatedly, for sentences in which eventuality types differed between clauses (olive and teal bars), participants chose the first reported situation first, but only when it was stative (olive bars, $\text{mean}_{\text{state_first}}=0.90$, $\text{SD}=0.31$).

For sentences in which the first situation was eventive (teal bars), participants chose the second situation to have happened first ($\text{mean}_{\text{event_first}}=0.23$, $\text{SD}=0.42$). This observation was statistically supported by a main effect of the eventuality in the first clause ($Df=1$, $\chi^2=28.33$, $p<0.001$), a significant interaction between eventuality type combination and the eventuality in

“What happened first?”

Mean response choices and standard errors (N=130)

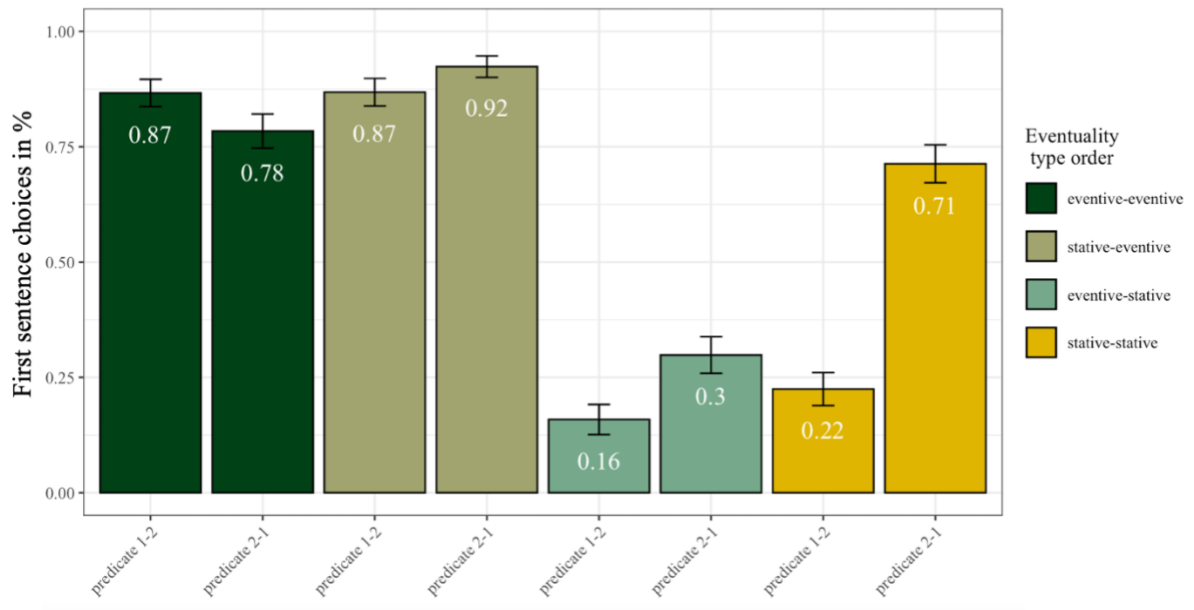


Figure 4 Mean proportions of first sentence choices across eight conditions, error bars represent standard errors. The x axis shows the reported order of the two situations.

the first clause ($Df=1$, $\chi^2=296.98$, $p<0.001$), as well as a significant difference between the stative-eventive and eventive-stative combinations, collapsed across reported orders (i.e., teal vs. olive bars, see Figure 5.II, $\beta=0.67$, $t=20.59$, $p<0.001$).

Within reported orders, when the first reported situation was eventive, participants were less likely to order it first in time when it was followed by a stative description (predicate order 1-2: $\text{mean}_{\text{event-state}}=0.16$, $SD=0.37$; predicate order 2-1: $\text{mean}_{\text{event-state}}=0.30$, $SD=0.46$), than by an eventive description (predicate order 1-2: $\text{mean}_{\text{event-event}}=0.87$, $SD=0.34$, predicate order 2-1: $\text{mean}_{\text{event-event}}=0.78$, $SD=0.41$), which was also confirmed by pairwise comparisons (predicate order 1-2, see Figure 5.III: $\beta=-0.71$, $t=-16.15$, $p<0.001$, predicate order 2-1, see Figure 5.IV: $\beta=0.49$, $t=-8.92$, $p<0.001$).

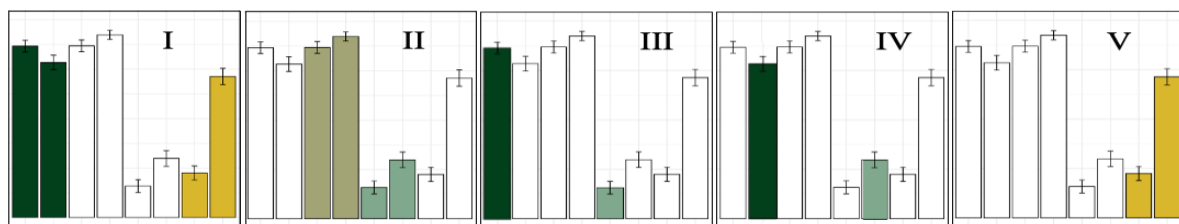


Figure 5 Outline of planned pairwise comparisons, colors of bars reflect compared conditions.

Unlike in Experiment 1, the interaction between reported order and eventuality type in the first clause reached significance ($Df=1$, $\chi^2=13.02$, $p=0.0003$), such that the reported order of the two sentences had a bigger effect on people's response choices when stative situations were reported first (predicate order 1-2: $\text{mean}_{\text{stative_first}}=0.54$, $SD=0.50$ vs. predicate order 2-1: $\text{mean}_{\text{stative_first}}=0.82$, $SD=0.38$) than when eventive situations were reported first (predicate order 1-2: $\text{mean}_{\text{eventive_first}}=0.52$, $SD=0.50$ vs. predicate order 2-1: $\text{mean}_{\text{eventive_first}}=0.53$, $SD=0.50$). That is, when states preceded events in reported order, participants were more likely to simply rely on this order of situations. Furthermore, in contrast to Experiment 1, we found a significant three-way interaction between the three predictors ($Df=1$, $\chi^2=19.80$, $p<0.001$).

As in Experiment 1, double-state descriptions differed in terms of their reported order: In Experiment 2, participants followed the reported order more frequently when the first reported situation was the relatively longer state (in the example above, being on holiday; $\text{mean}_{\text{predicate order 2-1}}=0.71$, $SD=0.45$) than when the first reported situation was the relatively shorter state (in the example above, being on a train; $\text{mean}_{\text{predicate order 1-2}}=0.22$, $SD=0.42$). This was statistically reflected in a main effect of reported order ($Df=1$, $\chi^2=21.84$, $p<0.001$) as well as in a significant difference between the double-stative conditions (see Figure 5.V, $\beta=-0.49$, $t=-9.02$, $p<0.001$).

We had expected that differences in the duration of a state might play a role when inferring temporal order. Therefore, we preregistered selected additional analyses for the double-stative conditions to probe selected items, and performed pairwise comparisons only for these items, where we expected durational differences to be more pronounced: First, we replicated the findings from Experiment 1, such that participants followed the reported order when marriage was described first ($\text{mean}_{\text{marriage_first}}=0.59$, $SD=0.50$), but not when pregnancy was described first ($\text{mean}_{\text{pregnancy_first}}=0.08$, $SD=0.28$); this was reflected in a significant difference for the double-stative conditions in marriage-pregnancy sentences ($\beta=-0.51$, $t=-3.27$, $p=0.003$): Similarly, a person being on holiday and sitting on a train, reported in this order, led to

“What happened first?”

Mean response choices and standard errors (N=130)

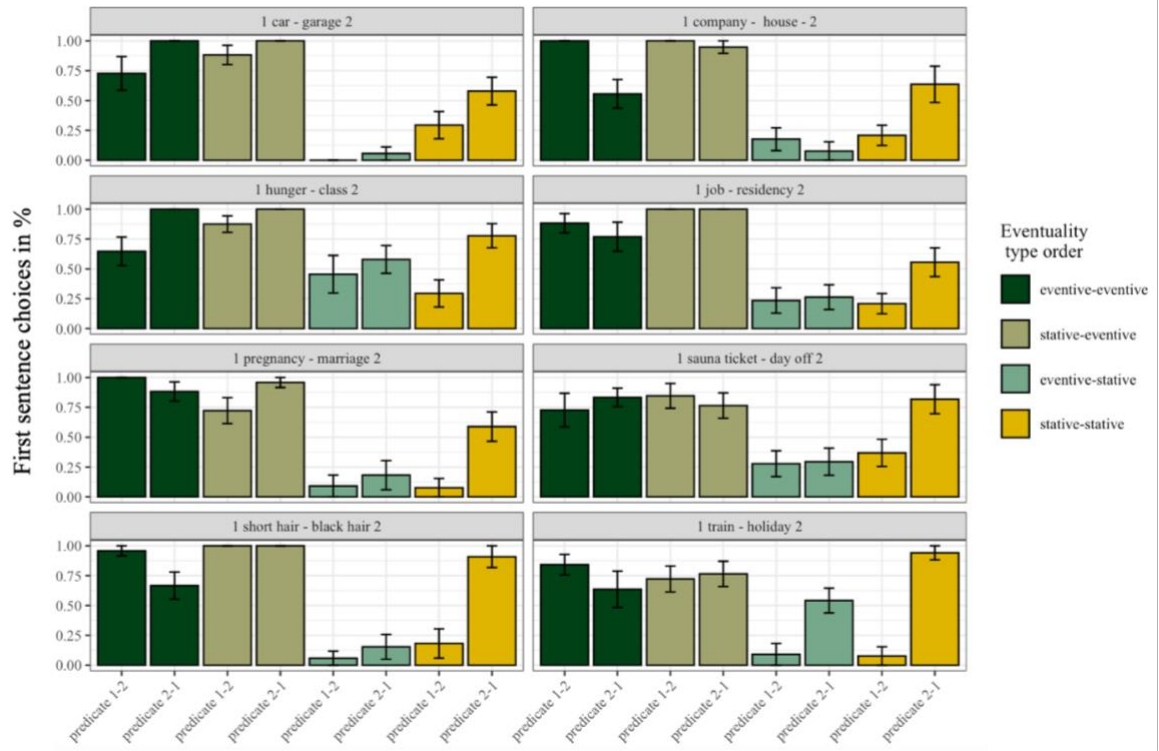


Figure 6 Mean proportions of first sentence choices across eight conditions item-by-item, error bars represent standard errors. The x axis shows the reported order of the two situations.

significantly more first sentence first choices ($\text{mean}_{\text{holiday_first}}=0.94$, $\text{SD}=0.24$) than when sentences reported the train sitting first ($\text{mean}_{\text{train_first}}=0.08$, $\text{SD}=0.28$, $\beta=0.86$, $t=9.09$, $p<0.001$). Visual inspection of the data, broken down by item (Figure 6), further revealed that this duration-driven asymmetry between double-stative descriptions was stable across items: For sentence pairs in which the longer state was reported first (i.e., predicate order 2-1), the first reported situation was numerically selected more often to have happened first than for sentence pairs in which the shorter state was reported first.

As in Experiment 1, the only interaction that did not reach significance was between reported order and eventuality type combination ($Df=1$, $\chi^2=0.50$, $p=0.48$). The main effect of event type combination did also not reach significance ($Df=1$, $\chi^2=2.49$, $p=0.11$).

4.3.3. Discussion of Experiment 2

Experiment 2 served as a conceptual replication of the first study; in addition, it broadened our empirical base by using a within-participants design. We replicated the pattern obtained in Experiment 1. That is, while participants iconically mapped reported order to temporal order in double-eventive sequences, this tendency was completely suppressed as soon as one of the events was encoded as a state: then, states were more likely to be ordered before events. In other words, eventuality type reliably predicted temporal order inferences in discourse comprehension.

Furthermore, the within-subjects design allowed us to account for noise by including items and participants as random intercepts, which increased the power of our statistical analysis. For instance, we found that the predicted main effect of event type in the first clause as well as the three-way interaction between reported order, eventuality in the first clause, and eventuality combination were significant in Experiment 2, but not in Experiment 1. As we have shown in Figure 4, there was limited variability in items that could have contributed to these effects. Therefore, Experiment 2 supports the generalizability of the conclusion that non-linguistic event structure is an important contributor to temporal inferences in discourse.

Experiment 2 also allowed us to touch upon a question that is related to the question of what makes a state a state, and an event an event: The question of temporal duration. We return to this point in Section 4.4.

4.4. General Discussion

In this paper, we asked whether linguistic eventuality type (i.e., stative vs. eventive descriptions) modulates the way in which people understand the temporal order between two situations reported in two independent sentences. More precisely, does a preference for ordering stative descriptions first override the widely reported observation that, in discourse

comprehension, people take the reported order of a narrative to map iconically onto temporal inferences?

In two experiments, we found that temporal inferences are reliably predicted by linguistic eventuality: When participants read two sentences containing eventive descriptions, they followed the reported order of the two sentences and inferred that the first reported situation also happened first in time. In this respect, our data confirm a classic observation in the literature on temporal inference in discourse, according to which the reported order of successive sentences is iconically mapped onto the temporal order of the described situations (Fleischman, 1990; Hinrichs, 1986; Oversteegen, 2005; Webber, 1988). However, when situation descriptions differed in their eventuality type, the reported order did not serve as a cue to temporal order. Instead, participants were more likely to locate stative descriptions before eventive descriptions, regardless of which of the two situations they read first (3-6).

These results suggest a systematic correspondence between linguistic eventuality type and the inferred temporal order between situations, confirming previous findings from other linguistic contexts (see Marx & Wittenberg, 2022, for relative clause constructions). Therefore, our results lend strength to event-structural accounts that link temporal inference in language comprehension to non-linguistic event cognition.

According to such accounts, deriving the temporal structure of a narrative requires that each described situation is temporally anchored by another time or situation. While multiple factors can in principle influence temporal anchoring relations (i.e., causality or explicit temporal marking), event structure – that is, the fact that states are stable and thus cognitively less salient than events (Clewett et al., 2020; Kurby & Zacks, 2008; Zacks et al., 2007) – seems to be a robust determinant for constructing temporal relations in language.

When both situations are encoded with stative descriptions, we found in Experiment 1, that people tended to order marriage before pregnancy, especially when marriage was linguistically

reported second. While this general preference for ordering marriage first may originate from a contingency asymmetry between the two situations in world knowledge, with people perhaps viewing pregnancy as more contingent on marriage than vice versa, the fact that this preference is more pronounced when it is opposite to the reported order may be explained in two ways: First, participants may have paid more attention to the temporal order of the two situations because of the markedness of the reported order. That is, people's expectation that marriage should occur first may have conflicted with their expectation that the reported order should reflect the order in which the situations occurred in the world (Behaghel, 1932; Tai, 1983).

However, there is also another possible explanation for the temporal inferences found in stative-stative conditions, which is independent of people's expectations about the contingency relations between pregnancy and marriage. Experiment 2 allowed us to further explore the possibility that this effect is driven by other dimensions of event structure: the estimated duration of states. We tentatively predicted that longer states should precede shorter states: If duration is a relevant distinction in event structure, shorter states should tend to pattern with events. This was the case in our data but has also been shown in linguistic tests.

Linguistic tests of stativity (Dowty, 1979; Katz, 2003; Lakoff, 1966) tend to vary in reliability. In general, they are quite robust for long-lived individual-level predicates (e.g., *to be French*), but short-lived stage-level predicates (e.g., *to be cooperative*) often resist. For instance, states are claimed to be incompatible with progressive tense, which is true for individual-level predicates (e.g., *?I am being French*), but not for stage-level predicates (e.g., *I am being cooperative*), which pattern with events (e.g., *I am singing*). Second, imperative morphology is marked with stative meaning (e.g., *??Be French!*, but see *Be cooperative!* or *Sing!*). Also, embedding states under *must* should elicit an epistemic interpretation, which we find in individual-level predicates (e.g. *You must be French*); however, stage-level predicates receive

a deontic, or at least ambiguous, interpretation (e.g., *You must be cooperative*) similar to events (e.g., *You must sing*).

Now, while there are manifold differences between individual-level and stage-level predicates (Carlson, 1977; Kratzer, 1995), for instance, that one has limited agency over being French, but full agency over being cooperative or singing, one very salient recurring difference between these different stative predicates is relevant to our project: time. Being French in most cases lasts from crib to grave, whereas being cooperative rarely does.

Therefore, we hypothesized that the estimated duration of a state could influence temporal order inferences in double-stative conditions; and we have argued that this explains the inferred order in these conditions in Experiment 1. Indeed, in Experiment 2, we found more generally that longer states tended to be ordered before shorter states. While a more nuanced exploration of this effect is beyond the scope of this paper and it should be confirmed by future work, we would argue that these findings further support a theory of language processing in which non-linguistic event knowledge reliably and predictably drives temporal inferences in discourse comprehension. In this sense, our study also touches on the more fundamental question of what makes a state a state, and an event an event in language and cognition.

Overall, these experiments are a first systematic attempt to empirically investigate the importance of linguistic eventuality in how people understand the temporal order of unrelated discourse entities. Our results suggest that whether we talk – and therefore likely think – about situations as static or dynamic is more relevant for our readers' inferences about what happened first than the linear order in which the situations are reported.

5. Temporal order inference across tenses – Evidence for a Figure-Ground bias

5.1. Introduction

The way in which comprehenders arrange situations in time is a crucial aspect of language comprehension, shaping the mental representation of any described complex scene (Radvansky et al., 1998; Zwaan et al., 1995): For example, hearing that the monkey was sitting in the tree while the tourist took a picture of it describes a situation in which the tree is part of the photograph. Conversely, hearing that the tourist took a picture of the monkey, which then went to sit in the tree, leads to a different mental construal: one in which the tree is absent from the picture.

Different temporal orders can, of course, be explicitly encoded through linguistic devices such as temporal connectives (e.g., *before*, *after*), temporal adverbials (e.g., *then*, *previously*), or contrasting verb tenses (e.g., *danced* – *will dance*). However, the availability and use of such devices varies across languages (Bohnenmeyer, 2009; Bybee & Dahl, 1989; Comrie, 1985; Haspelmath, Dryer, Gil, & Comrie, 2005; Lin, 2003, 2006; Tonhauser, 2015). Even in languages with mandatory explicit markers, such as the English tense system, temporal relations between events are often left ambiguous. For example, in a sentence like *The tourist took a picture of the monkey that sat in the tree*, it is unclear whether the monkey sat in the tree before, during, or after the picture was taken. This raises the question: How do comprehenders infer temporal order when explicit markers are absent or ambiguous?

In previous work, we demonstrated that, in temporally ambiguous descriptions, people systematically rely on the dynamic properties of predicates to infer the temporal relations between situations (Marx & Wittenberg, 2022, 2024, 2025; Marx, Iwan & Wittenberg, 2025): static states are interpreted to happen before dynamic events (the ‘states-before-events

inference’). Applying this state-before-event inference to the example above, comprehenders are likely to construe the monkey’s state of sitting in the tree as the temporal background to the event of the tourist taking a picture.

Crucially, our prior studies exclusively examined past tense. However, it remains unclear whether the state-before-event inference is specific to the simple past, or whether this pattern holds across tenses, specifically future tense, which intuitively relates to the utterance time as the past does – just mirrored into the opposite direction. In this paper, we therefore shift our focus to the future tense, to explore the generalizability of the states-before-events inference. As before, we use relative clause constructions, which link two propositions together without signaling a systematic causal, temporal, or conditional relationship (Thomson & Martinet, 1985).

Future contexts furthermore offer a way to evaluate two competing theoretical explanations for the states-before-events inference, examining the role of tense meaning more generally: a temporal anchoring account, which posits that temporal construal is mediated via tense meaning, and a conceptual Figure-Ground principle that operates independently of tense.

According to the temporal anchoring view (Carroll & von Stutterheim, 2010; Klein, 1994, 2009; von Stutterheim et al., 2003), people construe temporal relations from language by linking situations to one another, based on the salience differences between static states and dynamic events. Therefore, less salient states should be systematically anchored by more salient events, with tense determining the temporal relationship. This relationship should be one of anteriority in the case of the past tense, and arguably one of posteriority in the case of the future tense. Consequently, while states should be interpreted as preceding events in past-under-past contexts, the temporal order should reverse in future-under-future contexts: That is, states should be inferred to follow events in future tense.

This hypothesis, however, assumes that the future functions primarily as a tense (Demirdache & Uribe-Etxebarria, 1996; Giorgi & Pianesi, 1997; Kissine, 2008; Klein, 1994; Prior, 1967) encoding a futurative relationship between a described situation and the utterance time – an assumption that has been rebutted by alternative accounts conceptualizing the simple future as a marker of modality, similar to modal auxiliaries (Enç, 1996; Huddleston & Pullum, 2005; Palmer, 1987).

In contrast, a Figure-Ground explanation does not rely on tense at all. Instead, such accounts predict that states serve as the temporal background for events, based on the salience differences that are associated with these dynamic properties alone: Static states should consistently serve as temporal backgrounds for dynamic events in both past and future contexts.

In two experiments presented in this paper, we test these predictions, studying the cognitive mechanisms driving temporal construal in language comprehension in the absence of explicit temporal order cues. Additionally, we aim to contribute to the broader theoretical debate regarding the nature of the future – whether it functions primarily as a tense marker or as a modal marker – and investigate the role of tense meaning in temporal construal more broadly. Our paper is structured as follows: We first review the tense-mediated temporal anchoring account in greater depth (Section 6.1.1), applying it both to the past and the future tense. We then discuss the conceptualization of the future as a marker of modality (Section 6.1.2), which challenges the applicability of tense anchoring in future contexts. Following this, we turn to alternative strategies for temporal construal in future contexts (Section 6.1.3), focusing on a Figure-Ground principle that likewise accounts for previous findings in the simple past. Finally, we outline the predictions that form the basis of our study (Section 6.1.4).

5.1.1. Event order through tense-mediated temporal anchoring

5.1.1.1. The case of simple past

How do comprehenders derive temporal order from temporally ambiguous linguistic descriptions? One account suggests a mechanism that links temporal construal closely to event structure and the meaning of tense: tense-mediated temporal anchoring. According to this view, tensed predicates encode complex temporal structures comprising three key parameters (Carroll & von Stutterheim, 2010; Klein, 1994, 2009; von Stutterheim et al., 2003): The first parameter, the situation time, refers to the entire time span of the situation described by the predicate. Thus, situation times always correspond to the non-finite predicate (e.g., *taking a picture*, *sitting*) and rely on general event knowledge about duration and structure (e.g. taking a picture is punctual and culminates at the time the picture is shot, whereas sitting is durative and static over time). Second, a topic time refers to a specific time, about which the speaker makes an assertion, relative to the situation time. The relation between topic time and situation time is conveyed through aspect: For the simple past, which carries perfective aspect, the topic time corresponds to a part of the situation time, including its culmination point (i.e., for achievements/accomplishments) or post-state (i.e., for states/activities). However, topic time is always linked to the time the assertion is uttered, a relation established through tense: For the past tense, the topic time (and consequently parts of the situation time, including its culmination or post-state) is located at a time before the utterance time. For example, in *The tourist took a picture of the monkey*, the topic time denotes a period before the utterance time during which the event of taking the picture was completed.

Third, a temporal anchor makes reference to another temporal interval, linking it to the topic time. In isolation, like in a sentence such as *The tourist took a picture of the monkey*, the temporal anchor refers to the utterance time. In more complex event descriptions, however, the temporal anchor picks out another time interval mentioned in the description: In *The tourist*

took a picture of the monkey that sat in the tree, either the time of the main clause situation (i.e., the tourist taking a picture) could serve as a temporal anchor for the relative clause situation (i.e., the monkey sitting in the tree) or the relationship could be reversed.

On a tense-anchoring account, two assumptions are important: First, temporal anchors are not chosen at random, but always correspond to the most salient temporal interval available. And given that the distinction between dynamic events and static states comes with a difference in saliency such that the former are cognitively more salient than the latter (Clewett et al., 2020; Kurby & Zacks, 2008; Zacks et al., 2007), eventive descriptions like *taking a picture*, which correspond to cognitively salient changes in the world, should serve as temporal anchors for static descriptions like *sitting*, which correspond to less salient, non-changing states.

Second, because the temporal relation between situation time and topic time is tense mediated in this account, so should be the relation between the situation time and the temporal anchor: For simple past, the topic time spans parts of the situation time, including its culmination point or the post-state, which also imposes a logical constraint on its starting point: In *The tourist took a picture of the monkey that sat in the tree*, the anchored situation (i.e. monkey sitting in

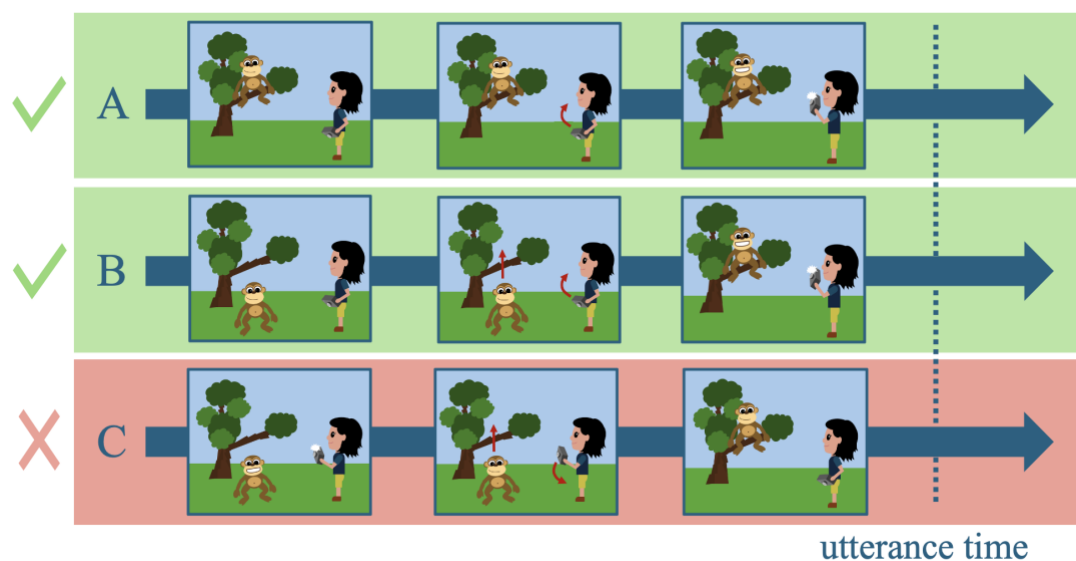


Figure 1 Possible temporal orders between two situations described by past-under-past relative clause constructions, according to a tense-mediated temporal anchoring account

the tree) might begin before (Figure 1A), simultaneously (Figure 1B) but not after the time of the temporal anchor (i.e., tourist taking a picture, Figure 1C).

5.1.1.2. Temporal anchoring and the future tense

Throughout the decades, future tense, and *will* in particular, has been analyzed by syntacticians and semanticists in (mainly) two ways: as a tense marker (Demirdache & Uribe-Etxebarria, 1996; Dowty, 1982; Giorgi & Pianesi, 1997; Kissine, 2008; Klein, 1994; Prior, 1967), or as a modal marker (Enç, 1996; Huddleston & Pullum, 2005; Palmer, 1987). The former regards the future as a tense, the mirror image of the past: According to this view, the future, like the past, denotes a temporal relation between the time of utterance and the topic time. This relation is represented in (1), for both tenses: the past locates the event in a time that precedes the utterance time (1a), whereas in the future this order is reversed (1b).

- (1) a. *The tourist took a picture of the monkey.*



- b. *The tourist will take a picture of the monkey.*



Deriving predictions about temporal order for future-under-future constructions is exceedingly difficult, since the literature either focuses on simple clauses (Demirdache & Uribe-Etxebarria, 1996; Dowty, 1982; Giorgi & Pianesi, 1997; Kissine, 2008; Klein, 1994; Ogihara, 1996; Prior, 1967) or complex utterances in which tenses are mixed (e.g. future-under-past, present-under-future, e.g., Abusch, 1998; Higginbotham, 2006; Ogihara, 1996). Therefore, we mainly base

our predictions on logical extensions of the temporal anchoring account, with a caveat that more precise theoretical accounts are rare.²⁵

If we apply the temporal anchoring approach to future-under-future constructions, the parameters of situation time, topic time, and temporal anchors as well as their interrelations are different from those in the simple past: For a sentence such as *The tourist will take a picture of the monkey that will sit in the tree*, the situation time corresponds to the non-finite part of the predicate (i.e., *take a picture, sit*), denoting both the duration and the internal structure of the situation described (i.e., dynamic, punctual *taking a picture* vs. static, durative *sitting*). This is similar to how situation time functions in the simple past.

The topic time, however, should span a different interval in future tense: With boundedness inherent to the simple future (as opposed to the future progressive, e.g. *will be sitting*), the topic time encompasses a time of the situation, including its culmination point or post-state. Crucially, the future tense locates the entire situation after the utterance time, including the situation's starting point. Put simply, while the simple past restricts the culmination point to a time before the utterance, the future tense topic time signals that the situation has not yet started, but will unfold and eventually be completed at a later time, thus including both the starting and the ending points of the situation.

This also has consequences for the temporal relationship between situation time and temporal anchor: In the example above, the temporal anchor is identified through saliency differences like in the simple past: The eventive description (i.e., *the tourist will take a picture*) establishes a salient reference time that anchors the stative description (i.e., *the monkey will sit in the tree*). However, since the topic time of the anchored situation now includes the entire situation time

²⁵ To our knowledge, one of the few analyses of future-under-future sentences is Dowty's (1982), whose conclusions align with the predictions we derive from the existing temporal anchoring theories. Dowty argues that a main clause future, functioning as a Priorian tense operator, shifts the reference time of any embedded tense. For example, in a sentence like *The tourist will take a picture of the monkey that will sit in the tree*, the relative clause takes the reference time of the main clause, situating it further into the future (see also Gennari, 2003).

(i.e., including starting and ending point), the anchored situation (i.e., *monkey sitting in the tree*) cannot logically begin before the temporal anchor (i.e., *tourist taking a picture*, Figure 2A). Instead, it should be interpreted as co-occurring with (Figure 2B) or following the anchor event (Figure 2C).

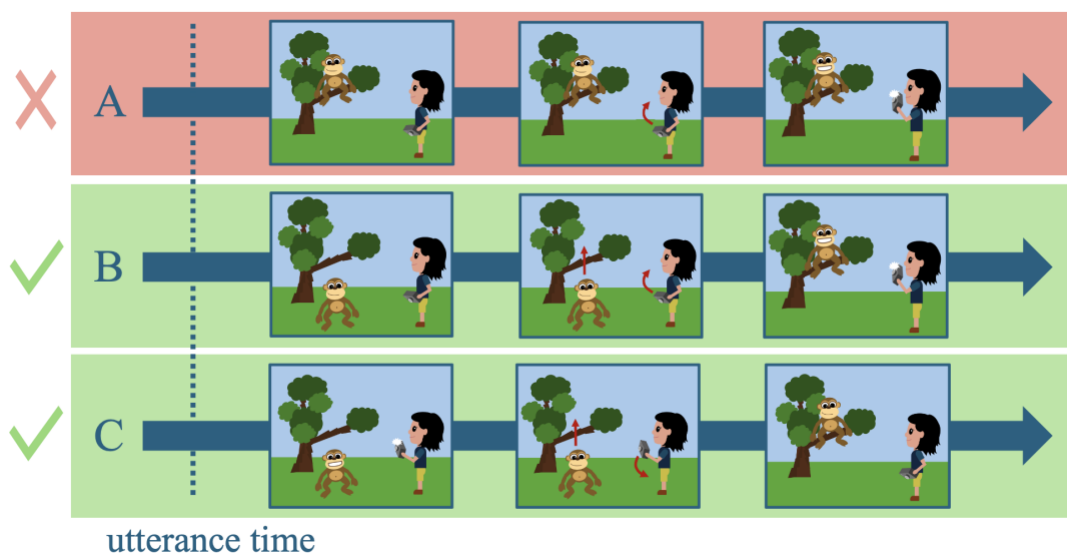


Figure 2 Possible temporal orders between two situations described by future-under-future relative clause constructions, derived from a tense-mediated temporal anchoring perspective

5.1.2. Future as a modal, not a tense

In addition to the future-as-a-tense view, a second approach takes the future to be a modal (Enç, 1996; Huddleston & Pullum, 2005; Palmer, 1987), emphasizing its non-deictic nature. The idea is that, unlike the past, the future does not refer to a time, but only to potential occurrences of an event. Modal accounts of the future emphasize an asymmetry between future and past in terms of truth conditions, which are the standard test for well-formedness in formal semantics: Whereas it is easy to establish if a statement is true or false in the past, the same proof becomes problematic when the statement is about an event “in the future” (Brogaard, 2008; Heck, 2006). For instance, in evaluating the truth of the future-tense sentence in (2), the evaluation point lies in the future as well: Only after tomorrow, we will know whether this sentence was true.

- (2) *We will submit this paper tomorrow.* (adapted from example (5), De Brabanter, Kissine, & Sharifzadeh, 2014)

Under a modal view of the future, the abovementioned asymmetry follows naturally if the function of the future in a sentence is not to locate an event in a time that has not occurred yet, but rather if it serves a variety of (a)temporal purposes. In cases such as (3-5), adapted from Palmer (1974), Lapaire and Rotgé (1998) and Downing and Locke (2006), it is argued that *will* does not refer to a “future time”, but rather to some other meaning: dispositional (3); epistemic (4); habitual (5); or volitional (6).

- (3) *Oil will float on water.*
(4) *John will be at the opera by now.*
(5) *In summer, Michael will(/would) always wear shorts.*
(6) *The child will eat the cake (if he gets the chance).*

The dispositional *will* in (3) states what is characteristic of oil, such that, according to its internal properties, and independently of time, it stands in a particular relationship to water; the epistemic *will* in (4) refers to how likely it is, according to what I know, that John is at the opera at that time. Again, it says nothing about John in a time following the present; the habitual (also dispositional) *will* in (5) takes the duration of the summer season as its temporal reference, and states that one can see Michael wearing shorts, or that Michael is a short-user, in that interval of the year; finally, the volitional *will* in (6) refers to the child’s intention or will (notice the word) to do something.

For those advocating a modal account of the future, the question becomes what kind of modality is at play: whether it is an instance of root modality, related to ability and volition, like *can* or *want* (Copley, 2002; Del Prete, 2011), or it is an instance of epistemic modality, along the lines of *must* (Giannakidou & Mari, 2012; Palmer, 1987). Independently of the specific modal flavor

of the future, the modal approach predicts underspecification in temporal construals in the future, as compared to the past.²⁶

Assuming that the simple future is not a true tense but a modal expression, this suggests that temporal order in future-tense contexts may rely more on alternative pragmatic cues than on tense-mediated temporal anchoring. The question is: What are alternative ways of deriving temporal order from language? Do dynamicity differences play a role in ordering events in time even when temporal reference is to the future rather than the past?

In the following chapter, we will outline three possible alternatives through which temporal order could be derived. One of the proposed mechanisms can also account for the patterns observed in past-under-past constructions by invoking a simpler mechanism based solely on event dynamicity: a Figure-Ground principle.

5.1.3. Alternative accounts to derive temporal order from language

5.1.3.1. Iconicity

One influential factor in temporal construal is iconicity – a principle suggesting that the order in which events are mentioned in language also reflects the order in which they occur in the world. This principle is well-supported across studies in both discourse comprehension (Fleischman, 1990; Hopper, 1979; Horn, 2022b; Kaiser, 2019; Kamp & Rohrer, 1983; Lascarides & Asher, 1991, 1993; Ohtsuka & Brewer, 1992; Schmerling, 1975; Tai, 1983; Xu & Kwok, 2019), and complex sentence comprehension (Blything & Cain, 2016; Blything, Davies, & Cain, 2015; Clark & Clark, 1968; Clark, 1971; de Ruiter, Theakston, Brandt, & Lieven, 2018;

²⁶ For our purposes, the term ‘underspecification’ does not imply a strong commitment to a specific formal semantic theory. Nevertheless, we point out that this view of temporal order, based on the future as a modal, is highly compatible with possible world semantics (Heim & Kratzer, 1998; Partee, 1973). What matters here is that the ordering relations between two situations are underspecified in the future under a modal account.

Münste, Schiltz, & Kutas, 1998; Ohtsuka & Brewer, 1992; Trosborg, 1982; Van Der Meer, Beyer, Heinze, & Badel, 2002; Xu & Kwok, 2019; Nieuwland, 2015; see also Dry, 1982).

In past-under-past relative clauses, iconicity seems to play a minor role. For instance, in Marx & Wittenberg (2024), participants read two sentences (e.g., *Mary got/was pregnant. She got/was married to John.*) which were crossed in dynamicity (i.e., eventive vs. stative) and in their linear order (i.e., pregnancy vs. marriage). Participants indicated which of the two situations had happened first. What determined temporal construal in this forced-choice task was dynamicity: States were consistently ordered before events. The same pattern was also found in relative clauses, using an act-out task (Marx, Iwan & Wittenberg, 2025) and sentence-video judgments (Marx & Wittenberg, 2022, 2025). Effects of iconicity were found, but they were second order, only relating to sentences that encoded two events. Furthermore, a systematic effect of iconicity in this condition across tasks was limited only to English (see Marx, Iwan & Wittenberg, 2025).

However, iconicity may play a larger role under certain conditions: If temporality is underspecified in future-under-future relative clauses, it may be underspecified overall, not just in terms of the temporal relation between the two described events. This might lead people to rely more on iconicity as a strategy for temporal construal. In this case, in a sentence in which the main clause precedes the relative clause, the event described in the main clause might be understood to precede the event described in the relative clause, irrespective of dynamicity differences between clauses.

5.1.3.2. Discourse-pragmatic function of relative clauses

Another factor in temporal order construal may be the pragmatic function of certain linguistic constructions which serve to highlight or contrast new information with what is already known or part of the background (Chafe, 1970; 1977; Givón, 1983; Gundel, 1988; Hopper &

Thompson, 1980; Kratzer, 2004; Krifka, 2008; Potts, 2005; Reinhart, 1981; Rizzi, 1997; Talmy, 1975; Tomlin, 1985). Relative clauses, in particular, have been described as syntactic tools that organize information pragmatically. Semantically, relative clauses either narrow the interpretative range of a noun phrase (i.e., restrictive relative clauses) or provide supplementary information about the referent (i.e., non-restrictive relative clauses, see Stockwell, Schacter, & Partee, 1973; Fox & Thompson, 1990; Lambrecht, 1988; Quirk, Greenbaum, Leech, & Svartvik, 1985; Thomson & Martinet, 1985; Levinson, 1983; Cinque, 2020).

Pragmatically, however, relative clauses often convey presupposed or given information, that is, details that the speaker assumes the listener is already familiar with or can readily infer. Especially within narratives, relative clauses have been associated with establishing the background of a story (c.f. Chierchia & McConnell-Ginet, 1990; Depraetere, 1996; Ehrlich, 1987; Hopper, 1979; Potts, 2005; Reinhart, 1984; see also Ramsay, 1987; Talmy, 1975 for other subordinate clauses). By contrast, main clauses advance the narrative by detailing the temporal progression of events. Although the distinction between background and foreground information is not inherently temporal, it has been argued to bear a temporal dimension (Hopper, 1979; Hopper & Thompson, 1980; Polanyi Bowditch, 1976; Wårvik, 2004): Backgrounds need to be established before the foreground events occur.

Similar to iconicity, previous work demonstrates that comprehenders do not consistently adhere to the discourse pragmatic function of relative clauses – at least in past-under-past tense sentences that encoded both a dynamic event and a static state (Marx & Wittenberg, 2022, 2025; Marx, Iwan & Wittenberg 2025). In such cases, dynamicity differences primarily guided temporal construal, with comprehenders showing little preference for backgrounding relative clause information. However, when dynamicity differences were absent, the pragmatic function of relative clauses did influence temporal interpretations, though this effect varied across languages, with English speakers displaying a weaker pattern (Marx, Iwan & Wittenberg 2025).

It is therefore plausible that the discourse-pragmatic function of relative clauses might exert a stronger influence in future-under-future relative clauses, if the future primarily conveys modal meaning: In this case, the backgrounding nature of relative clauses could bias comprehenders toward interpreting the information conveyed by these clauses as temporally preceding the events described in the main clauses.

5.1.3.3. Figure-Ground principle

While the two alternative principles discussed so far – iconicity and the pragmatic function of relative clauses – offer plausible ways of interpreting relative clause constructions in the future, neither can fully account for previous findings in the past tense (Marx & Wittenberg, 2022, 2024, 2025; Marx, Iwan & Wittenberg 2025).

However, there is a third possible explanation for the state-before-event inference: a general mechanism that structures representations into Figure and Ground in (nonlinguistic) cognition. This account would derive the state-before-event inference from dynamicity, independent of syntactic construction or tense.

Drawing on principles from Gestalt psychology, the Figure-Ground distinction originally described in perception highlights how a focal object (i.e., the Figure) at the center of attention is defined relative to a less focal or stable background (i.e., the Ground, Koffka, 1935). Conceptually applied to language, this distinction has been used to explain how syntax encodes spatial, temporal, and conceptual relationships (Gleitman, Gleitman, Miller, & Ostrin, 1996; Levinson, 1996; Talmy, 1975, 2000; Zlatev, 2010): Specifically, the Figure is typically conceptualized as a "(conceptually) movable" entity that is defined relative to a more stable Ground entity (Gleitman et al., 1996, p. 358): In a sentence such as “The bike stood next to the house”, the bike, corresponding to the Figure, is more mobile, whereas the house, corresponding to the Ground, is both stable and large.

Gleitman et al. (1996) go on to say that the converse sentence “The house stood next to the bike” is odd precisely because some objects, in relation to others, make better Grounds, and others, better Figures. For instance, houses are large, and bikes are small. However, size alone does not determine what constitutes a better Figure or Ground, given that bikes are more prototypical Figures compared to other, smaller objects as well: “The bike stood next to the fire hydrant” sounds more natural than “The fire hydrant stood next to the bike”, despite the fire hydrant being smaller than the bike. This is, as we argue, and in line with Gleitman et al. (1996), because Grounds are ideally static, or at least relatively more static than Figures.

The conceptual structuring of scenes into agile Figure and static Ground can easily be extended into event cognition as well: States – being static and stable – are more prototypical Grounds, while events, characterized by motion or change, are more prototypical Figures. Thus, comprehenders may use a contrast in dynamicity to infer temporal relationships, independent of tense meaning. Interestingly, this view also aligns with conceptions in the discourse comprehension literature, which emphasize the role of dynamic properties in arranging information into situational foreground and background of a narrative (Hopper, 1979; Reinhart, 1984; Wårvik, 2004). Although these accounts use different terminology, they share the intuition that event dynamicity contributes to the organization of temporal and conceptual relations during language comprehension.²⁷

A Figure-Ground account not only provides a coherent mechanism for temporal construal in future-under-future contexts, but also explains all previous findings for the past tense: In past-under-past relative clauses, people inferred that static states must precede dynamic events because states serve as a stable Ground for Figure events in the unfolding narrative. In future-

²⁷ See also Sun, Firestone, & Hafri (2023), who demonstrate that the influence of the Figure-Ground distinction on temporality extends beyond the event domain to relations between event participants. In a picture-recognition task, participants more quickly matched linguistic descriptions to visual scenes when the Ground entity (e.g., a table) was displayed first, before the Figure entity (e.g., a vase on top of it), consistent with a Ground's stable, background-like role.

under-future constructions, this mechanism should remain unchanged: Dynamic events should likewise be construed as Figures against the background of static states, so that they are reliably arranged after states in time. In both cases, the cognitive salience of events relative to states should drive the inference of temporal order, regardless of whether the temporal relation is located in the past or shifted into the future.

5.1.4. The current studies and predictions

The current studies were designed to test how wide the reach of the state-before-event inference is; and at the same time, as we argued above, our results should contribute to the debate on whether to conceive of the future as a tense, or as a modal.

The first, temporal anchoring account states that the state-before-event inference is derived from tense, based on the more salient event being an anchor for the less salient state. Therefore, it should predict that states precede events in the past tense, and states follow events in the future tense.

However, if the future tense is better thought of as a modal, we may expect effects driven by iconicity, whereby the first mentioned situation is always ordered first; effects driven by discourse-pragmatics in which relative clauses should be ordered first regardless of dynamicity; or, finally, effects driven by a Figure-Ground strategy, in which static states should precede dynamic events, regardless of order or syntax.

We conducted two preregistered experiments that used relative clause constructions to test the predictions derived from these two theoretical explanations as well as the formalizations of future tense in itself. Experiment 1 (Section 6.2) served as a conceptual replication of prior findings on past-under-past relative clauses (Marx & Wittenberg, 2022, 2025; Marx, Iwan & Wittenberg 2025), employing an explicit temporal order judgment task (Marx & Wittenberg, 2024) to directly assess participants' preferences of temporal order. This paradigm was chosen

for two reasons: First, to ensure comparability with previous research, and second, due to its applicability in both past and future tense. In Experiment 2 (Section 6.3), we extended the design to future-under-future relative clauses to investigate the influence of tense meaning on temporal order construal. The results of these two experiments were compared in order to distinguish between the predictions outlined above (Section 6.3.3).

5.2. Experiment 1: Past tense

5.2.1. Participants

48 native English speakers were recruited through testable.com to participate in the experiment, which lasted approximately 10 minutes. We included only complete datasets in the statistical analysis; furthermore, datasets needed to meet our pre-registered attention criterion of at least 50% accuracy on filler trials. Based on these criteria, three participants were excluded, resulting in a final sample of 45 datasets for analysis.

5.2.2. Materials

We constructed 16 critical sentence quadruplets similar to Marx & Wittenberg (2022, 2025) and Marx, Iwan & Wittenberg (2025). All our stimuli were active, transitive and intransitive relative clause constructions with both predicates in the main and relative clause in the simple past.²⁸ The event types of main clause (MC) and the relative clause (RC) were systematically manipulated throughout the four sentences: In (7), the main clause described a dynamic event while the relative clause described a static state; and in (8), the main clause described a state while the relative clause described a dynamic event. To investigate the role of alternative strategies of temporal construal in the absence of dynamicity differences in more detail, we

²⁸ We exclusively used restrictive relative clauses, however, for a comparison with non-restrictive relative clauses, see Marx & Wittenberg (2025). Overall, the same pattern of results emerged for the two relative clause types, suggesting that this distinction does not critically impact temporal construal across these sentence constructions.

included sentences like (9), where both the main and relative clause encoded stative descriptions, and sentences like (10), where both clauses encoded eventive descriptions.

- (7) eventive MC – stative RC

*The tourist [**took a picture**]_{EVENT} of the monkey that [**sat**]_{STATE} in the tree.*

- (8) stative MC – eventive RC

*The tourist [**stood**]_{STATE} next to the monkey that [**swung**]_{EVENT} in the tree.*

- (9) eventive MC – eventive RC

*The tourist [**took a picture**]_{EVENT} of the monkey that [**swung**]_{EVENT} in the tree.*

- (10) stative MC – stative RC

*The tourist [**stood**]_{STATE} next to the monkey that [**sat**]_{STATE} in the tree.*

In addition to the critical sentences, we created 24 complex filler sentences that explicitly encoded temporal order using various linguistic means: Sentences contained either temporal connectives (*before, after*), temporal adverbs (*and then, previously*), tense contrasts between the main and the relative clause (e.g., simple past vs. past perfect), or a combination of two of these strategies. In half of the filler sentences, the linear order of the sentence matched the encoded temporal order. In the other half, the temporal order was reversed, creating a mismatch between the temporal and linear order. All critical linguistic materials can be found at <https://osf.io/sfdx9>.

5.2.3. Procedure

In each trial, participants were presented with complex sentences such as those in examples (7-10) and were asked to make a temporal order judgment. Their task was to select one of two labeled buttons (e.g., *picture taking, standing*) to indicate which of the two described situations occurred first. The task prompt was “*What happened first?*” and the presentation side of the labels was randomized across trials to control for any potential side biases.

Participants completed a total of 41 trials: At the beginning of the experiment, participants completed a single practice trial to familiarize them with the procedure. Practice was followed by 16 critical trials (comprising 2 types of events in the main clause and 2 types in the relative clause, with 4 trials per condition) and 24 filler trials, all presented in a randomized order. To counterbalance the conditions across subjects, participants were randomly assigned to one of four lists using a Latin-square design. Experiment 1 was programmed using PsychoPy and conducted online via Pavlovia.org.

5.2.4. Statistical analysis

We fitted a logistic mixed-effects regression model (i.e., *glmer* from the lme4 package, Bates, Mächler, Bolker, & Walker, 2015) in the R statistics environment (R Core Team, 2014), including event type of the main clause (stative vs. eventive) and event type combination (mixed: eventive-stative/stative-eventive vs. same: eventive-eventive/stative-stative) as sum contrast-coded fixed effects, and participants and items as random intercepts to account for variability associated with these factors. As dependent variable, we coded whether participants chose the main clause to have happened first in time (main clause choices). We chose a minimal intercept-only structure with no random slopes to ensure model convergence and performed likelihood ratio tests to compare the full model with reduced models, excluding one of the predictors or the interaction at a time.

Furthermore, we conducted planned pairwise comparisons to further explore the interaction effects identified in the linear mixed-effects model. A preregistration of both experiments is available at <https://osf.io/sfdx9>.

5.2.5. Results

Participants' temporal judgments are displayed in Figure 3. Overall, we observed a general tendency for participants to order the relative clause first across conditions ($M_{\text{OVERALL}}=0.15$, $SD=0.36$). However, the strength of this tendency varied depending on the dynamic properties of the main and the relative clauses: For one thing, participants were more likely to judge the main clause as having happened first when both clauses encoded events ($M_{\text{EVENT-EVENT}}=0.18$, $SD=0.38$, light green third boxplot). More importantly, this pattern also held – and was even stronger – when the main clause encoded a state and the relative clause an event ($M_{\text{STATE-EVENT}}=0.29$, $SD=0.46$, dark green second boxplot).

The statistical analysis revealed no significant main effects of either event type of the main clause ($Df=1$, $\chi^2=3.07$, $p=0.08$) or event type combination ($Df=1$, $\chi^2=2.53$, $p=0.11$), but the interaction between both predictor variables turned out significant ($Df=1$, $\chi^2=41.76$, $p<0.001$): Stative main clauses were more often judged to have happened first when the relative clause

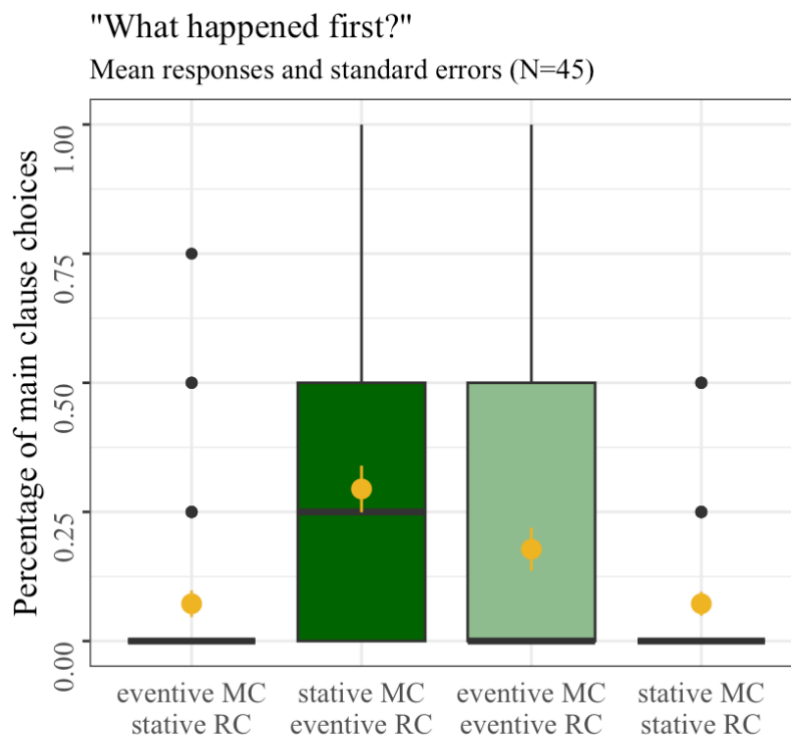


Figure 3 Mean main clause choices in Experiment 1 with boxplot summarizing means per participants in each condition. Yellow dots represent overall means, and yellow error bars the standard error of the mean.

encoded an event, compared to when the relative clause encoded a state as well ($M_{\text{STATE-EVENT}}=0.29$, $SD=0.46$ vs. $M_{\text{STATE-STATE}}=0.07$, $SD=0.26$; $\beta=0.11$, $t=-5.67$, $p<0.001$). At the same time, eventive main clauses were less often ordered first, according to participants' temporal order judgments, when the relative clause encoded a state rather than when both clauses encoded an event ($M_{\text{EVENT-STATE}}=0.07$, $SD=0.26$ vs. $M_{\text{EVENT-EVENT}}=0.18$, $SD=0.38$; $\beta=-0.05$, $t=-3.06$, $p=0.002$).

5.2.6. Discussion of Experiment 1

Our results replicate findings from previous studies using a temporal judgment design (Marx & Wittenberg, 2022, 2025; Marx, Iwan & Wittenberg, 2025): First, and foremost, the dynamic properties of the two situations described significantly influenced temporal construal in relative clause comprehension such that participants were more likely to order states before events.

However, participants also showed a strong tendency to interpret relative clauses as temporal backgrounds in all conditions – a pattern of results we predicted as a possible outcome for future-under-future, but not for past-under-past relative clauses based on the lack of it in previous studies. We suspect that the difference between Experiment 1 and previous findings may stem from the task design, which explicitly asked participants to reason about temporal relations within complex situations, rather than allowing temporal interpretation to emerge as an implicit by-product. This aspect, along with potential implications for experimental design in temporal comprehension, will be discussed in more detail in the General Discussion.

Finally, it should be noted that the relative-clause-first strategy was also weaker when both situations were framed as events – a finding that replicates a previously observed effect of iconicity, particularly for these conditions among English speakers (Marx, Iwan, & Wittenberg, 2025; Marx & Wittenberg, 2024).

In Experiment 2, we investigated the influence of tense on temporal construal using future-under-future relative clause constructions. If tense meaning primarily determines the relationship between a situation and its temporal reference point, we predicted a reversal of the states-before-events effect, particularly in the mixed event-type conditions where the distinction between the temporal anchor and the anchored situation is clear-cut. Alternatively, if participants relied on a general Figure-Ground principle to construe temporal relations, we should find a replication of the pattern observed in Experiment 1. To assess the relative strength of these effects and to further clarify the role of future tense in temporal construal, we conducted a statistical comparison between Experiments 1 and 2.

5.3. Experiment 2: Future tense

5.3.1. Participants

We recruited 48 native English speakers via testable.com, ensuring none had previously participated in Experiment 1. As in Experiment 1, we only included participants with complete datasets and accuracy rates above 50% on the filler trials. Four participants had to be excluded based on these criteria, resulting in 44 submitted datasets for analysis. Participation took approximately 10 minutes.

5.3.2. Materials, procedure and analysis

In Experiment 2, we used the same sentence structures as in Experiment 1, with the only modification being the change of tense to simple future across all critical sentences. The manipulation of event type in both the main and relative clauses remained the same as in Experiment 1 (see 11-14). For the filler trials, the only change made was the use of simple future tense. The same was true for the task prompt (i.e., *"What will happen first?"*).

- (11) eventive MC – stative RC

*The tourist [**will take a picture**]_{EVENT} of the monkey that [**will sit**]_{STATE} in the tree.*

- (12) stative MC – eventive RC

*The tourist [**will stand**]_{STATE} next to the monkey that [**will swing**]_{EVENT} in the tree.*

- (13) eventive MC – eventive RC

*The tourist [**will take a picture**]_{EVENT} of the monkey that [**will swing**]_{EVENT} in the tree.*

- (14) stative MC – stative RC

*The tourist [**will stand**]_{STATE} next to the monkey that [**will sit**]_{STATE} in the tree.*

The experimental procedure and statistical analysis in Experiment 2 were identical to those in Experiment 1 (i.e., mixed effects model with minimal intercept-only structure, planned pairwise comparisons). For the between-experiment comparison, we added tense as an additional predictor variable to the regression model to capture any difference in temporal construal between the simple past and the simple future tense.

5.3.3. Results

The temporal judgments in Experiment 2, as shown in Figure 4, largely mirrored the patterns observed in Experiment 1. Participants were more likely to judge the main clause as happening first in two conditions: when both the main and the relative clause described an event ($M_{\text{EVENT-EVENT}}=0.60$, $SD=0.49$, teal third boxplot), and when the main clause described a state, and the relative clause described an event ($M_{\text{STATE-EVENT}}=0.65$, $SD=0.48$, slate second boxplot). This pattern was statistically supported by a significant interaction between the main clause event type and event type combination ($Df=1$, $\chi^2=45.86$, $p<0.001$), with pairwise comparisons revealing that: (a) stative main clauses were judged to occur first more often when combined with an eventive relative clause compared to a stative relative clause ($M_{\text{STATE-EVENT}}=0.65$, $SD=0.48$ vs. $M_{\text{STATE-STATE}}=0.44$, $SD=0.50$; $\beta=0.11$, $t=4.04$, $p<0.001$); and (b) eventive main

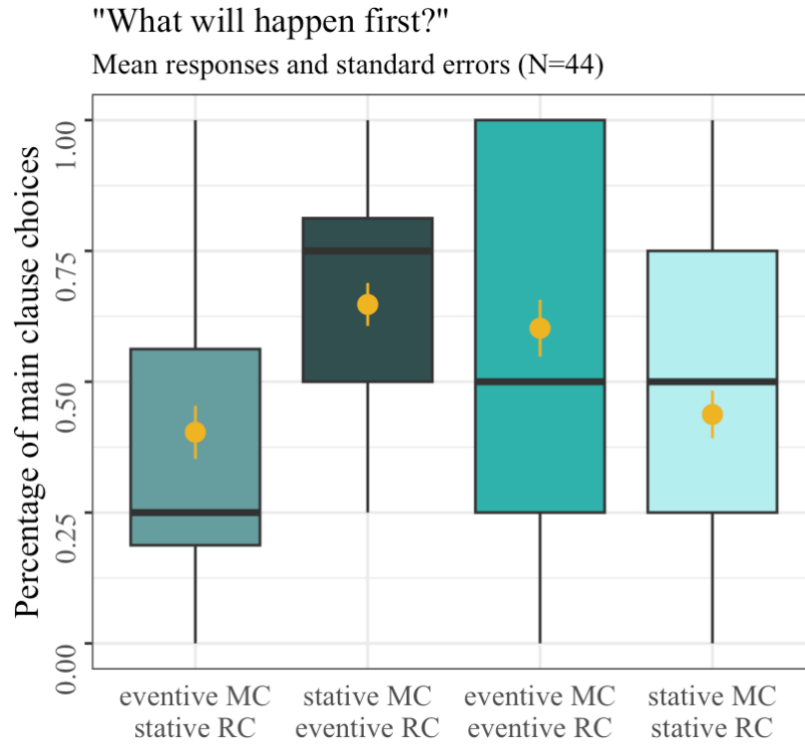


Figure 4 Mean main clause choices in Experiment 2 with boxplot summarizing means per participants in each condition. Yellow dots represent overall means, and yellow error bars the standard error of the mean.

clauses were less likely to be judged as occurring first when combined with a stative relative clause compared to an eventive relative clause ($M_{\text{EVENT-STATE}}=0.40$, $SD=0.49$ vs. $M_{\text{EVENT-EVENT}}=0.60$, $SD=0.49$; $\beta=-0.10$, $t=-3.80$, $p<0.001$). There were no main effects, neither for the event type in the main clause ($Df=1$, $\chi^2=2.15$, $p=0.14$) nor for event type combination ($Df=1$, $\chi^2=0.21$, $p=0.65$).

A notable difference between Experiments 1 and 2 was the absence of a relative clause-first strategy in Experiment 2 ($M_{\text{OVERALL}}=0.52$, $SD=0.50$). This was reflected by a significant main effect of tense in the cross-experimental analysis ($Df=1$, $\chi^2=58.73$, $p<0.001$). Additionally, a significant main effect of the event type in the main clause was observed ($Df=1$, $\chi^2=5.22$, $p=0.02$), along with a significant interaction between main clause event type and event type combination, consistent with the single-experiment results ($Df=1$, $\chi^2=81.33$, $p<0.001$). All other effects remained non-significant (i.e., main effect of event type combination: $Df=1$, $\chi^2=1.03$,

$p=0.31$; interaction between main clause event type and tense: $Df=1$, $\chi^2=0.49$, $p=0.49$; interaction between event type combination and tense: $Df=1$, $\chi^2=1.72$, $p=0.19$).

5.3.4. Discussion of Experiment 2

Experiment 2 largely replicated the pattern of results from Experiment 1, with participants judging situations to happen first based on the dynamic properties of the two predicates: For future-under-future relative clause constructions, they were more likely to order static states before dynamic events, as predicted by only by a Figure-Ground account of temporal construal. Similarly, we replicated the effect of iconicity from Experiment 1, in the absence of event dynamic differences: When both situations described events, participants tended to follow the linear order of the sentence to order situations in time.

In contrast to Experiment 1, we found no overall preference for locating relative clauses first in time: Choices between main and relative clauses were on average at chance level, with differences between conditions only due to differences in dynamicity encoding. While participants in Experiment 1 seemed to adhere to the discourse-pragmatic function of relative clauses as situational backgrounds – potentially influenced by the task design – participants in Experiment 2 did not incorporate this function into their temporal judgments. We propose that this difference between experiments supports the modal theory of the future: Since the individual temporal locations of each event as well as their relationship are underspecified, comprehenders have greater flexibility in determining temporal order. Interestingly, this flexibility appears to be mediated by a Figure-Ground principle and, in the absence of dynamicity differences, by iconicity.

5.4. General Discussion

The aim of this paper was to investigate the cognitive mechanisms underlying a robust phenomenon in the construal of temporally ambiguous language – state-before-event inference – and to test two alternative accounts. We asked: Does the state-before-event inference arise because people rely on temporal anchoring, using the temporal relationships encoded by grammatical tense? Or do comprehenders infer temporal order based on the dynamic properties of the described situations, following a Figure-Ground principle rooted in cognitive salience? In order to answer this question, we also explored questions about the meaning of the future tense: Is the future primarily a tense marker, encoding temporal relationships, or is it a modal marker, more connected to notions of future possibility?

In two experiments, we found a consistent pattern that not only replicated our previous findings regarding the state-before-event inference (Marx & Wittenberg, 2022, 2024, 2025; Marx, Iwan & Wittenberg, 2025) but also provided clear evidence for a Figure-Ground principle (Gleitman et al., 1996; Levinson, 1996; Talmy, 1975; Zlatev, 2010), extended to the event domain (see Hopper, 1979; Reinhart, 1984; Wårvik, 2004): Dynamicity influenced temporal order, regardless of any effect of syntactic encoding or tense. That is, dynamic events served as Figures against the Ground of static states, both when they were described as occurring in the past (Experiment 1) or in the future (Experiment 2).

Furthermore, in both experiments we found second-order effects of iconicity in cases where there were no differences in event dynamicity: Participants ordered events in time to match the reported order. Importantly, however, this pattern also demonstrates participants' sensitivity to the difference between states and events: Only in sentences describing two eventive situations did iconicity shift temporal order construals to orders that followed the reported order. This replicates our previous findings (Marx & Wittenberg, 2024; Marx, Iwan & Wittenberg, 2025), but also suggests that, even in the absence of event dynamic differences, people use their

knowledge about event structure and infer that events, rather than states, drive narratives forward in time (see Dry, 1982).

With regard to the states-before-events inference, one aspect of our data that requires further discussion is the observed preference for ordering relative clauses first in time in Experiment 1. This finding was unexpected, as it contradicts our initial hypotheses about past-under-past relative clause constructions and diverges from our previous findings using similar materials (see Marx & Wittenberg, 2022, 2025; Marx, Iwan & Wittenberg, 2025). Instead, it aligns with theories emphasizing the pragmatic function of relative clauses in discourse to establish situational (and temporal) background (Chierchia & McConnell-Ginet, 1990; Depraetere, 1996; Ehrlich, 1987; Hopper, 1979; Potts, 2005; Reinhart, 1984).

Why might this be the case? We suspect that this preference to order relative clauses first is driven by the design of the current study: In both experiments, participants were asked to reason about the temporal dimension of the sentences, providing explicit temporal order judgments. In contrast to previous methods (i.e., video-sentence matching or act-out tasks), which allowed for more implicit reasoning, the explicit focus on temporal order in the present study may have led participants to weigh linguistic and structural factors differently than in tasks that do not make temporal order the explicit question under discussion.

Similar shifts in processing strategies depending on how the question under discussion is framed have been observed in causal reasoning as well as in scope phenomena, presupposition and information structure, and other pragmatic inferences (Beaver, Roberts, Simons, & Tonhauser, 2017; Evcen & Wittenberg, 2022; Gualmini, Hulsey, Hacquard, & Fox, 2008; Skordos & Barner, 2019). Even more broadly, the specifics of how linguistic experiments are designed may shift results in important ways. Various task characteristics – such as how stimuli are presented, how many fillers are used, how many response choices are available, or how they

are labeled – can greatly affect the sensitivity of an experimental setup (Arehalli & Wittenberg, 2021; Marty, Chemla, & Sprouse, 2020; Schütze, 2016; Schütze & Sprouse, 2013).

However, unlike in these other studies, our main data pattern does not fundamentally change between tasks, populations, languages or, here, tense: There is a robust preference to order states before events. While the baseline preference may vary, we have consistently found people to draw this inference across all experiments we have ever run. As we have argued elsewhere, of course we would expect to find variance, and the effect should be overridden when stronger forces, such as causality, prevail, like in (15):

(15) *John planted persimmons that were delicious.*

In order to be delicious, persimmons need to be planted first: When causality is strong enough, the state-before-event inference should be overridden (although perhaps being delicious could be attributed to any persimmons, past and future). The mechanisms and determining factors for such a possible disappearance of the state-before-event inference should be the topic of future research.

Another open question is how variation within the class of stative predicates may influence temporal construal (for discussion see also Marx & Wittenberg 2025; Marx, Iwan & Wittenberg 2025b):²⁹ In this study, we primarily examined stative descriptions of agents' locations, relying mostly on posture verbs such as *to lean*, *to sit*, and *to stand*. However, stative predicates encompass a much broader range of meanings, including mental, physical, and possessive states, which could potentially affect how people infer temporal relations between described situations during language comprehension. While this possibility requires further investigation, initial evidence from Marx & Wittenberg (2024) suggests that this may not be the case: Testing

²⁹ We thank an anonymous reviewer for this suggestion.

a variety of stative predicates, we found a consistent pattern, indicating that the states-before-events-inference observed here extends beyond stative posture predicates.

Finally, our findings contribute to a long-standing debate in theoretical linguistics about theories of the future, which intersects with discussions in cognitive science about how people can mentally travel through time (Debus, 2014; Gautam, Suddendorf, Henry, & Redshaw, 2019; Redshaw, 2024; Suddendorf, Addis, & Corballis, 2009; Tulving, 1985). Both fields address the question of whether the future is simply another point on a timeline or refers to a non-deictic spectrum of possibilities. This distinction would also render reasoning about the future fundamentally different from reasoning about the past (Belnap, Perloff, & Xu, 2001; Perry, 2006). While other psycholinguistic research has investigated how people process the future (e.g., Biondo, Soilemezidi, & Mancini, 2022; Dickey, 2000), relatively little attention has been paid to temporal construal in complex constructions (but see Gennari, 2003; Dowty, 1982 for a theoretical discussion). The present study is a first step towards filling this gap.

5.5. Conclusion

This study examined how comprehenders derive temporal order from ambiguous relative clause constructions, revealing a robust state-before-event inference as predicted by a Figure-Ground account of temporal construal. In two experiments, we show that salience differences between static states and dynamic events predict temporal construal across tenses in English, independent of tense meaning: Static states serve as the background for dynamic events. Importantly, our findings shed light on theories of future tense, suggesting that it is the modal nature of the future that introduces greater flexibility in temporal order judgments compared to the past. Overall, our results may advance our understanding of the origins of the state-before-event inference and illuminate how the interplay between tense, modality and event dynamicity shapes the construal of temporal relations from language.

6. General discussion of the thesis and future directions

The guiding research question of this thesis was: How do the dynamic properties of described situations influence the mental construal of complex temporal structure during language comprehension? In answering this question, the thesis bridged two lines of research that have largely developed in parallel with limited exchange: on the one hand, linguistic research on the lexical semantics of verbs, a literature that has established fine-grained distinctions with regard to the internal structure of the situations verbal predicates they encode; and on the other hand, event cognition research, a literature that has identified the notion of change as a key element in how people parse and sequence their perceptual experience into discrete event units. Their possible point of convergence to generate new insights is how each tradition conceptualizes the notions of dynamicity and change:

In linguistic research, the primary goal has been to develop formal systems for verbal predicate types that can predict and explain the semantic and grammatical behavior of verbs (Carlson, 1981; Dowty, 1977, 1986; Hinrichs, 1986; Jackendoff, 1991; Mourelatos, 1978; Potts & Taylor, 1965; Pustejovsky, 1991; Rappaport Hovav & Levin, 1998; Smith, 1999; Taylor, 1977; Vendler, 1967; Verkuyl, 1972). Perhaps due to its self-understanding as a formal discipline, however, this line of work has often bracketed questions that ask whether such predicate types also systematically give rise to distinct mental representations – that is, whether the proposed categories are psychologically real, extending beyond the confines of a linguistic system. Even within psycholinguistics, where the mapping from language to mental representations is a central focus, research has often concentrated on individual predicates and how their internal structure shapes the mental construal of isolated, simplex situations (Barner et al., 2008; Becker et al., 2013; Bott, 2010; Bott & Gattnar, 2015; Coll-Florit & Gennari, 2011; Flecken & Gerwien, 2013; Foppolo et al., 2021; Magliano & Schleich, 2000; Milczarski, 2021; Minor et al., 2022; Paczynski et al., 2014; Piñango et al., 2006; Radvansky et al., 1998; Sakarias & Flecken, 2019;

Strickland et al., 2015; Wittenberg & Levy, 2017; Zwaan & Madden, 2005). In this respect, the emphasis has often been placed on event-structural dimensions such as telicity and durativity – which categorize different types of eventive verbs (i.e., activity vs. accomplishment vs. achievement verbs) – with comparatively little attention paid to more perceptually grounded features, such as dynamicity, and the real experience of perception, which is concerned, at any point, with complex situations.

Event cognition research, by contrast, has focused squarely on how people mentally represent and organize events over time, particularly across sequences of events and within more complex situations (Altmann & Ekves, 2019; Zacks et al., 2007, Baldwin & Kosie, 2021). Much of this work has treated dynamic change as a central cue for event segmentation, and consequently, for how the mind represents events from perceptual input. However, a closer examination of how change is conceptualized in this literature reveals an important divergence from the linguistic perspective: Change is typically considered a relative property – such as contextual discontinuities (Huff et al., 2014; Kurby & Zacks, 2008; Rouhani et al., 2020; Shin & DuBrow, 2021; Siefke et al., 2019) – rather than as a categorical distinction between dynamic and stative situations. As a result, the concept of stativity, as a contrast to dynamicity, has remained under-explored in the event perception literature, along with questions about how stative situations are represented mentally and which role they play in shaping the temporal structure of complex situations.

This thesis has brought together insights from both of these research traditions, by adopting the linguistic literature’s detailed attention to event-structural properties in the interpretation of linguistic descriptions – specifically its distinction between dynamic and non-dynamic predicates – while also acknowledging the central role that dynamicity plays in building mental representations of complex situations in perception, as demonstrated by the event cognition literature.

6.1. Mapping visual temporal order onto relative clause constructions

In Chapter 2, I used past-under-past relative clauses as a testing case to assess the role of event dynamicity on temporal construal in language comprehension. Relative clause constructions are particularly well-suited for this purpose because they do not explicitly specify the temporal relationship between main and subordinate clause. Indeed, previous formal work on temporal interpretation has emphasized the structural independence of relative clauses from their main clause, predicting no particular preference of temporal order (Abusch, 1997; Enç, 1986, 1987; Ogihara, 1996; Stowell, 2007; von Stechow, 1995): as long as both described situations are located past relative to the utterance time, any ordering between them should be semantically acceptable. From a discourse pragmatic perspective, however, it has been argued that relative clauses typically convey background information (Cinque, 2020; Fox & Thompson, 1990; Levinson, 1983; Quirk et al., 1985; Thomson & Martinet, 1985). And since the background of a scene should be in place before its foregrounded content unfolds, a discourse pragmatic principle would systematically favor temporal interpretations where situations described by a relative clause happened earlier in time than main clause situations. I contrasted these two well-established perspectives with a third possibility, motivated by the event-structural distinction between dynamic events and static states in linguistics, and informed by saliency asymmetries observed in event cognition research: If event dynamicity, cued by inherent differences in cognitive saliency (Clewett et al., 2020; Huff et al., 2014; Kurby & Zacks, 2008), guides temporal order interpretation, then less salient states should be more likely to be interpreted as temporally preceding more salient, dynamic events, irrespective of whether they are structurally encoded in the main or the relative clause. This prediction follows directly from the assumption that temporal construal in language involves linking situations to a temporal reference time – a temporal anchor – relative to which they are interpreted (Carroll & von Stutterheim, 2010; Klein, 1994, 2000, 2008a; von Stutterheim et al., 2003). The choice

of this anchor is not arbitrary but shaped by the (relative) saliency of the described situations: more salient, dynamic events are more likely to serve as the temporal reference time of less salient, static states. The linguistic encoding of tense and aspect then determines how the anchored state is temporally related to its anchor: past relative to it, in the case of two past tense situations.

To test these predictions, I conducted seven video-to-sentence matching experiments, each in both English and German: On each trial, participants watched a video clip, showing two independent situations, presented in one of two possible temporal orders. They then rated the adequacy of a past-under-past relative clause description for the scene they had just viewed. The key manipulation was which of the two situations was encoded by the main and the relative clause, and how each situation was linguistically framed: as a dynamic event or a static state. Across experiments, and languages, we found that participants showed clear temporal order preferences: Videos were rated as more acceptable when the situation described as a state preceded the situation described as an event. Crucially, this pattern held regardless of which clause encoded the state: whether it was described by the relative clause (Experiments 1) or the main clause (Experiments 2). The preference was further amplified when the linear order of the linguistic description matched the linear order of the video (Experiment 2c). At the same time, when both situations were framed with the same event-structural type, no order preference emerged (Experiment 3): Participants rated both sequences as equally acceptable when both situations were described as events and rated them equally unacceptable when both situations were described as states.

Overall, Chapter 2 marked a first empirical step towards examining the influence of event dynamicity on temporal construal in language comprehension. It introduced and systematically tested preferences of temporal order in a linguistic context – past-under-past relative clauses – that had not traditionally been considered as a central test case for temporal interpretations but

rather served as a counterexample for other types of complex sentences, precisely because it underspecifies temporal relations (but see Arregui & Kusumoto, 1998; Ogihara, 2015; von Stechow & Grønn, 2013a). For the purposes of this thesis, I leveraged this temporal underspecification to focus on and manipulate only the event-structural properties of the described situations.

More broadly, Chapter 2 represents a first attempt to meaningfully connect insights from linguistic theory on event structure with those from event cognition, in order to derive testable predictions: On the linguistic side, I built on the core hypothesis from event-structural accounts (Carroll & von Stutterheim, 2010; Klein, 1994, 2008a, 2009; von Stutterheim et al., 2003, 2009) according to which comprehenders track a set of parameters for each temporal expression, and integrate them into a complex temporal structure of a described situation (cf. Kehler, 1994; Radvansky et al., 1998; Zwaan et al., 1995 for a related point). Crucially, how these complex temporal structures look like depends on how language encodes these parameters and their interrelations, ranging from lexicalized event structure of predicates, to grammatical devices like tense and aspect, to adverbial modifiers.

At the same time, my approach departs from linguistic accounts in two crucial ways, both inspired by event cognitive research: First, while the linguistic accounts have concentrated on grammatical aspect and thus, in languages like English, on questions of telicity,³⁰ I focused instead on dynamicity as a more perceptually grounded property of situations. In the languages I tested, dynamicity is encoded in the lexical meaning of predicates. Second, I extended the notion of saliency: In linguistic event-structural approaches, temporal anchoring is generally explained via linguistic saliency, in other words, how accessible a potential reference time is given the context, usually provided via prior discourse or as sentence-initial cues. By contrast,

³⁰ One may speculate that this emphasis on telicity follows partly from Klein's original framework, where he introduced the notion of topic time into a Reichenbachian system of tense, specifically to accommodate grammatical aspect. On this account, aspectual distinctions are captured through the relation between topic and situation time.

I adopted a perceptual definition of saliency based on the well-established finding in event cognition research: dynamic events attract more attentional resources than static states (Clewett et al., 2020; Huff et al., 2014; Kurby & Zacks, 2008; Ongchoco et al., 2023; Swallow et al., 2011; Zacks et al., 2007). I hypothesized that people recruit this general event knowledge, rooted in their everyday perceptual experience, to identify likely temporal anchors and construe the temporal structure of complex situation descriptions accordingly: dynamicity functions as a cognitively grounded cue of saliency and can thus predict how people understand the temporal order within complex situations.

In sum, in Chapter 2, I have proposed – and tested – that people combine both their linguistic and general knowledge about the dynamic properties of situations when they map temporal order onto linguistic descriptions. One caveat of the empirical study, however, is that participants always watched temporal orders and then matched them to a complex sentence. While this method revealed reliable temporal order preferences, these might only emerge in response to already established orders, as determined by the structure of the visual materials. As such, the findings in Chapter 2 do not speak directly to how people *actively construe* temporal order in language comprehension, nor whether they rely on the same underlying mechanisms to do so. This question motivated my investigation in Chapter 3, which shifted focus from *preferred* temporal order to *inferred* temporal order.

6.2. Inferring temporal order from relative clause constructions across languages

In Chapter 3, I examined how people construe temporal order from linguistic input alone, to directly tap into their inferences of temporal order. The question was: How do people understand the temporal order between two described situations when visual input is lacking?

Without temporal order options in the real world, do they rely on the dynamic properties of the situations in the same way?

Chapter 3 also investigated whether temporal order inferences generalize across languages – a crucial step in understanding whether temporal construal depends on language-specific encoding. Specifically, I focused on two significant cross-linguistic differences and posed the following questions: (a) How does typological variation in encoding of temporal relations in embedded sentences influence dynamicity-driven temporal order inference? (b) To what extent is the role of dynamicity in temporal order inferences mediated by the linguistic level at which it is encoded?

These two foci stem from central observations in the linguistic literature: First, complex sentences differ across languages in how they encode temporal relations – a phenomenon commonly known as Sequence of Tense (SoT): In SoT languages such as English (Abusch, 1997; Declerck, 1988; Stowell, 2007), and to a certain extent German (Boskovic, 2012; but see von Stechow, 2005), past-under-past complement clauses are typically ambiguous: they can be interpreted as simultaneous, meaning the embedded situation overlaps with the main clause situation, or as sequential, meaning the embedded situation must precede the main clause. In non-SoT languages, such as Polish (Arregui & Kusumoto, 1998; Mucha et al., 2023), only the latter interpretation is available. Although SoT primarily concerns complement clauses rather than relative clauses, this typological distinction influences how comprehenders infer temporal relations between clauses in relative clause constructions (Ogihara, 2015, but see Arregui & Kusumoto, 1998).

Second, across languages, the correspondence between event-structural properties (i.e., telicity, durativity and dynamicity) and their linguistic expression is not one-to-one (cf. Bott & Gattnar, 2015; Flecken, 2011; Gerwien & von Stutterheim, 2018; Michaelis, 2011): While languages like English and German (explored in Chapter 2) encode dynamicity lexically, other languages,

such as Slavic languages, mark similar distinctions grammatically through aspect marking (Bogdan & Sullivan, 2009; Comrie, 1976; Czochralski, 1972). Overall, the goal was to investigate temporal order inferences across such cross-linguistic differences – relating both to the encoding of temporal relations or event-structural properties – to gain deeper insights into their linguistic and cognitive underpinnings.

I conducted three act-out studies – one with English, one with German, and one with Polish native speakers – to address these research questions: Participants used toys to act out complex situations described by relative clause constructions, similar to those in the study in Chapter 2. That is, descriptions were manipulated with regard to whether they framed the situations in the main and relative clause as a static state or a dynamic event. The question was which situation participants would enact first, based on their event-structural framing. Across all three languages, participants showed a consistent pattern: When one clause described an event and the other a state, participants tended to enact the state first. Sentences that encoded the same dynamic properties elicited more variation in temporal order inference between languages.

The study in Chapter 3 yielded two new insights: First, I replicated the pattern of temporal order preference found in Chapter 2, when participants had to infer the temporal order solely based on linguistic descriptions: static states were consistently inferred as preceding dynamic events. As such, Chapter 3 provided further evidence for the robustness of the temporal order inference that states precede events. It emerged as a reliable preference in Chapter 2 and extended to enacted construal in Chapter 3.

Second, I showed that this temporal order inference, driven by event dynamicity, emerges independently of typological differences across languages: Dynamicity served as a reliable cue for inferring temporal order in both SoT languages (English, German) and non-SoT languages (Polish), when the two described situations differed in dynamicity. Likewise, the level at which dynamicity was encoded – whether in the lexicon, as in English and German, or as part of the

grammatical system of a language, as in Polish – did not affect participants’ performance: across all languages, participants tended to order states before events. Thus, Chapter 3 showed that the inference is broadly generalizable across languages and that typological differences may not substantially modulate this inferential pattern.

6.3. Temporal order inference in discourse comprehension

In Chapter 4, I addressed a different type of generalizability, by taking a closer look at temporal order inferences in another linguistic context. This shift in context is important because one might reasonably question how frequently the construction type I investigated so far – namely, past-under-past relative clauses – actually occurs in everyday language use. If these constructions are relatively rare and contextually marked, the observed state-before-event inference may be a niche effect, tied to the specific syntactic context. In other temporally underspecified constructions, however, other principles might guide temporal construal and potentially override this inference.

A natural next step was to investigate temporal order inferences in discourse comprehension: Similar to the relative clauses in Chapters 2 and 3, consecutive sentences can be temporally underspecified. In such cases, previous research has made a strong case for an iconicity principle (Diessel, 2008; Fleischman, 1990; Horn, 2022; Schmerling, 1975; Tai, 1983): Comprehenders start from the assumption that the order of mention iconically reflects the order of situations in time. At the same time, formal accounts of discourse comprehension (cf., DRT) have long acknowledged that differences in event structure, such as the distinction between dynamic events and static states, influence temporal interpretation in discourse (Hinrichs, 1986; Kamp & Rohrer, 1983; Lascarides & Asher, 1993, see also Dowty, 1986; Partee, 1984). These accounts argue that stative situation descriptions usually function as elaborations, and thus temporal backgrounds, of the unfolding narrative. However, they offer little explanation for

why these differences should affect temporal construal, leaving the cognitive basis for such inferences an open question that Chapter 4 set out to explore.

In two forced-choice experiments, participants read pairs of consecutive sentences in past tense that varied both in their dynamic properties (event vs. state) and their reported order. They were then asked to indicate which of the two situations, as described by each sentence, had happened first in time. As in the previous chapters, participants' temporal inferences were predicted by the dynamic properties of the sentences: When one sentence encoded a state and the other an event, states tended to be ordered first in time, regardless of their reported order. Furthermore, when the sentences encoded two states, differing in their typical duration, participants systematically inferred that the longer state had happened before the shorter state. When the sentences encoded two events, the iconicity principle bore out: Inferences of temporal order were driven by the order in which situations were reported.

Chapter 4 expanded on the previous findings in the two important ways: First, it replicated the states-before-events inference, previously observed in past-under-past relative clause constructions, in a new linguistic context: consecutive past tense sentences forming minimal units of discourse. Broadly speaking, these findings are notable as they demonstrate that dynamicity-driven temporal order inferences are not bound to specific syntactic constructions but generalize across a range of temporally underspecified linguistic descriptions. More specifically, however, the study in Chapter 4 moved beyond structurally linked situation descriptions and tested temporal order inferences in structures that were linked via contingency relations: In the previous studies, syntactic embedding explicitly signaled two sub-situations belonged to a single complex situation, whereas in Chapter 4, the sub-situations were encoded in two separate sentences, linked only by (weakly) causal or contingent relations between the described situations that could plausibly unfold in either temporal order. Despite this linking difference – and even though people may have held expectations about the typical direction of

these relations – people inferred states to happen before events, thus aligning their causality or contingency interpretation with this temporal inference (i.e., the Causal Law, Lascarides & Asher, 1993; see also Briner et al., 2011; Dery & Bittner, 2016). While we did not obtain causal judgments, the findings in Chapter 4 suggest that temporal order inferences, grounded in event dynamicity, are not only central to complex situation comprehension but potentially also to how people reason about causality or construe contingency relations (see Section 6.5 for future research on this topic).

Second, Chapter 4 broadened the range of predicate types under investigation: Whereas the studies in the Chapters 2 and 3 examined temporal order inferences, using posture and position predicates as stative situation descriptions, in Chapter 4, I incorporated a more diverse set of stative predicates, including physical and possessive stative descriptions, that also varied in their duration (Dery & Koenig, 2015; Dowty, 1979; Kratzer, 1995). The findings show that, when combined with an eventive descriptions, these different stative predicates behave similarly: states were consistently inferred to have happened first. Notably, however, when two stative predicates were combined, durational differences systematically affected temporal construal: longer states were more likely to be ordered before shorter states – possibly because they were construed as the more prototypical state. At large, this suggests that duration adds nuanced differences to the internal structure of states, which become relevant when dynamicity differences are absent. To this effect, Chapter 4 is also a first empirical step to shed light on how dynamicity interacts with other event-structural dimensions in temporal construal in language.

6.4. Temporal order inferences across tenses

In Chapter 5, I shifted focus from tracing dynamicity-driven temporal order inferences across different languages, linguistic contexts and experimental tasks, to exploring their underlying

origins. This involved a critical re-evaluation of the working hypothesis I pursued in the previous chapters, questioning its theoretical explanation for how the observed temporal order inferences are mentally construed: Up to this point, I had argued that, in complex situation descriptions, people distinguish between a reference time – which is often identified with a more salient, dynamic event – and an anchored situation – which is often identified with a less salient state – to systematically infer temporal relations between the two (Carroll & von Stutterheim, 2010; Klein, 1994, 2009; von Stutterheim et al., 2003, 2009). The temporal relationship between reference time and anchored situation is determined linguistically through tense (cf. Kameyama et al., 1993): In the case of two past tense descriptions, this entails a constraint of anteriority, such that an anchored situation cannot follow its reference time, resulting in the inference that states happen before events. The reverse prediction follows for the future tense which imposes a constraint of posteriority: Here, anchored situations should be inferred to happen after their reference time (Dowty, 1982; Gennari, 2003).

This explanation, however, depends on a clearly defined tense meaning – a notion that remains a longstanding debate in linguistic theory (Ogihara, 2007; Partee, 1984; Prior, 1967; von Stechow, 2015), especially with regard to the semantics of the future tense (Giorgi & Pianesi, 1997; Huddleston & Pullum, 2005; Klein, 1994; Palmer, 1987). The aim in Chapter 5 was therefore to scrutinize the explanatory role of tense meaning in relation to the states-before-events-inference: To this end, I tested the tense-anchoring account against an alternative explanation – one that sets aside specific linguistic encoding of temporal structure, but instead draws on the conceptual difference between static states and dynamic events alone. The central research question then was whether dynamicity-driven temporal order inferences are grounded in language, as assumed by a tense-anchoring explanation, or whether they reflect broader cognitive principles – such as a Figure-Ground mechanism – that structure people’s mental representations.

According to a Figure-Ground account, spatial, temporal, and conceptual relationships are structured in mental representations – and consequently systematically encoded in language – along a conceptual saliency asymmetry between smaller, more moveable Figure entities and larger, more stable Ground entities (Gleitman et al., 1996; Levinson, 1996; Talmy, 1975; Zlatev, 2010). While this framework has often been applied to explain the syntactic encoding of spatial relations between objects, I propose that temporal relations within complex situation descriptions may be mentally organized in similar ways: On such a view, the dynamic parts of a complex situation function as Figure entities that unfold against the stable Ground, provided by the static parts of the situation. Crucially, this principle should apply across tenses and independently of their specific semantic content.

I tested these predictions through two forced-choice experiments in English – one with past-under-past and one with future-under-future relative clauses. Similar to Chapter 3, participants read complex situation descriptions, crossed in the dynamic properties of the clauses, and decided which of the sub-situations they thought had happened first (in the past tense experiment) or will happen first (in the future tense experiment). Results clearly supported the Figure-Ground mechanism: Participants consistently preferred to order states before events, regardless of tense meaning.

The findings in Chapter 5 are noteworthy in three ways: First, they provided another replication of the temporal order inference, showing that comprehenders reliably infer that states were more likely to precede events. This strengthens the overall hypothesis that, when the temporal order within complex situations is underspecified, event-structural properties – specifically, the distinction between dynamic events and static states – serve as an important cue to temporal construal during language comprehension.

Second, differences between the two experiments shed light the semantic status of the future tense: in the past tense, participants tended to assume that relative clauses provide

backgrounded, temporally prior information, with this preference mediated by the states-before-events inference. By contrast, this baseline preference was absent when situations were described in the future tense. Here, comprehenders showed greater flexibility in ordering the two sub-situations relative to each other, supporting the theoretical view that future tense functions more like a modal expression than a temporal marker (Enç, 1996; Huddleston & Pullum, 2005; Palmer, 1987): In other words, future tense only hints at the potential occurrence of a situation while leaving both the individual temporal locations of events and their relationship to one another underspecified.

Third, and most importantly, Chapter 5 represents one of the first empirical investigations of a Figure-Ground-mechanism in the event domain, locating it at the origin of how people draw temporal inferences from event-structural differences. Comprehenders use the conceptual distinction between dynamic and non-dynamic situations to mentally construct representations of complex situations based on inherent differences in saliency: more salient, dynamic events are construed as Figures unfolding against less salient, stable states that function as Grounds. Crucially, this Figure-Ground organization seems to operate independently of linguistic encoding, shifting the explanatory focus away from language-specific mechanisms such as tense-anchoring – which requires temporal structure to be constructed based on how languages encode temporal parameters and their interrelations – towards more general cognitive principles that shape how situations are mentally represented on the basis of their internal event-structural properties.

6.5. Future research

The chapters of this thesis have painted a clear empirical picture of the role of event dynamicity in temporal construal from language: When the temporal order within complex situation descriptions is underspecified, people consistently infer that non-dynamic states precede

dynamic events; and they do so reliably across different linguistic contexts and languages. We have called this the states-before-events inference. In this thesis, I have not only demonstrated the robustness of this inferential pattern but further investigated its underlying origins: Specifically, is the state-before-event inference specific to the way in which language maps onto (temporal) meaning or does it reflect broader cognitive mechanisms that organize and structure people's mental representations?

In Chapter 5 of this thesis, I provided a preliminary answer to this question, by offering evidence in favor of a broader cognitive mechanism, that aligns with a Figure-Ground principle. This interpretation is, of course, tentative and calls for further careful and systematic investigations that would need to probe the underlying mechanism in greater depth and evaluate its predictive power. The following sections outline how such research might proceed, thus further developing the Figure-Ground account by elaborating on how the temporal order inferences observed in my data might be situated within Figure-Ground-based reasoning more broadly.

6.5.1. Temporal order inference in development

One important direction for future research concerns the developmental trajectory of the states-before-events inference. In terms of temporal construal, most developmental work has focused on how children come to understand explicitly encoded temporal relations (e.g., Amidon & Carey, 1972; Blything & Cain, 2016; Blything et al., 2015; Clark, 1971; Coker, 1978; de Ruiter et al., 2018; Feagans, 1980; Gorrell et al., 1989; Pyykkönen & Järvikivi, 2012; Trosborg, 1982), linking this development to their temporal reasoning abilities and linguistic capacities (Cromer, 1971; McCormack & Hoerl, 2008; Pawlak, Oehrich, & Weist, 2006; Weist, 2014; Weist, Wysocka, & Lyytinen, 1991; Zhang & Hudson, 2018). Thus, investigating how children infer temporal order in the absence of explicit temporal order information might offer further insights into this interplay between cognitive development and linguistic acquisition more broadly.

From a Figure-Ground perspective in particular, however, tracing temporal order inferences in childhood adds another piece to the puzzle of their origins: If the state-before-event inference reflects a general cognitive strategy rather than a learned mapping between linguistic form and temporal relations, then it should emerge early in childhood — potentially before children have fully mastered explicitly encoded temporal relations such as those in *before/after*-sentences (Blything & Cain, 2016; Blything et al., 2015; Clark, 1971; de Ruiter et al., 2018; Pyykkönen & Järvikivi, 2012; Tillman, Tulagan, & Sullivan, 2020). For these sentences, children up to the age of five heavily rely on a linear order strategy, systematically mapping the order in which situations are mentioned onto real-world temporal order.

A Figure-Ground account would also tie in with broader theories of predictive learning in childhood where learning is generally driven by surprisal or prediction error (Huettig, 2025; Nagai, 2019). Against this backdrop, children need to discriminate early on between predictable and less-predictable aspects of the world, to efficiently direct their attention to what can be learned from: Stable Grounds provide a stable, predictable context, while dynamic events introduce less predictable change, and it is precisely this change that supports learning and cognitive development.

In an ongoing study, we adapted the act-out task in Chapter 3 for preschoolers: A total of 32 English speaking 4-year-olds (target sample size = 36) were recruited at the Harvard LDS to enact sentences with toys ($n = 8 + 8$ filler + 4 post-test trials with *before/after* sentences). On critical trials, sentences were relative clauses constructions that varied in the event dynamic properties of the main and subordinate clause (i.e., eventive-stative vs. stative-eventive). Filler trials tested children's comprehension of simplex sentences. As in previous experiments, we analyzed the order in which children enacted the sentences, as a function of the encoded dynamic properties, using a mixed-effects logistic regression model. Materials, the design as well as the analysis plan are pre-registered at <https://osf.io/kg9v8/>.

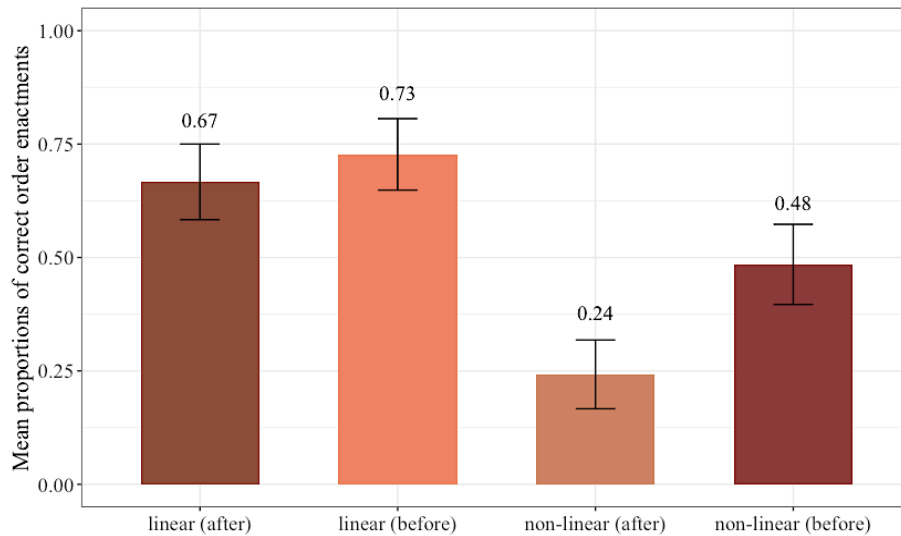


Figure 1 Mean accuracy in the post-test with error bars representing the standard error of the mean.

Consistent with previous findings, children's performance in the post-test trials was notably low ($M_{\text{ACCURACY}}=0.53$, $SD=0.50$, see Figure 1), mostly due to the fact that they failed to understand the correct temporal order: *Before/after* sentences were performed significantly more accurately when the linear order matched the described temporal order ($M_{\text{LINEAR}}=0.70$, $SD=0.46$) than sentences that encoded a non-linear order ($M_{\text{NON-LINEAR}}=0.36$, $SD=0.48$, $b=0.73$, $z=3.66$, $p<0.001$).

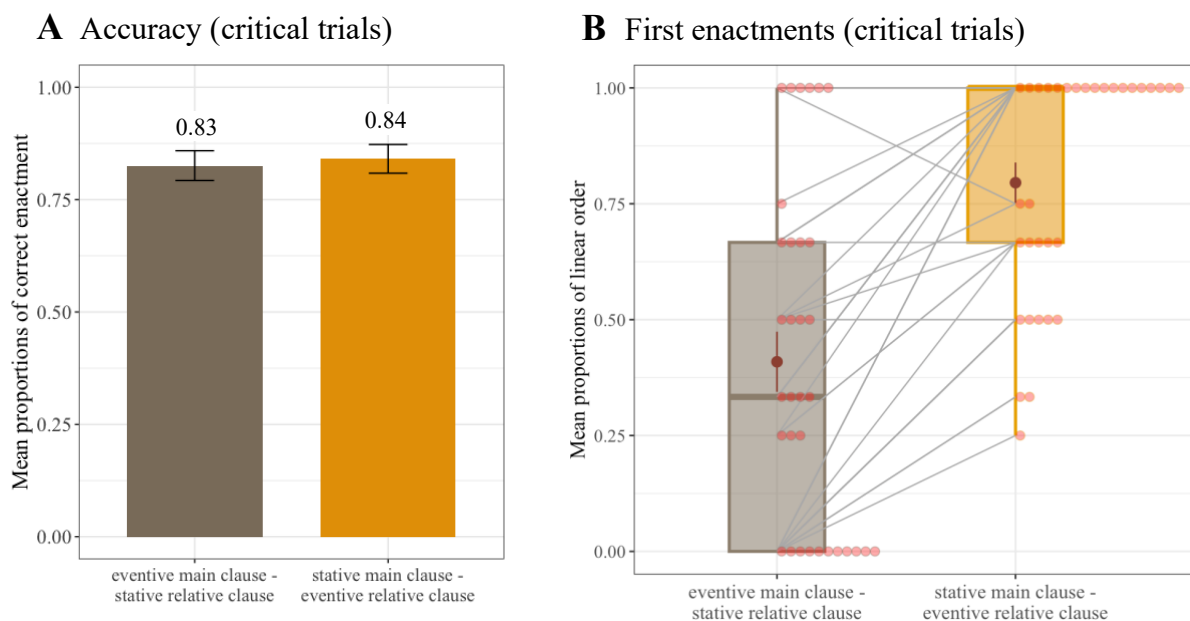


Figure 2 Mean accuracy in the critical conditions (A) with error bars representing the standard error of the mean. Mean first enactments of main clauses in both conditions (B), boxplots summarize means per participants, individually plotted as light red dots and grey lines. Dark red dots and error bars represent overall means and the standard error of the mean.

On critical trials, however, children's accuracy was overall high ($M_{\text{ACCURACY}}=0.83$, $SD=0.37$), such that they reliably chose the appropriate characters for the appropriate actions (see Figure 2A). Regarding the temporal order of their enactments, children were more likely to perform the main clause first when it encoded a state ($M=0.80$, $SD=0.40$), but they reversed the order when the relative clause encoded a state ($M=0.36$, $SD=0.48$). This pattern resulted in a significant main effect ($b=-1.27$, $z=-5.45$, $p<0.001$, Figure 2B).

In sum, 4-year-olds reliably use event dynamicity as a cue to infer temporal order. Like adults, they seem to interpret static states as the temporal backgrounds of dynamic events. This inferential pattern seems to emerge prior to the comprehension of *before/after* clauses and is not overridden by another strong temporal order bias in development: iconicity. Overall, these developmental findings point to a general cognitive mechanism, at the core of the states-before-events inference, which can operate independent of sophisticated understanding of how temporal relations are linguistically encoded.

6.5.2. Temporal order inferences across languages

Chapter 3 of this thesis investigated the generalizability of the states-before-events inference, using Germanic and Slavic languages, which differ along two relevant typological dimensions, as a first test case: Speakers of both language families consistently inferred that states precede events, despite these typological differences. While this marks a first step in investigating cross-linguistic stability, the pool of languages was far too limited to support strong claims of generalizability, as it covered only a narrow subset of the Indo-European language family. Therefore, Chapter 3 was unable to address the broader typological variation in the encoding of temporal information and event structure across the world's languages — a variation that becomes particularly relevant in light of a Figure-Ground account: As a language-independent cognitive mechanism, the Figure-Ground account predicts universality – namely, that event

dynamicity should drive temporal order inferences across languages: non-dynamic states should be mentally construed as stable backgrounds to dynamic events, regardless of how temporal and dynamic properties are linguistically encoded. To test this prediction, future research should thus adopt a large-scale cross-linguistic approach, which would also enable assessing to what extent variation in the inference pattern stems from other linguistic or from cultural differences in how time is conceptualized.

We have started to pursue this direction, by developing an online version of the act-out task described in Chapter 3, to effectively test the role of event dynamicity across a wide range of languages. So far, we have collected preliminary data in Kyrgyz, a Turkic language spoken in Central Asia, to explore whether the observed temporal inference generalizes beyond Indo-European languages. Participants (N=100) were native speakers, tested at the American University of Central Asia (AUCA) in Bishkek, Kyrgyzstan. In the coming years, we plan to extend data collection to ten additional languages (i.e., English, French, German, Czech, Slovak, Russian, Hungarian, Turkish, Arabic, Wolof) as part of an ÖAD-funded project.

The task is designed as follows: Participants read linguistic descriptions of complex situations on a screen (n = 20 critical + 10 control trials) and were then asked to recreate them by dragging images of referents into a central display (Figure 3A). The rationale, similar to that in Chapter 3, was that participants' performance should reflect the temporal structure of their mental

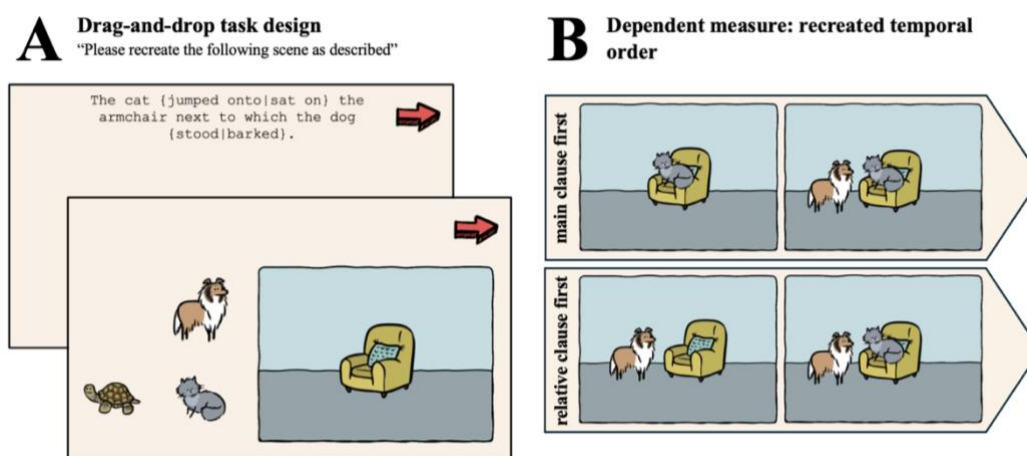


Figure 3 Outline of an experimental trial as an example in English. (A) shows a target sentence and the panel after a few seconds (B) shows its recreation, with first enactments as dependent measure: main clause first (up) or relative clause first (down).

representations: referents involved in events that happen earlier in time should be moved into the display first (Figure 3B).

As linguistic materials, we used past-under-past relative clause constructions where main and relative clauses each described two agents (e.g., dog, cat) in relation to a location (e.g., armchair). Clauses were crossed in their event dynamic properties: one clause always encoding a stative position (see the English example in Figure 1A: *sit on, stand next to*) and the other an eventive activity or change of position (see the English example in Figure 1A: *bark, jump onto*), resulting two event-structural conditions (i.e., state in the relative clause vs. state in the relative clause main clause). Similar to English, and as in other Turkic languages, Kyrgyz distinguishes between static states and dynamic events on a lexical level (Johanson, 1998; Landmann, 2011; Wurm, 1949).

A crucial typological difference, however, concerns the syntax of relative clauses which, in Kyrgyz, reflect the head-final structure and the relatively free word order of the language (see examples 1-2): relative clauses usually precede the noun they modify, and predicates in both the main and subordinate clauses appear in clause-final position. One particularity relates to the position of the main clause subject, which can either appear directly before its predicate at the end of the sentence (1), or in the very front of the sentence (2), resulting in a main clause bracket around a centered relative clause. The first structure, with fronted relative clauses, is thus the exact reverse linear order as English stimuli in most of this dissertation which relativized the object NP of the main clause (for an exception see sections 2.2.6. and 2.2.7.). This typological difference could arguably lead to different strategies in moving the referents.

(1) a. MC_{EVENTIVE} – RC_{STATIVE}

[Иттин жанында турган]RC [креслого мышык чыгып алды.]MC

dog.GEN beside stand.PART chair.DAT cat.NOM climb.PAST take.PERF

The cat climbed onto the armchair next to which the dog stood.

b. MC_{STATIVE} – RC_{EVENTIVE}

[Мышык чыгып алган]_{RC} [креслонун жанында ит турду.]_{MC}

cat.NOM climb.PAST take.PART chair.GEN beside dog.NOM stand.PAST

The dog stood next to the armchair onto which the cat climbed.

(2) a. MC_{EVENTIVE} – RC_{STATIVE}

[Коён]_{MC} [жанындагы уй турган]_{RC} [тосмонун үстүнөн секирип

rabbit.NOM beside cow.NOM stand.PART fence.GEN over jump.PAST

өттү.]_{MC}

pass.PERF

The rabbit jumped over the fence beside which the cow stood.

b. MC_{STATIVE} – RC_{EVENTIVE}

[Уй]_{MC} [коён секирип өткөн]_{RC} [тосмонун жанына турду.]_{MC}

cow.NOM rabbit.NOM jump.PAST pass.PART fence.GEN beside stand.PAST

The cow stood next to the fence over which the rabbit jumped.

In the control trials, participants read temporal adverb clause constructions as well as sentences encoding object relations (i.e., containment and support), to tie temporal construal more closely to other Figure-Ground distinctions (see Sun, Firestone, & Hafri, 2023, Experiment 1). As dependent measure, we analyzed whether participants moved the referent of the main clause or the referent of the relative clause first. A pre-registration including the planned statistical analysis plan is available at <https://osf.io/xt4rh>.

Figure 4 depicts the preliminary results from Kyrgyz: In syntactic constructions that were the reversed mirror images of English object modifying relative clause constructions (see example 1), we found an effect of event dynamicity as predicted: While participants showed an overall bias for linear order, thus dragging the relative clause agent first into the panel when it was also first mentioned, they were more likely to do so when the relative clause agent was involved in

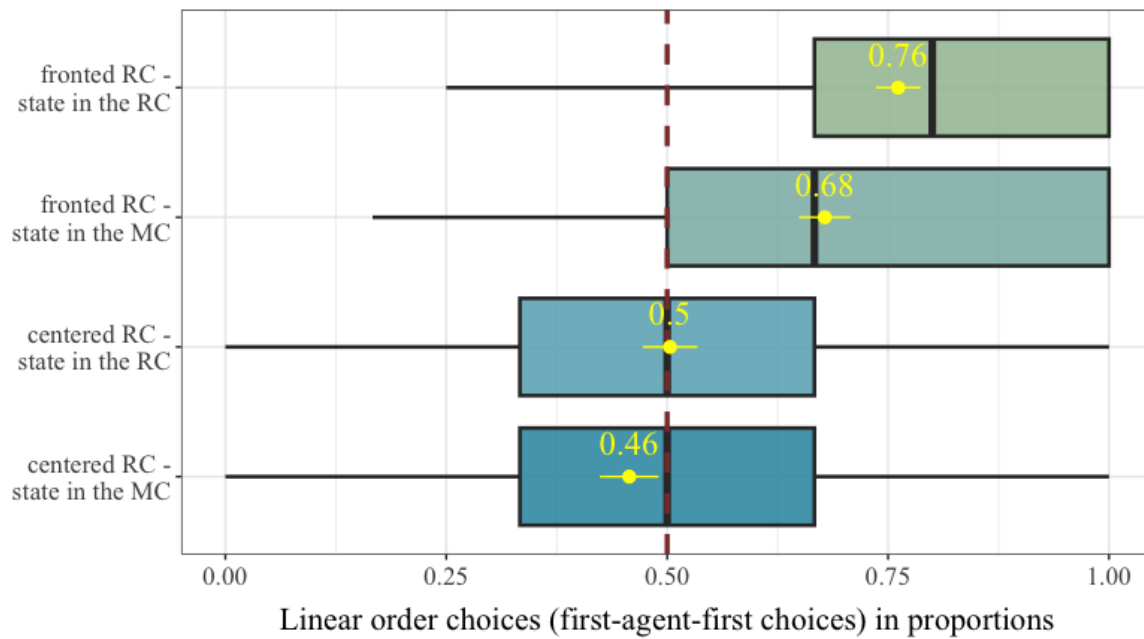


Figure 4 Mean first-agent-first placements in conditions that are syntactical reversals English relative clause constructions (first and second boxplots) and bracketed constructions with centered relative clauses (third and fourth boxplot). Boxplots summarize means per participants. Yellow dots and error bars represent overall means and the standard error of the mean.

a static situation (first boxplot, Figure 4) than when it performed an eventive action (second boxplot, Figure 4). However, when relative clauses were syntactically centered (see example 2), bracketed by the main clause agent and predicate, participants performed at chance level (statistical analysis is ongoing). Overall, this data thus suggests that event dynamicity is a robust cue to temporal order inference across languages, while also highlighting potential interactions with typological differences in syntactic encoding.

6.5.3. Temporal order inference across domains

Another central prediction, derived from a Figure-Ground account, is that the state-before-event inference should not be specific to language, but should generalize as a broader cognitive pattern across modalities. While prior work has provided supporting evidence for object relations (cf. Sun et al., 2023), we tested these predictions with a non-linguistic paradigm on the event domain, modeled on the drag-and-drop task in Section 6.1. This paradigm thus investigates whether the temporal construal of situations in the visual domain is similarly

grounded in event dynamicity. Importantly, this non-linguistic line of work is also part of our ongoing ÖAD-project, aiming to assess the cognitive basis of the state-before event inference across different cultures. To this end, we have already collected data in English (online) and Kyrgyz (in the lab).

For both studies, participants (N=60, native English, N=90 native Kyrgyz) were asked to build two-panel visual scenes, by dragging two images – each depicting the same item (e.g., a weightlifter) – from the side of the screen into a comic strip format (see Figure 5A). Temporal progression was cued both by the left-to-right layout of the panels and by the task structure: participants had to begin with the first panel, images could only be placed in one panel each, and placing an image in the first panel caused it to automatically appear in the same position in the second panel, implying persistence time (Figure 5B). The key manipulation between the two images concerned the depiction of dynamicity. One image showed the item in motion, using motion lines to convey dynamicity (see Cohn & Maher, 2015), while the other showed the item in a stative pose.³¹ The question was whether participants would first establish a stable

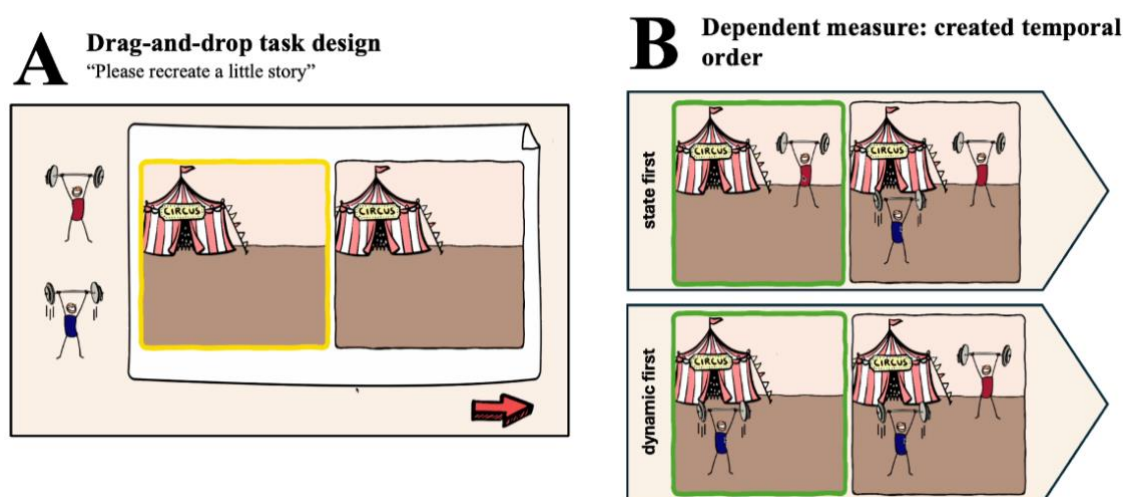


Figure 5 Outline of an experimental trial as an example. (A) depicts the visual display of the items and the comic strip, (B) shows dragged images and their temporal order as the dependent measure: stative first (up) or dynamic first (down).

³¹ We also ran a follow-up experiment that more tightly controlled for visual complexity: motion lines from the dynamic images were fragmented and repositioned beneath the static images, ensuring that any observed effects were attributable to perceived dynamicity. The results replicated the pattern reported in the two experiments here.

background by placing the static image in the first panel, and then the more foregrounded, dynamic image in the second, as predicted by a domain-general Figure-Ground strategy. Backgrounds were held constant across panels and included a spatial anchor (e.g., a circus tent), to avoid creating the impression of a shifting viewpoint. Control trials required participants to arrange containment or support relations in the panels (i.e., between a cup and a saucer). These trials ensured that people would temporally sequence Figure-Ground relations that operate without dynamicity as predicted, building on prior work on object relations (cf. Sun et al., 2023). Participants' image placements were analyzed as a function of dynamicity depicted by the images, using a mixed-effects logistic regression model. The intercept of the model was transformed into baseline probabilities. Materials, experimental design and a detailed description of the analysis plan are outlined in the preregistration, accessible at <https://osf.io/wymgb>.

Results of the both experiments are depicted in Figure 6: Consistent with the Figure-Ground predictions, participants were significantly more likely to place the static image in the first panel and the dynamic image in the second (English: $\beta=0.45$, $z=3.06$, $p=0.002$, Kyrgyz: $\beta=0.28$, $z=2.85$, $p=0.004$). That is, in the absence of linguistic descriptions, participants temporally

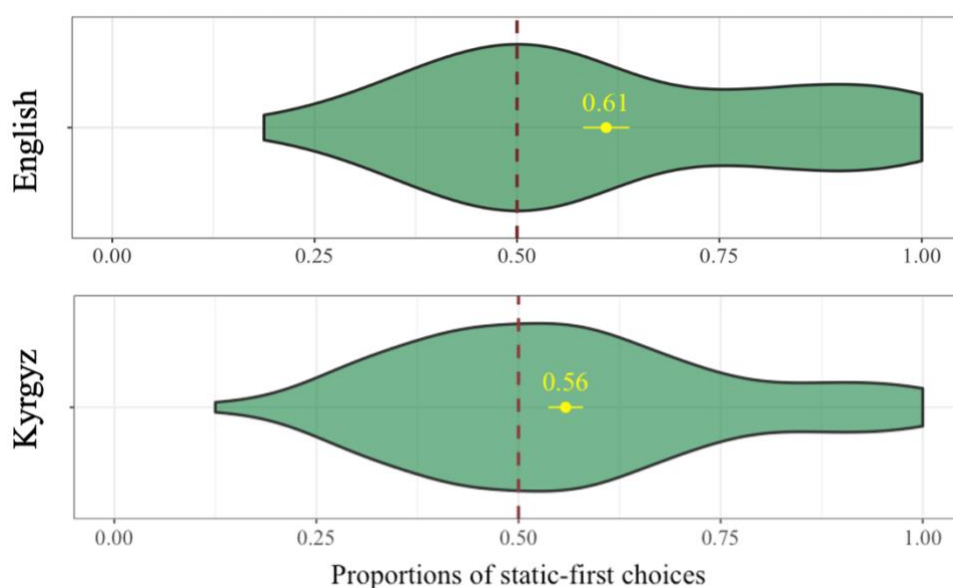


Figure 6 Proportion of static images moved first into the first panel by English (up) and Kyrgyz participants (down). Dashed line depicts chance level performance, width of violin plots corresponds to means by participant, and yellow point and line denote the overall mean and standard error.

structured visual scenes, systematically aligning the static parts with the stable background, against which the dynamic parts unfold as temporally foregrounded events. This pattern mirrors the temporal inference pattern observed in the linguistic evidence provided in this thesis, thus providing preliminary evidence for the hypothesis that the state-before-event inference reflects a broader cognitive strategy rather than a language-specific phenomenon.

6.5.4. Linking temporal order inference and causal reasoning

The state-before-event inference may be part of a larger set of inferential processes that call for further investigation: In Chapter 4, we found that people ordered states before events, even when two described situations were not independent but plausibly connected by weak causal or contingency relations and could have happened in either temporal direction. Given that temporal and causal structure is closely linked (see also Briner et al., 2011; Dery & Bittner, 2016; Oversteegen, 2005), in some sense, the temporal order judgments may have also reflected people's underlying causal assumptions. This link, however, has not been directly investigated in the context of this thesis. A key question, then, is: Does the states-before-events inference, proposed here to result from Figure-Ground reasoning more broadly, also co-occur with dynamicity-driven causal inferences? In other words, when people construe one situation as the stable, stative background and another as the dynamic center of attention, do they also implicitly align their causal judgments with this distinction. Addressing this question is particularly relevant in light of prior work connecting the Figure-Ground distinction with the linguistic encoding of cause-effect structure (Li, Du, & Wolff, 2015; Talmy, 2000).

We turned to conditionals to examine this question in more detail, as they are often associated – at least intuitively – with causal reasoning: People commonly interpret a sentence such as *If it rains today, the street gets wet* such that the *if*-clause (the antecedent) provides the cause or the conditions under which the main clause (i.e., the consequent) is true. However, not all

conditionals follow this interpretation pattern: In biscuit or relevance conditionals such as *If you get hungry, there are biscuits in the fridge*, a causal or conditional link is absent: Neither does the hunger cause the biscuits to be in the fridge, nor do the biscuits in the fridge provide a necessary condition for getting hungry. Cases like these have led linguists to argue that causal interpretations are inferences and not tied to the semantics of conditionals themselves (Bhatt & Pancheva, 2006; Lassiter, 2023).

Assuming that the state-before-event inference reliably orders states before events, then, similar to Copley (2023), we would predict that the dynamic properties of the consequent –whether it is framed as a dynamic event or a static state – should determine causal interpretations: When the consequent encodes an event, people should draw the common causal inference that the antecedent causes the consequent. When the consequent encodes a state, on the other hand, this inference should be absent or even reversed. That is, given that causes always precede their effects, people should be more likely to infer that the static consequent causes the dynamic antecedent.

We tested this prediction in two forced-choice experiments, each with 100 English native speakers: In both experiments, participants read 20 conditional sentences (along with 30 filler and controls) that either encoded an event or a state in the consequent main clause (i.e., *If Daria uses makeup, she gets/has a pimple*). In Experiment 1, participants were asked to make a temporal inference, choosing which part of the sentence they believed would happen first: the antecedent (e.g., makeup) or consequent (e.g. pimple). In Experiment 2, participants were asked to make a causal judgment, choosing whether the sentence meant that the antecedent caused the consequent (e.g., makeup leads to pimple) or the reverse (e.g., pimple leads to makeup). Forced-choice responses were analyzed using mixed-effects logistic regression models. Materials, design details, and the statistical analysis plan are available in the preregistration at <https://osf.io/35ywq/>.

Results for both experiments are shown in Figure 7: In Experiment 1, as predicted, participants chose the antecedent to happen first when the consequent encoded an event, but choices significantly shifted towards selecting the consequent when it encoded a state ($\beta=1.60, z=15.95, p<0.001$). In Experiment 2, participants reliably interpreted sentences to mean that the antecedent caused the consequent when the consequent encoded an event, in line with the common causal inference. By contrast, when the consequent encoded a static state, participants were more likely to reverse causality, interpreting the consequent significantly more often as the cause of the antecedent ($\beta=1.20, z=16.53, p<0.001$).

All in all, the results of both experiments show that temporal inferences through event dynamicity extend to the causal domain: Canonical antecedent-causes-consequent interpretations can be reversed, simply by using a stative predicate in the consequent in conditional constructions. From a Figure–Ground perspective, this pattern should follow naturally from its core principles: Grounds, expressed by stative predicates, do not convey causal meaning but simply elaborate on the background or preconditions of a complex situation.

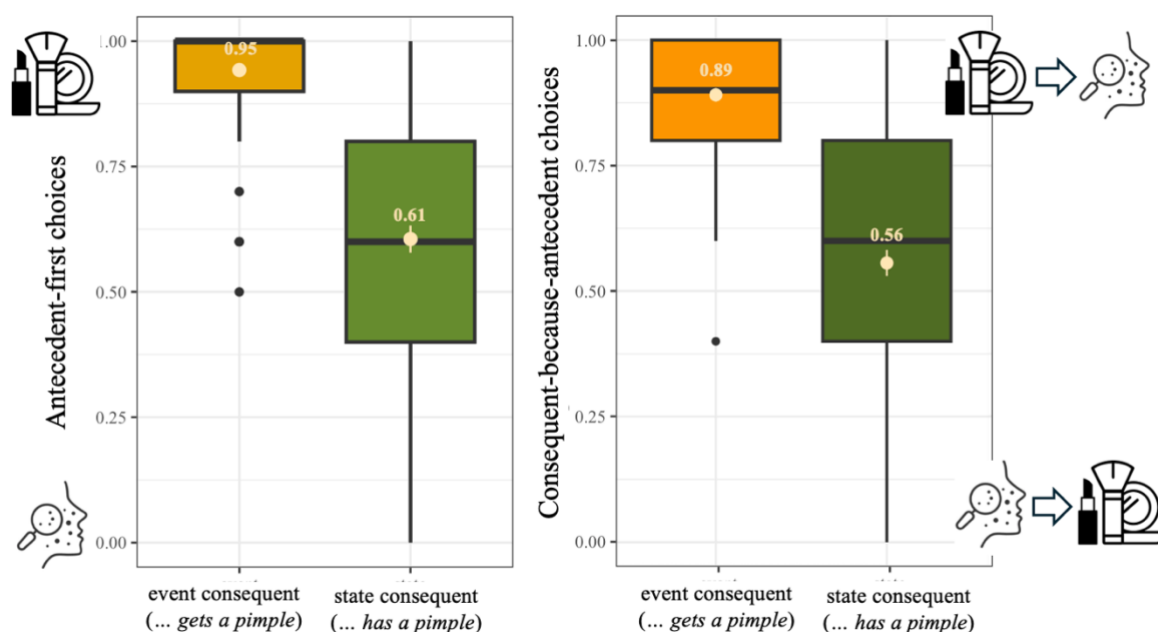


Figure 7 Mean proportions of antecedent-first choices in the temporal judgment task in Experiment 1 (left), mean proportions of consequent-because-antecedent choices in the causal judgment task in Experiment 2. Boxplots depict means per participants. Yellow dots and error bars represent overall means and the standard error of the mean.

Figures, by contrast, which are expressed by dynamic predicates, form the temporal and causal foreground, aligning with the canonical conditional interpretation in which the consequent follows causally and temporally from the antecedent.

7. Conclusion

This thesis used experimental data to bridge two research traditions which have placed events at the heart of their inquiries into human cognition and language: linguistic theories on verbal predicates and cognitive theories on event perception. To this end, it focused on a dimension that both accounts have acknowledged as a central property in how events are represented in language and the mind: event dynamicity as the presence or absence of change over time.

The empirical studies presented in this thesis examined how comprehenders infer temporal order between described situations when such relations are not explicitly encoded. In four lines of experiments, it is demonstrated that temporal order inferences are systematically driven by event dynamicity: when one situation is described as a static state and the other one as a dynamic event, states tend to be interpreted as preceding to events.

Taken together, the findings of this thesis offer a new perspective on event construal at the interface of language and cognition. They suggest that the way in which more complex situations are construed during language comprehension relies on broader cognitive mechanisms, rooted in event perception, where dynamic change serves as a cognitively salient cue for structuring perceptual experience over time. In doing so, this thesis contributes to a more integrated understanding of how a central perceptual principle, namely our sensitivity to change, extends into linguistic event construal and temporal inference.

8. References

- Abrahamsen, E. P., & Gilbert, J. L. (1990). Temporal clause comprehension in fluent and nonfluent aphasics. *Journal of Psycholinguistic Research*, 19(5), 281–292.
<https://doi.org/10.1007/BF01074361>
- Abusch, D. (1997). Sequence of Tense and Temporal De Re. *Linguistics and Philosophy*, 20, 1–50.
- Abusch, D. (1998). Generalizing Tense Semantics for Future Contexts. In S. Rothstein (Ed.), *Events and Grammar. Studies in Linguistics and Philosophy* (pp. 13–33).
https://doi.org/10.1007/978-94-011-3969-4_2
- Altmann, G. T. M., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review*, 126(6), 817–840.
<https://doi.org/10.1037/rev0000154>
- Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Verbal Learning and Verbal Behavior*, 11(4), 417–423.
[https://doi.org/10.1016/S0022-5371\(72\)80022-7](https://doi.org/10.1016/S0022-5371(72)80022-7)
- Anand, P., & Hacquard, V. (2007). When the Present is all in the Past. In L. de Saussure, J. Moeschler, & G. Puskás (Eds.), *Recent Advances in the Syntax and Semantics of Tense, Aspect and Modality* (pp. 209–228). De Gruyter Mouton.
- Angela Downing, P., & Locke, P. (2006). *English Grammar: A University Course*. Abingdon, New York: Routledge.
- Arehalli, S., & Wittenberg, E. (2021). Experimental filler design influences error correction rates in a word restoration paradigm. *Linguistics Vanguard*, 7(1), 1–15.
<https://doi.org/10.1515/lingvan-2020-0052>
- Aristotle (1941). *The basic works of Aristotle*. (R. McKeon, Ed.). New York: Random House.
- Arregui, A., & Kusumoto, K. (1998). Tense in Temporal Adjunct Clauses. *Semantics and Linguistic Theory*, 8, 1–18. <https://doi.org/10.3765/salt.v8i0.2814>
- Bach, E. (1986). The algebra of events. *Linguistics and Philosophy*, 9(1), 5–16.
- Bailey, H. R., Kurby, C. A., Giovannetti, T., & Zacks, J. M. (2013). Action perception predicts action performance. *Neuropsychologia*, 51(11), 2294–2304.
<https://doi.org/10.1016/j.neuropsychologia.2013.06.022>
- Baldwin, D. A., & Kosie, J. E. (2021). How Does the Mind Render Streaming Experience as Events? *Topics in Cognitive Science*, 13(1), 79–105. <https://doi.org/10.1111/tops.12502>
- Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2018). Dynamic adjustment of lexical processing in the lexical decision task: Cross-trial sequence effects. *Quarterly Journal of Experimental Psychology*, 71(1), 37–45.
<https://doi.org/https://doi.org/10.1080/17470218.2016.1240814>
- Barner, D., Wagner, L., & Snedeker, J. (2008). Events and the ontology of individuals: Verbs as a source of individuating mass and count nouns. *Cognition*, 106(2), 805–832.
<https://doi.org/10.1016/j.cognition.2007.05.001>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for

- confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Barrie, M., & Spreng, B. (2009). Noun incorporation and the progressive in German. *Lingua*, 119(2), 374–388. <https://doi.org/10.1016/j.lingua.2007.10.020>
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577–660. <https://doi.org/10.1017/S0140525X99002149>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bauer, P. J., & Mandler, J. M. (1992). Putting the Horse Before the Cart: The Use of Temporal Order in Recall of Events by One-Year-Old Children. *Developmental Psychology*, 28(3), 441–452. <https://doi.org/10.1037/0012-1649.28.3.441>
- Beaver, D. I., Roberts, C., Simons, M., & Tonhauser, J. (2017). Questions under discussion: Where information structure meets projective content. *Annual Review of Linguistics*, 3, 265–284. <https://doi.org/10.1146/annurev-linguistics-011516-033952>
- 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. <https://doi.org/10.1016/j.cognition.2013.06.014>
- Behaghel, O. (1932). *Deutsche Syntax. Eine geschichtliche Darstellung*. Heidelberg: Carl Winters.
- Belnap, N., Perloff, M., & Xu, M. (2001). *Facing the future. Agents and choices in our indeterminist world*. Oxford University Press.
- Bennett, M., & Partee, B. H. (1972). Toward the Logic of Tense and Aspect in English. In *Compositionality in Formal Semantics* (pp. 59–109). <https://doi.org/10.1002/9780470751305.ch4>
- Berman, R. A., & Slobin, D. I. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale, NJ: Lawrence Erlbaum.
- Bertinetto, P. M. (1994). Statives, progressives, and habituais: Analogies and differences. *Linguistics*, 32(3), 391–424.
- Bertinetto, P. M., & Delfitto, D. (2000). Aspect vs. Actionality: Why they should be kept apart. *Empirical Approaches to Language Typology*, 6, 189–226. <https://doi.org/10.1515/9783110197099.1.189>
- Bhatt, R., & Pancheva, R. (2006). Conditionals. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell Companion to Syntax* (pp. 638–687). Malden, MA: Blackwell Publishing Inc.
- Binnick, Robert. (1991). *Time and the verb: A guide to tense and aspect*. Oxford, UK: Oxford University Press.
- Biondo, N., Soilemezidi, M., & Mancini, S. (2022). Yesterday Is History, Tomorrow Is a Mystery: An Eye-Tracking Investigation of the Processing of Past and Future Time Reference During Sentence Reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, 48(7), 1001–1018. <https://doi.org/10.1037/xlm0001053>

- Bittner, M. (2005). Future discourse in a tenseless language. *Journal of Semantics*, 22(4), 339–387. <https://doi.org/10.1093/jos/ffh029>
- Bittner, M. (2014). *Temporality: Universals and variation*. Malden, MA: Blackwell.
- Blything, L. P., & Cain, K. (2016). Children's Processing and Comprehension of Complex Sentences Sontaining Temporal Connectives: The Influence of Memory on the Time Course of Accurate Responses. *Developmental Psychology*, 52(10), 1517–1529. <https://doi.org/10.1037/dev0000201>
- Blything, L. P., Davies, R., & Cain, K. (2015). Young Children's Comprehension of Temporal Relations in Complex Sentences: The Influence of Memory on Performance. *Child Development*, 86(6), 1922–1934. <https://doi.org/10.1111/cdev.12412>
- Bogdan, D., & Sullivan, W. J. (2009). *The tense aspect system of Polish narrative: A discourse and cognitive approach*. Munich: LINCOM Europa.
- Bohnenmeyer, J. (2009). Temporal anaphora in a tenseless language. In W. Klein & P. Li (Eds.), *The Expression of Time* (pp. 83–128). <https://doi.org/10.1515/9783110199031.83>
- Borer, H. (1991). The causative-inchoative alternation: A case study in parallel morphology. *The Linguistic Review*, (8), 119–158.
- Borer, H. (2003). Exo-Skeletal vs. Endo-Skeletal Explanations: Syntactic Projections and the Lexicon. In M. Polinsky & J. Moore (Eds.), *The Nature of Explanation in Linguistic Theory* (pp. 31–67). Chicago: Chicago University Press.
- Borer, H. (2005). *Structuring Sense Volume I: In Name Only*. <https://doi.org/10.1093/acprof:oso/9780199263905.001.0001>
- Boskovic, Z. (2012). On NPs and Clauses. *Discourse and Grammar: From Sentence Types to Lexical Categories*, 2(1), 179–242. <https://doi.org/10.1515/9781614511601.179>
- Bott, O. (2010). *The Processing of Events*. <https://doi.org/10.1075/la.162>
- Bott, O., & Gattnar, A. (2015). The cross-linguistic processing of aspect: An eyetracking study on the time course of aspectual interpretation in Russian and German. *Language, Cognition and Neuroscience*, 30(7), 877–898. <https://doi.org/10.1080/23273798.2015.1029499>
- Brandt, S., Diessel, H., & Tomasello, M. (2008). The acquisition of German relative clauses: A case study. *Journal of Child Language*, 35(2), 325–348. <https://doi.org/10.1017/S0305000907008379>
- Bresnan, J. (2008). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. *Roots: Linguistics in Search of Its Evidential Base*, (1), 75–96. <https://doi.org/10.1515/9783110198621.75>
- Briner, S. W., Virtue, S., & Kurby, C. A. (2011). Processing causality in narrative events: Temporal order matters. *Discourse Processes*, 49(1), 61–77. <https://doi.org/10.1080/0163853X.2011.607952>
- Brogaard, B. (2008). Sea Battle Semantics. *Philosophical Quarterly*, 58(231).
- Brown, H. D. (1971). Children's Comprehension of Relativized English Sentences. *Child Development*, 42(6), 1923–1936.

- Bunger, A., Papafragou, A., & Trueswell, J. C. (2013). Event Structure Influences Language Production: Evidence from Structural Priming in Motion Event Description. *Journal of Memory and Language*, 69(3), 299–323. <https://doi.org/10.1016/j.jml.2013.04.002>
- Bunger, A., Skordos, D., Trueswell, J. C., & Papafragou, A. (2016). How children and adults encode causative events cross-linguistically: implications for language production and attention. *Language, Cognition and Neuroscience*, 31(8), 1015–1037. <https://doi.org/10.1080/23273798.2016.1175649>
- Bybee, J., & Dahl, Ö. (1989). The creation of tense and aspect systems in the languages of the world. *Studies in Language*, 13(1), 51–103. <https://doi.org/10.2307/415749>
- Can Mutlu, M., Canbeyli, R., & Saybaşı, H. (2023). Functional near-infrared spectroscopy shows that object relative clauses are more difficult to process than subject relative clauses in Turkish. *European Journal of Neuroscience*, 57(6), 951–961. <https://doi.org/10.1111/ejn.15930>
- Candan Şimşek, A., Aydın, T., & Huff, M. (2025). How does it end? Endpoints of boundaries lead to completion in macro-events. *Memory and Cognition*, 53(5), 1380–1395. <https://doi.org/10.3758/s13421-024-01657-x>
- Carlson, G. (1977). A unified analysis of the English bare plural. *Linguistics and Philosophy*, 1(3), 413–457. <https://doi.org/10.1007/BF00353456>
- Carlson, L. (1981). Aspect and Quantification. In P. Tedeschi & A. Zaenen (Eds.), *Syntax and Semantics* (pp. 31–64). Bingley, UK: Emerald.
- Carroll, M., & von Stutterheim, C. (2010). Event representation, time event relations, and clause structure: A crosslinguistic study of English and German. In J. Bohnemeyer & E. Pederson (Eds.), *Event Representation in Language and Cognition* (pp. 68–83). <https://doi.org/10.1017/CBO9780511782039.004>
- Chafe, W. L. (1970). *Meaning and the Structure of Language*. University of Chicago Press.
- Chen, B., Ning, A., Bi, H., & Dunlap, S. (2008). Chinese subject-relative clauses are more difficult to process than the object-relative clauses. *Acta Psychologica*, 129(1), 61–65. <https://doi.org/10.1016/j.actpsy.2008.04.005>
- Chierchia, G., & McConnell-Ginet, S. (1990). *Meaning and Grammar: An Introduction to Semantics*. Cambridge, MA: MIT Press.
- Cinque, G. (2020). *The Syntax of Relative Clauses: A Unified Analysis*. <https://doi.org/10.1017/9781108856195>
- Clark, H. H. (1977). Inferences in comprehension. In D. LaBerge & S. J. Samuels (Eds.), *Basic Processes in Reading: Perception and Comprehension* (pp. 243–264). https://doi.org/10.4324/9781315467610_15
- Clark, H. H., & Clark, E. V. (1968). Semantic distinctions and memory for complex sentences. *The Quarterly Journal of Experimental Psychology*, 20(2), 129–138. <https://doi.org/10.1080/14640746808400141>
- Clark, E. V. (1971). On the Acquisition of the Meaning of Before and After. *Journal of Verbal Learning and Verbal Behavior*, 10(3), 266–275. [https://doi.org/10.1016/S0022-5371\(71\)80054-3](https://doi.org/10.1016/S0022-5371(71)80054-3)
- Claus, B., & Kelter, S. (2006). Comprehending narratives containing flashbacks: Evidence for

- temporally organized representations. *Journal of Experimental Psychology: Learning Memory and Cognition*, 32(5), 1031–1044. <https://doi.org/10.1037/0278-7393.32.5.1031>
- Clewett, D., Gasser, C., & Davachi, L. (2020). Pupil-linked arousal signals track the temporal organization of events in memory. *Nature Communications*, 11(1), 1–14. <https://doi.org/10.1038/s41467-020-17851-9>
- Cohn, N., & Maher, S. (2015). The notion of the motion: The neurocognition of motion lines in visual narratives. *Brain Research*, 1601(1), 73–84. <https://doi.org/10.1177/0022146515594631>.Marriage
- Coker, P. L. (1978). Syntactic and semantic factors in the acquisition of before and after. *Journal of Child Language*, 5(2), 261–277. <https://doi.org/10.1017/S0305000900007467>
- Coll-Florit, M., & Gennari, S. P. (2011). Time in language: Event duration in language comprehension. *Cognitive Psychology*, 62(1), 41–79. <https://doi.org/10.1016/j.cogpsych.2010.09.002>
- Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge, MA: Cambridge University Press.
- Comrie, B. (1985). *Tense*. Cambridge, MA: Cambridge University Press.
- Comrie, B. (1995). German Perfekt and Präteritum: Speculations on meaning and interpretation. *Grammar and Meaning*, 148–161. <https://doi.org/10.1017/cbo9780511620638.009>
- Copley, B. (2002). *The semantics of the future*. MIT.
- Copley, Bridget. (2023). *Events are the source of causal readings in the simplest English conditionals*. 459–498. https://doi.org/10.1007/978-3-031-05682-6_15
- Costa, J., Lobo, M., & Silva, C. (2011). Subject-object asymmetries in the acquisition of Portuguese relative clauses: Adults vs. children. *Lingua*, 121(6), 1083–1100. <https://doi.org/10.1016/j.lingua.2011.02.001>
- Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience*, 18(18), 7426–7435. <https://doi.org/10.1523/jneurosci.18-18-07426.1998>
- Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks, CA: SAGE Publications.
- Cromer, R. F. (1971). The development of the ability to decenter in time. *British Journal of Psychology*, 62(3), 353–365.
- Czochralski, J. (1972). *Verbalaspekt und Tempussystem im Deutschen und Polnischen: Eine konfrontative Darstellung*. Wydawn: Uniwersytetu Warszawskiego.
- De Brabanter, P., Kissine, M., & Sharifzadeh, S. (2014). Future times, future tenses. In *Future Times, Future Tenses*. <https://doi.org/10.1093/acprof:oso/9780199679157.001.0001>
- De Knop, S. (2016). German causative events with placement verbs. *Lege Artis*, 1(1), 75–115. <https://doi.org/10.1515/lart-2016-0002>
- de Ruiter, L. E., Theakston, A. L., Brandt, S., & Lieven, E. V. M. (2018). Iconicity affects

- children's comprehension of complex sentences: The role of semantics, clause order, input and individual differences. *Cognition*, 171, 202–224.
<https://doi.org/10.1016/j.cognition.2017.10.015>
- de Swart, H. (2012). Verbal Aspect. In Robert Binnick (Ed.), *Oxford Handbook of Tense and Aspect* (pp. 752–780). Oxford, UK: Oxford University Press.
- de Swart, H., & Verkuyl, H. (1999). Tense and aspect in sentence and discourse. *ESSLLI Summer School*, (August). Retrieved from
<http://folli.loria.fr/cds/1999/library/pdf/hhh.pdf>
- de Villiers, J. G., Tager Flusberg, H. B., Hakuta, K., & Cohen, M. (1979). Children's comprehension of relative clauses. *Journal of Psycholinguistic Research*, 8(5), 499–518.
<https://doi.org/10.1007/BF01067332>
- Debus, D. (2014). 'Mental Time Travel': Remembering the Past, Imagining the Future, and the Particularity of Events. *Review of Philosophy and Psychology*, 5(3), 333–350.
<https://doi.org/10.1007/s13164-014-0182-7>
- Declerck, R. (1988). Sequence of tenses in English. *Folia Linguistica*, 14(3–4), 513–544.
- Del Prete, F. (2011). Indefinites in future tensed sentences. A novel argument for the modality of will. In M. Kissine, P. de Brabanter, & S. Sharifzadeh (Eds.), *Future Times, Future Tenses*. Oxford University Press.
- Demberg, V., & Sayeed, A. (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PLoS ONE*, 11(1), 1–29.
<https://doi.org/10.1371/journal.pone.0146194>
- Demirdache, H., & Lungu, O. (2008). Sequence of tense in (French) child language. *Linguistic Variation Yearbook*, 8, 101–130. <https://doi.org/10.1075/livy.8.04dem>
- Demirdache, H., & Uribe-Etxebarria, M. (1996). The syntax of temporal relations: A uniform approach to tense and aspect. *Linguistics Colloquium, University of British Columbia*.
- Depraetere, I. (1995). On the necessity of distinguishing between (un)boundedness and (a)telicity. *Linguistics and Philosophy*, 18(1), 1–19. <https://doi.org/10.1007/BF00984959>
- Depraetere, I. (1996). Foregrounding in relative clauses. *Linguistics*, 34(4), 699–731.
- Dery, J. E., & Bittner, D. (2016). Time and Causation in Discourse: Temporal Proximity, Implicit Causality, and Re-mention Biases. *Journal of Psycholinguistic Research*, 45(4), 883–899. <https://doi.org/10.1007/s10936-015-9382-2>
- Dery, J. E., & Koenig, J. P. (2015). A Narrative-Expectation-Based Approach to Temporal Update in Discourse Comprehension. *Discourse Processes*, 52(7), 559–584.
<https://doi.org/10.1080/0163853X.2014.966293>
- Diamond, N. B., & Levine, B. (2020). Linking Detail to Temporal Structure in Naturalistic-Event Recall. *Psychological Science*, 31(12), 1557–1572.
<https://doi.org/10.1177/0956797620958651>
- Dickey, M. W. (2000). *The processing of tense*. University of Massachusetts Amherst.
- Diessel, H. (2005). Competing motivations for the ordering of main and adverbial clauses. *Linguistics*, 43(3), 449–470. <https://doi.org/10.1515/ling.2005.43.3.449>

- Diessel, H. (2008). Iconicity of sequence: A corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics*, 19(3), 465–490. <https://doi.org/10.1515/COGL.2008.018>
- Diessel, H., & Tomasello, M. (2005). A New Look at the Acquisition of Relative Clauses. *Language Acquisition. Palgrave Advances in Linguistics*, 81(4), 882–906.
- Downes, J. J., Mayes, A. R., MacDonald, C., & Hunkin, N. M. (2002). Temporal order memory in patients with Korsakoff's syndrome and medial temporal amnesia. *Neuropsychologia*, 40(7), 853–861. [https://doi.org/10.1016/S0028-3932\(01\)00172-5](https://doi.org/10.1016/S0028-3932(01)00172-5)
- Dowty, D. R. (1972). *Studies in the logic of verb aspect and time reference in English*. University of Texas at Austin.
- Dowty, D. R. (1977). Toward a semantic analysis of verb aspect and the English “imperfective” progressive. *Linguistics and Philosophy*, 1(1), 45–77. <https://doi.org/10.1007/BF00351936>
- Dowty, D. R. (1979). *Word meaning and Montague Grammar: The semantics of verbs and times in Generative Semantics and in Montague's PTQ*. Dordrecht: Springer Dordrecht.
- Dowty, D. R. (1982). Tenses, Time Adverbs, and Compositional Semantic Theory. *Linguistics and Philosophy*, 5(1), 23–55.
- Dowty, D. R. (1986). The effects of aspectual class on the temporal structure of discourse: Semantics or pragmatics? *Linguistics and Philosophy*, 9(1), 37–61.
- Dowty, D. R. (1991). Thematic Proto-Roles and Argument Selection. *Language*, 67(3), 547. <https://doi.org/10.2307/415037>
- Dry, H. A. (1982). The Movement of Narrative Time. *Journal of Literary Semantics*, 12(2), 19–53. <https://doi.org/10.1515/jls-2021-2043>
- Eckardt, R. (2021). The Parameters of Indirect Speech. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, & T. E. Zimmermann (Eds.), *The Wiley Blackwell Companion to Semantics* (pp. 2213–2237). <https://doi.org/10.1002/9781118788516.sem032>
- Ehrlich, S. (1987). Aspect, foregrounding and point of view. *Text*, 7(4), 363–376. <https://doi.org/10.1515/text.1.1987.7.4.363>
- Eisenberg, S. (2002). Interpretation of relative clauses by young children: Another look. *Journal of Child Language*, 29(1), 177–188. <https://doi.org/10.1017/S0305000901004986>
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211. [https://doi.org/10.1016/0364-0213\(90\)90002-E](https://doi.org/10.1016/0364-0213(90)90002-E)
- Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, 33(4), 547–582. <https://doi.org/10.1111/j.1551-6709.2009.01023.x>
- Enç, M. (1986). Towards a Referential Analysis of Temporal Expressions. *Linguistics and Philosophy*, 9(4), 405–426.
- Enç, M. (1987). Anchoring Conditions for Tense. *Linguistic Inquiry*, 18(4), 633–657.
- Enç, M. (1996). Tense and Modality. In S. Lappin (Ed.), *The Handbook of Contemporary*

Semantic Theory. Blackwell Reference.

- Evcken, E., & Wittenberg, E. (2022). Making the Question Under Discussion explicit shifts counterfactual interpretation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44(44), 1855–1862.
- Evers-Vermeul, J., Hoek, J., & Scholman, M. C. J. (2017). On temporality in discourse annotation: Theoretical and practical considerations. *Dialogue and Discourse*, 8(2), 1–20. <https://doi.org/10.5087/dad.2017.201>
- Fagan, S. M. B. (1991). The semantics of the positional predicates liegen/legen, sitzen/setzen. *Die Unterrichtspraxis. Teaching German*, 24(2), 136–145. <https://doi.org/10.2307/3531018>
- Feagans, L. (1980). Children's understanding of some temporal terms denoting order, duration, and simultaneity. *Journal of Psycholinguistic Research*, 9(1), 41–57. <https://doi.org/10.1007/BF01067301>
- Filip, H. (1999). *Aspect, eventuality types and nominal reference*. Garland Publishing Inc.
- Filip, H. (2019). Aspectual class and Aktionsart. In P. Portner, K. von Stechow, & C. Maienborn (Eds.), *Semantics. Noun Phrases and Verb Phrases* (pp. 274–312). <https://doi.org/10.1515/9783110589443-009>
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, 8(10). <https://doi.org/10.1371/journal.pone.0077661>
- Flecken, M. (2011). Event conceptualization by early Dutch-German bilinguals: Insights from linguistic and eye-tracking data. *Bilingualism*, 14(1), 61–77. <https://doi.org/10.1017/S1366728910000027>
- Flecken, M., & Gerwien, J. (2013). Grammatical Aspect influences Event Duration Estimations: Evidence from Dutch. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 35. Retrieved from <https://escholarship.org/uc/item/4f50f5h0>
- Fleischman, S. (1990). *Tense and narrativity: From medieval performance to modern fiction*. <https://doi.org/10.7560/780903>
- Foppolo, F., Bosch, J. E., Greco, C., Carminati, M. N., & Panzeri, F. (2021). Draw a Star and Make it Perfect: Incremental Processing of Telicity. *Cognitive Science*, 45(10), 1–24. <https://doi.org/10.1111/cogs.13052>
- Fox, B. A., & Thompson, S. A. (1990). A Discourse Explanation of the Grammar of Relative Clauses in English Conversation. *Language*, 66(2), 297. <https://doi.org/10.2307/414888>
- Friedmann, N., & Costa, J. (2010). The child heard a coordinated sentence and wondered: On children's difficulty in understanding coordination and relative clauses with crossing dependencies. *Lingua*, 120(6), 1502–1515. <https://doi.org/10.1016/j.lingua.2009.10.006>
- Garey, H. B. (1957). Verbal Aspect in French. *Language*, 33(2), 91–110.
- Gautam, S., Suddendorf, T., Henry, J. D., & Redshaw, J. (2019). A taxonomy of mental time travel and counterfactual thought: Insights from cognitive development. *Behavioural Brain Research*, 374, 112108. <https://doi.org/10.1016/j.bbr.2019.112108>
- Gennari, S. P. (2003). Tense meanings and temporal interpretation. *Journal of Semantics*,

- 20(1), 35–71. <https://doi.org/10.1093/jos/20.1.35>
- Gennari, S. P. (2004). Temporal references and temporal relations in sentence comprehension. *Journal of Experimental Psychology: Learning Memory and Cognition*, 30(4), 877–890. <https://doi.org/10.1037/0278-7393.30.4.877>
- Gennari, S. P., & Poeppel, D. (2002). Events versus States: Empirical Correlates of Lexical Classes. *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society*, 1–6.
- Gennari, S. P., & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. *Cognition*, 89(1), B27–B41. [https://doi.org/10.1016/S0010-0277\(03\)00069-6](https://doi.org/10.1016/S0010-0277(03)00069-6)
- Gerwien, J., & von Stutterheim, C. (2018). Event segmentation: Cross-linguistic differences in verbal and non-verbal tasks. *Cognition*, 180, 225–237. <https://doi.org/10.1016/j.cognition.2018.07.008>
- Giannakidou, A., & Mari, A. (2012). Italian and Greek futures as epistemic operators. *Proceedings of CLS*, 48, 247–262.
- Gibson, E., & Fedorenko, E. (2010). Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences*, 14(6), 233–234. <https://doi.org/10.1016/j.tics.2010.03.005>
- Gibson, J. J. (1975). Events are perceivable but time is not. In J. T. Fraser & N. Lawrence (Eds.), *The Study of Time II* (pp. 295–301). https://doi.org/10.1007/978-3-642-50121-0_22
- Gill, K. (1993). On the Metaphysical Distinction Between Processes and Events. *Canadian Journal of Philosophy*, 23(3), 365–384. <https://doi.org/10.1080/00455091.1993.10717326>
- Gillis, M. M., Quinn, K. M., Phillips, P. A. T., & Hampstead, B. M. (2013). Impaired retention is responsible for temporal order memory deficits in mild cognitive impairment. *Acta Psychologica*, 143(1), 88–95. <https://doi.org/10.1016/j.actpsy.2013.03.001>
- Giorgi, A., & Pianesi, F. (1997). *Tense and aspect: From semantics to morphosyntax*. New York: Oxford University Press.
- Giorgi, A., & Pianesi, F. (2000). Sequence of Tense phenomena in Italian: A morphosyntactic analysis. *Probus*, 12(1), 1–32. <https://doi.org/10.1515/prbs.2000.12.1.1>
- Givón, T. (1983). *Topic continuity in discourse: The functional domain of switch-reference*. <https://doi.org/10.1075/tsl.2.06giv>
- Gleitman, L., & Papafragou, A. (2012). Relations between language and thought. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 504–523). Oxford, UK: Oxford University Press.
- Gleitman, L. R., Gleitman, H., Miller, C., & Ostrin, R. (1996). Similar, and similar concepts. *Cognition*, 58(3), 321–376. [https://doi.org/10.1016/0010-0277\(95\)00686-9](https://doi.org/10.1016/0010-0277(95)00686-9)
- Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, 26(1), 69–83. [https://doi.org/10.1016/0749-596X\(87\)90063-5](https://doi.org/10.1016/0749-596X(87)90063-5)
- Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of

- events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America*, 105(27), 9163–9168. <https://doi.org/10.1073/pnas.0710060105>
- Gomez, S., Jeffrey, M., Viviana, L., Gomez, S., Louis, S., & Dennis, S. (2025). More Than Boundaries: Exploring the Characteristics and Attributes of Daily Life Events. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 47, 3948–3956.
- Gorrell, P., Crain, S., & Fodor, J. D. (1989). Contextual information and temporal terms. *Journal of Child Language*, 16(3), 623–632. <https://doi.org/10.1017/S0305000900010758>
- Gouldthorp, B., Katsipis, L., & Mueller, C. (2018). An Investigation of the Role of Sequencing in Children’s Reading Comprehension. *Reading Research Quarterly*, 53(1), 91–106. <https://doi.org/10.1002/rrq.186>
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371–395. <https://doi.org/10.1037/0033-295x.101.3.371>
- Gualmini, A., Hulsey, S., Hacquard, V., & Fox, D. (2008). The question-answer requirement for scope assignment. *Natural Language Semantics*, 16(3), 205–237. <https://doi.org/10.1007/s11050-008-9029-z>
- Gundel, J. K. (1988). Universals of topic-comment structure. *Studies in Syntactic Typology/John Benjamins*.
- Gunkel, L. (2007). Demonstrativa als Antezedentien von Relativsätzen. *Deutsche Sprache*, 35(3), 213–238. <https://doi.org/10.37307/j.1868-775x.2007.03.04>
- Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition*, 175, 36–52. <https://doi.org/10.1016/j.cognition.2018.02.011>
- Hamburger, H., & Crain, S. (1982). Relative acquisition. In S. Kuczaj (Ed.), *Language development* (Vol. 1). Hillsdale, NJ: Erlbaum.
- Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology*, 110, 70–104. <https://doi.org/10.1016/j.cogpsych.2019.01.001>
- Han, J., Sun, L., Hu, X., Han, J., & Shao, L. (2014). Spatial and temporal visual attention prediction in videos using eye movement data. *Neurocomputing*, 145, 140–153. <https://doi.org/10.1016/j.neucom.2014.05.049>
- Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General*, 140(4), 586–604. <https://doi.org/10.1037/a0024310>
- Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, 28(10), 1474–1508. <https://doi.org/10.1080/01690965.2012.689305>
- Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (2005). *The World Atlas of Language Structures*. Oxford, UK: Oxford University Press.
- Heck, R. G. (2006). MacFarlane on relative truth. *Philosophical Issues*, 16(1).

- Helbig, G., & Buscha, J. (2001). *Deutsche Grammatik*. Berlin, Munich: Langenscheidt KG.
- Hermann, M. (2019). The semantics of German posture and placement verbs in noun-verb phrases. *Yearbook of the German Cognitive Linguistics Association*, 7(1), 93–114. <https://doi.org/10.1515/gcla-2019-0007>
- Herweg, M. (1991). Perfective and imperfective aspect and the theory of events and states. *Linguistics*, 29(6), 969–1010. <https://doi.org/10.1515/ling.1991.29.6.969>
- Hespos, S. J., Grossman, S. R., & Saylor, M. M. (2010). Infants' Ability to Parse Continuous Actions: Further Evidence. *Neural Networks*, 23(8–9), 1026–1032. <https://doi.org/10.1016/j.neunet.2010.07.010>
- Hespos, S. J., Saylor, M. M., & Grossman, S. R. (2009). Infants' Ability to Parse Continuous Actions. *Developmental Psychology*, 45(2), 575–585. <https://doi.org/10.1037/a0014145>
- Higginbotham, J. (2006). The Anaphoric Theory of Tense. *Semantics and Linguistic Theory*, 16, 59. <https://doi.org/10.3765/salt.v16i0.2949>
- Hindy, N. C., Altmann, G. T. M., Kalenik, E., & Thompson-Schill, S. L. (2012). The effect of object state-changes on event processing: Do objects compete with themselves? *Journal of Neuroscience*, 32(17), 5795–5803. <https://doi.org/10.1523/jneurosci.6294-11.2012>
- Hinrichs, E. (1986). Temporal anaphora in discourses of English. *Linguistics and Philosophy*, 9(1), 63–82.
- Hoek, J., Rohde, H., Evers-Vermeul, J., & Sanders, T. J. M. (2021). Expectations from relative clauses: Real-time coherence updates in discourse processing. *Cognition*, 210, 1–18. <https://doi.org/10.1016/j.cognition.2020.104581>
- Hoekstra, T. (1992). Aspect and Theta Theory. In I. M. Roca (Ed.), *Its Role in Grammar* (pp. 145–174). <https://doi.org/doi:10.1515/9783110872613.145>
- Hofherr, P. C. (2014). Reduced definite articles with restrictive relative clauses. In *Syntax and Semantics* (Vol. 39). https://doi.org/10.1163/9789004260825_008
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24(5), 849–878. <https://doi.org/10.1017/S0140525X01000103>
- Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*.
- Hopper, P. J. (1979). Aspect and foregrounding in discourse. In Talmy Givón (Ed.), *Discourse and Syntax* (pp. 211–241). https://doi.org/10.1163/9789004368897_010
- Hopper, P. J., & Thompson, S. A. (1980). Transitivity in Grammar and Discourse. *Language*, 56(2), 251–299.
- Horchak, O. V., & Garrido, M. V. (2022). Simulating background settings during spoken and written sentence comprehension. *Psychonomic Bulletin and Review*, 29(4), 1426–1439. <https://doi.org/10.3758/s13423-022-02061-9>
- Horn, L. R. (2022a). Contrast and clausal order: Beyond Behaghel. *Language*, 98(4), 812–843. <https://doi.org/10.1353/lan.2022.0022>
- Horn, L. R. (2022b). Contrast and clausal order: On beyond Behaghel. *Language*, 98(4), 812–

843. <https://doi.org/10.1353/lan.2022.0022>
- Huddleston, R., & Pullum, G. (2005). The Cambridge grammar of the English language. *Zeitschrift Für Anglistik Und Amerikanistik*, 53(2), 193–194.
- Huettig, F. (2025). *Looking Ahead: The New Science of the Predictive Mind*. Cambridge University Press.
- Huff, M., Meitz, T. G. K., & Papenmeier, F. (2014). Changes in Situation Models Modulate Gaze Behavior, Memory and Prediction Performance in Motion Pictures Conference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/dx.doi.org/10.1037/a0036780>
- Husband, E. M. (2010). *On the compositional nature of stativity*. Michigan State University.
- Jackendoff, R. (1990). *Semantic structures*. The MIT Press.
- Jackendoff, R. (1991). Parts and boundaries. *Cognition*, 41(1–3), 9–45. [https://doi.org/10.1016/0010-0277\(91\)90031-X](https://doi.org/10.1016/0010-0277(91)90031-X)
- Jakobson, R. (1965). Quest for the essence of language. *Diogenes*, 13(51), 21–37. <https://doi.org/10.1177/039219216501305103>
- Jardón, N., Marx, E., & Wittenberg, E. (2025). *Is there a perfect state ? Experimental evidence from English and Spanish for the perfect-as-state hypothesis*. 10(1), 1–34.
- Ji, Y., & Papafragou, A. (2020). Is there an end in sight? Viewers' sensitivity to abstract event structure. *Cognition*, 197(April 2019). <https://doi.org/10.1016/j.cognition.2020.104197>
- Ji, Y., & Papafragou, A. (2022). Boundedness in event cognition: Viewers spontaneously represent the temporal texture of events. *Journal of Memory and Language*, 127(June), 104353. <https://doi.org/10.1016/j.jml.2022.104353>
- Johanson, L. (1998). The Structure of Turkic. In L. Johanson & É. Á. Csató (Eds.), *The Turkic Languages* (pp. 30–66). <https://doi.org/10.4324/9780203066102.ch3>
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Cambridge University Press.
- Jou, J., & Harris, R. J. (1990). Event order versus syntactic structure in recall of adverbial complex sentences. *Journal of Psycholinguistic Research*, 19(1), 21–42. <https://doi.org/10.1007/BF01068183>
- 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. <https://doi.org/10.1080/0163853X.2018.1522913>
- Kameyama, M., Passonneau, R., & Poesio, M. (1993). Temporal centering. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 70–77. <https://doi.org/10.3115/981574.981584>
- Kamp, H. (1980). Some remarks on the logic of change, Part I. In C. Rohrer (Ed.), *Time, Tense, and Quantifiers: Proceedings of the Stuttgart Conference on the Logic of Tense and Quantification*, (pp. 135–180). Berlin, New York: Max Niemeyer Verlag.
- Kamp, H., & Reyle, U. (1993). Tense and aspect. In H. Kamp & U. Reyle (Eds.), *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language*,

- Formal Logic and Discourse Representation Theory* (pp. 483–689). Dordrecht: Springer.
- Kamp, H., & Rohrer, C. (1983). Tense in texts. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, Use, and Interpretation of Language* (pp. 250–269). <https://doi.org/10.1515/9783110852820.250>
- Kang, X., Eerland, A., Joergensen, G. H., Zwaan, R. A., & Altmann, G. T. M. (2020). The influence of state change on object representations in language comprehension. *Memory and Cognition*, 48(3), 390–399. <https://doi.org/10.3758/s13421-019-00977-7>
- Katz, G. (2003). On the stativity of the English perfect. In A. Alexiadou, M. Rathert, & A. von Stechow (Eds.), *Perfect Explorations* (pp. 205–234). Berlin: Mouton de Gruyter.
- Kausler, D. H., Salthouse, T. A., Sauls, J. S., The, S., Journal, A., & Summer, N. (1988). Temporal memory over the adult lifespan. *The American Journal of Psychology*, 101(2), 207–215.
- Kearns, K. (2000). *Semantics*. New York: St. Martin's Press.
- Kehler, A. (1994). Temporal relations: Reference or discourse coherence? *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 319–321. <https://doi.org/10.3115/981732.981779>
- Kehler, A. (2002). *Coherence, reference and the theory of grammar*. CSLI Publications.
- Kehler, A. (2006). Discourse coherence. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 241–265). <https://doi.org/10.1002/9780470756959.ch11>
- Keller, F. (2000). *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. (University of Edinburgh). <https://doi.org/10.7282/T3GQ6WMS>
- Kenny, A. (1963). *Action, Emotion And Will*. London, New York: Routledge.
- Kepser, S., & Reis, M. (2005). *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*. <https://doi.org/10.1515/9783110197549>
- Kissine, M. (2008). From predictions to promises: How to derive deontic commitment. *Pragmatics & Cognition*, 16(3), 471–491.
- Klein, W. (1994). *Time in language*. New York, NY: Routledge.
- Klein, W. (2000). An analysis of the German perfekt. *Language*, 76(2), 358–382. <https://doi.org/10.1353/lan.2000.0140>
- Klein, W. (2008a). The topic situation. In *Empirische Forschung und Theoriebildung: Beiträge aus Soziolinguistik, Gesprochene-Sprache-und Zweitspracherwerbsforschung: Festschrift für Norbert Dittmar* (pp. 287–305).
- Klein, W. (2008b). Time in Language, Language in Time. In P. Indefrey & M. Gullberg (Eds.), *Time to Speak: Cognitive and Neural Prerequisites for Time in Language* (pp. 1–12). <https://doi.org/10.1002/9781444309645.ch1>
- Klein, W. (2009). How time is encoded. *The Expression of Time*, (1), 39–82. <https://doi.org/10.1515/9783110199031.39>
- Klimek-Jankowska, D., & Błaszczak, J. (2020). How Incremental is the Processing of Perfective and Imperfective Aspect in Polish ? An Exploratory Event-Related Potential

- Study. *Journal of Slavic Languages*, 28(1), 23–70.
- Koenig, J. P. (2015). Event semantics. In G. Russell & D. Graff Fara (Eds.), *The Routledge Handbook of Semantics* (pp. 387–402). https://doi.org/10.4324/9780203206966.ch2_8
- Koffka, K. (1935). *Principles of Gestalt psychology*. London, UK: Routledge.
- Kozłowska-Raś, R. (1987). The Categories of Tense, Aspect, and Aktionsart in Polish and Swedish. *Nordic Journal of Linguistics*, 10(2), 165–179. <https://doi.org/10.1017/S0332586500001645>
- Kratzer, A. (1995). Stage-level and Individual-level predicates. In G. Carlson & F. J. Pelletier (Eds.), *The Generic Book* (pp. 125–175). Chicago: University of Chicago Press.
- Kratzer, A. (2004). Interpreting focus: Presupposed or expressive meanings? *Theoretical Linguistics*, 30(1), 1–17.
- Krause, O. (2002). Progressiv im Deutschen. Eine empirische Untersuchung im Kontrast mit Niederländisch und Englisch. In H. Altmann, P. Blumenthal, H. J. Heringer, I. Plag, H. Vater, & R. Wiese (Eds.), *Linguistische Arbeiten* (Vol. 4). <https://doi.org/10.30965/25890859-00403010>
- Krifka, M. (1992). Thematic Relations as Links between Nominal Reference and Temporal Constitution. *Lexical Matters*, 29–53.
- Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, 55(3–4), 243–276. <https://doi.org/10.1556/ALing.55.2008.3-4.2>
- Kubota, Y., Lee, J., Smirnova, A., & Tonhauser, J. (2009). On the Cross-Linguistic Interpretation of Embedded Tenses. In A. Riester & T. Solstad (Eds.), *Proceedings of Sinn und Bedeutung* (Vol. 13, pp. 307–320). Universität Stuttgart.
- Kuperberg, G. R. (2021). Tea With Milk? A Hierarchical Generative Framework of Sequential Event Comprehension. *Topics in Cognitive Science*, 13(1), 256–298. <https://doi.org/10.1111/tops.12518>
- Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12(2), 72–79. <https://doi.org/10.1016/j.tics.2007.11.004>
- Kurby, C. A., & Zacks, J. M. (2011). Age differences in the perception of hierarchical structure in events. *Memory and Cognition*, 39(1), 75–91. <https://doi.org/10.3758/s13421-010-0027-2>
- Kusumoto, K. (1999). *Tense in embedded contexts*. University of Massachusetts Amherst.
- Lakoff, G. (1966). Stative adjectives and verbs in English. *Harvard Computational Laboratory Report NSF-17*. Cambridge, MA.
- Lambrecht, K. (1988). There Was a Farmer Had a Dog: Syntactic Amalgams Revisited. *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics*, 319–339. <https://doi.org/10.3765/bls.v14i0.1796>
- Landmann, A. (2011). *Kirgisisch. Kurzgrammatik*. Harrassowitz Verlag.
- Lapaire, J.-R., & Rotgé, W. (1998). *Linguistique et grammaire de l'anglais*. Presses Univ. du Mirail.
- Lascarides, A., & Asher, N. (1991). Discourse relations and defeasible knowledge.

- Proceedings of the Annual Meeting of the Association for Computational Linguistics*, (1), 55–62. <https://doi.org/10.3115/981344.981352>
- Lascarides, A., & Asher, N. (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy*, 16, 437–493. <https://doi.org/10.1007/BF00986208>
- Lashley, K. S. (1951). The problem of serial order in behavior. *Cerebral Mechanisms in Behavior*, (7), 112–136.
- Lassiter, D. (2023). Decomposing relevance in conditionals. *Mind and Language*, 38(3), 644–668. <https://doi.org/10.1111/mila.12418>
- Laurinavichyute, A., & von der Malsburg, T. (2024). Agreement attraction in grammatical sentences and the role of the task. *Journal of Memory and Language*, 137, 104525. <https://doi.org/10.31234/osf.io/n75vc>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.
- Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- Levin, B., & Hovav, M. R. (1994). *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge, MA: The MIT Press.
- Levin, B., & Rappaport Hovav, M. (2005). *Argument Realization*. Cambridge, MA: Cambridge University Press.
- Levine, D., Buchsbaum, D., Hirsh-Pasek, K., & Golinkoff, R. M. (2019). Finding Events in a Continuous World: A Developmental Account. *Developmental Psychobiology*, 61(3), 376–389. <https://doi.org/10.1002/dev.21804>
- Levine, D. F., Hirsh-Pasek, K., Pace, A. E., & Golinkoff, R. M. (2017). A Goal Bias in Action: The Boundaries Adults Perceive in Events Align With Sites of Actor Intent. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(6), 916–927. <https://doi.org/doi.org/10.1037/xlm0000364>
- Levinson, S. C. (1983). *Pragmatics*. London, UK: Cambridge University Press.
- Levinson, S. C. (1996). Language and space. *Annual Review of Anthropology*, 25(1), 353–382.
- Li, F., Du, J., & Wolff, P. (2015). The Linguistic Representations of Causing Events and Caused Events in Narrative Discourse. *Cognitive Semantics*, 1(1), 45–76. <https://doi.org/10.1163/23526416-00101002>
- Lin, J. W. (2003). Temporal Reference in Mandarin Chinese. *Journal of East Asian Linguistics*, 12(3), 259–311. <https://doi.org/10.1023/A:1023665301095>
- Lin, J. W. (2006). Time in a language without tense: The case of Chinese. *Journal of Semantics*, 23(1), 1–53. <https://doi.org/10.1093/jos/ffh033>
- Lyons, J. (1977). *Semantics: Volume 2*. London, UK: Cambridge University Press.
- Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory and Cognition*, 31(5), 663–672.

<https://doi.org/10.3758/BF03196106>

- Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing Space and Time in Film Understanding. *Applied Cognitive Psychology*, 15(5), 533–545.
<https://doi.org/10.1002/acp.724>
- Magliano, J. P., Radvansky, G. A., Forsythe, J. C., & Copeland, D. E. (2014). Event segmentation during first-person continuous events. *Journal of Cognitive Psychology*, 26(6), 649–661. <https://doi.org/10.1080/20445911.2014.930042>
- Magliano, J. P., & Schleich, M. C. (2000). Verb aspect and situation models. *Discourse Processes*, 29(2), 83–112. https://doi.org/10.1207/S15326950dp2902_1
- Maienborn, Claudia. (2007). On Davidsonian and Kimian States. In I. Comorovski & K. von Heusinger (Eds.), *Existence: Semantics and Syntax* (pp. 107–130). Springer.
- Maienborn, Claudia. (2011). Event semantics. In C. Maienborn, K. Heusinger, & P. Portner (Eds.), *Semantics* (Vol. 1, pp. 802–829). https://doi.org/10.4324/9780203206966.ch2_8
- Malaia, E., Wilbur, R. B., & Weber-Fox, C. (2012). Effects of Verbal Event Structure on Online Thematic Role Assignment. *Journal of Psycholinguistic Research*, 41(5), 323–345. <https://doi.org/10.1007/s10936-011-9195-x>
- Mandler, J. M. (1986). On the comprehension of temporal order. *Language and Cognitive Processes*, 1(2), 309–320.
- Marty, P., Chemla, E., & Sprouse, J. (2020). The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa: A Journal of General Linguistics*, 5(1), 72.
- Marx, E., Iwan, O., & Wittenberg, E. (2023). When you say that, do you mean it? Crosslinguistic patterns of anaphor resolution in English, German, and Polish. *Proceedings of the 45th Annual Meeting of the Cognitive Science Society*, 2581–2588.
- Marx, E., Iwan, O., & Wittenberg, E. (2025). Dynamicity predicts inferred temporal order in complex sentences: Evidence from English, German, and Polish. *Cognitive Science*, 49, e70038. <https://doi.org/10.1111/cogs.70038>
- Marx, E., & Wittenberg, E. (2022). Event structure predicts temporal interpretation of English and German past-under-past relative clauses. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 439–445.
- Marx, E., & Wittenberg, E. (2024). Eventuality type predicts temporal order inferences in discourse comprehension. *Glossa Psycholinguistics*, 3, 1–24.
<https://doi.org/10.5070/G60116579>
- Marx, E., & Wittenberg, E. (2025). Temporal construal in sentence comprehension depends on linguistically encoded event structure. *Cognition*, 254, 105975.
<https://doi.org/10.1016/j.cognition.2024.105975>
- Mathis, A., & Papafragou, A. (2022). Agents' goals affect construal of event endpoints. *Journal of Memory and Language*, (27), 104373.
- Mayrhofer, R., & Waldmann, M. R. (2014). Indicators of causal agency in physical interactions: The role of the prior context. *Cognition*, 132(3), 485–490.
<https://doi.org/10.1016/j.cognition.2014.05.013>

- McCormack, T., & Hoerl, C. (2008). Temporal Decentering and the Development of Temporal Concepts. *Language Learning*, 58, 89–113. <https://doi.org/10.1002/9781444309645.ch8>
- McKee, C., Snedeker, J., & McDaniel, D. (1998). Relatives children say. *Journal of Psycholinguistic Research*, 27(5), 573–596.
- McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Linguistics and Language Compass*, 3(6), 1417–1429. <https://doi.org/10.1111/j.1749-818X.2009.00174.x>
- Michaelis, L. A. (2011). Stative by construction. *Linguistics*, 49(6), 1359–1399. <https://doi.org/10.1515/LING.2011.038>
- Milczarski, W. (2021). Grammatical aspect in Polish and the perception of event duration. In *Beiträge zur allgemeinen und vergleichenden Sprachwissenschaft* (Vol. 10, pp. 71–96). <https://doi.org/10.23817/bzspr.10-4>
- Miller, D., DeLuca, V., Berndt, D., Iverson, M., & Rothman, J. (2017). Subtle aspectual differences in the L2 acquisition of German. In E. Blom, L. Cornips, & J. Schaeffer (Eds.), *Cross-linguistic Influence in Bilingualism: In Honor of Aafke Hulk*. (pp. 207–230). <https://doi.org/10.1075/sibil.52.10mil>
- Milsark, G. L. (1974). *Existential Sentences in English*. Massachusetts Institute of Technology.
- Minor, S., Mitrofanova, N., Guajardo, G., Vos, M., & Ramchand, G. (2023). Aspect processing across languages: A visual world eye-tracking study. *Frontiers in Language Sciences*, 1, 1–14. <https://doi.org/10.3389/flang.2022.1052205>
- Minor, S., Mitrofanova, N., & Ramchand, G. (2022). Fine-grained time course of verb aspect processing. *PLoS ONE*, 17, 1–20. <https://doi.org/10.1371/journal.pone.0264132>
- Misersky, J., Slivac, K., Hagoort, P., & Flecken, M. (2021). The state of the onion: Grammatical aspect modulates object representation during event comprehension. *Cognition*, 214, 104744. <https://doi.org/10.1016/j.cognition.2021.104744>
- Młynarczyk, A. (2004). *Aspectual Pairing in Polish* (Universiteit Utrecht). Retrieved from <https://api.semanticscholar.org/CorpusID:119008542>
- Moens, M. (1987). *Tense, Aspect and Temporal Reference*. University of Edinburgh.
- Moens, M., & Steedman, M. (1988). Temporal onotology and temporal reference. *Computational Linguistics*, 14(2), 15–28.
- Monroy, C. D., Gerson, S. A., & Hunnius, S. (2018). Translating visual information into action predictions: Statistical learning in action and nonaction contexts. *Memory and Cognition*, 46(4), 600–613. <https://doi.org/10.3758/s13421-018-0788-6>
- Morgan, A. M., von der Malsburg, T., Ferreira, F., & Wittenberg, E. (2020). Shared syntax between comprehension and production: Multi-paradigm evidence that resumptive pronouns hinder comprehension. *Cognition*, 205, 104417. <https://doi.org/10.1016/j.cognition.2020.104417>
- Mourelatos, A. P. D. (1978). Events, processes, and states. *Linguistics and Philosophy*, 2, 415–434.

- Mucha, A. (2012). Temporal reference in a genuinely tenseless language: The case of Hausa. *Semantics and Linguistic Theory*, 22, 188–207. <https://doi.org/10.3765/salt.v22i0.3084>
- Mucha, A., Renans, A., & Romoli, J. (2023). Sequence of tense and cessation implicatures: Evidence from Polish. *Natural Language and Linguistic Theory*, 41(1), 267–346. <https://doi.org/10.1007/s11049-022-09545-2>
- Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, 395, 71–73. <https://doi.org/10.1038/25731>
- Nagai, Y. (2019). Predictive learning: Its key role in early cognitive development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1771). <https://doi.org/10.1098/rstb.2018.0030>
- Natsopoulos, D., Mentenopoulos, G., Bostantzopoulou, S., Katsarou, Z., Grouios, G., & Logothetis, J. (1991). Understanding of relational time terms before and after in Parkinsonian patients. *Brain and Language*, 40(4), 444–458. [https://doi.org/10.1016/0093-934x\(91\)90142-n](https://doi.org/10.1016/0093-934x(91)90142-n)
- Ness, T., & Meltzer-Asscher, A. (2021). Rational adaptation in lexical prediction: The influence of prediction strength. *Frontiers in Psychology*, 12, 1–15. <https://doi.org/10.3389/fpsyg.2021.622873>
- Newton, D. (1973). Attribution and the Unit of Perception of Ongoing Behavior. *Journal of Personality and Social Psychology*, 28(1), 28–38.
- Nieuwland, M. S. (2015). The Truth Before and After: Brain Potentials Reveal Automatic Activation of Event Knowledge during Sentence Comprehension. *Journal of Cognitive Neuroscience*, 27(11), 2215–2228. <https://doi.org/10.1162/jocn>
- Nobre, A. C., & Van Ede, F. (2018). Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience*, 19(1), 34–48. <https://doi.org/10.1038/nrn.2017.141>
- Núñez, R., & Cooperrider, K. (2013). The tangle of space and time in human cognition. *Trends in Cognitive Sciences*, 17(5), 220–229. <https://doi.org/10.1016/j.tics.2013.03.008>
- Nuthmann, A., & Van Der Meer, E. (2005). Time's arrow and pupillary response. *Psychophysiology*, 42(3), 306–317. <https://doi.org/10.1111/j.1469-8986.2005.00291.x>
- O'Reilly, J. X., McCarthy, K. J., Capizzi, M., & Nobre, A. C. (2008). Acquisition of the temporal and ordinal structure of movement sequences in incidental learning. *Journal of Neurophysiology*, 99(5), 2731–2735. <https://doi.org/10.1152/jn.01141.2007>
- Ogihara, T. (1995). The Semantics of Tense in Embedded Clauses. *Linguistic Inquiry*, 26(4), 663–679.
- Ogihara, T. (1996). *Tense, Attitudes and Scope*. Springer Dordrecht.
- Ogihara, T. (2007). Tense and aspect in truth-conditional semantics. *Lingua*, 117(2), 392–418. <https://doi.org/10.1016/j.lingua.2005.01.002>
- Ogihara, T. (2015). Relative tense in relative clauses. *Journal of Japanese Linguistics*, 31(1), 31–54. <https://doi.org/10.1515/jjl-2015-0103>
- Ogihara, T., & Sharvit, Y. (2012). Embedded Tenses. In R. Binnick (Ed.), *The Oxford handbook of Tense and Aspect* (pp. 638–668). <https://doi.org/doi.org/10.1093/oxfordhb/9780195381979.001.0001>

- Ohtsuka, K., & Brewer, W. F. (1992). Discourse Organization in the Comprehension of Temporal Order in Narrative Texts. *Discourse Processes*, 15(3), 317–336. <https://doi.org/10.1080/01638539209544815>
- Olsen, M. B. (1994). *The Semantics and Pragmatics of Lexical and Grammatical Aspect*. <https://doi.org/10.4324/9781315052267>
- Olson, I. R., & Chun, M. M. (2001). Temporal Contextual Cuing of Visual Attention. *Journal of Experimental Psychology: Learning Memory and Cognition*, 27(5), 1299–1313. <https://doi.org/10.1037/0278-7393.27.5.1299>
- Ongchoco, J. D. K., Yates, T. S., & Scholl, B. J. (2023). Event segmentation structures temporal experience: Simultaneous dilation and contraction in rhythmic reproductions. *Journal of Experimental Psychology: General*, 152(11), 3266–3276. <https://doi.org/10.1037/xge0001447>
- Oversteegen, L. (2005). Causality and tense - Two temporal structure builders. *Journal of Semantics*, 22(3), 307–337. <https://doi.org/10.1093/jos/ffh021>
- Pace, A., Levine, D. F., Golinkoff, R. M., Carver, L. J., & Hirsh-Pasek, K. (2020). Keeping the End in Mind: Preliminary Brain and Behavioral Evidence for Broad Attention to Endpoints in Pre-Linguistic Infants. *Infant Behavior and Development*, 58, 101425. <https://doi.org/10.1016/j.infbeh.2020.101425>
- Paczynski, M., Jackendoff, R., & Kuperberg, G. (2014). When Events Change Their Nature: The Neurocognitive Mechanisms underlying Aspectual Coercion. *Journal of Cognitive Neuroscience*, 26(9), 1905–1917. https://doi.org/10.1162/jocn_a_00638
- Palmer, F. R. (1974). *The English Verb*. London: Longman.
- Palmer, F. R. (1987). *Mood and modality*. Cambridge University Press.
- Parsons, T. (1989). The Progressive in English: Events, States and Processes. *Linguistics*, 12(2), 213–241.
- Parsons, T. (1990). *Events in the Semantics of English: A Study in Subatomic Semantics*. MIT Press.
- Parsons, T. (2002). Eventualities and narrative progression. *Linguistics and Philosophy*, 25(5–6), 681–699. <https://doi.org/10.1023/a:1020863700064>
- Partee, B. H. (1973). Some structural analogies between tenses and pronouns in English. *The Journal of Philosophy*, 70(18), 601–609. <https://doi.org/10.2307/2025024>
- Partee, B. H. (1984). Nominal and temporal anaphora. *Linguistics and Philosophy*, 7(3), 243–286.
- Pawlak, A., Oehlrach, J. S., & Weist, R. M. (2006). Reference Time in Child English and Polish. *First Language*, 26(3), 281–297. <https://doi.org/10.1177/0142723706059447>
- Perry, J. (2006). How real are future events? In F. Stadler & M. Stoltzner (Eds.), *Time and History* (pp. 13–30). Frankfurt: Ontos verlag.
- Pettijohn, K. A., & Radvansky, G. A. (2016). Narrative event boundaries, reading times, and expectation. *Memory and Cognition*, 44(7), 1064–1075. <https://doi.org/10.3758/s13421-016-0619-6>

- Pickering, M. J., McElree, B., Frisson, S., Chen, L., & Traxler, M. J. (2006). Underspecification and aspectual coercion. *Discourse Processes*, 42(2), 131–155. https://doi.org/10.1207/s15326950dp4202_3
- Piñango, M. M., Winnick, A., Ullah, R., & Zurif, E. (2006). Time-course of semantic composition: The case of aspectual coercion. *Journal of Psycholinguistic Research*, 35(3), 233–244. <https://doi.org/10.1007/s10936-006-9013-z>
- Piñón, C. (1995). *An ontology of event semantics*. Stanford University.
- Piñón, C. (1997). Achievements in an Event Semantics. *Semantics and Linguistic Theory*, 7, 276. <https://doi.org/10.3765/salt.v7i0.2781>
- Polanyi Bowditch, L. (1976). Why the Whats are When: Mutually Contextualizing Realms of Narrative. *Proceedings of the 2nd Annual Meeting of the Berkeley Linguistics*, 59–77.
- Potts, C. (2005). *The Logic of Conventional Implicatures*. <https://doi.org/doi.org/10.1093/acprof:oso/9780199273829.001.0001>
- Potts, T., & Taylor, C. C. (1965). States, Activities and Performances. *Proceeding of the Aristotelian Society*, 39, 65–102.
- Pregla, D., Lissón, P., Vasishth, S., Burchert, F., & Stadie, N. (2021). Variability in sentence comprehension in aphasia in German. *Brain and Language*, 222. <https://doi.org/10.1016/j.bandl.2021.105008>
- Prior, A. N. (1967). *Time and modality*. Oxford, UK: Oxford University Press.
- Pustejovsky, J. (1991). The syntax of event structure. *Cognition*, 41, 47–81. [https://doi.org/10.1016/0010-0277\(91\)90032-Y](https://doi.org/10.1016/0010-0277(91)90032-Y)
- Pyykkönen, P., & Järvikivi, J. (2012). Children and Situation Models of Multiple Events. *Developmental Psychology*, 48(2), 521–529. <https://doi.org/10.1037/a0025526>
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Grammar of Contemporary English* (Vol. 58). Essex: Pearson Longman.
- R. Core, T. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Radvansky, G. A. (2012). Across the Event Horizon. *Current Directions in Psychological Science*, 21(4), 269–272. <https://doi.org/10.1177/0963721412451274>
- Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: Situation models and experienced space. *Memory and Cognition*, 34(5), 1150–1156. <https://doi.org/10.3758/BF03193261>
- 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), 210–216. <https://doi.org/10.1037/a0017258>
- Radvansky, G. A., Tamplin, A. K., & Krawietz, S. A. (2010). Walking through doorways causes forgetting: Environmental integration. *Psychonomic Bulletin and Review*, 17(6), 900–904. <https://doi.org/10.3758/PBR.17.6.900>
- Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(6), 608–620. <https://doi.org/10.1002/wcs.133>

- Radvansky, G. A., & Zacks, J. M. (2014). *Event Cognition*.
<https://doi.org/10.1093/acprof:oso/9780199898138.001.0001>
- Radvansky, G. A., Zwaan, R. A., Federico, T., & Franklin, N. (1998). Retrieval From Temporally Organized Situation Models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(5), 1224–1237.
- Ramchand, G. C. (2008). Verb Meaning and the Lexicon. *Verb Meaning and the Lexicon*, (March). <https://doi.org/10.1017/cbo9780511486319>
- Ramsay, V. (1987). The functional distribution of preposed and postposed ‘if’ and ‘when’ clauses in written discourse. *Coherence and Grounding in Discourse*, 11, 383–409.
- Rappaport Hovav, M., & Levin, B. (1998). Building verb meanings. In M. Butt & W. Geuder (Eds.), *The projection of arguments: Lexical and compositional factors* (pp. 97–134). Stanford, CA: CSLI Publications.
- Real, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57(1), 1–23.
<https://doi.org/10.1016/j.jml.2006.08.014>
- Redshaw, J. (2024). Remembered Futures and Anticipated Pasts: The Recursive Grammar of Mental Time Travel. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 46, 1433–1441.
- Reichenbach, H. (1947). *Elements of Symbolic Logic*. New York, NY: The Macmillan Company.
- Reinhart, T. (1981). Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, 1(1), 53–94.
- Reinhart, T. (1984). Principles of gestalt perception in the temporal organization of narrative texts. *Linguistics*, 22(6), 779–810. <https://doi.org/10.1515/ling.1984.22.6.779>
- Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, 7(1–3), 17–42. <https://doi.org/10.1080/135062800394667>
- Rinck, M., Gámez, E., Díaz, J. M., & De Vega, M. (2003). Processing of temporal information: Evidence from eye movements. *Memory and Cognition*, 31(1), 77–86.
<https://doi.org/10.3758/BF03196084>
- Rinck, M., Hähnel, A., & Becker, G. (2001). Using Temporal Information to Construct, Update, and Retrieve Situation Models of Narratives. *Journal of Experimental Psychology: Learning Memory and Cognition*, 27(1), 67–80.
<https://doi.org/10.1037/0278-7393.27.1.67>
- Rinck, M., & Weber, U. (2003). Who when where: An experimental test of the event-indexing model. *Memory & Cognition*, 31(8), 1284–1292.
<https://doi.org/10.3758/bf03195811>
- Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar*. Kluwer International Handbooks of Linguistics. Dordrecht: Springer.
- Rohde, H., Levy, R., & Kehler, A. (2011). Anticipating explanations in relative clause processing. *Cognition*, 118(3), 339–358. <https://doi.org/10.1016/j.cognition.2010.10.016>

- Rosen, S. (1999). The syntactic representation of linguistic events. *Glott International*, 323–366. <https://doi.org/10.1515/9783110890952.323>
- Rothstein, S. (2004). *Structuring Events*. Malden, MA: Blackwell Publishing Inc.
- Rothstein, S. (2013). Two puzzles for a theory of lexical aspect: Semelfactives and degree achievements. In J. Dölling, T. Heyde-Zybatow, & M. Schäfer (Eds.), *Event Structures in Linguistic Form and Interpretation* (pp. 175–198). <https://doi.org/10.1515/9783110925449.175>
- Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2020). Reward prediction errors create event boundaries in memory. *Cognition*, 203, 104269. <https://doi.org/10.1016/j.cognition.2020.104269>
- Royden, C. S., Wolfe, J. M., & Klempe, N. (2001). Visual search asymmetries in motion and optic flow fields. *Perception and Psychophysics*, 63(3), 436–444. <https://doi.org/10.3758/BF03194410>
- Ryle, G. (1949). *The Concept of Mind*. London, New York: Routledge.
- Sacks, O. (1985). The Lost Mariner. In O. Sacks (Ed.), *The man who mistook his wife for a hat and other clinical tales*. New York: Summit Books.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928. <https://doi.org/10.1126/science.274.5294.1926>
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1), 27–52. [https://doi.org/10.1016/S0010-0277\(98\)00075-4](https://doi.org/10.1016/S0010-0277(98)00075-4)
- 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). <https://doi.org/10.1111/cogs.12708>
- Santin, M., van Hout, A., & Flecken, M. (2021). Event endings in memory and language. *Language, Cognition and Neuroscience*, 36(5), 625–648. <https://doi.org/10.1080/23273798.2020.1868542>
- 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. <https://doi.org/10.1016/j.cognition.2013.07.002>
- Sasanuma, S., & Kamio, A. (1976). Aphasic's comprehension of sentences expressing temporal order of events. *Brain and Language*, 3(4), 495–506. [https://doi.org/10.1016/0093-934x\(76\)90044-4](https://doi.org/10.1016/0093-934x(76)90044-4)
- Sasse, H. J. (2002). Recent activity in the theory of aspect: Accomplishments, achievements, or just non-progressive state? *Linguistic Typology*, 6(2), 199–271. <https://doi.org/10.1515/lity.2002.007>
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.
- Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-, N. B., Botvinick, M. M., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural Representations of Events Arise from Temporal Community Structure. *Nature Neuroscience*, 16(4), 486–492. <https://doi.org/10.1038/nn.3331>

- Schmerling, S. F. (1975). Asymmetric conjunction and rules of conversation. In P. Cole & J. L. Morgan (Eds.), *Speech Acts* (pp. 211–231). https://doi.org/10.1163/9789004368811_009
- Schütze, C. (2016). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Language Science Press.
- Schütze, C., & Sprouse, J. (2013a). Judgement data. In R. Podesva & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 27–50). Cambridge, MA: Cambridge University Press.
- Schütze, C., & Sprouse, J. (2013b). Judgment data. In R. Podesva & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 27–50). <https://doi.org/10.1017/cbo9781139013734.004>
- Sharvit, Y. (2014). On the universal principles of tense embedding: The lesson from before. *Journal of Semantics*, 31(2), 263–313. <https://doi.org/10.1093/jos/ffs024>
- Shimamura, A. P., Janowsky, J. S., & Squire, L. R. (1990). Memory for the temporal order of events in patients with frontal lobe lesions and amnesic patients. *Neuropsychologia*, 28(8), 803–813. [https://doi.org/10.1016/0028-3932\(90\)90004-8](https://doi.org/10.1016/0028-3932(90)90004-8)
- Shin, Y. S., & DuBrow, S. (2021). Structuring Memory Through Inference-Based Event Segmentation. *Topics in Cognitive Science*, 13(1), 106–127. <https://doi.org/10.1111/tops.12505>
- Siefke, B. M., Smith, T. A., & Sederberg, P. B. (2019). A context-change account of temporal distinctiveness. *Memory and Cognition*, 47(6), 1158–1172. <https://doi.org/10.3758/s13421-019-00925-5>
- Skordos, D., & Barner, D. (2019). Language comprehension, inference, and alternatives. In C. Cummins & N. Katsos (Eds.), *The Oxford Handbook of Experimental Semantics and Pragmatics* (pp. 7–20). <https://doi.org/doi.org/10.1093/oxfordhb/9780198791768.013.1>
- Slobin, D. I. (1996). From “thought and language” to “thinking for speaking.” In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity*. (pp. 70–96). New York, NY, US: Cambridge University Press.
- Smith, C. S. (1986). A Speaker-Based Approach to Aspect. *Linguistics and Philosophy*, 9(1), 97–115.
- Smith, C. S. (1999). Activities: States or Events? *Linguistics and Philosophy*, 22(5), 479–508.
- Smith, C. S. (2008). Time With and Without Tense. In J. Guéron, J., Lecarme (Ed.), *Time and Modality. Studies in Natural Language and Linguistic Theory* (pp. 227–249). https://doi.org/10.1007/978-1-4020-8354-9_10
- Solstad, T., & Bott, O. (2022). On the nature of implicit causality and consequentiality: The case of psychological verbs. *Language, Cognition and Neuroscience*, 37(10), 1311–1340. <https://doi.org/10.1080/23273798.2022.2069277>
- Sonne, T., Kingo, O. S., & Krøjgaard, P. (2017). Bound to remember: Infants show superior memory for objects presented at event boundaries. *Scandinavian Journal of Psychology*, 58(2), 107–113. <https://doi.org/10.1111/sjop.12351>
- Speer, N. K., Swallow, K. M., & Zacks, J. M. (2003). Activation of human motion processing areas during event perception. *Cognitive, Affective and Behavioral Neuroscience*, 3(4),

- 335–345. <https://doi.org/10.3758/CABN.3.4.335>
- Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language*, 53(1), 125–140. <https://doi.org/10.1016/j.jml.2005.02.009>
- Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, 116(1), 71–86. <https://doi.org/10.1016/j.cognition.2010.04.002>
- Stockwell, R. P., Schacter, P., & Partee, B. H. (1973). *The major syntactic structures of English*. New York: Holt, Rinehart and Winston.
- Stojić, S., & Nadasdy, Z. (2024). Event as the central construal of psychological time in humans. *Frontiers in Psychology*, 15. <https://doi.org/10.3389/fpsyg.2024.1402903>
- Stowell, T. (2007). The syntactic expression of tense. *Lingua*, 117(2), 437–463. <https://doi.org/10.1016/j.lingua.2005.08.003>
- Strawson, P. F. (1952). *Introduction to logical theory*. London: Methuen.
- Strickland, B., Geraci, C., Chemla, E., Schlenker, P., Kelepir, M., & Pfau, R. (2015). Event representations constrain the structure of language: Sign language as a window into universally accessible linguistic biases. *Proceedings of the National Academy of Sciences of the United States of America*, 112(19), 5968–5973. <https://doi.org/10.1073/pnas.1423080112>
- Suddendorf, T., Addis, D. R., & Corballis, M. C. (2009). Mental time travel and the shaping of the human mind. *Philosophical Transactions of the Royal Society: Biological Sciences*, 364(1521), 1317–1324. <https://doi.org/10.1098/rstb.2008.0301>
- Sun, Z., Firestone, C., & Hafri, A. (2023). How to build a scene: Relational representations are constructed in a canonical order. *Journal of Vision*, 23(9), 4750. <https://doi.org/10.1167/jov.23.9.4750>
- Swallow, K. M., Barch, D. M., Head, D., Maley, C. J., Holder, D., & Zacks, J. M. (2011). Changes in events alter how people remember recent information. *Journal of Cognitive Neuroscience*, 23(5), 1052–1064. <https://doi.org/10.1162/jocn.2010.21524>
- 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(236), 1–7. <https://doi.org/10.1037/a0015631>.Event
- Tai, J. H. Y. (1983). Temporal sequence and Chinese word order. In J. Haiman (Ed.), *Iconicity in Syntax: Proceedings of a Symposium on Iconicity in Syntax* (pp. 49–72). Stanford: John Benjamins Publishing Company.
- Talmy, L. (1975). Figure and Ground in Complex Sentences. *Annual Meeting of the Berkeley Linguistics Society*, 1, 419–430. <https://doi.org/10.3765/bls.v1i0.2322>
- Talmy, L. (1988). The Relation of Grammar to Cognition. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (Vol. 2, pp. 165–205). <https://doi.org/10.3115/980262.980266>
- Talmy, L. (2000). *Toward a Cognitive Semantics, Volume 1: Concept Structuring Systems*. <https://doi.org/10.7551/mitpress/6847.001.0001>
- Taylor, B. (1977). Tense and Continuity. *Linguistics and Philosophy*, 1(2), 199–220. Retrieved from <http://www.jstor.org/stable/25000953>

- Therriault, D. J., & Raney, G. E. (2007). Processing and Representing Temporal Information in Narrative Text. *Discourse Processes*, 43(2), 173–200. <https://doi.org/10.1080/01638530709336897>
- Thomson, A. J., & Martinet, A. V. (1985). *A Practical English Grammar*. Oxford, UK: Oxford University Press.
- Thothathiri, M., & Snedeker, J. (2008). Syntactic priming during language comprehension in three- and four-year-old children. *Journal of Memory and Language*, 58(2), 188–213. <https://doi.org/10.1016/j.jml.2007.06.012>
- Tillman, K. A., Tulagan, N., & Sullivan, J. (2020). Children’s spontaneous inferences about time and causality in narrative. *Proceedings for the 42nd Annual Meeting of the Cognitive Science Society*, 3451–3457.
- Todorova, M., Straub, K., Badecker, W., & Org, E. (2000). Aspectual Coercion and the Online Computation of Sentential Aspect. *Proceedings of the Annual Meeting of the Cognitive Science Society*, (22), 22.
- Tomlin, R. S. (1985). Foreground-background information and the syntax of subordination. *Text*, 5(1–2), 85–122. <https://doi.org/10.1515/text.1.1985.5.1-2.85>
- Tonhauser, J. (2015). Cross-Linguistic Temporal Reference. *Annual Review of Linguistics*, 1(1), 129–154. <https://doi.org/10.1146/annurev-linguist-030514-124923>
- Townsend, D. J. (1983). Thematic processing in sentences and texts. *Cognition*, 13(2), 223–261. [https://doi.org/10.1016/0010-0277\(83\)90023-9](https://doi.org/10.1016/0010-0277(83)90023-9)
- Trosborg, A. (1982). Children’s comprehension of “before” and “after” reinvestigated. *Journal of Child Language*, 9(2), 381–402.
- Trosborg, Anna. (1982). Children’s comprehension of ‘before’ and ‘after’ reinvestigated. *Journal of Child Language*, 9(2), 381–402. <https://doi.org/10.1017/S0305000900004773>
- Tulving, E. (1984). Précis of Elements of episodic memory. *The Behavioral and Brain Sciences*, 7, 223–268.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26(1), 1–12.
- Ünal, E., Ji, Y., & Papafragou, A. (2021). From Event Representation to Linguistic Meaning. *Topics in Cognitive Science*, 13(1), 224–242. <https://doi.org/10.1111/tops.12475>
- 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), 770–779. <https://doi.org/10.1037/0278-7393.28.4.770>
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.
- van Hout, A. (2000). Event semantics in the lexicon-syntax interface: Verb frame alternations in Dutch and their acquisition. In C. Tenny & J. Pustejovsky (Eds.), *Events as grammatical objects* (pp. 239–282). Stanford, CA: CSLI.
- van Pottelberge, J. (2004). *Der am-Progressiv: Struktur und parallele Entwicklung in den kontinentalwestgermanischen Sprachen*. Tübingen: Gunter Narr Verlag.
- Van Valin, R., & LaPolla, R. (1997). *Syntax: Structure, Meaning, and Function*. Cambridge,

- U.K. ; New York, NY: Cambridge University Press.
- Vater, H. (1995). The Tense System of Polish. In R. Thieroff (Ed.), *Tense Systems in European Languages II* (pp. 153–166). <https://doi.org/doi:10.1515/9783110958911.153>
- Vega, M. de, Rinck, M., Díaz, J. M., & León, I. (2007). Figure and Ground in Temporal Sentences: The Role of the Adverbs When and While. *Discourse Processes*, 43(1), 1–23. <https://doi.org/10.1080/01638530709336891>
- Vendler, Z. (1967). Verbs and times. In Z. Vendler (Ed.), *Linguistics and Philosophy* (pp. 97–121). <https://doi.org/10.7591/9781501743726-005>
- Verfaillie, K., & Daems, A. (1996). The priority of the Agent in Visual Event Perception: On the Cognitive Basis of Grammatical Agent-Patient Asymmetries. *Cognitive Linguistics*, 7(2), 131–148. <https://doi.org/10.1515/cogl.1996.7.2.131>
- Verkuyl, H. (1972). *On the Compositional Nature of the Aspects*. <https://doi.org/10.1007/978-94-017-2478-4>
- von der Malsburg, T., Poppels, T., & Levy, R. P. (2020). Implicit gender bias in linguistic descriptions for expected events: The cases of the 2016 United States and 2017 United Kingdom elections. *Psychological Science*, 31(2), 115–128. <https://doi.org/10.1177/0956797619890619>
- von Stechow, A. (1995). On the Proper Treatment of Tense. *Semantics and Linguistic Theory*, 5, 362. <https://doi.org/10.3765/salt.v5i0.2700>
- von Stechow, A. (2005). Semantisches und morphologisches Tempus. Zur temporalen Orientierung von Einstellungen und Modalen. *Neue Beiträge Zur Germanistik*, (124), 9–54.
- von Stechow, A. (2015). On the Proper Treatment of Tense. *Semantics and Linguistic Theory*, 5, 362. <https://doi.org/10.3765/salt.v5i0.2700>
- von Stechow, A., & Grønn, A. (2013a). Tense in Adjuncts Part 1: Relative Clauses. *Linguistics and Language Compass*, 7(5), 295–310. <https://doi.org/10.1111/lnc3.12020>
- von Stechow, A., & Grønn, A. (2013b). Tense in Adjuncts Part 2: Temporal Adverbial Clauses. *Linguistics and Language Compass*, 7(5), 311–327. <https://doi.org/10.1111/lnc3.12019>
- von Stutterheim, C., Carroll, M., & Klein, W. (2003). Two ways of construing complex temporal structures. In F. Lenz (Ed.), *Deictic conceptualisation of time, space and person* (pp. 97–133). <https://doi.org/10.1075/pbns.112.07stu>
- von Stutterheim, C., Carroll, M., & Klein, W. (2009). New perspectives in analyzing aspectual distinctions across languages. *The Expression of Time*, (May 2014), 195–216. <https://doi.org/10.1515/9783110199031.195>
- Wang, Y. C., Adcock, R. A., & Egner, T. (2024). Toward an integrative account of internal and external determinants of event segmentation. *Psychonomic Bulletin and Review*, 31(2), 484–506. <https://doi.org/10.3758/s13423-023-02375-2>
- Wang, Y. C., & Egner, T. (2022). Switching task sets creates event boundaries in memory. *Cognition*, 221, 104992. <https://doi.org/10.1016/j.cognition.2021.104992>
- Wårvik, B. (2004). What is foregrounded in narratives? Hypotheses for the cognitive basis of

- foregrounding. In T. Virtanen (Ed.), *Approaches to cognition through text and discourse* (pp. 99–148). <https://doi.org/10.1515/9783110892895.99>
- Webber, B. L. (1988). Tense as discourse anaphor. *Computational Linguistics*, 14(2), 61–73.
- Weist, R. M. (2014). Developing temporal systems. *Psychology of Language and Communication*, 18(2), 126–142. <https://doi.org/10.2478/plc-2014-0010>
- Weist, R. M., Wysocka, H., & Lyytinen, P. (1991). A Cross-Linguistic Perspective on the Development of Temporal Systems. *Journal of Child Language*, 18(1), 67–92. <https://doi.org/10.1017/S0305000900013301>
- Wiese, H., Öncü, M. T., Müller, H. G., & Wittenberg, E. (2020). Verb Third in spoken German: A natural order of information. In S. Wolfe & R. Woods (Eds.), *Rethinking Verb Second* (pp. 682–699). <https://doi.org/10.1093/oso/9780198844303.003.0029>
- Winskel, H. (2004). The acquisition of temporal reference cross-linguistically using two acting-out comprehension tasks. *Journal of Psycholinguistic Research*, 33(4), 333–355. <https://doi.org/10.1023/B:JOPR.0000035105.65440.b8>
- Wittenberg, E., & Levy, R. (2017). If you want a quick kiss, make it count: How choice of syntactic construction affects event construal. *Journal of Memory and Language*, 94, 254–271. <https://doi.org/10.1016/j.jml.2016.12.001>
- Wolfe, M. B. W., Magliano, J. P., & Larsen, B. (2005). Causal and semantic relatedness in discourse understanding and representation. *Discourse Processes*, 39(2–3), 165–187. <https://doi.org/10.1080/0163853x.2005.9651678>
- Wurm, S. (1949). The (Kara-)Kirghiz Language. *Bulletin of the School of Oriental and African Studies*, 13(1), 97–120.
- Xu, X., & Kwok, S. C. (2019). Temporal-order iconicity bias in narrative event understanding and memory. *Memory*, 27(8), 1079–1090. <https://doi.org/10.1080/09658211.2019.1622734>
- Yates, T. S., Sherman, B. E., & Yousif, S. R. (2023). More than a moment: What does it mean to call something an ‘event’? *Psychonomic Bulletin and Review*, 30, 2067–2082. <https://doi.org/10.3758/s13423-023-02311-4>
- Zacks, J. M. (2004). Using Movement and Intentions to Understand Simple Events. *Cognitive Science*, 28(6), 979–1008. <https://doi.org/10.1016/j.cogsci.2004.06.003>
- Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in Reading and Film Comprehension. *Journal of Experimental Psychology: General*, 138(2), 307–327. <https://doi.org/10.1037/a0015305>
- 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), 273–293. <https://doi.org/10.1037/0033-2909.133.2.273>
- Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain’s cutting-room floor: Segmentation of narrative cinema. *Frontiers in Human Neuroscience*, 4, 1–15. <https://doi.org/10.3389/fnhum.2010.00168>
- Zacks, J. M., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006). Event understanding and memory in healthy aging and dementia of the alzheimer type. *Psychology and Aging*, 21(3), 466–482. <https://doi.org/10.1037/0882-7974.21.3.466>

- 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), 29–58. <https://doi.org/10.1037/0096-3445.130.1.29>
- Zarcone, A., & Demberg, V. (2021). Interaction of Script Knowledge and Temporal Discourse Cues in a Visual World Study. *Discourse Processes*, 58(9), 804–819. <https://doi.org/10.1080/0163853X.2021.1930807>
- Zhang, M., & Hudson, J. A. (2018). The Development of Temporal Concepts: Linguistic Factors and Cognitive Processes. *Frontiers in Psychology*, 9, 1–14. <https://doi.org/10.3389/fpsyg.2018.02451>
- Zlatev, J. (2010). Spatial Semantics. In D. Geeraerts & H. Cuyckens (Eds.), *The Oxford Handbook of Cognitive Linguistics* (pp. 318–350). <https://doi.org/10.1093/oxfordhb/9780199738632.013.0013>
- Zucchi, S. (1998). Aspect shift. In S. Rothstein (Ed.), *Events and grammar. Studies in Linguistics and Philosophy* (pp. 349–370). Dordrecht: Springer.
- Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning Memory and Cognition*, 22(5), 1196–1207. <https://doi.org/10.1037/0278-7393.22.5.1196>
- Zwaan, R. A., & Madden, C. J. (2005). Embodied Sentence Comprehension. In D. Pecher & R. A. Zwaan (Eds.), *Grounding Cognition. The Role of Perception and Action in Memory, Language, and Thinking* (pp. 224–245). <https://doi.org/10.1017/CBO9780511499968.010>
- Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of Situation Model Construction in Narrative Comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21(2), 286–397.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185. <https://doi.org/10.1037//0033-2909.123.2.162>
- 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. https://doi.org/10.1207/s1532799xssr0203_2