Spike-based acquisition of grammatical structure: on the importance of low-level brain mechanisms for high-level cognitive skills

Dissertation
zur Erlangung des Doktorgrades
des Fachbereichs Humanwissenschaften
der Universität Osnabrück

vorgelegt
von

Sophie Lehfeldt

aus

Hannover

Osnabrück, 2022

Erklärung über die Eigenständigkeit der erbrachten wissenschaftlichen Leistung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise ~~entgeltlich~~/ unentgeltlich geholfen.

1. Jutta L. Mueller*: Konzeptualisierung, Schreiben (kritische Beurteilung, Kommentare, Überarbeitung)
2. Gordon Pipa*: Konzeptualisierung, Schreiben (kritische Beurteilung, Kommentare)
3. *Manuskript "Spike-based statistical learning explains human grammar acquisition performance"

Weitere Personen waren an der inhaltlichen materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

....

(Ort. Datum) (Unterschrift)

Acknowledgements

First of all, I would like to thank my supervisors Prof. Dr. Jutta L. Mueller (University of Vienna) and Prof. Dr. Gordon Pipa (University of Osnabrück) for providing such a kind and helpful guidance throughout the time course of my PhD project. I enjoyed very much the creative atmosphere in our brainstorming and discussion sessions and highly appreciate the scientific and personal support provided at any time. Moreover, having been part of the RTG Computational Cognition at the Institute of Cognitive Science was very special as it allowed me to study cognition at an unique interface between the fields of Psycho- and Neurolinguistics and Neuroinformatics. Thanks for providing this framework that was the ideal place for me to develop an independent and interdisciplinary research agenda. And then, of course, thanks to my family and friends for accompanying me throughout this exciting time. Though we all pursue different goals or careers, sharing a mutual interest in each other's projects and passions was the most valuable support throughout.

Zusammenfassung

Kognition kann entlang verschiedener Schnittstellen betrachtet werden. Hierbei hebt die ‘Gehirn-Kognition’-Schnittstelle das Ziel hervor, erklärende Modelle der Implementierung von Kognition im biologischen Gehirn zu formulieren. Darüber hinaus stellen die ‘niedrig-hoch’- und ‘subsymbolisch-symbolisch’-Schnittstellen die Bandbreite möglicher Arten kognitiver Repräsentation und Verarbeitung dar. Ziel der vorliegenden Thesis war es, mögliche Verbindungen zwischen diesen Schnittstellen zu identifizieren. Hierfür wurden die zugrunde liegenden Lernmechanismen des Erwerbs menschlicher Grammatik, einer hohen und potenziell symbolischen kognitiven Fähigkeit, tiefergehend betrachtet. Insbesondere für frühkindliche Phasen des Spracherwerbs werden assoziative und statistische Lernmechanismen als charakteristisch eingestuft. Diese Hypothese motivierte die Anwendung eines assoziativen, statistischen Lernmodells in einem event-basierten neuronalen Netzwerk, um das Erlernen isolierter und geschachtelter Grammatiken mit diskontinuierlichen Abhängigkeiten abzubilden. Die Lernleistung des Modells wurde anhand der resultierenden synaptischen Verbindungsgewichte ermittelt. Hierbei zeigte sich, dass dessen Verteilungen einer Reihe menschlicher Lernmuster entsprachen. Daher scheint assoziatives, statistisches Lernen einen wesentlichen Beitrag zum erfolgreichen Erlernen von Grammatik und möglicherweise auch von menschlicher Sprache zu leisten. Da es sich bei dem Lernmodell um einen subsymbolischen Mechanismus auf niedriger Ebene handelte, könnten hohes und symbolisches Lernen menschlicher Grammatik eine potenzielle subsymbolische Basis haben. Durch die Berücksichtigung von grundlegenden Komponenten und Verarbeitungsprinzipien des Gehirns, d.h. durch die Verwendung eines generischen rekurrenten neuronalen Netzwerks sowie verteilter Verarbeitung und unüberwachten Lernens, konnte ein mögliches minimales Set neurobiologischer Bedingungen für den Erwerb diskontinuierlicher Grammatiken identifiziert werden. Zusammengenommen wurde in dieser Thesis eine Verbindung zwischen den Forschungsfeldern der Psycho- und Neurolinguistik und der Neuroinformatik gefunden. Diese Verbindung wurde anhand der hervorstehenden Bedeutung von assoziativem, statistischem Lernen in beiden Feldern identifiziert. Das Modell beschreibt den speziellen Fall einer naiven, lernenden Person, die durch passive Exposition möglicherweise erste Erkenntnisse über die Statistiken der Umgebung erlangen könnte. Andere kognitive Mechanismen, die eventuell auf das so erlangte Wissen aufbauen, gehen wahrscheinlich über reines assoziatives, statistisches Lernen hinaus. Diese Thesis stellt deshalb einen möglichen Startpunkt für zukünftige Forschung dar, die einen Fokus auf das Zusammenspiel verschiedener kognitiver Strategien legt: angefangen bei subsymbolischem Lernen auf niedriger Ebene bis hin zu hoher, symbolischer Kognition.

Abstract

Cognition can be studied across a variety of interfaces. While the 'brain-cognition' interface highlights the aim to build explanatory models for the implementation of cognition in the biological brain, 'low-high' and 'subsymbolic-symbolic' interfaces demonstrate the scope of possible representation and computation types underlying cognition. This thesis aimed at contributing potential connecting links across these interfaces by taking a closer look at the underlying learning mechanisms of human grammar acquisition (i.e. a high-level and possibly symbolic cognitive skill). Motivated by the hypothesis that especially early childhood language acquisition phases are characterised by associative, statistical learning types, the acquisition of isolated and nested non-adjacent grammars was modelled by associative, statistical learning in a spiking recurrent neural network (i.e. a low-level and subsymbolic mechanism). By demonstrating that grammar learning outcomes of the model (i.e. distributed synapse assembly strengths) complied with a variety of grammar acquisition performance patterns of human learners, associative, statistical learning was identified as an essential contributor for successful grammar learning and potentially also language acquisition. Moreover, given that a low-level and subsymbolic mechanism accounted for aspects of high-level human grammar learning, this cognitive skill might potentially be grounded on a subsymbolic basis. Finally, by including core brain components and processing principles into the model (i.e. a generic recurrent neural network, distributed encoding and unsupervised learning), a potential minimal set of neurobiological requirements for non-adjacent grammar learning was identified. Taken together, this thesis provided a connecting link between the fields of psycho- and neurolinguistics and neuroinformatics that was identified by a salient commonality in both, namely the central role of associative, statistical learning. The model especially described how a naive learner might acquire first knowledge about the statistics of the surrounding environment by passive exposure. However, other cognitive mechanisms that potentially build up on the formed low-level knowledge representations might go beyond mere associative, statistical learning. This thesis therefore provided a potential starting point for future research with a focus on understanding in more detail the potential interplay of different cognitive strategies: starting from low-level and subsymbolic learning towards reaching high-level and symbolic cognition.

Contents

Chapter 1

From low- to high-level cognition

1.1 Cognition at several interfaces

1.1.1 The ‘brain-cognition’ interface

The identification of explanatory links between low-level neural mechanisms and high-level cognitive skills is a historically long-standing (Churchland and Sejnowski 1988; Milner et al. 1998; Albright et al. 2000) and until today one of the most challenging research endeavours of cognitive neuroscience (Kandel et al. 2013; Piccinini and Shagrir 2014; Amunts et al. 2016; Rolls 2021; Torres and Marro 2022; Urai et al. 2022). The ambitious aim to bridge the ‘brain-cognition’ interface has brought fourth major lines of research that are concerned with problems ranging from:

1. ‘Cognitive areas’, i.e. localising brain regions that are specific to cognitive functions (Strotzer 2009; Zilles and Amunts 2010; Judaš et al. 2012; Amunts and Zilles 2015; Snow 2016; Glasser et al. 2016; Fan et al. 2016),

2. ‘Functional networks’, i.e. characterising the interaction of cognitive brain regions within functional networks (Horwitz et al. 1999; Hagmann et al. 2008; Bressler and Menon 2010; Sporns 2013; Park and Friston 2013; Sporns 2014; Pulvermüller et al. 2014; Petersen and Sporns 2015; Smith 2016; Avena-Koenigsberger et al. 2018),

3. ‘Activity signatures’, i.e. correlating brain activity with behavioural and cognitive operations (Polich 1993; Buzsáki and Draguhn 2004; Poldrack et al. 2011; Arnal and Giraud 2012; Einevoll et al. 2013; Krakauer et al. 2017; Duncan et al. 2020; Urai et al. 2022), to

4. ‘Underlying mechanisms’, i.e. identifying neural and synaptic mechanisms that underlie information processing principles of the brain and ultimately also cognition (Sejnowski et al. 1988; Friston 2005; Markram 2013; Harris et al. 2019).

Creating a holistic understanding about the implementation of cognition in the brain certainly requires an integration across all these levels of investigation. Such integration might be outlined taking the neural underpinnings of language as an example. With regard to ‘Cognitive areas’, Broca’s area is well established as a central language processing and production region of the cerebral cortex (Geschwind 1970; Novick et al. 2010; Flinker and Knight 2018). Concerning ‘Functional networks’, Broca’s area is embedded in a larger scale language network (Hickok and Poeppel 2007; Friederici 2011; Fedorenko and Thompson-Schill 2014; Flinker et al. 2015; Ardila et al. 2016; Friederici et al. 2017) and thus interacts with other brain areas during the performance of linguistic operations like the parsing of syntax (Friederici et al. 2006; Chen et al. 2021). With respect to ‘Activity signatures’, oscillations or event-related potentials are correlated to the online dynamics of linguistic computation (Lau et al. 2008), as for example the mismatch negativity (Näätänen and Winkler 1999; Näätänen et al. 2007) that indicates the detection of grammatical rule violations (Mueller et al. 2012; Winkler et al. 2018a). Finally, research dedicated to ‘Underlying mechanisms’ provides conceptual (Garrido et al. 2009; May and Tiitinen 2010), probabilistic (Lieder et al. 2013) or mechanistic models (Wacongne et al. 2012) that explain how the mismatch negativity might arise. Despite these advances, the great ‘brain-cognition’ interface puzzle undoubtedly still misses a great amount of pieces for a further completion. This thesis aimed at contributing such a missing piece by investigating the relevance of low-level brain mechanisms for high-level cognitive skills. Being attributed to the above sketched research line ‘Underlying mechanisms’, this work was concerned with modelling the acquisition of grammatical structure of human language, i.e. a high-level cognitive skill, with spike-based statistical learning, i.e. a low-level brain mechanism. The methodology of this thesis was motivated by the idea that neuroscience has reached the so called ‘simulation phase’ (Fan and Markram 2019). This signifies that theoretical models of neural and synaptic information processing principles are now applicable for simulating brain function (Markram 2006; Kass et al. 2018; Einevoll et al. 2019) and, importantly, for testing their suitability as underlying computational modules of cognition (Kriegeskorte and Douglas 2018). Especially the continuous development of neural network simulation tools has paved the way towards this new ‘simulation phase’ (Diesmann and Gewaltig 2001; Natschläger et al. 2003; Migliore et al. 2006; Brette et al. 2007; Goodman and Brette 2008; Goodman and Brette 2009; Bekolay et al. 2014; Stimberg et al. 2014; Plotnikov et al. 2016; Tikidji-Hamburyan et al. 2017; Blundell et al. 2018; Stimberg et al. 2019). Their availability further advances the aim to formulate and test neurobiologically motivated cognitive models (Gerstner et al. 2012; Palm 2016) and thus, ultimately contributes towards make bridging the ‘brain-cognition’ interface possible.

1.1.2 The ‘low-high’ and ‘subsymbolic-symbolic’ debates

The ‘brain-cognition’ interface is not the only challenge to overcome. The aim to formulate neurobiologically motivated models of cognition inevitably leads to facing the ‘low-high’ and ‘subsymbolic-symbolic’ debates (Smolensky 1987; Smolensky 1988; Harnad 1990; Kelley 2003; Simen and Polk 2010; Smolensky 2012; Kriete et al. 2013; König et al. 2013; Werning et al. 2013; Dehaene et al. 2022). These deal with the central problem of elucidating the levels (i.e. ‘low-high’) or natures (i.e. 'subsymbolic-symbolic') of knowledge representations and cognitive operations. Here, mappings across a scale from low- to high-level are a common approach for distinguishing primarily sensory processes from cognitive operations (König et al. 2013; Dehaene et al. 2022). Across a scale from subsymbolic to symbolic, usually distributed and parallel or centralised and serial representation and operation schemes are contrasted (McClelland and Rumelhart 1985; Kelley 2003). Taken together, while neural representations of sensory inputs are primarily considered low-level and subsymbolic, cognitive skills like thought, reasoning or language are considered high-level and symbolic (Fig. 1.1). However, when trying to explain the emergence of symbolic cognition from distributed sensory or neural representations (Kriete et al. 2013; König et al. 2013; Butz 2016), the question arises if a mapping at theses extremes is mandatory? Or might aspects of high-level, symbolic cognition as well be explained by low-level and subsymbolic mechanisms? Especially when considering that cognition takes place in the neural substrate of the brain (Houdé and Tzourio-Mazoyer 2003; Barsalou 2008), it is nothing but reasonable to expect that low-level operation principles of neurons and synapses set important representational constraints and potentially also influence the outcome of subsequent, high-level cognitive performances. Therefore, next to the 'brain-cognition' interface, this thesis additionally dealt with understanding better to which degree low-level and subsymbolic brain principles might be relevant for high-level cognitive operations that are usually considered to be symbolic.

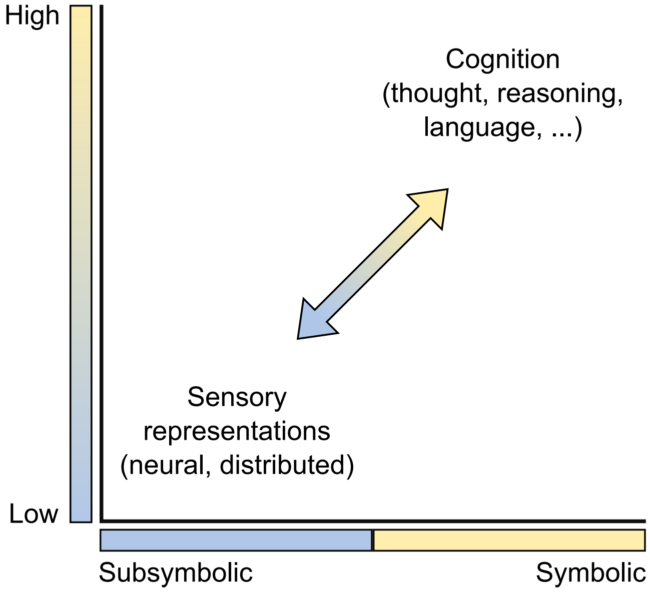


Figure 1.1: Cognition across 'low-high' and 'subsymbolic-symbolic' scales. Continuous 'low-high' scale and binary 'subsymbolic-symbolic' scale. Low-level and subsymbolic (blue) sensory representations (i.e. neural or distributed) contrast high-level and symbolic (yellow) cognition (i.e. thought, reasoning or language). Shaded double arrow indicates the missing link between the extremes of scales.

1.1.3 Challenges for a neuroscience of linguistics

While the 'brain-cognition' interface exists in any cognitive domain, it has specifically been described for the research areas of linguistics and neuroscience entitled as the 'interface problem' (Poeppel and Embick 2005). Here, the challenge to formulate a neuroscientific theory of linguistics was defined as a mapping problem between the 'elements' and 'operations' across domains (Fig. 1.2). According to the 'interface problem', elements and operations within domains can be related in causal ways while relations across domains remain random (Poeppel and Embick 2005; Embick and Poeppel 2015). To illustrate, in linguistics, the constituting elements of a language (i.e. its phonemes, syllables or words) are known to be subject to specific linguistic operations (i.e. parsing, generalisation or production). Equivalently, in neuroscience, core brain elements (i.e. neurons, synapses or networks) are known to perform certain operations (i.e. action potentials, oscillations or plasticity). However, linking 'elements' and 'operations' across domains is not trivial. Specifically, it is unclear to which degree or at which level of detail linguistic and neural elements and operations are related to each other (Poeppel 2012; Embick and Poeppel 2015; Friederici et al. 2017). For example, it is largely unknown at which elementary level a linguistic entity is neurally encoded (i.e. is a word reflected by individual neurons, dendrites or populations of cells?). Moreover, how linguistic operations relate to neural ones is mainly unidentified as well (i.e. is the production of an utterance supported by neural oscillations or individual action potentials?). These considerations reveal a general need for developing computational models of language that take neurobiological constraints of the brain into account (Poeppel and Embick 2005). Thus, by focusing on the implementation of language and speech in the brain (Petersson and Hagoort 2012; Poeppel 2014; Zuidema and Fitz 2019), the possibility to integrate basic neural principles with linguistic theory might arise (Jackendoff 2011). Finally, the 'interface problem' by Poeppel and Embick (2005) is a useful methodological scaffold for approaching the ambitious aim to overcome the 'brain-cognition' interface. It provides a strategy for identifying connecting links between neuroscience and cognition by motivating to search for shared principles across 'elements' and 'operations'.

1.2 From statistical regularity detection to language acquisition

1.2.1 Associative, statistical learning in early childhood

During early language acquisition, the primary source of information for a naive learner is the speech signal (Jusczyk and Luce 2002; Mueller et al. 2012; Mueller et al. 2018b; Frost et al. 2020). To illustrate, a pre-verbal newborn or child is constantly exposed to its first language by the utterances present in the surrounding environment. Moreover, as the sense of hearing develops before birth, both the exposure to spoken language and preliminary forms of language acquisition already start in utero (Gómez and Gerken 2000; Skeide and Friederici 2016; Minai et al. 2017). The speech signal is rich of temporal cues that are informative about specific

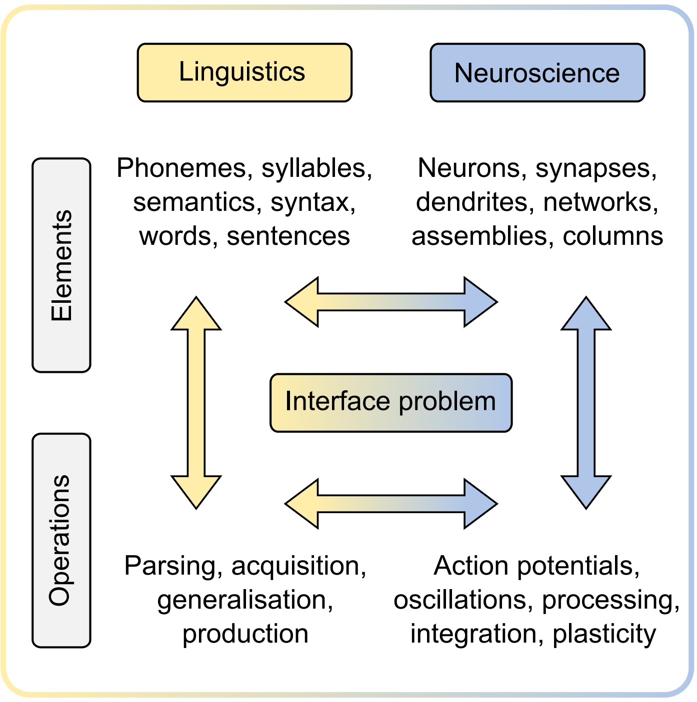


Figure 1.2: The 'interface problem' between linguistics and neuroscience. Elements and operations within each field can be related in explanatory ways (vertical arrows) while relations across the fields are random (horizontal arrows). Compare Fig. 7.1. in Poeppel and Embick (2005).

sound features and structures of a language (Rosen 1992). Further, the statistics of speech comprise a variety of distributional and conditional cues that are indicative for meaningful linguistic entities, i.e. words, or structural regularities, i.e. grammars (Thiessen and Erickson 2013b; Thiessen and Erickson 2013a; Saffran and Kirkham 2018; Frost and Monaghan 2020). Thus, the speech environment provides a bandwidth of information from which a language can be learned from scratch. Children show a remarkable sensitivity to the sound structures and statistics of speech, an ability which is therefore thought to help them bootstrapping a language's organisational principles (Johnson and Jusczyk 2001; Jusczyk 2002; Jusczyk and Luce 2002; de Diego-Balaguer et al. 2016). After their first year of life, children start to understand the communicative purpose of speech (Martin et al. 2012) and develop the ability to articulate words (O'Kane and Goldbart 1998). However, before speech production, there is a prolonged pre-verbal period in which children are exposed to their first language in a seemingly 'passive listening' way (Mueller et al. 2018a). Nonetheless, the importance of this phase becomes apparent when considering the variety of language-specific abilities that develop in this pre-verbal period. For example, newborns can discriminate different languages by their sound structures (Nazzi et al. 1998) and detect structural violations like backward pitch contours (Martinez-Alvarez et al. 2022). Further, pre-verbal children show preferences for repeating sound patters in their environment like their own names (Mandel et al. 1995). Moreover, also in foreign natural languages, pre-verbal children are able to discriminate familiar from novel words (Pelucchi et al. 2009) and detect grammatical rule violations (Friederici et al. 2011b). As the setting in which the language acquisition process begins is dominated by 'mere exposure' and 'passive listening' situations (Aslin 2017), underlying learning mechanisms were hypothesised to be 'implicit', 'automatic', 'statistical' or 'associative' (Perruchet and Pacton 2006; Erickson and Thiessen 2015; Saffran and Kirkham 2018; Mueller et al. 2018a). Summarised by the broad term 'statistical learning', a set of unsupervised learning principles is thought to enable the detection, extraction and ultimately the acquisition of statistical linguistic structures from the raw language signal (Romberg and Saffran 2010; Thiessen 2017). Of course, beyond 'mere exposure' and 'passive listening', pre-verbal children constantly interact with their environment, for example via the production of communication sounds like crying or babbling (Oller et al. 1976; Locke 1989; Cecchini et al. 2007; ter Haar et al. 2021). Yet, these types of verbal communication represent preliminary forms preceding the production of articulated speech (ter Haar et al. 2021). Presumably, when children start their first attempts to produce articulated speech, a transition from seemingly passive to interactive and feedback-guided acquisition phases is initiated. Here, for example, reinforcing feedback might be experienced intrinsically during the rewarding success of communicating a thought in a meaningful way (keyword 'reinforcement learning'). Or, in more supervised scenarios, external feedback by a teacher might guide language learning by providing a verbal correction of an otherwise ungrammatical utterance (keyword 'supervised learning'). However, such interactive learning settings and the possibility to refine language proficiency by specific feedback only occur after having passed through the pre-verbal and 'passive listening' developmental phase. Moreover, this transition of learning settings aligns to the developmental trajectory of cognitive control and the accompanying development of learning strategies. In early childhood, prefrontal cortex, a brain region responsible for cognitive control (Miller and Cohen 2001; Koechlin et al. 2003; Thompson-Schill et al. 2005; Novick et al. 2005), is immature and hence, uncontrolled and implicit learning strategies are hypothesised to dominate (Thompson-Schill et al. 2009). However, during prefrontal cortex maturation, a shift towards more goal-directed and controlled learning strategies occurs (Thompson-Schill et al. 2009). Applied to language acquisition, this developmental trajectory of prefrontal cortex is further hypothesised to explain observed shifts from automatic, associative language learning in infants and controlled learning in adults (Friederici et al. 2013; Mueller et al. 2018a). Recent findings further indicate that this shift presumably starts after the second year of life (Mueller et al. 2019; van der Kant et al. 2020; Paul et al. 2021). Even though infants outperform adults in passive listening language acquisition tasks (Mueller et al. 2018a), implicit learning is also present in adults (Friederici et al. 2002; Mueller et al. 2009; De Diego Balaguer et al. 2007; Morgan-Short et al. 2012). Thus, even though cognitively controlled learning strategies might become increasingly important in adulthood, statistical learning seems to be a mechanisms available throughout a lifetime. Taken together, the central role of associative, statistical learning during initial stages of language acquisition is underscored by pre-verbal children's sensitivity to statistical and linguistically meaningful information contained in the speech signal. Moreover, the predominance of 'passive listening' and 'mere exposure' learning scenarios in early developmental stages gives rise to the need of an unsupervised learning strategy.

Therefore, this thesis aimed at investigating more closely the scope and limits of cognitively low-level, associative, statistical learning in the specific context of early grammar acquisition.

1.2.2 Insights from artificial grammar learning

The role of associative, statistical language learning is commonly studied using the so called 'artificial grammar learning' (AGL) paradigm (Pothos 2007). A major motivation for the development of AGL was to study the processing and acquisition of specific linguistic entities or structures in isolation, i.e. without the biasing effect of a learner's individual level of language experience and proficiency (Petersson et al. 2012; Wilson et al. 2018; Uddén and Männel 2018). Though being criticised for not scaling up to the full complexity of natural languages (Beckers et al. 2012; Chen et al. 2021; Rawski et al. 2021), artificial grammars make use of the basic building blocks of language and therefore allow for a specific investigation of their processing and acquisition (Zuidema et al. 2020). Moreover, also in cognitive domains other than language, the application of artificial stimuli or tasks that contain isolated real-world features of interest is a well-established method for investigating central computational principles of cognition. For example, core mechanisms of vision and movement perception are studied using moving bars or animated images (Castet and Masson 2000; Cracco et al. 2021). Here, basic principles of real-world vision were elucidated by low-dimensional, artificial screen stimuli. But also higher-order cognitive functions like problem solving are investigated by well-defined, artificial toy problems (Dörner and Funke 2017). In the case of language research, artificial grammars are constructed 'miniature languages' composed of linguistic acoustic items, for examples syllables that do not exist in natural language (Wilson et al. 2018). Training procedures according to the AGL paradigm usually comprise the exposition of a passive listener to an acoustic stimulus stream that contains the grammar. Here, a common technique is the so called 'oddball training' where invalid or ungrammatical exemplars, i.e. low-frequent deviants, are interspersed randomly into a stream of otherwise valid or grammatical exemplars, i.e. high-frequent standards (Sussman 2007). Implicit learning performances are measured by behavioural responses like head-turn preferences in children (Kemler Nelson et al. 1995), even though a large variety of preference directions along the developmental trajectory has to be taken into account (Culbertson et al. 2016; Santolin et al. 2020). However, also instructed forms of grammaticality judgements (Chen et al. 2021) and free recall techniques exist (Kidd et al. 2020; Isbilen et al. 2020). Further, brain signatures like the mismatch negativity in response to deviant stimuli (Näätänen and Winkler 1999; Näätänen et al. 2005; Näätänen et al. 2007; Kuhl and Rivera-Gaxiola 2008) are interpreted as a sign of error detection (Sussman 2007; May and Tiitinen 2010). While in adults, mismatch potentials are characterised by negativity, in children both negativity or positivity occur (Trainor et al. 2003; He et al. 2007). The AGL paradigm provided major insights into early statistical language acquisition. In preverbal children, observed learning abilities ranged from the segmentation of words (Saffran et al. 1996a; Aslin et al. 1998), tone sequences (Saffran et al. 1999) to the acquisition of grammatical rules or syntax (Marcus et al. 1999; Saffran and Wilson

2003). However, also verbal children and adults showed implicit learning successes in word segmentation (Saffran et al. 1996b; Saffran et al. 1997) and grammar acquisition (Reber 1967; Thompson and Newport 2007; Kidd 2012). Taken together, these findings by AGL demonstrate the importance of implicit and statistical learning for non-proficient but also proficient language learners to acquire knowledge about the basic units of language, namely its words and grammars. Moreover, statistical learning was also demonstrated for artificial visual scenes in both children (Fiser and Aslin 2002b; Kirkham et al. 2002) and adults (Fiser and Aslin 2002a; Fiser and Aslin 2005; Lazartigues et al. 2021) but also for tactile sequences (Conway and Christiansen 2005). The demonstration of statistical learning across modalities therefore suggests the existence of domain-general, statistical learning mechanisms for general sensory scene understanding (Conway and Christiansen 2005; Frost et al. 2015) beyond the scope of language acquisition.

1.2.3 Non-adjacent dependency acquisition

Language is rich of discontinuous relations among its constituting units. In this regard, non-adjacent dependencies (NADs) are prominent discontinuous grammars where related elements are separated by intervening material (Wilson et al. 2018). For example, in tense agreement, the '*is, -ing*' frame represents an isolated NAD that is combinable with arbitrary verbs, resulting in the grammatical structure '*is* [verb]*ing*' (Wilson et al. 2018). Moreover, mastering NADs is important for understanding nested sentences (de Vries et al. 2012), as for example '*the boy* [the girl chased] *kicked the ball*', where the outer sentence, i.e. 'the boy kicked the ball', is wrapped around the inner embedded clause, i.e. 'the girl chased' (Winkler et al. 2018a). Both isolated and nested NADs pose the challenging task for a learner to detect the grammatical relation between the non-adjacent elements while omitting the intervening material (Wilson et al. 2018). Despite this complexity of the problem, NAD learning in natural language has been observed already in 4-month-olds (Friederici et al. 2011b) and 18-month-olds (Santelmann and Jusczyk 1998; Höhle et al. 2006). Moreover, also nonlocal rules of syntax were already acquired by 18-month-olds (Perkins and Lidz 2021). Using artificial grammars, the ability to acquire NADs was shown for 8-month-olds (Kabdebon et al. 2015), 15-month-olds (Gómez and Maye 2005), 17-month-olds (Frost et al. 2020) and 18-month-olds (Gómez 2002). Further, NAD learning was not restricted to the language domain but could be demonstrated for other modalities like tone sequences in 5-month-olds (Winkler et al. 2018a; Winkler et al. 2018b), human action sequences in 9-month-olds (Lu and Mintz 2021) and visual sequences in 13- to 15-month-olds (Bettoni et al. 2021). But also in adults, the implicit acquisition of linguistic NADs could be evoked by passive listening training procedures (Perruchet et al. 2004; De Diego Balaguer et al. 2007; van den Bos et al. 2012; Romberg and Saffran 2013; Grama et al. 2016; Frost and Monaghan 2016; Isbilen et al. 2022). Again, also adults could learn NADs in other modalities like tone sequences (Gebhart et al. 2009; Endress 2010) and visual symbols (Bahlmann et al. 2009; Iao et al. 2021). Throughout, a variety of stimulus conditions for the facilitation of artificial NAD acquisition has been identified ranging from statistical, i.e. distributional and conditional, to perceptual stimulus features (Wilson et al. 2018). For example, a high variability of the intervening material was beneficial for NAD learning by facilitating a detection of the statistically more persistent, invariant non-adjacent relations (Gómez 2002; Gómez and Maye 2005). But also in general, high frequent linguistic structures or words seem to be valuable cues for learning and efficient processing (Mintz 2003; Brysbaert et al. 2018). Further, training procedures with an incremental transition from adjacent to non-adjacent dependencies were beneficial (Lany and Gómez 2008). In this context, learning grammatical dependencies of short length is generally thought to be easier because less intervening material has to be processed (van den Bos and Poletiek 2008). Next, pauses between grammar exemplars were beneficial as they introduced segmental cues that facilitated their detection (Peña et al. 2002; Mueller et al. 2010; de Diego-Balaguer et al. 2015). But also in natural language, pauses were valuable cues for the detection of phrase boundaries (Männel and Friederici 2009). Further, a bandwidth of structural features were identified to facilitate NAD learning by emphasising the discontinuous relations. For example, NAD acquisition was facilitated when grammatical elements were perceptually similar (Newport and Aslin 2004), shared phonological (Onnis et al. 2005) or prosodic features (Grama et al. 2016) or were marked by prosodic cues (Mueller et al. 2010). Finally, evoked directed attention on NADs facilitated their acquisition (Pacton and Perruchet 2008; de Diego-Balaguer et al. 2016). Despite these insights, it remained unclear whether improved NAD acquisition performances in response to this variety of stimulus conditions could be explained by a shared underlying learning mechanism. Here, the claim that statistical, associative learning is a predominant learning type for the acquisition of language would gain further support if it could explain this variety of observations. Motivated by the need to further evaluate the performance and explanatory power of associative, statistical learning (Frost et al. 2019; Bogaerts et al. 2021), this thesis took the identified stimulus variations ranging from variability, dependency length, stimulus pauses to emphasis on grammars (Wilson et al. 2018) as benchmark conditions for this endeavour. Further, beyond isolated NADs (Wilson et al. 2018) also the acquisition of nested NADs (Winkler et al. 2018a) was studied under these stimulus constraints.

1.3 Brain principles for neuro-cognitive models

1.3.1 Which brain features to consider?

A major question for the development of neurobiologically grounded models of cognition is the identification of indispensable brain components and operation principles. In other words, what are the neural constraints that are essential for building mechanistic models of cognition (O'Reilly 1998; Eliasmith 2007; Maass 2016; Pulvermüller et al. 2021; Saxe et al. 2021)? The bandwidth of potential architectural structures, encoding and learning schemes is large. To illustrate, should neuro-cognitive models consider the columnar layout of cortex (Markram 2006; DeFelipe et al. 2012), include several neural layers (LeCun et al. 2015; Richards et al. 2019) or implement recurrent connectivity schemes (Kietzmann et al. 2019; van Bergen and Kriegeskorte 2020; Prince et al. 2022)? Are specific large-scale networks (Eliasmith and Trujillo

2014; Markram et al. 2015; Abdellah et al. 2017) required or can generic small-scale networks (Miner and Triesch 2016) be cognitively functional? Can the morphology of biological neurons be compressed to point models (Izhikevich 2004) or do we miss out an essential part of neural computation by neglecting the complex structures of dendritic trees (Frick and Johnston 2005; Spruston 2008; Stuart and Spruston 2015; Hawkins and Ahmad 2016; Doron et al. 2017; Bono and Clopath 2017; Kanari et al. 2022)? Is the precise timing of individual action potentials of relevance (Gerstner et al. 1993; Gerstner et al. 1997; Borst and Theunissen 1999; VanRullen et al. 2005; Panzeri et al. 2014; Fiorillo et al. 2014; Nolte et al. 2019) or are firing rates (Stemmler and Koch 1999; Heiberg et al. 2018; Keeley et al. 2019; Qian and Zhang 2020) and oscillations (Buzsáki and Draguhn 2004; Siegel et al. 2012; Kayser et al. 2012; Heusser et al. 2016; Kühn and Helias 2017) the main carriers of information? And finally, are bottom-up local learning rules (Markram et al. 2011) or top-down error- and reward-driven learning schemes (Khaligh-Razavi and Kriegeskorte 2014; Yamins and DiCarlo 2016; Kriegeskorte and Golan 2019) at the heart of cognition? An expedient approach to overcome this multitude of possibilities and options is to concentrate on the nature of the task of interest, followed by a subsequent search of a minimal set of brain features that are potentially supportive (Abbott 2008). For example, if the question is how brain signatures like the mismatch negativity arise, models with continuous encoding schemes potentially represent reasonable substrates as they produce oscillatory outputs that correspond to the experimentally observed signal (Lieder et al. 2013). Further, if the question is how the interaction of a learner with the environment shapes knowledge representations, reinforcement schemes are probably the learning type of choice (Clay et al. 2021). Thus, decisions about included biological detail in model architectures and about the applied encoding or learning schemes ideally relate to the constraints given by the research question. However, this approach inevitably introduces strong biases by considering specific brain and hence model features to be appropriate or supportive a priori. Still, a successful reproduction of the phenomenon of interest by the chosen feature set can be viewed as a major milestone for a further refinement of indispensable neuro-cognitive model properties in the specifically defined context (Abbott 2008; Stöckel et al. 2021). Finally, models at all levels of biological detail ranging from single neuron outputs to mass signals have their reasons as they merely reflect the large diversity of brain features and operations that exist and hence, the bandwidth of analysis levels at which brain dynamics and cognition can be described and studied (Deco et al. 2008; Panzeri et al. 2015). In this thesis, the search of an appropriate neural modelling substrate was guided and facilitated by considering a top-down constraint from experimental sciences (Eliasmith and Trujillo 2014), namely the associative language learning hypothesis (Mueller et al. 2018a; Thompson-Schill et al. 2009). Interestingly, associative learning in the form of spike-timing dependent plasticity (STDP) is considered a central unsupervised learning scheme of the brain (Markram et al. 1997; Bi and Poo 1998; Bi and Poo 1999; Markram et al. 2011). This salient commonality between significant learning types in linguistics and neuroscience motivated to explore the power of STDP for potentially explaining the associative acquisition of grammatical structures like NADs. Choosing associative, statistical learning by STDP as the algorithm of interest further entailed the use of a spike-based recurrent neural architecture for its implementation (Gilson et al. 2010; Kunkel et al. 2011). Taken together, in this thesis, the minimal set of potentially supportive brain features for associative grammar acquisition therefore consisted of spike-based statistical learning in a recurrent neural network.

1.3.2 Spike-based encoding and processing

Biological brains encode and transmit information in a striking way that relies on the interplay between continuously evolving neuronal or synaptic parameters and discontinuous all-or-nothing events, i.e. action potentials or spikes (Stevens and Zador 1995; Strong et al. 1998; Borst and Theunissen 1999; Aur et al. 2006). Here, continuous variables like the membrane potential of a neuron or the plastic weight of a synapse are thought to reflect the integrated input received over time (Maass and Markram 2002). Complementary, spikes encode currently processed stimuli or reflect the current activity output of neurons (Nádasdy 2000). Next to the internal dynamics of continuous variables, for example decaying (i.e. leaking) or consolidation behaviours (Zenke et al. 2013), they undergo local and instantaneous spike-triggered changes that drive activity, processing and ultimately also learning (Stimberg et al. 2014). Considering this dynamic interaction of variables at different temporal scales, the importance of time for neural encoding, processing and computation becomes highlighted (Maass 1997; Belatreche et al. 2007). Many temporal codes are observed in neurons, ranging from bursting behaviours (Izhikevich 2003; Izhikevich 2006; Berger et al. 2010), synchrony (Engel et al. 1991; Engel et al. 1992; König et al. 1996; Stevens and Zador 1998; Uhlhaas et al. 2009; Pipa and Munk 2011; Singer and Lazar 2016; Palacios et al. 2019; Lewis et al. 2021), asynchrony (Ostojic 2014), sparse coding (Olshausen and Field 2004; Hromádka et al. 2008; Koulakov et al. 2009; Kloppenburg and Nawrot 2014) to distributed spatio-temporal patterns (Gerstner et al. 1993; Nádasdy 2000; Shamma 2001; Fellous et al. 2004; Plenz and Thiagarajan 2007; Pillow et al. 2008; Kayser et al. 2009; Buonomano and Maass 2009; De Zeeuw et al. 2011; Ju et al. 2015; Takahashi et al. 2015). Potentially, this flexibility of a spike-based neural code is the brain's way to represent the rich temporal structures encountered in its sensory environment (Gerstner et al. 1997; deCharms and Zador 2000; Panzeri et al. 2010). Speech is a prime example of such a temporal environmental signal, unfolding sequentially and embedding a diversity of structured cues (Rosen 1992). Notably, the encoding of sound and speech comprises a locking of neurons' excitability or oscillatory activity to the temporal contours of the acoustic envelope in humans (Peelle and Davis 2012; Giraud and Poeppel 2012; Ding and Simon 2014; Doelling et al. 2014; Ding et al. 2016; Teng et al. 2017; Keitel et al. 2018; Kaufeld et al. 2020). Further, sampling windows at different time scales underlie online speech understanding (Poeppel 2003; Luo and Poeppel 2007; van Wassenhove et al. 2007; Poeppel et al. 2008; Norman-Haignere et al. 2022; Schmitt and Obleser 2022). But also in monkeys, this temporal locking to the acoustic stimulus is observed (Lakatos et al. 2005). Moreover, at the spike-based level, the temporal tracking of acoustic stimulus properties is a universal neural processing feature as demonstrated in rats (Machens et al. 2004), songbirds (Narayan et al. 2006) monkeys (Belitski et al. 2010; Brasselet et al. 2012; Ng et al. 2013) and bats

(García-Rosales et al. 2018). The spike-based temporal coding of biological neurons has a long history of being simulated by the model class of leaky integrate-and-fire (LIF) neurons (Abbott 1999; Izhikevich 2004; Brette and Gerstner 2005; La Camera et al. 2006; Brunel and van Rossum 2007; Pospischil et al. 2008). This class implements the above sketched hybrid system of a continuously evolving variable, i.e. the leaky integrator, in combination with the discontinuous generation of spike events (Stimberg et al. 2014). Importantly, LIF neurons do not model the voltage trace of action potentials explicitly but rather define spikes by their time points of occurrence (Maass 1997). Beyond modelling the processing dynamics of individual cells, LIF neurons allow to investigate spike-based activity dynamics and processing principles also at the network level (Izhikevich 2003; Scarpetta and de Candia 2013; Abbott et al. 2016; Sahasranamam et al. 2016; Duarte and Morrison 2019). Therefore, they are an important research tool for understanding the details of neural computation (Kass et al. 2018) but also for testing the implementation of high-level cognitive skills in a low-level and distributed neural substrate (Vogels and Abbott 2005; Vertes and Duke 2010; Duarte et al. 2018).

1.3.3 Local synaptic and neural plasticity

Time is also essential for local learning at individual synapses (Fuhrmann et al. 2002). Here, the relative timing between pre- or postsynaptic input events is decisive for the direction and degree of plasticity changes as evidenced by electrophysiological measurements in rat cortical neurons (Markram et al. 1997; Bi and Poo 1998; Debanne et al. 1998; Bi and Poo 1999; Feldman 2000; Sjöström et al. 2001; Froemke and Dan 2002). The collected findings of these experiments further gave rise to the development of computational models of spike-timing dependent plasticity (Song et al. 2000; Rubin et al. 2001; Izhikevich and Desai 2003) that include the following learning principles. First, STDP induces a strengthening for temporally positive correlations (i.e. potentiation by *pre-post* events) and a weakening for temporally negative correlations (i.e. depression by *post-pre* events) at synapses (Bi and Poo 1998). This plasticity profile of STDP is widely thought to support the theory of 'Hebbian learning' (Abbott and Nelson 2000; Markram et al. 2011; Markram et al. 2012) according to which positively correlated event sequences between two neurons elicit a strengthening of their synaptic connection (Hebb 1949). However, also a bandwidth of other temporal windows exists, demonstrating the variety of temporal learning algorithms that can be expressed locally at biological synapses (Kepecs et al. 2002; Shouval et al. 2010; Graupner and Brunel 2012a). Second, temporal windows for the induction of potentiation and depression are characterised by exponential contours (Dan and Poo 2006; Markram et al. 2011). Thus, the extent of plasticity changes increases with temporal proximity of events and respectively decreases with temporal distance. And finally, STDP is frequency dependent (Markram and Tsodyks 1996; Markram et al. 1998b; Markram et al. 1998a). In STDP models, this is expressed by a positive correlation of weight changes to stimulus frequency (Graupner and Brunel 2012a; Litwin-Kumar and Doiron 2014; Zenke et al. 2015). Model implementations of STDP usually comprise spike-triggered, exponentially decaying memory traces whose time constants define the temporal windows of plasticity induction. These memory traces are a computationally affordable way for implementing the STDP windows. However, they are rather theoretical constructs, i.e. do not reflect a specific parameter of biological synapses. This highlights the phenomenological character of STDP models (Morrison et al. 2008) that reproduce the plasticity outcome at synapses but not necessarily the underlying, biophysical signalling dynamics. Still, memory traces based on biophysical pathways exist, as for example in calcium-based STDP models (Shouval et al. 2002; Graupner and Brunel 2010; Graupner and Brunel 2012a; Graupner and Brunel 2012b). Taken together, STDP is a phenomenological, associative and statistical learning model that describes how the connection strengths, i.e. synaptic weights, between neurons are shaped during online spike-based processing. It represents a domain-general learning algorithm of the brain and allows to study the emergence of synaptic weight memory (Martin et al. 2000; Abbott and Regehr 2004; Hu et al. 2016) and cortical map formation (Song and Abbott 2001; Kepecs et al. 2002). However, synaptic weights are not the only brain parameters that can undergo plasticity changes. Also structural plasticity, i.e. the growth and vanishing of synaptic connections (Le Bé and Markram 2006) is considered a possible mechanism of learning and memory in the brain (Miner and Triesch 2016; Gallinaro et al. 2022). Moreover, also neurons demonstrate an activity-dependent intrinsic plasticity (IP) of their excitability thresholds (Turrigiano et al. 1994; Desai et al. 1999; Daoudal and Debanne 2003). In IP models, the voltage threshold for spike generation usually increases in response to each emitted spike and decreases during periods of neuronal silence (Lazar et al. 2009; Stimberg et al. 2014). Thereby, as highly active neurons subsequently have to reach higher sub-threshold activity states in order to emit a spike, IP contributes to firing rate homeostatis and the stabilisation of network activity (Lazar et al. 2007). Further, the decay of excitability thresholds during silence allows for a redundant utilisation of a network's resources as it facilitates the participation of previously inactive neurons (Toutounji and Pipa 2014). Taken together, the activity-dependent history of processed stimuli can also be imprinted neurally by the plasticity of excitability thresholds (Daoudal and Debanne 2003; Zhang and Linden 2003). However, IP is commonly applied as a complementary mechanism next to learning via synaptic weight plasticity (Lazar et al. 2009; Triesch 2004; Lazar et al. 2007; Triesch 2007; Toutounji and Pipa 2014; Leugering and Pipa 2018). In this thesis, the combination of synaptic STDP and neural IP therefore represented the scaffold of choice for exploring associative, statistical learning in a spike-based neural network model.

1.3.4 Recurrent assemblies for cognition

Recurrent connectivity of excitatory and inhibitory neurons is a predominant architectural feature of the cortex (Douglas et al. 1989; Thomson and Deuchars 1994; Douglas et al. 1995a; Thomson and Bannister 2003; Binzegger et al. 2004; Markram et al. 2004; Douglas and Martin 2004; Douglas and Martin 2007a; Reimann et al. 2017a; Siu et al. 2022). Given its ubiquitous presence, generic recurrent networks are considered basic computational units of the brain (Shu et al. 2003; Kalisman et al. 2005; Douglas and Martin 2007b; Perin et al. 2011). Further, in combination with associative STDP, recurrent excitatory networks are viewed as suitable neural substrates for the formation and representation of distributed memory (Mishra et al. 2016; Kossio et al. 2021). Here, the plasticity-driven emergence of strongly connected subgroups in the excitatory network is thought to underlie the formation of functional circuits, also referred to as ‘Hebbian cell assemblies’, that serve the purpose of stimulus-specific feature encoding (Pulvermüller 1996; Chady 2001; Reilly 2001; Sommer and Wennekers 2003; Lansner 2009; Huyck and Passmore 2013; Palm et al. 2014; Poo et al. 2016). At the activity level, cell assemblies demonstrate correlated dynamics (Peyrache et al. 2010; Reimann et al. 2017b) or attractor states (Amit et al. 1997). However, when structural plasticity is subject to learning, also silent memories can emerge (Gallinaro et al. 2022). Beyond being a potential substrate for neural information encoding (Wehr and Laurent 1996; Sakurai 1999), assemblies are also thought to support higher cognitive function (Buzsáki 2010; Pulvermüller et al. 2014; Müller et al. 2020). For example, in language they are hypothesised to underlie the encoding and processing of words (Wennekers et al. 2006) or syntax (Wennekers and Palm 2009; Papadimitriou et al. 2020). Building up on these ideas of assembly-based cognition, this thesis investigated the plasticity-driven emergence of synapse assemblies in the excitatory-to-excitatory connections of a recurrent neural network. By reading from the synaptic weight distribution (Barbour et al. 2007) after learning, the ability of the model to encode NADs by synapse assemblies was evaluated.

Chapter 2

The potential of spiking neural networks

2.1 From brain principles to cognition

2.1.1 Spike-based models of brain properties and dynamics

The motivation to consider spiking neural networks as promising modelling substrates for explaining cognitive computations is based on the bandwidth of brain properties and dynamics that they reproduce. Spiking models predominantly cover the previously identified abstraction levels ‘Functional networks’, ’Activity signatures’ and ‘Underlying mechanisms’ of the ‘brain-cognition’ interface. For example, regarding ‘Functional networks’, the dynamics of the thalamo-cortical system were modelled at a large scale (Izhikevich and Edelman 2008) and the filter properties of the thalamus further classified (Connelly et al. 2016). At the ‘Activity signature’ level, spiking networks reproduced cortical resting activity (Deco and Jirsa 2012; Deco et al. 2013), generated a large bandwidth of cortical oscillation types (Izhikevich 2006) or provided explanatory links between spiking activity and cortical oscillations (Vincent-Lamarre et al. 2020). Further, they modelled local field potentials (Reimann et al. 2013) and electroencephalography (EEG) data (Doborjeh et al. 2019; Schwab et al. 2021). At the ‘Underlying mechanisms’ level, spiking models explained basic brain properties like the self-organisation into receptive fields (Izhikevich et al. 2004), the emergence of synaptic inhomogeneities (Effenberger et al. 2015) and principles of information transfer (Zajzon et al. 2019). But also binary models, i.e. approximations of spiking models, could further elucidate principles of neural information transmission (Del Papa et al. 2017) and computation performance (Toutounji and Pipa 2014). Beyond, cortical rewiring was modelled by binary and spike-based structural plasticity (Zheng et al. 2013; Miner and Triesch 2016). Additionally, the role of spike-based structural plasticity for stable memory formation (Gallinaro et al. 2022) expanded research concerned with spiking learning and memory models grounded in synaptic weight plasticity (Zenke et al. 2015). Spiking models further modelled working memory (Szatmáry and Izhikevich 2010; Mi et al. 2017) and reproduced cortical sequence learning (Klos et al. 2018) observed in rats (Xu et al. 2012). In the context of sensory perception and cognitive science research, spiking approaches achieved the modelling of auditory adaptation in general (Kudela et al. 2018) and during oddball training (Mill et al. 2011) as well as the modelling of mismatch negativity (Wacongne et al. 2012). Further, spiking models were used to elucidate how the discrepancy between the different time scales of behaviours, i.e. taking place at the order of seconds, and neural dynamics, i.e. taking place at the order of milliseconds, might be bridged (Melamed et al. 2004; Setareh et al. 2018). Taken together, the spiking models collected here demonstrate the great flexibility of this model class for reproducing and explaining brain properties and dynamics in a variety of contexts.

2.1.2 Spike-based models of cognition

Even though a large bandwidth of brain properties and dynamics were reproduced by spiking models, spike-based approaches for the modelling of high-level cognitive skills remain rare. However, a small-scale model for the implementation of spike-based numerical cognition was proposed (Rapp et al. 2020) as well as a very large-scale spiking model for solving several cognitive tasks (Eliasmith 2005; Eliasmith et al. 2012; Stewart et al. 2012). These approaches provide first indications about a general feasibility of implementing cognitive skills in spiking neural networks. However, the extent to which this model class is applicable to explaining high-level cognition in a broader sense has not been demonstrated in a satisfying manner yet. Taken together, a bottom-up approach for accomplishing the transition from 'spiking models of brain dynamics' to 'spiking models of cognition' seems to remain a major open research problem. However, potentially, a top-down transition approach from 'non-spiking models of cognition' to spiking ones might provide valuable contributions. For example, especially the class of deep neural networks successfully implement cognitively relevant functions ranging from pattern detection in speech and vision (LeCun et al. 2015) to cognitive planning and navigation (Cichy and Kaiser 2019). Interestingly, a large bandwidth of studies already explored the possibility to consider spike-based encoding schemes of the brain also in deep neural architectures (O'Connor and Welling 2016; Neil et al. 2016; Lee et al. 2016; Bellec et al. 2018; Park et al. 2019; Wu et al. 2019; Sengupta et al. 2019; Bellec et al. 2020; Han and Roy 2020; Kaiser et al. 2020; Wu et al. 2020; Zhang et al. 2020a; Zhang and Li 2020; Jia et al. 2021; Ma et al. 2021; Tang et al. 2022). Here, spike-based processing and learning were considered potential solutions for solving efficiency problems in deep architectures that are introduced by resource extensive update cycles (Neil et al. 2016; Lee et al. 2016; Park et al. 2019; Wu et al. 2020; Zhang et al. 2020a). Further, spiking neurons were widely viewed as an important model feature to improve 'biological realism' (Pfeiffer and Pfeil 2018; Sengupta et al. 2019; Jia et al. 2021; Tang et al. 2022). Finally, these approaches already reached benchmark (Wu et al. 2019; Ma et al. 2021) or even better learning performances as compared to the state-of-the-art (Sengupta et al. 2019; Jia et al. 2021). Taken together, the challenge to build spiking models of higher-order cognition might profit from integrating key insights gained by both modelling fields that either deal with primarily biologically motivated or deep artificial neural networks (Pfeiffer and Pfeil 2018; Taherkhani et al. 2020; Rapp and Nawrot 2020; Macpherson et al. 2021; Eshraghian et al. 2022).

2.2 From sequence to language learning models

2.2.1 Sequence learning as a prerequisite for language

The sequential nature of language poses the challenging task to extract grammatical structures during online processing (Mueller et al. 2012; Christiansen and Chater 2016; Mueller et al. 2018a). Further, linguistic dependencies are considered to be sequential to a great extent so that, as a consequence, language understanding, learning and production might be well described as sequence solving problems (Frank et al. 2012). Going in line with this idea, also this thesis hypothesised that language acquisition might substantially rely on general temporal sequence processing and learning abilities. However, language can also be described as highly hierarchical with tree-like dependency structures amongst linguistic elements or categories that go beyond mere sequential relations (Makuuchi et al. 2009; Friederici et al. 2011a; Jäger and Rogers 2012; Fitch and Friederici 2012; Beekhuizen et al. 2013; Yang et al. 2017; Uddén et al. 2020; Chen et al. 2021). Acknowledging this, sequence learning can nonetheless be considered an essential underlying skill of hierarchical linguistic parsing. Moreover, it potentially is a crucial prerequisite for mastering the complexity of human language. For example, while the parsing of syntax in a sentence might require hierarchical analysis strategies (Dehaene et al. 2015; Doumas and Martin 2018; Chen et al. 2021), sequential processing abilities are indispensable for the extraction of the linguistic structures of interest that are embedded in the temporally unfolding language signal (Frank et al. 2012; Martin and Doumas 2017; Uddén et al. 2020). Taken together, understanding better how sequential and hierarchical aspects of language and the respective cognitive analysis strategies for parsing combine represents a major open research challenge (Dehaene et al. 2015). As the aim of this thesis was to elucidate possible low-level and neural mechanisms underlying grammar learning during online processing, the focus was primarily set on sequence processing, learning or production skills. Further, these were not necessarily required to be language-specific, but could rather be domain- or modality-general. To illustrate, they might include behavioural or cognitive capacities ranging from movement generation (Doyon 2008), olfactory (Johnson et al. 2013), tactile (Bale et al. 2017) and visual sequence learning (Bettoni et al. 2021) to counting (Zaleznik and Park 2021). Though modality-specific constraints were identified (Conway and Christiansen 2005), especially integration in visuo-auditory (Bröhl et al. 2022) and auditory-tactile domains (Bresciani et al. 2005) highlighted tight links between sequence processing abilities across the senses. Thus, by looking for mechanisms beyond the mere scope of language, the possibility of a common shared set of principles for domain- and modality-general sequence processing and learning was explored (Dominey 2005; Johnson et al. 2013).

2.2.2 State-of-the-art modelling approaches

The following literature review provided an overview about state-of-the-art models for the processing, learning and production of language-specific or domain-general sequences. Especially for the research question of this thesis, models that performed the learning of grammars or the extraction of statistical structure from sequential input were of particular interest. Moreover, a special focus was laid on models that implemented such skills in recurrent neural networks, with distributed encodings and purely unsupervised learning schemes. In order to compare language-specific (Tab. 2.1) and domain-general (Tab. 2.2) models, they were categorised according to their (a) architecture, (b) connectivity type, (c) encoding scheme, (d) learning mechanism and (e) learning task (Fig. 2.1):

(a) The sub-categories of the feature 'architecture' ranged from feed forward neural networks (FNNs), simple recurrent networks (SRNs), recurrent neural networks (RNNs) to reservoir computing networks (RCs). FNNs consisted of several neuronal layers that were connected unidirectionally so that the input processing occurred from input to output layers (Svozil et al. 1997; LeCun et al. 2015). SRNs were as well layered neural networks that additionally contained backward projections to earlier layers so that a complementary recurrent processing loop was achieved (Elman 1990). RNNs were single layers of neurons that reciprocally projected onto each other so that the input processing was predominantly recurrent (Douglas and Martin 2007b). RCs were combinations of a recurrently connected neural layer, i.e. the reservoir, and a readout layer (Maass et al. 2002; Lukoševičius and Jaeger 2009; Tanaka et al. 2019; Gauthier et al. 2021). Thus, the input processing occurred recurrently and feed forwardly. Finally, the sub-category 'Miscellaneous' comprised other types of neural architectures that did not fall into the previously defined sub-categories. Or further, miscellaneous models were no neural networks, for example probabilistic or Bayesian models, vector-based approaches or entropy models (Lampinen and Vehtari 2001; Kelly and West 2012).

(b) The sub-categories of the feature 'connectivity type' ranged from full, structured, random to mixed connection schemes. Full schemes were all-to-all connections of neurons. Structured schemes comprised pre-defined connectivity patterns or, in larger-scale networks, pre-defined connectivities between subareas. Random schemes had randomised neural connections that were established with pre-defined probabilities. Mixed schemes were random connectivities in combination with structured sub-networks. While structured schemes biased network dynamics, for example by supporting the convergence towards specific activity states or computations (Maass 2000), random schemes were generic and thus, implemented no specific features for the fulfilment of a certain computational task (Papadimitriou et al. 2020).

(c) The sub-categories of the feature 'encoding scheme' ranged from symbolic, continuous, binary to spiking. Further, encoding schemes of 'input' and 'processing' were distinguished. Symbolic input encodings were based on letters, digits or geometric icons. Symbolic processing schemes included algebraic manipulations of the received symbols (i.e. input letters, digits but also continuous decimal numbers of model architectures that were no neural networks). Continuous input encodings comprised time-varying signals (i.e. oscillation waves or input currents) or representations of decimal numbers that encoded the graded strength of stimulus attributes (i.e. graded feature vectors or maps).

Continuous processing schemes applied to neural networks architectures. Here, they comprised activation functions of artificial neurons (Hopfield and Tank 1986; Maass 1997) or reflected time-varying neural activities by differentiable variables (i.e. rate-based models). Binary input encodings were combinations of integer zeroes and ones where ‘one’ represented the presence and ‘zero’ the absence of a stimulus attribute or input activity. In binary processing schemes of neural networks, ‘one’ represented the spiking and ‘zero’ the silence of neurons (Hopfield 1982; Maass 1997). Finally, in spiking input encodings, stimuli were (spatio-temporal) spike trains. Spiking processing schemes were based on the discontinuous time points of neural action potential generation (Maass 1997; Izhikevich 2004). Both, binary and spiking processing schemes were considered distributed encodings.

(d) The sub-categories of the feature ‘learning mechanism’ ranged from algebraic, supervised, reinforcement learning, semi-supervised to unsupervised. Algebraic models performed learning or cognitive computations by mathematical operations (i.e. in a centralised fashion). In neural network models (i.e. distributed approaches), supervised learning (Knudsen 1994; Raymond and Medina 2018) included a teacher signal that shaped the weight changes in an error-driven process, for example by error-backpropagation (Rumelhart et al. 1986). Here, learning outcomes were judged according to a ‘right-wrong’ principle. During reinforcement learning (Kaelbling et al. 1996; Dayan and Niv 2008; Niv 2009; Arulkumaran et al. 2017; Neftci and Averbeck 2019), a reward signal shaped the weight changes whereby beneficial outputs were promoted. Thus, learning followed a goal-driven principle without the need to formulate ‘right’ or ‘wrong’ outputs (Neftci and Averbeck 2019). Semi-supervised learning was a combination of local plasticity in neural networks and the training of a readout layer via supervised or reinforcement learning. Here, both the stimulus as well as additional ‘right’, ’wrong’ or reward cues shaped the weight changes. Finally, during unsupervised learning (Barlow 1989), weight changes were purely driven by the stimulus (Doya 1999; Doya 2000) and the plasticity type implemented at synapses. Thus, learning occurred locally and in an online-fashion during the processing of the input.

(e) The sub-categories of the final feature ‘learning task’ ranged from ‘pattern or word learning’, ‘temporal structure or grammar learning’ to ‘sequence or sentence learning’. These sub-categories were motivated by the possibility that aspects of language processing and acquisition across ‘word to grammar to sentence’ levels might be explained by domain-general abilities like the learning of patterns, temporal structures or sequences. ‘Pattern or word learning’ tasks dealt with the extraction of meaningful entities from a (noisy) environment. For example, these could be visual patterns in a picture or words from the language signal. While in language the input is naturally always sequential, other domains might include temporally static patterns. Nevertheless, patterns and words were grouped into the same sub-category because of the ‘entity’-character of individual words. ‘Temporal structure or grammar learning’ tasks dealt with the sequential extraction of relevant patterns that were combinable with other entities, as for example grammatical NADs in combination with arbitrary intervening verbs. Thus, these tasks tested the ability to extract repeating or statistically significant structures from a stimulus stream. Finally, ‘sequence or sentence learning’ tasks dealt with the understanding, learning or production of complete sequences or sentences. As this capacity built up on the ability to learn and/or produce patterns or words as well as temporal structures or grammars, it was considered the most complex learning task reviewed here.

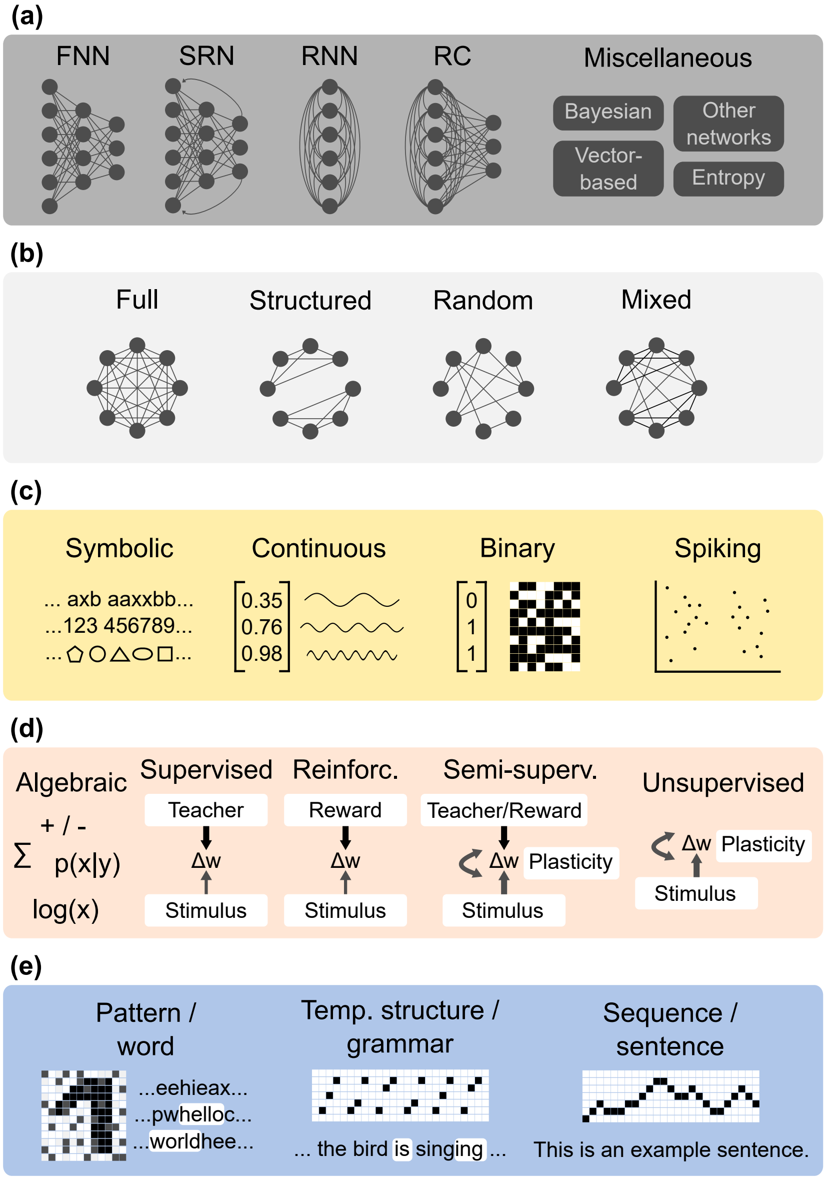


Figure 2.1: Model feature categories. (a) Architecture; FNN = feed forward network, SRN = simple recurrent network, RNN = recurrent neural network, RC = reservoir computing network. (b) Connectivity type. (c) Encoding scheme. (d) Learning mechanism; arrow widths indicate degree of contribution to weight changes (Δw). (e) Learning task.

Table 2.1: Language-specific learning models. **(a)** Architecture: miscellaneous, feed forward neural network, simple recurrent network, recurrent neural network, reservoir computing network, i.e. Misc. FNN SRN RNN RC. **(b)** Connectivity type: no connectivity, full/all-to-all, structured, random, mixed, i.e. None Full Struc. Rand. Mix.. **(c-i)** Encoding scheme of input and **(c-p)** encoding scheme of processing: symbolic, continuous, binary, spiking, i.e. Symb. Cont. Bin. Spike. **(d)** Learning mechanism: algebraic, supervised, reinforcement, semi-supervised, unsupervised, i.e. Alg. Super. Reinf. Semi. Unsu.. **(e)** Learning task: miscellaneous, 'pattern or word learning', 'temporal structure or grammar learning', 'sequence or sentence learning', i.e. Misc. P-W St-Gr Sq-Sn. Throughout: n.d. = not defined, s.n.a. = sub-categories not applicable.

Model	**a**	**b**	**c-i**	**c-p**	**d**	**e**
Hupkes et al. (2020)	Misc.	n.d.	Symb.	n.d.	Super.	Misc.
Doumas et al. (2008)	Misc.	n.d.	Symb.	Cont.	s.n.a.	
Yang and Piantadosi (2022)	Misc.	None	Symb.	Symb.	Alg.	
Heinrich and Wermter (2018)	Misc.	Rand.	Cont.	Cont.	Reinf.	
Jones and Macken (2018)	Misc.	None	Symb.	Symb.	n.d.	P-W
Perruchet and Vinter (1998)	Misc.	None	Symb.	Symb.	Alg.	
Thiessen and Pavlik Jr. (2013)	Misc.	None	Cont.	Symb.	Alg.	
Ueno et al. (2011)	Misc.	Struc.	Cont.	Cont.	Super.	
Garagnani et al. (2008)	Misc.	Struc.	Bin.	Cont.	Unsu.	
Garagnani et al. (2017)	Misc.	Struc.	n.d.	Bin.	Unsu.	
Tomasello et al. (2018)	Misc.	Struc.	Bin.	Bin.	Unsu.	
Tomasello et al. (2019)	Misc.	Struc.	Bin.	Bin.	Unsu.	
French et al. (2011)	FNN	n.d.	Symb.	Cont.	Super.	
French and Cottrell (2014)	FNN	n.d.	Symb.	Cont.	Super.	
Elman and Zipser (1988)	FNN	Full	Cont.	Cont.	Super.	
Christiansen et al. (1998)	SRN	n.d.	Cont.	Cont.	Super.	
Elman (1990)	SRN	Struc.	Bin.	Cont.	Super.	
Endress and Johnson (2021)	RNN	Full	Cont.	Cont.	Unsu.	
Fukino et al. (2016)	RC	Struc.	Cont.	Cont.	Super.	
Servan-Schreiber and Anderson (1990)	Misc.	None	Symb.	n.d.	n.d.	St-Gr
Uddén et al. (2012)	Misc.	None	Symb.	n.d.	n.d.	
Pothos (2010)	Misc.	None	Symb.	Symb.	Alg.	
Wang et al. (2019)	Misc.	None	Symb.	Symb.	Alg.	
Lieder et al. (2013)	Misc.	None	Cont.	Symb.	Alg.	
Thiessen and Pavlik Jr. (2013)	Misc.	None	Cont.	Symb.	Alg.	
Pulvermüller and Knoblauch (2009)	Misc.	Struc.	Cont.	Cont.	Unsu.	
Tong et al. (2007)	Misc.	Struc.	Bin.	Cont.	Super.	
Wacongne et al. (2012)	Misc.	Struc.	Spike	Spike	Unsu.	
Shultz and Bale (2001)	FNN	n.d.	Symb.	Cont.	Super.	
Shultz (1999)	FNN	n.d.	Cont.	Cont.	Super.	
Boucher and Dienes (2003)	FNN	n.d.	Bin.	Bin.	Alg.	
Elman and Zipser (1988)	FNN	Full	Cont.	Cont.	Super.	

Christiansen and Curtin (1999)	SRN	n.d.	n.d.	n.d.	Super.	St-Gr
Boucher and Dienes (2003)	SRN	n.d.	Cont.	n.d.	Super.	
Petersson et al. (2005)	SRN	n.d.	Cont.	Cont.	Super.	
Cleeremans and McClelland (1991)	SRN	Full	Bin.	n.d.	Super.	
Dominey and Ramus (2000)	SRN	Struc.	Cont.	Cont.	Semi.	
Fitz (2011)	RC	Rand.	Bin.	Cont.	Super.	
Duarte et al. (2014)	RC	Rand.	Bin.	Bin.	Semi.	
Weber et al. (2021)	Misc.	n.d.	Symb.	n.d.	n.d.	Sq-Sn
Goodman and Stuhlmüller (2013)	Misc.	None	Cont.	Symb.	Alg.	
Mitropolsky et al. (2021)	Misc.	Struc.	n.d.	Bin.	s.n.a	
Kriete et al. (2013)	Misc.	Struc.	Symb.	n.d.	s.n.a.	
Chang and Fitz (2014)	Misc.	Struc.	Symb.	n.d.	Super.	
Lakretz et al. (2021)	Misc.	Struc.	Symb.	Cont.	Super.	
Martin and Doumas (2017)	Misc.	Struc.	Cont.	Cont.	s.n.a.	
ten Oever and Martin (2021)	Misc.	Struc.	Cont.	Cont.	Super.	
Pitti et al. (2019)	Misc.	Struc.	Cont.	Spike	Semi.	
McClelland et al. (1989)	Misc.	Struc.	Bin.	Cont.	Super.	
Rabovsky et al. (2018)	Misc.	Struc.	Bin.	Cont.	Super.	
Fitz et al. (2020)	RNN	Rand.	Spike	Spike	s.n.a.	
Hinaut et al. (2015)	RC	Rand.	Symb.	Cont.	Super.	
Juven and Hinaut (2020)	RC	Rand.	Cont.	Cont.	Super.	
Dinh and Hinaut (2020)	RC	Rand.	Cont.	Cont.	Super.	
Hinaut and Dominey (2013)	RC	Rand.	Bin.	Cont.	Super.	

In the following, reviewed models were sorted by the increasing complexity of the learning tasks, shortly mentioning the particularly solved problem:

Misc. Miscellaneous In both language-specific and domain-general models, a fraction of studies could not be attributed to the previously defined sub-categories of learning tasks, giving rise to the 'Miscellaneous' task sub-category. In language-specific models (Tab. 2.1, see 'Misc.'), these miscellaneous tasks ranged from relational concept learning (Doumas et al. 2008), formal language acquisition (Yang and Piantadosi 2022) to the investigation of linguistic compositionality (Hupkes et al. 2020) and the implementation of embodied language acquisition in a humanoid robot (Heinrich and Wermter 2018). In domain-general models (Tab. 2.2, see 'Misc.'), miscellaneous tasks comprised the performance of goal-directed planning (Basanisi et al. 2020) and solving a bandwidth of cognitive tasks (Eliasmith 2005; Eliasmith et al. 2012; Stewart et al. 2012).

P-W Pattern or word learning 'Pattern or word learning' tasks in language-specific models (Tab. 2.1, see 'P-W') ranged from syllable recognition (Elman and Zipser 1988), pitch pattern perception (Fukino et al. 2016), phonemic category learning (Thiessen and Pavlik Jr. 2013), word prediction (Elman 1990), word chunking or segmentation (Perruchet and Vinter 1998; Christiansen et al.

1998; French et al. 2011; French and Cottrell 2014; Jones and Macken 2018; Endress and Johnson 2021) to word learning (Garagnani et al. 2008; Ueno et al. 2011; Garagnani et al. 2017; Tomasello et al. 2018; Tomasello et al. 2019). 'Pattern or word learning' tasks in domain-general models (Tab. 2.2, see 'P-W') dealt with exemplar learning and classification (Hintzman and Ludlam 1980), frequency judgements, generalisations and pair learning (Hintzman 1984), image classification (Fahlman and Lebiere 1989; Vinyals et al. 2015; Xu et al. 2015), visual chunking (Mareschal and French 2017), arbitrary pattern learning (Litwin-Kumar and Doiron 2014) and separation (Löffler and Gupta 2022) as well as the learning of written digits (Brader et al. 2007; Diehl and Cook 2015; Panda et al. 2018), spoken digits (Klampfl and Maass 2013) and geometric shapes (Zenke et al. 2015).

St-Gr **Temporal structure or grammar learning** 'Temporal structure or grammar learning' tasks in language-specific models (Tab. 2.1, see 'St-Gr') ranged from feature learning from continuous speech (Elman and Zipser 1988), serial structure learning (Dominey and Ramus 2000; Tong et al. 2007), grammatical rule learning (Cleeremans and McClelland 1991; Christiansen and Curtin 1999; Shultz 1999; Shultz and Bale 2001; Petersson et al. 2005; Pulvermüller and Knoblauch 2009; Pothos 2010; Duarte et al. 2014), adjacent dependency learning (Servan-Schreiber and Anderson 1990; Boucher and Dienes 2003) to non-adjacent dependency acquisition (Fitz 2011; Uddén et al. 2012; Thiessen and Pavlik Jr. 2013; Wang et al. 2019) and mismatch negativity production during oddball training (Wacongne et al. 2012; Lieder et al. 2013). 'Temporal structure or grammar learning' tasks in domain-general models (Tab. 2.2, see 'St-Gr') ranged from structure learning in sequential stimulus streams (Lazar et al. 2009; Lazar et al. 2011; Toutounji and Pipa 2014; Aswolinskiy and Pipa 2015; Hartmann et al. 2015; Cui et al. 2016) to pattern onset detection (Masquelier et al. 2008) and sequence chunking (Fonollosa et al. 2015).

Sq-Sn **Sequence or sentence learning** 'Sequence or sentence learning' tasks in language-specific models (Tab. 2.1, see 'Sq-Sn') ranged from oscillation-based tracking of hierarchical linguistic structure (Martin and Doumas 2017) and speech (ten Oever and Martin 2021), speech perception and production (Pitti et al. 2019), sentence processing (Fitz et al. 2020), sentence parsing (Mitropolsky et al. 2021), sentence comprehension (McClelland et al. 1989; Hinaut and Dominey 2013; Kriete et al. 2013; Rabovsky et al. 2018; Juven and Hinaut 2020; Dinh and Hinaut 2020), sentence production (Chang and Fitz 2014; Hinaut et al. 2015) to nested dependency learning in sentences (Lakretz et al. 2021), parsing of negative polarity items (Weber et al. 2021) and language understanding in social situations (Goodman and Stuhlmüller 2013). 'Sequence or sentence learning' tasks in domain-general models (Tab. 2.2, see 'Sq-Sn') ranged from the generation of pre-movement signals (Sussillo and Abbott 2009), sensorimotor control tasks (Tetzlaff et al. 2015; Rueckert et al. 2016; Enel et al. 2016; Rapp and Nawrot 2020), time series classification (Aswolinskiy et al. 2018), sequence processing (Duarte and Morrison 2014;

Setareh et al. 2018), sequence learning (Hoerzer et al. 2014; Liang et al. 2020; Klos et al. 2020; Maes et al. 2020; Vincent-Lamarre et al. 2020; Maes et al. 2021), sequence recall or replay (Dominey et al. 1995; Fonollosa et al. 2015; Tully et al. 2016; Fiebig and Lansner 2017; Klos et al. 2018; Martinez et al. 2019; Löffler and Gupta 2020) to sequence production (Panda and Roy 2017; Cazin et al. 2019; Fang et al. 2021).

Table 2.2: Domain-general learning models. **(a)** Architecture: miscellaneous, feed forward neural network, simple recurrent network, recurrent neural network, reservoir computing network, i.e. Misc. FNN SRN RNN RC. **(b)** Connectivity type: no connectivity, full/all-to-all, structured, random, mixed, i.e. None Full Struc. Rand. Mix.. **(c-i)** Encoding scheme of input and **(c-p)** encoding scheme of processing: symbolic, continuous, binary, spiking, i.e. Symb. Cont. Bin. Spike. **(d)** Learning mechanism: algebraic, supervised, reinforcement, semi-supervised, unsupervised, i.e. Alg. Super. Reinf. Semi. Unsu.. **(e)** Learning task: miscellaneous, 'pattern or word learning', 'temporal structure or grammar learning', 'sequence or sentence learning', i.e. Misc. P-W St-Gr Sq-Sn. Throughout: n.d. = not defined, s.n.a. = sub-categories not applicable.

Model	**a**	**b**	**c-i**	**c-p**	**d**	**e**
Eliasmith (2005)	Misc.	Struc.	Cont.	Spike	s.n.a.	Misc.
Basanisi et al. (2020)	Misc.	Struc.	Bin.	Bin.	Semi.	
Eliasmith et al. (2012)	Misc.	Struc.	Spike	Spike	s.n.a.	
Stewart et al. (2012)	Misc.	Struc.	Spike	Spike	s.n.a.	
Hintzman (1984)	Misc.	None	Symb.	Symb.	Alg.	P-W
Hintzman and Ludlam (1980)	Misc.	None	Cont.	Symb.	Alg.	
Vinyals et al. (2015)	Misc.	Struc.	Cont.	Cont.	Super.	
Xu et al. (2015)	Misc.	Struc.	Cont.	Cont.	Super.	
Brader et al. (2007)	Misc.	Struc.	Spike	Spike	Semi.	
Diehl and Cook (2015)	Misc.	Struc.	Spike	Spike	Unsu.	
Löffler and Gupta (2022)	Misc.	Mix.	Spike	Spike	s.n.a	
Mareschal and French (2017)	FNN	n.d.	n.d.	Cont.	Super.	
Fahlman and Lebiere (1989)	FNN	Full	Symb.	Cont.	Super.	
Panda et al. (2018)	FNN	Full	Spike	Spike	Unsu.	
Zenke et al. (2015)	RNN	Rand.	Spike	Spike	Unsu.	
Litwin-Kumar and Doiron (2014)	RNN	Rand.	Spike	Spike	Unsu.	
Klampfl and Maass (2013)	RC	Mix.	Spike	Spike	Semi.	
Fonollosa et al. (2015)	Misc.	Struc.	Bin.	Cont.	Unsu.	St-Gr
Cui et al. (2016)	Misc.	Struc.	Bin.	Bin.	Unsu.	
Masquelier et al. (2008)	Misc.	Struc.	Spike	Spike	Unsu.	
Lazar et al. (2009)	RC	Rand.	Bin.	Bin.	Semi.	
Lazar et al. (2011)	RC	Rand.	Bin.	Bin.	Semi.	
Toutounji and Pipa (2014)	RC	Rand.	Bin.	Bin.	Semi.	
Aswolinskiy and Pipa (2015)	RC	Rand.	Bin.	Bin.	Semi.	
Hartmann et al. (2015)	RC	Rand.	Bin.	Bin.	Semi.	

Dominey et al. (1995)	Misc.	Struc.	Cont.	Cont.	Reinf.	Sq-Sn
Martinez et al. (2019)	Misc.	Struc.	Cont.	Cont.	Unsu.	
Rapp and Nawrot (2020)	Misc.	Struc.	Cont.	Spike	Reinf.	
Liang et al. (2020)	Misc.	Struc.	Cont.	Spike	Semi.	
Fang et al. (2021)	Misc.	Struc.	Cont.	Spike	Semi.	
Fonollosa et al. (2015)	Misc.	Struc.	Bin.	Cont.	Unsu.	
Maes et al. (2021)	Misc.	Struc.	Spike	Spike	Semi.	
Löffler and Gupta (2020)	Misc.	Struc.	Spike	Spike	Unsu.	
Tully et al. (2016)	Misc.	Struc.	Spike	Spike	Unsu.	
Fiebig and Lansner (2017)	Misc.	Struc.	Spike	Spike	Unsu.	
Setareh et al. (2018)	Misc.	Struc.	Spike	Spike	Unsu.	
Panda and Roy (2017)	RNN	Rand.	Spike	Spike	Unsu.	
Klos et al. (2018)	RNN	Mix.	Spike	Spike	Unsu.	
Sussillo and Abbott (2009)	RC	Rand.	Cont.	Cont.	Super.	
Enel et al. (2016)	RC	Rand.	Cont.	Cont.	Super.	
Cazin et al. (2019)	RC	Rand.	Cont.	Cont.	Super.	
Aswolinskiy et al. (2018)	RC	Rand.	Cont.	Cont.	Super.	
Klos et al. (2020)	RC	Rand.	Cont.	Cont.	Super.	
Hoerzer et al. (2014)	RC	Rand.	Cont.	Cont.	Reinf.	
Tetzlaff et al. (2015)	RC	Rand.	Cont.	Cont.	Semi.	
Vincent-Lamarre et al. (2020)	RC	Rand.	Cont.	Spike	Super.	
Rueckert et al. (2016)	RC	Rand.	Bin.	Bin.	Reinf.	
Duarte and Morrison (2014)	RC	Rand.	Spike	Spike	Semi.	
Maes et al. (2020)	RC	Mix.	Spike	Spike	Semi.	

In this collection of state-of-the art sequence learning models, the class of interest for this thesis (i.e. single layers of randomly connected recurrent neural networks) was strongly underrepresented. Only a small fraction of studies used exclusively simple RNNs as a modelling substrate (Litwin-Kumar and Doiron 2014; Zenke et al. 2015; Panda and Roy 2017; Klos et al. 2018; Fitz et al. 2020; Endress and Johnson 2021). The majority of remaining neural networks either implemented specifically structured architectures (i.e. miscellaneous networks), FNNs, SRNs or RCs. However, specifically structured network models incorporated recurrent network modules throughout (Dominey et al. 1995; Garagnani et al. 2008; Pulvermüller and Knoblauch 2009; Wacongne et al. 2012; Fonollosa et al. 2015; Vinyals et al. 2015; Xu et al. 2015; Tully et al. 2016; Fiebig and Lansner 2017; Garagnani et al. 2017; Tomasello et al. 2018; Heinrich and Wermter 2018; Setareh et al. 2018; Tomasello et al. 2019; Pitti et al. 2019; Martinez et al. 2019; Basanisi et al. 2020; Liang et al. 2020; Lakretz et al. 2021; Maes et al. 2021; Mitropolsky et al. 2021; ten Oever and Martin 2021). Further, also RC models relied on a recurrent reservoir as an essential network component (Lazar et al. 2009; Sussillo and Abbott 2009; Lazar et al. 2011; Fitz 2011; Klampfl and Maass 2013; Hinaut and Dominey 2013; Duarte et al. 2014; Duarte and Morrison 2014; Hoerzer et al. 2014; Toutounji and Pipa 2014; Tetzlaff et al. 2015; Aswolinskiy and Pipa 2015; Hartmann et al. 2015; Hinaut et al. 2015; Fukino et al. 2016; Rueckert et al. 2016; Enel et al. 2016; Aswolinskiy et al. 2018; Cazin et al. 2019; Klos et al. 2020; Juven and Hinaut 2020; Dinh and Hinaut 2020; Vincent-Lamarre et al. 2020; Maes et al. 2020). Therefore, generic recurrent network modules can be seen as important features amongst other, more specific architectural structures (Dominey 2013). However, combinations of different module types inevitably lead to the difficulty of disentangling their individual contributions to the learning outcome. For example, are RC models successful sequence learners because of the recurrent reservoir or because of the supervised training of the readout layer? Here, increased RC learning performances by plastic instead of static reservoirs already highlighted the importance of learning in the recurrent layer (Lazar et al. 2009; Toutounji and Pipa 2014). This thesis therefore further investigated the idea that an essential proportion of the learning capacity might already be implemented in generic recurrent layers of neurons, i.e. in a small-scale recurrent building block of cortex without any further architectural pre-definitions. Next to architecture, the reviewed collection of language-specific models further revealed an underrepresentation of both distributed, i.e. binary or spiking, processing schemes and unsupervised learning. Throughout, a fraction of models was based on symbolic processing schemes that went together with algebraic implementations of learning (Perruchet and Vinter 1998; Pothos 2010; Goodman and Stuhlmüller 2013; Lieder et al. 2013; Thiessen and Pavlik Jr. 2013; Wang et al. 2019; Yang and Piantadosi 2022). Among distributed networks, semi-supervised learning (Duarte et al. 2014; Pitti et al. 2019) and fully unsupervised learning were occasionally applied (Wacongne et al. 2012; Garagnani et al. 2017; Tomasello et al. 2018; Tomasello et al. 2019). Even though neural networks in the form of FNNs, SRNs and RCs were widely implemented, their predominantly continuous processing schemes were primarily combined with semi- or fully supervised learning (Elman and Zipser 1988; Christiansen et al. 1998; Elman 1990; Shultz 1999; Shultz and Bale 2001; Dominey and Ramus 2000; Petersson et al. 2005; Fitz 2011; French et al. 2011; Hinaut and Dominey 2013; Hinaut et al. 2015; French and Cottrell 2014; Fukino et al. 2016; Juven and Hinaut 2020; Dinh and Hinaut 2020). Thus, the question how unsupervised, associative learning of language might be implemented in the brain remained largely unaddressed by theses state-of-the-art language learning models. On the contrary, in domain-general learning models, binary or spiking processing schemes were commonly applied and often endowed with unsupervised learning (Masquelier et al. 2008; Litwin-Kumar and Doiron 2014; Diehl and Cook 2015; Zenke et al. 2015; Cui et al. 2016; Tully et al. 2016; Fiebig and Lansner 2017; Panda and Roy 2017; Klos et al. 2018; Setareh et al. 2018; Panda et al. 2018; Löffler and Gupta 2020). Fully supervised learning mechanisms were again primarily implemented in neural networks that were based on continuous processing schemes (Fahlman and Lebiere 1989; Sussillo and Abbott 2009; Vinyals et al. 2015; Xu et al. 2015; Enel et al. 2016; Mareschal and French 2017; Aswolinskiy et al. 2018; Cazin et al. 2019; Klos et al. 2020). Taken together and throughout, unsupervised learning was mainly applied in distributed networks. Distributed and unsupervised domain-general models further had a wide applicability in sequence learning problems across the defined task categories. However, distributed networks remained underrepresented in language-specific sequence learning problems. Thus, their potential to model aspects of unsupervised language acquisition has to a large extent not been explored yet.

2.2.3 Unsupervised grammar learning models are rare

A closer look at language-specific neural network models in the 'temporal structure or grammar learning' task category revealed only two studies that applied fully unsupervised learning mechanisms (Pulvermüller and Knoblauch 2009; Wacongne et al. 2012). As of special interest to this thesis, their implementation of unsupervised learning was 'Hebbian' (Pulvermüller and Knoblauch 2009) or even based on spike-timing dependent plasticity (Wacongne et al. 2012). However, in the 'pattern or word learning' task category, some more unsupervised neural network approaches existed. Here, 'Hebbian' learning found application in a small scale recurrent network (Endress and Johnson 2021) and in more detailed cortical networks (Garagnani et al. 2008; Garagnani et al. 2017; Tomasello et al. 2018; Tomasello et al. 2019). In the most complex 'sequence or sentence learning' task category, however, no unsupervised learning models could be identified. Taken together, even though not widely applied yet, the identified unsupervised models provided important first hints towards a general applicability of unsupervised learning in neural networks for language-specific problems. Further, some symbolic or algebraic models implemented statistical learning with principles resembling unsupervised learning schemes. For example, the frequency effect of repeatedly co-occurring stimulus elements played a central role in a vector-based chunking model (Perruchet and Vinter 1998). But also principles of long-term memory like stimulus activation and decay were implemented as underlying mechanisms of learning (Thiessen and Pavlik Jr. 2013). Interestingly for this thesis, the effect of variability on NAD acquisition has been investigated already to some extent. Here, a beneficial effect of large variabilities in the intervening material has been demonstrated in a vector-based model of statistical learning (Thiessen and Pavlik Jr. 2013). Further, a liquid state machine model reproduced human performance curves across variability conditions (Fitz 2011). Thus, similar NAD benchmark variations like those identified for this thesis (Wilson et al. 2018) have already been applied for testing the performance of other language learning model classes. Taken together, some modelling approaches for the unsupervised acquisition of statistical structure and human language learning performances existed (Frank et al. 2010; Perruchet and Tillmann 2010; Rohrmeier and Cross 2014). However, the model review revealed that neurobiologically motivated approaches primarily dealt with other learning problems than grammar acquisition, as for example word learning (Garagnani et al. 2008; Garagnani et al. 2017; Tomasello et al. 2018; Tomasello et al. 2019). Models that dealt with the specific problem of NAD acquisition were, however, either vector-based (Thiessen and Pavlik Jr. 2013) or supervised (Fitz 2011). Therefore, neurobiologically motivated (i.e. distributed) networks for the unsupervised acquisition of non-adjacent grammars have not been studied to a satisfactory degree yet. This thesis therefore further aimed at filling this gap and at extending existing lines of (statistical) grammar learning modelling research (Chang et al. 2012; Zuidema et al. 2020).

Chapter 3

Modelling the spike-based acquisition of grammars

3.1 Grammar learning experiments

3.1.1 Preliminary remarks

The manuscript ‘Spike-based statistical learning explains human grammar acquisition performance’ contains the main materials and results of this thesis about modelling the spike-based acquisition of grammatical structure in form of isolated or nested non-adjacent dependencies (NADs) with a spiking recurrent neural network. The manuscript, from now on referred to as Lehfeldt et al. (n.d.), shows a selection of learning performance measures (i.e. so called separability measures) and the respectively underlying synapse parameters (i.e. synaptic weights). These demonstrated an accordance between modelled and human grammar acquisition performance patterns. The supplementary material of Lehfeldt et al. (n.d.) further contains the full collection of underlying synapse assembly weights and the grammatical training sequence statistics of all experiments and across all recurrent networks tested. Additional material for demonstrating basic dynamics of the model and further result data from the grammar learning experiments are presented in this thesis under ‘Model demonstration’ (see Sec. 3.2) and ‘Grammar learning experiments: additional material’ (see Sec. 3.3).

3.1.2 Manuscript and supplementary material

After thesis submission, a peer-reviewed version of the manuscript ‘Spike-based statistical learning explains human grammar acquisition performance’ (Lehfeldt et al. n.d.) has been published:

Lehfeldt, S., Mueller, J. L., and Pipa, G. (2022). “Spike-based statistical learning explains human performance in non-adjacent dependency learning tasks”. In: *Frontiers in Cognition* 1, 1026819. DOI: 10.3389/fcogn.2022.1026819.

This thesis contains numerous references to the original version of the manuscript. For reviewing the referenced content, please refer to the above mentioned publication.

3.2 Model demonstration

3.2.1 Spiking neural network simulator

The simulations of this thesis were conducted in Python using the 'Brian2' spiking neural network simulator (Stimberg et al. 2014; Stimberg et al. 2019). This is the follow-up version of the original 'Brian' simulator (Goodman and Brette 2008; Goodman and Brette 2009). 'Brian2' uses a mathematical string notation format for the definition of neuron and synapse models, whereby continuously evolving state variables are described by differential equations in combination with spike-triggered update routines (Stimberg et al. 2014). In the following, neuron and synapse models of the RNN were therefore presented by demonstrating their state variable and spike-triggered update dynamics during the online-processing of external spiking input. Throughout, simulations were conducted with a time step resolution of 0.1 ms. Neuron and synapse model equations were solved using the 'euler' method.

3.2.2 RNN connectivity

The network architecture of the recurrent neural network (RNN) model was described in detail in Sec. 3.1.2, 'Network architecture' (Lehfeldt et al. n.d.). Here, the synaptic connectivity of an example RNN was shown for demonstrating the randomised projections within and between the excitatory (E) and inhibitory (I) population as well as the one-to-one connection scheme between the input and excitatory neurons (Fig. 3.9).

3.2.3 Neuron model dynamics

3.2.3.1 LIF neuron

Leaky integrate-and-fire (LIF) neuron models of the RNN were derived from the LIF-SORN by Miner and Triesch (2016) and comprised the following set of variables and parameters: a membrane voltage (V_m), a membrane voltage time constant (τ_{V_m}), a resting voltage ($V_{\hookrightarrow}$), a reset voltage ($V_{\perp}$) and a threshold voltage for spiking (V_θ). While the activity state of a LIF neuron was modelled by the state variable V_m, the parameters τ_{V_m}, $V_{\hookrightarrow}$, $V_{\perp}$ and V_θ represented the parameter space within which LIF neuron dynamics evolved. Here, neuron model dynamics were demonstrated in small scale networks where an excitatory or inhibitory LIF neuron with a fixed V_θ received a Poisson spike train as input (Fig. 3.10). For a description of the LIF neuron model, see equations (3.1a) to (3.1c) in Sec. 3.1.2 'Leaky integrate-and-fire neurons (LIF) with intrinsic plasticity (IP)' (Lehfeldt et al. n.d.). Each input spike to a LIF neuron caused an increase of the neuron's V_m by the voltage value of the connecting synapse. For details, see equation (3.3) in Sec. 3.1.2 'Synaptic transmission' (Lehfeldt et al. n.d.). In the absence of input spikes V_m decayed, i.e. leaked, back to $V_{\hookrightarrow}$. However, under a repeated input activation, V_m integrated the received input spikes in an accumulative way. Thus, once it's value crossed V_θ, the emission of a spike was triggered. After the emission of a spike, V_m was reset to $V_{\perp}$ and a new input integration cycle started.

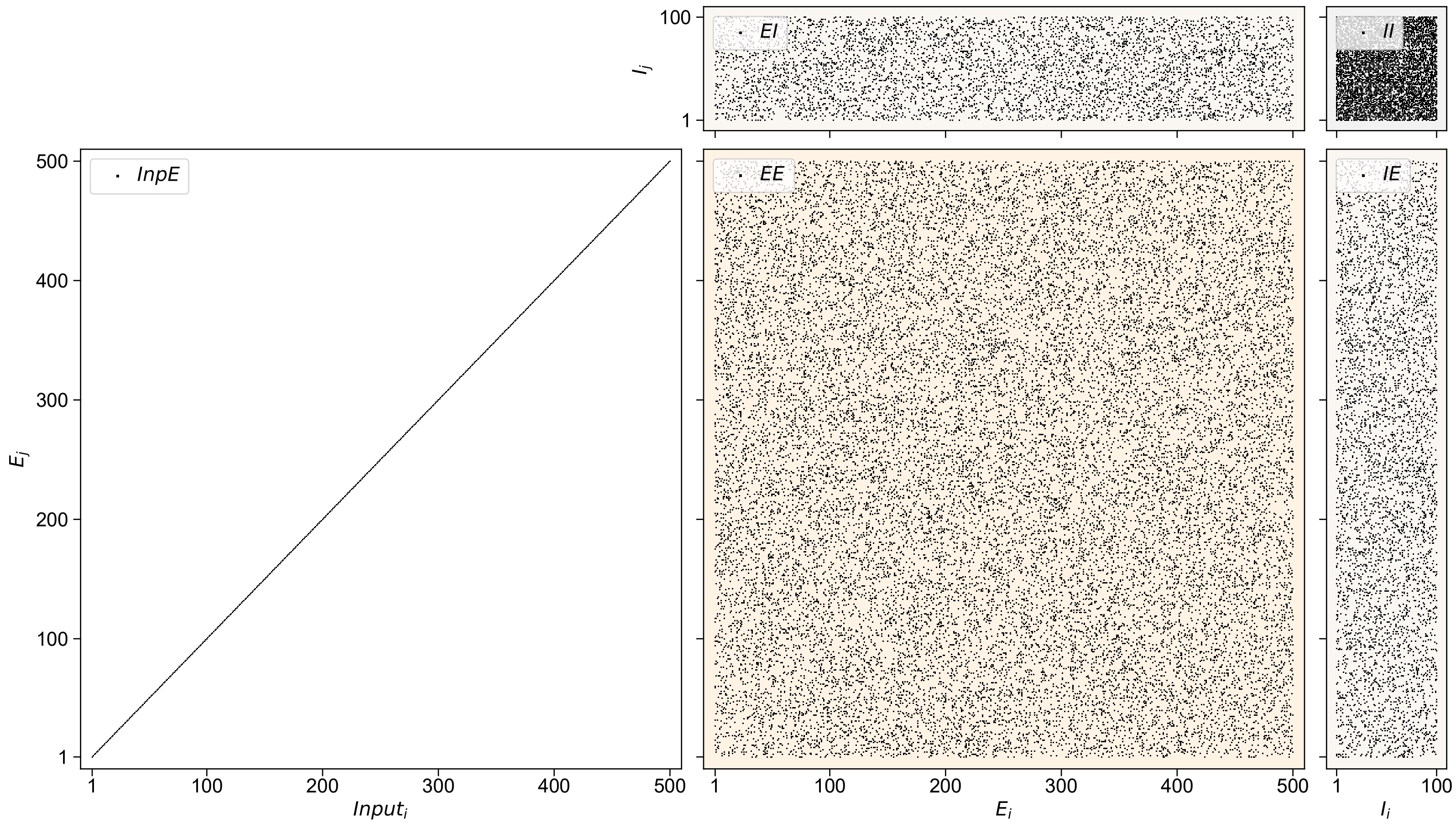


Figure 3.9: Connectivity of an example RNN from the grammar learning experiment series. Each dot represents a synaptic connection between a presynaptic (i) and postsynaptic (j) neuron. Synaptic connections ($p =$ connection probability): $InpE$ = input-to-excitatory (one-to-one connection scheme), EE = excitatory-to-excitatory ($p_{EE} = 0.1$), EI = excitatory-to-inhibitory ($p_{EI} = 0.1$), IE = inhibitory-to-excitatory ($p_{IE} = 0.1$), II = inhibitory-to-inhibitory ($p_{II} = 0.5$).

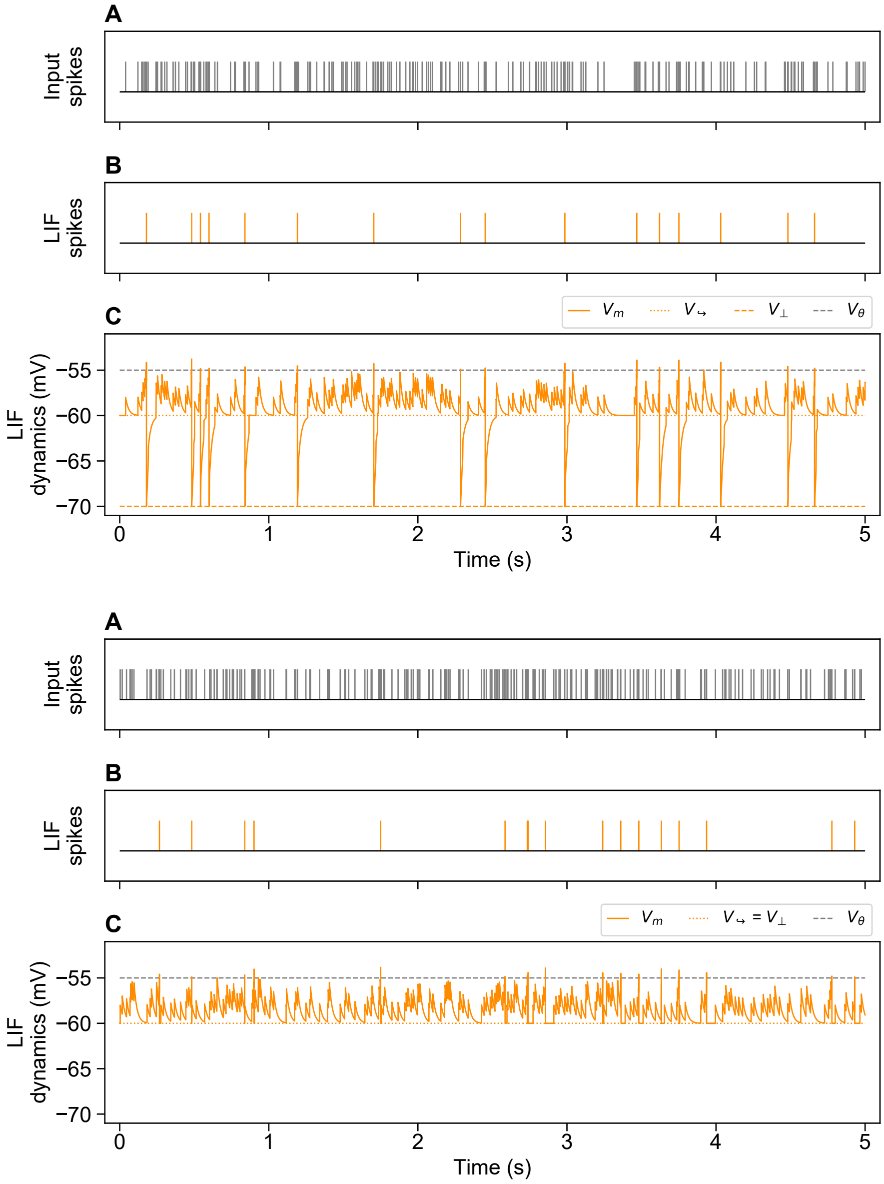


Figure 3.10: LIF neuron dynamics with fixed threshold for spiking. Excitatory (**top**) and inhibitory (**bottom**) LIF neuron. **A** Poisson input spikes. **B** LIF neuron output spikes. **C** LIF neuron variables; V_m = membrane voltage; $V_{\hookrightarrow}$ = resting voltage; $V_{\perp}$ = reset voltage; V_{θ} = threshold voltage. Neuron parameter space as in Lehfeldt et al. (n.d.).

3.2.3.2 LIF neuron with IP

As in the SORN (Lazar et al. 2009) and LIF-SORN (Miner and Triesch 2016), LIF neurons further comprised an intrinsic plasticity (IP) of V_{θ}. Here, the implementation of IP was derived from Stimberg et al. (2014). For a description of the IP model, see equations (3.2a) and (3.2b) in Sec. 3.1.2 'Leaky integrate-and-fire neurons (LIF) with intrinsic plasticity (IP)' (Lehfeldt et al. n.d.). The IP model converted V_{θ} into a neuronal state variable and introduced three additional parameters: a threshold

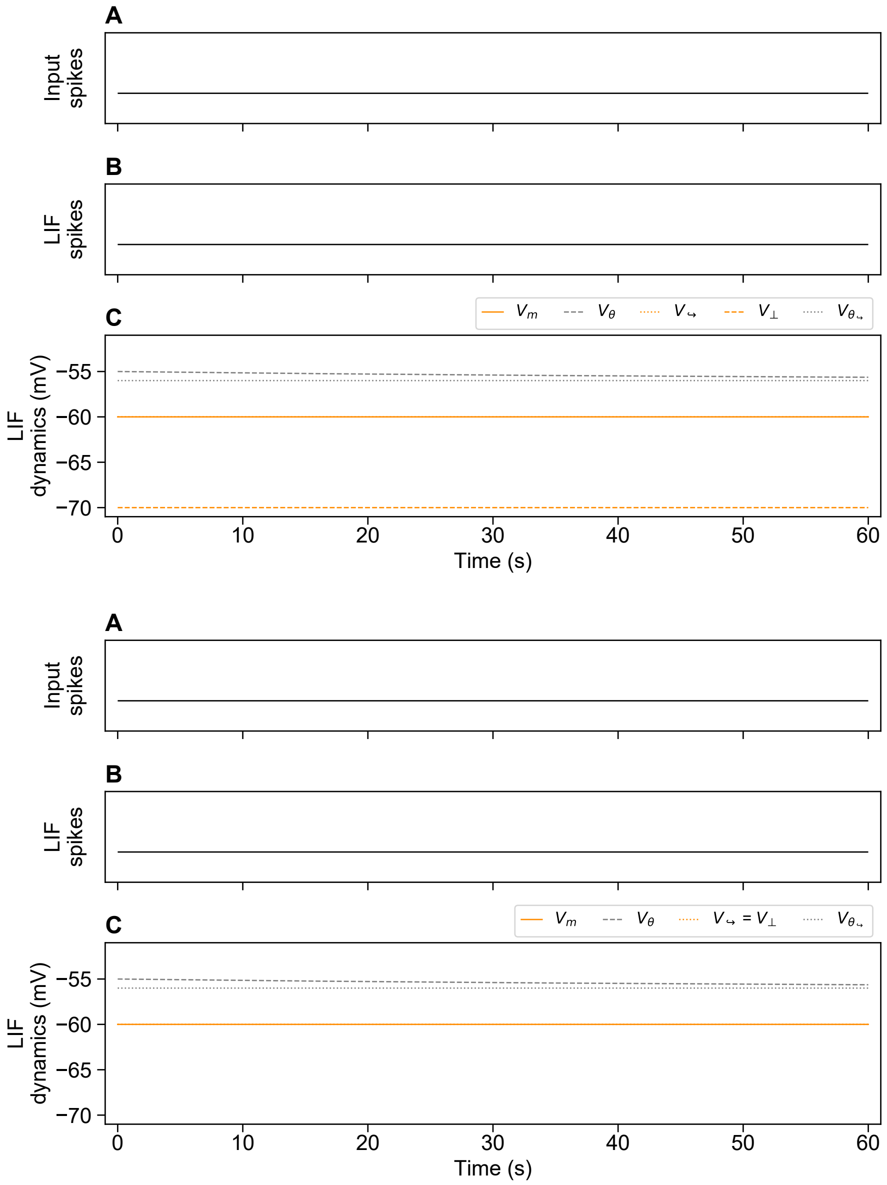


Figure 3.11: IP dynamics in the absence of LIF neuron spiking activity. Excitatory (**top**) and inhibitory (**bottom**) LIF neuron. **A** No input spikes. **B** No LIF neuron output spikes. **C** LIF neuron variables; V_m = membrane voltage; V_θ = threshold voltage; $V_\hookrightarrow$ = resting voltage; $V_\perp$ = reset voltage; $V_{\theta_\hookrightarrow}$ = threshold resting voltage. Neuron parameter space as in Lehfeldt et al. (n.d.).

resting voltage ($V_{\theta_\hookrightarrow}$), a threshold voltage time constant (τ_{V_θ}) and a threshold voltage increase (V_{θ_Δ}). Thus, instead of being fixed at a constant value, V_θ decayed to $V_{\theta_\hookrightarrow}$ in the absence of LIF neuron spiking activity (Fig. 3.11). After the emission of a LIF neuron spike, V_θ was increased by V_{θ_Δ}. Therefore, V_θ successively increased during a repeated emission of spikes that occurred within the temporal window of the V_θ decay as defined by τ_{V_θ} (Fig. 3.12).

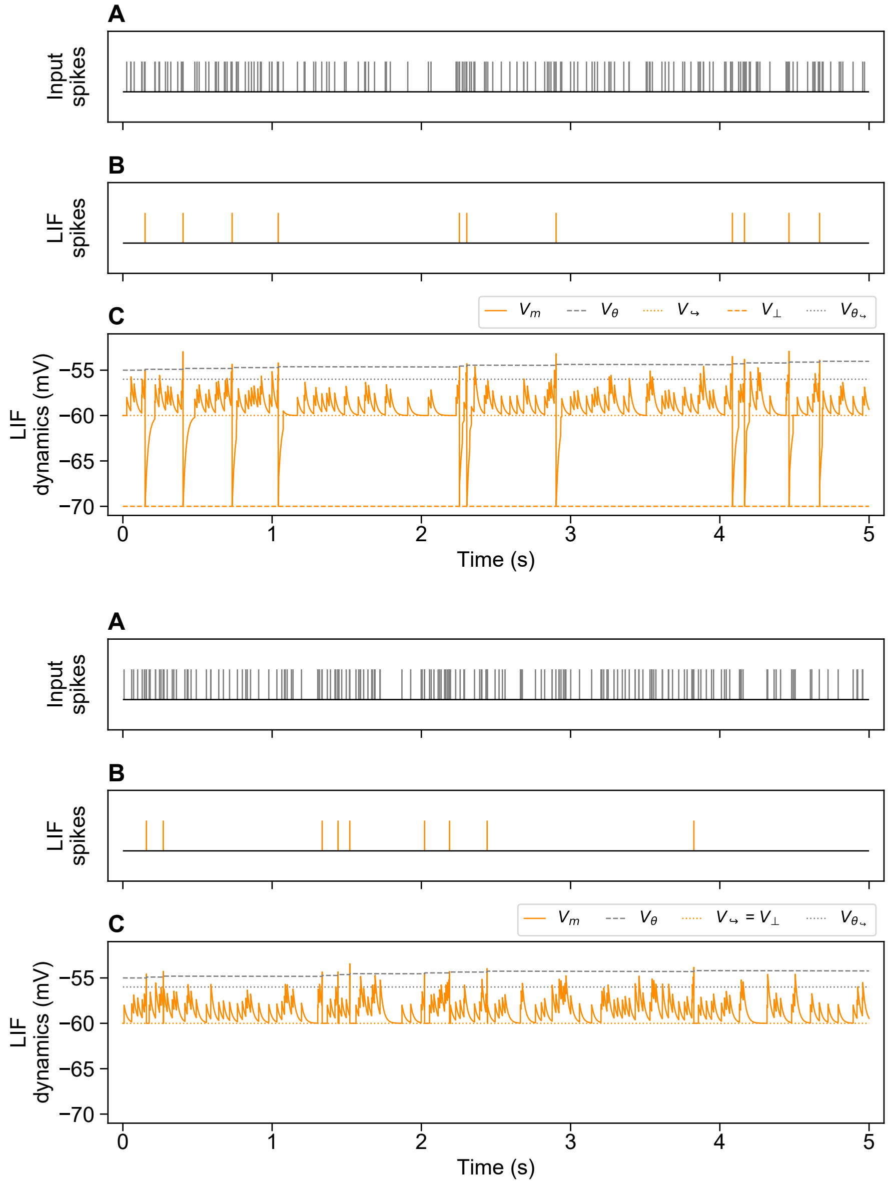


Figure 3.12: LIF neuron dynamics with IP. Excitatory (**top**) and inhibitory (**bottom**) LIF neuron. **A** Poisson input spikes. **B** LIF neuron output spikes. **C** LIF neuron variables; V_m = membrane voltage; V_θ = threshold voltage; $V_\hookrightarrow$ = resting voltage; $V_\perp$ = reset voltage; $V_{\theta_\hookrightarrow}$ = threshold resting voltage. Neuron parameter space as in Lehfeldt et al. (n.d.).

3.2.4 Synapse model dynamics

3.2.4.1 STDP synapse

As in the SORN (Lazar et al. 2009) and LIF-SORN (Miner and Triesch 2016), learning occurred in plastic excitatory-to-excitatory (EE) synapses of the RNN. Here, spike-timing dependent plasticity (STDP) was combined with a short- to long-term memory transition (SLMT) mechanism that was developed in this thesis. Before introducing SLMT, the dynamics of a STDP synapse were demonstrated in a small

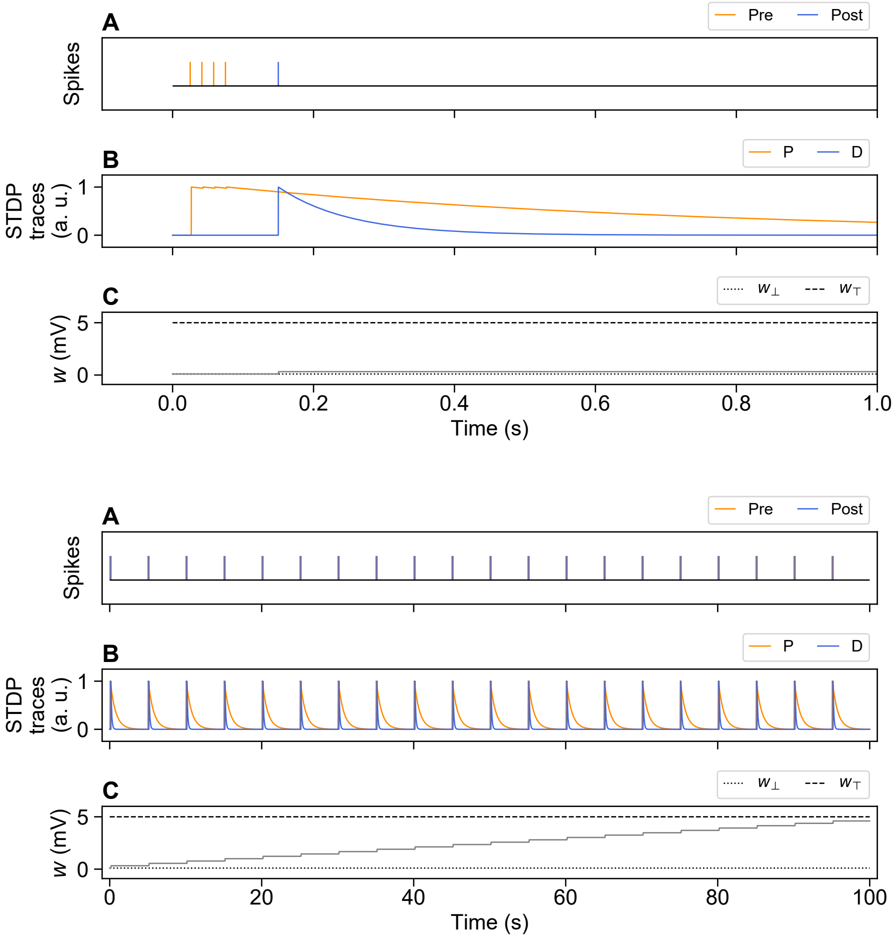


Figure 3.13: STDP synapse experiencing a series of *pre-post* spike pairings. Zoom into the first second (**top**) and full simulation run (**bottom**). **A** Spikes at the synapse; Pre = presynaptic spikes; Post = postsynaptic spikes. **B** STDP traces; P = potentiation trace; D = depression trace. **C** Synaptic weight; $w_\perp$ = lower weight boundary; $w_\top$ = upper weight boundary. Synapse parameter space as in Lehfeldt et al. (n.d.).

scale network with single pre- and postsynaptic neurons that were connected by a single plastic synapse (Fig. 3.13 and Fig. 3.14). For a description of the STDP synapse model, see equations (3.4a) to (3.4c) in Sec. 3.1.2 'Spike-timing dependent plasticity (STDP) with short- to long-term memory transition (SLMT)' (Lehfeldt et al. n.d.). A plastic STDP synapse was described by three state variables: a synaptic weight (w), a potentiation trace (P) and a depression trace (D). Further, the parameter space comprised a lower ($w_\perp$) and upper ($w_\top$) weight boundary, a potentiation trace time constant (τ_P), a depression trace time constant (τ_D), a weight potentiation maximum ($P_{\top w}$), a weight depression maximum ($D_{\perp w}$) and a trace maximum ($T_\top$). The temporal correlation of pre- and postsynaptic spikes at the synapse triggered weight changes by the following update routines. While the occurrence of presynaptic spikes elicited a maximisation of P, the occurrence of postsynaptic spikes elicited a maximisation of D. In the absence of pre- or postsynaptic

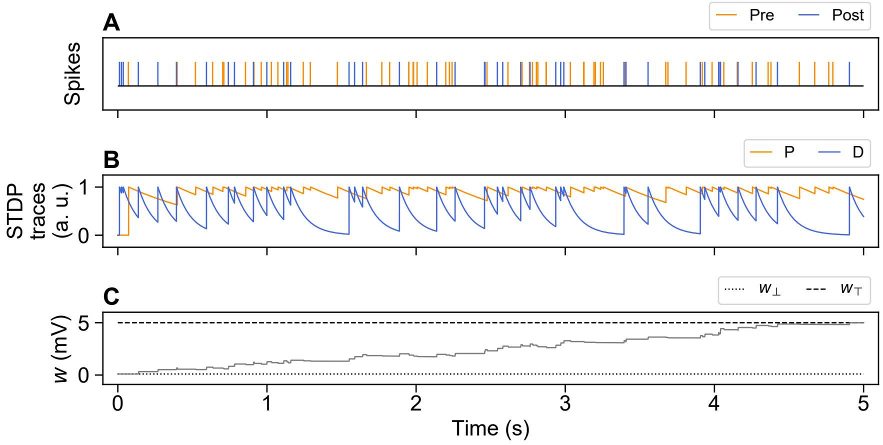


Figure 3.14: STDP synapse experiencing a Poisson pre- and postsynaptic input. **A** Spikes at the synapse; Pre = presynaptic spikes; Post = postsynaptic spikes. **B** STDP traces; P = potentiation trace; D = depression trace. **C** Synaptic weight; $w_\perp$ = lower weight boundary; $w_\top$ = upper weight boundary. Synapse parameter space as in Lehfeldt et al. (n.d.).

spikes at the synapse, both P and D decayed back to zero within temporal windows defined by τ_P and τ_D, respectively. After a *pre-post* spike pairing, i.e. after the occurrence of a postsynaptic spike, w was increased by the potentiation maximum value $P_{\top_w}$ that was scaled by the current value of P. Equivalently, after a *post-pre* spike pairing, i.e. after the occurrence of a presynaptic spike, w was decreased by the depression maximum value $D_{\perp_w}$ that was scaled by the current value of D. Thus, while positively correlated spike pairings induced potentiation, negatively correlated spike pairings induced depression. Further, larger weight changes were induced for temporally close spike pairings while smaller weight changes occurred for temporally distant spike pairings. Figure 3.13 demonstrated a constantly growing w at a STDP synapse that experienced a series of *pre-post* spike pairings. Figure 3.14 demonstrated the dynamics of w during a Poisson pre- and postsynaptic input where both *pre-post* and *post-pre* spike pairings occurred concurrently and thus, also both potentiation and depression. However, as STDP was skewed towards potentiation by $\tau_P > \tau_D$ and $P_{\top_w} > D_{\perp_w}$, w gradually grew towards $w_\top$.

3.2.4.2 STDP synapse with SLMT

The STDP synapse model was extended by a short- to long-term memory transition (SLMT) mechanism that introduced a decay of w with a plastic weight time constant (τ_w). The SLMT mechanism developed in this thesis bears similarity to the 'adaptive synaptic plasticity (ASP)' mechanism (Panda and Roy 2017; Panda et al. 2018) as both included a weight decay with adaptive τ_w. However, SLMT was symmetric and its plasticity fully based on the STDP traces. For a full description of the SLMT model, see equations (3.5a) to (3.5c) in Sec. 3.1.2 'Spike-timing dependent plasticity (STDP) with short- to long-term memory transition (SLMT)' (Lehfeldt et al. n.d.). A STDP synapse with SLMT included the plastic τ_w as an additional synaptic state

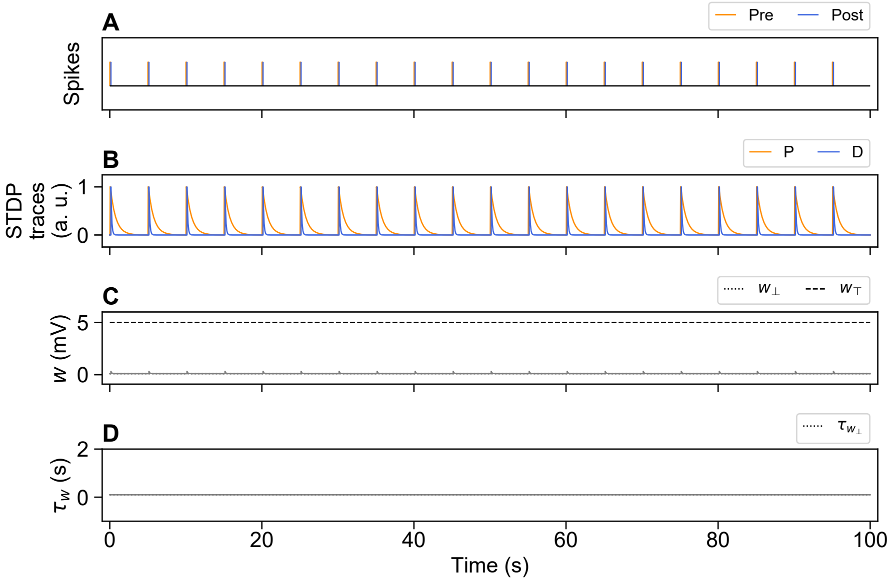


Figure 3.15: STDP synapse with SLMT experiencing a series of *pre-post* spike pairings while τ_w was non-plastic and short. **A** Spikes at the synapse; Pre = presynaptic spikes; Post = postsynaptic spikes. **B** STDP traces; P = potentiation trace; D = depression trace. **C** Synaptic weight; $w_\perp$ = lower weight boundary; $w_\top$ = upper weight boundary. **D** Synaptic weight time constant; $\tau_{w_\perp}$ = lower weight time constant boundary. Synapse parameter space as in Lehfeldt et al. (n.d.), except for $D_{\perp\tau_w} = 0$ s and $P_{\top\tau_w} = 0$ s.

variable. Further, the synaptic parameter space was extended by a resting weight ($w_\perp$), a weight time constant potentiation maximum ($P_{\top\tau_w}$), a weight time constant depression maximum ($D_{\perp\tau_w}$) and a lower ($\tau_{w_\perp}$) and upper ($\tau_{w_\top}$) weight time constant boundary. Weight time constant changes followed the same spike-triggered update routines as w in STDP. Thus, after the occurrence of a postsynaptic spike at the synapse, τ_w was potentiated, i.e. increased by the potentiation maximum $P_{\top\tau_w}$ that was scaled by the current value of P. Equivalently, after the occurrence of a presynaptic spike, τ_w was decreased by the depression maximum value $D_{\perp\tau_w}$ that was scaled by the current value of D. In the following, the dynamics of a STDP synapse with SLMT were demonstrated again in a small scale network with single pre- and postsynaptic neurons that were connected by a single plastic synapse (Fig. 3.15, Fig. 3.16, Fig. 3.17, Fig. 3.18 and Fig. 3.19). Figure 3.15 demonstrated the dynamics of w with a non-plastic and short τ_w during the repeated experience of *pre-post* spike pairings. As τ_w was short and remained constant at its initialisation value, plasticity-induced weight changes were not permanent and w decayed quickly back to $w_\perp$. Thus, despite a repeated experience of *pre-post* spike pairings, the synaptic weight did not grow on a temporally consistent scale. Figure 3.16 demonstrated how the growth of w with a plastic τ_w became possible under a repeated experience of *pre-post* spike pairings. In the beginning, τ_w was short and w decayed quickly towards $w_\perp$. However, throughout the repeated experience of *pre-post* spike pairings, τ_w constantly grew so that w gradually decayed on slower time scales. In turn, this also

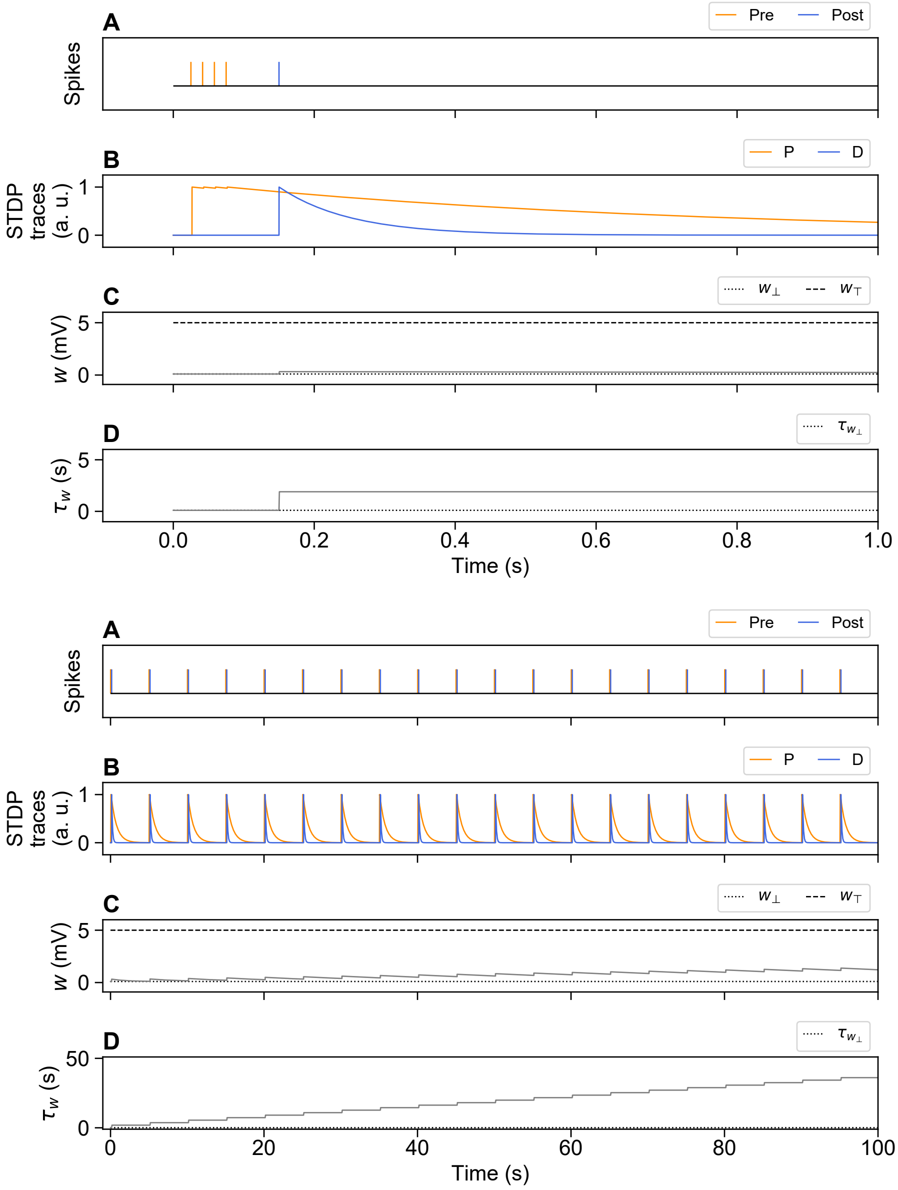


Figure 3.16: STDP synapse with SLMT experiencing a series of *pre-post* spike pairings. Zoom into the first second (**top**) and full simulation run (**bottom**). **A** Spikes at the synapse; Pre = presynaptic spikes; Post = postsynaptic spikes. **B** STDP traces; P = potentiation trace; D = depression trace. **C** Synaptic weight; $w_\perp$ = lower weight boundary; $w_\top$ = upper weight boundary. **D** Synaptic weight time constant; $\tau_{w_\perp}$ = lower weight time constant boundary. Synapse parameter space as in Lehfeldt et al. (n.d.), except for $D_{\perp\tau_w}$ = -1 s and $P_{\top\tau_w}$ = 2 s.

allowed w to grow as soon as the *pre-post* spike pairings, i.e. the potentiation events, occurred within temporal windows that were smaller than τ_w. As the growth of w was constantly hindered by the decay and depended on the growth of τ_w, learning at STDP synapses with SLMT took comparably longer as at STDP synapses. While w of a STDP synapse already approached $w_\top$ within the first 100 s of simulation (Fig. 3.13), w of a STDP synapse with SLMT still remained in a lower regime after the first

100 s of simulation (Fig. 3.16). Figure 3.17 further demonstrated that the growth of w in a STDP synapse with SLMT required substantially more time to approach $w_\top$, i.e. here approximately four times longer. Figure 3.18 demonstrated the dynamics of w during a Poisson pre- and postsynaptic input and showed that w in a STDP synapse with SLMT reached an elevated strength level but not the maximum, while w in a STDP synapse already saturated at $w_\top$ (Fig. 3.14). Finally, the influence of temporal gaps between *pre-post* spike pairings but also stimulus pauses between *post-pre* spike pairings was illustrated in Figure 3.19. With short temporal gaps, P was stronger at the onset of the postsynaptic spikes as compared to P with long temporal gaps. As a result, τ_w and w grew comparably stronger during the first potentiation event with a short gap. Further, with short stimulus pauses w had a shorter decay period until the occurrence of the next *pre-post* spike paring. Thus, w was at a comparably higher level when the second potentiation event occurred and consequently grew even stronger. Taken together, two consecutive *pre-post* input spike pairings with short gaps and pauses induced comparably stronger τ_w and w increases as compared to an input with long gaps and pauses. Even though with short pauses, also a notable decrease of τ_w and w occurred during the first *post-pre* spike pairings, this depression event did not outweigh the enhancing effect of short gaps and pauses.

3.2.5 Neuron and synapse variables at a glance

Next to the detailed demonstrations of individual model dynamics, all variables of excitatory LIF neurons with IP and plastic synapses were demonstrated at a glance in Figure 3.20 (STDP synapse) and Figure 3.21 (STDP synapse with SLMT). Here, again, the dynamics of variables were demonstrated in a small scale network where a pre- and postsynaptic excitatory LIF neuron both received individual Poisson spiking input. Further, pre- and postsynaptic LIF neurons were connected by a single plastic synapse.

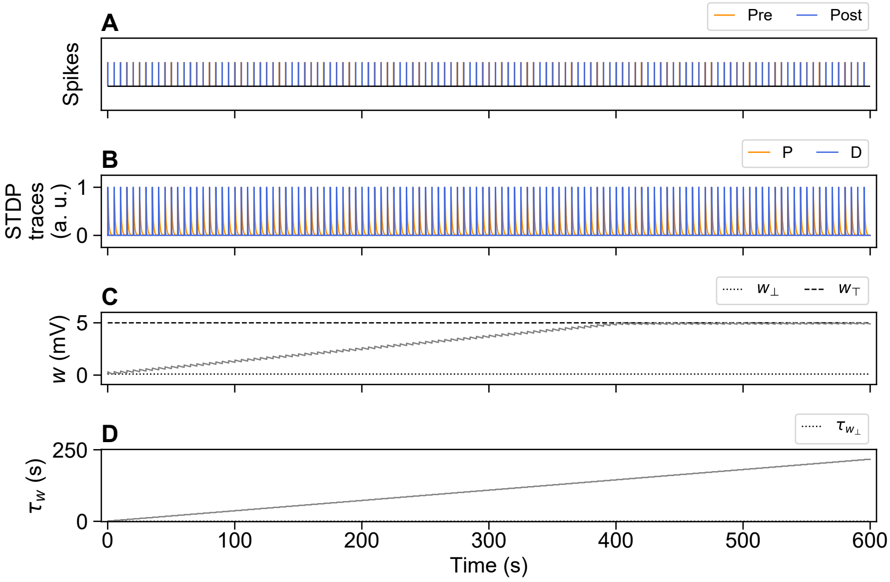


Figure 3.17: STDP synapse with SLMT experiencing a series of *pre-post* spike pairings. **A** Spikes at the synapse; Pre = presynaptic spikes; Post = postsynaptic spikes. **B** STDP traces; P = potentiation trace; D = depression trace. **C** Synaptic weight; $w_\perp$ = lower weight boundary; $w_\top$ = upper weight boundary. **D** Synaptic weight time constant; $\tau_{w_\perp}$ = lower weight time constant boundary. Synapse parameter space as in Lehfeldt et al. (n.d.), except for $D_{\perp\tau_w}$ = -1 s and $P_{\top\tau_w}$ = 2 s.

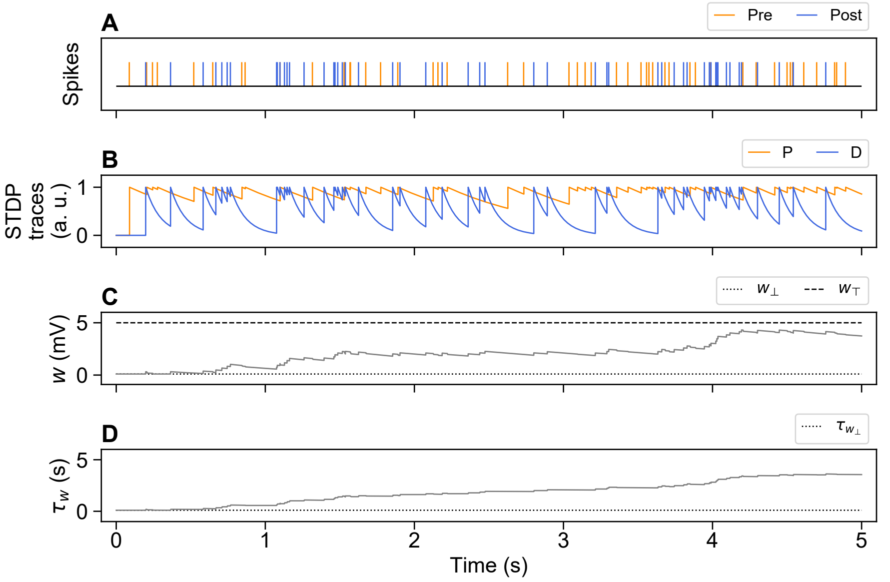


Figure 3.18: STDP synapse with SLMT experiencing a Poisson pre- and postsynaptic input. **A** Spikes at the synapse; Pre = presynaptic spikes; Post = postsynaptic spikes. **B** STDP traces; P = potentiation trace; D = depression trace. **C** Synaptic weight; $w_\perp$ = lower weight boundary; $w_\top$ = upper weight boundary. **D** Synaptic weight time constant; $\tau_{w_\perp}$ = lower weight time constant boundary. Synapse parameter space as in Lehfeldt et al. (n.d.), except for $D_{\perp\tau_w}$ = -0.05 s and $P_{\top\tau_w}$ = 0.1 s.

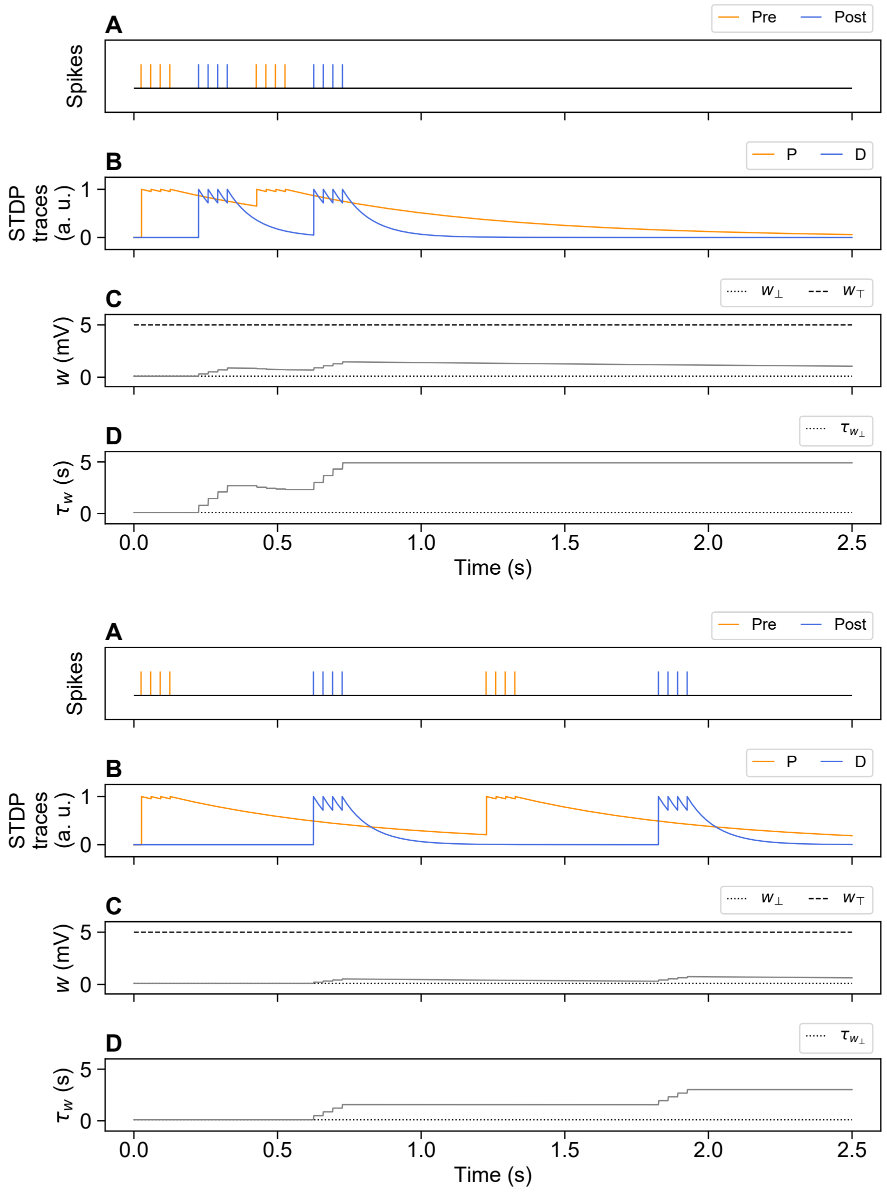


Figure 3.19: STDP synapse with SLMT experiencing *pre-post* and *post-pre* spike pairings with short and long temporal gaps/pauses. Comparably short (**top**) and long (**bottom**) temporal gaps/pauses. **A** Spikes at the synapse; Pre = presynaptic spikes; Post = postsynaptic spikes. **B** STDP traces; P = potentiation trace; D = depression trace. **C** Synaptic weight; $w_\perp$ = lower weight boundary; $w_\top$ = upper weight boundary. **D** Synaptic weight time constant; $\tau_{w_\perp}$ = lower weight time constant boundary. Synapse parameter space as in Lehfeldt et al. (n.d.), except for $D_{\perp\tau_w}$ = -0.4 s and $P_{\top\tau_w}$ = 0.8 s.

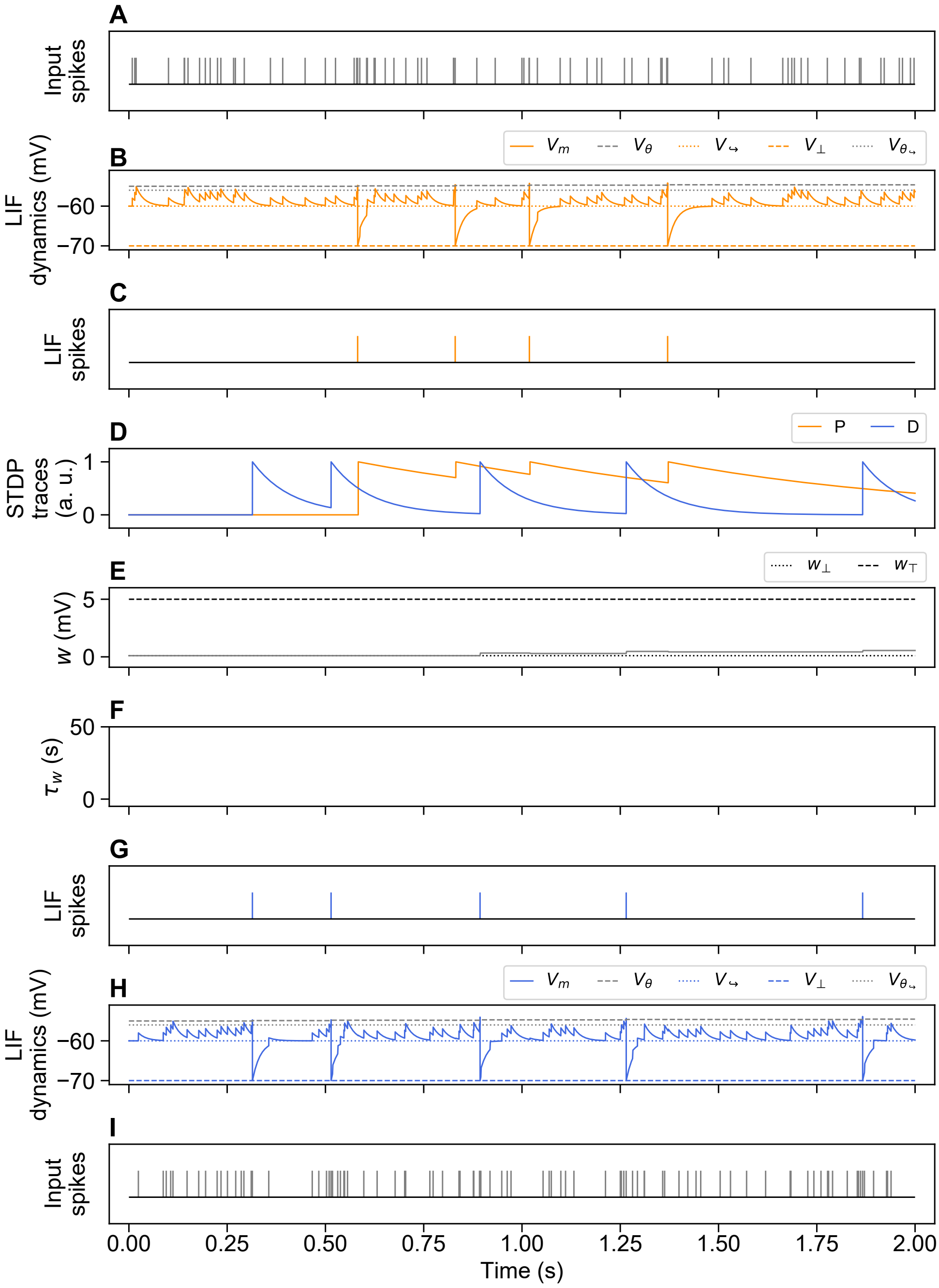


Figure 3.20: Variables of excitatory LIF neurons with IP and a STDP synapse. **A** Poisson input spikes to presynaptic LIF neuron. **B** Presynaptic LIF neuron variables. **C** Presynaptic LIF neuron output spikes. **D** Synaptic STDP traces. **E** Synaptic weight. **F** Synaptic weight time constant (empty here, as τ_w was not existent in a STDP synapse). **G** Postsynaptic LIF neuron output spikes. **H** Postsynaptic LIF neuron variables. **I** Poisson input spikes to postsynaptic LIF neuron. V_m = membrane voltage; V_θ = threshold voltage; $V_\hookrightarrow$ = resting voltage; $V_\perp$ = reset voltage; $V_{\theta_\hookrightarrow}$ = threshold resting voltage; P = potentiation trace; D = depression trace; $w_\perp$ = lower weight boundary; $w_\top$ = upper weight boundary. Neuron and synapse parameter spaces as in Lehfeldt et al. (n.d.).

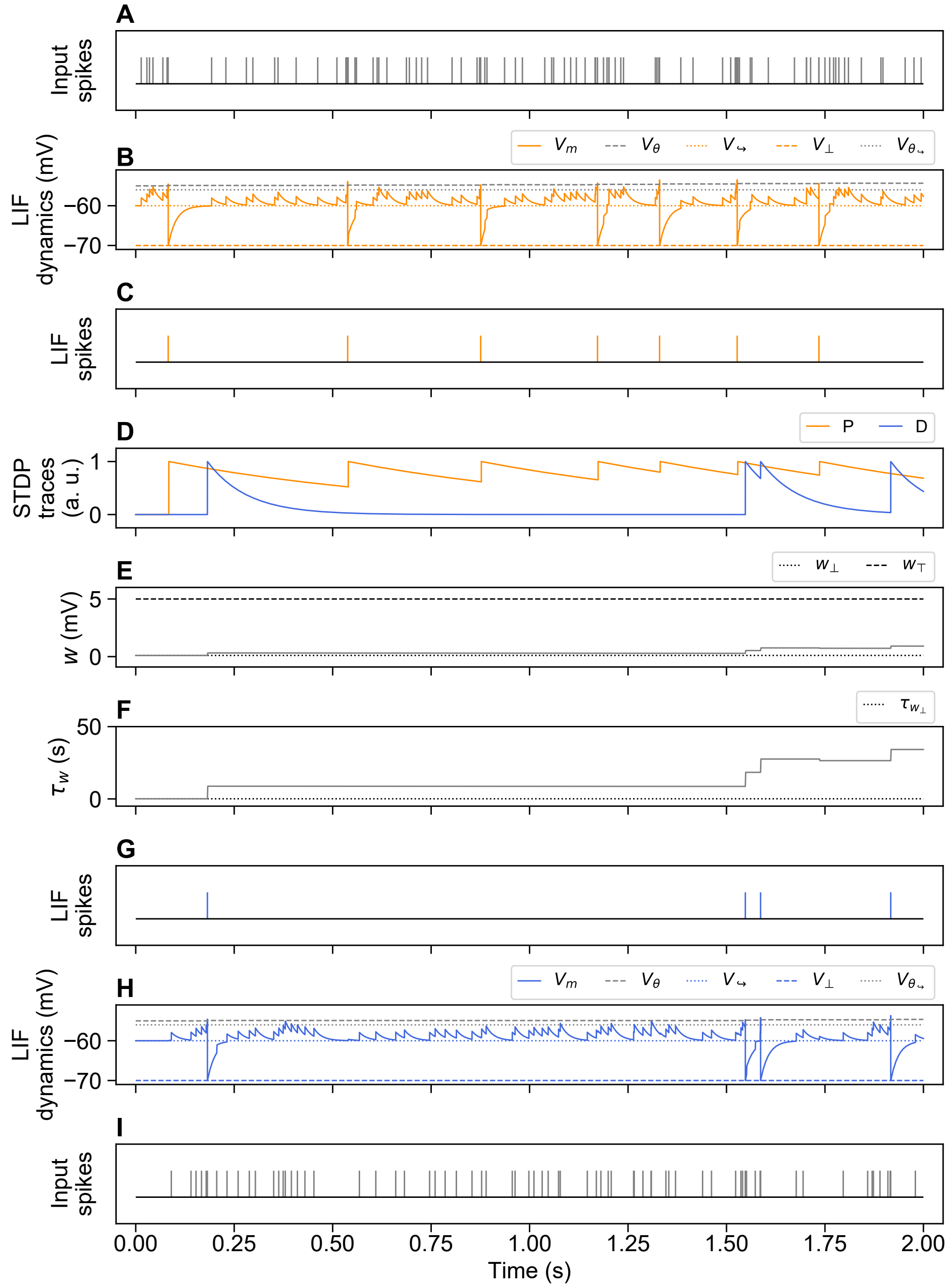


Figure 3.21: Variables of excitatory LIF neurons with IP and a STDP synapse with SLMT. **A** Poisson input spikes to presynaptic LIF neuron. **B** Presynaptic LIF neuron variables. **C** Presynaptic LIF neuron output spikes. **D** Synaptic STDP traces. **E** Synaptic weight. **F** Synaptic weight time constant. **G** Postsynaptic LIF neuron output spikes. **H** Postsynaptic LIF neuron variables. **I** Poisson input spikes to postsynaptic LIF neuron. V_m = membrane voltage; V_θ = threshold voltage; $V_{\hookrightarrow}$ = resting voltage; $V_\perp$ = reset voltage; $V_{\theta_{\hookrightarrow}}$ = threshold resting voltage; P = potentiation trace; D = depression trace; $w_\perp$ = lower weight boundary; $w_\top$ = upper weight boundary; $\tau_{w_\perp}$ = lower weight time constant boundary. Neuron and synapse parameter spaces as in Lehfeldt et al. (n.d.).

3.3 Grammar learning experiments: additional material

3.3.1 Spiking grammar sequences

In the grammar learning experiments (i.e. ‘X variability’, ‘X chunk size’, ‘Pause variability’ and ‘Emphasis’), isolated and nested NADs were presented as spiking sequences to the RNN. For details, see Sec. 3.1.2 ‘Spiking training sequences’ (Lehfeldt et al. n.d.). In the following, examples of spiking RNN activities during the grammar learning experiments were shown. Throughout, a zoom into the first 10 seconds of training was presented. Examples of experiment ‘X variability’ were shown for the conditions NL $=1$ and $N_X = 1$ (Fig. 3.22), NL $=2$ and $N_X = 5$ (Fig. 3.23) as well as NL $=3$ and $N_X = 15$ (Fig. 3.24). Examples of experiment ‘X chunk size’ were shown for the conditions NL $=1$ and $n_X = 0$ (Fig. 3.25), NL $=2$ and $n_X = 3$ (Fig. 3.26) as well as NL $=3$ and $n_X = 7$ (Fig. 3.27). Examples for experiment ‘Pause variability’ were shown for the conditions NL $=1$ and $\cup_\Delta = 0$ ms (Fig. 3.28), NL $=2$ and $\cup_\Delta = 300$ ms (Fig. 3.29) as well as NL $=3$ and $\cup_\Delta = 700$ ms (Fig. 3.30). An example for experiment ‘Emphasis’ was shown for the NL $=3$ condition (Fig. 3.31). For details on the set of grammatical stimuli, see Sec. 3.1.2 ‘Grammatical stimuli’ and for details on the grammar learning experiment conditions see Sec. 3.1.2 ‘Grammar learning experiments’ (Lehfeldt et al. n.d.). These figures illustrated how symbolic training sequences of AXB grammars were represented by the sequential and spike-based activation of stimulus element encoding input neurons. Further, they depicted the stimulus-driven spiking activity of the excitatory population which, in turn, triggered learning at the plastic EE synapses.

3.3.2 Synapse assembly types

For evaluation of the learning outcome, the strengths of stimulus element and transition encoding synapse assemblies in EE were analysed after training. An illustration of the different assembly types can be found in Fig. 3.1**E** (Lehfeldt et al. n.d.). Here, example snapshots of EE weight time constants and weights after training were shown, highlighting in colour the different assembly types. The stimulus element encoding synapse assemblies (i.e. A, X and B) were shown in Fig. 3.32. Grammatical A to B (i.e. $\lambda : A \rightarrow B$) and non-grammatical A to B (i.e. $\neg\lambda : A \rightarrow B$) and B to A (i.e. $B \rightarrow A$) transition assemblies were shown in Fig. 3.33. Non-grammatical A to A (i.e. $A \rightarrow A$), X to X (i.e. $X \rightarrow X$) and B to B (i.e. $B \rightarrow B$) transition assemblies were shown in Fig. 3.34. Finally, non-grammatical B to X (i.e. $B \rightarrow X$), X to A (i.e. $X \rightarrow A$) as well as non-assembly synapses (i.e. None) were shown in Fig. 3.35. The procedure to calculate the measure of successful learning from the assembly strength values was described in Sec. 3.1.2 ‘Separability measures’ (Lehfeldt et al. n.d.). The procedure for deriving separability measures of synaptic weights (w) applied equivalently to the synaptic weight time constants (τ_w).

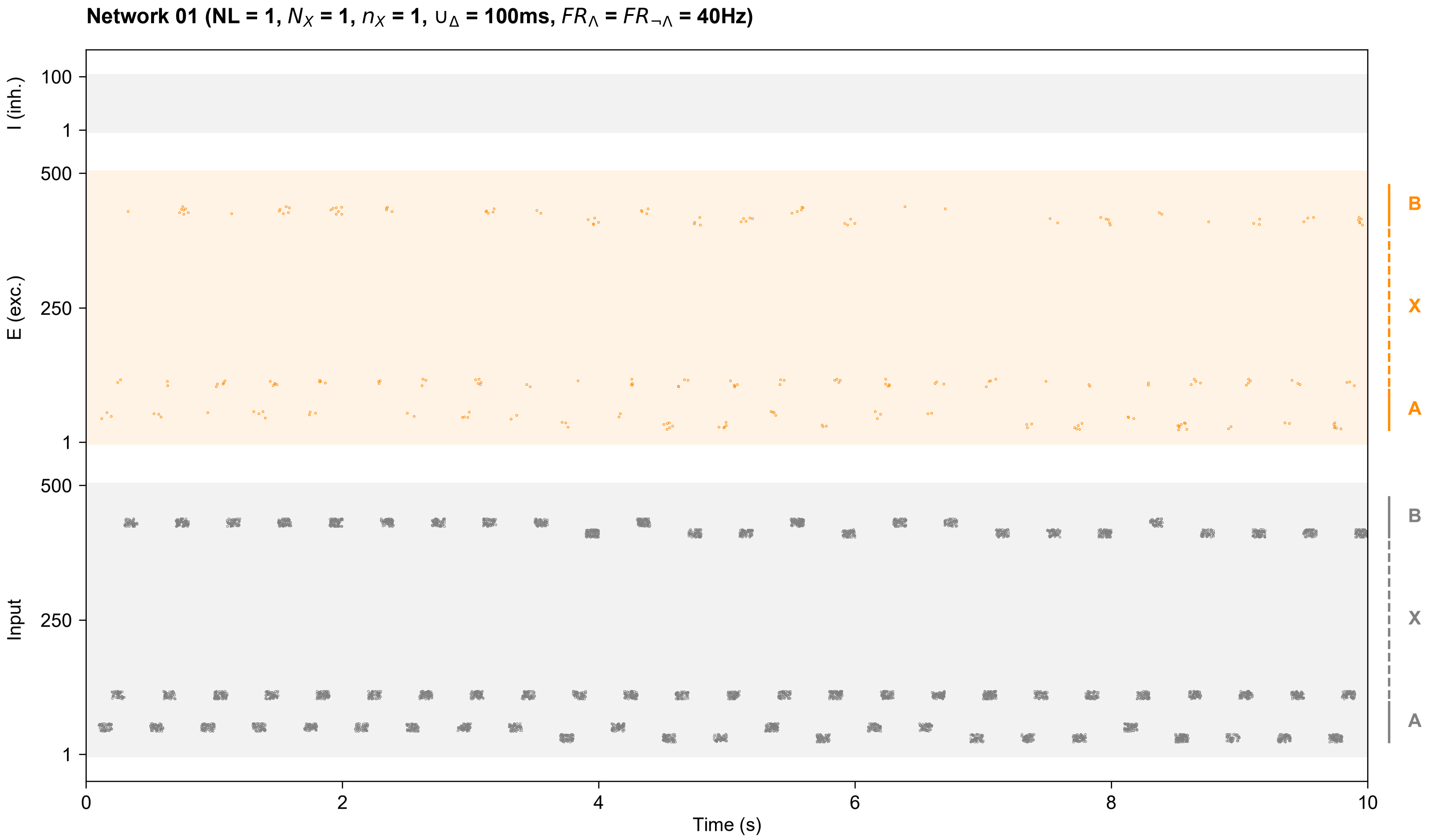


Figure 3.22: Example spiking network activity during training in experiment 'X variability' ($N_X = 1$). Zoom into the first 10 s of training; NL = nesting level; N_X = number of available X elements; n_X = number of X elements per grammar sample; $\cup_\Delta$ pause duration between grammar samples; FR = Poisson stimulus firing rate; Λ = grammatical neuron subgroups; $\neg\Lambda$ = non-grammatical neuron subgroups; Input = input population; E = excitatory population; I = inhibitory population; A = A element encoding subgroups; X = X element encoding subgroups; B = B element encoding subgroups.

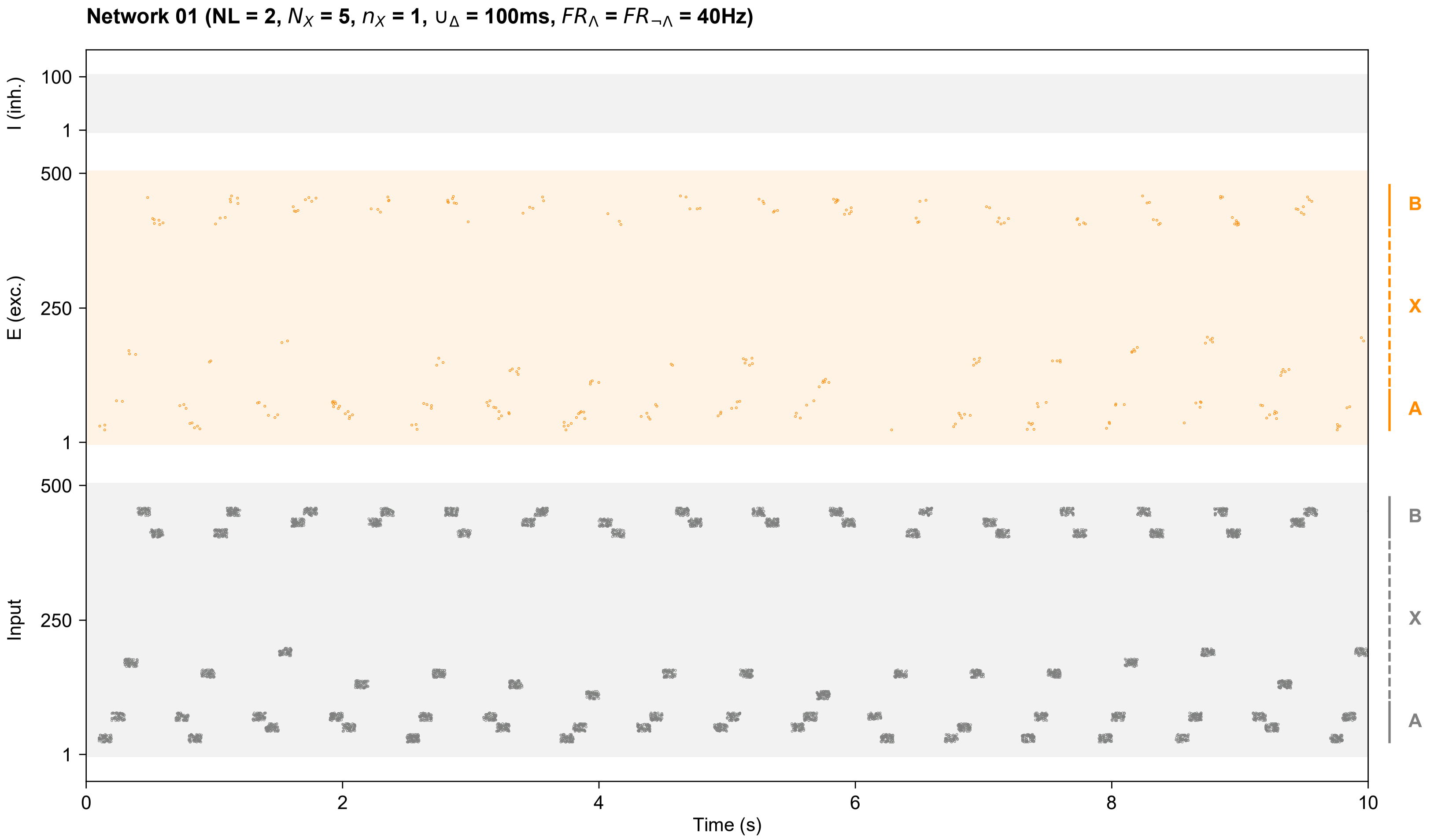


Figure 3.23: Example spiking network activity during training in experiment 'X variability' ($N_X = 5$). Zoom into the first 10 s of training; NL = nesting level; N_X = number of available X elements; n_X = number of X elements per grammar sample; $\cup_\Delta$ pause duration between grammar samples; FR = Poisson stimulus firing rate; Λ = grammatical neuron subgroups; $\neg\Lambda$ = non-grammatical neuron subgroups; Input = input population; E = excitatory population; I = inhibitory population; A = A element encoding subgroups; X = X element encoding subgroups; B = B element encoding subgroups.

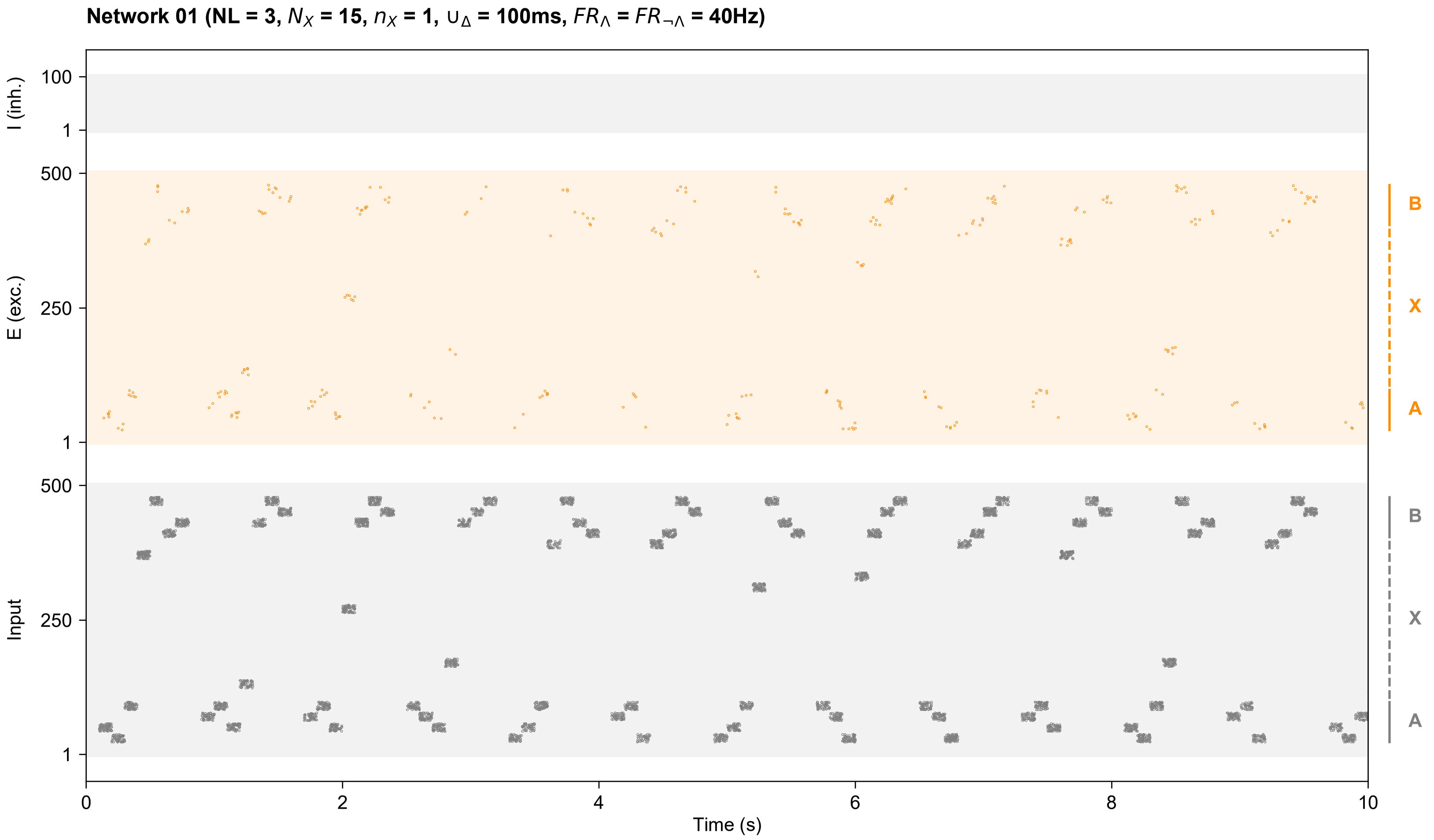


Figure 3.24: Example spiking network activity during training in experiment 'X variability' ($N_X = 15$). Zoom into the first 10 s of training; NL = nesting level; N_X = number of available X elements; n_X = number of X elements per grammar sample; $\cup_\Delta$ pause duration between grammar samples; FR = Poisson stimulus firing rate; Λ = grammatical neuron subgroups; $\neg\Lambda$ = non-grammatical neuron subgroups; Input = input population; E = excitatory population; I = inhibitory population; A = A element encoding subgroups; X = X element encoding subgroups; B = B element encoding subgroups.

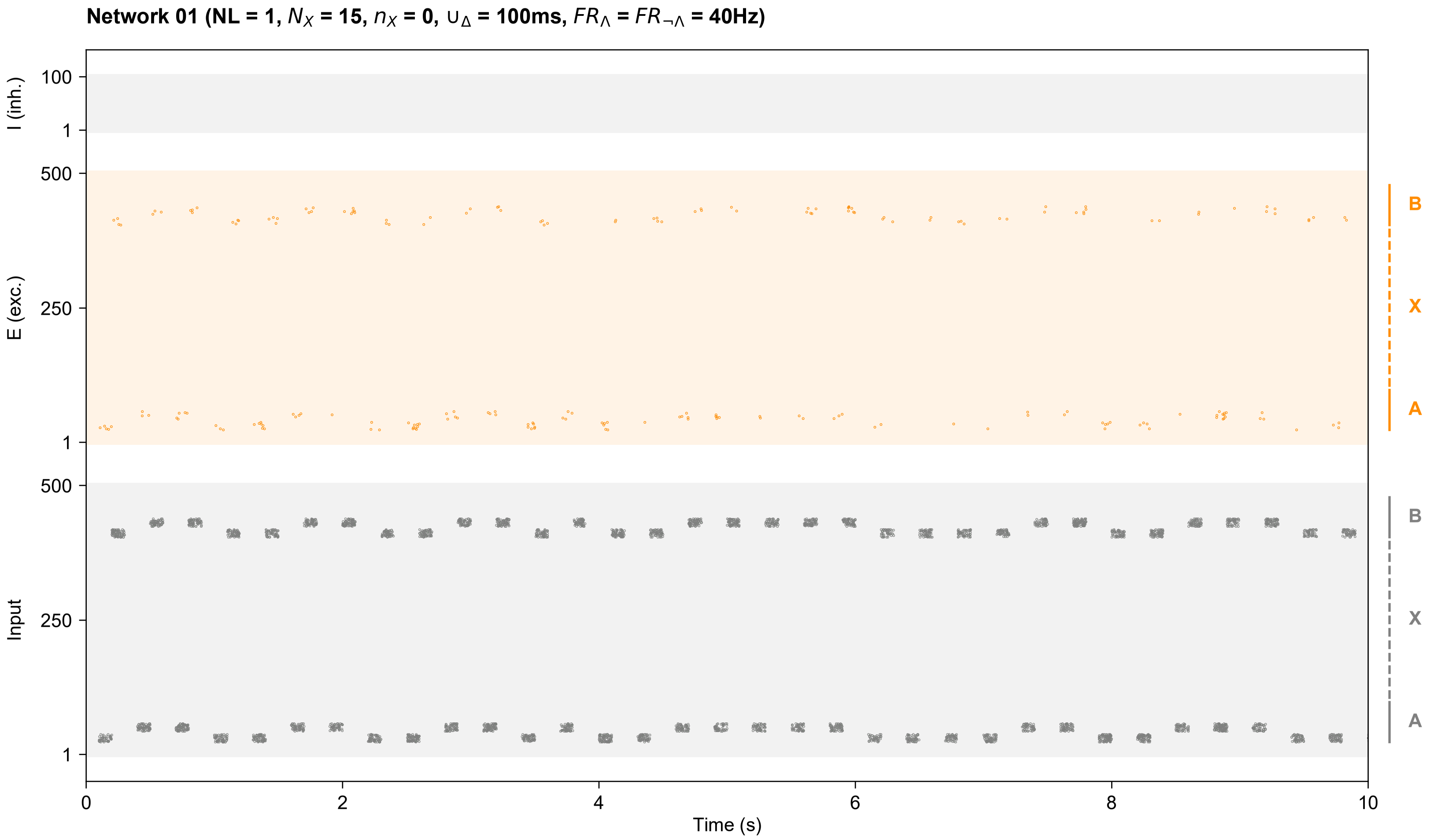


Figure 3.25: Example spiking network activity during training in experiment 'X chunk size' ($n_X = 0$). Zoom into the first 10 s of training; NL = nesting level; N_X = number of available X elements; n_X = number of X elements per grammar sample; $\cup_\Delta$ pause duration between grammar samples; FR = Poisson stimulus firing rate; Λ = grammatical neuron subgroups; $\neg\Lambda$ = non-grammatical neuron subgroups; Input = input population; E = excitatory population; I = inhibitory population; A = A element encoding subgroups; X = X element encoding subgroups; B = B element encoding subgroups.

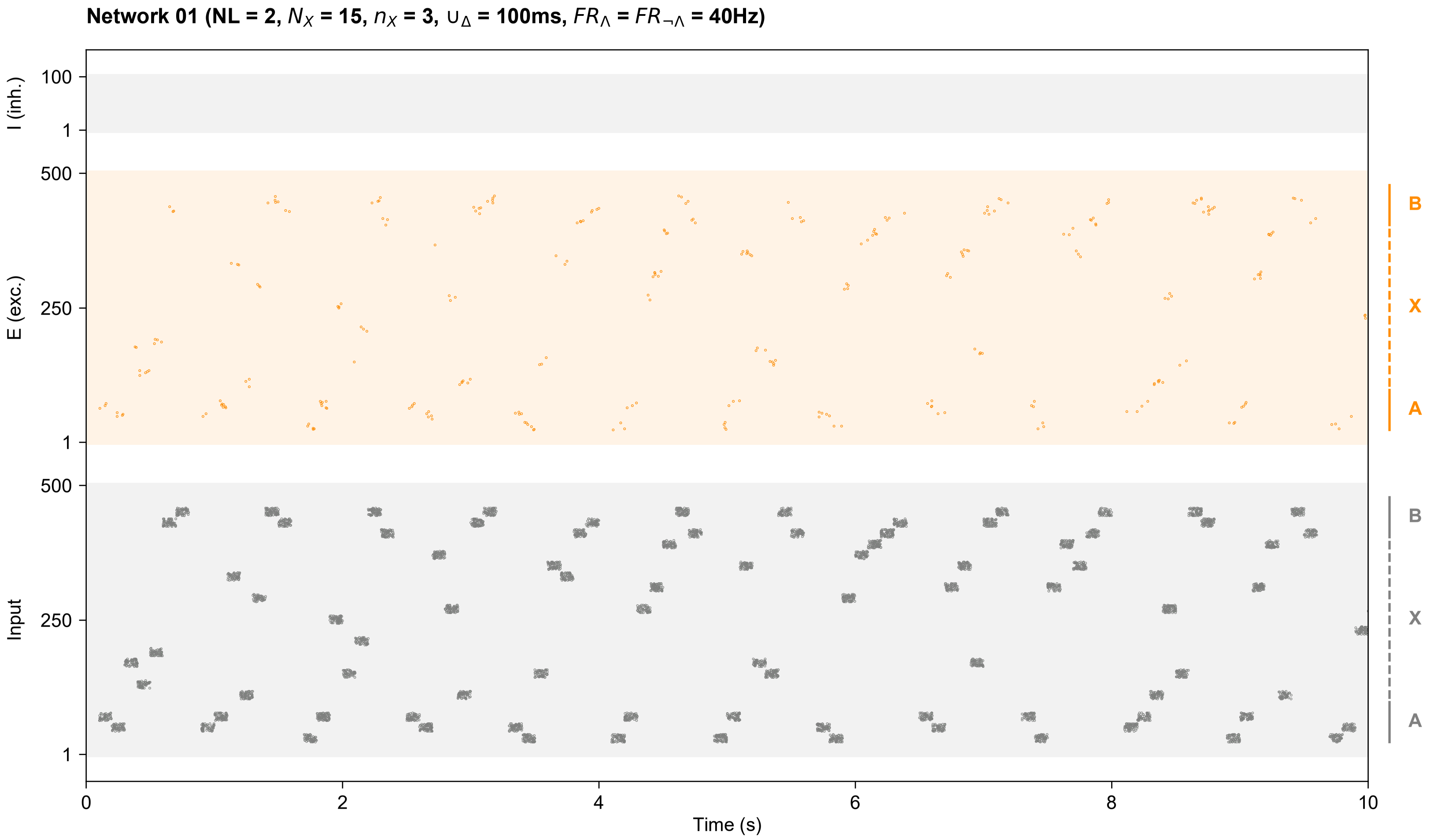


Figure 3.26: Example spiking network activity during training in experiment ‘X chunk size’ ($n_X = 3$). Zoom into the first 10 s of training; NL = nesting level; N_X = number of available X elements; n_X = number of X elements per grammar sample; $\cup_\Delta$ pause duration between grammar samples; FR = Poisson stimulus firing rate; Λ = grammatical neuron subgroups; $\neg\Lambda$ = non-grammatical neuron subgroups; Input = input population; E = excitatory population; I = inhibitory population; A = A element encoding subgroups; X = X element encoding subgroups; B = B element encoding subgroups.

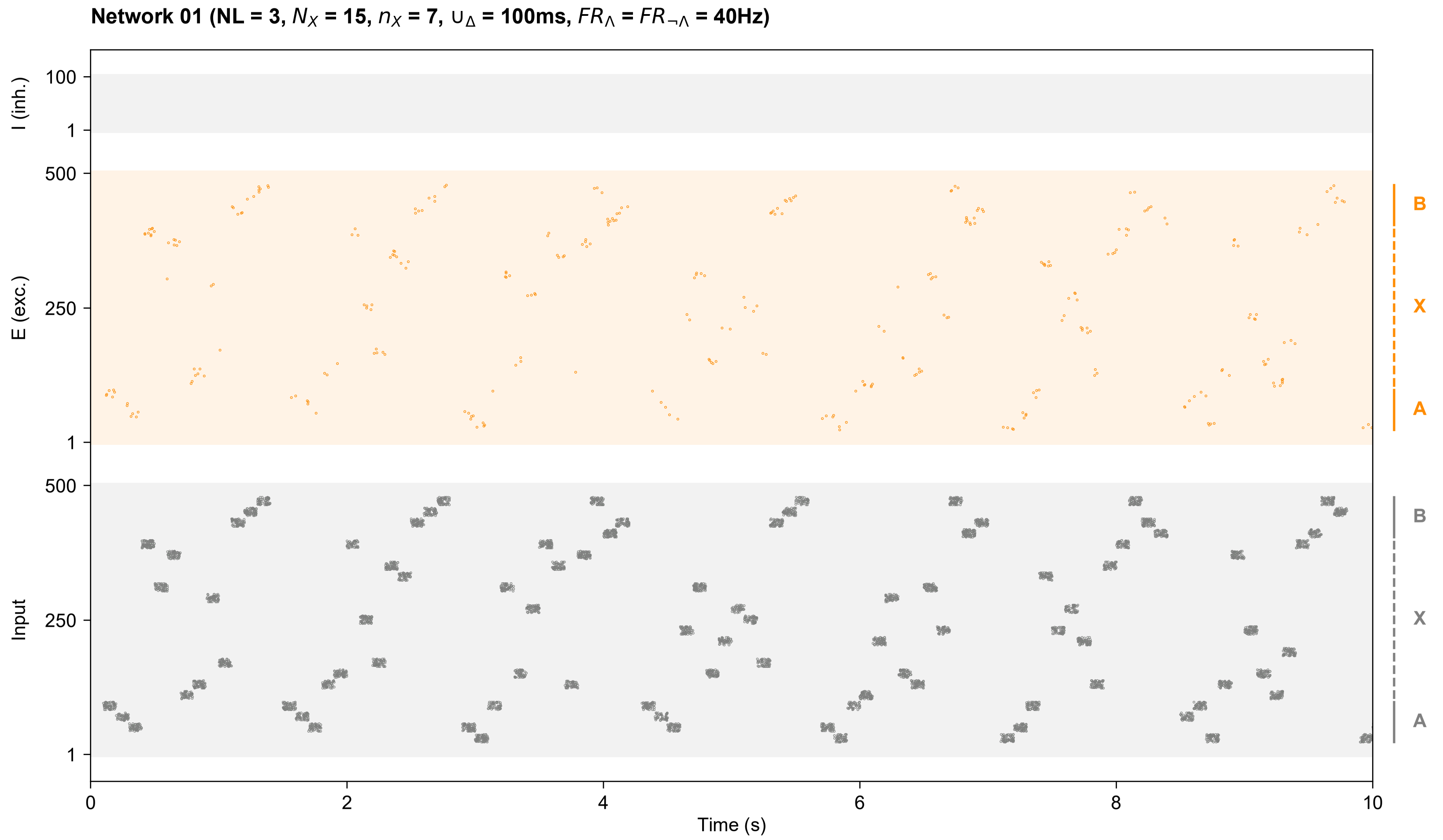


Figure 3.27: Example spiking network activity during training in experiment 'X chunk size' (n_X = 7). Zoom into the first 10 s of training; NL = nesting level; N_X = number of available X elements; n_X = number of X elements per grammar sample; $\cup_\Delta$ pause duration between grammar samples; FR = Poisson stimulus firing rate; Λ = grammatical neuron subgroups; $\neg\Lambda$ = non-grammatical neuron subgroups; Input = input population; E = excitatory population; I = inhibitory population; A = A element encoding subgroups; X = X element encoding subgroups; B = B element encoding subgroups.

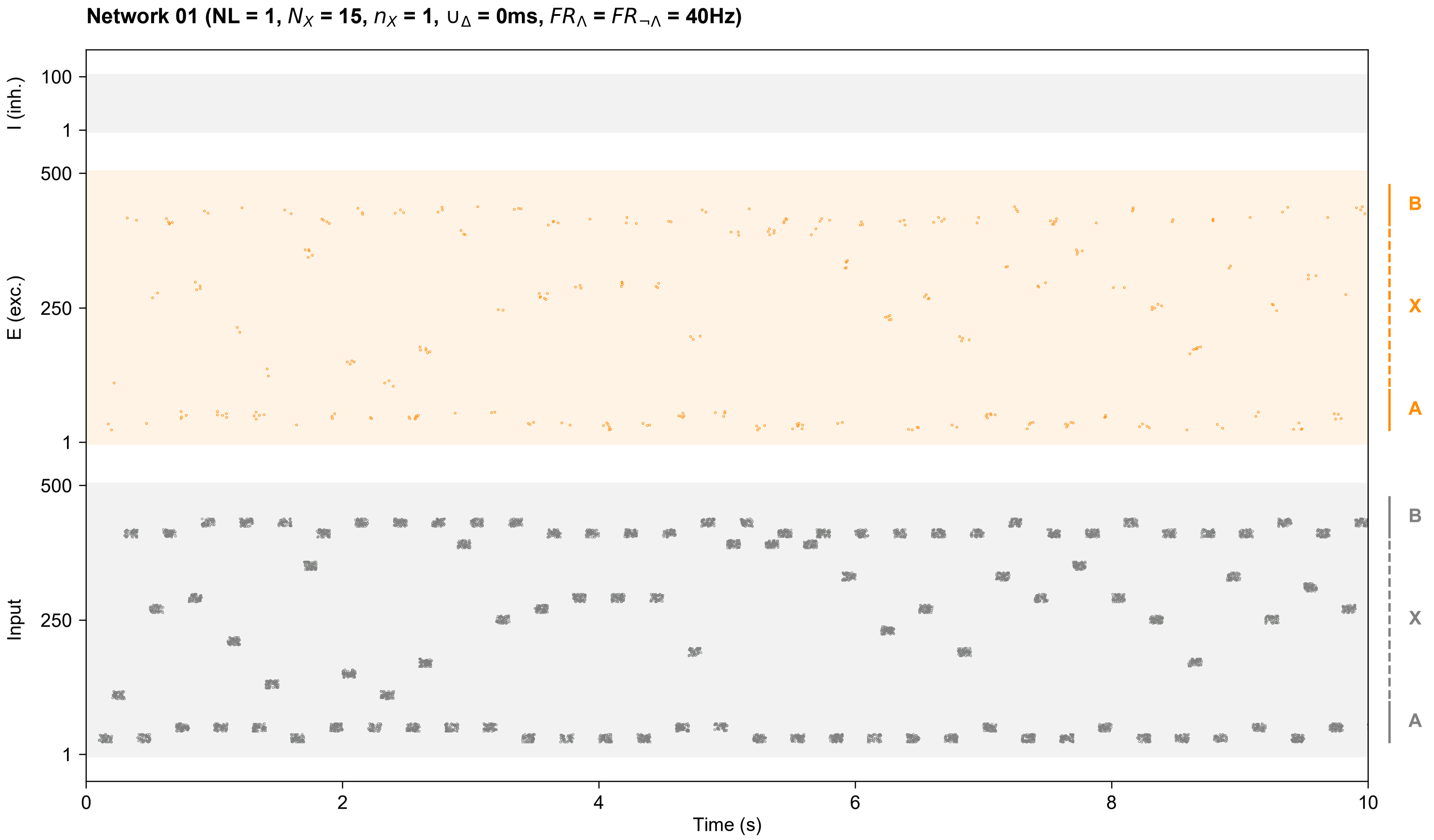


Figure 3.28: Example spiking network activity during training in experiment 'Pause variability' ($\cup_\Delta = 0$ ms). Zoom into the first 10 s of training; NL = nesting level; N_X = number of available X elements; n_X = number of X elements per grammar sample; $\cup_\Delta$ pause duration between grammar samples; FR = Poisson stimulus firing rate; Λ = grammatical neuron subgroups; $\neg\Lambda$ = non-grammatical neuron subgroups; Input = input population; E = excitatory population; I = inhibitory population; A = A element encoding subgroups; X = X element encoding subgroups; B = B element encoding subgroups.

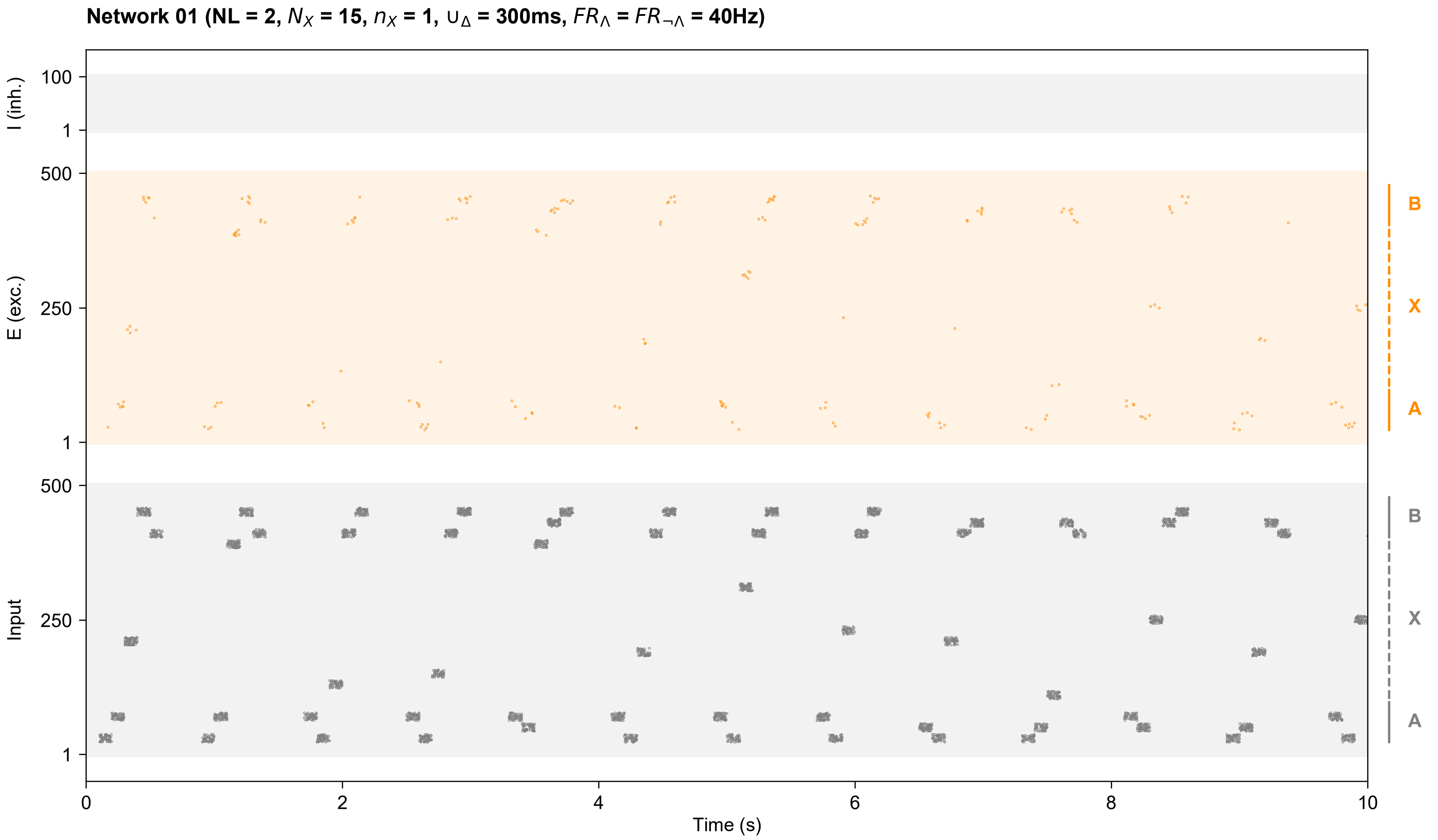


Figure 3.29: Example spiking network activity during training in experiment 'Pause variability' ($\cup_\Delta = 300$ ms). Zoom into the first 10 s of training; NL = nesting level; N_X = number of available X elements; n_X = number of X elements per grammar sample; $\cup_\Delta$ pause duration between grammar samples; FR = Poisson stimulus firing rate; Λ = grammatical neuron subgroups; $\neg\Lambda$ = non-grammatical neuron subgroups; Input = input population; E = excitatory population; I = inhibitory population; A = A element encoding subgroups; X = X element encoding subgroups; B = B element encoding subgroups.

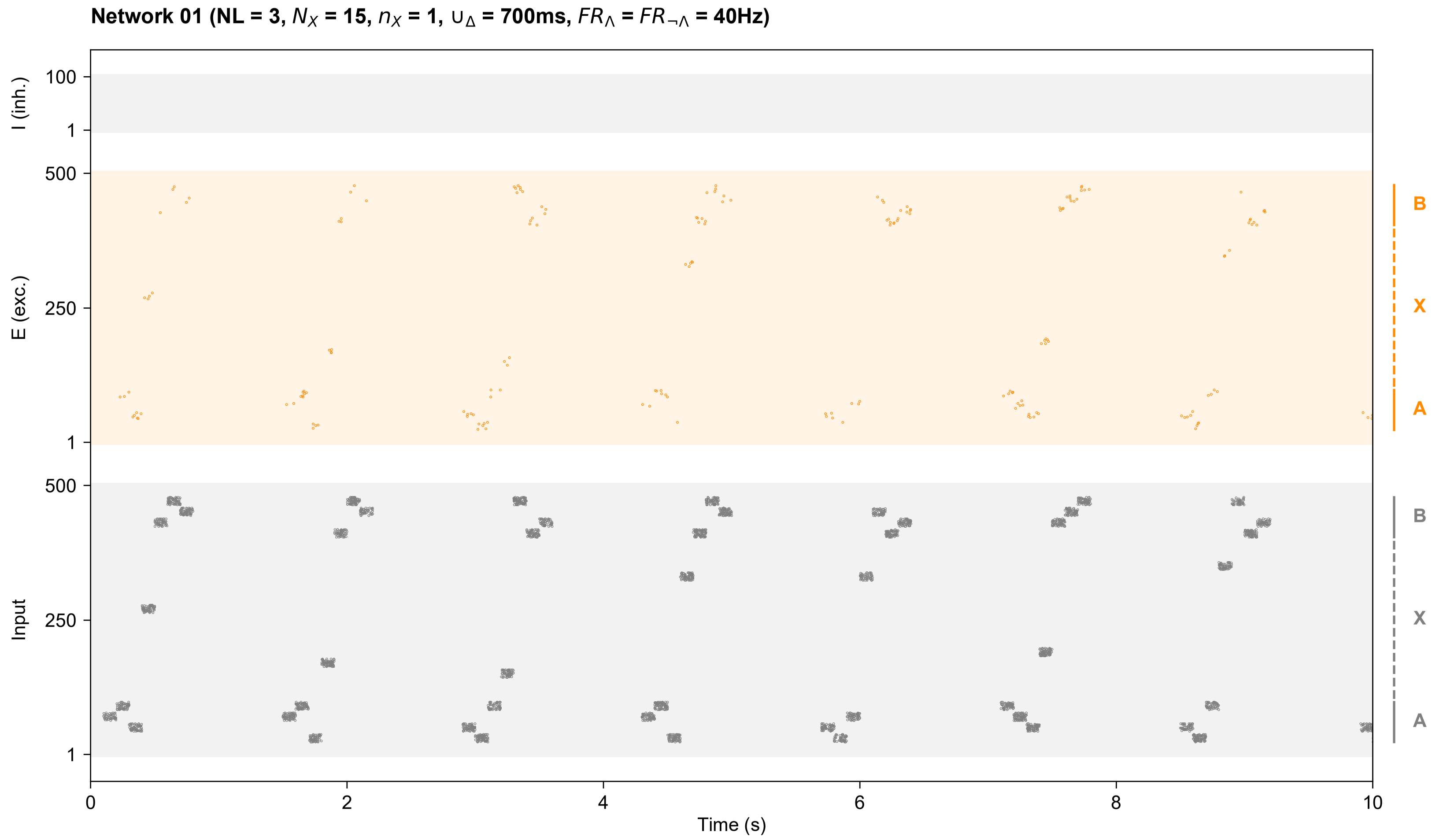


Figure 3.30: Example spiking network activity during training in experiment 'Pause variability' ($\cup_\Delta = 700$ ms). Zoom into the first 10 s of training; NL = nesting level; N_X = number of available X elements; n_X = number of X elements per grammar sample; $\cup_\Delta$ pause duration between grammar samples; FR = Poisson stimulus firing rate; Λ = grammatical neuron subgroups; $\neg\Lambda$ = non-grammatical neuron subgroups; Input = input population; E = excitatory population; I = inhibitory population; A = A element encoding subgroups; X = X element encoding subgroups; B = B element encoding subgroups.

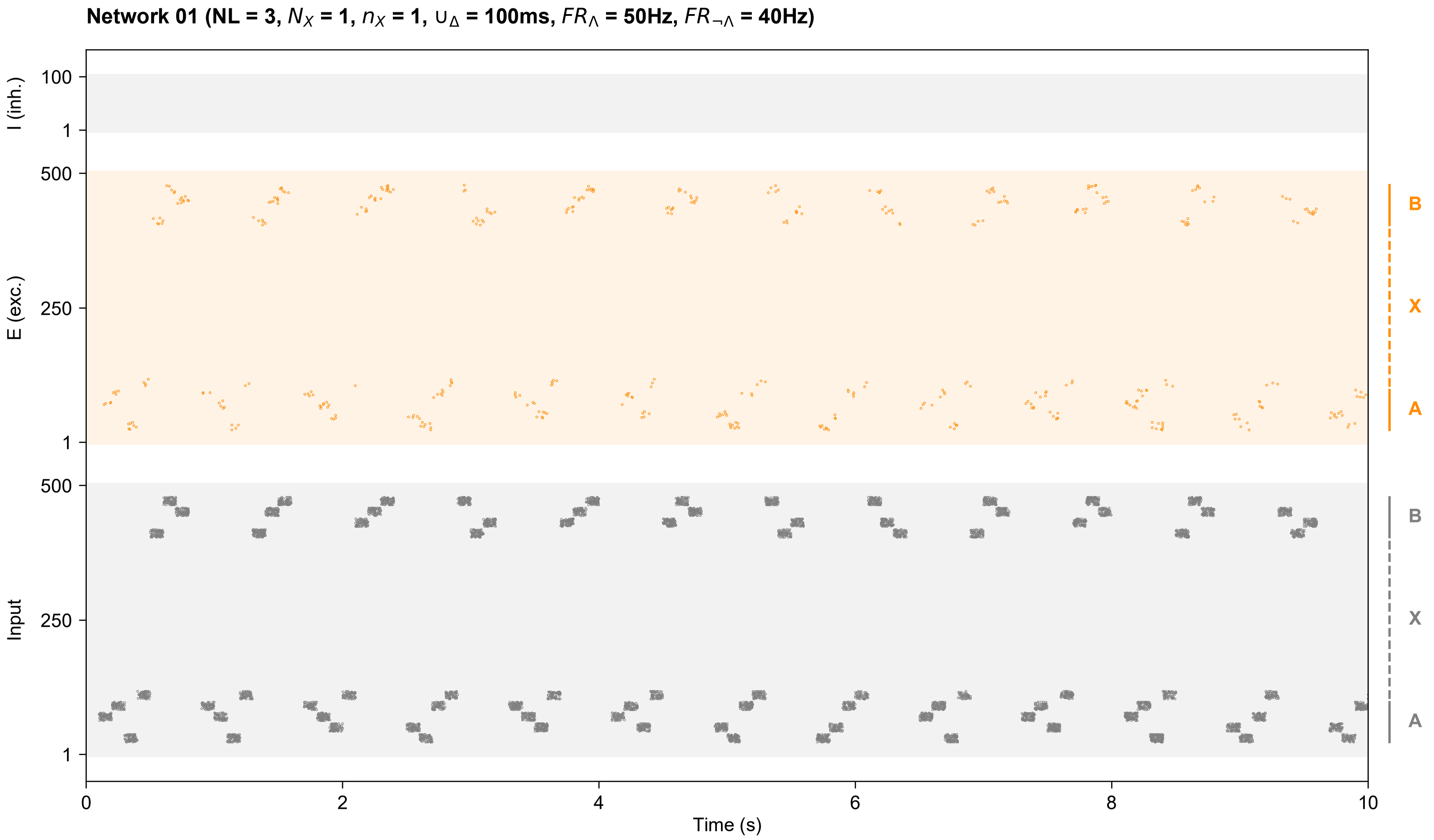


Figure 3.31: Example spiking network activity during training in experiment ‘Emphasis’. Zoom into the first 10 s of training; NL = nesting level; N_X = number of available X elements; n_X = number of X elements per grammar sample; $\cup_\Delta$ pause duration between grammar samples; FR = Poisson stimulus firing rate; Λ = grammatical neuron subgroups; $\neg\Lambda$ = non-grammatical neuron subgroups; Input = input population; E = excitatory population; I = inhibitory population; A = A element encoding subgroups; X = X element encoding subgroups; B = B element encoding subgroups.

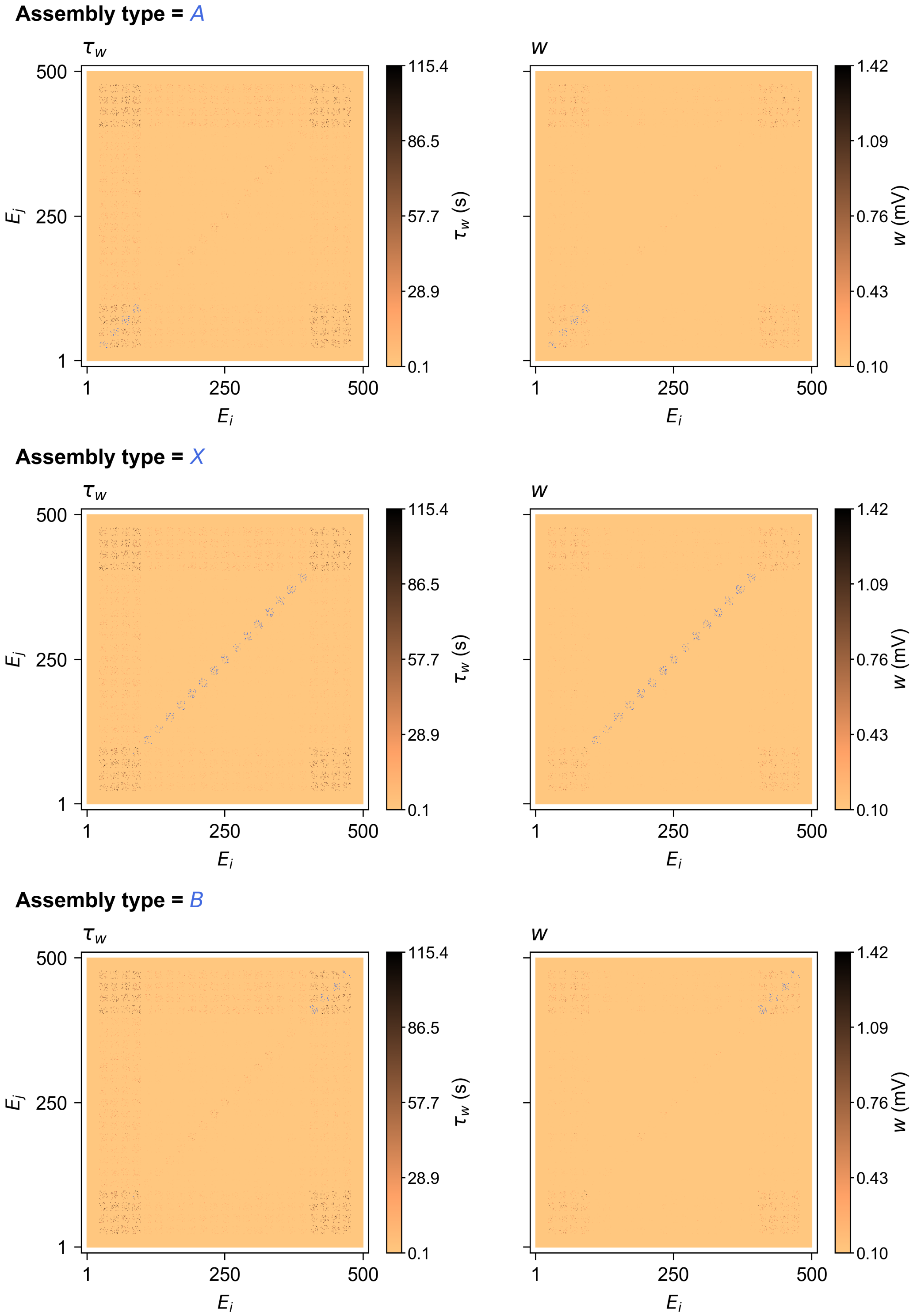


Figure 3.32: Element encoding synapse assemblies A, X and B. Example synaptic EE snapshot (E_i = source/pre, E_j = target/post) after training (NL = 3, experiment 'X variability', $N_X = 15$) with highlighted assembly types by colour (blue); assembly type indicated above each row; left column: weight time constant (τ_w); right column: weight (w).

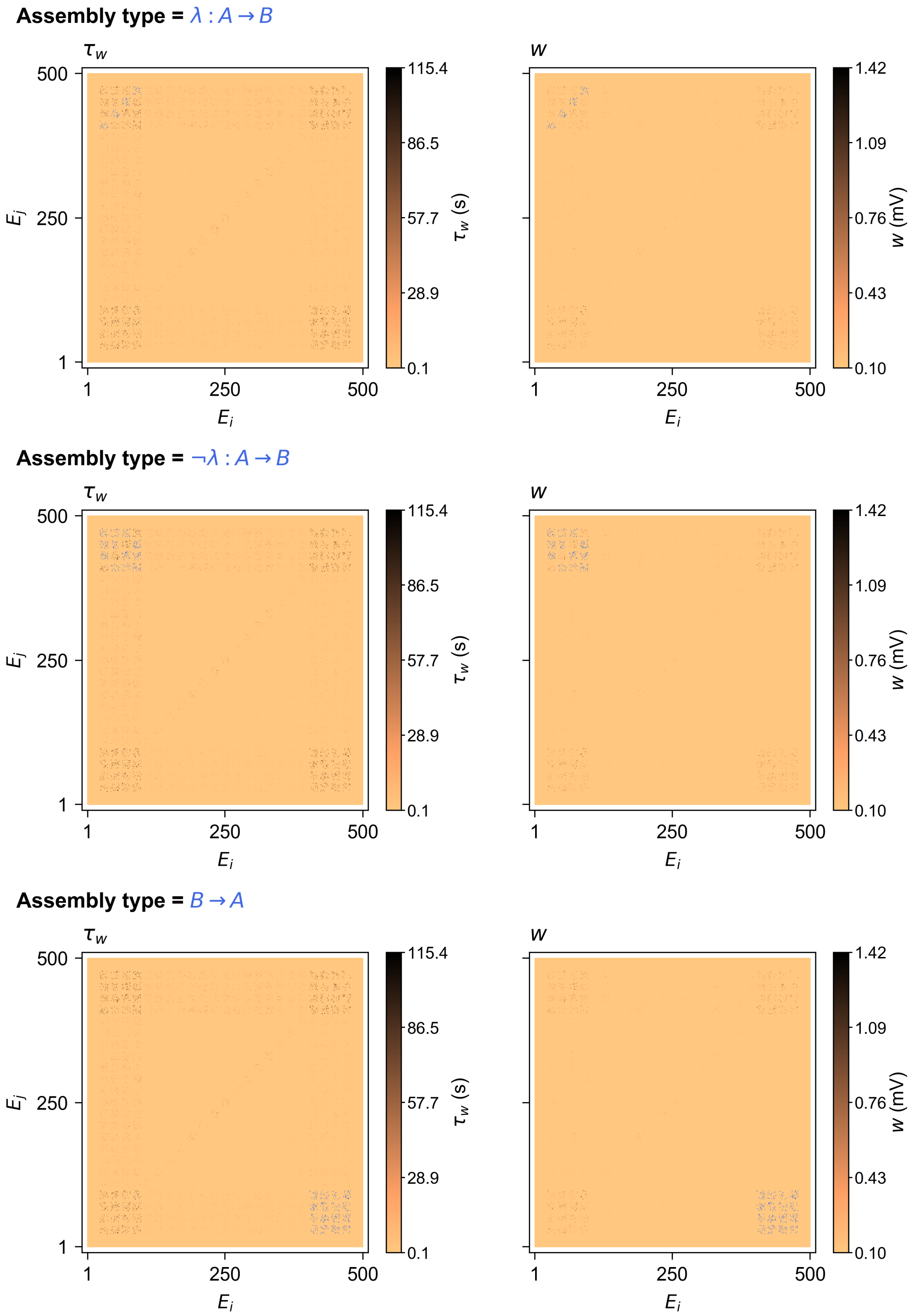


Figure 3.33: Transition encoding synapse assemblies $\lambda : A \rightarrow B$, $\neg\lambda : A \rightarrow B$ and $B \rightarrow A$. Example synaptic EE snapshot (E_i = source/pre, E_j = target/post) after training (NL = 3, experiment 'X variability', $N_X = 15$) with highlighted assembly types by colour (blue); assembly type indicated above each row; left column: weight time constant (τ_w); right column: weight (w).

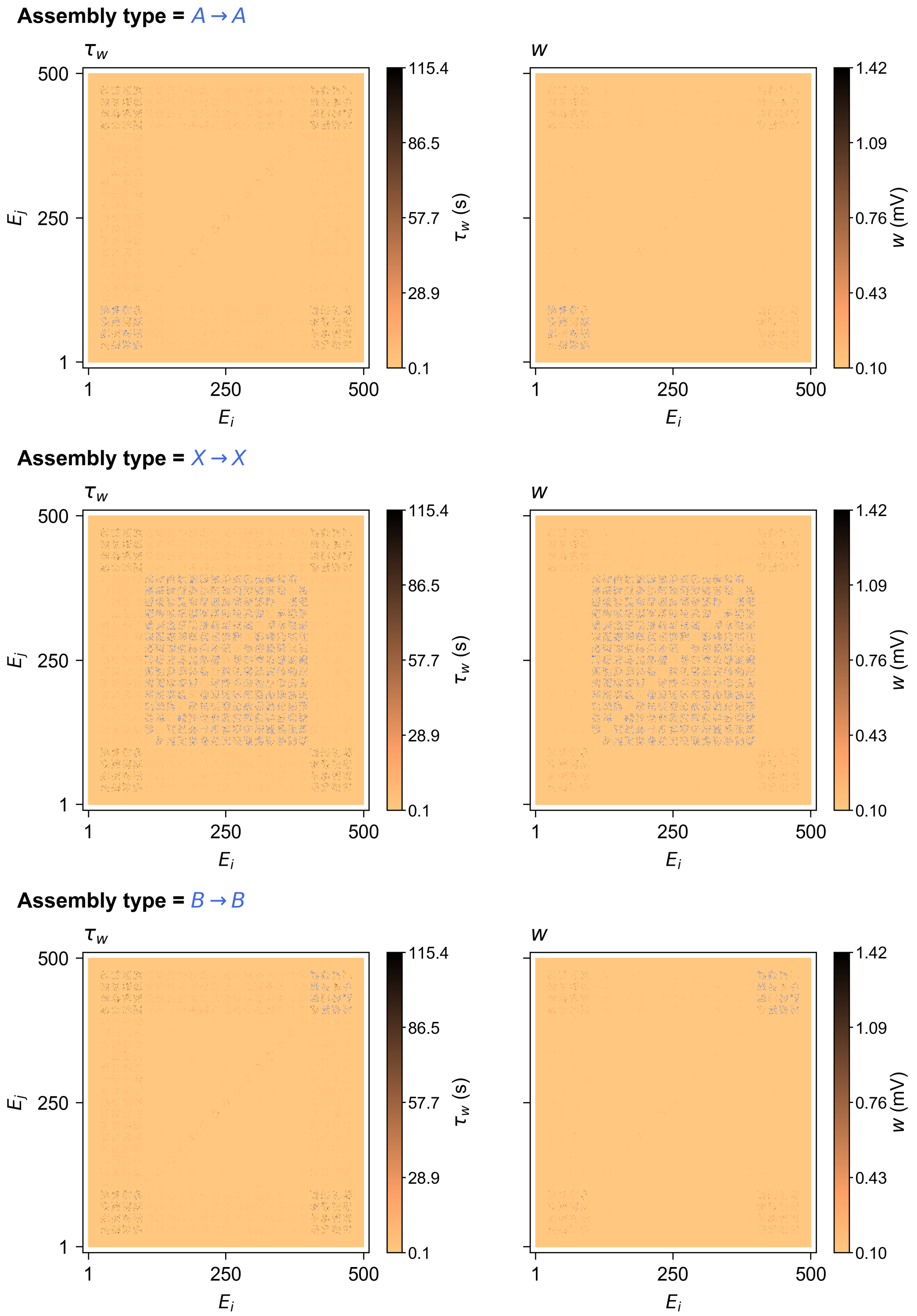


Figure 3.34: Transition encoding synapse assemblies $A \rightarrow A$, $X \rightarrow X$ and $B \rightarrow B$. Example synaptic EE snapshot (E_i = source/pre, E_j = target/post) after training (NL = 3, experiment 'X variability', $N_X = 15$) with highlighted assembly types by colour (blue); assembly type indicated above each row; left column: weight time constant (τ_w); right column: weight (w).

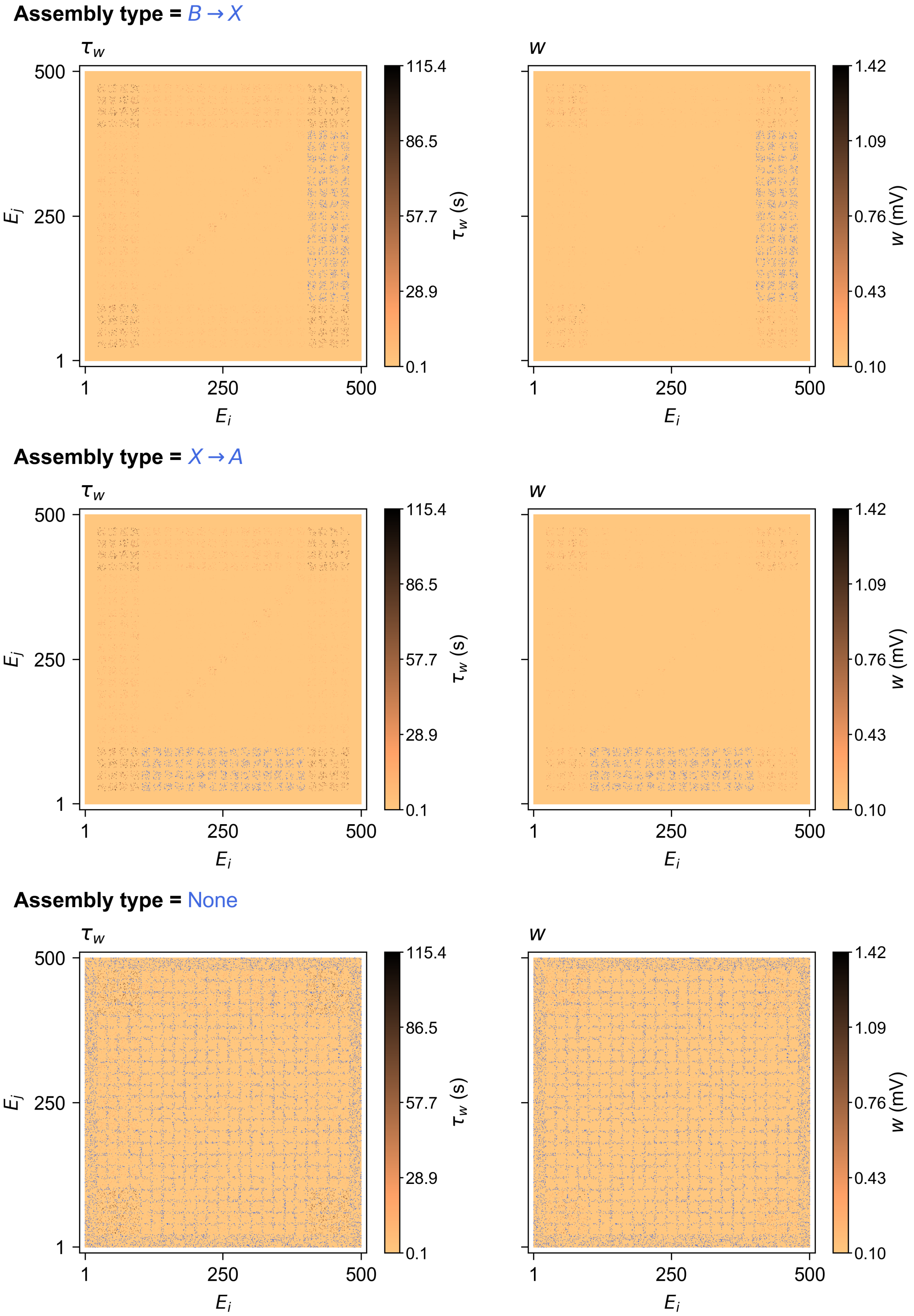


Figure 3.35: Transition encoding synapse assemblies $B \to X$, $X \to A$ and *None*. Example synaptic EE snapshot (E_i = source/pre, E_j = target/post) after training (NL = 3, experiment 'X variability', $N_X = 15$) with highlighted assembly types by colour (blue); assembly type indicated above each row; left column: weight time constant (τ_w); right column: weight (w).

3.3.3 Assembly dynamics

3.3.3.1 Complete collection of assembly dynamics

The full collection of synaptic weight (w) dynamics across experiments and RNNs was shown in Fig. 3.5 to Fig. 3.8 (Lehfeldt et al. n.d.). Here, additionally the corresponding and complete collection of assembly weight time constant (τ_w) dynamics were presented (Fig. 3.36, Fig. 3.37, Fig. 3.38 and Fig. 3.39). The synaptic state variable τ_w followed the same spike-triggered update routines as w, though with a different parameter space. For details, see Sec. 3.1.2 'Spike-timing dependent plasticity (STDP) with short- to long-term memory transition (SLMT)' (Lehfeldt et al. n.d.). Thus, qualitatively similar τ_w dynamics across the experiment conditions occurred when compared to w. However, as τ_w was temporally infinite, i.e. had no decay itself, it occasionally showed different dynamics for stimulus conditions that modified temporal gaps between stimulus element occurrences and transitions. In the following, τ_w dynamics were thus shortly described and explained in comparison to w, highlighting shared and different dynamics between them. Again, assembly dynamics across grammars and experimental conditions were fully explained by the sensitivity of STDP and SLMT to stimulus element order, temporal proximity, occurrence frequency and strength (Lehfeldt et al. n.d.). For a detailed description and explanation of w dynamics, see Sec. 3.1.2 'Results' and Sec. 3.1.2 'Discussion' (Lehfeldt et al. n.d.). Taken together, the full collection of w and τ_w dynamics provided a complete view onto the modelled learning dynamics. This further highlighted the benefit of the model, namely the possibility to study the full bandwidth of stimulus element and transition dynamics.

3.3.3.2 Experiment 'X variability': assembly dynamics

In experiment 'X variability', both τ_w and w showed qualitatively equal dynamics across the N_X conditions (τ_w: Fig. 3.36 and w: Fig. 3.5). While those assemblies that were independent from the X variability remained constant, dependent assemblies (i.e. the sets $\chi : \{A \rightarrow X,\, X,\, X \rightarrow B\}$ and $\chi' : \{X \rightarrow X,\, B \rightarrow X,\, X \rightarrow A\}$) decreased over increasing X variabilities. The decrease of χ and χ' assembly weights was explained by decreased occurrence frequencies of individual X elements under high X variabilities (Lehfeldt et al. n.d.). This effect equally applied to τ_w (τ_w: Fig. 3.36**E**,**F** and w: Fig. 3.5**E**,**F**). However, different dynamics in w and τ_w occurred for the remaining assembly types and across grammar complexities, i.e. nesting levels (NLs). Across NLs, w of $\lambda : A \rightarrow B$, A, B and $B \rightarrow A$ assemblies remained largely at the same strength levels (Fig. 3.5**A**,**B**,**G**). In $\neg\lambda : A \rightarrow B$ and the set $\widehat{AA}\widehat{BB} : \{A \rightarrow A,\, B \rightarrow B\}$, w showed an increase between isolated and nested NADs (Fig. 3.5**C**,**D**). However, τ_w had consistently increasing strength levels across NLs and in all theses assembly types (Fig. 3.36**A**-**D**, **G**). These increases were explained by consistently increasing A and B element occurrence frequencies across NLs (Fig. 3.4**A**). While the temporally infinite τ_w represented different occurrence frequencies in the training sequences accordingly, the decay of w compensated this frequency effect. As a result, w assembly strengths remained largely constant across NLs or only increased between isolated and nested NADs.

3.3.3.3 Experiment 'X chunk size': assembly dynamics

In experiment 'X chunk size', τ_w and w showed qualitatively equal dynamics for $\lambda : A \rightarrow B$, $\neg\lambda : A \rightarrow B$, χ and χ' (τ_w: Fig. 3.37 and w: Fig. 3.6). In both, $\lambda : A \rightarrow B$ and $\neg\lambda : A \rightarrow B$ decreased over the n_X conditions, while χ and χ' increased (τ_w: Fig. 3.37**A**,**C**,**E**,**F** and w: Fig. 3.6**A**,**C**,**E**,**F**). These effects in w were explained by increased within-sample temporal gaps of A to B element transitions and increased occurrence frequencies of X elements (Lehfeldt et al. n.d.). These effects equally applied to τ_w. However, different dynamics in w and τ_w occurred for the remaining assemblies. While A and B in w decreased over the n_X conditions, they remained constant or slightly increased in τ_w (τ_w: Fig. 3.37**B** and w: Fig. 3.6**B**). Here, increased temporal gaps between A and B element occurrences resulted in prolonged decay periods of w, while the temporally infinite τ_w was not affected. The $\widehat{AABB}$ assemblies decreased consistently in w while in τ_w, they decreased the strongest for NL = 1 grammars, decreased comparably weaker for NL = 2 grammars and remained constant for NL = 3 grammars (τ_w: Fig. 3.37**D** and w: Fig. 3.6**D**). The consistent decrease of w was explained by prolonged decay periods across n_X conditions. However, the dynamics of temporally infinite τ_w were explained by the structural composition of grammars. In isolated NADs, $A \rightarrow A$ and $B \rightarrow B$ transitions could only grow due to cross-sample interference (Lehfeldt et al. n.d.). Thus, their growth was negatively affected by increased temporal gaps induced by large n_X. This effect, however, decreased with increasing grammar complexities given that in nested NADs, $A \rightarrow A$ and $B \rightarrow B$ transitions already existed within individual grammar samples. Finally, $B \rightarrow A$ assemblies in w decreased for isolated NADs but remained largely constant for nested NADs while in τ_w, they consistently increased across n_X conditions (τ_w: Fig. 3.37**G** and w: Fig. 3.6**G**). These dynamics in τ_w were explained by a decreased amount of cross-sample depression in $B \rightarrow A$. In general, during the experience of an A to B transition, potentiation occurred in $\lambda : A \rightarrow B$ while depression occurred in $B \rightarrow A$ (Lehfeldt et al. n.d.). Thus, for large n_X, the amount of depression was reduced by longer decay periods of A element depression traces in $B \rightarrow A$, resulting in a stronger $B \rightarrow A$ growth. In w, however, the weight decay again compensated for this effect in nested NADs. Further, the decrease of $B \rightarrow A$ in isolated NADs was explained by prolonged decay periods of w between potentiation events (i.e. cross-sample B to A transitions), that were induced by longer grammar samples. As isolated NADs consisted of only one grammatical A and B pair, this effect was the strongest here.

3.3.3.4 Experiment 'Pause variability': assembly dynamics

In experiment 'Pause variability', τ_w and w showed qualitatively equal dynamics for $\lambda : A \rightarrow B$, χ and $B \rightarrow A$ (τ_w: Fig. 3.38 and w: Fig. 3.7). In both, $\lambda : A \rightarrow B$ increased over increasing pause durations while this effect was stronger in τ_w (τ_w: Fig. 3.38**A** and w: Fig. 3.7**A**). The consistent increase was explained by decreased amounts of cross-sample depression (Lehfeldt et al. n.d.). As τ_w had no decay, this effect was more pronounced here. In χ, τ_w and w were constant across pause durations, whereby assemblies in w remained at the baseline level (τ_w: Fig. 3.38**E** and w: Fig. 3.7**E**). In τ_w, the consistency of within-sample χ was explained by the temporal infinity of τ_w and hence, by its independence from between-sample pause durations. In w, χ assemblies remained at the baseline level as the 'Pause variability' experiment was conducted in the $N_X = 15$ variability condition. Cross-sample $B \rightarrow A$ in w and τ_w decreased consistently across pause durations (τ_w: Fig. 3.38**G** and w: Fig. 3.7**G**). This effect was explained by prolonged decay periods of potentiation traces induced by long pauses and hence, reduced amounts of cross-sample potentiation. However, different dynamics in w and τ_w occurred for the remaining assembly types. In A and B, strengths remained largely constant across pause durations in τ_w while in w, they consistently decreased (τ_w: Fig. 3.38**B** and w: Fig. 3.7**B**). Further, A and B reached higher strength levels across grammar complexities in τ_w, which was explained by increased occurrence frequencies of A and B elements in nested NADs (Fig. 3.4**C**). The consistency of A and B in τ_w was again explained by its temporal infinity while in w, the decrease was explained by prolonged decay periods. In $\neg\lambda : A \rightarrow B$, both w and τ_w decreased in isolated NADs while in nested NADs, w remained largely constant and τ_w increased (τ_w: Fig. 3.38**C** and w: Fig. 3.7**C**). The decrease of $\neg\lambda : A \rightarrow B$ in isolated NADs was explained by decreased amounts of cross-sample potentiation due to long stimulus pauses and hence, long decay periods of potentiation traces. In τ_w, the consistent increase across pause durations was explained by decreased amounts of cross-sample depression whereby in w, this effect was counteracted by the weight decay. In $\widehat{AA}\widehat{BB}$, w consistently decreased over pause durations while τ_w decreased the most for isolated NADs and subsequently less for nested NADs (τ_w: Fig. 3.38**D** and w: Fig. 3.7**D**). The decrease in w was again explained by prolonged decay periods due to long pauses. In τ_w, dynamics were explained by the structural composition of grammars, whereby in isolated NADs the effect of reduced amounts of cross-sample potentiation was the strongest. In χ', assembly strengths slightly decreased or remained at baseline level in τ_w, while they fully remained at baseline level in w (τ_w: Fig. 3.38**F** and w: Fig. 3.7**F**). χ' assemblies in w consistently remained at baseline given that 'Pause variability' experiments were conducted in the $N_X = 15$ variability condition. In τ_w, also $X \rightarrow X$ was at baseline while $B \rightarrow X$ and $X \rightarrow A$ slightly decreased across pause durations. This was again explained by reduced amounts of cross-sample potentiation due to long pauses.

3.3.3.5 Experiment 'Emphasis': assembly dynamics

In experiment 'Emphasis', τ_w and w showed qualitatively equal dynamics in all synapse assembly types (τ_w: Fig. 3.39 and w: Fig. 3.8). In the 'On' emphasis condition, assembly strengths were consistently stronger than in the 'Off' emphasis condition. These dynamics were explained by increased stimulus firing rates in the 'On' condition and thus, by the increased amount of induced potentiation (Lehfeldt et al. n.d.). Only X largely remained constant across emphasis conditions (τ_w: Fig. 3.39**E** and w: Fig. 3.7**E**). This consistency was explained by the fact that X was independent from emphasis, given that only the grammatical A and B element encoding neural subgroups received increased stimulus firing rates (Lehfeldt et al. n.d.).

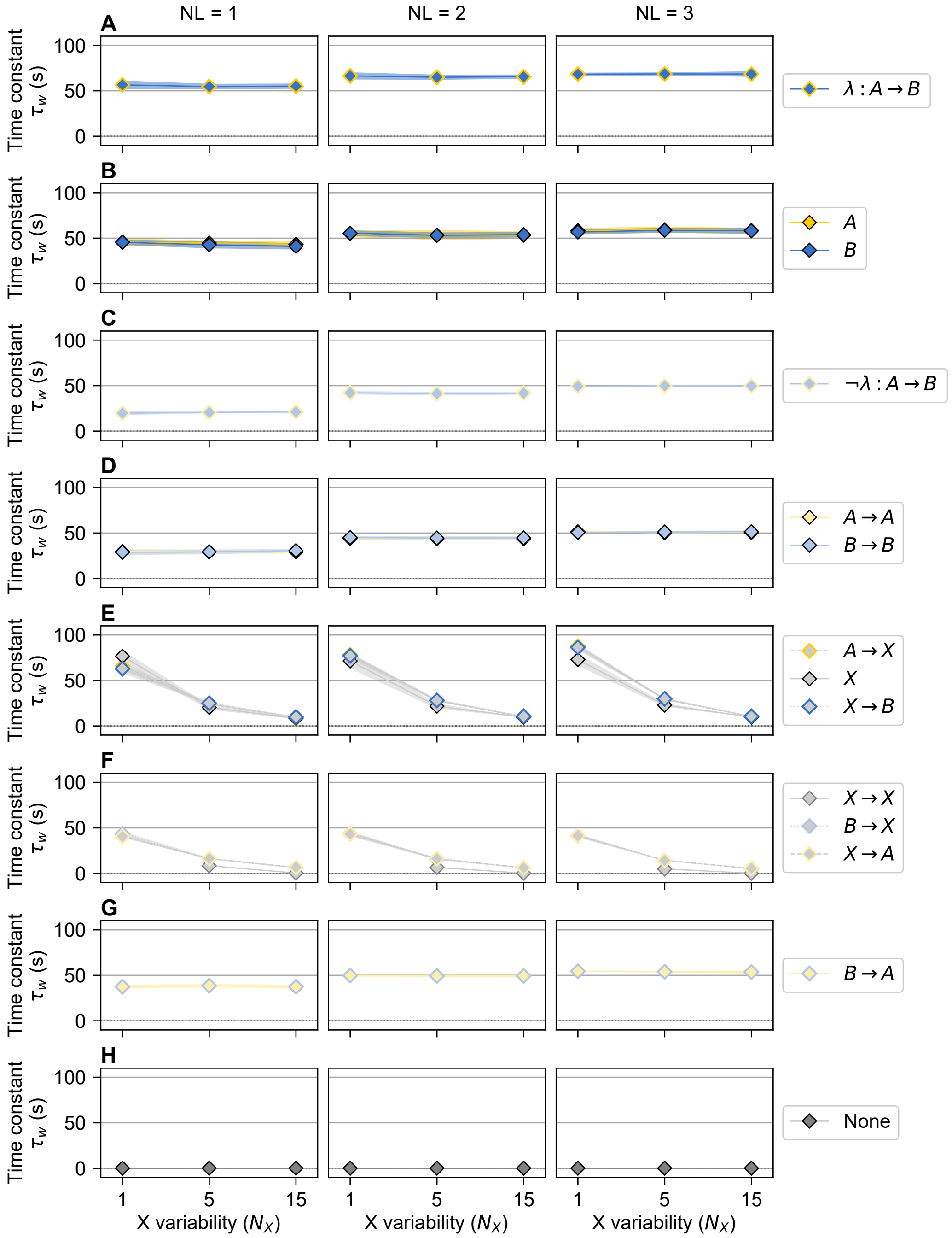


Figure 3.36: Assembly medians (synaptic weight time constant τ_w) in experiment 'X variability'. Each individual data point (= n) indicates mean ± standard deviation (SD) across the RNNs (N = 10). Columns indicate grammar (NL = 1: isolated NAD, NL = 2 and NL = 3: nested NADs). N_X = number of available X elements. Dotted line = lower weight boundary. **A** Grammatical A to B transitions (n = 10). **B** A and B elements (n = 10). **C** Non-grammatical A to B transitions (n = 10). **D** A to A and B to B transitions ($\widehat{AA}\widehat{BB}$, n = 10). **E** X elements, A to X and X to B transitions (χ, n = 10). **F** X to X, B to X and X to A transitions (χ', n = 10). **G** B to A transitions (n = 10). **H** Non-assembly synapses (None, n = 10).

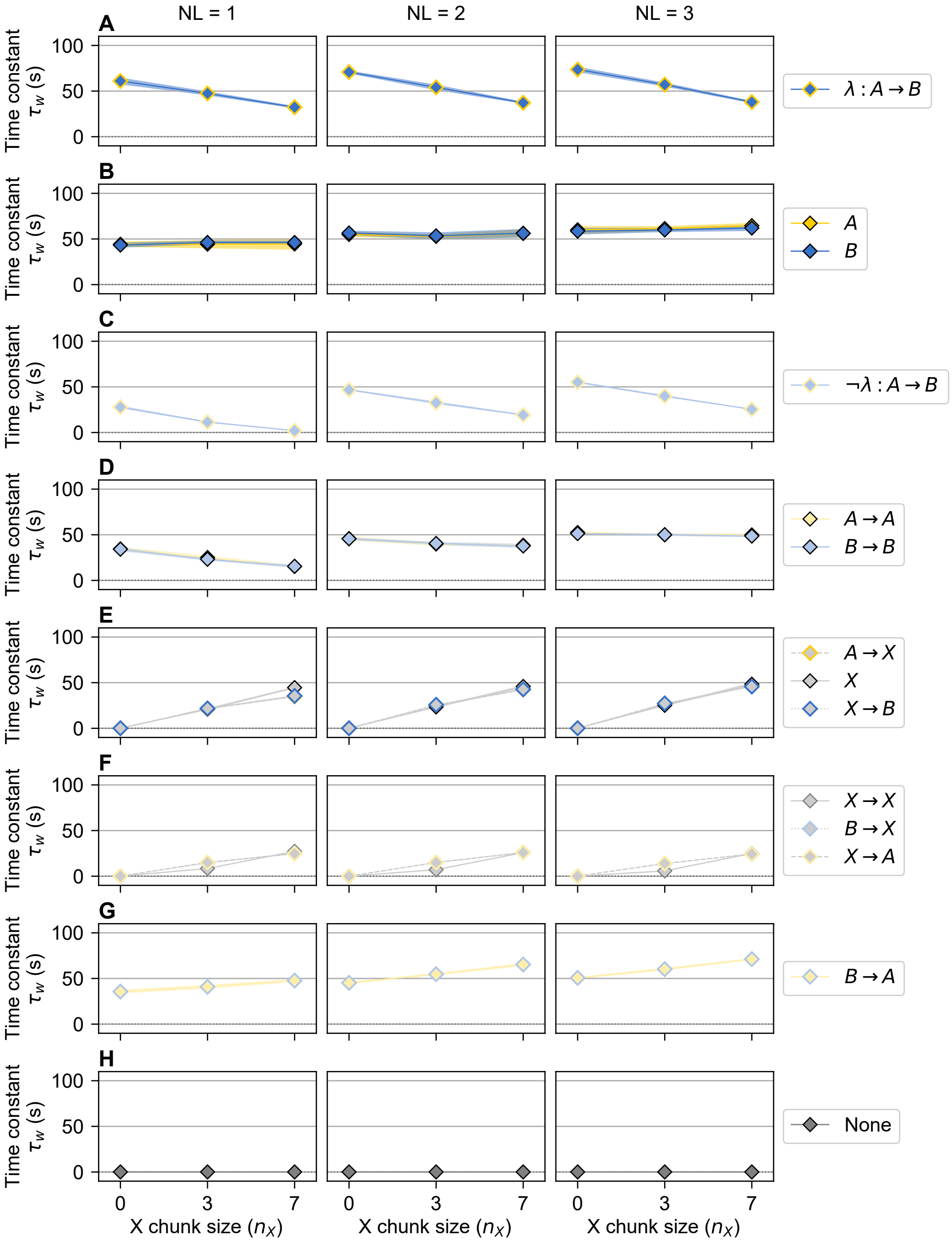


Figure 3.37: Assembly medians (synaptic weight time constant τ_w) in experiment 'X chunk size'. Each individual data point (= n) indicates mean ± standard deviation (SD) across the RNNs (N = 10). Columns indicate grammar (NL = 1: isolated NAD, NL = 2 and NL = 3: nested NADs). N_X = number of available X elements. Dotted line = lower weight boundary. **A** Grammatical A to B transitions (n = 10). **B** A and B elements (n = 10). **C** Non-grammatical A to B transitions (n = 10). **D** A to A and B to B transitions ($\widehat{AA}\widehat{BB}$, n = 10). **E** X elements, A to X and X to B transitions (χ, n = 10). **F** X to X, B to X and X to A transitions (χ', n = 10). **G** B to A transitions (n = 10). **H** Non-assembly synapses (None, n = 10).

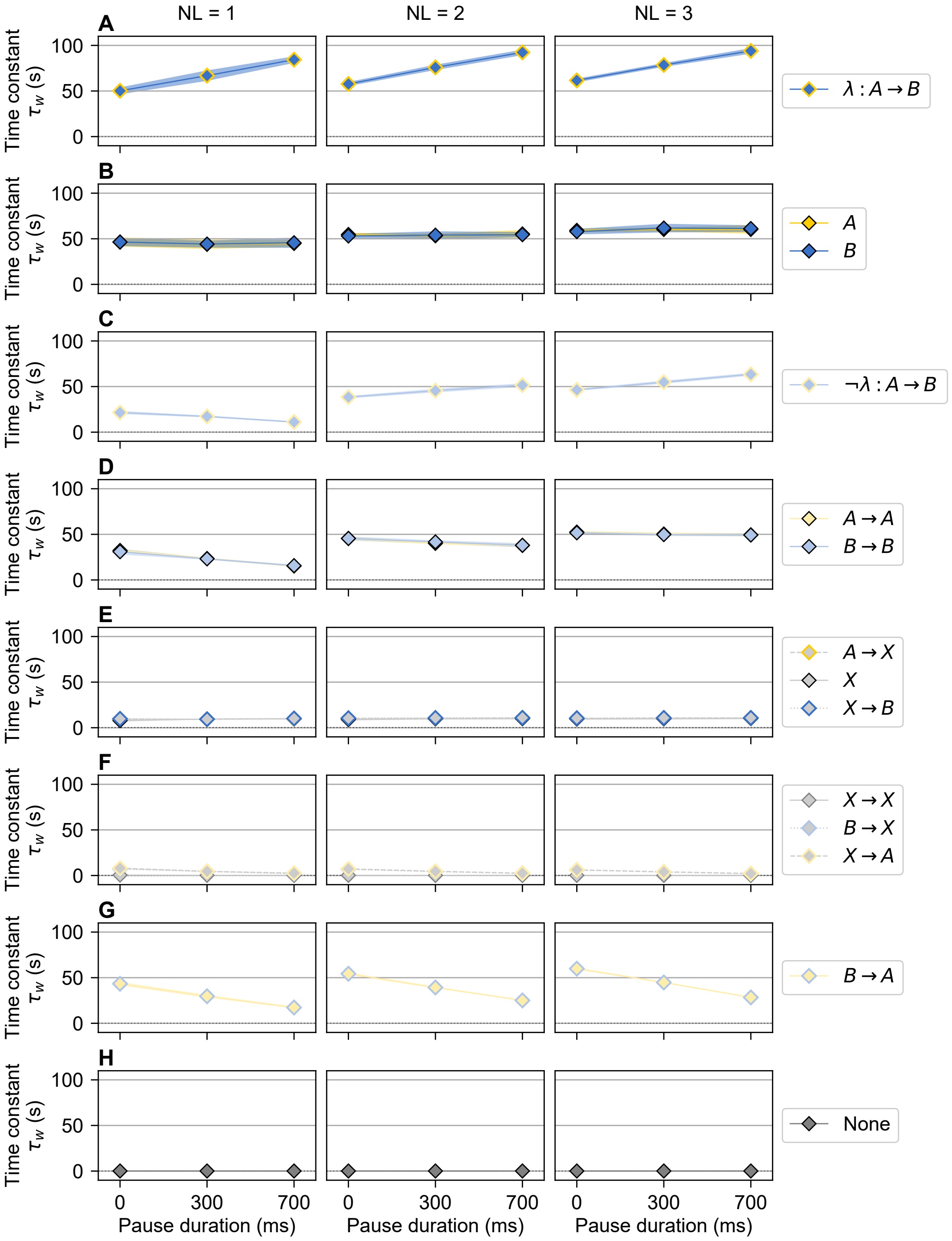


Figure 3.38: Assembly medians (synaptic weight time constant τ_w) in experiment 'Pause variability'. Each individual data point (= n) indicates mean ± standard deviation (SD) across the RNNs (N = 10). Columns indicate grammar (NL = 1: isolated NAD, NL = 2 and NL = 3: nested NADs). N_X = number of available X elements. Dotted line = lower weight boundary. **A** Grammatical A to B transitions (n = 10). **B** A and B elements (n = 10). **C** Non-grammatical A to B transitions (n = 10). **D** A to A and B to B transitions ($\widehat{AA}\widehat{BB}$, n = 10). **E** X elements, A to X and X to B transitions (χ, n = 10). **F** X to X, B to X and X to A transitions (χ', n = 10). **G** B to A transitions (n = 10). **H** Non-assembly synapses (None, n = 10).

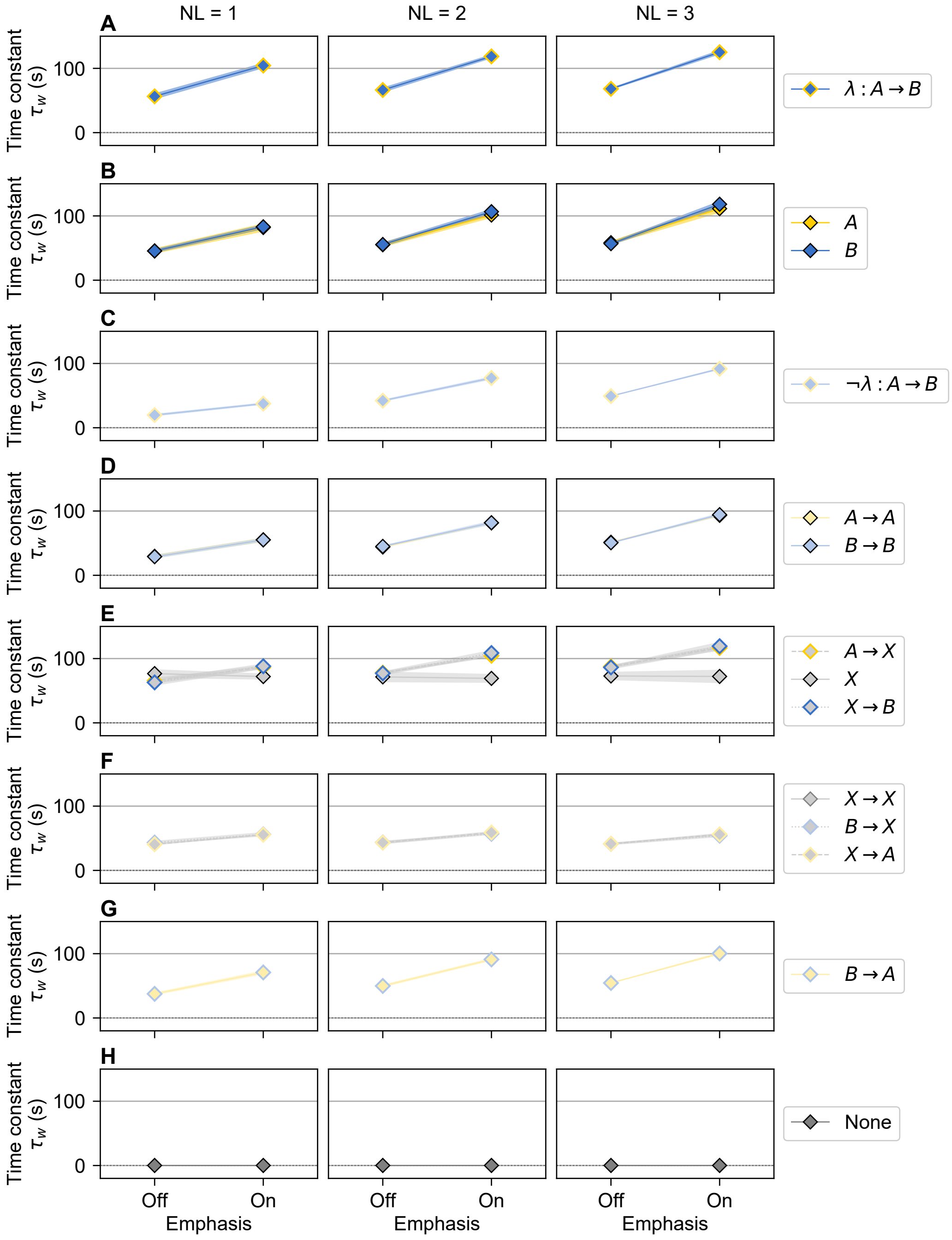


Figure 3.39: Assembly medians (synaptic weight time constant τ_w) in experiment 'Emphasis'. Each individual data point (= n) indicates mean ± standard deviation (SD) across the RNNs (N = 10). Columns indicate grammar (NL = 1: isolated NAD, NL = 2 and NL = 3: nested NADs). N_X = number of available X elements. Dotted line = lower weight boundary. **A** Grammatical A to B transitions (n = 10). **B** A and B elements (n = 10). **C** Non-grammatical A to B transitions (n = 10). **D** A to A and B to B transitions ($\widehat{AA}\widehat{BB}$, n = 10). **E** X elements, A to X and X to B transitions (χ, n = 10). **F** X to X, B to X and X to A transitions (χ', n = 10). **G** B to A transitions (n = 10). **H** Non-assembly synapses (None, n = 10).

3.3.4 Separability measures

3.3.4.1 Complete collection of separability measures and meta analysis

Next to the investigation of synaptic w and τ_w dynamics, the modelled grammar learning performance was further characterised by so called ‘separability measures’. These were defined as the ratios of grammatical (Λ) and non-grammatical ($\neg\Lambda$) synapse assemblies, i.e. $\frac{\Lambda}{\neg\Lambda}$. For details, see Sec. 3.1.2 ‘Separability measures’ (Lehfeldt et al. n.d.). The manuscript by Lehfeldt et al. (n.d.) showed $\frac{\lambda:A\rightarrow B}{\neg\lambda:A\rightarrow B}$ and $\frac{\lambda:A\rightarrow B}{\chi}$ as the selection of separability measures that demonstrated an accordance between modelled and human NAD learning performances. For completeness, the appendix of this thesis included the full set of separability measures across experiments (‘X variability’: Fig. A.1 and Fig. A.2, ‘X chunk size’: Fig. A.3 and Fig. A.4, ‘Pause variability’: Fig. A.5 and Fig. A.6, ‘Emphasis’: Fig. A.7 and Fig. A.8). In Lehfeldt et al. (n.d.), the analysis was focused on separability measures taking the grammatical A to B transition assemblies as a reference. This was motivated by the fact that $\lambda : A \rightarrow B$ was the model equivalent to the grammatical dependency that participants are usually tested for in psycho- and neurolinguistic experiments (Mueller et al. 2012; Winkler et al. 2018a; Wilson et al. 2018). However, next to $\lambda : A \rightarrow B$, the correct acquisition of grammatical A and B elements was as well important for demonstrating an overall successful grammar learning. Thus, also the separability measures taking A and B as references were demonstrated in the appendix of this thesis. Interested readers might take a closer look at the details of all separability measure dynamics (Fig. A.1 to Fig. A.8). However, in order to gain a full view on the grammar learning performance, a meta analysis of the complete set of separability measures was performed (Fig. 3.40 to Fig. 3.47). This analysis indicated whether ratios were above the critical mark of one or below, i.e. provided a binary measure of learning success. Investigating all possible separability measures was of particular importance for demonstrating the general reliability of the learning outcome, whereby ideally all $\frac{\Lambda}{\neg\Lambda}$ ratios resulted in values larger than one. For analysis, the mean values of separability measures minus the respective standard deviations across RNNs (N = 10) were used. This subtraction from the means was done in order to obtain a success measure that also considered the standard deviations of values.

3.3.4.2 Experiment ‘X variability’: meta analysis

In experiment ‘X variability’, separability measures with $\lambda : A \rightarrow B$ as a reference showed equal success distributions in w and τ_w (Fig. 3.40**A** and Fig. 3.41**A**). For both variables, learning was successful in all measures and conditions except the N_X = 1 variability condition. Further, the $\frac{\lambda:A\rightarrow B}{X\rightarrow X}$ separability was NaN in this variability condition given that during training with only one available X element, no X to X transitions occurred (this equivalently applied to the $\frac{A}{X\rightarrow X}$ and $\frac{B}{X\rightarrow X}$ measures). Separability measures with A and B as references showed overall equal dynamics in w and τ_w, however with some variations (Fig. 3.40**B**,**C** and Fig. 3.41**B**,**C**) Here, different success measures in w and τ_w occurred for the experiment conditions NL = 1: $\frac{A}{X\rightarrow A}$ with $N_X = 1$, NL = 2: $\frac{A}{B\rightarrow A}$ with $N_X = 5$, NL = 1: $\frac{B}{B\rightarrow A}$ with $N_X = 15$ and

NL $= 2$: $\frac{B}{B\to A}$ with $N_X = 5$. However, when investigating the belonging separability measure values, differences occurred for those ratios that were concentrated around the critical mark of one (Fig. A.2). Thus, observed differences in success measures were not considered to reflect a systematic effect but where rather interpreted as a result of data variability in the vicinity of the critical mark. Still, these measures indicated that a separation of grammatical from non-grammatical assemblies was not very pronounced here. Across grammars and X variability conditions, $\frac{A}{\lambda:A\to B}$ and $\frac{B}{\lambda:A\to B}$ were below the critical mark of one, indicating that A and B grammar element representations were weaker than grammatical A to B transition assemblies (Fig. A.2). As a final observation, $\frac{B}{A}$ and $\frac{A}{B}$ were both below the critical mark of one, which seemed odd at first. However, as the belonging separability measures were concentrated around the critical mark of one (Fig. A.2), the subtraction of the respective standard deviations apparently returned unsuccessful measures in both. Taken together, the learning performance across grammars and X variability conditions was the most robust in $\lambda : A \to B$. A and B showed more fluctuations in the learning success and interestingly, they were consistently weaker than $\lambda : A \to B$. However, fluctuations did not concern the separability from χ or χ' in $N_X > 1$ variability conditions, so that the beneficial effect of large X variabilities for NAD learning (Lehfeldt et al. n.d.) was further supported.

3.3.4.3 Experiment 'X chunk size': meta analysis

In experiment 'X chunk size', separability measures with $\lambda : A \to B$ as a reference showed overall equal success distributions in w and τ_w, however with one variation in the experiment condition NL $= 2$: $\frac{\lambda:A\to B}{B\to B}$ with $n_x = 7$ (Fig. 3.42**A** and Fig. 3.43**A**). Despite this, all measures were successful for the $n_X = 0$ chunk size condition, i.e. 'Starting small'. For large X chunks (i.e. with $n_X = 3$ and $n_X = 7$), unsuccessful measures occurred in $\frac{\lambda:A\to B}{A}$ and $\frac{\lambda:A\to B}{B}$, indicating that A and B were slightly or prominently stronger than $\lambda : A \to B$ (Fig. A.3). Further, for the largest X chunk size tested, $\frac{\lambda:A\to B}{\chi}$ was unsuccessful across all grammars. Here, also $\frac{\lambda:A\to B}{\widehat{AA}\widehat{BB}}$ was unsuccessful for nested NADs, except for NL $= 2$: $\frac{\lambda:A\to B}{B\to B}$ in w. Thus, long X chunks added the additional difficulty to separate $\lambda : A \to B$ from non-grammatical $A \to A$ and $B \to B$ transitions in nested NADs. Separability measures with A and B as references showed equal success dynamics in w and τ_w, however with one exception for $\frac{A}{\lambda:A\to B}$ in the NL $= 3$ and $n_X = 3$ condition (Fig. 3.42**B**,**C** and Fig. 3.43**B**,**C**). This deviation and also the unsuccessful $\frac{B}{A}$ and $\frac{A}{B}$ measures were, however, again explained by a concentration of the belonging separability measures around the critical mark of one (Fig. A.4). For X chunks with $n_X > 0$, $\frac{A}{B\to A}$ and $\frac{B}{B\to A}$ were unsuccessful, though with an exception for $\frac{B}{B\to A}$ in the NL $= 1$ and $n_X = 3$ condition. Thus, in overall, long X chunks added the additional difficulty to separate A and B from non-grammatical and cross-sample $B \to A$ transitions. Further, $\frac{A}{X}$ and $\frac{B}{X}$ were not successful in the experiment condition NL $= 1$ and $n_X = 7$, though see Fig. A.4. With regard to $\frac{A}{\lambda:A\to B}$ and $\frac{B}{\lambda:A\to B}$, long X chunks were however beneficial. Here, only in the $n_X = 7$ chunk size condition, $\frac{A}{\lambda:A\to B}$ and $\frac{B}{\lambda:A\to B}$ were successful throughout. Finally, $\frac{B}{\neg\lambda:A\to B}$ was not successful in the NL $= 3$ and $n_X = 0$ chunk size condition, though see Fig. A.4. As a final observation, A and B

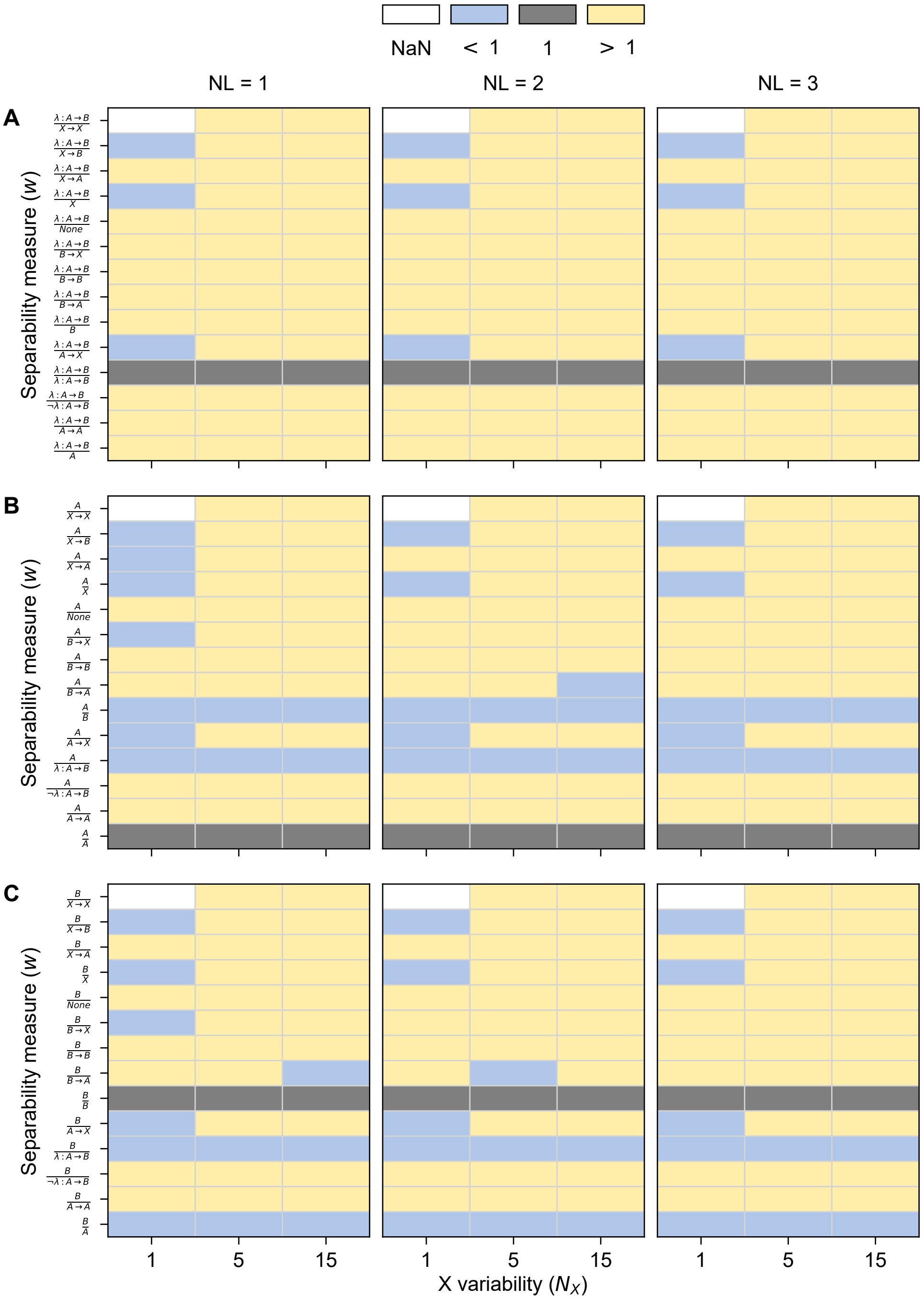


Figure 3.40: Separability measure meta analysis of w in experiment 'X variability'. Each individual tile indicates by colour if the respective separability measure was NaN (white), below one (blue), equal to one (grey) or above one (yellow). **A** Reference assembly: grammatical A to B transitions ($\lambda : A \rightarrow B$). **B** Reference assembly: grammatical A elements (A). **C** Reference assembly: grammatical B elements (B).

were more robust against large X chunks than $\lambda : A \rightarrow B$, as indicated by largely successful $\frac{A}{X}$ and $\frac{B}{X}$ measures across n_X conditions. Taken together, large X chunk

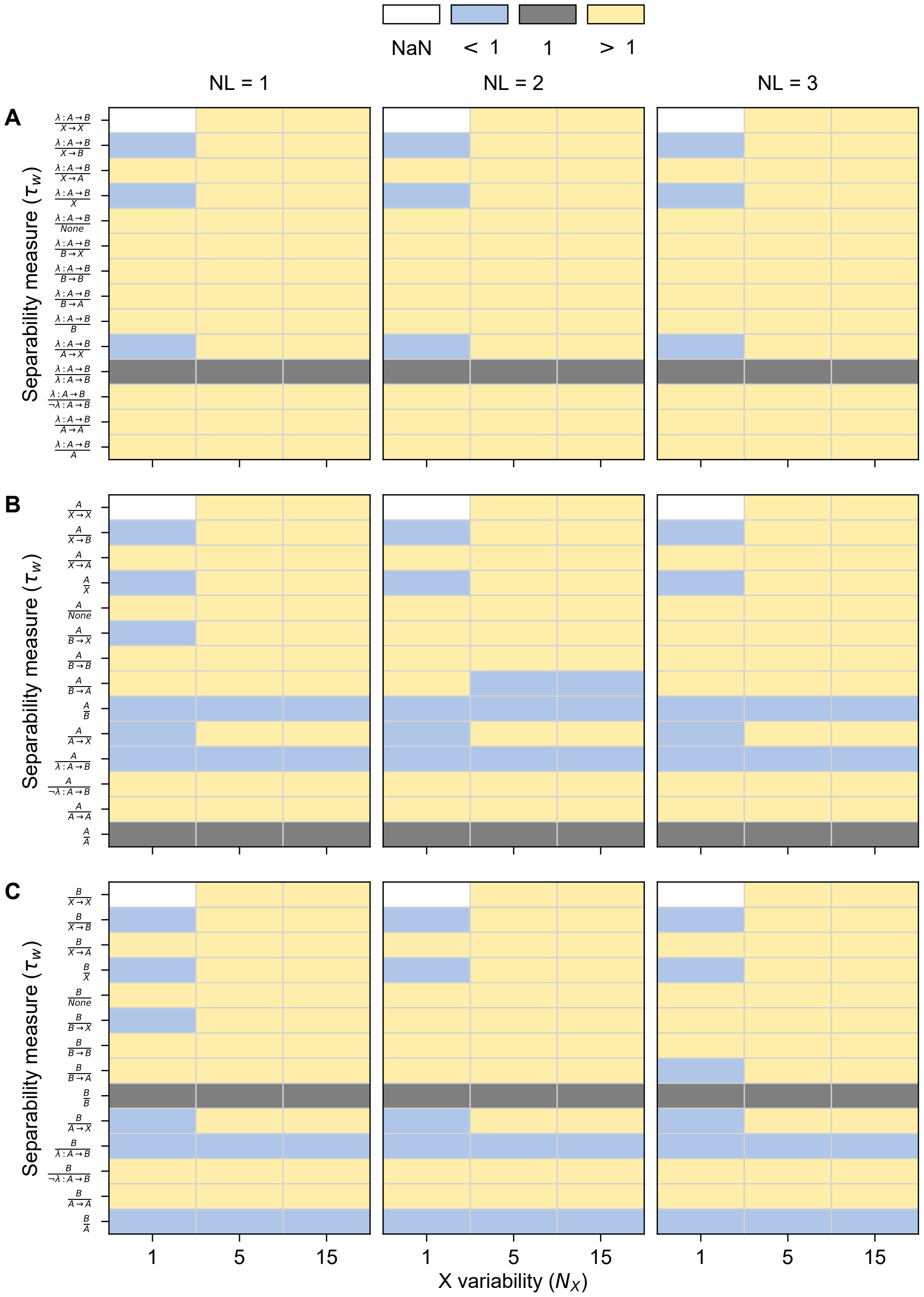


Figure 3.41: Separability measure meta analysis of τ_w in experiment 'X variability'. Each individual tile indicates by colour if the respective separability measure was NaN (white), below one (blue), equal to one (grey) or above one (yellow). **A** Reference assembly: grammatical A to B transitions ($\lambda : A \to B$). **B** Reference assembly: grammatical A elements (A). **C** Reference assembly: grammatical B elements (B).

sizes introduced additional difficulties for the separation from $\widehat{AA}\widehat{BB}$ and $B \to A$. Thus, the beneficial effect of 'Starting small' for NAD learning (Lehfeldt et al. n.d.) was further supported.

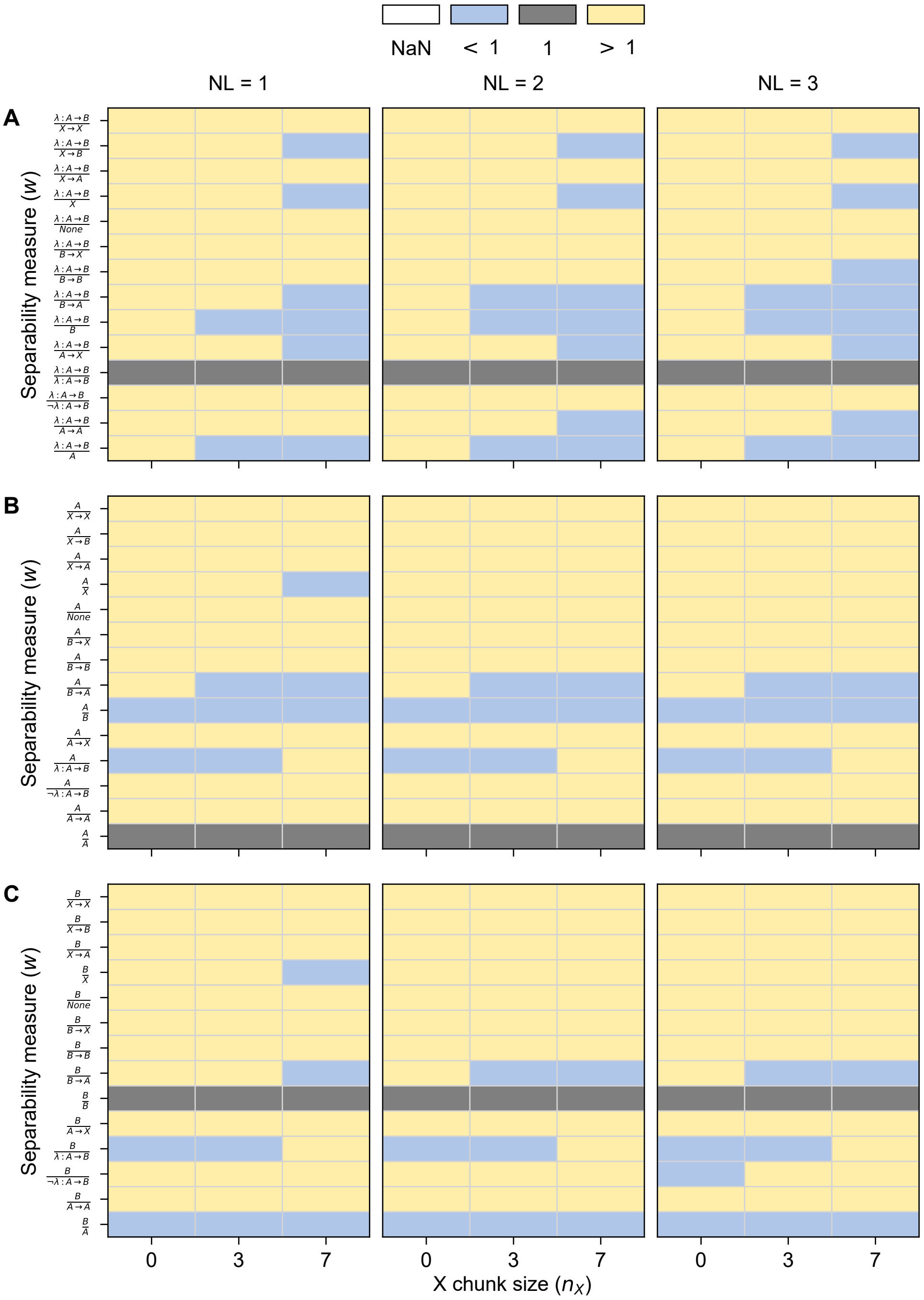


Figure 3.42: Separability measure meta analysis of w in experiment 'X chunk size'. Each individual tile indicates by colour if the respective separability measure was NaN (white), below one (blue), equal to one (grey) or above one (yellow). **A** Reference assembly: grammatical A to B transitions ($\lambda : A \to B$). **B** Reference assembly: grammatical A elements (A). **C** Reference assembly: grammatical B elements (B).

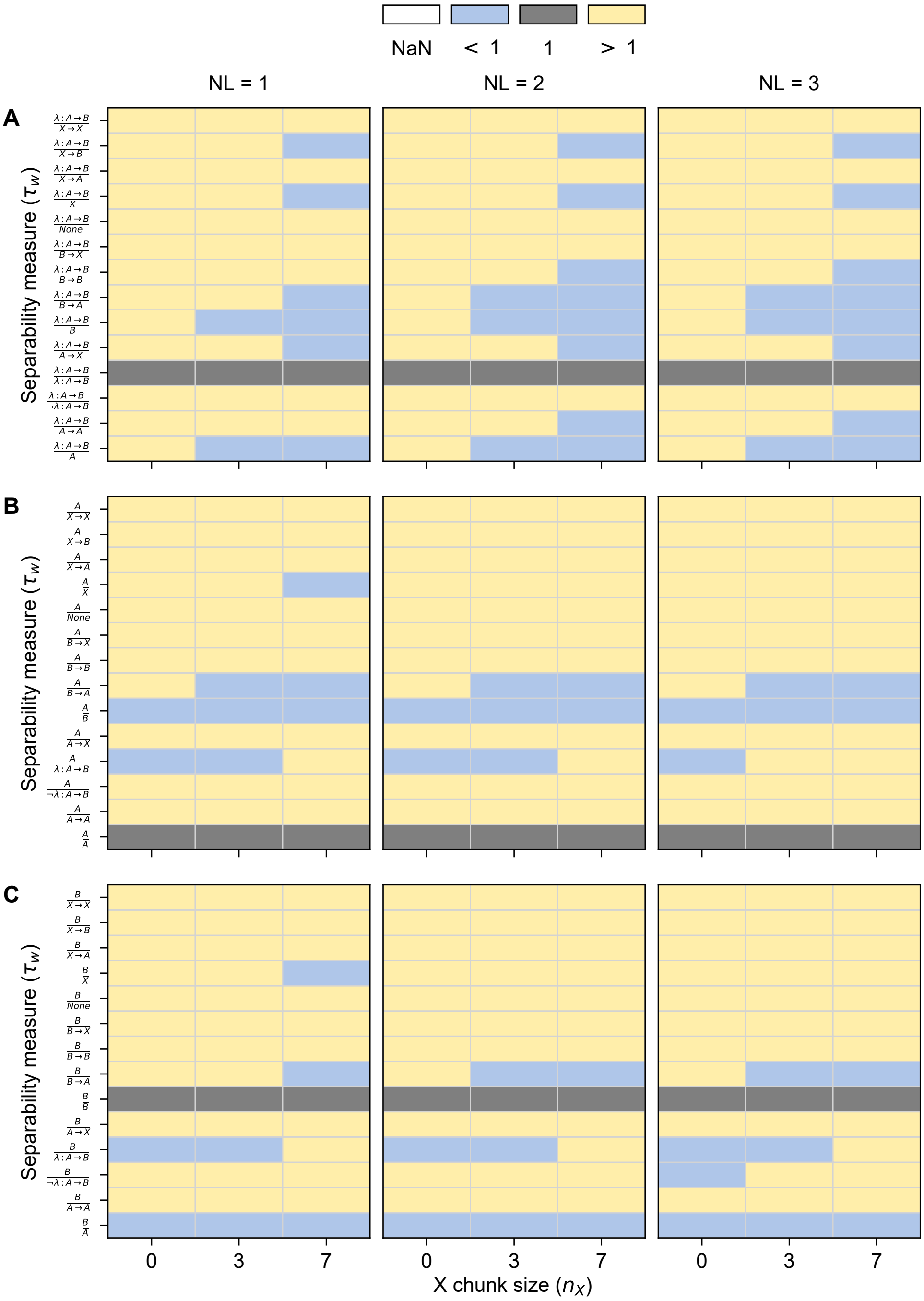


Figure 3.43: Separability measure meta analysis of τ_w in experiment 'X chunk size'. Each individual tile indicates by colour if the respective separability measure was NaN (white), below one (blue), equal to one (grey) or above one (yellow). **A** Reference assembly: grammatical A to B transitions ($\lambda : A \to B$). **B** Reference assembly: grammatical A elements (A). **C** Reference assembly: grammatical B elements (B).

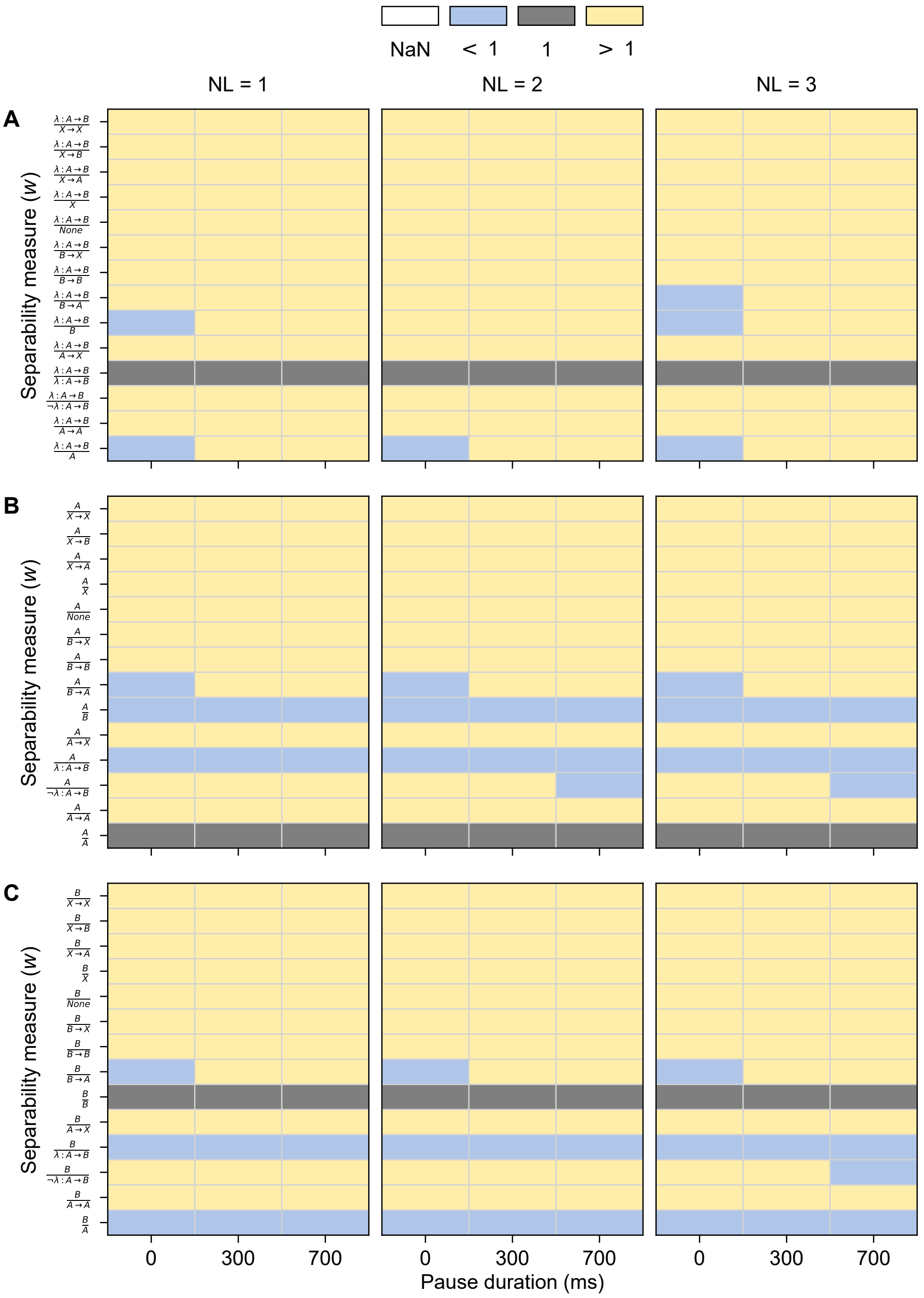


Figure 3.44: Separability measure meta analysis of w in experiment 'Pause variability'. Each individual tile indicates by colour if the respective separability measure was NaN (white), below one (blue), equal to one (grey) or above one (yellow). **A** Reference assembly: grammatical A to B transitions ($\lambda : A \rightarrow B$). **B** Reference assembly: grammatical A elements (A). **C** Reference assembly: grammatical B elements (B).

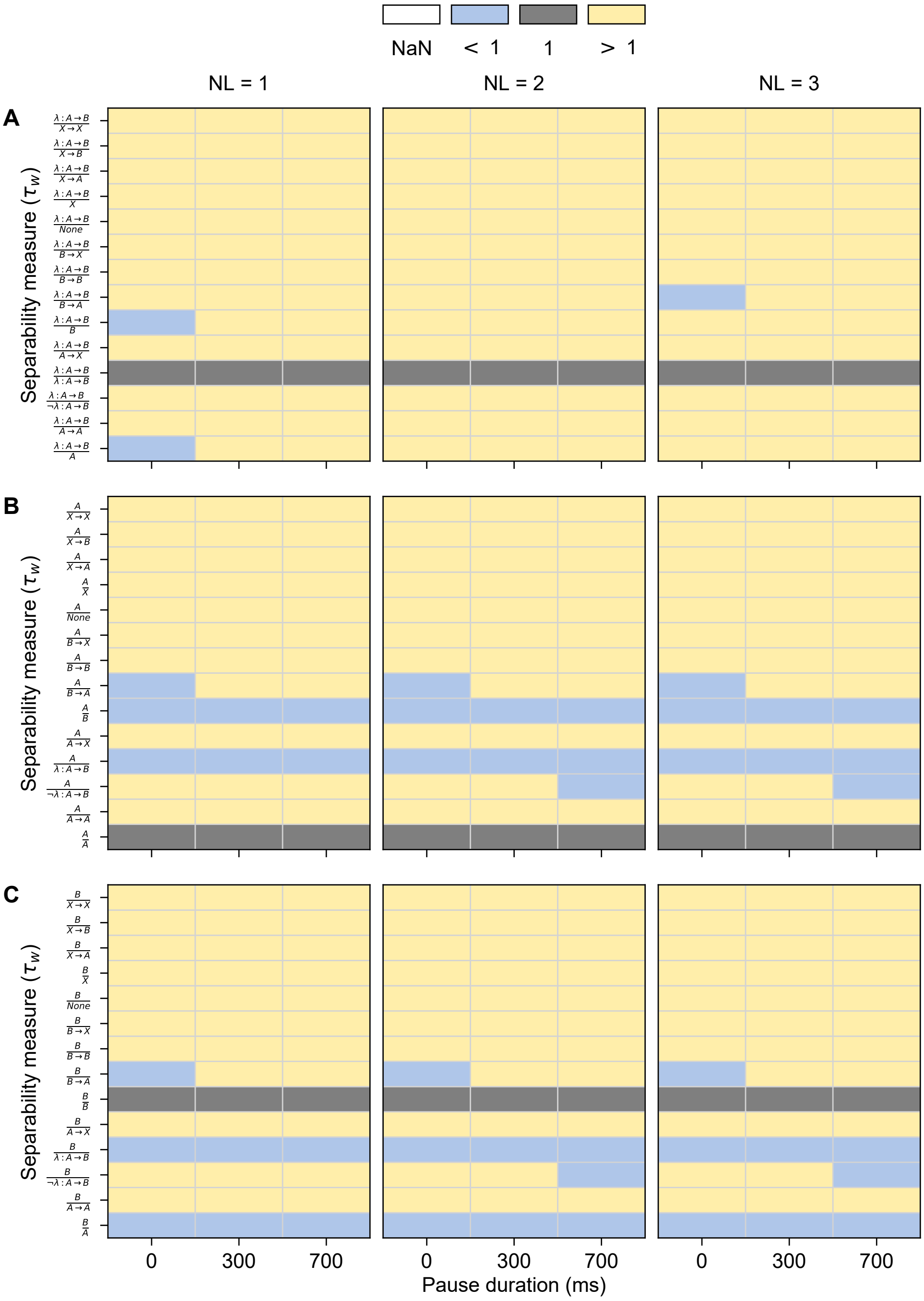


Figure 3.45: Separability measure meta analysis of τ_w in experiment 'Pause variability'. Each individual tile indicates by colour if the respective separability measure was NaN (white), below one (blue), equal to one (grey) or above one (yellow). **A** Reference assembly: grammatical A to B transitions ($\lambda : A \rightarrow B$). **B** Reference assembly: grammatical A elements (A). **C** Reference assembly: grammatical B elements (B).

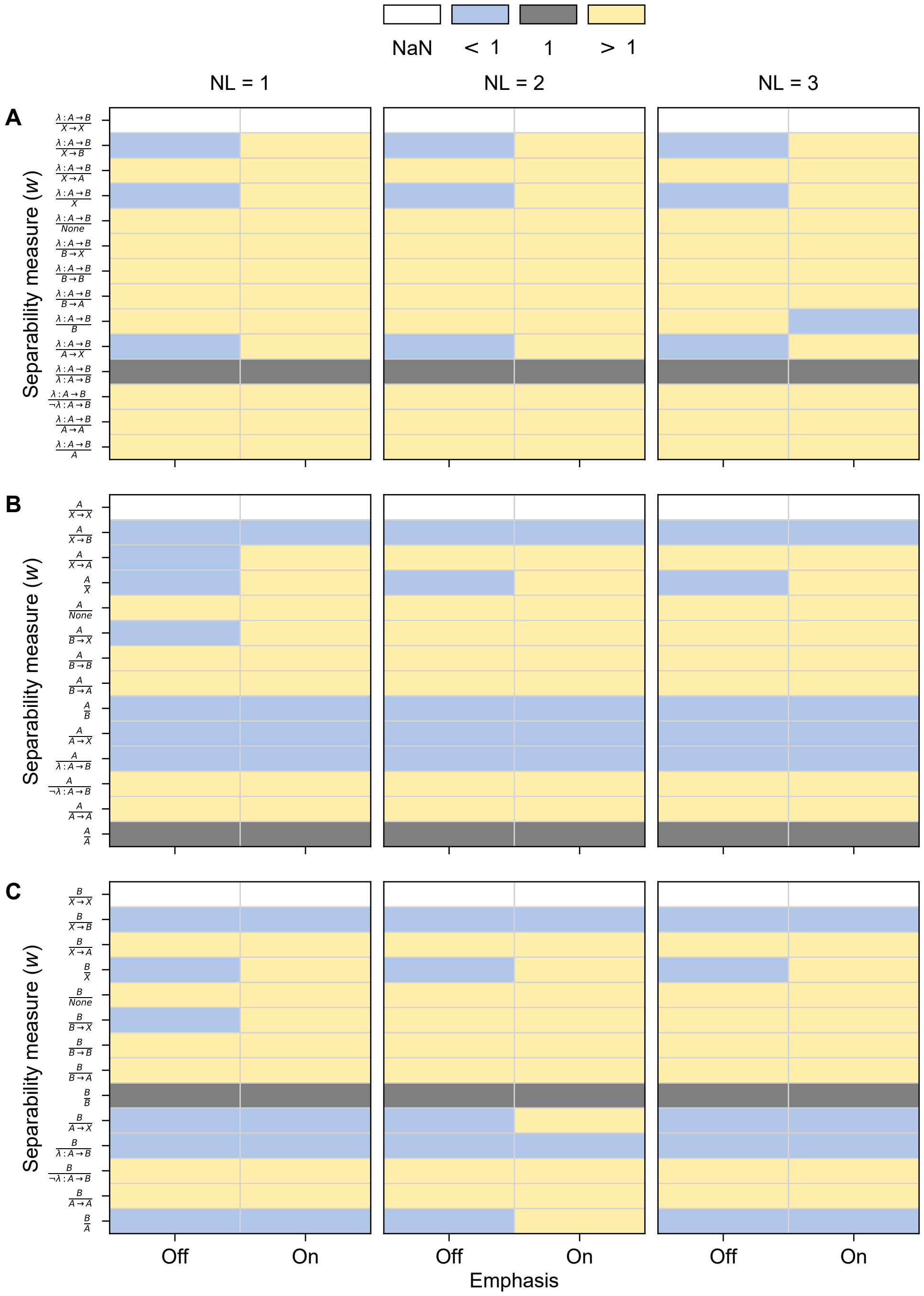


Figure 3.46: Separability measure meta analysis of w in experiment 'Emphasis'. Each individual tile indicates by colour if the respective separability measure was NaN (white), below one (blue), equal to one (grey) or above one (yellow). **A** Reference assembly: grammatical A to B transitions ($\lambda : A \rightarrow B$). **B** Reference assembly: grammatical A elements (A). **C** Reference assembly: grammatical B elements (B).

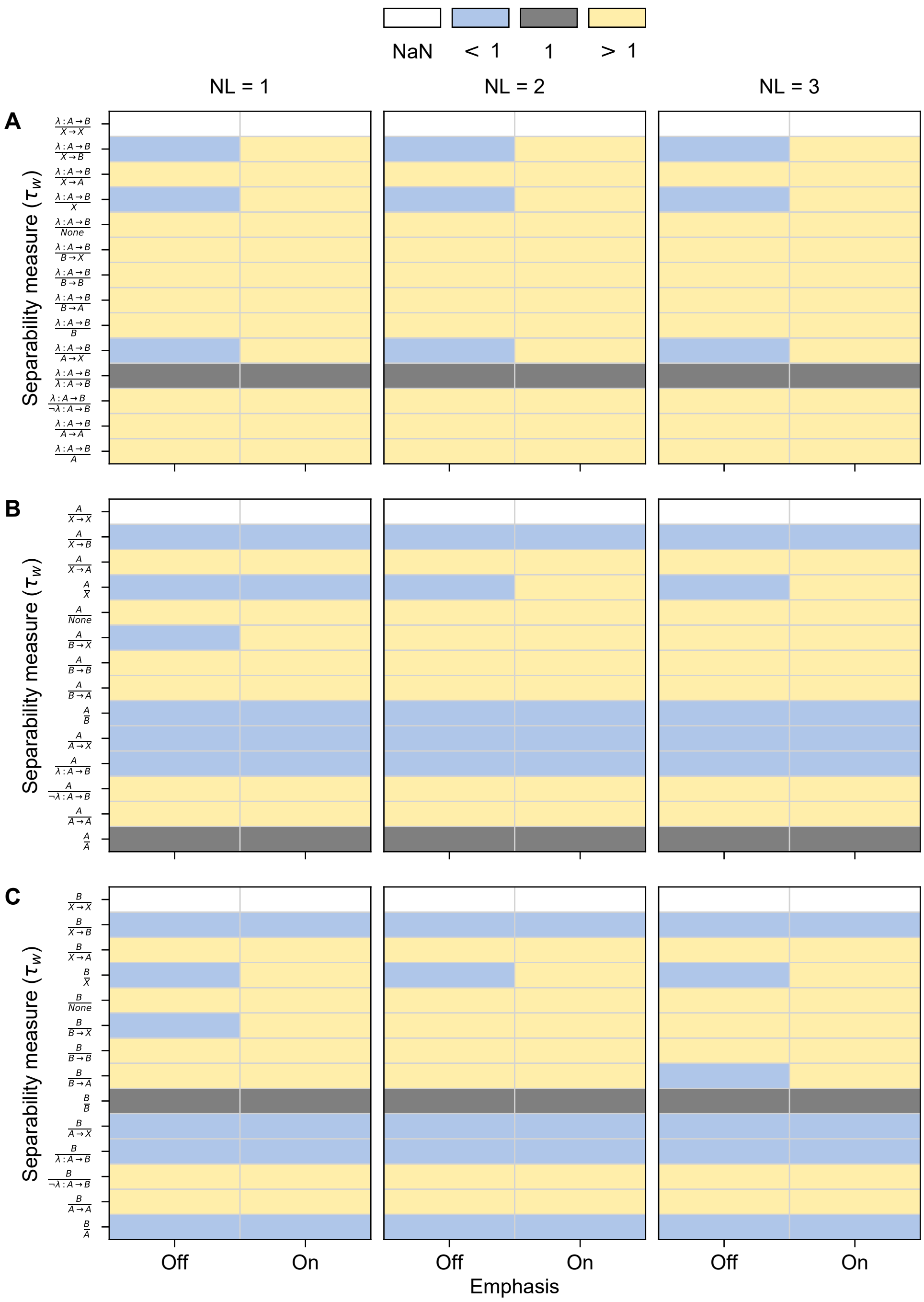


Figure 3.47: Separability measure meta analysis of τ_w in experiment 'Emphasis'. Each individual tile indicates by colour if the respective separability measure was NaN (white), below one (blue), equal to one (grey) or above one (yellow). **A** Reference assembly: grammatical A to B transitions ($\lambda : A \to B$). **B** Reference assembly: grammatical A elements (A). **C** Reference assembly: grammatical B elements (B).

3.3.4.4 Experiment 'Pause variability': meta analysis

In experiment 'Pause variability', separability measures with $\lambda : A \rightarrow B$ as a reference showed overall equal success distributions in w and τ_w, however with some variations (Fig. 3.44**A** and Fig. 3.45**A**). Here, different success measures in w and τ_w occurred for the experiment conditions NL = 2: $\frac{\lambda:A\rightarrow B}{A}$ with $\cup_\Delta = 0$ ms, NL = 3: $\frac{\lambda:A\rightarrow B}{A}$ with $\cup_\Delta = 0$ ms and NL = 3: $\frac{\lambda:A\rightarrow B}{B}$ with $\cup_\Delta = 0$ ms. Throughout, unsuccessful measures only occurred during training without stimulus pauses, i.e. in the $\cup_\Delta = 0$ ms condition. Separability measures with A and B as references showed overall equal success dynamics in w and τ_w, however with one exception for $\frac{B}{\neg\lambda:A\rightarrow B}$ in the NL = 2 and $\cup_\Delta = 700$ ms condition (Fig. 3.44**B**,**C** and Fig. 3.45**B**,**C**). The separation from χ and χ' was successful across all pause conditions. However, in the $\cup_\Delta = 0$ ms condition, $\frac{A}{B\rightarrow A}$ and $\frac{B}{B\rightarrow A}$ were not successful. Further, in nested NADs, $\frac{A}{\neg\lambda:A\rightarrow B}$ and $\frac{B}{\neg\lambda:A\rightarrow B}$ were largely not successful during training with the longest pause duration tested, i.e. $\cup_\Delta = 700$ ms. Again, $\frac{B}{A}$ and $\frac{A}{B}$ were consistently both below the critical mark of one, though see Fig. A.6. Finally, also $\frac{A}{\lambda:A\rightarrow B}$ and $\frac{B}{\lambda:A\rightarrow B}$ were consistently unsuccessful, indicating that $\lambda : A \rightarrow B$ was slightly or prominently stronger than A and B (Fig. A.6). Taken together, the beneficial effect of long stimulus pauses for NAD learning (Lehfeldt et al. n.d.) was further supported. However, in nested NADs, the separability of A and B from $\neg\lambda : A \rightarrow B$ suffered from pause conditions $\cup_\Delta > 300$ ms. Thus, the beneficial effect of long stimulus pauses has to be considered within this identified constraint (Fig. A.6).

3.3.4.5 Experiment 'Emphasis': meta analysis

In experiment 'Emphasis', separability measures with $\lambda : A \rightarrow B$ as references showed overall equal success distributions in w and τ_w, however with one exception for the experiment condition NL = 3: $\frac{\lambda:A\rightarrow B}{B}$ with 'On' emphasis (Fig. 3.46**A** and Fig. 3.47**A**), though see Fig. A.7. Despite this, all remaining measures were consistently successful in the 'On' emphasis condition. Separability measures with A and B as references showed overall equal success distributions in w and τ_w, however with exceptions for the experiment conditions NL = 1: $\frac{A}{X\rightarrow A}$ with 'Off' emphasis, NL = 1: $\frac{A}{X}$ with 'On' emphasis, NL = 2: $\frac{B}{A\rightarrow X}$ with 'On' emphasis, NL = 2: $\frac{B}{A}$ with 'On' emphasis and NL = 3: $\frac{B}{B\rightarrow A}$ with 'Off' emphasis (Fig. 3.46**B**,**C** and Fig. 3.47**B**,**C**), though see Fig. A.8. However, the separability measures $\frac{A}{A\rightarrow X}$, $\frac{B}{A\rightarrow X}$ and $\frac{A}{X\rightarrow B}$, $\frac{B}{X\rightarrow B}$ remained largely unsuccessful in the 'On' emphasis condition. Thus, in contrast to $\lambda : A \rightarrow B$, the separability of A and B from χ did not increase consistently by emphasis. Again, $\frac{B}{A}$ and $\frac{A}{B}$ were largely both below the critical mark of one, though see Fig. A.8. Throughout, $\frac{\lambda:A\rightarrow B}{X\rightarrow X}$, $\frac{A}{X\rightarrow X}$ and $\frac{B}{X\rightarrow X}$ were NaN, given that training was conducted in the $N_X = 1$ variability condition. Finally, also $\frac{A}{\lambda:A\rightarrow B}$ and $\frac{B}{\lambda:A\rightarrow B}$ were consistently unsuccessful, indicating that $\lambda : A \rightarrow B$ was slightly stronger than A and B across the emphasis conditions (Fig. A.8). Taken together, the beneficial effect of emphasis on the $\frac{\lambda:A\rightarrow B}{X}$ separability (Lehfeldt et al. n.d.) was largely also present in $\frac{A}{X}$ and $\frac{B}{X}$. However, as only the separability measure $\frac{\lambda:A\rightarrow B}{\chi}$ consistently benefited from emphasis, its facilitating effect has to be considered within the here identified constraints.

3.3.5 SLMT for robust learning

A major problem in the research area of spike-based statistical learning is the emergence of unstable network states, also referred to as ‘runaway activity’ (Abbott and Nelson 2000; Gütig et al. 2003; Chistiakova et al. 2014; Chistiakova et al. 2015; Keck et al. 2017). Runaway states are characterised by strong ‘on-off’ oscillations at maximal firing rates (Lazar et al. 2009; Zenke et al. 2013; Litwin-Kumar and Doiron 2014) and by saturated synaptic weights (Chen et al. 2013). As a consequence, networks are prevented from proper functioning and previously established synaptic memories are destroyed. Underlying to runaway activity is a positive feedback loop introduced by the associativity of spike-based statistical learning (Zenke et al. 2013; Keck et al. 2017). During training, the continuous strengthening of synapses between neurons that share temporally correlated activities leads to continuously stronger excitatory activations and hence, increased neural firing. In turn, this promotes the further growth of synapse strengths and thereby causes an even further increase of spiking activity. At some point, this positive feedback loop becomes so strong that previously silent neurons, i.e. those that were not part of the memory subset, also become activated. Consequently, the growth of synapses external to the actual memory is initiated, adding extra spiking activity and potentiation to the network state. As a final result of this spreading, networks inevitably converge towards unstable runaway dynamics. Solving this issue is still a topic of ongoing research, whereby the main idea for preventing runaway states is to combine purely associative learning with other compensatory plasticity mechanisms (Zenke et al. 2013; Chistiakova et al. 2014; Chistiakova et al. 2015; Keck et al. 2017; Zenke and Gerstner 2017). For example, stability was achieved by extending associative learning with neuronal intrinsic plasticity and normalisation (Lazar et al. 2009), fast rate detectors (Zenke et al. 2013), heterosynaptic plasticity (Chen et al. 2013), a combination of inhibitory plasticity and normalisation (Litwin-Kumar and Doiron 2014), a combination of short-term and inhibitory plasticity with a consolidation mechanism (Zenke et al. 2015) or by using structural plasticity for learning (Gallinaro et al. 2022). Further approaches included additional dynamics of synaptic variables such as bistability (Fusi 2002; Graupner and Brunel 2012a; Graupner and Brunel 2012b) or adaptive weight decays (Zuo et al. 2017a; Zuo et al. 2017b; Panda and Roy 2017; Panda et al. 2018). Here, the idea was that synaptic variables are not temporally infinite but rather exhibit bistable or decaying behaviours during the absence of input at the synapse. Especially weight decay adaptivity was interpreted as a way to implement forgetting and has already been shown to stabilise spike-based learning (Panda and Roy 2017) and to increase performance by preventing catastrophic forgetting (Panda et al. 2018). The concept of forgetting by weight decay recently also found application in models of auditory predictions (Harrison et al. 2020) or statistical sequence learning (Endress and Johnson 2021). In this thesis, stable learning was achieved by a combination of spike-timing dependent plasticity (STDP), neural intrinsic plasticity (IP) and a short- to long-term memory transition (SLMT) mechanism that implemented an activity dependent weight decay adaptivity. For details on the SLMT implementation see Sec. 3.1.2 ‘Spike-timing dependent plasticity (STDP) with short- to long-term memory transition (SLMT)’ (Lehfeldt et al. n.d.). In order to demonstrate the stabilising effect of SLMT, the experiments

'X variability' and 'Emphasis' were run with STDP and IP as only plasticity types, i.e. without SLMT. In response to the same experimental conditions as in the grammar learning experiment series (compare Fig. 3.48 and Fig. 3.4), runaway activities occurred during training without SLMT. While during learning with STDP, IP and SLMT no unstable dynamics were observed (300 simulations in total), during training with STDP and IP only, runaway occurred in 37 out of 120 simulations. In experiment 'X variability', learning of NL = 1 and NL = 2 grammars was still stable while learning of NL = 3 grammars resulted in runaway dynamics (Fig. 3.49). This was demonstrated by saturation of synaptic weights and overall larger standard deviations of weight medians. In experiment 'Emphasis', runaway occurred in NL = 1 and NL = 2 grammars for the 'On' condition and in NL = 3 grammars for both 'Off' and 'On' conditions (Fig. 3.50). Examples of training runs with runaway spiking activity were demonstrated in Fig. 3.51 and Fig. 3.52. Throughout, the stabilising effect of SLMT was explained by the adaptive weight decay, as this constantly counteracted the positive feedback loop induced by associative learning. Due to decaying weights, the unstable spreading of excitatory activation and hence, the neural firing outside of the memory subsets was successfully prevented. Within the training conditions of this thesis, SLMT was therefore sufficient to prevent unstable runaway dynamics.

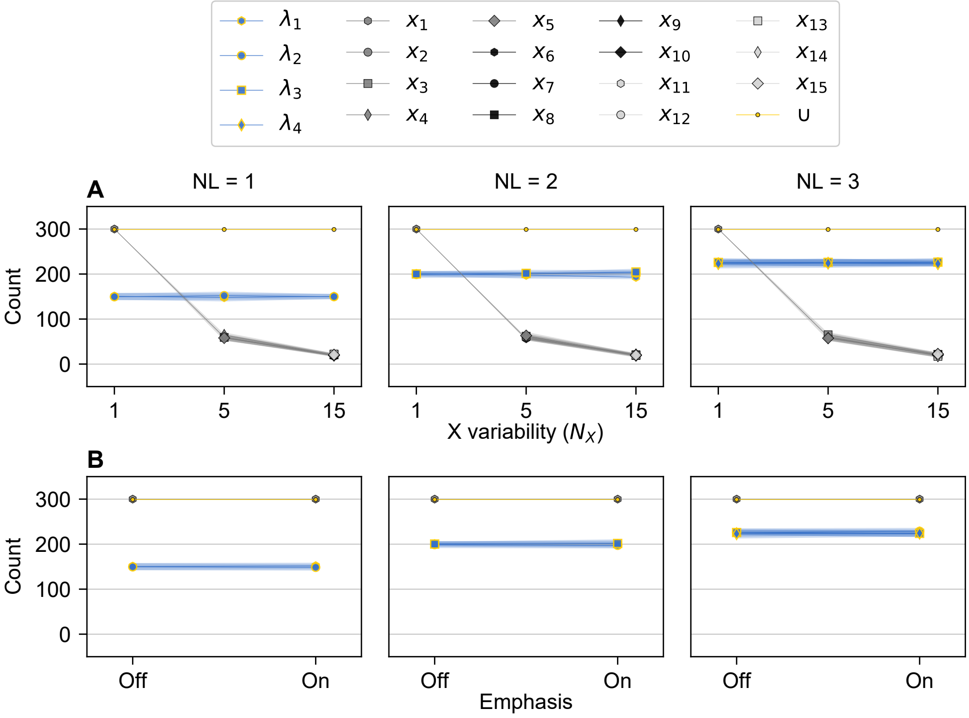


Figure 3.48: Stimulus element counts in all training sequences of experiments without SLMT. Each individual data point (= n) indicates mean count ± standard deviation (SD) of grammatical A and B pairs (λ), non-grammatical X elements (x) and pause elements ($\cup$) across the RNNs (N = 10). **A** Counts in experiment 'X variability' (n = 10); N_X = number of available X elements. **B** Counts in experiment 'Emphasis' (n = 10).

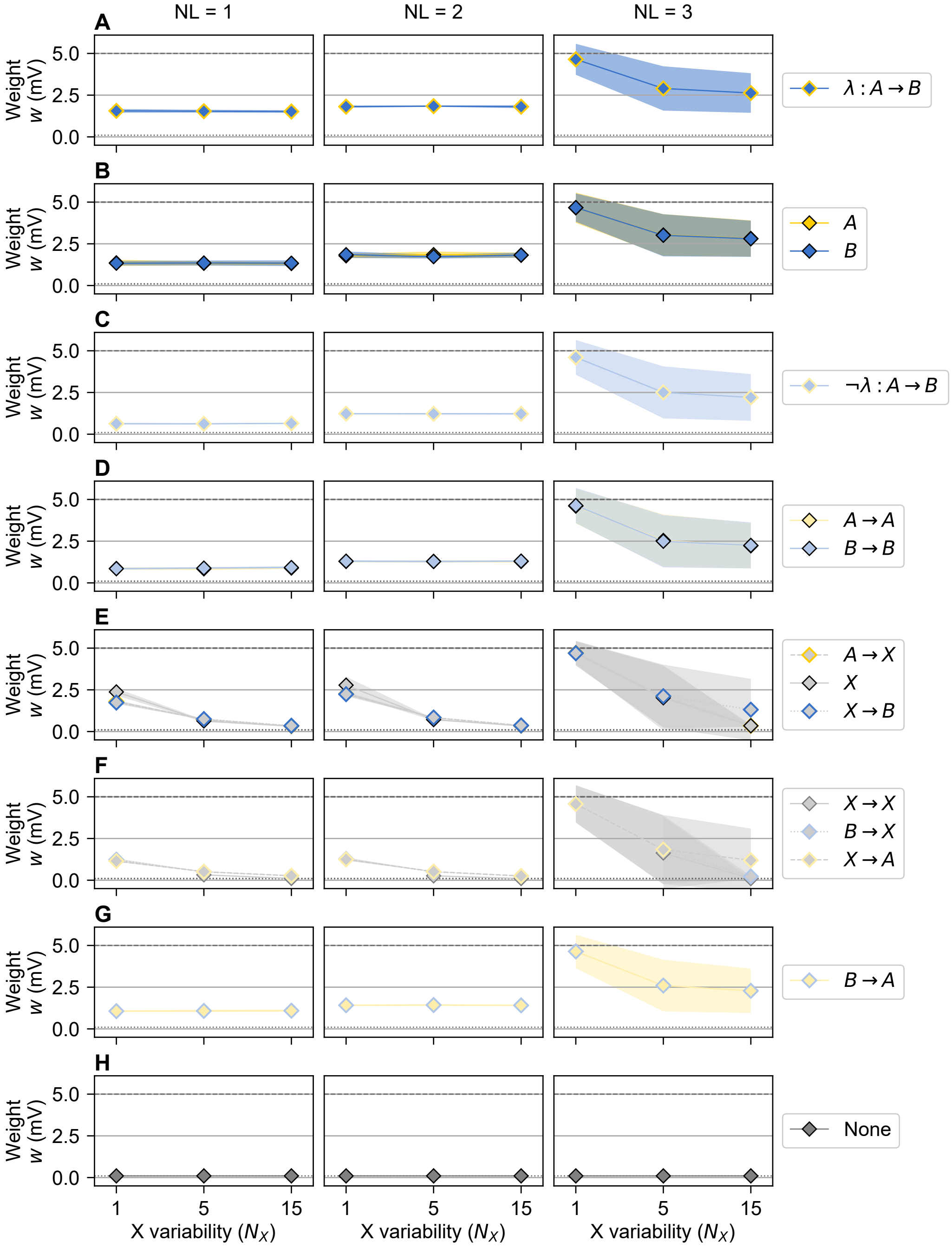


Figure 3.49: Assembly medians (synaptic weight w) in experiment 'X variability' and without SLMT. Each individual data point (= n) indicates mean ± standard deviation (SD) across the RNNs (N = 10). Columns indicate grammar (NL = 1: isolated NAD, NL = 2 and NL = 3: nested NADs). N_X = number of available X elements. Dotted line = lower weight boundary. **A** Grammatical A to B transitions (n = 10). **B** A and B elements (n = 10). **C** Non-grammatical A to B transitions (n = 10). **D** A to A and B to B transitions ($\widehat{AA}\widehat{BB}$, n = 10). **E** X elements, A to X and X to B transitions (χ, n = 10). **F** X to X, B to X and X to A transitions (χ', n = 10). **G** B to A transitions (n = 10). **H** Non-assembly synapses (None, n = 10).

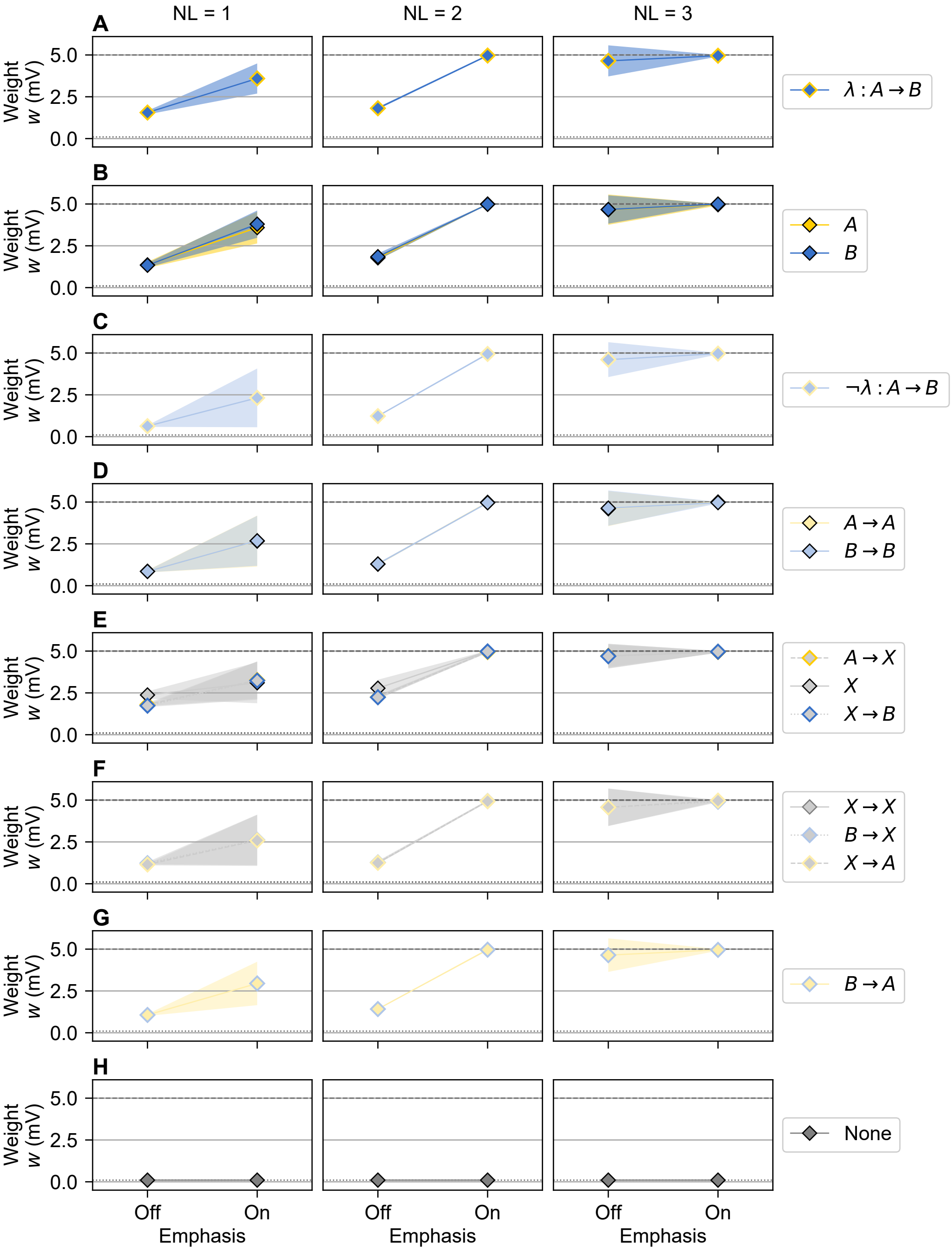


Figure 3.50: Assembly medians (synaptic weight w) in experiment 'Emphasis' and without SLMT. Each individual data point (= n) indicates mean ± standard deviation (SD) across the RNNs (N = 10). Columns indicate grammar (NL = 1: isolated NAD, NL = 2 and NL = 3: nested NADs). N_X = number of available X elements. Dotted line = lower weight boundary. **A** Grammatical A to B transitions (n = 10). **B** A and B elements (n = 10). **C** Non-grammatical A to B transitions (n = 10). **D** A to A and B to B transitions ($\widehat{AA}\widehat{BB}$, n = 10). **E** X elements, A to X and X to B transitions (χ, n = 10). **F** X to X, B to X and X to A transitions (χ', n = 10). **G** B to A transitions (n = 10). **H** Non-assembly synapses (None, n = 10).

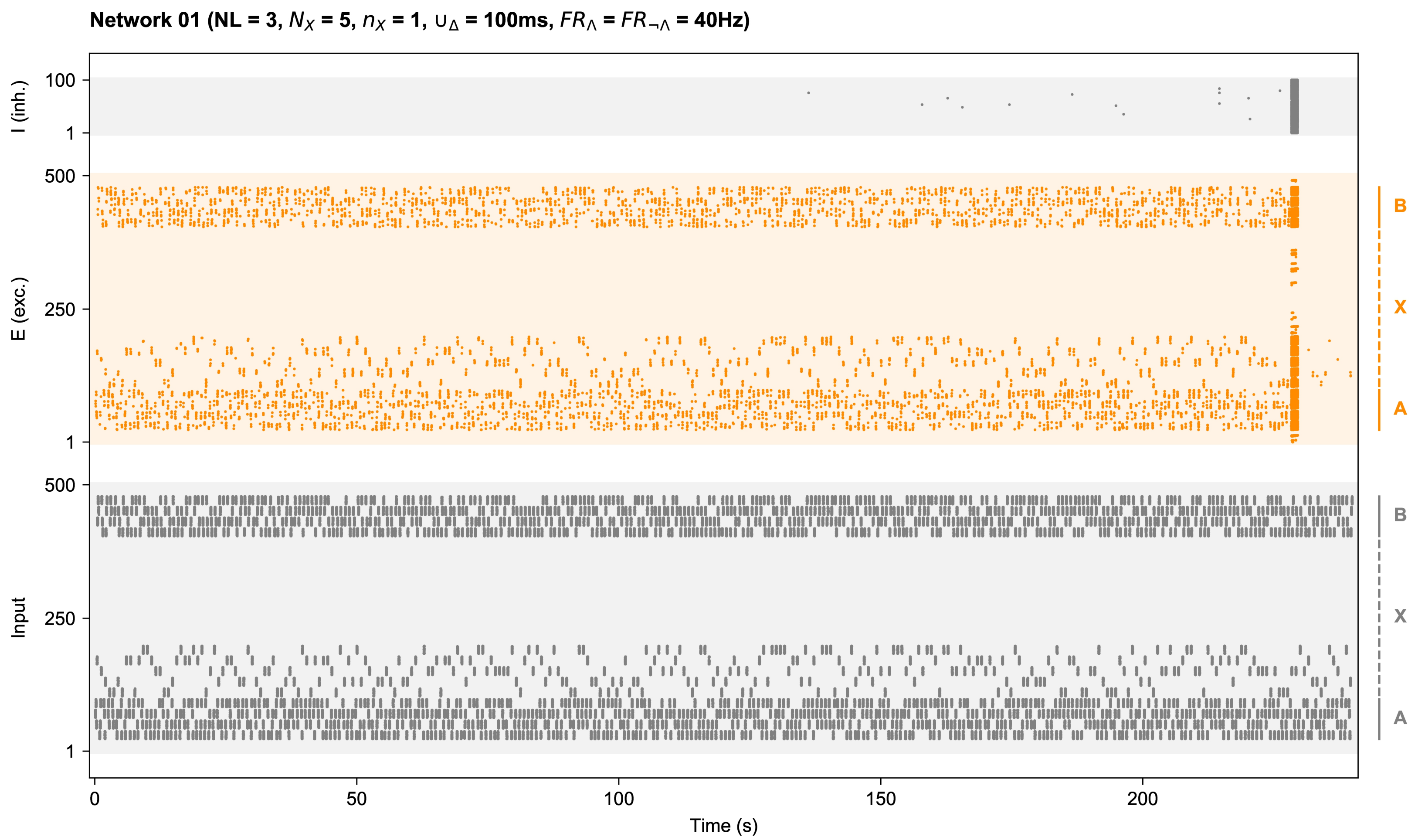


Figure 3.51: Example spiking network activity during training in experiment 'X variability' and without SLMT. NL = nesting level; N_X = number of available X elements; n_X = number of X elements per grammar sample; $\cup_\Delta$ pause duration between grammar samples; FR = Poisson stimulus firing rate; Λ = grammatical neuron subgroups; $\neg\Lambda$ = non-grammatical neuron subgroups; Input = input population; E = excitatory population; I = inhibitory population; A = A element encoding subgroups; X = X element encoding subgroups; B = B element encoding subgroups.

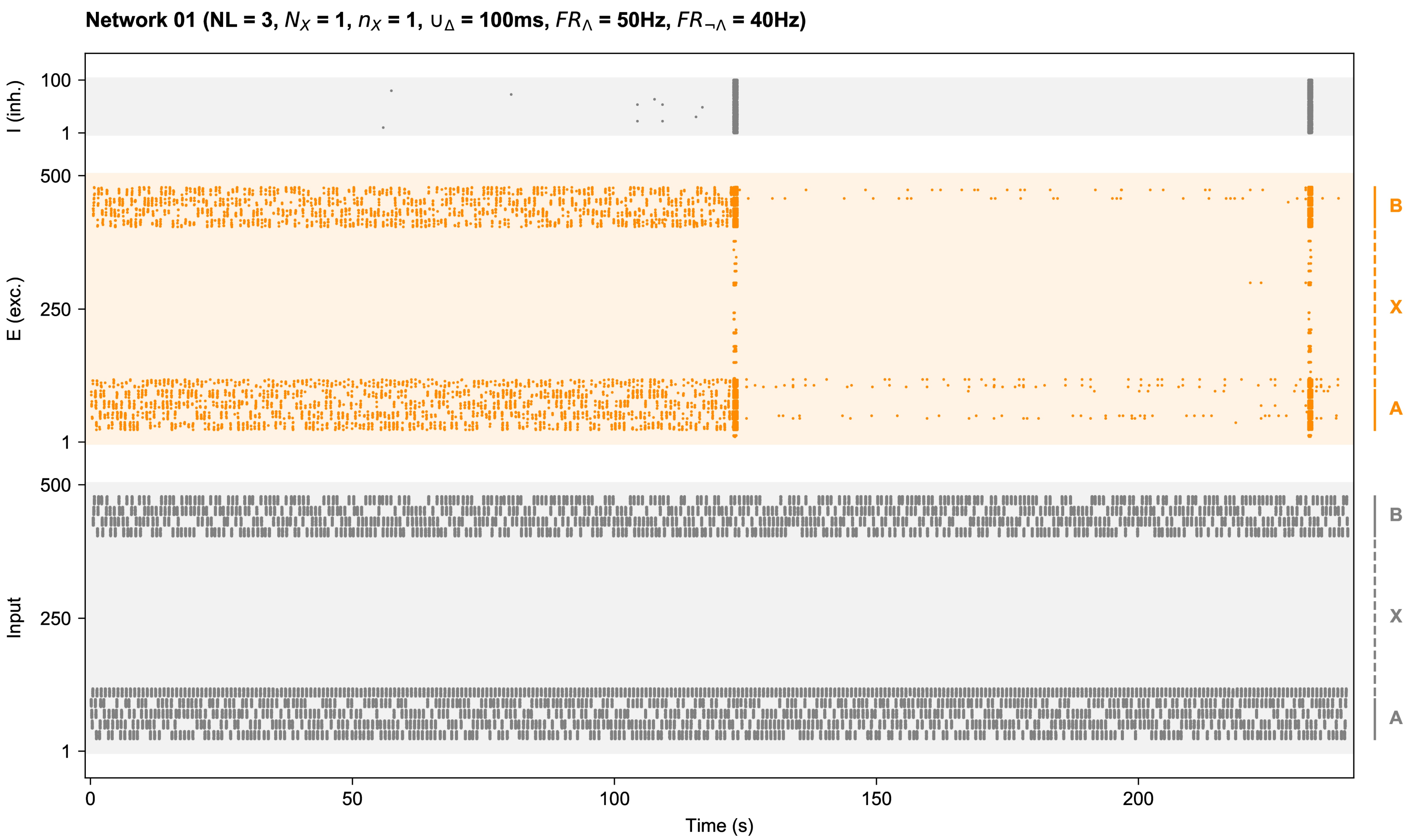


Figure 3.52: Example spiking network activity during training in experiment 'Emphasis' and without SLMT. NL = nesting level; N_X = number of available X elements; n_X = number of X elements per grammar sample; $\cup_\Delta$ pause duration between grammar samples; FR = Poisson stimulus firing rate; Λ = grammatical neuron subgroups; $\neg\Lambda$ = non-grammatical neuron subgroups; Input = input population; E = excitatory population; I = inhibitory population; A = A element encoding subgroups; X = X element encoding subgroups; B = B element encoding subgroups.

Chapter 4

Achievements and future directions

4.1 Overcoming the ‘interface problem’

4.1.1 A link between linguistics and neuroscience

This thesis provided a solution for overcoming the ‘interface problem’ between linguistics and neuroscience (Poeppel and Embick 2005). For the specific problem of nested non-adjacent dependency acquisition, the connecting link between the fields was found across the ‘operations’ of each domain by associative, statistical learning and plasticity (Fig. 4.1, yellow/blue arrow at the bottom). The application of this type of learning was motivated be the observation that it played a central role in both fields. While the acquisition of non-adjacent grammars during passive listening was hypothesised to be supported by unsupervised associative learning (Mueller et al. 2018a), STDP was identified as a central associative learning mechanism in the brain (Dan and Poo 2006; Markram et al. 2011). Importantly, the application of STDP in a recurrent neural network model made it possible to also draw an explanatory link across the ‘elements’ of linguistics and neuroscience (Fig. 4.1, yellow/blue arrow on top). Here, the emerged synapse assemblies were a natural way for encoding the stimulus elements and transitions of nested non-adjacent dependencies. This encoding included that synapse assemblies reflected the accumulated stimulus statistics by their strengths, a hypothesised way of neural networks to encode environmental statistics in general (Bogaerts et al. 2022). Taken together, the ‘interface problem’ framework for linguistics and neuroscience (Poeppel and Embick 2005) provided a very useful scaffold for bridging the gap between these fields. As the ‘interface problem’ is present across all cognitive domains (Pulvermüller et al. 2014), the approach to focus on potential connections between ‘operations’, ‘computations’ (Carandini 2012) or ‘algorithms’ (Love 2015; Griffiths et al. 2015) and ‘elements’ of the domains of interest might be essential for the identification of even more ‘brain-cognition’ interfaces.

4.1.2 Implications for linguistics

Next to establishing an explanatory link between linguistics and neuroscience, the findings of this thesis further have important implications for reconsidering within-

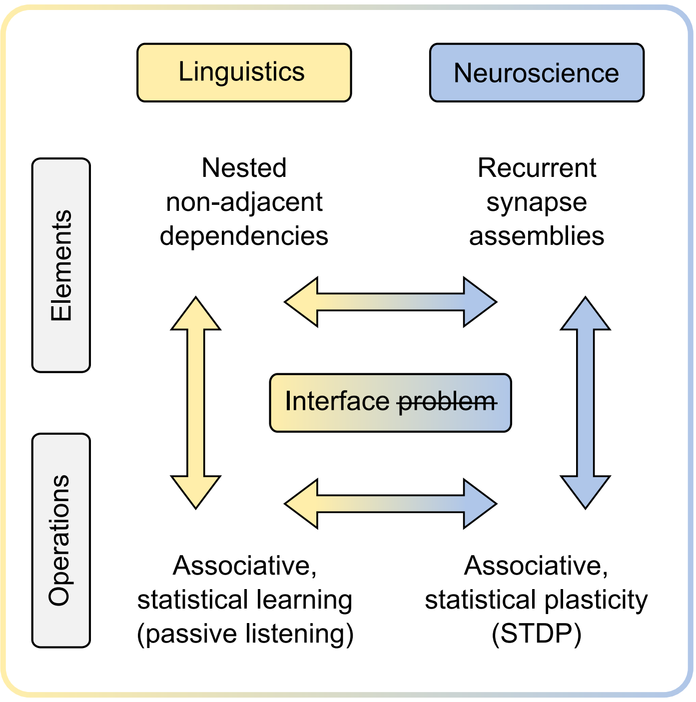


Figure 4.1: Overcoming the 'interface problem' between linguistics and neuroscience for nested non-adjacent dependency acquisition. Arrows indicate explanatory relations of elements and operations within (vertical) and across (horizontal) disciplines. Compare Fig. 7.1. in Poeppel and Embick (2005).

field links between 'operations' and 'elements'. In this thesis, the acquisition of nested NADs was successfully modelled by associative statistical learning with a training procedure where grammars were merely presented to the model, thereby mimicking the passive listening paradigm. In linguistics, the role of statistical, associative learning for grammar and language acquisition (Fig. 4.1, yellow vertical arrow) is however controversially discussed. While statistical learning is widely considered a potential underlying mechanism for language acquisition (Perruchet and Pacton 2006; Romberg and Saffran 2010; Erickson and Thiessen 2015; Thiessen 2017; Saffran and Kirkham 2018; Mueller et al. 2018a), it is also criticised to be insufficient for explaining the mastering of language and its rich hierarchical structures (Chen et al. 2021; Rawski et al. 2021). Moreover, the role of associative, statistical learning is also challenged by the so called 'poverty of the stimulus' argument (Pullum and Scholz 2002; Berwick et al. 2011) and the 'universal grammar' hypothesis (Yang 2004; Lidz and Gagliardi 2015; Friederici et al. 2017). Both lines of thought assume that the raw speech signal does not contain the necessary information for learning language from scratch. Rather, 'innate language knowledge' is seen as a guiding framework during the acquisition process (Lidz and Gagliardi 2015; Shi et al. 2020). However, more differentiated views consider 'universal grammar' as one essential factor next to others, like language experience and domain-general learning mechanisms (Yang et al. 2017). In support of critical views on the 'poverty of the stimulus' argument and the 'universal grammar' hypothesis (Lewis and Elman 2002; Perfors et al. 2006; Dąbrowska 2015; Chater and Christiansen 2018), this thesis contributed towards elucidating further the potential scope of associative, statistical learning. Here, the finding that a purely associative learning mechanism successfully modelled human grammar learning performances under various stimulus conditions was taken as a strong argument for considering it as a central contributor to language acquisition (Lehfeldt et al. n.d.). Moreover, a generic and randomly connected network, i.e. a purely naive architecture, was the substrate of NAD learning. Further, the network only had access to the grammars during sequential presentation and subsequent online processing, so that learning was purely stimulus-driven. Successful NAD learning in such a linguistically unbiased network architecture and with a training procedure resembling the passive-listening paradigm further strengthened the assumption that associative, statistical learning might be an important mechanism especially for early language acquisition phases. Taken together, the major role of associative, statistical learning for language acquisition might be characterised as allowing for the acquisition of statistically relevant structures of the perceived language input and thereby, also allowing for the formation of early knowledge representations.

4.1.3 Implications for neuroscience

The results of this thesis were conform with the idea of assembly-based cognition in neuroscience (Pulvermüller et al. 2014; Papadimitriou et al. 2020; Kossio et al. 2021), i.e. that STDP leads to the emergence of distributed stimulus representations in form of synapse assemblies (Fig. 4.1, blue vertical arrow). Further, this thesis demonstrated that STDP is not only a brain mechanism of basic memory formation problems (Dan and Poo 2006), but was also applicable to higher-order cognitive tasks like human grammar acquisition. This further strengthens STDP as a central learning mechanism of the brain and expands the model's scope of applicability from explaining low-level brain dynamics to high-level cognition. Moreover, a central challenge in the field of neuroscience is to understand how different time scales of neural encoding, learning and behaviour ranging from milliseconds, seconds to minutes and hours combine (Melamed et al. 2004; Drew and Abbott 2006; Setareh et al. 2018; Gerstner et al. 2018; Brzosko et al. 2019; Cone and Shouval 2021). Here, the implemented spiking RNN demonstrated that currently active stimuli can be represented by the instantaneous spiking activity of neurons. Learning, however, took place on slower time scales by the continuously evolving synaptic weights that were updated according to the processed spiking activity. Thus, considering the (spiking) brain as a hybrid system with both very fast and slow variables interacting with each other (Stimberg et al. 2014) might explain how arbitrary time scales are complemented via different brain variables (Kiebel et al. 2008; Castellano and Pipa 2013). In this thesis, temporal STDP windows for the induction of potentiation and depression were further examples for such an interaction on different time scales. While the temporal windows were defined by the decay of phenomenological memory traces taking place on comparably slow time scales, instantaneous spikes occurring within theses windows triggered plasticity changes. The time constants applied in this thesis (i.e. $\tau_P = 700$ ms and $\tau_D = 100$ ms), however, exceeded the electrophysiologically measured time windows in rats (Markram et al. 1997; Bi and

Poo 1998; Debanne et al. 1998; Bi and Poo 1999; Feldman 2000; Sjöström et al. 2001; Froemke and Dan 2002) which are usually applied accordingly in STDP models (Miner and Triesch 2016; Klos et al. 2018). Though this might be viewed as a critical deviation from the biological evidence, the rationale behind this extension of time constants was to explore if the core STDP algorithm was applicable to learning problems taking place on larger time scales (Testa-Silva et al. 2010). As the evidence of temporal STDP windows in humans is scarce (Koch et al. 2013), testing parameter spaces 'fitted' to the time scales of higher-level cognitive learning tasks currently remains a practicable solution for this endeavour. Next to the interaction of slow and fast brain variables, also memory itself is expressed at different time scales ranging from short- to long-term memory (Zucker and Regehr 2002; Tetzlaff et al. 2012; Benna and Fusi 2016). In this thesis, time scales of memory were expressed by an activity dependent weight decay, i.e. a short- to long-term memory transition (SLMT) mechanism (Lehfeldt et al. n.d.). Here, memory on different time scales was modelled in a conceptual fashion, implementing its transition according to the STDP learning protocol equivalently acting on a plastic weight decay. On the one hand, SLMT served as a practicable solution in order to prevent unstable runaway dynamics during grammar learning experiments. However, a further motivation was to include the concept of forgetting into the learning model (Richards and Frankland 2017). Studies implementing different forms of (adaptive) weight decay already acknowledged forgetting as a beneficial property of learning models (Panda and Roy 2017; Panda et al. 2018; Harrison et al. 2020; Endress and Johnson 2021). Throughout, a main contribution of forgetting was interpreted as de-noising the formed memory (Panda et al. 2018), i.e. as reducing the impact of outdated information (Hintzman and Ludlam 1980; Servan-Schreiber and Anderson 1990; Hintzman 1984; Yalnizyan-Carson and Richards 2022). Taken together, the ability of STDP to detect invariant structure in the input (Sprekeler et al. 2007) apparently can further be supported by memory decay or forgetting. Also in this thesis, this beneficial effect was observed as learning with SLMT led to strong and persistent grammar encodings while non-grammar encodings were comparably weaker and less persistent.

4.1.4 Revisiting the 'low-high' and 'subsymbolic-symbolic' debates

The finding that a low-level brain mechanism of associative, statistical learning implemented in a distributed neural network was successful in forming meaningful grammar representations made revisiting the 'low-high' and 'subsymbolic-symbolic' debates very intriguing. When integrating the results of this thesis into the previously identified 'low-high' and 'subsymbolic-symbolic' scales (Fig. 1.1), a shift of high-level human NAD learning performances from the symbolic into the subsymbolic domain became possible (Fig. 4.2). This was motivated by the fact that low-level and subsymbolic grammar representations, expressed via recurrent synapse assembly weight distributions, already matched human NAD learning performances. Apparently, the low-level and subsymbolic STDP mechanism was sufficient to achieve a cognitively meaningful encoding of NADs. This, in turn, further

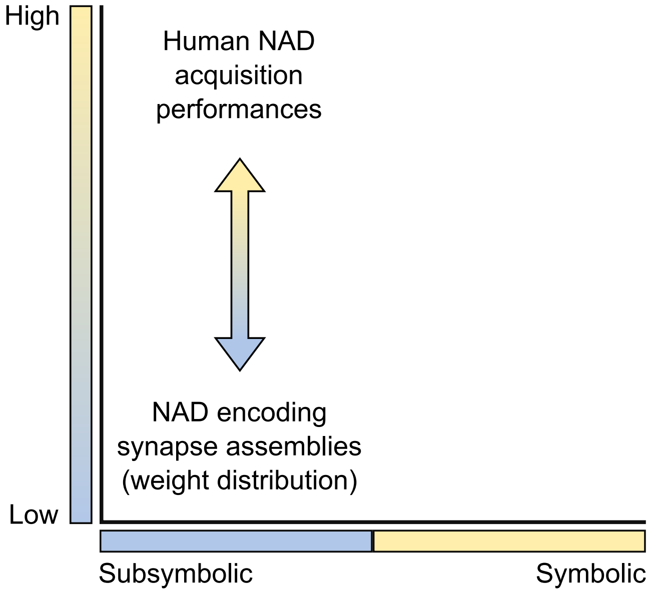
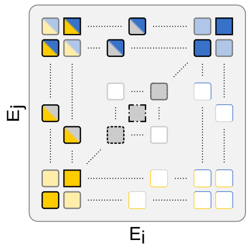
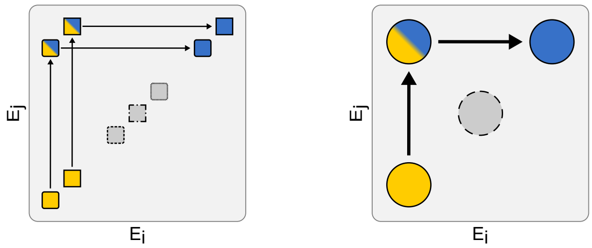


Figure 4.2: Integration of grammar learning experiment results into 'low-high' and 'subsymbolic-symbolic' scales. Low-level and subsymbolic NAD encoding synapse assemblies (i.e. their weight distribution) complied with high-level human NAD acquisition performances. This motivated a potential shift to the subsymbolic/high domain.

led to the possibility that higher-order grammar learning performances are not necessarily fully based on purely symbolic encodings and operations. Even though the nature of additional operation mechanisms impacting the human learning outcome remains elusive, the modelling results of this thesis are at least an indicator that such high-level cognitive skills might as well be performed in the subsymbolic domain. Further, subsymbolic knowledge representations acquired via spike-based statistical learning in the recurrent synapses might represent a potential basis for further linguistic operations (Fig. 4.3(a)). This thesis can be considered as modelling an early stage during language learning that primarily was concerned with the acquisition of distributional statistical knowledge. However, later phases of the language acquisition process possibly also include an evaluation of the structural and temporal composition of grammars, i.e. an analysis of the nesting levels. This might require that, a priori, correct pair associations between grammatical A and B elements have been learned. Further, a nesting level analysis probably also relies on knowledge about the correct temporal within-pair order, i.e. that a B element will follow after experiencing its grammatical A element. In this regard, the grammatical A to B transition assemblies (i.e. $\lambda : A \rightarrow B$) encoded both the correct pair associations of grammatical A and B elements as well as their temporal within-pair orders (Fig. 4.3(b)). A further linguistic operation might deal with the extraction of the general tendency of experienced stimuli, i.e. a generalisation of rules. This might require that, a priori, a sufficiently high number of individual grammar samples following a



(a) NAD encoding synapse assemblies



(b) Template for nesting level analysis?

(c) Template for generalisation?

Figure 4.3: NAD encoding synapse assemblies as templates for further linguistic operations? (a) NAD encoding synapse assemblies: strengths encode the accumulated statistics of all stimulus elements and transitions. (b) Potential template for nesting level analysis: $\lambda : A \to B$ encode correct associations and temporal within-pair orders. (c) Potential template for generalisation: extraction of general tendency from all individual NAD samples.

shared rule has been experienced and learned. Further, some abstraction or summation mechanism, or probably a higher-order brain region, could subsequently extract the general tendency of the individually learned examples and derive a generalised representation (Fig. 4.3(c)). Here, the recurrent synapse assemblies encoded the strengths of all individual element pairs and transitions, such that A to B transition assemblies (i.e. $\lambda : A \to B$) might be detected as a strong and generally persistent structure. Thus, the learned synapse representations might be viewed as an informative template that a potential generalisation mechanism could rely upon. That low-level knowledge representations might support subsequent analysis steps of higher-order cognitive skills is conform with the idea of an 'integrated' view on cognition where subsymbolic representations underlie symbolic ones (Kelley 2003). However, understanding how symbolic cognitive computations arise from neural operations is still an ongoing process (Smolensky 2012; Alhama and Zuidema 2019; Mastrogiuseppe et al. 2022). Still, the findings of this thesis support the notion that higher-level brain regions might compute over low-level sensory representations (Näätänen et al. 2001) and further, that (statistical) language acquisition might rely on the extraction of distributed cues and subsequent, potentially more symbolic-like, integration processes (Endress and Bonatti 2007; Thiessen et al. 2013; Erickson and Thiessen 2015).

4.2 Possible future directions

4.2.1 Including the 'perception-cognition' interface

Cognition is tightly interlinked with the representations formed at the sensory periphery and therefore with perception (Mesulam 1998; Näätänen and Winkler 1999; Munkong and Juang 2008; Wolff et al. 2021). The importance of low-level brain mechanisms becomes especially apparent in auditory cortex, where 'primitive' forms of cognition are hypothesised to exist as reflected by pre-attentive forms of sensory input transformations (Näätänen et al. 2001). Further, language acquisition performance relies on auditory sensory processes (Mueller et al. 2012). Moreover, with regard to grammar learning, the stimulus properties of the speech signal itself have been shown to influence the learnability (Nespor et al. 2003; Bonatti et al. 2005; Toro et al. 2008a; Toro et al. 2008b; Fogerty and Humes 2012), though see Weyers and Mueller (2022). However, as the link between perceptual and cognitive processes is not fully understood, the importance to include sensory encoding schemes in the design of cognitive (language learning) models becomes highlighted. In this context, the distributed encoding scheme of the RNN model of this thesis opens the possibility to approach this problem further by studying language learning based on auditory spike representations (Fontaine et al. 2011). To illustrate, spiking training sequences implemented grammar feature variations ranging from the temporal structure, occurrence frequency to stimulus strengths in a low-level and distributed manner. However, in order to understand the 'perception-cognition' interface better, ecologically more informative stimuli could be used. In particular, the applied Poisson spike patterns can be viewed as ecologically 'meaningless' as they primarily encoded the strength of a stimulus reflected by the firing rate. In contrast to these Poisson spike trains, spatio-temporal spike representations derived from sound or speech signals might be more informative stimuli (Holmberg and Hemmert 2004; Holmberg et al. 2005; Liu et al. 2010; Klampfl and Maass 2013; Yu et al. 2019). These are also referred to as 'spikegrams' (Zai et al. 2015) and have the great potential to extend training stimuli by essential sound or speech signal properties, as for example temporal information (Purwins et al. 2000; Rankin and Rinzel 2019). In summary, cognitive models that include subsymbolic and distributed encoding schemes are candidates to implement details about sensory stimulus representations in their inputs. Thus, they have the potential to advance understanding the formation of meaningful percepts and to elucidate further the influence of low-level brain mechanisms on higher-order cognitive skills.

4.2.2 Precursor linguistic abilities across species?

Despite the uniqueness of the human language faculty (Berwick et al. 2013), the discovery of shared processing or learning principles across humans and animal species is a promising approach for identifying low-level brain mechanisms that possibly support high-level cognitive skills such as language (van Horik and Emery 2011; Steinschneider et al. 2013; Lattenkamp and Vernes 2018; Levelt 2018; Corballis 2020; Suzuki 2021). A bandwidth of potential evidence about precursor linguistic abilities in non-human animals exists. For example, rodents showed an ability to discriminate syllables based on voicing contrast (Kuhl and Miller 1975). Further, rats could discriminate languages (Toro et al. 2003), segment speech (Toro and Trobalón 2005), generalise acoustic patterns (Toro and Hoeschele 2017) and learn acoustic rules (de la Mora and Toro 2013). Moreover, in birds, visual artificial grammar learning (Stobbe et al. 2012) and auditory syntactic rule discrimination (Abe and Watanabe 2011) has been shown. In primates, finite state grammar learning (Fitch and Hauser 2004), adjacent dependency learning (Englund et al. 2022), but also NAD processing (Watson et al. 2020), associative learning of center embedded grammars (Rey et al. 2012) and the production of embedded movement sequences (Jiang et al. 2018) has been demonstrated. However, potential evidence about preliminary forms of linguistic proficiency in non-human animals is hotly debated (Beckers et al. 2012; Poletiek et al. 2016; Rawski et al. 2021; Watson et al. 2021). Mostly, the design of comparative language acquisition studies is considered inadequately for drawing far-reaching conclusions such as the existence of precursor linguistic abilities in animals. For example, the discrepancy between non-adjacent grammar stimuli and human syntax (Rawski et al. 2021) or between animal sequence production and human language production (Poletiek et al. 2016) is highlighted. Further, claims about precursor linguistic abilities are viewed as premature because reported learning abilities do not compare to the compositionality of human language (Beckers et al. 2012). Next to critically evaluating the findings of comparative experimental studies, computational modelling can contribute towards the identification of potentially shared processing principles (Yildiz et al. 2013) or organisational features of animal vocalisations and human language (Kershenbaum et al. 2014; Sainburg et al. 2019). Here, especially the general ability to process long-range dependencies is currently investigated as an important precursor ability for language (Sainburg et al. 2022). Moreover, the performance of possibly underlying learning mechanisms can be studied, such as done in this thesis by testing domain-general, associative learning. Notably, the STDP learning model was originally derived based on electrophysiological evidence in rats (Markram et al. 1997; Bi and Poo 1998; Debanne et al. 1998; Bi and Poo 1999; Feldman 2000; Sjöström et al. 2001; Froemke and Dan 2002). Despite being a non-human learning algorithm, its success in modelling a variety of human-specific grammar acquisition patterns was interpreted as an indicator that STDP might not only represent an important low-level prerequisite for language but is presumably also shared across species (Lehfeldt et al. n.d.). Further, the applicability of STDP for modelling the acquisition of human grammatical structure on a low-level conforms with the idea that language acquisition might rely on both evolutionary 'old' and more recent cognitive strategies (Toro 2016). Especially the proficiency in serial analysis is viewed as one important skill for language, next to segmentation, categorisation and abstraction (Kiggins et al. 2012), whereby STDP might be a potential candidate for such sequence processing and learning abilities. Taken together, the results of this thesis showed that neural modelling studies can contribute new insights into the linguistic capacities of low-level (and non-human) brain mechanisms. The insights gained here might be worthy of discussion in comparative lines of research dedicated to the evolutionary origins of language (Hauser et al. 2002; Sherwood et al. 2008; ten Cate and Okanoya 2012; Kiggins et al. 2012; Bowling and Fitch 2015; Wilson et al. 2018; Mueller et al. 2018b).

4.2.3 A possible template for neuromorphic applications

Software models of spiking neural networks serve as valuable templates for developing hardware implementations of brain circuits and function (Zenke et al. 2021; Guo et al. 2021), i.e. so called neuromorphic models (Mead 1990; Douglas et al. 1995b). The aim to develop computing devices that make use of the information processing and learning principles of biological brains has gained increasing attention over the last years (Indiveri and Horiuchi 2011; Poon and Zhou 2011; Monroe 2014; Hamilton et al. 2014; Furber 2016; Knight et al. 2016; Thakur et al. 2018; Neftci 2018; Strukov et al. 2019; Roy et al. 2019; Kendall and Kumar 2020; Zhang et al. 2020b; Ielmini et al. 2021; Kim et al. 2021; Bartolozzi et al. 2022; Mehonic and Kenyon 2022; Venkatesan and Williams 2022). So far, the field of neuromorphic engineering has brought fourth a variety of spike-based hardware applications based on principles of sensory processing ranging from the visual (Mead and Mahowald 1988; Chen et al. 2020; Gallego et al. 2022), auditory (Liu et al. 2010; Jiménez-Fernández et al. 2017), olfactory (Jürgensen et al. 2021) to the tactile domain (Rongala et al. 2017; Osborn et al. 2018; Sengupta et al. 2022). But also spike-based learning systems have been developed giving rise to the possibility of implementing cognitive skills in neuromorphic hardware (Indiveri et al. 2009; Neftci et al. 2013; Indiveri and Douglas 2014; Chicca et al. 2014). Here, for example, learning applications range from pattern classification (Mitra et al. 2009; Schmuker et al. 2014; Patiño-Saucedo et al. 2020; Kugele et al. 2020), pattern learning and recall (Milo et al. 2018), pattern learning and retrieval (Fang et al. 2022) to prediction (Debat et al. 2021) and decision making (Corradi et al. 2015; Liang and Indiveri 2019; Pedretti et al. 2020). As neuromorphic systems rely on event-based neural and synaptic dynamics, exploring the possibility to apply the developed training stimuli and implemented learning model of this thesis in such hardware applications might be a promising approach for extending state-of-the-art neuromorophic systems by grammar learning performances. Taken together, studying cognitive computation in spike-based models not only contributes towards a better understanding of its low-level implementation in the neural substrate. It also contributes towards the potential transfer of modelled cognitive skills into computing devices that are based on brain principles.

4.3 Conclusion

4.3.1 Diverse strategies for navigating a diverse world

A large variety of theories about how the brain ‘makes sense’ of the world, learns its principles and rules, adapts to its constraints or foresees future scenarios has been developed in the field of cognitive neuroscience. While this thesis laid a special focus on associative, statistical learning and its potential low-level implementation in the neural substrate of the brain, it is clear that a huge bandwidth of complementary learning principles and strategies exists in order to understand and, ultimately, to navigate the world. Thus, beyond ‘associative’ (Hebb 1949), ‘unsupervised’ (Barlow 1989) or ‘statistical’ (Thiessen 2020; Sherman et al. 2020) learning, additional classes of learning range from ‘reinforcement learning’ (Kaelbling et al. 1996; Dayan and Niv

2008; Niv 2009; Arulkumaran et al. 2017; Neftci and Averbeck 2019), 'supervised learning' (Knudsen 1994; Raymond and Medina 2018), to 'predictive coding' (Bubic et al. 2010; de-Wit et al. 2010; Huang and Rao 2011; Clark 2013; Hull 2020) and 'free energy' principles (Friston 2005; Friston 2009; Friston 2010; Friston et al. 2018). Further, action-based learning and exploration strategies are described by theories of 'embodied cognition' (Anderson 2003; Adams 2010; Engel et al. 2013; Foglia and Wilson 2013; Wilson and Golonka 2013; Badcock et al. 2019) and 'curiosity-driven cognition' (Gottlieb et al. 2013; Oudeyer and Smith 2016; Gottlieb and Oudeyer 2018; Schwartenbeck et al. 2019; Ten et al. 2021). This (certainly still incomplete) collection of learning classes highlights the rich repertoire of cognitive mechanisms and strategies that the brain (and active agents) can rely upon when faced with a task or problem to solve. However, theses mechanisms potentially do not work in perfect isolation but rather interact in dynamic ways. Therefore, a major challenge for future research is to understand how these different learning mechanisms and strategies work together (Orpella et al. 2021; Lang et al. 2022) in order to give rise to intelligent behaviour and cognition. Taking unsupervised, reinforcement and supervised learning schemes as an example, they are hypothesised to be predominant mechanism of cortex, basal ganglia and cerebellum, respectively (Doya 1999; Doya 2000). However, to which degree these isolated learning types interact and support each other is still subject of recent investigations (Caligiore et al. 2019; Badakva et al. 2020). Moreover, it remains to be shown if they are activated in a demand-driven way, i.e. corresponding to the encountered learning scenario that either provides no feedback, reinforcing signals or teacher cues that might be exploited. But finally, also understanding better to which extent (cognitive) behaviours are pre-defined genetically (Zador 2019) or by evolutionary processes (Hasson et al. 2020) will contribute towards a holistic picture of the underlying mechanisms of cognition. Taken together, a full understanding of cognition certainly requires an integration across a variety of scales, from low- to high levels of analysis, subsymbolic to symbolic encoding or processing schemes but also comparatively across species. Moreover, by considering the diversity of tasks encountered in the world, a diverse set of cognitive mechanisms needs to be understood in relation to each other (Wörgötter and Porr 2005). By demonstrating the importance of low-level brain mechanisms for high-level cognitive skills, the present thesis contributed to this endeavour and furthermore, provided a missing piece to the great 'brain-cognition' interface puzzle. To which extent this contribution gives rise to the discovery of more connecting links between brain and cognition remains to be explored in the future.

Appendix A

Complete collection of separability measures: figures

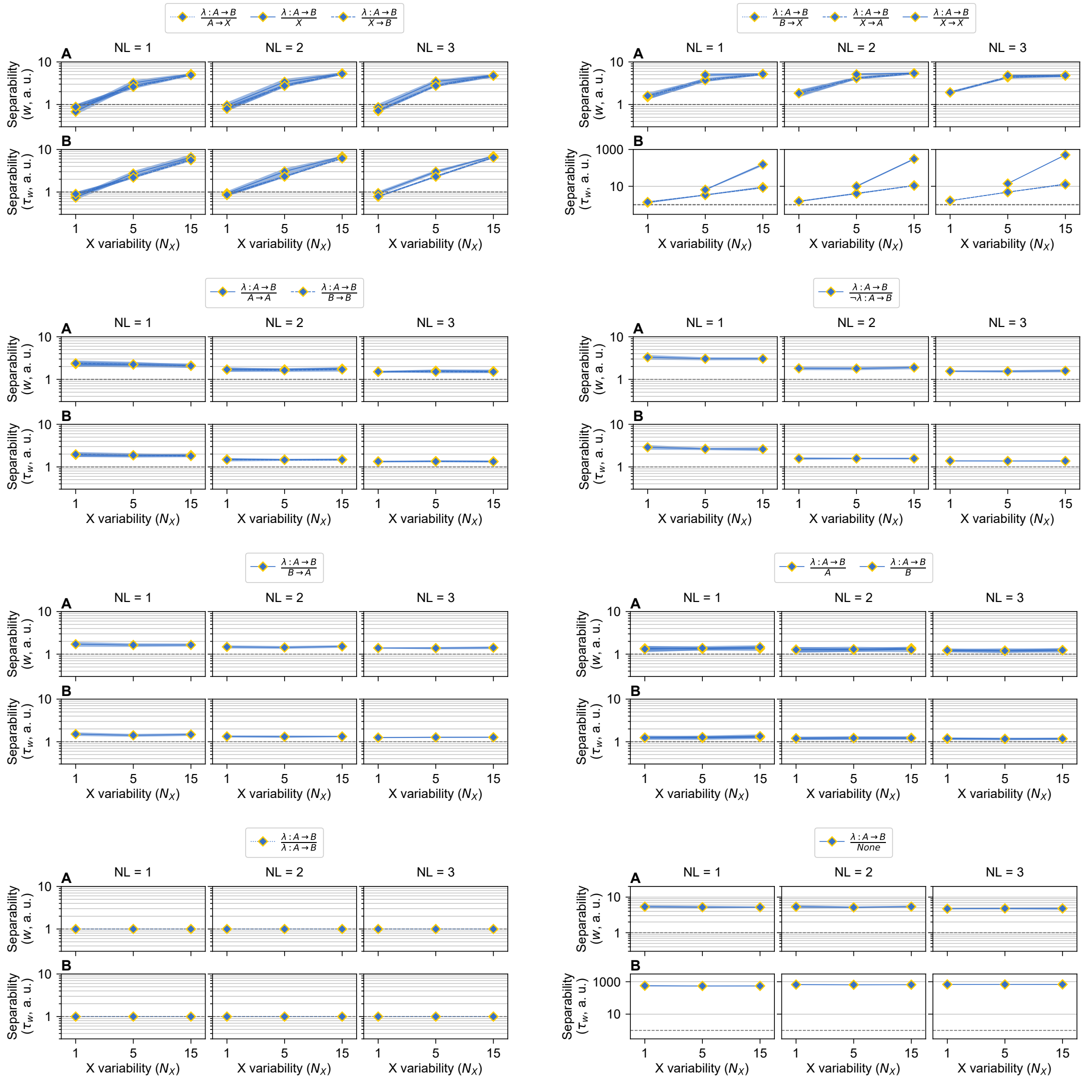


Figure A.1: Separability measures of w and τ_w in the experiment 'X variability' (reference $\lambda : A \to B$).

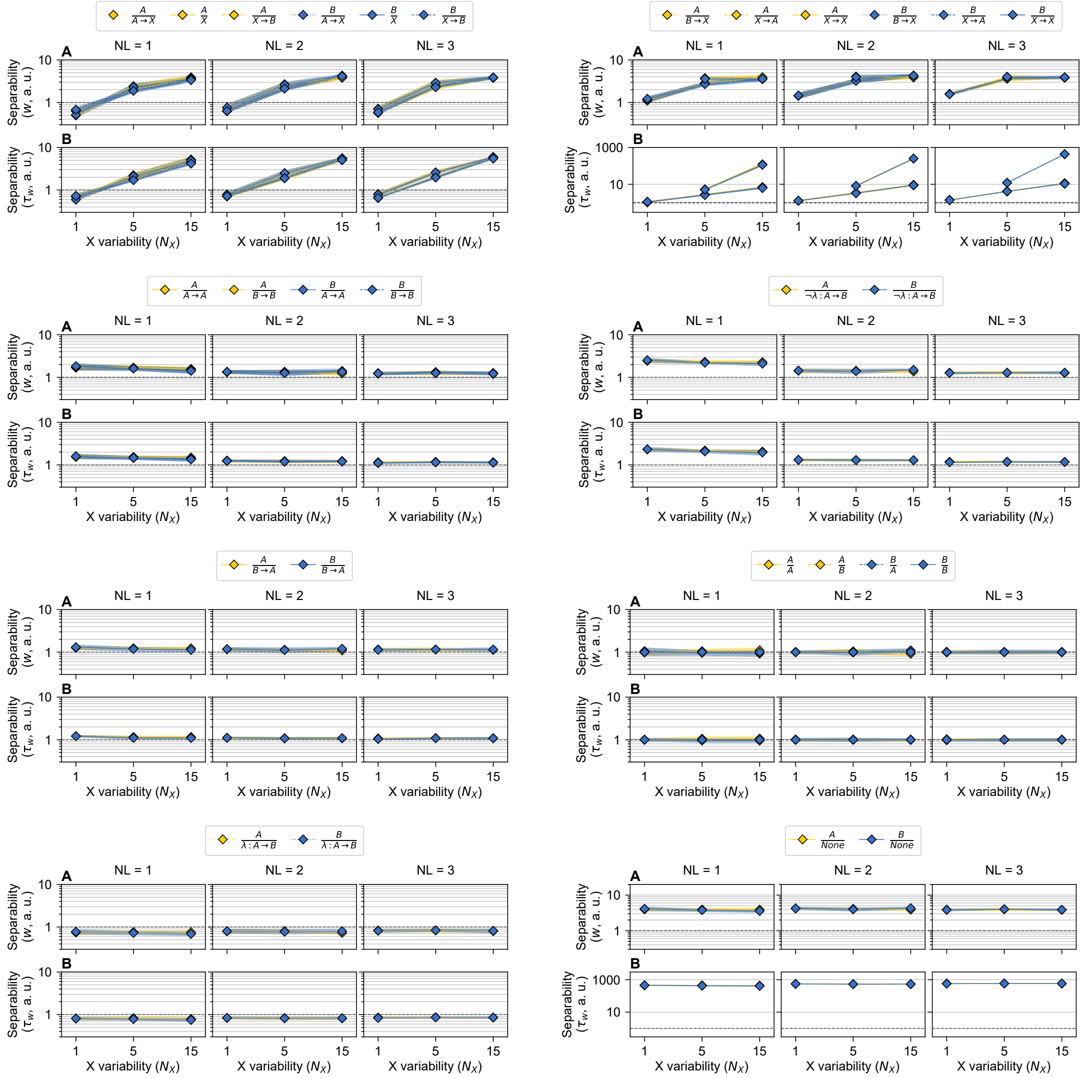


Figure A.2: Separability measures of w and τ_w in the experiment 'X variability' (references A and B).

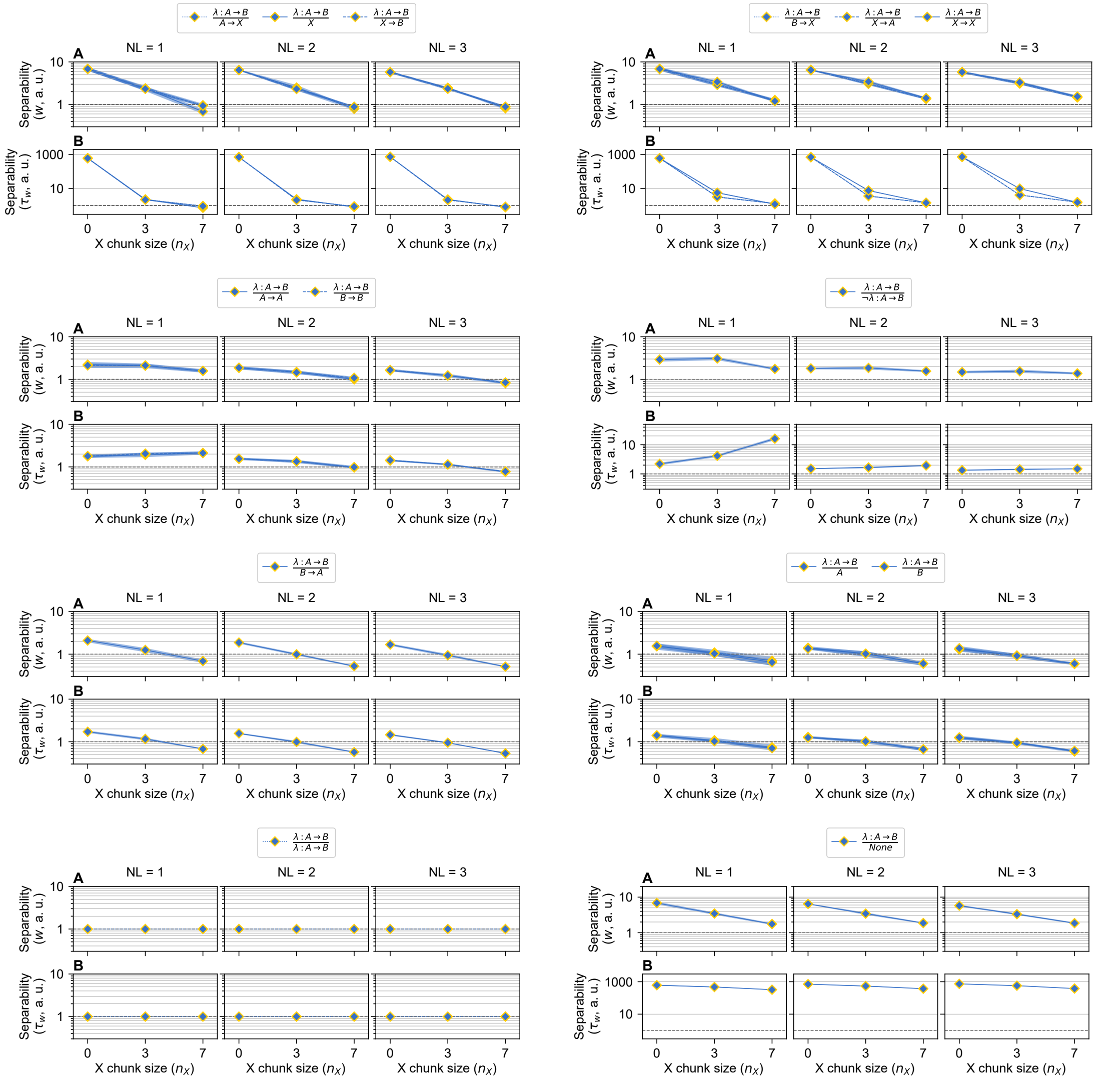


Figure A.3: Separability measures of w and τ_w in the experiment 'X chunk size' (reference $\lambda : A \to B$).

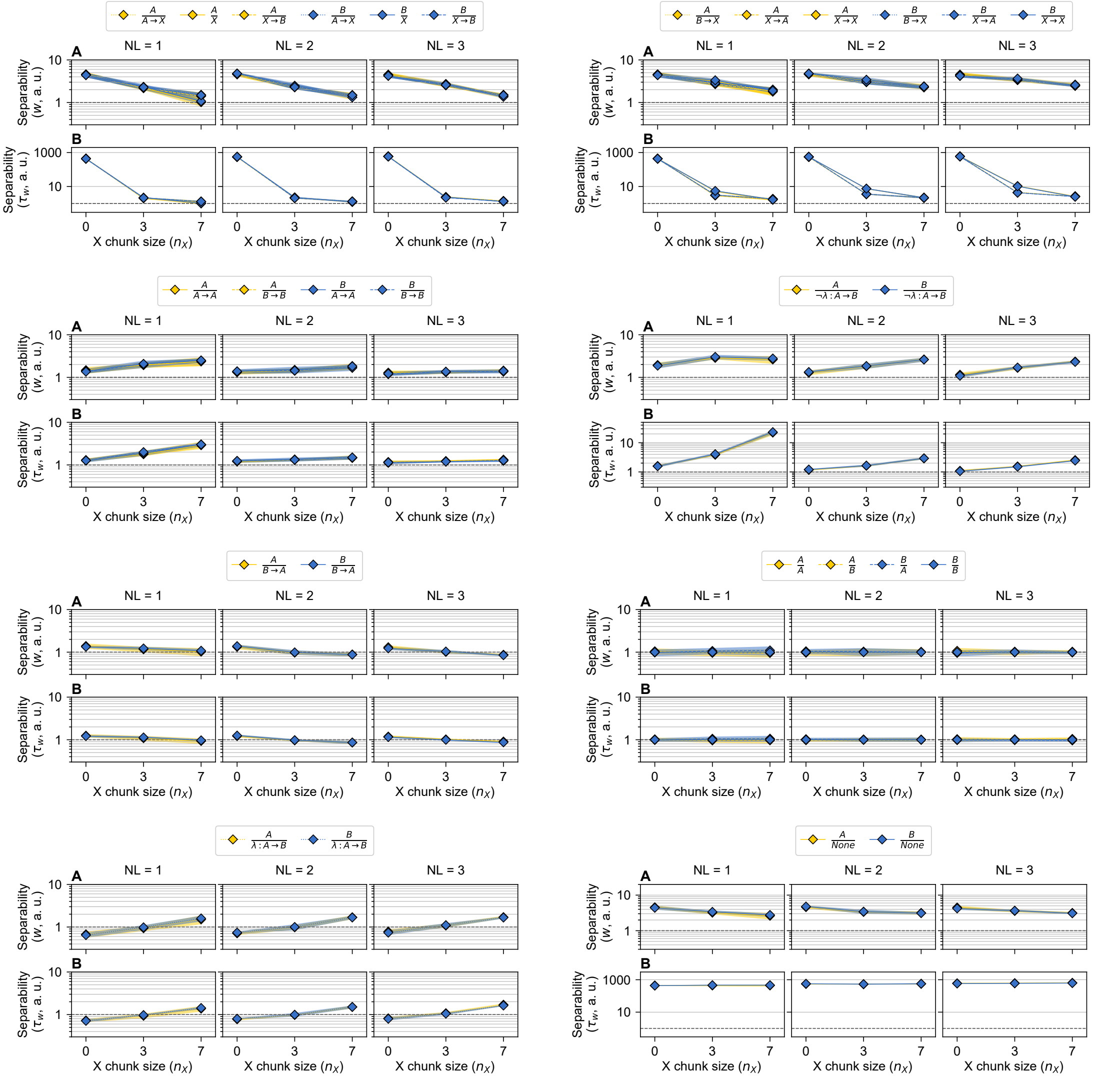


Figure A.4: Separability measures of w and τ_w in the experiment 'X chunk size' (references A and B).

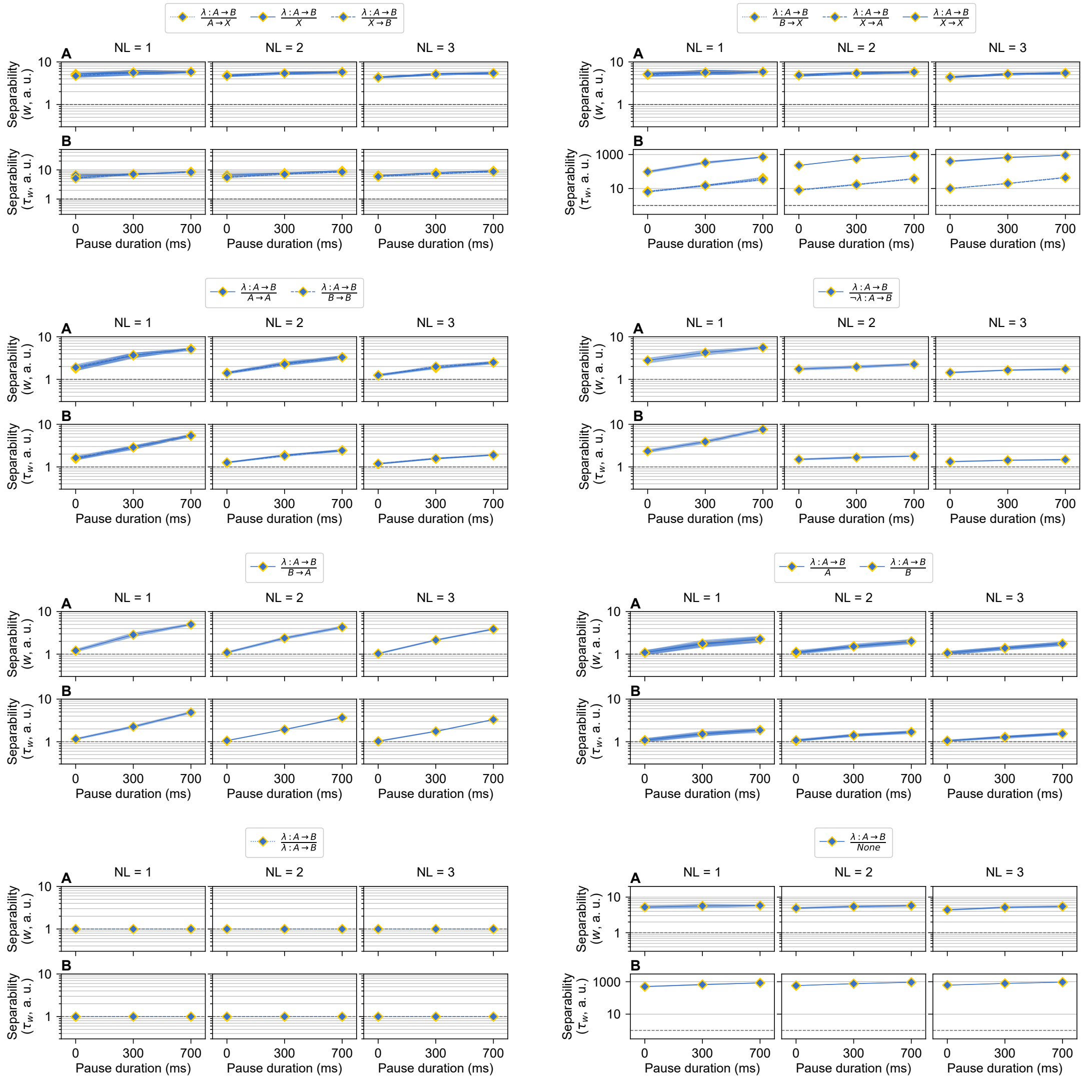


Figure A.5: Separability measures of w and τ_w in the experiment 'Pause variability' (reference $\lambda : A \to B$).

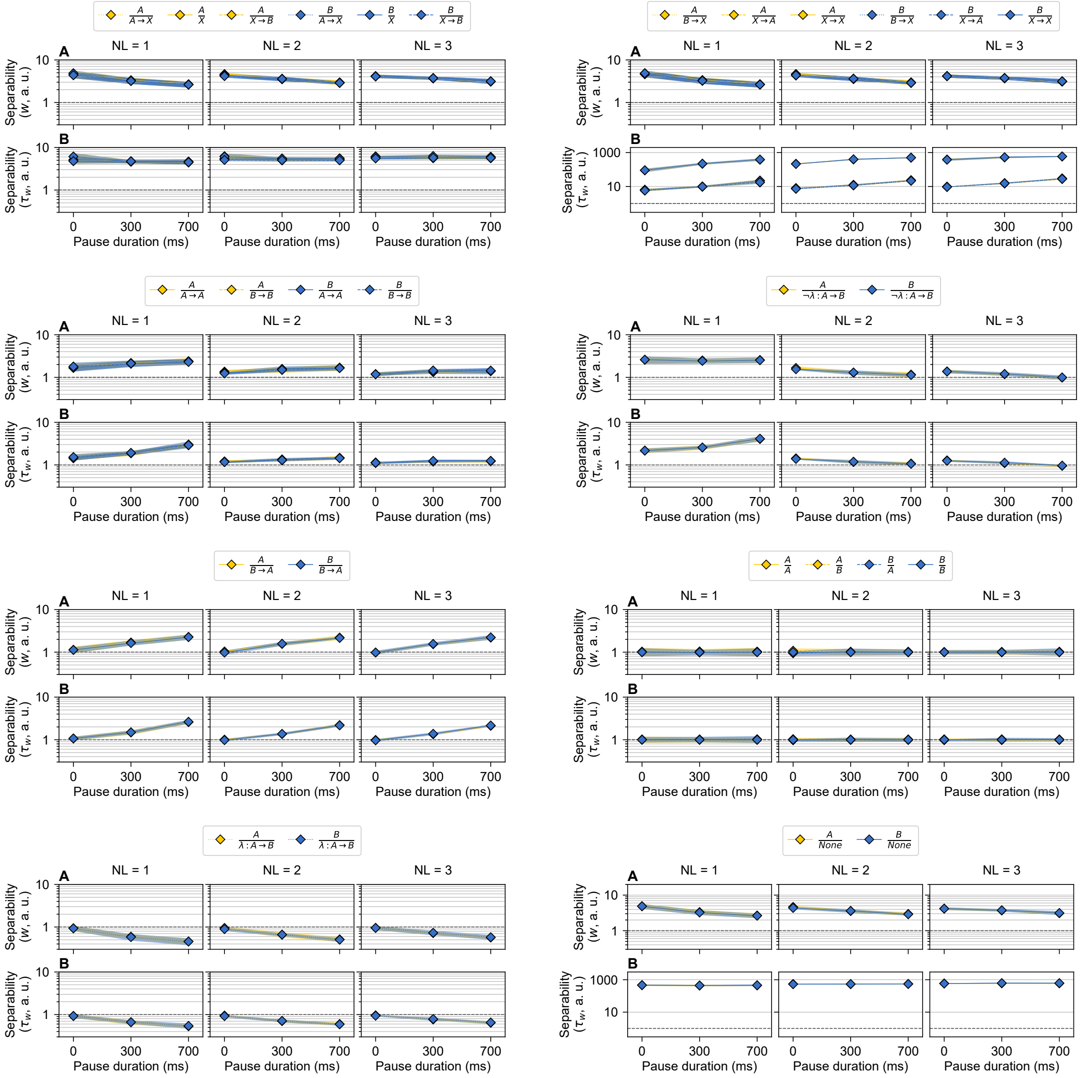


Figure A.6: Separability measures of w and τ_w in the experiment 'Pause variability' (references A and B).

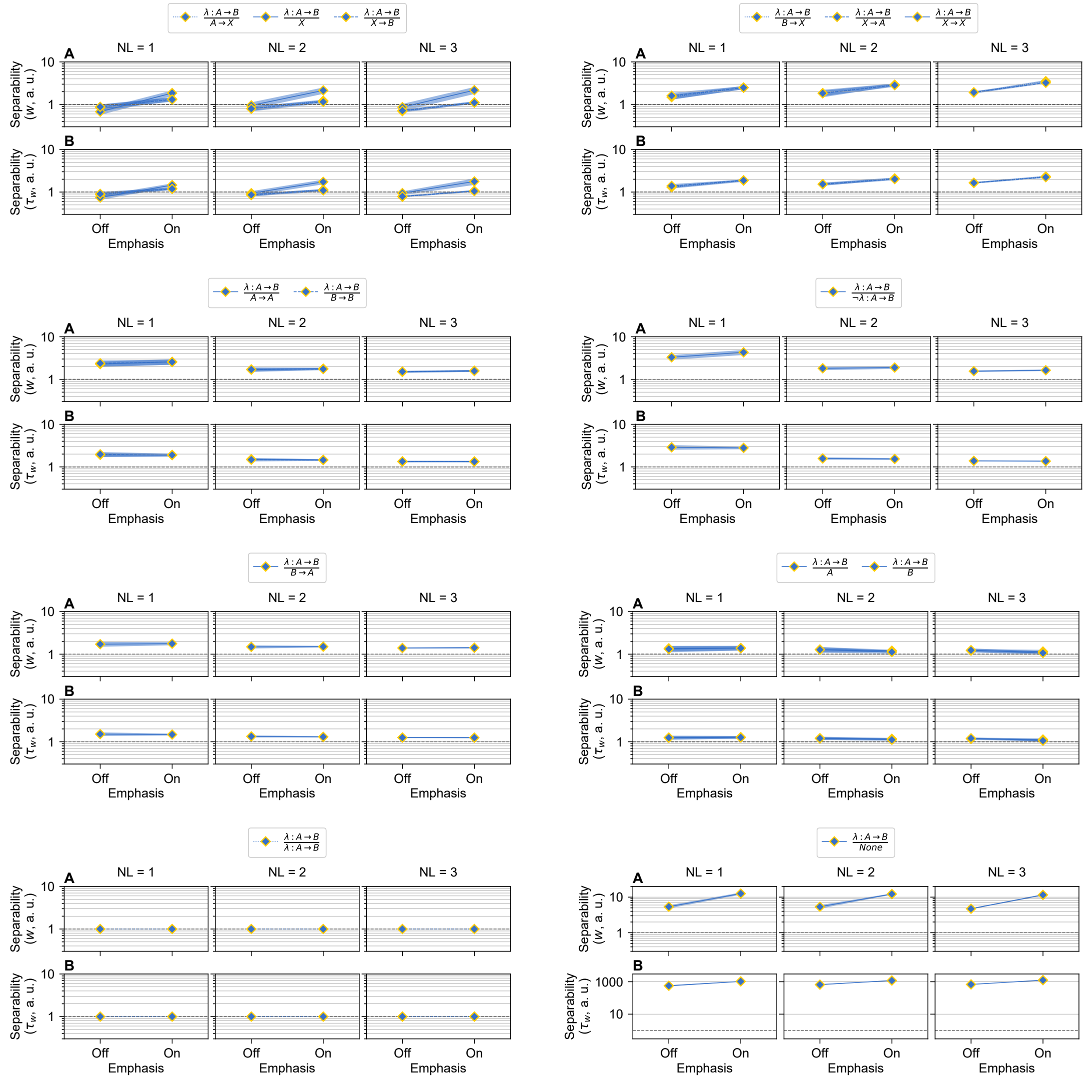


Figure A.7: Separability measures of w and τ_w in the experiment ‘Emphasis’ (reference $\lambda : A \to B$).

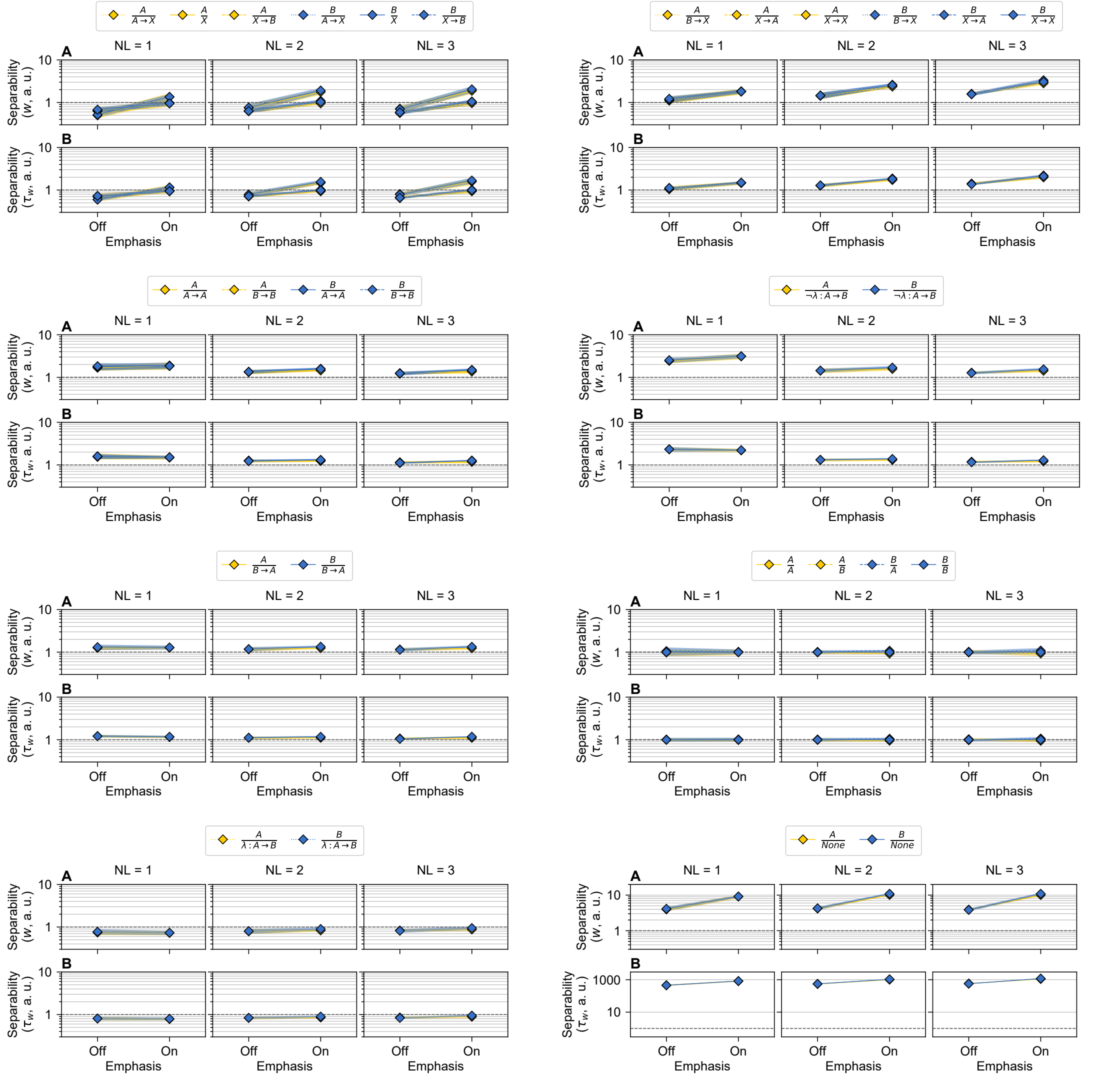


Figure A.8: Separability measures of w and τ_w in the experiment ‘Emphasis’ (references A and B).

List of Figures

List of Tables

Literature

Abbott, L. F. (1999). “Lapicque’s introduction of the integrate-and-fire model neuron (1907)”. In: *Brain Research Bulletin* 50.5/6, pp. 303–304.

Abbott, L. F. (2008). “Theoretical neuroscience rising”. In: *Neuron* 60, pp. 489–495. DOI: 10.1016/j.neuron.2008.10.019.

Abbott, L. F., DePasquale, B., and Memmesheimer, R.-M. (2016). “Building functional networks of spiking model neurons”. In: *Nature Neuroscience* 19.3, pp. 350–355. DOI: 10.1038/nn.4241.

Abbott, L. F. and Nelson, S. B. (2000). “Synaptic plasticity: taming the beast”. In: *Nature Neuroscience* 3, pp. 1178–1183. DOI: 10.1038/81453.

Abbott, L. F. and Regehr, W. G. (2004). “Synaptic computation”. In: *Nature* 431, pp. 796–803. DOI: 10.1038/nature03010.

Abdellah, M., Hernando, J., Antille, N., Eilemann, S., Markram, H., and Schürmann, F. (2017). “Reconstruction and visualization of large-scale volumetric models of neocortical circuits for physically-plausible in silico optical studies”. In: *BMC Bioinformatics* 18(Suppl 10).402, pp. 39–50. DOI: 10.1186/s12859-017-1788-4.

Abe, K. and Watanabe, D. (2011). “Songbirds possess the spontaneous ability to discriminate syntactic rules”. In: *Nature Neuroscience* 14.8, pp. 1067–1074. DOI: 10.1038/nn.2869.

Adams, F. (2010). “Embodied cognition”. In: *Phenomenology and the Cognitive Sciences* 9, pp. 619–628. DOI: 10.1007/s11097-010-9175-x.

Albright, T. D., Kandel, E. R., and Posner, M. I. (2000). “Cognitive neuroscience”. In: *Current Opinion in Neurobiology* 10, pp. 612–624. DOI: 10.1016/S0959-4388(00)00132-X.

Alhama, R. G. and Zuidema, W. (2019). “A review of computational models of basic rule learning: the neural-symbolic debate and beyond”. In: *Psychonomic Bulletin and Review* 26, pp. 1174–1194. DOI: 10.3758/s13423-019-01602-z.

Amit, D. J., Fusi, S., and Yakovlev, V. (1997). “Paradigmatic working memory (attractor) cell in IT cortex”. In: *Neural Computation* 9, pp. 1071–1092. DOI: 10.1162/neco.1997.9.5.1071.

Amunts, K., Ebell, C., Muller, J., Telefont, M., Knoll, A., and Lippert, T. (2016). "The human brain project: creating a European research infrastructure to decode the human brain". In: *Neuron* 92, pp. 574–581. DOI: 10.1016/j.neuron.2016.10.046.

Amunts, K. and Zilles, K. (2015). "Architectonic mapping of the human brain beyond Brodmann". In: *Neuron* 88, pp. 1086–1107. DOI: 10.1016/j.neuron.2015.12.001.

Anderson, M. L. (2003). "Embodied cognition: a field guide". In: *Artificial Intelligence* 149, pp. 91–130. DOI: 10.1016/S0004-3702(03)00054-7.

Ardila, A., Bernal, B., and Rosselli, M. (2016). "How localized are language brain areas? A review of Brodmann areas involvement in oral language". In: *Archives of Clinical Neuropsychology* 31, pp. 112–122. DOI: 10.1093/arclin/acv081.

Arnal, L. H. and Giraud, A.-L. (2012). "Cortical oscillations and sensory predictions". In: *Trends in Cognitive Sciences* 16.7, pp. 390–398. DOI: 10.1016/j.tics.2012.05.003.

Arulkumaran, K., Deisenroth, M. P., Brundage, M., and Bharath, A. A. (2017). "Deep reinforcement learning: a brief survey". In: *IEEE Signal Processing Magazine* 34.6, pp. 26–38. DOI: 10.1109/MSP.2017.2743240.

Aslin, R. N. (2017). "Statistical learning: a powerful mechanism that operates by mere exposure". In: *WIREs Cognitive Science* 8, e1373. DOI: 10.1002/wcs.1373.

Aslin, R. N., Saffran, J. R., and Newport, E. L. (1998). "Computation of conditional probability statistics by 8-month-old infants". In: *Psychological Science* 9.4, pp. 321–324. DOI: 10.1111/1467-9280.00063.

Aswolinskiy, W. and Pipa, G. (2015). "RM-SORN: a reward-modulated self-organizing recurrent neural network". In: *Frontiers in Computational Neuroscience* 9.36, pp. 1–15. DOI: 10.3389/fncom.2015.00036.

Aswolinskiy, W., Reinhart, R. F., and Steil, J. (2018). "Time series classification in reservoir- and model-space". In: *Neural Processing Letters* 48, pp. 789–809. DOI: 10.1007/s11063-017-9765-5.

Aur, D., Conolly, C. I., and Jog, M. S. (2006). "Computing information in neuronal spikes". In: *Neural Processing Letters* 23, pp. 183–199. DOI: 10.1007/s11063-006-6266-3.

Avena-Koenigsberger, A., Misic, B., and Sporns, O. (2018). "Communication dynamics in complex brain networks". In: *Nature Reviews Neuroscience* 19, pp. 17–33. DOI: 10.1038/nrn.2017.149.

Badakva, A. M., Miller, N. V., and Zobova, L. N. (2020). "New concepts on the implementation of motor and cognitive functions in the brain: facts and hypotheses". In: *Human Physiology* 46.3, pp. 343–350. DOI: 10.1134/S0362119720030020.

Badcock, P. B., Friston, K. J., Ramstead, M. J. D., Ploeger, A., and Hohwy, J. (2019). "The hierarchically mechanistic mind: an evolutionary systems theory of the human brain, cognition, and behavior". In: *Cognitive, Affective, & Behavioral Neuroscience* 19, pp. 1319–1351. DOI: 10.3758/s13415-019-00721-3.

Bahlmann, J., Schubotz, R. I., Mueller, J. L., Koester, D., and Friederici, A. D. (2009). “Neural circuits of hierarchical visuo-spatial sequence processing”. In: *Brain Research* 1298, pp. 161–170. DOI: 10.1016/j.brainres.2009.08.017.

Bale, M. R., Bitzidou, M., Pitas, A., Brebner, L. S., Khazim, L., Anagnou, S. T., Stevenson, C. D., and Maravall, M. (2017). “Learning and recognition of tactile temporal sequences by mice and humans”. In: *eLife* 6, e27333. DOI: 10.7554/eLife.27333.

Barbour, B., Brunel, N., Hakim, V., and Nadal, J.-P. (2007). “What can we learn from synaptic weight distributions?” In: *Trends in Neurosciences* 30.12, pp. 622–629. DOI: 10.1016/j.tins.2007.09.005.

Barlow, H. B. (1989). “Unsupervised learning”. In: *Neural Computation* 1, pp. 295–311. DOI: 10.1162/neco.1989.1.3.295.

Barsalou, L. W. (2008). “Grounded cognition”. In: *Annual Review of Psychology* 59, pp. 617–645. DOI: 10.1146/annurev.psych.59.103006.093639.

Bartolozzi, C., Indiveri, G., and Donati, E. (2022). “Embodied neuromorphic intelligence”. In: *Nature Communications* 13.1024, pp. 1–14. DOI: 10.1038/s41467-022-28487-2.

Basanisi, R., Brovelli, A., Cartoni, E., and Baldassarre, G. (2020). “A generative spiking neural-network model of goal-directed behaviour and one-step planning”. In: *PLoS Computational Biology* 16.12, e1007579. DOI: 10.1371/journal.pcbi.1007579.

Beckers, G. J. L., Bolhuis, J. J., Okanoya, K., and Berwick, R. C. (2012). “Birdsong neurolinguistics: songbird context-free grammar claim is premature”. In: *NeuroReport* 23.3, pp. 139–145. DOI: 10.1097/WNR.0b013e32834f1765.

Beekhuizen, B., Bod, R., and Zuidema, W. (2013). “Three design principles of language: the search for parsimony in redundancy”. In: *Language and Speech* 56.3, 265–290. DOI: 10.1177/0023830913484897.

Bekolay, T., Bergstra, J., Hunsberger, E., DeWolf, T., Stewart, T. C., Rasmussen, D., Choo, X., Voelker, A. R., and Eliasmith, C. (2014). “Nengo: a Python tool for building large-scale functional brain models”. In: *Frontiers in Neuroinformatics* 7.48, pp. 1–13. DOI: 10.3389/fninf.2013.00048.

Belatreche, A., Maguire, L. P., and McGinnity, M. (2007). “Advances in design and application of spiking neural networks”. In: *Soft Computing* 11, pp. 239–248. DOI: 10.1007/s00500-006-0065-7.

Belitski, A., Panzeri, S., Magri, C., Logothetis, N. K., and Kayser, C. (2010). “Sensory information in local field potentials and spikes from visual and auditory cortices: time scales and frequency bands”. In: *Journal of Computational Neuroscience* 29, pp. 533–545. DOI: 10.1007/s10827-010-0230-y.

Bellec, G., Salaj, D., Subramoney, A., Legenstein, R., and Maass, W. (2018). “Long short-term memory and learning-to-learn in networks of spiking neurons”. In:

Advances in Neural Information Processing Systems 31. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Curran Associates, Inc.

Bellec, G., Scherr, F., Subramoney, A., Hajek, E., Salaj, D., Legenstein, R., and Maass, W. (2020). "A solution to the learning dilemma for recurrent networks of spiking neurons". In: *Nature Communications* 11.3625, pp. 1–15. DOI: 10.1038/s41467-020-17236-y.

Benna, M. K. and Fusi, S. (2016). "Computational principles of synaptic memory consolidation". In: *Nature Neuroscience* 19.12, pp. 1697–1706. DOI: 10.1038/nn.4401.

Berger, T. K., Silberberg, G., Perin, R., and Markram, H. (2010). "Brief bursts self-inhibit and correlate the pyramidal network". In: *PLoS Biology* 8.9, e1000473. DOI: 10.1371/journal.pbio.1000473.

Berwick, R. C., Friederici, A. D., Chomsky, N., and Bolhuis, J. J. (2013). "Evolution, brain, and the nature of language". In: *Trends in Cognitive Sciences* 17.2, pp. 89–98. DOI: 10.1016/j.tics.2012.12.002.

Berwick, R. C., Pietroski, P., Yankama, B., and Chomsky, N. (2011). "Poverty of the stimulus revisited". In: *Cognitive Science* 35, pp. 1207–1242. DOI: 10.1111/j.1551-6709.2011.01189.x.

Bettoni, R., Bulf, H., Brady, S., and Johnson, S. P. (2021). "Infants' learning of non-adjacent regularities from visual sequences". In: *Infancy* 26, pp. 319–326. DOI: 10.1111/infa.12384.

Bi, G.-q. and Poo, M.-m. (1998). "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type". In: *The Journal of Neuroscience* 18.24, pp. 10464–10472. DOI: 10.1523/JNEUROSCI.18-24-10464.1998.

Bi, G.-q. and Poo, M.-m. (1999). "Distributed synaptic modification in neural networks induced by patterned stimulation". In: *Nature* 401, pp. 792–796. DOI: 10.1038/44573.

Binzegger, T., Douglas, R. J., and Martin, K. A. C. (2004). "A quantitative map of the circuit of cat primary visual cortex". In: *The Journal of Neuroscience* 24.39, pp. 8441–8453. DOI: 10.1523/JNEUROSCI.1400-04.2004.

Blundell, I., Brette, R., Cleland, T. A., Close, T. G., Coca, D., Davison, A. P., Diaz-Pier, S., Musoles, C. F., Gleeson, P., Goodman, D. F. M., Hines, M., Hopkins, M. W., Kumbhar, P., Lester, D. R., Marin, B., Morrison, A., Müller, E., Nowotny, T., Peyser, A., Plotnikov, D., Richmond, P., Rowley, A., Rumpe, B., Stimberg, M., Stokes, A. B., Tomkins, A., Trensch, G., Woodman, M., and Eppler, J. M. (2018). "Code generation in computational neuroscience: a review of tools and techniques". In: *Frontiers in Neuroinformatics* 12.68, pp. 1–35. DOI: 10.3389/fninf.2018.00068.

Bogaerts, L., Siegelman, N., Christiansen, M. H., and Frost, R. (2022). “Is there such a thing as a ‘good statistical learner’?” In: *Trends in Cognitive Sciences* 26.1, pp. 25–37. DOI: 10.1016/j.tics.2021.10.012.

Bogaerts, L., Siegelman, N., and Frost, R. (2021). “Statistical learning and language impairments: toward more precise theoretical accounts”. In: *Perspectives on Psychological Science* 16.2, 319–337. DOI: 10.1177/1745691620953082.

Bonatti, L. L., Peña, M., Nespor, M., and Mehler, J. (2005). “Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing”. In: *Psychological Science* 16.6, pp. 451–459. DOI: 10.1111/j.0956-7976.2005.01556.x.

Bono, J. and Clopath, C. (2017). “Modeling somatic and dendritic spike mediated plasticity at the single neuron and network level”. In: *Nature Communications* 8.706, pp. 1–17. DOI: 10.1038/s41467-017-00740-z.

Borst, A. and Theunissen, F. E. (1999). “Information theory and neural coding”. In: *Nature Neuroscience* 2.11, pp. 947–957. DOI: 10.1038/14731.

Boucher, L. and Dienes, Z. (2003). “Two ways of learning associations”. In: *Cognitive Science* 27, pp. 807–842. DOI: 10.1016/j.cogsci.2003.03.001.

Bowling, D. L. and Fitch, W. T. (2015). “Do animal communication systems have phonemes?” In: *Trends in Cognitive Sciences* 19.10, pp. 555–557. DOI: 10.1016/j.tics.2015.08.011.

Brader, J. M., Senn, W., and Fusi, S. (2007). “Learning real-world stimuli in a neural network with spike-driven synaptic dynamics”. In: *Neural Computation* 19, pp. 2881–2912. DOI: 10.1162/neco.2007.19.11.2881.

Brasselet, R., Panzeri, S., Logothetis, N. K., and Kayser, C. (2012). “Neurons with stereotyped and rapid responses provide a reference frame for relative temporal coding in primate auditory cortex”. In: *The Journal of Neuroscience* 32.9, pp. 2998–3008. DOI: 10.1523/JNEUROSCI.5435-11.2012.

Bresciani, J.-P., Ernst, M. O., Drewing, K., Bouyer, G., Maury, V., and Kheddar, A. (2005). “Feeling what you hear: auditory signals can modulate tactile tap perception”. In: *Experimental Brain Research* 162, pp. 172–180. DOI: 10.1007/s00221-004-2128-2.

Bressler, S. L. and Menon, V. (2010). “Large-scale brain networks in cognition: emerging methods and principles”. In: *Trends in Cognitive Sciences* 14.6, pp. 277–290. DOI: 10.1016/j.tics.2010.04.004.

Brette, R. and Gerstner, W. (2005). “Adaptive exponential integrate-and-fire model as an effective description of neuronal activity”. In: *Journal of Neurophysiology* 94, pp. 3637–3642. DOI: 10.1152/jn.00686.2005.

Brette, R., Rudolph, M., Carnevale, T., Hines, M., Beeman, D., Bower, J. M., Diesmann, M., Morrison, A., Goodman, P. H., Harris, Jr., F. C., Zirpe, M., Natschläger, T., Pecevski, D., Ermentrout, B., Djurfeldt, M., Lansner, A., Rochel, O., Vieville, T., Muller, E., Davison, A. P., El Boustani, S., and Destexhe, A. (2007). “Simulation of networks of spiking neurons: a review of tools and strategies”. In: *Journal of Computational Neuroscience* 23, pp. 349–398. DOI: 10.1007/s10827-007-0038-6.

Bröhl, F., Keitel, A., and Kayser, C. (2022). “MEG activity in visual and auditory cortices represents acoustic speech-related information during silent lip reading”. In: *eNeuro* 9.3, pp. 1–15. DOI: 10.1523/ENEURO.0209-22.2022.

Brunel, N. and van Rossum, M. C. W. (2007). “Lapicque’s 1907 paper: from frogs to integrate-and-fire”. In: *Biological Cybernetics* 97, pp. 337–339. DOI: 10.1007/s00422-007-0190-0.

Brysbaert, M., Mandera, P., and Keuleers, E. (2018). “The word frequency effect in word processing: an updated review”. In: *Current Directions in Psychological Science* 27.1, 45–50. DOI: 10.1177/0963721417727521.

Brzosko, Z., Mierau, S. B., and Paulsen, O. (2019). “Neuromodulation of spike-timing-dependent plasticity: past, present, and future”. In: *Neuron* 103, pp. 563–581. DOI: 10.1016/j.neuron.2019.05.041.

Bubic, A., von Cramon, D. Y., and Schubotz, R. I. (2010). “Prediction, cognition and the brain”. In: *Frontiers in Human Neuroscience* 4.25, pp. 1–15. DOI: 10.3389/fnhum.2010.00025.

Buonomano, D. V. and Maass, W. (2009). “State-dependent computations: spatiotemporal processing in cortical networks”. In: *Nature Reviews Neuroscience* 10, pp. 113–125. DOI: 10.1038/nrn2558.

Butz, M. V. (2016). “Toward a unified sub-symbolic computational theory of cognition”. In: *Frontiers in Psychology* 7.925, pp. 1–19. DOI: 10.3389/fpsyg.2016.00925.

Buzsáki, G. (2010). “Neural syntax: cell assemblies, synapsembles, and readers”. In: *Neuron* 68, pp. 362–385. DOI: 10.1016/j.neuron.2010.09.023.

Buzsáki, G. and Draguhn, A. (2004). “Neuronal oscillations in cortical networks”. In: *Science* 304.5679, pp. 1926–1929. DOI: 10.1126/science.1099745.

Caligiore, D., Arbib, M. A., Miall, R. C., and Baldassarre, G. (2019). “The super-learning hypothesis: integrating learning processes across cortex, cerebellum and basal ganglia”. In: *Neuroscience and Biobehavioral Reviews* 100, pp. 19–34. DOI: 10.1016/j.neubiorev.2019.02.008.

Carandini, M. (2012). “From circuits to behavior: a bridge too far?” In: *Nature Neuroscience* 15.4, pp. 507–509. DOI: 10.1038/nn.3043.

Castellano, M. and Pipa, G. (2013). “Memory trace in spiking neural networks”. In: *Artificial Neural Networks and Machine Learning – ICANN 2013. ICANN 2013. Lecture Notes in Computer Science, vol 8131.* Ed. by Mladenov V., Koprinkova-Hristova P., Palm G., Villa A. E. P., Appollini B., and Kasabov N. Springer, Berlin, Heidelberg, pp. 264–271. DOI: 10.1007/978-3-642-40728-4_33.

Castet, E. and Masson, G. S. (2000). “Motion perception during saccadic eye movements”. In: *Nature Neuroscience* 3.2, pp. 177–183. DOI: 10.1038/72124.

Cazin, N., Llofriu Alonso, M., Scleidorovich Chiodi, P., Pelc, T., Harland, B., Weitzenfeld, A., Fellous, J.-M., and Dominey, P. F. (2019). “Reservoir computing model of prefrontal cortex creates novel combinations of previous navigation sequences from hippocampal place-cell replay with spatial reward propagation”. In: *PLoS Computational Biology* 15.7, e1006624. DOI: 10.1371/journal.pcbi.1006624.

Cecchini, M., Lai, C., and Langher, V. (2007). “Communication and crying in newborns”. In: *Infant Behavior & Development* 30, pp. 655–665. DOI: 10.1016/j.infbeh.2007.03.002.

Chady, M. (2001). “Modelling higher cognitive functions with Hebbian cell assemblies”. In: *Emergent Neural Computational Architectures Based on Neuroscience: Towards Neuroscience-Inspired Computing. Lecture Notes in Computer Science, vol 2036*. Ed. by S. Wermter, J. Austin, and D. Willshaw. Springer, Berlin, Heidelberg, pp. 398–406. DOI: 10.1007/3-540-44597-8_29.

Chang, F. and Fitz, H. (2014). “Computational models of sentence production: a dual-path approach”. In: *The Oxford Handbook of Language Production*. Ed. by M. Goldrick, V. S. Ferreira, and M. Miozzo. Oxford University Press. DOI: 10.1093/oxfordhb/9780199735471.013.019.

Chang, F., Janciauskas, M., and Fitz, H. (2012). “Language adaptation and learning: getting explicit about implicit learning”. In: *Language and Linguistics Compass* 6.5, pp. 259–278. DOI: 10.1002/lnc3.337.

Chater, N. and Christiansen, M. H. (2018). “Language acquisition as skill learning”. In: *Current Opinion in Behavioral Sciences* 21, pp. 205–208. DOI: 10.1016/j.cobeha.2018.04.001.

Chen, G., Cao, H., Conradt, J., Tang, H., Röhrbein, F., and Knoll, A. (2020). “Event-based neuromorphic vision for autonomous driving: a paradigm shift for bio-inspired visual sensing and perception”. In: *IEEE Signal Processing Magazine* 37.4, pp. 34–49. DOI: 10.1109/MSP.2020.2985815.

Chen, J.-Y., Lonjers, P., Lee, C., Chistiakova, M., Volgushev, M., and Bazhenov, M. (2013). “Heterosynaptic plasticity prevents runaway synaptic dynamics”. In: *The Journal of Neuroscience* 33.40, pp. 15915–15929. DOI: 10.1523/JNEUROSCI.5088-12.2013.

Chen, L., Goucha, T., Männel, C., Friederici, A. D., and Zaccarella, E. (2021). “Hierarchical syntactic processing is beyond mere associating: functional magnetic resonance imaging evidence from a novel artificial grammar”. In: *Human Brain Mapping* 42, pp. 3253–3268. DOI: 10.1002/hbm.25432.

Chicca, E., Stefanini, F., Bartolozzi, C., and Indiveri, G. (2014). “Neuromorphic electronic circuits for building autonomous cognitive systems”. In: *Proceedings of the IEEE* 102.9, pp. 1367–1388. DOI: 10.1109/JPROC.2014.2313954.

Chistiakova, M., Bannon, N. M., Bazhenov, M., and Volgushev, M. (2014). "Heterosynaptic plasticity: multiple mechanisms and multiple roles". In: *The Neuroscientist* 20.5, 483–498. DOI: 10.1177/1073858414529829.

Chistiakova, M., Bannon, N. M., Chen, J.-Y., Bazhenov, M., and Volgushev, M. (2015). "Homeostatic role of heterosynaptic plasticity: models and experiments". In: *Frontiers in Computational Neuroscience* 9.89, pp. 1–22. DOI: 10.3389/fncom.2015.00089.

Christiansen, M. H., Allen, J., and Seidenberg, M. S. (1998). "Learning to segment speech using multiple cues: a connectionist model". In: *Language and Cognitive Processes* 13.2–3, pp. 221–268. DOI: 10.1080/016909698386528.

Christiansen, M. H. and Chater, N. (2016). "The now-or-never bottleneck: a fundamental constraint on language". In: *Behavioral and Brain Sciences* 39, e62. DOI: 10.1017/S0140525X1500031X.

Christiansen, M. H. and Curtin, S. L. (1999). "The power of statistical learning: no need for algebraic rules". In: *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society*. Ed. by M. Hahn and S. C. Stoness. Psychology Press. DOI: 10.4324/9781410603494.

Churchland, P. S. and Sejnowski, T. J. (1988). "Perspectives on cognitive neuroscience". In: *Science* 242, pp. 741–745. DOI: 10.1126/science.3055294.

Cichy, R. M. and Kaiser, D. (2019). "Deep neural networks as scientific models". In: *Trends in Cognitive Sciences* 23.4, pp. 305–317. DOI: 10.1016/j.tics.2019.01.009.

Clark, A. (2013). "Whatever next? Predictive brains, situated agents, and the future of cognitive science". In: *Behavioral and Brain Sciences* 36.3, pp. 181–204. DOI: 10.1017/S0140525X12000477.

Clay, V., König, P., Kühnberger, K.-U., and Pipa, G. (2021). "Learning sparse and meaningful representations through embodiment". In: *Neural Networks* 134, pp. 23–41. DOI: 10.1016/j.neunet.2020.11.004.

Cleeremans, A. and McClelland, J. L. (1991). "Learning the structure of event sequences". In: *Journal of Experimental Psychology: General* 120.3, pp. 235–253. DOI: 10.1037/0096-3445.120.3.235.

Cone, I. and Shouval, H. Z. (2021). "Behavioral time scale plasticity of place fields: mathematical analysis". In: *Frontiers in Computational Neuroscience* 15.640235, pp. 1–11. DOI: 10.3389/fncom.2021.640235.

Connelly, W. M., Laing, M., Errington, A. C., and Crunelli, V. (2016). "The thalamus as a low pass filter: filtering at the cellular level does not equate with filtering at the network level". In: *Frontiers in Neural Circuits* 9.89, pp. 1–10. DOI: 10.3389/fncir.2015.00089.

Conway, C. M. and Christiansen, M. H. (2005). "Modality-constrained statistical learning of tactile, visual, and auditory sequences". In: *Journal of Experimental*

Psychology: Learning, Memory, and Cognition 31.1, pp. 24–39. DOI: 10.1037/0278-7393.31.1.24.

Corballis, M. C. (2020). "Crossing the rubicon: behaviorism, language, and evolutionary continuity". In: *Frontiers in Psychology* 11.653, pp. 1–10. DOI: 10.3389/fpsyg.2020.00653.

Corradi, F., You, H., Giulioni, M., and Indiveri, G. (2015). "Decision making and perceptual bistability in spike-based neuromorphic VLSI systems". In: *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2708–2711. DOI: 10.1109/ISCAS.2015.7169245.

Cracco, E., Lee, H., van Belle, G., Quenon, L., Haggard, P., Rossion, B., and Orgs, G. (2021). "EEG frequency tagging reveals the integration of form and motion cues into the perception of group movement". In: *Cerebral Cortex* 32.13, pp. 2843–2857. DOI: 10.1093/cercor/bhab385.

Cui, Y., Ahmad, S., and Hawkins, J. (2016). "Continuous online sequence learning with an unsupervised neural network model". In: *Neural Computation* 28, pp. 2474–2504. DOI: 10.1162/NECO_a_00893.

Culbertson, J., Koulaguina, E., Gonzalez-Gomez, N., Legendre, G., and Nazzi, T. (2016). "Developing knowledge of nonadjacent dependencies". In: *Developmental Psychology* 52.12, pp. 2174–2183. DOI: 10.1037/dev0000246.

Dąbrowska, E. (2015). "What exactly is universal grammar, and has anyone seen it?" In: *Frontiers in Psychology* 6.852, pp. 1–17. DOI: 10.3389/fpsyg.2015.00852.

Dan, Y. and Poo, M.-M. (2006). "Spike timing-dependent plasticity: from synapse to perception". In: *Physiological Reviews* 86, pp. 1033–1048. DOI: 10.1152/physrev.00030.2005.

Daoudal, G. and Debanne, D. (2003). "Long-term plasticity of intrinsic excitability: learning rules and mechanisms". In: *Learning & Memory* 10, pp. 456–465. DOI: 10.1101/lm.64103.

Dayan, P. and Niv, Y. (2008). "Reinforcement learning: the good, the bad and the ugly". In: *Current Opinion in Neurobiology* 18, pp. 185–196. DOI: 10.1016/j.conb.2008.08.003.

De Diego Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., and Bachoud-Lévi, A.-C. (2007). "Different neurophysiological mechanisms underlying word and rule extraction from speech". In: *PLoS ONE* 2.11, e1175. DOI: 10.1371/journal.pone.0001175.

de Diego-Balaguer, R., Martinez-Alvarez, A., and Pons, F. (2016). "Temporal attention as a scaffold for language development". In: *Frontiers in Psychology* 7.44, pp. 1–15. DOI: 10.3389/fpsyg.2016.00044.

de Diego-Balaguer, R., Rodríguez-Fornells, A., and Bachoud-Lévi, A.-C. (2015). "Prosodic cues enhance rule learning by changing speech segmentation mechanisms". In: *Frontiers in Psychology* 6.1478, pp. 1–11. DOI: 10.3389/fpsyg.2015.01478.

de la Mora, D. M. and Toro, J. M. (2013). "Rule learning over consonants and vowels in a non-human animal". In: *Cognition* 126, pp. 307–312. DOI: 10.1016/j.cognition.2012.09.015.

de Vries, M. H., Petersson, K. M., Geukes, S., Zwitserlood, P., and Christiansen, M. H. (2012). "Processing multiple non-adjacent dependencies: evidence from sequence learning". In: *Philosophical Transactions of the Royal Society B* 367, pp. 2065–2076. DOI: 10.1098/rstb.2011.0414.

De Zeeuw, C. I., Hoebeek, F. E., Bosman, L. W. J., Schonewille, M., Witter, L., and Koekkoek, S. K. (2011). "Spatiotemporal firing patterns in the cerebellum". In: *Nature Reviews Neuroscience* 12, pp. 327–344. DOI: 10.1038/nrn3011.

Debanne, D., Gähwiler, B. H., and Thompson, S. M. (1998). "Long-term synaptic plasticity between pairs of individual CA3 pyramidal cells in rat hippocampal slice cultures". In: *Journal of Physiology* 507.1, pp. 237–247. DOI: 10.1111/j.1469-7793.1998.237bu.x.

Debat, G., Chauhan, T., Cottereau, B. R., Masquelier, T., Paindavoine, M., and Baures, R. (2021). "Event-based trajectory prediction using spiking neural networks". In: *Frontiers in Computational Neuroscience* 15.658764, pp. 1–16. DOI: 10.3389/fncom.2021.658764.

deCharms, R. C. and Zador, A. (2000). "Neural representation and the cortical code". In: *Annual Review of Neuroscience* 23, pp. 613–647. DOI: 10.1146/annurev.neuro.23.1.613.

Deco, G. and Jirsa, V. K. (2012). "Ongoing cortical activity at rest: criticality, multistability, and ghost attractors". In: *The Journal of Neuroscience* 32.10, pp. 3366–3375. DOI: 10.1523/JNEUROSCI.2523-11.2012.

Deco, G., Jirsa, V. K., and McIntosh, A. R. (2013). "Resting brains never rest: computational insights into potential cognitive architectures". In: *Trends in Neurosciences* 36.5, pp. 268–274. DOI: 10.1016/j.tins.2013.03.001.

Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., and Friston, K. (2008). "The dynamic brain: from spiking neurons to neural masses and cortical fields". In: *PLoS Computational Biology* 4.8, e1000092. DOI: 10.1371/journal.pcbi.1000092.

DeFelipe, J., Markram, H., and Rockland, K. S. (2012). "The neocortical column". In: *Frontiers in Neuroanatomy* 6.22, pp. 1–2. DOI: 10.3389/fnana.2012.00022.

Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., and Sablé-Meyer, M. (2022). "Symbols and mental programs: a hypothesis about human singularity". In: *Trends in Cognitive Sciences* 26.9, pp. 751–766. DOI: 10.1016/j.tics.2022.06.010.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). "The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees". In: *Neuron* 88, pp. 2–19. DOI: 10.1016/j.neuron.2015.09.019.

Del Papa, B., Priesemann, V., and Triesch, J. (2017). “Criticality meets learning: criticality signatures in a self-organizing recurrent neural network”. In: *PLoS ONE* 12.5, e0178683. DOI: 10.1371/journal.pone.0178683.

Desai, N. S., Rutherford, L. C., and Turrigiano, G. G. (1999). “Plasticity in the intrinsic excitability of cortical pyramidal neurons”. In: *Nature Neuroscience* 2.6, pp. 515–520. DOI: 10.1038/9165.

Diehl, P. U. and Cook, M. (2015). “Unsupervised learning of digit recognition using spike-timing-dependent plasticity”. In: *Frontiers in Computational Neuroscience* 9.99, pp. 1–9. DOI: 10.3389/fncom.2015.00099.

Diesmann, M. and Gewaltig, M.-O. (2001). “NEST: an environment for neural systems simulations”. In: *Forschung und wissenschaftliches Rechnen: Beiträge zum Heinz-Billing-Preis 2001*. Ed. by T. Plesser and V. Macho. Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen, pp. 43–70.

Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). “Cortical tracking of hierarchical linguistic structures in connected speech”. In: *Nature Neuroscience* 19.1, pp. 158–164. DOI: doi:10.1038/nn.4186.

Ding, N. and Simon, J. Z. (2014). “Cortical entrainment to continuous speech: functional roles and interpretations”. In: *Frontiers in Human Neuroscience* 8.311, pp. 1–7. DOI: 10.3389/fnhum.2014.00311.

Dinh, T. T. and Hinaut, X. (2020). “Language acquisition with echo state networks: towards unsupervised learning”. In: *2020 Joint IEEE 10th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pp. 1–6. DOI: 10.1109/ICDL-EpiRob48136.2020.9278041.

Doborjeh, Z., Doborjeh, M., Taylor, T., Kasabov, N., Wang, G. Y., Siegert, R., and Sumich, A. (2019). “Spiking neural network modelling approach reveals how mindfulness training rewires the brain”. In: *Scientific Reports* 9.6367, pp. 1–15. DOI: 10.1038/s41598-019-42863-x.

Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). “Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing”. In: *NeuroImage* 85, pp. 761–768. DOI: 10.1016/j.neuroimage.2013.06.035.

Dominey, P. F. (2005). “From sensorimotor sequence to grammatical construction: evidence from simulation and neurophysiology”. In: *Adaptive Behavior* 13.4, pp. 347–361. DOI: 10.1177/105971230501300401.

Dominey, P. F. (2013). “Recurrent temporal networks and language acquisition—from corticostriatal neurophysiology to reservoir computing”. In: *Frontiers in Psychology* 4.500, pp. 1–14. DOI: 10.3389/fpsyg.2013.00500.

Dominey, P. F. and Ramus, F. (2000). “Neural network processing of natural language: I. sensitivity to serial, temporal and abstract structure of language in the infant”. In: *Language and Cognitive Processes* 15.1, pp. 87–127. DOI: 10.1080/016909600386129.

Dominey, P., Arbib, M., and Joseph, J.-P. (1995). "A model of corticostriatal plasticity for learning oculomotor associations and sequences". In: *Journal of Cognitive Neuroscience* 7.3, pp. 311–336. DOI: 10.1162/jocn.1995.7.3.311.

Dörner, D. and Funke, J. (2017). "Complex problem solving: what it is and what it is not". In: *Frontiers in Psychology* 8.1153, pp. 1–11. DOI: 10.3389/fpsyg.2017.01153.

Doron, M., Chindemi, G., Muller, E., Markram, H., and Segev, I. (2017). "Timed synaptic inhibition shapes NMDA spikes, influencing local dendritic processing and global I/O properties of cortical neurons". In: *Cell Reports* 21, pp. 1550–1561. DOI: 10.1016/j.celrep.2017.10.035.

Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. C., and Suarez, H. H. (1995a). "Recurrent excitation in neocortical circuits". In: *Science* 269, pp. 981–985. DOI: 10.1126/science.7638624.

Douglas, R. J. and Martin, K. A. C. (2004). "Neuronal circuits of the neocortex". In: *Annual Review of Neuroscience* 27, pp. 419–451. DOI: 10.1146/annurev.neuro.27.070203.144152.

Douglas, R. J. and Martin, K. A. C. (2007a). "Mapping the matrix: the ways of neocortex". In: *Neuron* 56, pp. 226–238. DOI: 10.1016/j.neuron.2007.10.017.

Douglas, R. J. and Martin, K. A. C. (2007b). "Recurrent neuronal circuits in the neocortex". In: *Current Biology* 17.13, R496–R500. DOI: 10.1016/j.cub.2007.04.024.

Douglas, R. J., Martin, K. A. C., and Whitteridge, D. (1989). "A canonical microcircuit for neocortex". In: *Neural Computation* 1, pp. 480–488. DOI: 10.1162/neco.1989.1.4.480.

Douglas, R., Mahowald, M., and Mead, C. (1995b). "Neuromorphic analogue VLSI". In: *Annual Review of Neuroscience* 18, pp. 255–281. DOI: 10.1146/annurev.ne.18.030195.001351.

Doumas, L. A. A., Hummel, J. E., and Sandhofer, C. M. (2008). "A theory of the discovery and predication of relational concepts". In: *Psychological Review* 115.1, pp. 1–43. DOI: 10.1037/0033-295X.115.1.1.

Doumas, L. A. A. and Martin, A. E. (2018). "Chapter Six - Learning structured representations from experience". In: *Psychology of Learning and Motivation*. Ed. by K. D. Federmeier. Vol. 69. Academic Press, pp. 165–203. DOI: 10.1016/bs.plm.2018.10.002.

Doya, K. (1999). "What are the computations of the cerebellum, the basal ganglia and the cerebral cortex?" In: *Neural Networks* 12, pp. 961–974. DOI: 10.1016/S0893-6080(99)00046-5.

Doya, K. (2000). "Complementary roles of basal ganglia and cerebellum in learning and motor control". In: *Current Opinion in Neurobiology* 10, pp. 732–739. DOI: 10.1016/S0959-4388(00)00153-7.

Doyon, J. (2008). “Motor sequence learning and movement disorders”. In: *Current Opinion in Neurology* 21, pp. 478–483. DOI: 10.1097/WCO.0b013e328304b6a3.

Drew, P. J. and Abbott, L. F. (2006). “Extending the effects of spike-timing-dependent plasticity to behavioral timescales”. In: *Proceedings of the National Academy of Sciences* 103.23, pp. 8876–8881. DOI: 10.1073/pnas.0600676103.

Duarte, R. C. F. and Morrison, A. (2014). “Dynamic stability of sequential stimulus representations in adapting neuronal networks”. In: *Frontiers in Computational Neuroscience* 8.124, pp. 1–20. DOI: 10.3389/fncom.2014.00124.

Duarte, R. and Morrison, A. (2019). “Leveraging heterogeneity for neural computation with fading memory in layer 2/3 cortical microcircuits”. In: *PLoS Computational Biology* 15.4, e1006781. DOI: 10.1371/journal.pcbi.1006781.

Duarte, R., Seriès, P., and Morrison, A. (2014). “Self-organized artificial grammar learning in spiking neural networks”. In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 36, pp. 427–432.

Duarte, R., Uhlmann, M., van den Broek, D., Fitz, H., Petersson, K. M., and Morrison, A. (2018). “Encoding symbolic sequences with spiking neural reservoirs”. In: *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. DOI: 10.1109/IJCNN.2018.8489114.

Duncan, J., Assem, M., and Shashidhara, S. (2020). “Integrated intelligence from distributed brain activity”. In: *Trends in Cognitive Sciences* 24.10, pp. 838–852. DOI: 10.1016/j.tics.2020.06.012.

Effenberger, F., Jost, J., and Levina, A. (2015). “Self-organization in balanced state networks by STDP and homeostatic plasticity”. In: *PLoS Computational Biology* 11.9, e1004420. DOI: 10.1371/journal.pcbi.1004420.

Einevoll, G. T., Destexhe, A., Diesmann, M., Grün, S., Jirsa, V., de Kamps, M., Migliore, M., Ness, T. V., Plesser, H. E., and Schürmann, F. (2019). “The scientific case for brain simulations”. In: *Neuron* 102, pp. 735–744. DOI: 10.1016/j.neuron.2019.03.027.

Einevoll, G. T., Kayser, C., Logothetis, N. K., and Panzeri, S. (2013). “Modelling and analysis of local field potentials for studying the function of cortical circuits”. In: *Nature Reviews Neuroscience* 14, pp. 770–785. DOI: 10.1038/nrn3599.

Eliasmith, C. (2005). “Cognition with neurons: a large-scale, biologically realistic model of the Wason task”. In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 27, pp. 624–629.

Eliasmith, C. (2007). “How to build a brain: from function to implementation”. In: *Synthese* 159, pp. 373–388. DOI: 10.1007/s11229-007-9235-0.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., and Rasmussen, D. (2012). “A large-scale model of the functioning brain”. In: *Science* 338, pp. 1202–1205. DOI: 10.1126/science.1225266.

Eliasmith, C. and Trujillo, O. (2014). “The use and abuse of large-scale brain models”. In: *Current Opinion in Neurobiology* 25, pp. 1–6. DOI: 10.1016/j.conb.2013.09.009.

Elman, J. L. (1990). “Finding structure in time”. In: *Cognitive Science* 14, pp. 179–211. DOI: 10.1207/s15516709cog1402_1.

Elman, J. L. and Zipser, D. (1988). “Learning the hidden structure of speech”. In: *The Journal of the Acoustical Society of America* 83.4, pp. 1615–1626. DOI: 10.1121/1.395916.

Embick, D. and Poeppel, D. (2015). “Towards a computational(ist) neurobiology of language: correlational, integrated and explanatory neurolinguistics”. In: *Language, Cognition and Neuroscience* 30.4, pp. 357–366. DOI: 10.1080/23273798.2014.980750.

Endress, A. D. (2010). “Learning melodies from non-adjacent tones”. In: *Acta Psychologica* 135, pp. 182–190. DOI: 10.1016/j.actpsy.2010.06.005.

Endress, A. D. and Bonatti, L. L. (2007). “Rapid learning of syllable classes from a perceptually continuous speech stream”. In: *Cognition* 105, pp. 247–299. DOI: 10.1016/j.cognition.2006.09.010.

Endress, A. D. and Johnson, S. P. (2021). “When forgetting fosters learning: a neural network model for statistical learning”. In: *Cognition* 213, p. 104621. DOI: 10.1016/j.cognition.2021.104621.

Enel, P., Procyk, E., Quilodran, R., and Dominey, P. F. (2016). “Reservoir computing properties of neural dynamics in prefrontal cortex”. In: *PLoS Computational Biology* 12.6, e1004967. DOI: 10.1371/journal.pcbi.1004967.

Engel, A. K., König, P., Kreiter, A. K., Schillen, T. B., and Singer, W. (1992). “Temporal coding in the visual cortex: new vistas on integration in the nervous system”. In: *Trends in Neurosciences* 15.6, pp. 218–226. DOI: 10.1016/0166-2236(92)90039-B.

Engel, A. K., König, P., and Singer, W. (1991). “Direct physiological evidence for scene segmentation by temporal coding”. In: *Proceedings of the National Academy of Sciences* 88, pp. 9136–9140. DOI: 10.1073/pnas.88.20.9136.

Engel, A. K., Maye, A., Kurthen, M., and König, P. (2013). “Where’s the action? The pragmatic turn in cognitive science”. In: *Trends in Cognitive Sciences* 17.5, pp. 202–209. DOI: 10.1016/j.tics.2013.03.006.

Englund, M., Whitham, W., Conway, C. M., Beran, M. J., and Washburn, D. A. (2022). “Nonhuman primates learn adjacent dependencies but fail to learn nonadjacent dependencies in a statistical learning task with a salient cue”. In: *Learning & Behavior* 50, pp. 242–253. DOI: 10.3758/s13420-021-00485-2.

Erickson, L. C. and Thiessen, E. D. (2015). “Statistical learning of language: theory, validity, and predictions of a statistical learning account of language acquisition”. In: *Developmental Review* 37, pp. 66–108. DOI: 10.1016/j.dr.2015.05.002.

Eshraghian, J. K., Ward, M., Neftci, E., Wang, X., Lenz, G., Dwivedi, G., Bennamoun, M., Jeong, D. S., and Lu, W. D. (2022). “Training spiking neural networks using lessons from deep learning”. In: *arXiv* 2109.12894v4, pp. 1–44. DOI: 10.48550/arXiv.2109.12894.

Fahlman, S. E. and Lebiere, C. (1989). “The cascade-correlation learning architecture”. In: *Advances in Neural Information Processing Systems 2*. Ed. by D. Touretzky. Morgan-Kaufmann, pp. 524–532.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., Fox, P. T., Eickhoff, S. B., Yu, C., and Jiang, T. (2016). “The human brainnetome atlas: a new brain atlas based on connectional architecture”. In: *Cerebral Cortex* 26.8, pp. 3508–3526. DOI: 10.1093/cercor/bhw157.

Fan, X. and Markram, H. (2019). “A brief history of simulation neuroscience”. In: *Frontiers in Neuroinformatics* 13.32, pp. 1–28. DOI: 10.3389/fninf.2019.00032.

Fang, H., Zeng, Y., and Zhao, F. (2021). “Brain inspired sequences production by spiking neural networks with reward-modulated STDP”. In: *Frontiers in Computational Neuroscience* 15.612041, pp. 1–13. DOI: 10.3389/fncom.2021.612041.

Fang, X., Duan, S., and Wang, L. (2022). “Memristive Izhikevich spiking neuron model and its application in oscillatory associative memory”. In: *Frontiers in Neuroscience* 16.885322, pp. 1–18. DOI: 10.3389/fnins.2022.885322.

Fedorenko, E. and Thompson-Schill, S. L. (2014). “Reworking the language network”. In: *Trends in Cognitive Sciences* 18.3, pp. 120–126. DOI: 10.1016/j.tics.2013.12.006.

Feldman, D. E. (2000). “Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex”. In: *Neuron* 27, pp. 45–56. DOI: 10.1016/S0896-6273(00)00008-8.

Fellous, J.-M., Tiesinga, P. H. E., Thomas, P. J., and Sejnowski, T. J. (2004). “Discovering spike patterns in neuronal responses”. In: *The Journal of Neuroscience* 24.12, pp. 2989–3001. DOI: 10.1523/JNEUROSCI.4649-03.2004.

Fiebig, F. and Lansner, A. (2017). “A spiking working memory model based on Hebbian short-term potentiation”. In: *The Journal of Neuroscience* 37.1, pp. 83–96. DOI: 10.1523/JNEUROSCI.1989-16.2016.

Fiorillo, C. D., Kim, J. K., and Hong, S. Z. (2014). “The meaning of spikes from the neuron’s point of view: predictive homeostasis generates the appearance of randomness”. In: *Frontiers in Computational Neuroscience* 8.49, pp. 1–17. DOI: 10.3389/fncom.2014.00049.

Fiser, J. and Aslin, R. N. (2002a). “Statistical learning of higher-order temporal structure from visual shape sequences”. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28.3, pp. 458–467. DOI: 10.1037//0278-7393.28.3.458.

Fiser, J. and Aslin, R. N. (2002b). "Statistical learning of new visual feature combinations by infants". In: *Proceedings of the National Academy of Sciences* 99.24, pp. 15822–15826. DOI: 10.1073/pnas.232472899.

Fiser, J. and Aslin, R. N. (2005). "Encoding multielement scenes: statistical learning of visual feature hierarchies". In: *Journal of Experimental Psychology: General* 134.4, pp. 521–537. DOI: 10.1037/0096-3445.134.4.521.

Fitch, W. T. and Friederici, A. D. (2012). "Artificial grammar learning meets formal language theory: an overview". In: *Philosophical Transactions of the Royal Society B* 367, pp. 1933–1955. DOI: 10.1098/rstb.2012.0103.

Fitch, W. T. and Hauser, M. D. (2004). "Computational constraints on syntactic processing in a nonhuman primate". In: *Science* 303, pp. 377–380. DOI: 10.1126/science.1089401.

Fitz, H. (2011). "A liquid-state model of variability effects in learning nonadjacent dependencies". In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 33, pp. 897–902.

Fitz, H., Uhlmann, M., van den Broek, D., Duarte, R., Hagoort, P., and Petersson, K. M. (2020). "Neuronal spike-rate adaptation supports working memory in language processing". In: *Proceedings of the National Academy of Sciences* 117.34, pp. 20881–20889. DOI: 10.1073/pnas.2000222117.

Flinker, A. and Knight, R. T. (2018). "Broca's area in comprehension and production, insights from intracranial studies in humans". In: *Current Opinion in Behavioral Sciences* 21, pp. 170–175. DOI: 10.1016/j.cobeha.2018.04.012.

Flinker, A., Korzeniewska, A., Shestyuk, A. Y., Franaszczuk, P. J., Dronkers, N. F., Knight, R. T., and Crone, N. E. (2015). "Redefining the role of Broca's area in speech". In: *Proceedings of the National Academy of Sciences* 112.9, pp. 2871–2875. DOI: 10.1073/pnas.1414491112.

Fogerty, D. and Humes, L. E. (2012). "The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences". In: *The Journal of the Acoustical Society of America* 131.2, pp. 1490–1501. DOI: 10.1121/1.3676696.

Foglia, L. and Wilson, R. A. (2013). "Embodied cognition". In: *WIREs Cognitive Science* 4, pp. 319–325. DOI: 10.1002/wcs.1226.

Fonollosa, J., Neftci, E., and Rabinovich, M. (2015). "Learning of chunking sequences in cognition and behavior". In: *PLoS Computational Biology* 11.11, e1004592. DOI: 10.1371/journal.pcbi.1004592.

Fontaine, B., Goodman, D. F. M., Benichoux, V., and Brette, R. (2011). "Brian hears: online auditory processing using vectorization over channels". In: *Frontiers in Neuroinformatics* 5.9, pp. 1–9. DOI: 10.3389/fninf.2011.00009.

Frank, M. C., Goldwater, S., Griffiths, T. L., and Tenenbaum, J. B. (2010). “Modeling human performance in statistical word segmentation”. In: *Cognition* 117, pp. 107–125. DOI: 10.1016/j.cognition.2010.07.005.

Frank, S. L., Bod, R., and Christiansen, M. H. (2012). “How hierarchical is language use?” In: *Proceedings of the Royal Society B* 279, pp. 4522–4531. DOI: 10.1098/rspb.2012.1741.

French, R. M., Addyman, C., and Mareschal, D. (2011). “TRACX: a recognition-based connectionist framework for sequence segmentation and chunk extraction”. In: *Psychological Review* 118.4, pp. 614–636. DOI: 10.1037/a0025255.

French, R. M. and Cottrell, G. W. (2014). “TRACX 2.0: a memory-based, biologically-plausible model of sequence segmentation and chunk extraction”. In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 36, pp. 2216–2221.

Frick, A. and Johnston, D. (2005). “Plasticity of dendritic excitability”. In: *Journal of Neurobiology* 64, pp. 100–115. DOI: 10.1002/neu.20148.

Friederici, A. D. (2011). “The brain basis of language processing: from structure to function”. In: *Physiological Reviews* 91, pp. 1357–1392. DOI: 10.1152/physrev.00006.2011.

Friederici, A. D., Bahlmann, J., Friedrich, R., and Makuuchi, M. (2011a). “The neural basis of recursion and complex syntactic hierarchy”. In: *Biolinguistics* 5.1–2, pp. 87–104. DOI: 10.5964/bioling.8833.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., and Anwander, A. (2006). “The brain differentiates human and non-human grammars: functional localization and structural connectivity”. In: *Proceedings of the National Academy of Sciences* 103.7, pp. 2458–2463. DOI: 10.1073/pnas.0509389103.

Friederici, A. D., Chomsky, N., Berwick, R. C., Moro, A., and Bolhuis, J. J. (2017). “Language, mind and brain”. In: *Nature Human Behaviour* 1, pp. 713–722. DOI: 10.1038/s41562-017-0184-4.

Friederici, A. D., Mueller, J. L., and Oberecker, R. (2011b). “Precursors to natural grammar learning: preliminary evidence from 4-month-old infants”. In: *PLoS ONE* 6.3, e17920. DOI: 10.1371/journal.pone.0017920.

Friederici, A. D., Mueller, J. L., Sehm, B., and Ragert, P. (2013). “Language learning without control: the role of the PFC”. In: *Journal of Cognitive Neuroscience* 25.5, pp. 814–821. DOI: 10.1162/jocn_a_00350.

Friederici, A. D., Steinhauer, K., and Pfeifer, E. (2002). “Brain signatures of artificial language processing: evidence challenging the critical period hypothesis”. In: *Proceedings of the National Academy of Sciences* 99.1, pp. 529–534. DOI: 10.1073/pnas.012611199.

Friston, K. (2005). “A theory of cortical responses”. In: *Philosophical Transactions of the Royal Society B* 360, pp. 815–836. DOI: 10.1098/rstb.2005.1622.

Friston, K. (2009). "The free-energy principle: a rough guide to the brain?" In: *Trends in Cognitive Sciences* 13.7, pp. 293–301. DOI: 10.1016/j.tics.2009.04.005.

Friston, K. (2010). "The free-energy principle: a unified brain theory?" In: *Nature Reviews Neuroscience* 11, pp. 127–138. DOI: 10.1038/nrn2787.

Friston, K. J., Rosch, R., Parr, T., Price, C., and Bowman, H. (2018). "Deep temporal models and active inference". In: *Neuroscience and Biobehavioral Reviews* 90, pp. 486–501. DOI: 10.1016/j.neubiorev.2018.04.004.

Froemke, R. C. and Dan, Y. (2002). "Spike-timing-dependent synaptic modification induced by natural spike trains". In: *Nature* 416, pp. 433–438. DOI: 10.1038/416433a.

Frost, R. L. A., Jessop, A., Durrant, S., Peter, M. S., Bidgood, A., Pine, J. M., Rowland, C. F., and Monaghan, P. (2020). "Non-adjacent dependency learning in infancy, and its link to language development". In: *Cognitive Psychology* 120, p. 101291. DOI: 10.1016/j.cogpsych.2020.101291.

Frost, R. L. A. and Monaghan, P. (2016). "Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech". In: *Cognition* 147, pp. 70–74. DOI: 10.1016/j.cognition.2015.11.010.

Frost, R. L. A. and Monaghan, P. (2020). "Insights from studying statistical learning". In: *Current Perspectives on Child Language Acquisition: How children use their environment to learn.* Ed. by C. F. Rowland, A. L. Theakston, B. Ambridge, and K. E. Twomey. John Benjamins Publishing Company, pp. 65–89. DOI: 10.1075/tilar.27.03fro.

Frost, R., Armstrong, B. C., and Christiansen, M. H. (2019). "Statistical learning research: a critical review and possible new directions". In: *Psychological Bulletin* 145.12, pp. 1128–1153. DOI: 10.1037/bul0000210.

Frost, R., Armstrong, B. C., Siegelman, N., and Christiansen, M. H. (2015). "Domain generality versus modality specificity: the paradox of statistical learning". In: *Trends in Cognitive Sciences* 19.3, pp. 117–125. DOI: 10.1016/j.tics.2014.12.010.

Fuhrmann, G., Segev, I., Markram, H., and Tsodyks, M. (2002). "Coding of temporal information by activity-dependent synapses". In: *Journal of Neurophysiology* 87, pp. 140–148. DOI: 10.1152/jn.00258.2001.

Fukino, M., Katori, Y., and Aihara, K. (2016). "A computational model for pitch pattern perception with the echo state network". In: *International Symposium on Nonlinear Theory and Its Applications, NOLTA2016*, pp. 271–274.

Furber, S. B. (2016). "Brain-inspired computing". In: *IET Computers & Digital Techniques* 10.6, pp. 299–305. DOI: 10.1049/iet-cdt.2015.0171.

Fusi, S. (2002). "Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates". In: *Biological Cybernetics* 87, pp. 459–470. DOI: 10.1007/s00422-002-0356-8.

Gallego, G., Delbrück, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., Leutenegger, S., Davison, A. J., Conradt, J., Daniilidis, K., and Scaramuzza, D. (2022). "Event-based vision: a survey". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.1, pp. 154–180. DOI: 10.1109/TPAMI.2020.3008413.

Gallinaro, J. V., Gašparović, N., and Rotter, S. (2022). "Homeostatic control of synaptic rewiring in recurrent networks induces the formation of stable memory engrams". In: *PLoS Computational Biology* 18.2, e1009836. DOI: 10.1371/journal.pcbi.1009836.

Garagnani, M., Lucchese, G., Tomasello, R., Wennekers, T., and Pulvermüller, F. (2017). "A spiking neurocomputational model of high-frequency oscillatory brain responses to words and pseudowords". In: *Frontiers in Computational Neuroscience* 10.145, pp. 1–19. DOI: 10.3389/fncom.2016.00145.

Garagnani, M., Wennekers, T., and Pulvermüller, F. (2008). "A neuroanatomically grounded Hebbian-learning model of attention–language interactions in the human brain". In: *European Journal of Neuroscience* 27, pp. 492–513. DOI: 10.1111/j.1460-9568.2008.06015.x.

García-Rosales, F., Beetz, M. J., Cabral-Calderin, Y., Kössl, M., and Hechavarria, J. C. (2018). "Neuronal coding of multiscale temporal features in communication sequences within the bat auditory cortex". In: *Communications Biology* 1.200, pp. 1–14. DOI: 10.1038/s42003-018-0205-5.

Garrido, M. I., Kilner, J. M., Stephan, K. E., and Friston, K. J. (2009). "The mismatch negativity: a review of underlying mechanisms". In: *Clinical Neurophysiology* 120, pp. 453–463. DOI: 10.1016/j.clinph.2008.11.029.

Gauthier, D. J., Bollt, E., Griffith, A., and Barbosa, W. A. S. (2021). "Next generation reservoir computing". In: *Nature Communications* 12.5564, pp. 1–8. DOI: 10.1038/s41467-021-25801-2.

Gebhart, A. L., Newport, E. L., and Aslin, R. N. (2009). "Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds". In: *Psychonomic Bulletin & Review* 16.3, pp. 486–490. DOI: 10.3758/PBR.16.3.486.

Gerstner, W., Kreiter, A. K., Markram, H., and Herz, A. V. M. (1997). "Neural codes: firing rates and beyond". In: *Proceedings of the National Academy of Sciences* 94, pp. 12740–12741. DOI: 10.1073/pnas.94.24.12740.

Gerstner, W., Lehmann, M., Liakoni, V., Corneil, D., and Brea, J. (2018). "Eligibility traces and plasticity on behavioral time scales: experimental support of neoHebbian three-factor learning rules". In: *Frontiers in Neural Circuits* 12.53, pp. 1–16. DOI: 10.3389/fncir.2018.00053.

Gerstner, W., Ritz, R., and van Hemmen, J. L (1993). "Why spikes? Hebbian learning and retrieval of time-resolved excitation patterns". In: *Biological Cybernetics* 69, pp. 503–515. DOI: 10.1007/BF00199450.

Gerstner, W., Sprekeler, H., and Deco, G. (2012). "Theory and simulation in neuroscience". In: *Science* 338, pp. 60–65. DOI: 10.1126/science.1227356.

Geschwind, N. (1970). “The organization of language and the brain”. In: *Science* 170, pp. 940–944. DOI: 10.1126/science.170.3961.940.

Gilson, M., Burkitt, A., and van Hemmen, J. L (2010). “STDP in recurrent neuronal networks”. In: *Frontiers in Computational Neuroscience* 4.23, pp. 1–15. DOI: 10.3389/fncom.2010.00023.

Giraud, A.-L. and Poeppel, D. (2012). “Cortical oscillations and speech processing: emerging computational principles and operations”. In: *Nature Neuroscience* 15.4, pp. 511–517. DOI: 10.1038/nn.3063.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., and Van Essen, D. C. (2016). “A multi-modal parcellation of human cerebral cortex”. In: *Nature* 536, pp. 171–178. DOI: 10.1038/nature18933.

Gómez, R. L. (2002). “Variability and detection of invariant structure”. In: *Psychological Science* 13.5, pp. 431–436. DOI: 10.1111/1467-9280.00476.

Gómez, R. L. and Gerken, L. (2000). “Infant artificial language learning and language acquisition”. In: *Trends in Cognitive Sciences* 4.5, pp. 178–186. DOI: 10.1016/S1364-6613(00)01467-4.

Gómez, R. and Maye, J. (2005). “The developmental trajectory of nonadjacent dependency learning”. In: *Infancy* 7.2, pp. 183–206. DOI: 10.1207/s15327078in0702_4.

Goodman, D. F. M. and Brette, R. (2009). “The Brian simulator”. In: *Frontiers in Neuroscience* 3.2, pp. 192–197. DOI: 10.3389/neuro.01.026.2009.

Goodman, D. and Brette, R. (2008). “Brian: a simulator for spiking neural networks in Python”. In: *Frontiers in Neuroinformatics* 2.5, pp. 1–10. DOI: 10.3389/neuro.11.005.2008.

Goodman, N. D. and Stuhlmüller, A. (2013). “Knowledge and implicature: modeling language understanding as social cognition”. In: *Topics in Cognitive Science* 5, pp. 173–184. DOI: 10.1111/tops.12007.

Gottlieb, J. and Oudeyer, P.-Y. (2018). “Towards a neuroscience of active sampling and curiosity”. In: *Nature Reviews Neuroscience* 19, pp. 758–770. DOI: 10.1038/s41583-018-0078-0.

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., and Baranes, A. (2013). “Information-seeking, curiosity, and attention: computational and neural mechanisms”. In: *Trends in Cognitive Sciences* 17.11, pp. 585–593. DOI: 10.1016/j.tics.2013.09.001.

Grama, I. C., Kerkhoff, A., and Wijnen, F. (2016). “Gleaning structure from sound: the role of prosodic contrast in learning non-adjacent dependencies”. In: *Journal of Psycholinguistic Research* 45, pp. 1427–1449. DOI: 10.1007/s10936-016-9412-8.

Graupner, M. and Brunel, N. (2010). “Mechanisms of induction and maintenance of spike-timing dependent plasticity in biophysical synapse models”. In: *Frontiers in Computational Neuroscience* 4.136, pp. 1–19. DOI: 10.3389/fncom.2010.00136.

Graupner, M. and Brunel, N. (2012a). “Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location”. In: *Proceedings of the National Academy of Sciences* 109.10, pp. 3991–3996. DOI: 10.1073/pnas.1109359109.

Graupner, M. and Brunel, N. (2012b). “Correction for Graupner and Brunel, Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location”. In: *Proceedings of the National Academy of Sciences* 109.52, pp. 21551–21552. DOI: 10.1073/pnas.1220044110.

Griffiths, T. L., Lieder, F., and Goodman, N. D. (2015). “Rational use of cognitive resources: levels of analysis between the computational and the algorithmic”. In: *Topics in Cognitive Science* 7, pp. 217–229. DOI: 10.1111/tops.12142.

Guo, W., Fouda, M. E., Eltawil, A. M., and Salama, K. N. (2021). “Neural coding in spiking neural networks: a comparative study for robust neuromorphic systems”. In: *Frontiers in Neuroscience* 15.638474, pp. 1–21. DOI: 10.3389/fnins.2021.638474.

Gütig, R., Aharonov, R., Rotter, S., and Sompolinsky, H. (2003). “Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity”. In: *The Journal of Neuroscience* 23.9, pp. 3697–3714. DOI: 10.1523/JNEUROSCI.23-09-03697.2003.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., and Sporns, O. (2008). “Mapping the structural core of human cerebral cortex”. In: *PLoS Biology* 6.7, e159. DOI: 10.1371/journal.pbio.0060159.

Hamilton, T. J., Afshar, S., van Schaik, A., and Tapson, J. (2014). “Stochastic electronics: a neuro-inspired design paradigm for integrated circuits”. In: *Proceedings of the IEEE* 102.5, pp. 843–859. DOI: 10.1109/JPROC.2014.2310713.

Han, B. and Roy, K. (2020). “Deep spiking neural network: energy efficiency through time based coding”. In: *Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science, vol 12355.* Ed. by Vedaldi A., Bischof H., Brox T., and Frahm J.-M. Springer, Cham, pp. 388–404. DOI: 10.1007/978-3-030-58607-2_23.

Harnad, S. (1990). “The symbol grounding problem”. In: *Physica D: Nonlinear Phenomena* 42, pp. 335–346. DOI: 10.1016/0167-2789(90)90087-6.

Harris, K. D., Groh, J. M., DiCarlo, J., Fries, P., Kaschube, M., Laurent, G., MacLean, J. N., McCormick, D. A., Pipa, G., Reynolds, J. H., Schwartz, A. B., Sejnowski, T. J., Singer, W., and Vinck, M. (2019). “Functional properties of circuits, cellular populations, and areas”. In: *The Neocortex.* Ed. by W. Singer, T. J. Sejnowski, and P. Rakic. Cambridge, MA: MIT Press, pp. 223–272.

Harrison, P. M. C., Bianco, R., Chait, M., and Pearce, M. T. (2020). “PPM-decay: a computational model of auditory prediction with memory decay”. In: *PLoS Computational Biology* 16.11, e1008304. DOI: 10.1371/journal.pcbi.1008304.

Hartmann, C., Lazar, A., Nessler, B., and Triesch, J. (2015). “Where’s the noise? Key features of spontaneous activity and neural variability arise through learning in a deterministic network”. In: *PLoS Computational Biology* 11.12, e1004640. DOI: 10.1371/journal.pcbi.1004640.

Hasson, U., Nastase, S. A., and Goldstein, A. (2020). “Direct fit to nature: an evolutionary perspective on biological and artificial neural networks”. In: *Neuron* 105, pp. 416–434. DOI: 10.1016/j.neuron.2019.12.002.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). “The faculty of language: what is it, who has it, and how did it evolve?” In: *Science* 298, pp. 1569–1579. DOI: 10.1126/science.298.5598.1569.

Hawkins, J. and Ahmad, S. (2016). “Why neurons have thousands of synapses, a theory of sequence memory in neocortex”. In: *Frontiers in Neural Circuits* 10.23, pp. 1–13. DOI: 10.3389/fncir.2016.00023.

He, C., Hotson, L., and Trainor, L. J. (2007). “Mismatch responses to pitch changes in early infancy”. In: *Journal of Cognitive Neuroscience* 19.5, pp. 878–892. DOI: 10.1162/jocn.2007.19.5.878.

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory.* 1st Edition 2002. Routledge Taylor & Francis Group, Psychology Press. DOI: 10.4324/9781410612403.

Heiberg, T., Kriener, B., Tetzlaff, T., Einevoll, G. T., and Plesser, H. E. (2018). “Firing-rate models for neurons with a broad repertoire of spiking behaviors”. In: *Journal of Computational Neuroscience* 45, pp. 103–132. DOI: 10.1007/s10827-018-0693-9.

Heinrich, S. and Wermter, S. (2018). “Interactive natural language acquisition in a multi-modal recurrent neural architecture”. In: *Connection Science* 30.1, pp. 99–133. DOI: 10.1080/09540091.2017.1318357.

Heusser, A. C., Poeppel, D., Ezzyat, Y., and Davachi, L. (2016). “Episodic sequence memory is supported by a theta–gamma phase code”. In: *Nature Neuroscience* 19.10, pp. 1374–1380. DOI: 10.1038/nn.4374.

Hickok, G. and Poeppel, D. (2007). “The cortical organization of speech processing”. In: *Nature Reviews Neuroscience* 8, pp. 393–402. DOI: 10.1038/nrn2113.

Hinaut, X. and Dominey, P. F. (2013). “Real-time parallel processing of grammatical structure in the fronto-striatal system: a recurrent network simulation study using reservoir computing”. In: *PLoS ONE* 8.2, e52946. DOI: 10.1371/journal.pone.0052946.

Hinaut, X., Lance, F., Droin, C., Petit, M., Pointeau, G., and Dominey, P. F. (2015). “Corticostriatal response selection in sentence production: insights from neural network simulation with reservoir computing”. In: *Brain & Language* 150, pp. 54–68. DOI: 10.1016/j.bandl.2015.08.002.

Hintzman, D. L. (1984). “MINERVA 2: a simulation model of human memory”. In: *Behavior Research Methods, Instruments, & Computers* 16.2, pp. 96–101. DOI: 10.3758/BF03202365.

Hintzman, D. L. and Ludlam, G. (1980). “Differential forgetting of prototypes and old instances: simulation by an exemplar-based classification model”. In: *Memory & Cognition* 8.4, pp. 378–382. DOI: 10.3758/BF03198278.

Hoerzer, G. M., Legenstein, R., and Maass, W. (2014). “Emergence of complex computational structures from chaotic neural networks through reward-modulated Hebbian learning”. In: *Cerebral Cortex* 24, pp. 677–690. DOI: 10.1093/cercor/bhs348.

Höhle, B., Schmitz, M., Santelmann, L. M., and Weissenborn, J. (2006). “The recognition of discontinuous verbal dependencies by German 19-month-olds: evidence for lexical and structural influences on children’s early processing capacities”. In: *Language Learning and Development* 2.4, pp. 277–300. DOI: 10.1207/s15473341lld0204_3.

Holmberg, M., Gelbart, D., Ramacher, U., and Hemmert, W. (2005). “Automatic speech recognition with neural spike trains”. In: *INTERSPEECH 2005*, pp. 1253–1256.

Holmberg, M. and Hemmert, W. (2004). “An auditory model for coding speech into nerve-action potentials”. In: *CFA/DAGA’04, Strasbourg*, pp. 773–774.

Hopfield, J. J. (1982). “Neural networks and physical systems with emergent collective computational abilities”. In: *Proceedings of the National Academy of Sciences* 79, pp. 2554–2558. DOI: 10.1073/pnas.79.8.2554.

Hopfield, J. J. and Tank, D. W. (1986). “Computing with neural circuits: a model”. In: *Science* 233, pp. 625–633. DOI: 10.1126/science.3755256.

Horwitz, B., Tagamets, M-A., and McIntosh, A. R. (1999). “Neural modeling, functional brain imaging, and cognition”. In: *Trends in Cognitive Sciences* 3.3, pp. 91–98. DOI: 10.1016/S1364-6613(99)01282-6.

Houdé, O. and Tzourio-Mazoyer, N. (2003). “Neural foundations of logical and mathematical cognition”. In: *Nature Reviews Neuroscience* 4, pp. 507–514. DOI: 10.1038/nrn1117.

Hromádka, T., DeWeese, M. R., and Zador, A. M. (2008). “Sparse representation of sounds in the unanesthetized auditory cortex”. In: *PLoS Biology* 6.1, e16. DOI: 10.1371/journal.pbio.0060016.

Hu, J., Tang, H., Tan, K. C., and Li, H. (2016). “How the brain formulates memory: a spatio-temporal model”. In: *IEEE Computational Intelligence Magazine* 11.2, pp. 56–68. DOI: 10.1109/MCI.2016.2532268.

Huang, Y. and Rao, R. P. N. (2011). “Predictive coding”. In: *WIREs Cognitive Science* 2, pp. 580–593. DOI: 10.1002/wcs.142.

Hull, C. (2020). “Prediction signals in the cerebellum: beyond supervised motor learning”. In: *eLife* 9, e54073. DOI: 10.7554/eLife.54073.

Hupkes, D., Dankers, V., Mul, M., and Bruni, E. (2020). “Compositionality decomposed: how do neural networks generalise?” In: *Journal of Artificial Intelligence Research* 67, pp. 757–795. DOI: 10.1613/jair.1.11674.

Huyck, C. R. and Passmore, P. J. (2013). “A review of cell assemblies”. In: *Biological Cybernetics* 107, pp. 263–288. DOI: 10.1007/s00422-013-0555-5.

Iao, L.-S., Roeser, J., Justice, L., and Jones, G. (2021). “Concurrent visual learning of adjacent and nonadjacent dependencies in adults and children”. In: *Developmental Psychology* 57.5, pp. 733–748. DOI: 10.1037/dev0000998.

Ielmini, D., Wang, Z., and Liu, Y. (2021). “Brain-inspired computing via memory device physics”. In: *APL Materials* 9, p. 050702. DOI: 10.1063/5.0047641.

Indiveri, G., Chicca, E., and Douglas, R. J. (2009). “Artificial cognitive systems: from VLSI networks of spiking neurons to neuromorphic cognition”. In: *Cognitive Computation* 1, pp. 119–127. DOI: 10.1007/s12559-008-9003-6.

Indiveri, G. and Douglas, R. (2014). “Neuromorphic cognition”. In: *Encyclopedia of Computational Neuroscience.* Ed. by Jaeger D. and Jung R. Springer, New York, NY. DOI: 10.1007/978-1-4614-7320-6_113-1.

Indiveri, G. and Horiuchi, T. K. (2011). “Frontiers in neuromorphic engineering”. In: *Frontiers in Neuroscience* 5.118, pp. 1–2. DOI: 10.3389/fnins.2011.00118.

Isbilen, E. S., Frost, R. L. A., Monaghan, P., and Christiansen, M. H. (2022). “Statistically based chunking of nonadjacent dependencies”. In: *Journal of Experimental Psychology: General*, pp. 1–18. DOI: 10.1037/xge0001207.

Isbilen, E. S., McCauley, S. M., Kidd, E., and Christiansen, M. H. (2020). “Statistically induced chunking recall: a memory-based approach to statistical learning”. In: *Cognitive Science* 44, e12848. DOI: 10.1111/cogs.12848.

Izhikevich, E. M. (2003). “Simple model of spiking neurons”. In: *IEEE Transactions on Neural Networks* 14.6, pp. 1569–1572. DOI: 10.1109/TNN.2003.820440.

Izhikevich, E. M. (2004). “Which model to use for cortical spiking neurons?” In: *IEEE Transactions on Neural Networks* 15.5, pp. 1063–1070. DOI: 10.1109/TNN.2004.832719.

Izhikevich, E. M. (2006). “Polychronization: computation with spikes”. In: *Neural Computation* 18, pp. 245–282. DOI: 10.1162/089976606775093882.

Izhikevich, E. M. and Desai, N. S. (2003). “Relating STDP to BCM”. In: *Neural Computation* 15, pp. 1511–1523. DOI: 10.1162/089976603321891783.

Izhikevich, E. M. and Edelman, G. M. (2008). “Large-scale model of mammalian thalamocortical systems”. In: *Proceedings of the National Academy of Sciences* 105.9, pp. 3593–3598. DOI: 10.1073/pnas.0712231105.

Izhikevich, E. M., Gally, J. A., and Edelman, G. M. (2004). “Spike-timing dynamics of neuronal groups”. In: *Cerebral Cortex* 14.8, pp. 933–944. DOI: 10.1093/cercor/bhh053.

Jackendoff, R. (2011). “What is the human language faculty? Two views”. In: *Language* 87.3, pp. 586–624. DOI: 10.1353/lan.2011.0063.

Jäger, G. and Rogers, J. (2012). “Formal language theory: refining the Chomsky hierarchy”. In: *Philosophical Transactions of the Royal Society B* 367, pp. 1956–1970. DOI: 10.1098/rstb.2012.0077.

Jia, S., Zhang, T., Cheng, X., Liu, H., and Xu, B. (2021). “Neuronal-plasticity and reward-propagation improved recurrent spiking neural networks”. In: *Frontiers in Neuroscience* 15.654786, pp. 1–11. DOI: 10.3389/fnins.2021.654786.

Jiang, X., Long, T., Cao, W., Li, J., Dehaene, S., and Wang, L. (2018). “Production of supra-regular spatial sequences by macaque monkeys”. In: *Current Biology* 28, 1851–1859.e4. DOI: 10.1016/j.cub.2018.04.047.

Jiménez-Fernández, A., Cerezuela-Escudero, E., Miró-Amarante, L., Domínguez-Morales, M. J., de Asís Gómez-Rodríguez, F., Linares-Barranco, A., and Jiménez-Moreno, G. (2017). “A binaural neuromorphic auditory sensor for FPGA: a spike signal processing approach”. In: *IEEE Transactions on Neural Networks and Learning Systems* 28.4, pp. 804–818. DOI: 10.1109/TNNLS.2016.2583223.

Johnson, A. J., Cauchi, L., and Miles, C. (2013). “Hebbian learning for olfactory sequences”. In: *Quarterly Journal of Experimental Psychology.* 66.6, pp. 1082–1089. DOI: 10.1080/17470218.2012.729068.

Johnson, E. K. and Jusczyk, P. W. (2001). “Word segmentation by 8-month-olds: when speech cues count more than statistics”. In: *Journal of Memory and Language* 44, pp. 548–567. DOI: 10.1006/jmla.2000.2755.

Jones, G. and Macken, B. (2018). “Long-term associative learning predicts verbal short-term memory performance”. In: *Memory & Cognition* 46, pp. 216–229. DOI: 10.3758/s13421-017-0759-3.

Ju, H., Dranias, M. R., Banumurthy, G., and VanDongen, A. M. J. (2015). “Spatiotemporal memory is an intrinsic property of networks of dissociated cortical neurons”. In: *The Journal of Neuroscience* 35.9, pp. 4040–4051. DOI: 10.1523/JNEUROSCI.3793-14.2015.

Judaš, M., Cepanec, M., and Sedmak, G. (2012). “Brodmann’s map of the human cerebral cortex – or Brodmann’s maps?” In: *Translational Neuroscience* 3.1, pp. 67–74. DOI: 10.2478/s13380-012-0009-x.

Jürgensen, A.-M., Khalili, A., Chicca, E., Indiveri, G., and Nawrot, M. P. (2021). “A neuromorphic model of olfactory processing and sparse coding in the Drosophila larva brain”. In: *Neuromorphic Computing and Engineering* 1, p. 024008. DOI: 10.1088/2634-4386/ac3ba6.

Jusczyk, P. W. (2002). “How infants adapt speech-processing capacities to native-language structure”. In: *Current Directions in Psychological Science* 11.1, pp. 15–18. DOI: 10.1111/1467-8721.00159.

Jusczyk, P. W. and Luce, P. A. (2002). “Speech perception and spoken word recognition: past and present”. In: *Ear & Hearing* 23.1, pp. 2–40.

Juven, A. and Hinaut, X. (2020). “Cross-situational learning with reservoir computing for language acquisition modelling”. In: *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. DOI: 10.1109/IJCNN48605.2020.9206650.

Kabdebon, C., Pena, M., Buiatti, M., and Dehaene-Lambertz, G. (2015). “Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants”. In: *Brain & Language* 148, pp. 25–36. DOI: 10.1016/j.bandl.2015.03.005.

Kaelbling, L. P., Littman, M. L., and Moore, A. W. (1996). “Reinforcement learning: a survey”. In: *Journal of Artificial Intelligence Research* 4, pp. 237–285. DOI: 10.1613/jair.301.

Kaiser, J., Mostafa, H., and Neftci, E. (2020). “Synaptic plasticity dynamics for deep continuous local learning (DECOLLE)”. In: *Frontiers in Neuroscience* 14.424, pp. 1–11. DOI: 10.3389/fnins.2020.00424.

Kalisman, N., Silberberg, G., and Markram, H. (2005). “The neocortical microcircuit as a tabula rasa”. In: *Proceedings of the National Academy of Sciences* 102.3, pp. 880–885. DOI: 10.1073/pnas.0407088102.

Kanari, L., Dictus, H., Chalimourda, A., Arnaudon, A., Van Geit, W., Coste, B., Shillcock, J., Hess, K., and Markram, H. (2022). “Computational synthesis of cortical dendritic morphologies”. In: *Cell Reports* 39, p. 110586. DOI: 10.1016/j.celrep.2022.110586.

Kandel, E. R., Markram, H., Matthews, P. M., Yuste, R., and Koch, C. (2013). “Neuroscience thinks big (and collaboratively)”. In: *Nature Reviews Neuroscience* 14, pp. 659–664. DOI: 10.1038/nrn3578.

Kass, R. E., Amari, S.-I., Arai, K., Brown, E. N., Diekman, C. O., Diesmann, M., Doiron, B., Eden, U. T., Fairhall, A. L., Fiddyment, G. M., Fukai, T., Grün, S., Harrison, M. T., Helias, M., Nakahara, H., Teramae, J.-n., Thomas, P. J., Reimers, M., Rodu, J., Rotstein, H. G., Shea-Brown, E., Shimazaki, H., Shinomoto, S., Yu, B. M., and Kramer, M. A. (2018). “Computational neuroscience: mathematical and statistical perspectives”. In: *Annual Review of Statistics and Its Application* 5, pp. 183–214. DOI: 10.1146/annurev-statistics-041715-033733.

Kaufeld, G., Bosker, H. R., ten Oever, S., Alday, P. M., Meyer, A. S., and Martin, A. E. (2020). “Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy”. In: *The Journal of Neuroscience* 40.49, pp. 9467–9475. DOI: 10.1523/JNEUROSCI.0302-20.2020.

Kayser, C., Ince, R. A. A., and Panzeri, S. (2012). “Analysis of slow (theta) oscillations as a potential temporal reference frame for information coding in sensory cortices”. In: *PLoS Computational Biology* 8.10, e1002717. DOI: 10.1371/journal.pcbi.1002717.

Kayser, C., Montemurro, M. A., Logothetis, N. K., and Panzeri, S. (2009). “Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns”. In: *Neuron* 61, pp. 597–608. DOI: 10.1016/j.neuron.2009.01.008.

Keck, T., Toyoizumi, T., Chen, L., Doiron, B., Feldman, D. E., Fox, K., Gerstner, W., Haydon, P. G., Hübener, M., Lee, H.-K., Lisman, J. E., Rose, T., Sengpiel, F., Stellwagen, D., Stryker, M. P., Turrigiano, G. G., and van Rossum, M. C. (2017). “Integrating Hebbian and homeostatic plasticity: the current state of the field and future research directions”. In: *Philosophical Transactions of the Royal Society B* 372, p. 20160158. DOI: 10.1098/rstb.2016.0158.

Keeley, S., Byrne, Á., Fenton, A., and Rinzel, J. (2019). “Firing rate models for gamma oscillations”. In: *Journal of Neurophysiology* 121, pp. 2181–2190. DOI: 10.1152/jn.00741.2018.

Keitel, A., Gross, J., and Kayser, C. (2018). “Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features”. In: *PLoS Biology* 16.3, e2004473. DOI: 10.1371/journal.pbio.2004473.

Kelley, T. D. (2003). “Symbolic and sub-symbolic representations in computational models of human cognition: what can be learned from biology?” In: *Theory & Psychology* 13.6, pp. 847–860. DOI: 10.1177/0959354303136005.

Kelly, M. A. and West, R. L. (2012). “From vectors to symbols to cognition: the symbolic and sub-symbolic aspects of vector-symbolic cognitive models”. In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 34, pp. 1768–1773.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., and Gerken, L. (1995). “The head-turn preference procedure for testing auditory perception”. In: *Infant Behavior and Development* 18, pp. 111–116. DOI: 10.1016/0163-6383(95)90012-8.

Kendall, J. D. and Kumar, S. (2020). “The building blocks of a brain-inspired computer”. In: *Applied Physics Reviews* 7, p. 011305. DOI: 10.1063/1.5129306.

Kepecs, A., van Rossum, M. C. W., Song, S., and Tegner, J. (2002). “Spike-timing-dependent plasticity: common themes and divergent vistas”. In: *Biological Cybernetics* 87, pp. 446–458. DOI: 10.1007/s00422-002-0358-6.

Kershenbaum, A., Bowles, A. E., Freeberg, T. M., Jin, D. Z., Lameira, A. R., and Bohn, K. (2014). “Animal vocal sequences: not the Markov chains we thought they were”. In: *Proceedings of the Royal Society B* 281, p. 20141370. DOI: 10.1098/rspb.2014.1370.

Khaligh-Razavi, S.-M. and Kriegeskorte, N. (2014). “Deep supervised, but not unsupervised, models may explain IT cortical representation”. In: *PLoS Computational Biology* 10.11, e1003915. DOI: 10.1371/journal.pcbi.1003915.

Kidd, E. (2012). "Implicit statistical learning is directly associated with the acquisition of syntax". In: *Developmental Psychology* 48.1, pp. 171–184. DOI: 10.1037/a0025405.

Kidd, E., Arciuli, J., Christiansen, M. H., Isbilen, E. S., Revius, K., and Smithson, M. (2020). "Measuring children's auditory statistical learning via serial recall". In: *Journal of Experimental Child Psychology* 200, p. 104964. DOI: 10.1016/j.jecp.2020.104964.

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). "A hierarchy of time-scales and the brain". In: *PLoS Computational Biology* 4.11, e1000209. DOI: 10.1371/journal.pcbi.1000209.

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., and Kriegeskorte, N. (2019). "Recurrence is required to capture the representational dynamics of the human visual system". In: *Proceedings of the National Academy of Sciences* 116.43, pp. 21854–21863. DOI: 10.1073/pnas.1905544116.

Kiggins, J. T., Comins, J. A., and Gentner, T. Q. (2012). "Targets for a comparative neurobiology of language". In: *Frontiers in Evolutionary Neuroscience* 4.6, pp. 1–13. DOI: 10.3389/fnevo.2012.00006.

Kim, S. J., Kim, S., and Jang, H. W. (2021). "Competing memristors for brain-inspired computing". In: *iScience* 24, p. 101889. DOI: 10.1016/j.isci.2020.101889.

Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). "Visual statistical learning in infancy: evidence for a domain general learning mechanism". In: *Cognition* 83, B35–B42. DOI: 10.1016/S0010-0277(02)00004-5.

Klampfl, S. and Maass, W. (2013). "Emergence of dynamic memory traces in cortical microcircuit models through STDP". In: *The Journal of Neuroscience* 33.28, pp. 11515–11529. DOI: 10.1523/JNEUROSCI.5044-12.2013.

Kloppenburg, P. and Nawrot, M. P. (2014). "Neural coding: sparse but on time". In: *Current Biology* 24.19, R957–R959. DOI: 10.1016/j.cub.2014.08.041.

Klos, C., Kossio, Y. F. K., Goedeke, S., Gilra, A., and Memmesheimer, R.-M. (2020). "Dynamical learning of dynamics". In: *Physical Review Letters* 125, p. 088103. DOI: 10.1103/PhysRevLett.125.088103.

Klos, C., Miner, D., and Triesch, J. (2018). "Bridging structure and function: a model of sequence learning and prediction in primary visual cortex". In: *PLoS Computational Biology* 14.6, e1006187. DOI: 10.1371/journal.pcbi.1006187.

Knight, J. C., Tully, P. J., Kaplan, B. A., Lansner, A., and Furber, S. B. (2016). "Large-scale simulations of plastic neural networks on neuromorphic hardware". In: *Frontiers in Neuroanatomy* 10.37, pp. 1–17. DOI: 10.3389/fnana.2016.00037.

Knudsen, E. I. (1994). "Supervised learning in the brain". In: *The Journal of Neuroscience* 14.7, pp. 3985–3997. DOI: 10.1523/JNEUROSCI.14-07-03985.1994.

Koch, G., Ponzo, V., Di Lorenzo, F., Caltagirone, C., and Veniero, D. (2013). "Hebbian and anti-Hebbian spike-timing-dependent plasticity of human cortico-cortical connections". In: *The Journal of Neuroscience* 33.23, pp. 9725–9733. DOI: 10.1523/JNEUROSCI.4988-12.2013.

Koechlin, E., Ody, C., and Kouneiher, F. (2003). "The architecture of cognitive control in the human prefrontal cortex". In: *Science* 302, pp. 1181–1185. DOI: 10.1126/science.1088545.

König, P., Engel, A. K., and Singer, W. (1996). "Integrator or coincidence detector? The role of the cortical neuron revisited". In: *Trends in Neurosciences* 19.4, pp. 130–137. DOI: 10.1016/S0166-2236(96)80019-1.

König, P., Kühnberger, K.-U., and Kietzmann, T. C. (2013). "A unifying approach to high- and low-level cognition". In: *Models, Simulations, and the Reduction of Complexity*. Ed. by U. Gähde, S. Hartmann, and J. H. Wolf. Berlin, Boston: De Gruyter, pp. 117–140. DOI: 10.1515/9783110313680.117.

Kossio, Y. F. K., Goedeke, S., Klos, C., and Memmesheimer, R.-M. (2021). "Drifting assemblies for persistent memory: neuron transitions and unsupervised compensation". In: *Proceedings of the National Academy of Sciences* 118.46, e2023832118. DOI: 10.1073/pnas.2023832118.

Koulakov, A. A., Hromádka, T., and Zador, A. M. (2009). "Correlated connectivity and the distribution of firing rates in the neocortex". In: *The Journal of Neuroscience* 29.12, pp. 3685–3694. DOI: 10.1523/JNEUROSCI.4500-08.2009.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., and Poeppel, D. (2017). "Neuroscience needs behavior: correcting a reductionist bias". In: *Neuron* 93, pp. 480–490. DOI: 10.1016/j.neuron.2016.12.041.

Kriegeskorte, N. and Douglas, P. K. (2018). "Cognitive computational neuroscience". In: *Nature Neuroscience* 21, pp. 1148–1160. DOI: 10.1038/s41593-018-0210-5.

Kriegeskorte, N. and Golan, T. (2019). "Neural network models and deep learning". In: *Current Biology* 29, R231–R236. DOI: 10.1016/j.cub.2019.02.034.

Kriete, T., Noelle, D. C., Cohen, J. D., and O'Reilly, R. C. (2013). "Indirection and symbol-like processing in the prefrontal cortex and basal ganglia". In: *Proceedings of the National Academy of Sciences* 110.41, pp. 16390–16395. DOI: 10.1073/pnas.1303547110.

Kudela, P., Boatman-Reich, D., Beeman, D., and Anderson, W. S. (2018). "Modeling neural adaptation in auditory cortex". In: *Frontiers in Neural Circuits* 12.72, pp. 1–11. DOI: 10.3389/fncir.2018.00072.

Kugele, A., Pfeil, T., Pfeiffer, M., and Chicca, E. (2020). "Efficient processing of spatio-temporal data streams with spiking neural networks". In: *Frontiers in Neuroscience* 14.439, pp. 1–13. DOI: 10.3389/fnins.2020.00439.

Kuhl, P. K. and Miller, J. D. (1975). "Speech perception by the Chinchilla: voiced-voiceless distinction in alveolar plosive consonants". In: *Science* 190, pp. 69–72. DOI: 10.1126/science.1166301.

Kuhl, P. and Rivera-Gaxiola, M. (2008). “Neural substrates of language acquisition”. In: *Annual Review of Neuroscience* 31, pp. 511–534. DOI: 10.1146/annurev.neuro.30.051606.094321.

Kühn, T. and Helias, M. (2017). “Locking of correlated neural activity to ongoing oscillations”. In: *PLoS Computational Biology* 13.6, e1005534. DOI: 10.1371/journal.pcbi.1005534.

Kunkel, S., Diesmann, M., and Morrison, A. (2011). “Limits to the development of feed-forward structures in large recurrent neuronal networks”. In: *Frontiers in Computational Neuroscience* 4.160, pp. 1–15. DOI: 10.3389/fncom.2010.00160.

La Camera, G., Rauch, A., Thurbon, D., Lüscher, H.-R., Senn, W., and Fusi, S. (2006). “Multiple time scales of temporal response in pyramidal and fast spiking cortical neurons”. In: *Journal of Neurophysiology* 96, pp. 3448–3464. DOI: 10.1152/jn.00453.2006.

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., and Schroeder, C. E. (2005). “An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex”. In: *Journal of Neurophysiology* 94, pp. 1904–1911. DOI: 10.1152/jn.00263.2005.

Lakretz, Y., Hupkes, D., Vergallito, A., Marelli, M., Baroni, M., and Dehaene, S. (2021). “Mechanisms for handling nested dependencies in neural-network language models and humans”. In: *arXiv* 2006.11098v2, pp. 1–70. DOI: 10.48550/arXiv.2006.11098.

Lampinen, J. and Vehtari, A. (2001). “Bayesian approach for neural networks – review and case studies”. In: *Neural Networks* 14, pp. 257–274. DOI: 10.1016/S0893-6080(00)00098-8.

Lang, E. A., van Geen, C., Tedeschi, E., Marvin, C. B., and Shohamy, D. (2022). “Learned temporal statistics guide information seeking and shape memory”. In: *Journal of Experimental Psychology: General* 151.5, pp. 986–995. DOI: 10.1037/xge0001122.

Lansner, A. (2009). “Associative memory models: from the cell-assembly theory to biophysically detailed cortex simulations”. In: *Trends in Neurosciences* 32.3, pp. 178–186. DOI: 10.1016/j.tins.2008.12.002.

Lany, J. and Gómez, R. L. (2008). “Twelve-month-old infants benefit from prior experience in statistical learning”. In: *Psychological Science* 19.12, pp. 1247–1252. DOI: 10.1111/j.1467-9280.2008.02233.x.

Lattenkamp, E. Z. and Vernes, S. C. (2018). “Vocal learning: a language-relevant trait in need of a broad cross-species approach”. In: *Current Opinion in Behavioral Sciences* 21, pp. 209–215. DOI: 10.1016/j.cobeha.2018.04.007.

Lau, E. F., Phillips, C., and Poeppel, D. (2008). “A cortical network for semantics: (de)constructing the N400”. In: *Nature Reviews Neuroscience* 9, pp. 920–933. DOI: 10.1038/nrn2532.

Lazar, A., Pipa, G., and Triesch, J. (2007). “Fading memory and time series prediction in recurrent networks with different forms of plasticity”. In: *Neural Networks* 20, pp. 312–322. DOI: 10.1016/j.neunet.2007.04.020.

Lazar, A., Pipa, G., and Triesch, J. (2009). “SORN: a self-organizing recurrent neural network”. In: *Frontiers in Computational Neuroscience* 3.23, pp. 1–9. DOI: 10.3389/neuro.10.023.2009.

Lazar, A., Pipa, G., and Triesch, J. (2011). “Emerging Bayesian priors in a self-organizing recurrent network”. In: *Artificial Neural Networks and Machine Learning – ICANN 2011. ICANN 2011. Lecture Notes in Computer Science, vol 6792.* Ed. by T. Honkela, W. Duch, M. Girolami, and S. Kaski. Springer, Berlin, Heidelberg, pp. 127–134. DOI: 10.1007/978-3-642-21738-8_17.

Lazartigues, L., Mathy, F., and Lavigne, F. (2021). “Statistical learning of unbalanced exclusive-or temporal sequences in humans”. In: *PLoS ONE* 16.2, e0246826. DOI: 10.1371/journal.pone.0246826.

Le Bé, J.-V. and Markram, H. (2006). “Spontaneous and evoked synaptic rewiring in the neonatal neocortex”. In: *Proceedings of the National Academy of Sciences* 103.35, pp. 13214–13219. DOI: 10.1073/pnas.0604691103.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). “Deep learning”. In: *Nature* 521, pp. 436–444. DOI: 10.1038/nature14539.

Lee, J. H., Delbruck, T., and Pfeiffer, M. (2016). “Training deep spiking neural networks using backpropagation”. In: *Frontiers in Neuroscience* 10.508, pp. 1–13. DOI: 10.3389/fnins.2016.00508.

Lehfeldt, S., Mueller, J. L., and Pipa, G. “Spike-based statistical learning explains human grammar acquisition performance (manuscript)”. In: *n.d.*

Leugering, J. and Pipa, G. (2018). “A unifying framework of synaptic and intrinsic plasticity in neural populations”. In: *Neural Computation* 30, pp. 945–986. DOI: 10.1162/NECO_a_01057.

Levelt, W. J. M. (2018). “Is language natural to man? Some historical considerations”. In: *Current Opinion in Behavioral Sciences* 21, pp. 127–131. DOI: 10.1016/j.cobeha.2018.04.003.

Lewis, C. M., Ni, J., Wunderle, T., Jendritza, P., Lazar, A., Diester, I., and Fries, P. (2021). “Cortical gamma-band resonance preferentially transmits coherent input”. In: *Cell Reports* 35, p. 109083. DOI: 10.1016/j.celrep.2021.109083.

Lewis, J. D. and Elman, J. L. (2002). “Learnability and the statistical structure of language: poverty of stimulus arguments revisited”. In: *Proceedings of the 26th Annual Boston University Conference on Language Development.*

Liang, D. and Indiveri, G. (2019). “A neuromorphic computational primitive for robust context-dependent decision making and context-dependent stochastic computation”. In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 66.5, pp. 843–847. DOI: 10.1109/TCSII.2019.2907848.

Liang, Q., Zeng, Y., and Xu, B. (2020). “Temporal-sequential learning with a brain-inspired spiking neural network and its application to musical memory”. In: *Frontiers in Computational Neuroscience* 14.51, pp. 1–17. DOI: 10.3389/fncom.2020.00051.

Lidz, J. and Gagliardi, A. (2015). “How nature meets nurture: universal grammar and statistical learning”. In: *Annual Review of Linguistics* 1, pp. 333–353. DOI: 10.1146/annurev-linguist-030514-125236.

Lieder, F., Stephan, K. E., Daunizeau, J., Garrido, M. I., and Friston, K. J. (2013). “A neurocomputational model of the mismatch negativity”. In: *PLoS Computational Biology* 9.11, e1003288. DOI: 10.1371/journal.pcbi.1003288.

Litwin-Kumar, A. and Doiron, B. (2014). “Formation and maintenance of neuronal assemblies through synaptic plasticity”. In: *Nature Communications* 5.5319, pp. 1–12. DOI: 10.1038/ncomms6319.

Liu, S.-C., Mesgarani, N., Harris, J., and Hermansky, H. (2010). “The use of spike-based representations for hardware audition systems”. In: *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, pp. 505–508. DOI: 10.1109/ISCAS.2010.5537588.

Locke, J. L. (1989). “Babbling and early speech: continuity and individual differences”. In: *First Language* 9, pp. 191–205. DOI: 10.1177/014272378900900606.

Löffler, H. and Gupta, D. S. (2020). “A model of memory linking time to space”. In: *Frontiers in Computational Neuroscience* 14.60, pp. 1–11. DOI: 10.3389/fncom.2020.00060.

Löffler, H. and Gupta, D. S. (2022). “A model of pattern separation by single neurons”. In: *Frontiers in Computational Neuroscience* 16.858353, pp. 1–13. DOI: 10.3389/fncom.2022.858353.

Love, B. C. (2015). “The algorithmic level is the bridge between computation and brain”. In: *Topics in Cognitive Science* 7, pp. 230–242. DOI: 10.1111/tops.12131.

Lu, H. S. and Mintz, T. H. (2021). “Learning non-adjacent rules and non-adjacent dependencies from human actions in 9-month-old infants”. In: *PLoS ONE* 16.6, e0252959. DOI: 10.1371/journal.pone.0252959.

Lukoševičius, M. and Jaeger, H. (2009). “Reservoir computing approaches to recurrent neural network training”. In: *Computer Science Review* 3, pp. 127–149. DOI: 10.1016/j.cosrev.2009.03.005.

Luo, H. and Poeppel, D. (2007). “Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex”. In: *Neuron* 54, pp. 1001–1010. DOI: 10.1016/j.neuron.2007.06.004.

Ma, C., Xu, J., and Yu, Q. (2021). “Temporal dependent local learning for deep spiking neural networks”. In: *2021 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7. DOI: 10.1109/IJCNN52387.2021.9534390.

Maass, W. (1997). “Networks of spiking neurons: the third generation of neural network models”. In: *Neural Networks* 10.9, pp. 1659–1671. DOI: 10.1016/S0893-6080(97)00011-7.

Maass, W. (2000). “On the computational power of winner-take-all”. In: *Neural Computation* 12, pp. 2519–2535. DOI: 10.1162/089976600300014827.

Maass, W. (2016). “Searching for principles of brain computation”. In: *Current Opinion in Behavioral Sciences* 11, pp. 81–92. DOI: 10.1016/j.cobeha.2016.06.003.

Maass, W. and Markram, H. (2002). “Synapses as dynamic memory buffers”. In: *Neural Networks* 15, pp. 155–161. DOI: 10.1016/S0893-6080(01)00144-7.

Maass, W., Natschläger, T., and Markram, H. (2002). “Real-time computing without stable states: a new framework for neural computation based on perturbations”. In: *Neural Computation* 14, pp. 2531–2560. DOI: 10.1162/089976602760407955.

Machens, C. K., Wehr, M. S., and Zador, A. M. (2004). “Linearity of cortical receptive fields measured with natural sounds”. In: *The Journal of Neuroscience* 24.5, pp. 1089–1100. DOI: 10.1523/JNEUROSCI.4445-03.2004.

Macpherson, T., Churchland, A., Sejnowski, T., DiCarlo, J., Kamitani, Y., Takahashi, H., and Hikida, T. (2021). “Natural and artificial intelligence: a brief introduction to the interplay between AI and neuroscience research”. In: *Neural Networks* 144, pp. 603–613. DOI: 10.1016/j.neunet.2021.09.018.

Maes, A., Barahona, M., and Clopath, C. (2020). “Learning spatiotemporal signals using a recurrent spiking network that discretizes time”. In: *PLoS Computational Biology* 16.1, e1007606. DOI: 10.1371/journal.pcbi.1007606.

Maes, A., Barahona, M., and Clopath, C. (2021). “Learning compositional sequences with multiple time scales through a hierarchical network of spiking neurons”. In: *PLoS Computational Biology* 17.3, e1008866. DOI: 10.1371/journal.pcbi.1008866.

Makuuchi, M., Bahlmann, J., Anwander, A., and Friederici, A. D. (2009). “Segregating the core computational faculty of human language from working memory”. In: *Proceedings of the National Academy of Sciences* 106.20, pp. 8362–8367. DOI: 10.1073/pnas.0810928106.

Mandel, D. R., Jusczyk, P. W., and Pisoni, D. B. (1995). “Infants’ recognition of the sound patterns of their own names”. In: *Psychological Science* 6.5, pp. 314–317. DOI: 10.1111/j.1467-9280.1995.tb00517.x.

Männel, C. and Friederici, A. D. (2009). “Pauses and intonational phrasing: ERP studies in 5-month-old German infants and adults”. In: *Journal of Cognitive Neuroscience* 21.10, pp. 1988–2006. DOI: 10.1162/jocn.2009.21221.

Marcus, G. F., Vijayan, S., Bandi Rao, S., and Vishton, P. M. (1999). “Rule learning by seven-month-old infants”. In: *Science* 283, pp. 77–80. DOI: 10.1126/science.283.5398.77.

Mareschal, D. and French, R. M. (2017). “TRACX2: a connectionist autoencoder using graded chunks to model infant visual statistical learning”. In: *Philosophical*

Transactions of the Royal Society B 372, p. 20160057. DOI: 10.1098/rstb.2016.0057.

Markram, H. (2006). “The Blue Brain Project”. In: *Nature Reviews Neuroscience* 7, pp. 153–160. DOI: 10.1038/nrn1848.

Markram, H. (2013). “Seven challenges for neuroscience”. In: *Functional Neurology* 28.3, pp. 145–151.

Markram, H., Gerstner, W., and Sjöström, P. J. (2011). “A history of spike-timing-dependent plasticity”. In: *Frontiers in Synaptic Neuroscience* 3.4, pp. 1–24. DOI: 10.3389/fnsyn.2011.00004.

Markram, H., Gerstner, W., and Sjöström, P. J. (2012). “Spike-timing-dependent plasticity: a comprehensive overview”. In: *Frontiers in Synaptic Neuroscience* 4.2, pp. 1–3. DOI: 10.3389/fnsyn.2012.00002.

Markram, H., Gupta, A., Uziel, A., Wang, Y., and Tsodyks, M. (1998a). “Information processing with frequency-dependent synaptic connections”. In: *Neurobiology of Learning and Memory* 70, pp. 101–112. DOI: 10.1006/nlme.1998.3841.

Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). “Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs”. In: *Science* 275, pp. 213–215. DOI: 10.1126/science.275.5297.213.

Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., Kahou, G. A. A., Berger, T. K., Bilgili, A., Buncic, N., Chalimourda, A., Chindemi, G., Courcol, J.-D., Delalondre, F., Delattre, V., Druckmann, S., Dumusc, R., Dynes, J., Eilemann, S., Gal, E., Gevaert, M. E., Ghobril, J.-P., Gidon, A., Graham, J. W., Gupta, A., Haenel, V., Hay, E., Heinis, T., Hernando, J. B., Hines, M., Kanari, L., Keller, D., Kenyon, J., Khazen, G., Kim, Y., King, J. G., Kisvarday, Z., Kumbhar, P., Lasserre, S., Le Bé, J.-V., Magalhães, B. R. C., Merchán-Pérez, A., Meystre, J., Morrice, B. R., Muller, J., Muñoz-Céspedes, A., Muralidhar, S., Muthurasa, K., Nachbaur, D., Newton, T. H., Nolte, M., Ovcharenko, A., Palacios, J., Pastor, L., Perin, R., Ranjan, R., Riachi, I., Rodríguez, J.-R., Riquelme, J. L., Rössert, C., Sfyrakis, K., Shi, Y., Shillcock, J. C., Silberberg, G., Silva, R., Tauheed, F., Telefont, M., Toledo-Rodriguez, M., Tränkler, T., Van Geit, W., Díaz, J. V., Walker, R., Wang, Y., Zaninetta, S. M., DeFelipe, J., Hill, S. L., Segev, I., and Schürmann, F. (2015). “Reconstruction and simulation of neocortical microcircuitry”. In: *Cell* 163, pp. 456–492. DOI: 10.1016/j.cell.2015.09.029.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004). “Interneurons of the neocortical inhibitory system”. In: *Nature Reviews Neuroscience* 5, pp. 793–807. DOI: 10.1038/nrn1519.

Markram, H. and Tsodyks, M. (1996). “Redistribution of synaptic efficacy between neocortical pyramidal neurons”. In: *Nature* 382, pp. 807–810. DOI: 10.1038/382807a0.

Markram, H., Wang, Y., and Tsodyks, M. (1998b). “Differential signaling via the same axon of neocortical pyramidal neurons”. In: *Proceedings of the National Academy of Sciences* 95, pp. 5323–5328. DOI: 10.1073/pnas.95.9.5323.

Martin, A. E. and Doumas, L. A. A. (2017). “A mechanism for the cortical computation of hierarchical linguistic structure”. In: *PLoS Biology* 15.3, e2000663. DOI: 10.1371/journal.pbio.2000663.

Martin, A., Onishi, K. H., and Vouloumanos, A. (2012). “Understanding the abstract role of speech in communication at 12 months”. In: *Cognition* 123, pp. 50–60. DOI: 10.1016/j.cognition.2011.12.003.

Martin, S. J., Grimwood, P. D., and Morris, R. G. M. (2000). “Synaptic plasticity and memory: an evaluation of the hypothesis”. In: *Annual Review of Neuroscience* 23, pp. 649–711. DOI: 10.1146/annurev.neuro.23.1.649.

Martinez, R. H., Lansner, A., and Herman, P. (2019). “Probabilistic associative learning suffices for learning the temporal structure of multiple sequences”. In: *PLoS ONE* 14.8, e0220161. DOI: 10.1371/journal.pone.0220161.

Martinez-Alvarez, A., Benavides-Varela, S., Lapillonne, A., and Gervain, J. (2022). “Newborns discriminate utterance-level prosodic contours”. In: *Developmental Science*, e13304. DOI: 10.1111/desc.13304.

Masquelier, T., Guyonneau, R., and Thorpe, S. J. (2008). “Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains”. In: *PLoS ONE* 3.1, e1377. DOI: 10.1371/journal.pone.0001377.

Mastrogiuseppe, F., Hiratani, N., and Latham, P. (2022). “Evolution of neural activity in circuits bridging sensory and abstract knowledge”. In: *bioRxiv*, pp. 1–61. DOI: 10.1101/2022.01.29.478317.

May, P. J. C. and Tiitinen, H. (2010). “Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained”. In: *Psychophysiology* 47, pp. 66–122. DOI: 10.1111/j.1469-8986.2009.00856.x.

McClelland, J. L. and Rumelhart, D. E. (1985). “Distributed memory and the representation of general and specific information”. In: *Journal of Experimental Psychology: General* 114.2, pp. 159–188. DOI: 10.1037/0096-3445.114.2.159.

McClelland, J. L., St. John, M., and Taraban, R. (1989). “Sentence comprehension: a parallel distributed processing approach”. In: *Language and Cognitive Processes* 4.3-4, SI287–SI335. DOI: 10.1080/01690968908406371.

Mead, C. (1990). “Neuromorphic electronic systems”. In: *Proceedings of the IEEE* 78.10, pp. 1629–1636. DOI: 10.1109/5.58356.

Mead, C. A. and Mahowald, M. A. (1988). “A silicon model of early visual processing”. In: *Neural Networks* 1, pp. 91–97. DOI: 10.1016/0893-6080(88)90024-X.

Mehonic, A. and Kenyon, A. J. (2022). “Brain-inspired computing needs a master plan”. In: *Nature* 604, pp. 255–260. DOI: 10.1038/s41586-021-04362-w.

Melamed, O., Gerstner, W., Maass, W., Tsodyks, M., and Markram, H. (2004). "Coding and learning of behavioral sequences". In: *Trends in Neurosciences* 27.1, pp. 11–14. DOI: 10.1016/j.tins.2003.10.014.

Mesulam, M.-M. (1998). "From sensation to cognition". In: *Brain* 121, pp. 1013–1052. DOI: 10.1093/brain/121.6.1013.

Mi, Y., Katkov, M., and Tsodyks, M. (2017). "Synaptic correlates of working memory capacity". In: *Neuron* 93, pp. 323–330. DOI: 10.1016/j.neuron.2016.12.004.

Migliore, M., Cannia, C., Lytton, W. W., Markram, H., and Hines, M. L. (2006). "Parallel network simulations with NEURON". In: *Journal of Computational Neuroscience* 21, pp. 119–129. DOI: 10.1007/s10827-006-7949-5.

Mill, R., Coath, M., Wennekers, T., and Denham, S. L. (2011). "A neurocomputational model of stimulus-specific adaptation to oddball and Markov sequences". In: *PLoS Computational Biology* 7.8, e1002117. DOI: 10.1371/journal.pcbi.1002117.

Miller, E. K. and Cohen, J. D. (2001). "An integrative theory of prefrontal cortex function". In: *Annual Review of Neuroscience* 24, pp. 167–202. DOI: 10.1146/annurev.neuro.24.1.167.

Milner, B., Squire, L. R., and Kandel, E. R. (1998). "Cognitive neuroscience and the study of memory". In: *Neuron* 20, pp. 445–468. DOI: 10.1016/s0896-6273(00)80987-3.

Milo, V., Chicca, E., and Ielmini, D. (2018). "Brain-inspired recurrent neural network with plastic RRAM synapses". In: *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5. DOI: 10.1109/ISCAS.2018.8351523.

Minai, U., Gustafson, K., Fiorentino, R., Jongman, A., and Sereno, J. (2017). "Fetal rhythm-based language discrimination: a biomagnetometry study". In: *NeuroReport* 28.10, pp. 561–564. DOI: 10.1097/WNR.0000000000000794.

Miner, D. and Triesch, J. (2016). "Plasticity-driven self-organization under topological constraints accounts for non-random features of cortical synaptic wiring". In: *PLoS Computational Biology* 12.2, e1004759. DOI: 10.1371/journal.pcbi.1004759.

Mintz, T. H. (2003). "Frequent frames as a cue for grammatical categories in child directed speech". In: *Cognition* 90, pp. 91–117. DOI: 10.1016/S0010-0277(03)00140-9.

Mishra, R. K., Kim, S., Guzman, S. J., and Jonas, P. (2016). "Symmetric spike timing-dependent plasticity at CA3–CA3 synapses optimizes storage and recall in autoassociative networks". In: *Nature Communications* 7.11552, pp. 1–11. DOI: 10.1038/ncomms11552.

Mitra, S., Fusi, S., and Indiveri, G. (2009). "Real-time classification of complex patterns using spike-based learning in neuromorphic VLSI". In: *IEEE Transactions on Biomedical Circuits and Systems* 3.1, pp. 32–42. DOI: 10.1109/TBCAS.2008.2005781.

Mitropolsky, D., Collins, M. J., and Papadimitriou, C. H. (2021). “A biologically plausible parser”. In: *Transactions of the Association for Computational Linguistics* 9, pp. 1374–1388.

Monroe, D. (2014). “Neuromorphic computing gets ready for the (really) big time”. In: *Communications of the ACM* 57.6, pp. 13–15. DOI: 10.1145/2601069.

Morgan-Short, K., Steinhauer, K., Sanz, C., and Ullman, M. T. (2012). “Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns”. In: *Journal of Cognitive Neuroscience* 24.4, pp. 933–947. DOI: 10.1162/jocn_a_00119.

Morrison, A., Diesmann, M., and Gerstner, W. (2008). “Phenomenological models of synaptic plasticity based on spike timing”. In: *Biological Cybernetics* 98, pp. 459–478. DOI: 10.1007/s00422-008-0233-1.

Mueller, J. L., Bahlmann, J., and Friederici, A. D. (2010). “Learnability of embedded syntactic structures depends on prosodic cues”. In: *Cognitive Science* 34, pp. 338–349. DOI: 10.1111/j.1551-6709.2009.01093.x.

Mueller, J. L., Friederici, A. D., and Männel, C. (2012). “Auditory perception at the root of language learning”. In: *Proceedings of the National Academy of Sciences* 109.39, pp. 15953–15958. DOI: 10.1073/pnas.1204319109.

Mueller, J. L., Friederici, A. D., and Männel, C. (2019). “Developmental changes in automatic rule-learning mechanisms across early childhood”. In: *Developmental Science* 22, e12700. DOI: 10.1111/desc.12700.

Mueller, J. L., Milne, A., and Männel, C. (2018a). “Non-adjacent auditory sequence learning across development and primate species”. In: *Current Opinion in Behavioral Sciences* 21, pp. 112–119. DOI: 10.1016/j.cobeha.2018.04.002.

Mueller, J. L., Oberecker, R., and Friederici, A. D. (2009). “Syntactic learning by mere exposure: an ERP study in adult learners”. In: *BMC Neuroscience* 10.89. DOI: 10.1186/1471-2202-10-89.

Mueller, J. L., ten Cate, C., and Toro, J. M. (2018b). “A comparative perspective on the role of acoustic cues in detecting language structure”. In: *Topics in Cognitive Science*, pp. 1–16. DOI: 10.1111/tops.12373.

Müller, M. G., Papadimitriou, C. H., Maass, W., and Legenstein, R. (2020). “A model for structured information representation in neural networks of the brain”. In: *eNeuro* 7.3, pp. 1–17. DOI: 10.1523/ENEURO.0533-19.2020.

Munkong, R. and Juang, B.-H. (2008). “Auditory perception and cognition”. In: *IEEE Signal Processing Magazine* 25.3, pp. 98–117. DOI: 10.1109/MSP.2008.918418.

Näätänen, R., Jacobsen, T., and Winkler, I. (2005). “Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence”. In: *Psychophysiology* 42, pp. 25–32. DOI: 10.1111/j.1469-8986.2005.00256.x.

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). "The mismatch negativity (MMN) in basic research of central auditory processing: a review". In: *Clinical Neurophysiology* 118, pp. 2544–2590. DOI: 10.1016/j.clinph.2007.04.026.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., and Winkler, I. (2001). "'Primitive intelligence' in the auditory cortex". In: *Trends in Neurosciences* 24.5, pp. 283–288. DOI: 10.1016/S0166-2236(00)01790-2.

Näätänen, R. and Winkler, I. (1999). "The concept of auditory stimulus representation in cognitive neuroscience". In: *Psychological Bulletin* 125.6, pp. 826–859. DOI: 10.1037/0033-2909.125.6.826.

Nádasdy, Z. (2000). "Spike sequences and their consequences". In: *Journal of Physiology-Paris* 94, pp. 505–524. DOI: 10.1016/S0928-4257(00)01103-7.

Narayan, R., Graña, G., and Sen, K. (2006). "Distinct time scales in cortical discrimination of natural sounds in songbirds". In: *Journal of Neurophysiology* 96, pp. 252–258. DOI: 10.1152/jn.01257.2005.

Natschläger, T., Markram, H., and Maass, W. (2003). "Computer models and analysis tools for neural microcircuits". In: *Neuroscience Databases*. Ed. by R. Kötter. Springer, Boston, MA. DOI: 10.1007/978-1-4615-1079-6_9.

Nazzi, T., Bertoncini, J., and Mehler, J. (1998). "Language discrimination by newborns: toward an understanding of the role of rhythm". In: *Journal of Experimental Psychology: Human Perception and Performance* 24.3, pp. 756–766.

Neftci, E. O. (2018). "Data and power efficient intelligence with neuromorphic learning machines". In: *iScience* 5, pp. 52–68. DOI: 10.1016/j.isci.2018.06.010.

Neftci, E. O. and Averbeck, B. B. (2019). "Reinforcement learning in artificial and biological systems". In: *Nature Machine Intelligence* 1, pp. 133–143. DOI: 10.1038/s42256-019-0025-4.

Neftci, E., Binas, J., Rutishauser, U., Chicca, E., Indiveri, G., and Douglas, R. J. (2013). "Synthesizing cognition in neuromorphic electronic systems". In: *Proceedings of the National Academy of Sciences* 110.37, E3468–E3476. DOI: 10.1073/pnas.1212083110.

Neil, D., Pfeiffer, M., and Liu, S.-C. (2016). "Learning to be efficient: algorithms for training low-latency, low-compute deep spiking neural networks". In: *Proceedings of the 31st Annual ACM Symposium on Applied Computing (SAC '16)*. Association for Computing Machinery, pp. 293–298. DOI: 10.1145/2851613.2851724.

Nespor, M., Peña, M., and Mehler, J. (2003). "On the different roles of vowels and consonants in speech processing and language acquisition". In: *Lingue e linguaggio* 2, pp. 203–230. DOI: 10.1418/10879.

Newport, E. L. and Aslin, R. N. (2004). "Learning at a distance I. Statistical learning of non-adjacent dependencies". In: *Cognitive Psychology* 48, pp. 127–162. DOI: 10.1016/S0010-0285(03)00128-2.

Ng, B. S. W., Logothetis, N. K., and Kayser, C. (2013). “EEG phase patterns reflect the selectivity of neural firing”. In: *Cerebral Cortex* 23, pp. 389–398. DOI: 10.1093/cercor/bhs031.

Niv, Y. (2009). “Reinforcement learning in the brain”. In: *Journal of Mathematical Psychology* 53, pp. 139–154. DOI: 10.1016/j.jmp.2008.12.005.

Nolte, M., Reimann, M. W., King, J. G., Markram, H., and Muller, E. B. (2019). “Cortical reliability amid noise and chaos”. In: *Nature Communications* 10.3792, pp. 1–15. DOI: 10.1038/s41467-019-11633-8.

Norman-Haignere, S. V., Long, L. K., Devinsky, O., Doyle, W., Irobunda, I., Merricks, E. M., Feldstein, N. A., McKhann, G. M., Schevon, C. A., Flinker, A., and Mesgarani, N. (2022). “Multiscale temporal integration organizes hierarchical computation in human auditory cortex”. In: *Nature Human Behaviour* 6, pp. 455–469. DOI: 10.1038/s41562-021-01261-y.

Novick, J. M., Trueswell, J. C., and Thompson-Schill, S. L. (2005). “Cognitive control and parsing: reexamining the role of Broca’s area in sentence comprehension”. In: *Cognitive, Affective, & Behavioral Neuroscience* 5.3, pp. 263–281. DOI: 10.3758/CABN.5.3.263.

Novick, J. M., Trueswell, J. C., and Thompson-Schill, S. L. (2010). “Broca’s area and language processing: evidence for the cognitive control connection”. In: *Language and Linguistics Compass* 4.10, pp. 906–924. DOI: 10.1111/j.1749-818x.2010.00244.x.

O’Connor, P. and Welling, M. (2016). “Deep spiking networks”. In: *arXiv* 1602.08323v2, pp. 1–16. DOI: 10.48550/arXiv.1602.08323.

O’Kane, J. C. and Goldbart, J. (1998). “Re-examining the development of early communication”. In: *Communication Before Speech: Development and Assessment.* David Fulton Publishers, pp. 1–9. DOI: 10.4324/9780203462157.

Oller, D. K., Wieman, L. A., Doyle, W. J., and Ross, C. (1976). “Infant babbling and speech”. In: *Journal of Child Language* 3, pp. 1–11. DOI: 10.1017/S0305000900001276.

Olshausen, B. A. and Field, D. J. (2004). “Sparse coding of sensory inputs”. In: *Current Opinion in Neurobiology* 14, pp. 481–487. DOI: 10.1016/j.conb.2004.07.007.

Onnis, L., Monaghan, P., Richmond, K., and Chater, N. (2005). “Phonology impacts segmentation in online speech processing”. In: *Journal of Memory and Language* 53, pp. 225–237. DOI: 10.1016/j.jml.2005.02.011.

O’Reilly, R. C. (1998). “Six principles for biologically based computational models of cortical cognition”. In: *Trends in Cognitive Sciences* 2.11, pp. 455–462. DOI: 10.1016/S1364-6613(98)01241-8.

Orpella, J., Mas-Herrero, E., Ripollés, P., Marco-Pallarés, J., and de Diego-Balaguer, R. (2021). “Language statistical learning responds to reinforcement learning principles rooted in the striatum". In: *PLoS Biology* 19.9, e3001119. DOI: 10.1371/journal.pbio.3001119.

Osborn, L. E., Dragomir, A., Betthauser, J. L., Hunt, C. L., Nguyen, H. H., Kaliki, R. R., and Thakor, N. V. (2018). "Prosthesis with neuromorphic multilayered e-dermis perceives touch and pain". In: *Science Robotics* 3, eaat3818. DOI: 10.1126/scirobotics.aat3818.

Ostojic, S. (2014). "Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons". In: *Nature Neuroscience* 17.4, pp. 594–600. DOI: 10.1038/nn.3658.

Oudeyer, P.-Y. and Smith, L. B. (2016). "How evolution may work through curiosity-driven developmental process". In: *Topics in Cognitive Science* 8, pp. 492–502. DOI: 10.1111/tops.12196.

Pacton, S. and Perruchet, P. (2008). "An attention-based associative account of adjacent and nonadjacent dependency learning". In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34.1, pp. 80–96. DOI: 10.1037/0278-7393.34.1.80.

Palacios, E. R., Isomura, T., Parr, T., and Friston, K. (2019). "The emergence of synchrony in networks of mutually inferring neurons". In: *Scientific Reports* 9.6412, pp. 1–14. DOI: 10.1038/s41598-019-42821-7.

Palm, G. (2016). "Neural information processing in cognition: we start to understand the orchestra, but where is the conductor?" In: *Frontiers in Computational Neuroscience* 10.3, pp. 1–6. DOI: 10.3389/fncom.2016.00003.

Palm, G., Knoblauch, A., Hauser, F., and Schüz, A. (2014). "Cell assemblies in the cerebral cortex". In: *Biological Cybernetics* 108, pp. 559–572. DOI: 10.1007/s00422-014-0596-4.

Panda, P., Allred, J. M., Ramanathan, S., and Roy, K. (2018). "ASP: learning to forget with adaptive synaptic plasticity in spiking neural networks". In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 8.1, pp. 51–64. DOI: 10.1109/JETCAS.2017.2769684.

Panda, P. and Roy, K. (2017). "Learning to generate sequences with combination of Hebbian and non-Hebbian plasticity in recurrent spiking neural networks". In: *Frontiers in Neuroscience* 11.693, pp. 1–11. DOI: 10.3389/fnins.2017.00693.

Panzeri, S., Brunel, N., Logothetis, N. K., and Kayser, C. (2010). "Sensory neural codes using multiplexed temporal scales". In: *Trends in Neurosciences* 33.3, pp. 111–120. DOI: 10.1016/j.tins.2009.12.001.

Panzeri, S., Ince, R. A. A., Diamond, M. E., and Kayser, C. (2014). "Reading spike timing without a clock: intrinsic decoding of spike trains". In: *Philosophical Transactions of the Royal Society B* 369, p. 20120467. DOI: 10.1098/rstb.2012.0467.

Panzeri, S., Macke, J. H., Gross, J., and Kayser, C. (2015). "Neural population coding: combining insights from microscopic and mass signals". In: *Trends in Cognitive Sciences* 19.3, pp. 162–172. DOI: 10.1016/j.tics.2015.01.002.

Papadimitriou, C. H., Vempala, S. S., Mitropolsky, D., Collins, M., and Maass, W. (2020). "Brain computation by assemblies of neurons". In: *Proceedings of the National Academy of Sciences* 117.25, pp. 14464–14472. DOI: 10.1073/pnas.2001893117.

Park, H.-J. and Friston, K. (2013). "Structural and functional brain networks: from connections to cognition". In: *Science* 342, p. 1238411. DOI: 10.1126/science.1238411.

Park, S., Kim, S., Choe, H., and Yoon, S. (2019). "Fast and efficient information transmission with burst spikes in deep spiking neural networks". In: *Proceedings of the 56th Annual Design Automation Conference 2019 (DAC '19)*. Association for Computing Machinery, Article 53, pp. 1–6. DOI: 10.1145/3316781.3317822.

Patiño-Saucedo, A., Rostro-Gonzalez, H., Serrano-Gotarredona, T., and Linares-Barranco, B. (2020). "Event-driven implementation of deep spiking convolutional neural networks for supervised classification using the SpiNNaker neuromorphic platform". In: *Neural Networks* 121, pp. 319–328. DOI: 10.1016/j.neunet.2019.09.008.

Paul, M., Männel, C., van der Kant, A., Mueller, J. L., Höhle, B., Wartenburger, I., and Friederici, A. D. (2021). "Gradual development of non-adjacent dependency learning during early childhood". In: *Developmental Cognitive Neuroscience* 50, p. 100975. DOI: 10.1016/j.dcn.2021.100975.

Pedretti, G., Milo, V., Hashemkhani, S., Mannocci, P., Melnic, O., Chicca, E., and Ielmini, D. (2020). "A spiking recurrent neural network with phase change memory synapses for decision making". In: *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5. DOI: 10.1109/ISCAS45731.2020.9180513.

Peelle, J. E. and Davis, M. H. (2012). "Neural oscillations carry speech rhythm through to comprehension". In: *Frontiers in Psychology* 3.320, pp. 1–17. DOI: 10.3389/fpsyg.2012.00320.

Pelucchi, B., Hay, J. F., and Saffran, J. R. (2009). "Statistical learning in a natural language by 8-month-old infants". In: *Child Development* 80.3, pp. 674–685. DOI: 10.1111/j.1467-8624.2009.01290.x.

Peña, M., Bonatti, L. L., Nespor, M., and Mehler, J. (2002). "Signal-driven computations in speech processing". In: *Science* 298, pp. 604–607. DOI: 10.1126/science.1072901.

Perfors, A., Tenenbaum, J. B., and Regier, T. (2006). "Poverty of the stimulus? A rational approach". In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 28, pp. 663–668.

Perin, R., Berger, T. K., and Markram, H. (2011). “A synaptic organizing principle for cortical neuronal groups”. In: *Proceedings of the National Academy of Sciences* 108.13, pp. 5419–5424. DOI: 10.1073/pnas.1016051108.

Perkins, L. and Lidz, J. (2021). “Eighteen-month-old infants represent nonlocal syntactic dependencies”. In: *Proceedings of the National Academy of Sciences* 118.41, e2026469118. DOI: 10.1073/pnas.2026469118.

Perruchet, P. and Pacton, S. (2006). “Implicit learning and statistical learning: one phenomenon, two approaches”. In: *Trends in Cognitive Sciences* 10.5, pp. 233–238. DOI: 10.1016/j.tics.2006.03.006.

Perruchet, P. and Tillmann, B. (2010). “Exploiting multiple sources of information in learning an artificial language: human data and modeling”. In: *Cognitive Science* 34, pp. 255–285. DOI: 10.1111/j.1551-6709.2009.01074.x.

Perruchet, P., Tyler, M. D., Galland, N., and Peereman, R. (2004). “Learning nonadjacent dependencies: no need for algebraic-like computations”. In: *Journal of Experimental Psychology: General* 133.4, pp. 573–583. DOI: 10.1037/0096-3445.133.4.573.

Perruchet, P. and Vinter, A. (1998). “PARSER: a model for word segmentation”. In: *Journal of Memory and Language* 39, pp. 246–263. DOI: 10.1006/jmla.1998.2576.

Petersen, S. E. and Sporns, O. (2015). “Brain networks and cognitive architectures”. In: *Neuron* 88, pp. 207–219. DOI: 10.1016/j.neuron.2015.09.027.

Petersson, K. M., Grenholm, P., and Forkstam, C. (2005). “Artificial grammar learning and neural networks”. In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 27, pp. 1726–1731.

Petersson, K. M. and Hagoort, P. (2012). “The neurobiology of syntax: beyond string sets”. In: *Philosophical Transactions of the Royal Society B* 367, pp. 1971–1983. DOI: 10.1098/rstb.2012.0101.

Petersson, K.-M., Folia, V., and Hagoort, P. (2012). “What artificial grammar learning reveals about the neurobiology of syntax”. In: *Brain & Language* 120, pp. 83–95. DOI: 10.1016/j.bandl.2010.08.003.

Peyrache, A., Benchenane, K., Khamassi, M., Wiener, S. I., and Battaglia, F. P. (2010). “Principal component analysis of ensemble recordings reveals cell assemblies at high temporal resolution”. In: *Journal of Computational Neuroscience* 29, pp. 309–325. DOI: 10.1007/s10827-009-0154-6.

Pfeiffer, M. and Pfeil, T. (2018). “Deep learning with spiking neurons: opportunities and challenges”. In: *Frontiers in Neuroscience* 12.774, pp. 1–18. DOI: 10.3389/fnins.2018.00774.

Piccinini, G. and Shagrir, O. (2014). “Foundations of computational neuroscience”. In: *Current Opinion in Neurobiology* 25, pp. 25–30. DOI: 10.1016/j.conb.2013.10.005.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli, E. P. (2008). "Spatio-temporal correlations and visual signalling in a complete neuronal population". In: *Nature* 454, pp. 995–999. DOI: 10.1038/nature07140.

Pipa, G. and Munk, M. H. J. (2011). "Higher order spike synchrony in prefrontal cortex during visual memory". In: *Frontiers in Computational Neuroscience* 5.23, pp. 1–13. DOI: 10.3389/fncom.2011.00023.

Pitti, A., Quoy, M., Lavandier, C., and Boucenna, S. (2019). "Feature and structural learning of memory sequences with recurrent and gated spiking neural networks using free-energy: application to speech perception and production II". In: *HAL* hal-02140049.

Plenz, D. and Thiagarajan, T. C. (2007). "The organizing principles of neuronal avalanches: cell assemblies in the cortex?" In: *Trends in Neurosciences* 30.3, pp. 101–110. DOI: 10.1016/j.tins.2007.01.005.

Plotnikov, D., Rumpe, B., Blundell, I., Ippen, T., Eppler, J. M., and Morrison, A. (2016). "NESTML: a modeling language for spiking neurons". In: *Modellierung 2016, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2016*. Ed. by A. Oberweis and R. Reussner. Bonner Köllen Verlag, pp. 93–108.

Poeppel, D. (2003). "The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'". In: *Speech Communication* 41, pp. 245–255. DOI: 10.1016/S0167-6393(02)00107-3.

Poeppel, D. (2012). "The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language". In: *Cognitive Neuropsychology* 29.1-2, pp. 34–55. DOI: 10.1080/02643294.2012.710600.

Poeppel, D. (2014). "The neuroanatomic and neurophysiological infrastructure for speech and language". In: *Current Opinion in Neurobiology* 28, pp. 142–149. DOI: 10.1016/j.conb.2014.07.005.

Poeppel, D. and Embick, D. (2005). "Defining the relation between linguistics and neuroscience". In: *Twenty-First Century Psycholinguistics: Four Cornerstones: Four Cornerstones (1st ed.)* Ed. by A. Cutler. Routledge, pp. 103–118. DOI: 10.4324/9781315084503.

Poeppel, D., Idsardi, W. J., and van Wassenhove, V. (2008). "Speech perception at the interface of neurobiology and linguistics". In: *Philosophical Transactions of the Royal Society B* 363, pp. 1071–1086. DOI: 10.1098/rstb.2007.2160.

Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., Parker, D. S., Sabb, F. W., and Bilder, R. M. (2011). "The cognitive atlas: toward a knowledge foundation for cognitive neuroscience". In: *Frontiers in Neuroinformatics* 5.17, pp. 1–11. DOI: 10.3389/fninf.2011.00017.

Poletiek, F. H., Fitz, H., and Bocanegra, B. R. (2016). "What baboons can (not) tell us about natural language grammars". In: *Cognition* 151, pp. 108–112. DOI: 10.1016/j.cognition.2015.04.016.

Polich, J. (1993). "Cognitive brain potentials". In: *Current Directions in Psychological Science* 2.6, pp. 175–179. DOI: 10.1111/1467-8721.ep10769728.

Poo, M-m., Pignatelli, M., Ryan, T. J., Tonegawa, S., Bonhoeffer, T., Martin, K. C., Rudenko, A., Tsai, L.-H., Tsien, R. W., Fishell, G., Mullins, C., Gonçalves, J. T., Shtrahman, M., Johnston, S. T., Gage, F. H., Dan, Y., Long, J., Buzsáki, G., and Stevens, C. (2016). "What is memory? The present state of the engram". In: *BMC Biology* 14.40, pp. 1–18. DOI: 10.1186/s12915-016-0261-6.

Poon, C.-S. and Zhou, K. (2011). "Neuromorphic silicon neurons and large-scale neural networks: challenges and opportunities". In: *Frontiers in Neuroscience* 5.108, pp. 1–3. DOI: 10.3389/fnins.2011.00108.

Pospischil, M., Toledo-Rodriguez, M., Monier, C., Piwkowska, Z., Bal, T., Frégnac, Y., Markram, H., and Destexhe, A. (2008). "Minimal Hodgkin-Huxley type models for different classes of cortical and thalamic neurons". In: *Biological Cybernetics* 99, pp. 427–441. DOI: 10.1007/s00422-008-0263-8.

Pothos, E. M. (2007). "Theories of artificial grammar learning". In: *Psychological Bulletin* 133.2, pp. 227–244. DOI: 10.1037/0033-2909.133.2.227.

Pothos, E. M. (2010). "An entropy model for artificial grammar learning". In: *Frontiers in Psychology* 1.16, pp. 1–13. DOI: 10.3389/fpsyg.2010.00016.

Prince, L. Y., Eyono, R. H., Boven, E., Ghosh, A., Pemberton, J., Scherr, F., Clopath, C., Costa, R. P., Maass, W., Richards, B. A., Savin, C., and Wilmes, K. A. (2022). "Current state and future directions for learning in biological recurrent neural networks: a perspective piece". In: *Neurons, Behavior, Data analysis, and Theory*, pp. 1–10. DOI: 10.48550/arXiv.2105.05382.

Pullum, G. K. and Scholz, B. C. (2002). "Empirical assessment of stimulus poverty arguments". In: *The Linguistic Review* 19, pp. 9–50. DOI: 10.1515/tlir.19.1-2.9.

Pulvermüller, F. (1996). "Hebb's concept of cell assemblies and the psychophysiology of word processing". In: *Psychophysiology* 33, pp. 317–333. DOI: 10.1111/j.1469-8986.1996.tb01057.x.

Pulvermüller, F., Garagnani, M., and Wennekers, T. (2014). "Thinking in circuits: toward neurobiological explanation in cognitive neuroscience". In: *Biological Cybernetics* 108, pp. 573–593. DOI: 10.1007/s00422-014-0603-9.

Pulvermüller, F. and Knoblauch, A. (2009). "Discrete combinatorial circuits emerging in neural networks: a mechanism for rules of grammar in the human brain?" In: *Neural Networks* 22, pp. 161–172. DOI: 10.1016/j.neunet.2009.01.009.

Pulvermüller, F., Tomasello, R., Henningsen-Schomers, M. R., and Wennekers, T. (2021). "Biological constraints on neural network models of cognitive function". In: *Nature Reviews Neuroscience* 22, pp. 488–502. DOI: 10.1038/s41583-021-00473-5.

Purwins, H., Blankertz, B., and Obermayer, K. (2000). "Computing auditory perception". In: *Organised Sound* 5.3, pp. 159–171. DOI: 10.1017/S1355771800005069.

Qian, N. and Zhang, J. (2020). “Neuronal firing rate as code length: a hypothesis”. In: *Computational Brain & Behavior* 3, pp. 34–53. DOI: 10.1007/s42113-019-00028-z.

Rabovsky, M., Hansen, S. S., and McClelland, J. L. (2018). “Modelling the N400 brain potential as change in a probabilistic representation of meaning”. In: *Nature Human Behaviour* 2, pp. 693–705. DOI: 10.1038/s41562-018-0406-4.

Rankin, J. and Rinzel, J. (2019). “Computational models of auditory perception from feature extraction to stream segregation and behavior”. In: *Current Opinion in Neurobiology* 58, pp. 46–53. DOI: 10.1016/j.conb.2019.06.009.

Rapp, H. and Nawrot, M. P. (2020). “A spiking neural program for sensorimotor control during foraging in flying insects”. In: *Proceedings of the National Academy of Sciences* 117.45, pp. 28412–28421. DOI: 10.1073/pnas.2009821117.

Rapp, H., Nawrot, M. P., and Stern, M. (2020). “Numerical cognition based on precise counting with a single spiking neuron”. In: *iScience* 23, p. 100852. DOI: 10.1016/j.isci.2020.100852.

Rawski, J., Idsardi, W., and Heinz, J. (2021). “Comment on “Nonadjacent dependency processing in monkeys, apes, and humans””. In: *Science Advances* 7, eabg0455. DOI: 10.1126/sciadv.abg0455.

Raymond, J. L. and Medina, J. F. (2018). “Computational principles of supervised learning in the cerebellum”. In: *Annual Review of Neuroscience* 41, pp. 233–253. DOI: 10.1146/annurev-neuro-080317-061948.

Reber, A. S. (1967). “Implicit learning of artificial grammars”. In: *Journal of Verbal Learning and Verbal Behavior* 6, pp. 855–863. DOI: 10.1016/S0022-5371(67)80149-X.

Reilly, R. G. (2001). “Collaborative cell assemblies: building blocks of cortical computation”. In: *Emergent Neural Computational Architectures Based on Neuroscience: Towards Neuroscience-Inspired Computing. Lecture Notes in Computer Science, vol 2036.* Ed. by S. Wermter, J. Austin, and D. Willshaw. Springer, Berlin, Heidelberg, pp. 161–173. DOI: 10.1007/3-540-44597-8_12.

Reimann, M. W., Anastassiou, C. A., Perin, R., Hill, S. L., Markram, H., and Koch, C. (2013). “A biophysically detailed model of neocortical local field potentials predicts the critical role of active membrane currents”. In: *Neuron* 79, pp. 375–390. DOI: 10.1016/j.neuron.2013.05.023.

Reimann, M. W., Horlemann, A.-L., Ramaswamy, S., Muller, E. B., and Markram, H. (2017a). “Morphological diversity strongly constrains synaptic connectivity and plasticity”. In: *Cerebral Cortex* 27, pp. 4570–4585. DOI: 10.1093/cercor/bhx150.

Reimann, M. W., Nolte, M., Scolamiero, M., Turner, K., Perin, R., Chindemi, G., Dłotko, P., Levi, R., Hess, K., and Markram, H. (2017b). “Cliques of neurons bound into cavities provide a missing link between structure and function”. In: *Frontiers in Computational Neuroscience* 11.48, pp. 1–16. DOI: 10.3389/fncom.2017.00048.

Rey, A., Perruchet, P., and Fagot, J. (2012). “Centre-embedded structures are a by-product of associative learning and working memory constraints: evidence from baboons (Papio papio)”. In: *Cognition* 123, pp. 180–184. DOI: 10.1016/j.cognition.2011.12.005.

Richards, B. A. and Frankland, P. W. (2017). “The persistence and transience of memory”. In: *Neuron* 94, pp. 1071–1084. DOI: 10.1016/j.neuron.2017.04.037.

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R. P., de Berker, A., Ganguli, S., Gillon, C. J., Hafner, D., Kepecs, A., Kriegeskorte, N., Latham, P., Lindsay, G. W., Miller, K. D., Naud, R., Pack, C. C., Poirazi, P., Roelfsema, P., Sacramento, J., Saxe, A., Scellier, B., Schapiro, A. C., Senn, W., Wayne, G., Yamins, D., Zenke, F., Zylberberg, J., Therien, D., and Kording, K. P. (2019). “A deep learning framework for neuroscience”. In: *Nature Neuroscience* 22, pp. 1761–1770. DOI: 10.1038/s41593-019-0520-2.

Rohrmeier, M. A. and Cross, I. (2014). “Modelling unsupervised online-learning of artificial grammars: linking implicit and statistical learning”. In: *Consciousness and Cognition* 27, pp. 155–167. DOI: 10.1016/j.concog.2014.03.011.

Rolls, E. T. (2021). “A neuroscience levels of explanation approach to the mind and the brain”. In: *Frontiers in Computational Neuroscience* 15.649679, pp. 1–8. DOI: 10.3389/fncom.2021.649679.

Romberg, A. R. and Saffran, J. R. (2010). “Statistical learning and language acquisition”. In: *WIREs Cognitive Science* 1, pp. 906–914. DOI: 10.1002/wcs.78.

Romberg, A. R. and Saffran, J. R. (2013). “All together now: concurrent learning of multiple structures in an artificial language”. In: *Cognitive Science* 37, pp. 1290–1318. DOI: 10.1111/cogs.12050.

Rongala, U. B., Mazzoni, A., and Oddo, C. M. (2017). “Neuromorphic artificial touch for categorization of naturalistic textures”. In: *IEEE Transactions on Neural Networks and Learning Systems* 28.4, pp. 819–829. DOI: 10.1109/TNNLS.2015.2472477.

Rosen, S. (1992). “Temporal information in speech: acoustic, auditory and linguistic aspects”. In: *Philosophical Transactions of the Royal Society of London. Series B* 336, pp. 367–373. DOI: 10.1098/rstb.1992.0070.

Roy, K., Jaiswal, A., and Panda, P. (2019). “Towards spike-based machine intelligence with neuromorphic computing”. In: *Nature* 575, pp. 607–617. DOI: 10.1038/s41586-019-1677-2.

Rubin, J., Lee, D. D., and Sompolinsky, H. (2001). “Equilibrium properties of temporally asymmetric Hebbian plasticity”. In: *Physical Review Letters* 86.2, pp. 364–367. DOI: 10.1103/PhysRevLett.86.364.

Rueckert, E., Kappel, D., Tanneberg, D., Pecevski, D., and Peters, J. (2016). “Recurrent spiking networks solve planning tasks”. In: *Scientific Reports* 6.21142, pp. 1–10. DOI: 10.1038/srep21142.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). “Learning representations by back-propagating errors”. In: *Nature* 323, pp. 533–536. DOI: 10.1038/323533a0.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996a). “Statistical learning by 8-month-old infants”. In: *Science* 274, pp. 1926–1928. DOI: 10.1126/science.274.5294.1926.

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999). “Statistical learning of tone sequences by human infants and adults”. In: *Cognition* 70, pp. 27–52. DOI: 10.1016/S0010-0277(98)00075-4.

Saffran, J. R. and Kirkham, N. Z. (2018). “Infant statistical learning”. In: *Annual Review of Psychology* 69, pp. 181–203. DOI: 10.1146/annurev-psych-122216-011805.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996b). “Word segmentation: the role of distributional cues”. In: *Journal of Memory and Language* 35, pp. 606–621. DOI: 10.1006/jmla.1996.0032.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., and Barrueco, S. (1997). “Incidental language learning: listening (and learning) out of the corner of your ear”. In: *Psychological Science* 8.2, pp. 101–105. DOI: 10.1111/j.1467-9280.1997.tb00690.x.

Saffran, J. R. and Wilson, D. P. (2003). “From syllables to syntax: multilevel statistical learning by 12-month-old infants”. In: *Infancy* 4.2, pp. 273–284. DOI: 10.1207/S15327078IN0402_07.

Sahasranamam, A., Vlachos, I., Aertsen, A., and Kumar, A. (2016). “Dynamical state of the network determines the efficacy of single neuron properties in shaping the network activity”. In: *Scientific Reports* 6.26029, pp. 1–16. DOI: 10.1038/srep26029.

Sainburg, T., Mai, A., and Gentner, T. Q. (2022). “Long-range sequential dependencies precede complex syntactic production in language acquisition”. In: *Proceedings of the Royal Society B* 289, p. 20212657. DOI: 10.1098/rspb.2021.2657.

Sainburg, T., Theilman, B., Thielk, M., and Gentner, T. Q. (2019). “Parallels in the sequential organization of birdsong and human speech”. In: *Nature Communications* 10.3636, pp. 1–11. DOI: 10.1038/s41467-019-11605-y.

Sakurai, Y. (1999). “How do cell assemblies encode information in the brain?” In: *Neuroscience and Biobehavioral Reviews* 23, pp. 785–796. DOI: 10.1016/S0149-7634(99)00017-2.

Santelmann, L. M. and Jusczyk, P. W. (1998). “Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space”. In: *Cognition* 69, pp. 105–134. DOI: 10.1016/S0010-0277(98)00060-2.

Santolin, C., Garcia-Castro, G., Zettersten, M., Sebastian-Galles, N., and Saffran, J. R. (2020). “Experience with research paradigms relates to infants’ direction of preference”. In: *Infancy* 26, pp. 39–46. DOI: 10.1111/infa.12372.

Saxe, A., Nelli, S., and Summerfield, C. (2021). “If deep learning is the answer, what is the question?” In: *Nature Reviews Neuroscience* 22, pp. 55–67. DOI: 10.1038/s41583-020-00395-8.

Scarpetta, S. and de Candia, A. (2013). “Neural avalanches at the critical point between replay and non-replay of spatiotemporal patterns”. In: *PLoS ONE* 8.6, e64162. DOI: 10.1371/journal.pone.0064162.

Schmitt, L.-M. and Obleser, J. (2022). “What auditory cortex is waiting for”. In: *Nature Human Behaviour* 6, pp. 324–325. DOI: 10.1038/s41562-021-01262-x.

Schmuker, M., Pfeil, T., and Nawrot, M. P. (2014). “A neuromorphic network for generic multivariate data classification”. In: *Proceedings of the National Academy of Sciences* 111.6, pp. 2081–2086. DOI: 10.1073/pnas.1303053111.

Schwab, B. C., König, P., and Engel, A. K. (2021). “Spike-timing-dependent plasticity can account for connectivity aftereffects of dual-site transcranial alternating current stimulation”. In: *NeuroImage* 237, p. 118179. DOI: 10.1016/j.neuroimage.2021.118179.

Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H. B., Kronbichler, M., and Friston, K. J. (2019). “Computational mechanisms of curiosity and goal-directed exploration”. In: *eLife* 8, e41703. DOI: 10.7554/eLife.41703.

Sejnowski, T. J., Koch, C., and Churchland, P. S. (1988). “Computational neuroscience”. In: *Science* 241, pp. 1299–1306. DOI: 10.1126/science.3045969.

Sengupta, A., Ye, Y., Wang, R., Liu, C., and Roy, K. (2019). “Going deeper in spiking neural networks: VGG and residual architectures”. In: *Frontiers in Neuroscience* 13.95, pp. 1–10. DOI: 10.3389/fnins.2019.00095.

Sengupta, D., Mastella, M., Chicca, E., and Kottapalli, A. G. P. (2022). “Skin-inspired flexible and stretchable electrospun carbon nanofiber sensors for neuromorphic sensing”. In: *ACS Applied Electronic Materials* 4, pp. 308–315. DOI: 10.1021/acsaelm.1c01010.

Servan-Schreiber, E. and Anderson, J. R. (1990). “Learning artificial grammars with competitive chunking”. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16.4, pp. 592–608. DOI: 10.1037/0278-7393.16.4.592.

Setareh, H., Deger, M., and Gerstner, W. (2018). “Excitable neuronal assemblies with adaptation as a building block of brain circuits for velocity-controlled signal propagation”. In: *PLoS Computational Biology* 14.7, e1006216. DOI: 10.1371/journal.pcbi.1006216.

Shamma, S. (2001). “On the role of space and time in auditory processing”. In: *Trends in Cognitive Sciences* 5.8, pp. 340–348. DOI: 10.1016/S1364-6613(00)01704-6.

Sherman, B. E., Graves, K. N., and Turk-Browne, N. B. (2020). “The prevalence and importance of statistical learning in human cognition and behavior”. In: *Current Opinion in Behavioral Sciences* 32, pp. 15–20. DOI: 10.1016/j.cobeha.2020.01.015.

Sherwood, C. C., Subiaul, F., and Zawidzki, T. W. (2008). “A natural history of the human mind: tracing evolutionary changes in brain and cognition”. In: *Journal of Anatomy* 212, pp. 426–454. DOI: 10.1111/j.1469-7580.2008.00868.x.

Shi, R., Legrand, C., and Brandenberger, A. (2020). “Toddlers track hierarchical structure dependence”. In: *Language Acquisition* 27.4, pp. 397–409. DOI: 10.1080/10489223.2020.1776010.

Shouval, H. Z., Bear, M. F., and Cooper, L. N. (2002). “A unified model of NMDA receptor-dependent bidirectional synaptic plasticity”. In: *Proceedings of the National Academy of Sciences* 99.16, pp. 10831–10836. DOI: 10.1073/pnas.152343099.

Shouval, H. Z., Wang, S. S.-H., and Wittenberg, G. M. (2010). “Spike timing dependent plasticity: a consequence of more fundamental learning rules”. In: *Frontiers in Computational Neuroscience* 4.19, pp. 1–13. DOI: 10.3389/fncom.2010.00019.

Shu, Y., Hasenstaub, A., and McCormick, D. A. (2003). “Turning on and off recurrent balanced cortical activity”. In: *Nature* 423, pp. 288–293. DOI: 10.1038/nature01616.

Shultz, T. R. (1999). “Rule learning by habituation can be simulated in neural networks”. In: *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society (1st ed.)* Ed. by M. Hahn and S. C. Stoness. Psychology Press, pp. 665–670. DOI: 10.4324/9781410603494.

Shultz, T. R. and Bale, A. C. (2001). “Neural network simulation of infant familiarization to artificial sentences: rule-like behavior without explicit rules and variables”. In: *Infancy* 2.4, pp. 501–536. DOI: 10.1207/S15327078IN0204_07.

Siegel, M., Donner, T. H., and Engel, A. K. (2012). “Spectral fingerprints of large-scale neuronal interactions”. In: *Nature Reviews Neuroscience* 13, pp. 121–134. DOI: 10.1038/nrn3137.

Simen, P. and Polk, T. (2010). “A symbolic/subsymbolic interface protocol for cognitive modeling”. In: *Logic Journal of the IGPL* 18.5, pp. 705–761. DOI: 10.1093/jigpal/jzp046.

Singer, W. and Lazar, A. (2016). “Does the cerebral cortex exploit high-dimensional, non-linear dynamics for information processing?” In: *Frontiers in Computational Neuroscience* 10.99, pp. 1–10. DOI: 10.3389/fncom.2016.00099.

Siu, P. H., Müller, E., Zerbi, V., Aquino, K., and Fulcher, B. D. (2022). “Extracting dynamical understanding from neural-mass models of mouse cortex”. In: *Frontiers in Computational Neuroscience* 16.847336, pp. 1–15. DOI: 10.3389/fncom.2022.847336.

Sjöström, P. J., Turrigiano, G. G., and Nelson, S. B. (2001). “Rate, timing, and cooperativity jointly determine cortical synaptic plasticity”. In: *Neuron* 32, pp. 1149–1164. DOI: 10.1016/S0896-6273(01)00542-6.

Skeide, M. A. and Friederici, A. D. (2016). “The ontogeny of the cortical language network”. In: *Nature Reviews Neuroscience* 17, pp. 323–332. DOI: 10.1038/nrn.2016.23.

Smith, S. (2016). “Linking cognition to brain connectivity”. In: *Nature Neuroscience* 19.1, pp. 7–9. DOI: 10.1038/nn.4206.

Smolensky, P. (1987). “Connectionist AI, symbolic AI, and the brain”. In: *Artificial Intelligence Review* 1, pp. 95–109. DOI: 10.1007/BF00130011.

Smolensky, P. (1988). “On the proper treatment of connectionism”. In: *Behavioral and Brain Sciences* 11.1, pp. 1–23. DOI: 10.1017/S0140525X00052432.

Smolensky, P. (2012). “Symbolic functions from neural computation”. In: *Philosophical Transactions of the Royal Society A* 370, pp. 3543–3569. DOI: 10.1098/rsta.2011.0334.

Snow, P. J. (2016). “The structural and functional organization of cognition”. In: *Frontiers in Human Neuroscience* 10.501, pp. 1–24. DOI: 10.3389/fnhum.2016.00501.

Sommer, F. T. and Wennekers, T. (2003). “Models of distributed associative memory networks in the brain”. In: *Theory in Biosciences* 122, pp. 55–69. DOI: 10.1007/s12064-003-0037-8.

Song, S. and Abbott, L. F. (2001). “Cortical development and remapping through spike timing-dependent plasticity”. In: *Neuron* 32, pp. 339–350. DOI: 10.1016/S0896-6273(01)00451-2.

Song, S., Miller, K. D., and Abbott, L. F. (2000). “Competitive Hebbian learning through spike-timing-dependent synaptic plasticity”. In: *Nature Neuroscience* 3.9, pp. 919–926. DOI: 10.1038/78829.

Sporns, O. (2013). “Structure and function of complex brain networks”. In: *Dialogues in Clinical Neuroscience* 15.3, pp. 247–262. DOI: 10.31887/DCNS.2013.15.3/osporns.

Sporns, O. (2014). “Contributions and challenges for network models in cognitive neuroscience”. In: *Nature Neuroscience* 17.5, pp. 652–660. DOI: 10.1038/nn.3690.

Sprekeler, H., Michaelis, C., and Wiskott, L. (2007). “Slowness: an objective for spike-timing-dependent plasticity?” In: *PLoS Computational Biology* 3.6, e112. DOI: 10.1371/journal.pcbi.0030112.

Spruston, N. (2008). “Pyramidal neurons: dendritic structure and synaptic integration”. In: *Nature Reviews Neuroscience* 9, pp. 206–221. DOI: 10.1038/nrn2286.

Steinschneider, M., Nourski, K. V., and Fishman, Y. I. (2013). “Representation of speech in human auditory cortex: is it special?” In: *Hearing Research* 305, pp. 57–73. DOI: 10.1016/j.heares.2013.05.013.

Stemmler, M. and Koch, C. (1999). “How voltage-dependent conductances can adapt to maximize the information encoded by neuronal firing rate”. In: *Nature Neuroscience* 2.6, pp. 521–527. DOI: 10.1038/9173.

Stevens, C. F. and Zador, A. (1995). “Information through a spiking neuron”. In: *Advances in Neural Information Processing Systems 8 (NIPS 1995)*. Ed. by D. Touretzky, M. C. Mozer, and M. Hasselmo. MIT Press, pp. 75–81.

Stevens, C. F. and Zador, A. M. (1998). “Input synchrony and the irregular firing of cortical neurons”. In: *Nature Neuroscience* 1.3, pp. 210–217. DOI: 10.1038/659.

Stewart, T. C., Choo, F.-X., and Eliasmith, C. (2012). “Spaun: a perception-cognition-action model using spiking neurons”. In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 34, pp. 1018–1023.

Stimberg, M., Brette, R., and Goodman, D. F. M. (2019). “Brian 2, an intuitive and efficient neural simulator”. In: *eLife* 8, e47314. DOI: 10.7554/eLife.47314.

Stimberg, M., Goodman, D. F. M., Benichoux, V., and Brette, R. (2014). “Equation-oriented specification of neural models for simulations”. In: *Frontiers in Neuroinformatics* 8.6, pp. 1–14. DOI: 10.3389/fninf.2014.00006.

Stobbe, N., Westphal-Fitch, G., Aust, U., and Fitch, W. T. (2012). “Visual artificial grammar learning: comparative research on humans, kea (Nestor notabilis) and pigeons (Columba livia)”. In: *Philosophical Transactions of the Royal Society B* 367, pp. 1995–2006. DOI: 10.1098/rstb.2012.0096.

Stöckel, A., Stewart, T. C., and Eliasmith, C. (2021). “Connecting biological detail with neural computation: application to the cerebellar granule–Golgi microcircuit”. In: *Topics in Cognitive Science* 13, pp. 515–533. DOI: 10.1111/tops.12536.

Strong, S. P., Koberle, R., de Ruyter van Steveninck, R. R., and Bialek, W. (1998). “Entropy and information in neural spike trains”. In: *Physical Review Letters* 80.1, pp. 197–200. DOI: 10.1103/PhysRevLett.80.197.

Strotzer, M. (2009). “One century of brain mapping using Brodmann areas*”. In: *Clinical Neuroradiology* 19, pp. 179–186. DOI: 10.1007/s00062-009-9002-3.

Strukov, D., Indiveri, G., Grollier, J., and Fusi, S. (2019). “Building brain-inspired computing”. In: *Nature Communications* 10.4838, pp. 1–6. DOI: 10.1038/s41467-019-12521-x.

Stuart, G. J. and Spruston, N. (2015). “Dendritic integration: 60 years of progress”. In: *Nature Neuroscience* 18.12, pp. 1713–1721. DOI: 10.1038/nn.4157.

Sussillo, D. and Abbott, L. F. (2009). “Generating coherent patterns of activity from chaotic neural networks”. In: *Neuron* 63, pp. 544–557. DOI: 10.1016/j.neuron.2009.07.018.

Sussman, E. S. (2007). “A new view on the MMN and attention debate: the role of context in processing auditory events”. In: *Journal of Psychophysiology* 21.3-4, pp. 164–175. DOI: 10.1027/0269-8803.21.34.164.

Suzuki, T. N. (2021). “Animal linguistics: exploring referentiality and compositionality in bird calls”. In: *Ecological Research* 36, pp. 221–231. DOI: 10.1111/1440-1703.12200.

Svozil, D., Kvasnička, V., and Pospíchal, J. (1997). “Introduction to multi-layer feed-forward neural networks”. In: *Chemometrics and Intelligent Laboratory Systems* 39, pp. 43–62. DOI: 10.1016/S0169-7439(97)00061-0.

Szatmáry, B. and Izhikevich, E. M. (2010). “Spike-timing theory of working memory”. In: *PLoS Computational Biology* 6.8, e1000879. DOI: 10.1371/journal.pcbi.1000879.

Taherkhani, A., Belatreche, A., Li, Y., Cosma, G., Maguire, L. P., and McGinnity, T. M. (2020). “A review of learning in biologically plausible spiking neural networks”. In: *Neural Networks* 122, pp. 253–272. DOI: 10.1016/j.neunet.2019.09.036.

Takahashi, K., Kim, S., Coleman, T. P., Brown, K. A., Suminski, A. J., Best, M. D., and Hatsopoulos, N. G. (2015). “Large-scale spatiotemporal spike patterning consistent with wave propagation in motor cortex”. In: *Nature Communications* 6.7169, pp. 1–11. DOI: 10.1038/ncomms8169.

Tanaka, G., Yamane, T., Héroux, J. B., Nakane, R., Kanazawa, N., Takeda, S., Numata, H., Nakano, D., and Hirose, A. (2019). “Recent advances in physical reservoir computing: a review”. In: *Neural Networks* 115, pp. 100–123. DOI: 10.1016/j.neunet.2019.03.005.

Tang, M., Yang, Y., and Amit, Y. (2022). “Biologically plausible training mechanisms for self-supervised learning in deep networks”. In: *Frontiers in Computational Neuroscience* 16.789253, pp. 1–15. DOI: 10.3389/fncom.2022.789253.

ten Cate, C. and Okanoya, K. (2012). “Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning”. In: *Philosophical Transactions of the Royal Society B* 367, pp. 1984–1994. DOI: 10.1098/rstb.2012.0055.

ten Oever, S. and Martin, A. E. (2021). “An oscillating computational model can track pseudo-rhythmic speech by using linguistic predictions”. In: *eLife* 10, e68066. DOI: 10.7554/eLife.68066.

Ten, A., Kaushik, P., Oudeyer, P.-Y., and Gottlieb, J. (2021). “Humans monitor learning progress in curiosity-driven exploration”. In: *Nature Communications* 12.5972, pp. 1–10. DOI: 10.1038/s41467-021-26196-w.

Teng, X., Tian, X., Rowland, J., and Poeppel, D. (2017). “Concurrent temporal channels for auditory processing: oscillatory neural entrainment reveals segregation of function at different scales”. In: *PLoS Biology* 15.11, e2000812. DOI: 10.1371/journal.pbio.2000812.

ter Haar, S. M., Fernandez, A. A., Gratier, M., Knörnschild, M., Levelt, C., Moore, R. K., Vellema, M., Wang, X., and Oller, D. K. (2021). “Cross-species parallels in babbling: animals and algorithms”. In: *Philosophical Transactions of the Royal Society B* 376, p. 20200239. DOI: 10.1098/rstb.2020.0239.

Testa-Silva, G., Verhoog, M. B., Goriounova, N. A., Loebel, A., Hjorth, J. J. J., Baayen, J. C., de Kock, C. P. J., and Mansvelder, H. D. (2010). “Human synapses show a wide temporal window for spike-timing-dependent plasticity”. In: *Frontiers in Synaptic Neuroscience* 2.12, pp. 1–11. DOI: 10.3389/fnsyn.2010.00012.

Tetzlaff, C., Dasgupta, S., Kulvicius, T., and Wörgötter, F. (2015). “The use of Hebbian cell assemblies for nonlinear computation”. In: *Scientific Reports* 5.12866, pp. 1–14. DOI: 10.1038/srep12866.

Tetzlaff, C., Kolodziejski, C., Markelic, I., and Wörgötter, F. (2012). “Time scales of memory, learning, and plasticity”. In: *Biological Cybernetics* 106, pp. 715–726. DOI: 10.1007/s00422-012-0529-z.

Thakur, C. S., Molin, J. L., Cauwenberghs, G., Indiveri, G., Kumar, K., Qiao, N., Schemmel, J., Wang, R., Chicca, E., Olson Hasler, J., Seo, J.-s., Yu, S., Cao, Y., van Schaik, A., and Etienne-Cummings, R. (2018). “Large-scale neuromorphic spiking array processors: a quest to mimic the brain”. In: *Frontiers in Neuroscience* 12.891, pp. 1–37. DOI: 10.3389/fnins.2018.00891.

Thiessen, E. D. (2017). “What’s statistical about learning? Insights from modelling statistical learning as a set of memory processes”. In: *Philosophical Transactions of the Royal Society B* 372, p. 20160056. DOI: 10.1098/rstb.2016.0056.

Thiessen, E. D. (2020). “How the demands of a variable environment give rise to statistical learning”. In: *Language and Concept Acquisition from Infancy Through Childhood: Learning from Multiple Exemplars*. Ed. by J. B. Childers. Springer, Cham, pp. 59–77. DOI: 10.1007/978-3-030-35594-4_4.

Thiessen, E. D. and Erickson, L. C. (2013a). “Beyond word segmentation: a two-process account of statistical learning”. In: *Current Directions in Psychological Science* 22.3, pp. 239–243. DOI: 10.1177/0963721413476035.

Thiessen, E. D. and Erickson, L. C. (2013b). “Discovering words in fluent speech: the contribution of two kinds of statistical information”. In: *Frontiers in Psychology* 3.590, pp. 1–10. DOI: 10.3389/fpsyg.2012.00590.

Thiessen, E. D., Kronstein, A. T., and Hufnagle, D. G. (2013). “The extraction and integration framework: a two-process account of statistical learning”. In: *Psychological Bulletin* 139.4, pp. 792–814. DOI: 10.1037/a0030801.

Thiessen, E. D. and Pavlik Jr., P. I. (2013). “iMinerva: a mathematical model of distributional statistical learning”. In: *Cognitive Science* 37, pp. 310–343. DOI: 10.1111/cogs.12011.

Thompson, S. P. and Newport, E. L. (2007). “Statistical learning of syntax: the role of transitional probability”. In: *Language Learning and Development* 3.1, pp. 1–42. DOI: 10.1207/s15473341lld0301_1.

Thompson-Schill, S. L., Bedny, M., and Goldberg, R. F. (2005). “The frontal lobes and the regulation of mental activity”. In: *Current Opinion in Neurobiology* 15, pp. 219–224. DOI: 10.1016/j.conb.2005.03.006.

Thompson-Schill, S. L., Ramscar, M., and Chrysikou, E. G. (2009). “Cognition without control: when a little frontal lobe goes a long way”. In: *Current Directions in Psychological Science* 18.5, pp. 259–263. DOI: 10.1111/j.1467-8721.2009.01648.x.

Thomson, A. M. and Bannister, A. P. (2003). “Interlaminar connections in the neocortex”. In: *Cerebral Cortex* 13, pp. 5–14. DOI: 10.1093/cercor/13.1.5.

Thomson, A. M. and Deuchars, J. (1994). “Temporal and spatial properties of local circuits in neocortex”. In: *Trends in Neurosciences* 17.3, pp. 119–126. DOI: 10.1016/0166-2236(94)90121-X.

Tikidji-Hamburyan, R. A., Narayana, V., Bozkus, Z., and El-Ghazawi, T. A. (2017). “Software for brain network simulations: a comparative study”. In: *Frontiers in Neuroinformatics* 11.46, pp. 1–16. DOI: 10.3389/fninf.2017.00046.

Tomasello, R., Garagnani, M., Wennekers, T., and Pulvermüller, F. (2018). “A neurobiologically constrained cortex model of semantic grounding with spiking neurons and brain-like connectivity”. In: *Frontiers in Computational Neuroscience* 12.88, pp. 1–17. DOI: 10.3389/fncom.2018.00088.

Tomasello, R., Wennekers, T., Garagnani, M., and Pulvermüller, F. (2019). “Visual cortex recruitment during language processing in blind individuals is explained by Hebbian learning”. In: *Scientific Reports* 9.3579, pp. 1–16. DOI: 10.1038/s41598-019-39864-1.

Tong, M. H., Bickett, A. D., Christiansen, E. M., and Cottrell, G. W. (2007). “Learning grammatical structure with echo state networks”. In: *Neural Networks* 20, pp. 424–432. DOI: 10.1016/j.neunet.2007.04.013.

Toro, J. M. (2016). “Something old, something new: combining mechanisms during language acquisition”. In: *Current Directions in Psychological Science* 25.2, pp. 130–134. DOI: 10.1177/0963721416629645.

Toro, J. M. and Hoeschele, M. (2017). “Generalizing prosodic patterns by a non-vocal learning mammal”. In: *Animal Cognition* 20, pp. 179–185. DOI: 10.1007/s10071-016-1036-8.

Toro, J. M., Nespor, M., Mehler, J., and Bonatti, L. L. (2008a). “Finding words and rules in a speech stream: functional differences between vowels and consonants”. In: *Psychological Science* 19.2, pp. 137–144. DOI: 10.1111/j.1467-9280.2008.02059.x.

Toro, J. M., Shukla, M., Nespor, M., and Endress, A. D. (2008b). “The quest for generalizations over consonants: asymmetries between consonants and vowels are not the by-product of acoustic differences”. In: *Perception & Psychophysics* 70.8, pp. 1515–1525. DOI: 10.3758/PP.70.8.1515.

Toro, J. M. and Trobalón, J. B. (2005). “Statistical computations over a speech stream in a rodent”. In: *Perception & Psychophysics* 67.5, pp. 867–875. DOI: 10.3758/BF03193539.

Toro, J. M., Trobalon, J. B., and Sebastián-Gallés, N. (2003). “The use of prosodic cues in language discrimination tasks by rats”. In: *Animal Cognition* 6, pp. 131–136. DOI: 10.1007/s10071-003-0172-0.

Torres, J. J. and Marro, J. (2022). “Physics clues on the mind substrate and attributes”. In: *Frontiers in Computational Neuroscience* 16.836532, pp. 1–10. DOI: 10.3389/fncom.2022.836532.

Toutounji, H. and Pipa, G. (2014). “Spatiotemporal computations of an excitable and plastic brain: neuronal plasticity leads to noise-robust and noise-constructive computations”. In: *PLoS Computational Biology* 10.3, e1003512. DOI: 10.1371/journal.pcbi.1003512.

Trainor, L., McFadden, M., Hodgson, L., Darragh, L., Barlow, J., Matsos, L., and Sonnadara, R. (2003). “Changes in auditory cortex and the development of mismatch negativity between 2 and 6 months of age”. In: *International Journal of Psychophysiology* 51, pp. 5–15. DOI: 10.1016/S0167-8760(03)00148-X.

Triesch, J. (2004). “Synergies between intrinsic and synaptic plasticity in individual model neurons”. In: *Advances in Neural Information Processing Systems 17 (NIPS 2004).* Ed. by L. Saul, Y. Weiss, and L. Bottou. MIT Press.

Triesch, J. (2007). “Synergies between intrinsic and synaptic plasticity mechanisms”. In: *Neural Computation* 19, pp. 885–909. DOI: 10.1162/neco.2007.19.4.885.

Tully, P. J., Lindén, H., Hennig, M. H., and Lansner, A. (2016). “Spike-based Bayesian-Hebbian learning of temporal sequences”. In: *PLoS Computational Biology* 12.5, e1004954. DOI: 10.1371/journal.pcbi.1004954.

Turrigiano, G., Abbott, L. F., and Marder, E. (1994). “Activity-dependent changes in the intrinsic properties of cultured neurons”. In: *Science* 264, pp. 974–977. DOI: 10.1126/science.8178157.

Uddén, J., de Jesus Dias Martins, M., Zuidema, W., and Fitch, W. T. (2020). “Hierarchical structure in sequence processing: how to measure it and determine its neural implementation”. In: *Topics in Cognitive Science* 12, pp. 910–924. DOI: 10.1111/tops.12442.

Uddén, J., Ingvar, M., Hagoort, P., and Petersson, K. M. (2012). “Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: investigating the push-down stack model”. In: *Cognitive Science* 36, pp. 1078–1101. DOI: 10.1111/j.1551-6709.2012.01235.x.

Uddén, J. and Männel, C. (2018). “Artificial grammar learning and its neurobiology in relation to language processing and development”. In: *The Oxford Handbook of Psycholinguistics (2 ed.)* Ed. by S.-A. Rueschemeyer and M. G. Gaskell. Oxford University Press. DOI: 10.1093/oxfordhb/9780198786825.013.33.

Ueno, T., Saito, S., Rogers, T. T., and Lambon Ralph, M. A. (2011). “Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways”. In: *Neuron* 72, pp. 385–396. DOI: 10.1016/j.neuron.2011.09.013.

Uhlhaas, P. J., Pipa, G., Lima, B., Melloni, L., Neuenschwander, S., Nikolić, D., and Singer, W. (2009). “Neural synchrony in cortical networks: history, concept and current status”. In: *Frontiers in Integrative Neuroscience* 3.17, pp. 1–19. DOI: 10.3389/neuro.07.017.2009.

Urai, A. E., Doiron, B., Leifer, A. M., and Churchland, A. K. (2022). “Large-scale neural recordings call for new insights to link brain and behavior”. In: *Nature Neuroscience* 25, pp. 11–19. DOI: 10.1038/s41593-021-00980-9.

van Bergen, R. S. and Kriegeskorte, N. (2020). “Going in circles is the way forward: the role of recurrence in visual inference”. In: *Current Opinion in Neurobiology* 65, pp. 176–193. DOI: 10.1016/j.conb.2020.11.009.

van den Bos, E., Christiansen, M. H., and Misyak, J. B. (2012). “Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration”. In: *Journal of Memory and Language* 67, pp. 507–520. DOI: 10.1016/j.jml.2012.07.008.

van den Bos, E. and Poletiek, F. H. (2008). “Effects of grammar complexity on artificial grammar learning”. In: *Memory & Cognition* 36.6, pp. 1122–1131. DOI: 10.3758/MC.36.6.1122.

van der Kant, A., Männel, C., Paul, M., Friederici, A. D., Höhle, B., and Wartenburger, I. (2020). “Linguistic and non-linguistic non-adjacent dependency learning in early development”. In: *Developmental Cognitive Neuroscience* 45, p. 100819. DOI: 10.1016/j.dcn.2020.100819.

van Horik, J. and Emery, N. J. (2011). “Evolution of cognition”. In: *WIREs Cognitive Science* 2, pp. 621–633. DOI: 10.1002/wcs.144.

van Wassenhove, V., Grant, K. W., and Poeppel, D. (2007). “Temporal window of integration in auditory-visual speech perception”. In: *Neuropsychologia* 45, pp. 598–607. DOI: 10.1016/j.neuropsychologia.2006.01.001.

VanRullen, R., Guyonneau, R., and Thorpe, S. J. (2005). “Spike times make sense”. In: *Trends in Neurosciences* 28.1, pp. 1–4. DOI: 10.1016/j.tins.2004.10.010.

Venkatesan, T. and Williams, S. (2022). “Brain inspired electronics”. In: *Applied Physics Reviews* 9, p. 010401. DOI: 10.1063/5.0078798.

Vertes, P. E. and Duke, T. (2010). “Effect of network topology on neuronal encoding based on spatiotemporal patterns of spikes”. In: *HFSP Journal* 4.3–4, pp. 153–163. DOI: 10.2976/1.3386761.

Vincent-Lamarre, P., Calderini, M., and Thivierge, J.-P. (2020). “Learning long temporal sequences in spiking networks by multiplexing neural oscillations”. In: *Frontiers in Computational Neuroscience* 14.78, pp. 1–16. DOI: 10.3389/fncom.2020.00078.

Vinyals, O., Toshev, A., Bengio, S., and Erhan, D. (2015). “Show and tell: a neural image caption generator”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3156–3164.

Vogels, T. P. and Abbott, L. F. (2005). “Signal propagation and logic gating in networks of integrate-and-fire neurons”. In: *The Journal of Neuroscience* 25.46, pp. 10786–10795. DOI: 10.1523/JNEUROSCI.3508-05.2005.

Wacongne, C., Changeux, J.-P., and Dehaene, S. (2012). “A neuronal model of predictive coding accounting for the mismatch negativity”. In: *The Journal of Neuroscience* 32.11, pp. 3665–3678. DOI: 10.1523/JNEUROSCI.5003-11.2012.

Wang, F. H., Zevin, J., and Mintz, T. H. (2019). “Successfully learning non-adjacent dependencies in a continuous artificial language stream”. In: *Cognitive Psychology* 113, p. 101223. DOI: 10.1016/j.cogpsych.2019.101223.

Watson, S. K., Burkart, J. M., Schapiro, S. J., Lambeth, S. P., Mueller, J. L., and Townsend, S. W. (2020). “Nonadjacent dependency processing in monkeys, apes, and humans”. In: *Science Advances* 6, eabb0725. DOI: 10.1126/sciadv.abb0725.

Watson, S. K., Burkart, J. M., Schapiro, S. J., Lambeth, S. P., Mueller, J. L., and Townsend, S. W. (2021). “Reply to comment on “Nonadjacent dependency processing in monkeys, apes, and humans””. In: *Science Advances* 7, eabj1517. DOI: 10.1126/sciadv.abj1517.

Weber, L., Jumelet, J., Bruni, E., and Hupkes, D. (2021). “Language modelling as a multi-task problem”. In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 2049–2060. DOI: 10.18653/v1/2021.eacl-main.176.

Wehr, M. and Laurent, G. (1996). “Odour encoding by temporal sequences of firing in oscillating neural assemblies”. In: *Nature* 384, pp. 162–166. DOI: 10.1038/384162a0.

Wennekers, T., Garagnani, M., and Pulvermüller, F. (2006). “Language models based on Hebbian cell assemblies”. In: *Journal of Physiology-Paris* 100, pp. 16–30. DOI: 10.1016/j.jphysparis.2006.09.007.

Wennekers, T. and Palm, G. (2009). “Syntactic sequencing in Hebbian cell assemblies”. In: *Cognitive Neurodynamics* 3, pp. 429–441. DOI: 10.1007/s11571-009-9095-z.

Werning, M., Tacca, M. C., and Mroczko-Wąsowicz, A. (2013). “High- vs low-level cognition and the neuro-emulative theory of mental representation”. In: *Models, Simulations, and the Reduction of Complexity*. Ed. by U. Gähde, S. Hartmann, and J. H. Wolf. Berlin, Boston: De Gruyter, pp. 141–152. DOI: 10.1515/9783110313680.141.

Weyers, I. and Mueller, J. L. (2022). “A special role of syllables, but not vowels or consonants, for nonadjacent dependency learning”. In: *Journal of Cognitive Neuroscience* 34.8, pp. 1467–1487. DOI: 10.1162/jocn_a_01874.

Wilson, A. D. and Golonka, S. (2013). “Embodied cognition is not what you think it is”. In: *Frontiers in Psychology* 4.58, pp. 1–13. DOI: 10.3389/fpsyg.2013.00058.

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., van der Kant, A., Smith, K., and Rey, A. (2018). “Non-adjacent dependency learning in humans and other animals”. In: *Topics in Cognitive Science*, pp. 1–16. DOI: 10.1111/tops.12381.

Winkler, M., Mueller, J. L., Friederici, A. D., and Männel, C. (2018a). “Infant cognition includes the potentially human-unique ability to encode embedding”. In: *Science Advances* 4, eaar8334. DOI: 10.1126/sciadv.aar8334.

Winkler, M., Mueller, J. L., Friederici, A. D., and Männel, C. (2018b). “Supplementary materials for “Infant cognition includes the potentially human-unique ability to encode embedding””. In: *Science Advances* 4, eaar8334. DOI: 10.1126/sciadv.aar8334.

de-Wit, L, Machilsen, B., and Putzeys, T. (2010). “Predictive coding and the neural response to predictable stimuli”. In: *The Journal of Neuroscience* 30.26, pp. 8702–8703. DOI: 10.1523/JNEUROSCI.2248-10.2010.

Wolff, M., Morceau, S., Folkard, R., Martin-Cortecero, J., and Groh, A. (2021). “A thalamic bridge from sensory perception to cognition”. In: *Neuroscience and Biobehavioral Reviews* 120, pp. 222–235. DOI: 10.1016/j.neubiorev.2020.11.013.

Wörgötter, F. and Porr, B. (2005). “Temporal sequence learning, prediction, and control: a review of different models and their relation to biological mechanisms”. In: *Neural Computation* 17, pp. 245–319. DOI: 10.1162/0899766053011555.

Wu, J., Chua, Y., Zhang, M., Yang, Q., Li, G., and Li, H. (2019). “Deep spiking neural network with spike count based learning rule”. In: *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6. DOI: 10.1109/IJCNN.2019.8852380.

Wu, J., Yılmaz, E., Zhang, M., Li, H., and Tan, K. C. (2020). “Deep spiking neural networks for large vocabulary automatic speech recognition”. In: *Frontiers in Neuroscience* 14.199, pp. 1–14. DOI: 10.3389/fnins.2020.00199.

Xu, K., Ba, J. L., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., Zemel, R. S., and Bengio, Y. (2015). “Show, attend and tell: neural image caption generation with visual attention”. In: *Proceedings of the 32nd International Conference on Machine Learning, PMLR* 37, pp. 2048–2057.

Xu, S., Jiang, W., Poo, M.-m., and Dan, Y. (2012). “Activity recall in a visual cortical ensemble”. In: *Nature Neuroscience* 15.3, pp. 449–455. DOI: 10.1038/nn.3036.

Yalnizyan-Carson, A. and Richards, B. A. (2022). “Forgetting enhances episodic control with structured memories”. In: *Frontiers in Computational Neuroscience* 16.757244, pp. 1–16. DOI: 10.3389/fncom.2022.757244.

Yamins, D. L. K. and DiCarlo, J. J. (2016). “Using goal-driven deep learning models to understand sensory cortex”. In: *Nature Neuroscience* 19.3, pp. 356 –365. DOI: 10.1038/nn.4244.

Yang, C. D. (2004). “Universal grammar, statistics or both?” In: *Trends in Cognitive Sciences* 8.10, pp. 451–456. DOI: 10.1016/j.tics.2004.08.006.

Yang, C., Crain, S., Berwick, R. C., Chomsky, N., and Bolhuis, J. J. (2017). “The growth of language: universal grammar, experience, and principles of computation”. In: *Neuroscience and Biobehavioral Reviews* 81, pp. 103–119. DOI: 10.1016/j.neubiorev.2016.12.023.

Yang, Y. and Piantadosi, S. T. (2022). “One model for the learning of language”. In: *Proceedings of the National Academy of Sciences* 119.5, e2021865119. DOI: 10.1073/pnas.2021865119.

Yildiz, I. B., von Kriegstein, K., and Kiebel, S. J. (2013). “From birdsong to human speech recognition: Bayesian inference on a hierarchy of nonlinear dynamical systems”. In: *PLoS Computational Biology* 9.9, e1003219. DOI: 10.1371/journal.pcbi.1003219.

Yu, Q., Yao, Y., Wang, L., Tang, H., and Dang, J. (2019). “A multi-spike approach for robust sound recognition”. In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 890–894. DOI: 10.1109/ICASSP.2019.8682963.

Zador, A. M. (2019). “A critique of pure learning and what artificial neural networks can learn from animal brains”. In: *Nature Communications* 10.3770, pp. 1–7. DOI: 10.1038/s41467-019-11786-6.

Zai, A. T., Bhargava, S., Mesgarani, N., and Liu, S.-C. (2015). “Reconstruction of audio waveforms from spike trains of artificial cochlea models”. In: *Frontiers in Neuroscience* 9.347, pp. 1–13. DOI: 10.3389/fnins.2015.00347.

Zajzon, B., Mahmoudian, S., Morrison, A., and Duarte, R. (2019). “Passing the message: representation transfer in modular balanced networks”. In: *Frontiers in Computational Neuroscience* 13.79, pp. 1–18. DOI: 10.3389/fncom.2019.00079.

Zaleznik, E. and Park, J. (2021). “The neural basis of counting sequences”. In: *NeuroImage* 237, p. 118146. DOI: 10.1016/j.neuroimage.2021.118146.

Zenke, F., Agnes, E. J., and Gerstner, W. (2015). “Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks”. In: *Nature Communications* 6.6922, pp. 1–13. DOI: 10.1038/ncomms7922.

Zenke, F., Bohté, S. M., Clopath, C., Comşa, I. M., Göltz, J., Maass, W., Masquelier, T., Naud, R., Neftci, E. O., Petrovici, M. A., Scherr, F., and Goodman, D. F. M. (2021). “Visualizing a joint future of neuroscience and neuromorphic engineering”. In: *Neuron* 109, pp. 571–575. DOI: 10.1016/j.neuron.2021.01.009.

Zenke, F. and Gerstner, W. (2017). “Hebbian plasticity requires compensatory processes on multiple timescales”. In: *Philosophical Transactions of the Royal Society B* 372, p. 20160259. DOI: 10.1098/rstb.2016.0259.

Zenke, F., Hennequin, G., and Gerstner, W. (2013). “Synaptic plasticity in neural networks needs homeostasis with a fast rate detector”. In: *PLoS Computational Biology* 9.11, e1003330. DOI: 10.1371/journal.pcbi.1003330.

Zhang, M., Wang, J., Zhang, Z., Belatreche, A., Wu, J., Chua, Y., Qu, H., and Li, H. (2020a). “Spike-timing-dependent back propagation in deep spiking neural networks”. In: *arXiv* 2003.11837v1, pp. 1–10.

Zhang, W. and Li, P. (2020). “Temporal spike sequence learning via backpropagation for deep spiking neural networks”. In: *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*. Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin. Curran Associates, Inc.

Zhang, W. and Linden, D. J. (2003). “The other side of the engram: experience-driven changes in neuronal intrinsic excitability”. In: *Nature Reviews Neuroscience* 4, pp. 885–900. DOI: 10.1038/nrn1248.

Zhang, Y., Qu, P., Ji, Y., Zhang, W., Gao, G., Wang, G., Song, S., Li, G., Chen, W., Zheng, W., Chen, F., Pei, J., Zhao, R., Zhao, M., and Shi, L. (2020b). “A system hierarchy for brain-inspired computing”. In: *Nature* 586, pp. 378–384. DOI: 10.1038/s41586-020-2782-y.

Zheng, P., Dimitrakakis, C., and Triesch, J. (2013). “Network self-organization explains the statistics and dynamics of synaptic connection strengths in cortex”. In: *PLoS Computational Biology* 9.1, e1002848. DOI: 10.1371/journal.pcbi.1002848.

Zilles, K. and Amunts, K. (2010). “Centenary of Brodmann’s map — conception and fate”. In: *Nature Reviews Neuroscience* 11, pp. 139–145. DOI: 10.1038/nrn2776.

Zucker, R. S. and Regehr, W. G. (2002). “Short-term synaptic plasticity”. In: *Annual Review of Physiology* 64, pp. 355–405. DOI: 10.1146/annurev.physiol.64.092501.114547.

Zuidema, W. and Fitz, H. (2019). “Key issues and future directions: models of human language and speech processing”. In: *Human Language: From Genes and Brains to Behavior*. Ed. by P. Hagoort. Cambridge, MA: MIT Press, pp. 353–358.

Zuidema, W., French, R. M., Alhama, R. G., Ellis, K., O’Donnell, T. J., Sainburg, T., and Gentner, T. Q. (2020). “Five ways in which computational modeling can help advance cognitive science: lessons from artificial grammar learning”. In: *Topics in Cognitive Science* 12, pp. 925–941. DOI: 10.1111/tops.12474.

Zuo, F., Panda, P., Kotiuga, M., Li, J., Kang, M., Mazzoli, C., Zhou, H., Barbour, A., Wilkins, S., Narayanan, B., Cherukara, M., Zhang, Z., Sankaranarayanan, S. K. R. S., Comin, R., Rabe, K. M., Roy, K., and Ramanathan, S. (2017a). “Habituation based synaptic plasticity and organismic learning in a quantum perovskite”. In: *Nature Communications* 8.240, pp. 1–7. DOI: 10.1038/s41467-017-00248-6.

Zuo, F., Panda, P., Kotiuga, M., Li, J., Kang, M., Mazzoli, C., Zhou, H., Barbour, A., Wilkins, S., Narayanan, B., Cherukara, M., Zhang, Z., Sankaranarayanan, S. K. R. S., Comin, R., Rabe, K. M., Roy, K., and Ramanathan, S. (2017b). "Supplementary information for "Habituation based synaptic plasticity and organismic learning in a quantum perovskite"". In: *Nature Communications* 8.240. DOI: 10.1038/s41467-017-00248-6.