The Principles of Self-Organization of Memories in Neural Networks for Generating and Performing Cognitive Strategies

---

Dissertation

in order to obtain the doctoral degree
"Doctor rerum naturalium"
of the Georg-August-Universität Göttingen

within the doctoral program

International Max Planck Research School on Physics of Biological and Complex Systems

of the

Göttingen Graduate School for Neurosciences, Biophysics, and Molecular Biosciences

submitted by

Juliane Herpich
born in Grevesmühlen

Göttingen, 2018

Thesis committee

Dr. Christian Tetzlaff,
III. Physikalisches Institut - Biophysik, Georg-August-Universität Göttingen

Prof. Dr. Stefan Klumpp,
Institut für Nichtlineare Dynamik, Georg-August-Universität Göttingen

Prof. Dr. Robert Gütig,
Berliner Institut für Gesundheitsforschung und Charité - Universitätsmedizin, Berlin

Members of the examination board

First Reviewer: Dr. Christian Tetzlaff,
III. Physikalisches Institut - Biophysik, Georg-August-Universität Göttingen

Second Reviewer: Prof. Dr. Stefan Klumpp,
Institut für Nichtlineare Dynamik, Georg-August-Universität Göttingen

Further members of the examination board

Prof. Dr. Robert Gütig,
Berliner Institut für Gesundheitsforschung und Charité - Universitätsmedizin, Berlin

Prof. Dr. Jörg Enderlein,
III. Physikalisches Institut - Biophysik, Georg-August-Universität Göttingen

Dr. Dieter Klopfenstein,
III. Physikalisches Institut - Biophysik, Georg-August-Universität Göttingen

Prof. Dr. Alexander Gail,
Deutsches Primatenzentrum, Göttingen

Date of the oral examination: 7$^{\text{th}}$ of December, 2018

ABSTRACT

Higher-order animals exhibit the remarkable ability to dynamically adapt to a changing environment. On the neuronal level, they have to form mental representations of specific stimuli, so-called memories. Furthermore, they abstract and arrange multiple context-related memories into a corresponding network that can also be adapted by changes in the environment. Such adaptive networks of interconnected memories are termed schemata and construct the mental representation guiding behavior. Considering two interconnected memories within a schema, we can define three different forms of functional organizations of memories dependent on the ability of the memories to either excite or inhibit each other: Two memories can either mutually excite each other, i.e. form an association, mutually inhibit each other i.e. form a discrimination or build up an asymmetric organization, where one memory excites and the other inhibits its interconnected memory, i.e. form a sequence. In order to adapt schema to external stimuli, all of these functional organizations must emerge from the same underlying neuronal mechanism. Experimental, computational and theoretical studies have shown that the underlying neuronal mechanism forming memory representations is activity-dependent synaptic plasticity. This mechanism leads to the formation of strongly interconnected groups of neurons, so-called cell assemblies, which decode memories. However, whether the same synaptic plasticity mechanism can account for the formation of large networks of memories is still unknown. In this thesis, we derive a theoretical model of interacting neuronal populations that enables to analytically study different synaptic plasticity mechanisms with respect to their ability to form all three functional organizations of memories. Two specific excitatory synaptic plasticity mechanisms, correlation-based and homeostatic plasticity, have already successfully been used to form individual cell assemblies in neuronal networks. Nevertheless, our analysis reveals that these two plasticity mechanisms are not sufficient to implement all different forms of functional memory organizations such that further mechanisms are necessary. In this thesis, three distinct strategies are proposed that enable the formation of diverse networks of memories. The first approach is to add a further excitatory synaptic plasticity mechanism based on the causality of neuronal firing, in particular, calculating the difference of pre- and postsynaptic neuronal activities. The second strategy is to allow for inhomogeneities in the time scale of the homeostatic synaptic plasticity mechanism, serving the consolidation of individual memories. The third solution is accomplished by inhibitory synaptic plasticity in addition to correlation-based and homeostatic excitatory synaptic plasticity. However, these three distinct implementations of synaptic plasticity mechanisms are capable to enable the input-dependent formation of all three functional organizations of memories. Therefore, implementing these strategies yield complex adaptive networks of memories, hence, enabling behavior. Finally, we strongly advocate that these synaptic plasticity mechanism can be used in an dynamically input-dependent manner to compute any algorithm that is complete with respect to structured program theory. Thus, the synaptic plasticity mechanisms proposed in this thesis could be extremely useful for technical and computational applications.

Contents

Contents

List of Acronyms

χ**APS** characteristic activity phase space

AMPA α-amino-3-hydroxy-5-mehtyl-4-isoxazolepropionic acid

AP *action potential*

asc *association*

aSPaSS *augmented learning rule on Synaptic Plasticity and Synaptic Scaling by Causality*

bAP *back-propagating action potential*

CA *cell assembly*

Ca^{2+} calcium

disc *discrimination*

EPSP *excitatory postsynaptic potential*

FO *functional organization*

IPSP *inhibitory postsynaptic potential*

LTD *long-term depression*

LTM *long-term memory*

LTP *long-term potentiation*

LTR *long-term representation*

Mg^{2+} magnesium

MR *memory representation*

Na$^+$ sodium

nm *no-memory state*

Terms and Abbreviations

NMDA N-methyl D-aspartate

ODE *ordinary differential equation*

PSD *postsynaptic density*

PSP *postsynaptic potential*

seq *sequence*

seq:rs *sequence from memory s to memory r*

SPaSS *learning rule on Synaptic Plasticity and Synaptic Scaling*

SPM *synaptic-plasticity-and-memory*

STDP *spike-timing dependent plasticity*

VSCC *voltage-sensitive calcium channels*

List of Figures

5 Discussion, Conclusions and Prospects

List of Tables

1

Introduction

In cognitive science, *memory* is understood as a record of events leading to experience. This experience, in turn, can lead to a change of behavior, which is defined as *learning* (Lieberman 2004; Hasselhorn and Gold 2013; Richter and Yang 2009) and provides an essential ability for the survival of living beings. In this framework, the changes in behavior are due to learning and adapting the corresponding mental representations and relations of memories in the human brain. Many biological experiments provide evidence that the underlying physical mechanism of learning is activity-dependent synaptic plasticity. This mechanism induces changes of the *synaptic weights* (strength of the connections) between neurons (Hebb 1949; Martin et al. 2000). Consequently, synaptic plasticity leads to the formation of strongly interconnected groups of neurons called *cell assemblies* (CAs). CAs are hypothesized to represent memory items in neuronal systems (Kandel et al. 2013; Palm et al. 2014; Pulvermüller et al. 2014; Buzsaki 2010). These relations between synaptic plasticity, cell assemblies, learning, and memory are summarized in the *synaptic-plasticity-and-memory* (SPM) hypothesis (Martin et al. 2000; Hebb 1949; Dayan and Abbott 2001; Eichenbaum 2012). In particular, the storage of a memory is associated with the creation of a *cell assembly* (CA) whilst the recall of a memory is linked to the activation of a CA by an external cue (Gagné 1965; Braitenberg 1978; Palm 1981; Wickelgren 1999; Kandel et al. 2013). As known from everyday life, the brain is able to store and recall single isolated memories. Furthermore, the brain can dynamically connect such memories with each other to perform complex cognitive strategies (e.g., problem solving), which are dependent on the specific learning situation. A cortical network of such functionally connected memories, that allow for context specific behavior, is described by a *schema* (Head and Holmes 1911; Piaget 1926; Bartlett 1932; Ghosh and Gilboa 2014). Among the multiple different cognitive structures, *schema* accounts for many of its properties, hence represents a generic cognitive structure (Head and Holmes 1911; Bartlett 1932; Rumelhart 1980; Cooper et al. 1995; Shea et al. 2008; Kumaran et al. 2009).

Although, it is well known that synaptic plasticity enables the formation of single *memory representation*s (MRs) in neuronal networks (Garagnani et al. 2009; Tetzlaff et al. 2013; Litwin-Kumar and Doiron 2014; Zenke et al. 2015) the underlying principles of the self-

organization of memories into a network of memories resulting in cognitive strategies are still unknown. Since synaptic plasticity is the underlying mechanism of learning and memory in neuronal systems, this work is based on the hypothesis that *complex cognitive strategies, dependent on the interaction and connection of numerous memories, are generated and performed based on the dynamics of synaptic plasticity.*

In order to prove this hypothesis, in Chapter 2, we provide an overview of the relevant experimental results and theoretical concepts that describe synaptic plasticity and the formation of single MRs in brain tissue. Although complex strategies require a multitude of connected memories, the smallest functional entity is that of two connected memories (Preston and Eichenbaum 2013; Ghosh and Gilboa 2014). Hence, in this thesis, we start by analyzing the brain's neuronal principles for the self-organization of two interconnected memories. For this, we study adaptive recurrent neuronal network models sensitive to environmental learning stimuli. Independent of the underlying neuronal principles guiding learning (i.e. synaptic plasticity), we derive in Chapter 3 a method that enables an analytic evaluation of the FO of two interconnected memories in a population model formalism. For this method, the dimensionality of the neuronal network is reduced by considering the mean equilibrium states of the neuronal populations. Here, each population represents the neuronal correlate of the specific environmental input, i.e. *cell assembly* (CA) or *memory representation* (MR). Considering two interconnected memories within a schema, we define three different forms of FOs of memories dependent on the ability of the memories to either excite or inhibit each other: Two memories can either mutually excite each other, i.e. form an *association* (asc), mutually inhibit each other, i.e. form a *discrimination* (disc), or form an asymmetric organization, where one memory excites and the other inhibits its interconnected memory, i.e. form a *sequence* (seq). These distinctions for the different forms of two functionally organized memories is based on the relation of excitatory and inhibitory synaptic weight strengths. Importantly, our derived mathematical framework is the first theoretical method that enables a comprehensive study on diverse neuronal and synaptic mechanisms to form diverse FOs of two interconnected memories.

Using the analytic method derived in Chapter 3 to describe the FOs of two interconnected memories, namely *association* (asc), *discrimination* (disc), and *sequence* (seq), in Chapter 4, we analyze the essential underlying synaptic plasticity mechanisms enabling the formation of all different types of FOs. Our results indicate that plain correlation-based synaptic plasticity is not sufficient to organize two interconnected memories in all different forms of FOs. Moreover, based on theoretical studies on the formation of individual MRs in neuronal networks (Tetzlaff et al. 2013; Litwin-Kumar and Doiron 2014; Zenke et al. 2015; Tetzlaff et al. 2015; Chenkov et al. 2017) which showed that correlation-based synaptic plasticity leads to *long-term potentiation* (LTP) is the main underlying mechanism and is

likely to be balanced by homeostatic mechanisms (Turrigiano et al. 1998; Dayan and Abbott 2001; Tetzlaff et al. 2011; Zenke et al. 2013), we add homeostatic plasticity. This also fails to organize two interconnected memories is all different forms of FOs. In particular, the formation of discriminated memories mainly challenges synaptic plasticity. Therefore, we further augment the adaptive model by additional synaptic plasticity mechanisms. Hereby, our generically valid theoretical investigations reveal that a further mechanism in synaptic plasticity has to either describe *long-term depression* (LTD), or, alternatively, to consider local differences in the synaptic plasticity mechanisms to enable all different forms of functionally organized memories. We underpin these theoretically proposed strategies by specific formulations for synaptic learning rules determining the dynamics of the adaptive neuronal population model. Specifically, we derive a learning rule for rate-coded neurons. This rule describes activity-dependent LTP, dependent on the correlation of pre- and postsynaptic neurons, and LTD, dependent on the difference in firing rates of pre- and postsynaptic neurons. This learning rule leads to stable synaptic weight dynamics. The stability of synaptic weight dynamics enables us to map the synaptic weight-dependent constraints for the different FOs of two interconnected memories onto the two-dimensional average activity phase space of the respective memories. Hence, we derived a method to analyze the system of two interconnected memories on its so-called synaptic learning rule-specific characteristic activity phase space (χAPS) (first introduced in Fig. 4.1 Aii, p. 63). This method reduces the analysis of two interconnected memories towards the two-dimensional χAPS of the neuronal populations. By means of the synaptic learning rule specific χAPS we validate that the synaptic learning rules considering LTP and LTD are able to arrange the system of two interconnected memories in all different FOs, dependent on the average activity levels of the memories involved. Alternatively, we provide another approach that account for all different types of FOs of two interconnected memories. For this, we consider local differences in the neuronal network dynamics by neuronal processes that describe memory consolidation or enable local inhibitory synaptic plasticity to occur.

Finally, in Chapter 5, we discuss the different approaches to describe activity-dependent synaptic plasticity in the context of two interacting memories. By means of the synaptic learning rule specific χAPSs, we interpolate the extracted principles on the self-organization of two functionally connected memories to the self-organization of multiple connected memories enabling cognitive strategies. Hence, we can conclude that a synaptic learning rule describing LTP and LTD will allow for more diverse schemata of functionally connected memories than the approach enabling local differences in the synaptic weight dynamics. Thus, we validate our hypothesis that activity-dependent synaptic plasticity reliably generates and performs complex cognitive strategies.

Our work strongly suggest that networks of self-organized memories are capable to form the three basic computational structures, such as sequence, selection and iteration. Hence, we conclude with the claim that such self-organized networks of memories can compute any algorithm that is complete with respect to *structured program theory* (Böhm and Jacopini 1966; Byrne and Huyck 2010).

To be able to study the relationship between a mental process and specific brain regions, we must first identify the components of the mental process that we are attempting to explain.

Eric R. Kandel

2

Experimental and Theoretical Background

Our research question states that complex cognitive functions underlie synaptic plasticity on the neuronal level. Thus, in this Chapter, we provide an overview of the relevant experimental results and theoretical concepts describing learning and memory based on synaptic plasticity.

2.1 The Brain and Behavior

The brain is an information processing machine. It is sensitive to sensory information from the environment, as well as the brain's internal state. By comparing the sensory stimuli with the previously stored mental representations it produces behavior.

The first attempts to bring together biological and psychological concepts of behavior are found in the theory of *phrenology*, proposed by Gall (1818). Without adequate evidence, Gall came up with the simplisitc view of specific brain regions of the cerebral cortex that are correlated towards specific characteristics of individual's personality. Therefore, Gall divided the cerebral cortex into 35 distinct *cerebral organs*, each corresponding to a specific mental function. With this approach, phrenologists viewed the cortex as a mosaic of functionally specific areas (Kandel et al. 2013). About a century later, studies by Brodmann (1909), based upon the severely limited approach of phrenology, distinguished 52 anatomically and functionally distinct areas in the cerebral cortex. By this, Brodmann established the *school of cortical localization* (Kandel et al. 2013).

A

B
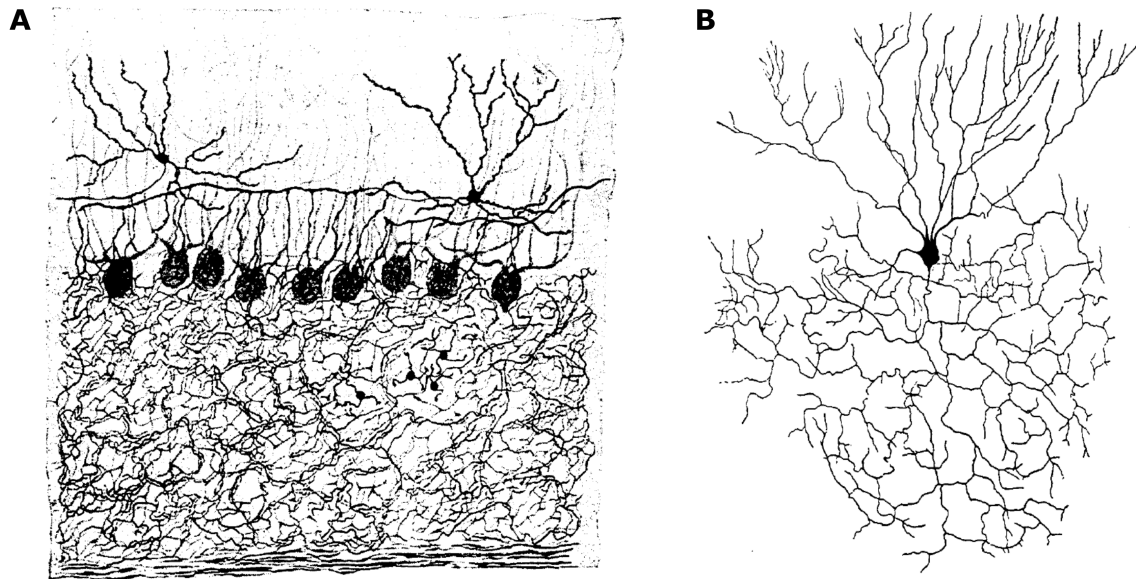
Figure 2.1: Visualization of neurons in brain tissue. (A) Nerve cells that interconnect form a neuronal network. **(B)** A single cell of the granular layer of the cerebellum (Figures taken from Golgi 1906).

An opposing view on the relation of the functionality of the brain and human behavior was given by the so-called, *aggregate-field view* (Flourens 1842). This view considers every mental function is homogeneously embedded within the entire cerebral cortex. By this, any part of the cerebral cortex is crucial for each behavioral function and, along this line, no single brain region is solely responsible for any specific behavior (Kandel et al. 2013).

Even though the principle of locally distinct functional regions, tenously proposed by phreologists, was redeemed by the aggregate-field view, it gained support by major scientific achievements. Therefore, the evolving *concept of localization of brain functions* become a cornerstone of modern brain science (Kandel et al. 2013).

2.1.1 The Brain's Structure and Function

First steps in unraveling the brain's structure were made by Waldeyer (1891). He stated that the nervous system is composed by assemblies of anatomically distinct cellular nerve units and called them *neurons*. His work provides the basis for an emerging central neuroscientific concept, so-called *neuron doctrine*, where neurons are thought to be the basic structural and functional signaling units of the nervous system (Shepherd 1991; Dierig 1994; Yuste 2015). This assumption gained support by the work of Golgi. In 1906, Golgi published a technique to stain neurons in brain tissue. By this, he revealed the basic

anatomical structure of neurons: the cell body, branching dendrites and a long cable-like axon as shown in Fig. 2.1 B (Golgi 1906; see Fig. 2.2 A for a schematic of a neuron). Based on this technique, Ramón y Cajal (1906) was able to stain individual neurons and observed physical discontinuities between single neurons. Hence, he confirmed that the nervous system is not a continuous tissue – appropriate to the aggregate-field view – but rather a network of interconnected, individual cells (Kandel et al. 2013; Yuste 2015). A schematic view on interconnected neurons and their components is given in Fig. 2.2 A.

Concurrently, the neuron doctrine was broadend towards a *neuronal network concept* already proposed by Exner (1894). Exner stated that neurons interconnect at their dendritic and axonal branches, thereby, forming a network-like architecture out of several neurons (Dierig 1994). At the same time, based on the idea that mental functions require a cellular control element, Schleich (1894) stated that the nervous system is not only composed of neurons but also of *glial cells*. Moreover, he proposed that both cell types are stongly associated and functionally interlinked with each other. Thus, he concludes that the brain's function depends on the activity of neurons as well as glial cells. However, glia cells are thought to mainly provide support and protection for neurons in their function of information transmission and storage (Jessen and Mirsky 1980). Furthermore, Schleich (1894) proposed that the information flow between connected neurons is transmitted via an internal gap, called *synaptic cleft* (Fig. 2.2 A). The existence of such structures were proven and visualized by the introduction of electron microscopy (Porter et al. 1945) decades later (Dierig 1994).

2.1.2 Neuronal Mechanisms of Information Processing

To understand how information is encoded and integrated across the brain is an important issue. As already mentioned, the brain is made up of neuronal networks that allow communication between neurons that act as information channels. Neurons are highly specialized for receiving input signal, processing signal information and transmitting the signal towards interconnected target cells. By this, single neurons are the basic units of information processing. A connection between two neurons, called *synapse* (Fig. 2.2 B), is made when the presynaptic neuron's terminal, located in its axon, adheres to any location of the postsynaptic *dendrite*, a tree-like extension that arise from the neuronal cell body (Stuart et al. 1997). The presynaptic's specific information, that is transmitted to the postsynaptic neuron, is encoded by the so-called *action potential*s (APs) or *spikes*. Whereby, a single AP is generated in an all-or-none fashion at the neuron's axon hillock. Whenever the presynaptic's potential is depolarized enough to reach a threshold value, an AP is emitted in a stereotypical form of a sharp electric depolarization of the cell. Those APs
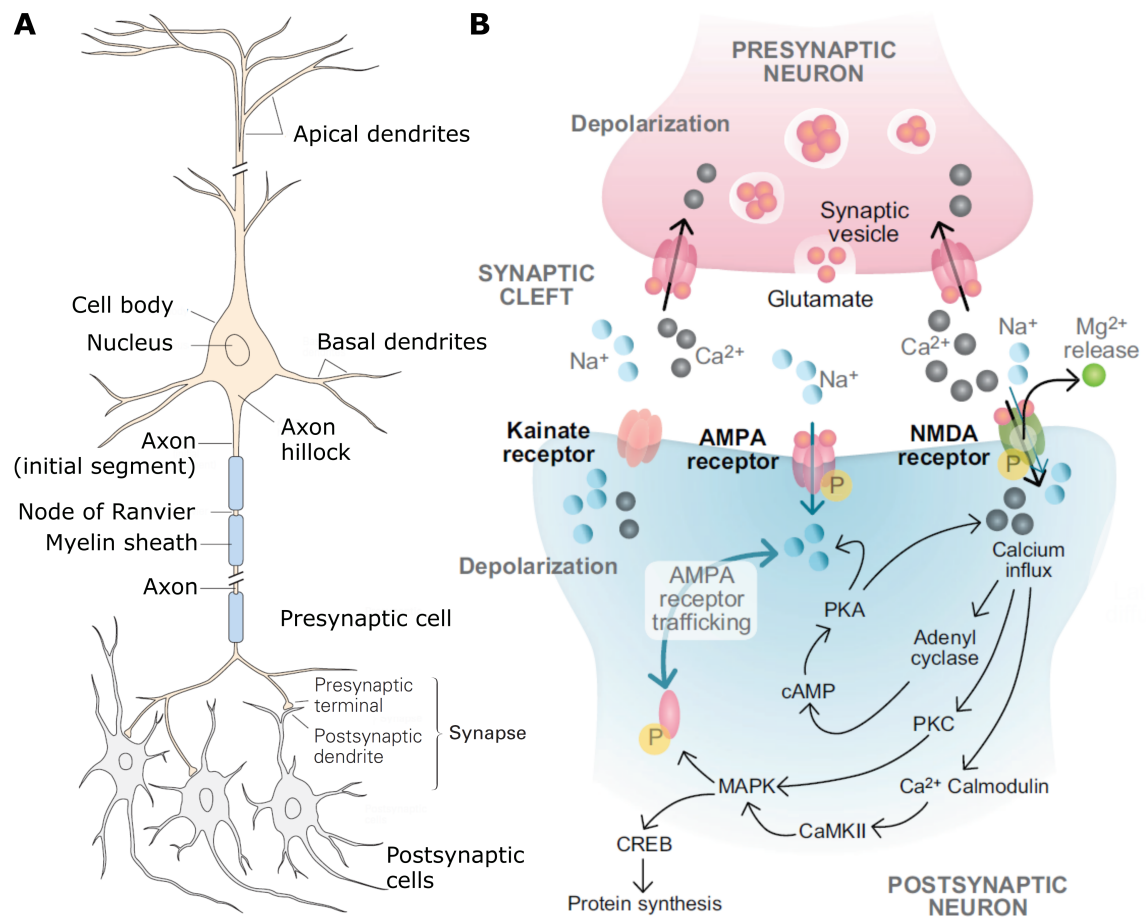
Figure 2.2: Structure of a neuron and its signaling pathways at synapses. (A) Structure of a neuron. Details see main text (Figure taken from Kandel et al. 2013). **(B)** Signaling pathways from a pre- to a postsynaptic neuron at a chemical synapse. Details see main text (Figure taken from Korte and Schmitz 2016).

are transmitted along the neuron's axon to its terminals, so-called *axon terminals*. Such axon terminals form the presynaptic part of the synapse (see Fig. 2.2).

There are two different types of synapses connecting the presynaptic with the postsynaptic nerve cell. Both neurons are interconnected via either an *electrical* or *chemical synapse*. At electrical synapses, the presynaptic signal is directly transmitted towards the postsynaptic cell. Here, specific ion channels, so-called gap junctions, connect the cytoplasm of both neurons. Hence, a presynaptic electrical depolarization leading to electrical currents, is directly transmitted via the gap junctions towards the postsynaptic cell (Bennett and Zukin 2004). The more complex signal transmission at chemical synapses converts the presynaptic electrical signal into a chemical signal within the synaptic cleft that again elicit an electrical postsynaptic potential. Here, the arriving action potentials at the presynaptic side of the synapse opens *voltage-sensitive calcium channels* (VSCC), nested in the

presynaptic neuron's membrane. Due to the intra- and intercellular electrochemical gradient, this opening of calcium (Ca^{2+}) channels causes a Ca^{2+} influx into the presynaptic neuron that triggers the fusion of vesicles with the presynaptic membrane. The vesicels contain molecules so-called *neurotransmitters*. Due to the fusion of the vesicles with the membrane, the neurotransmitters are released into the synaptic cleft and further diffuse to the postsynaptic membrane. There, the released molecules bind neurotransmitter-specific receptors. These receptors are coupled to ion channels that are nested in the postsynaptic membrane. This binding causes the ion channels to open (or close) and, thus, cause different types of ions to flow into or out of the postsynaptic cell (Fig. 2.2 B). Thereby, the chemical signal is translated into an electrical one, as the postsynaptic membrane potential is altered. The specific evoked signal is called *postsynaptic potential* (PSP) (Dayan and Abbott 2001; Trappenberg 2010; Kandel et al. 2013). The evoked voltage amplitude or slope, of the postsynaptic neuron's response to a presynaptic action potential is correlated to synapse specific parameters and defined as the *synaptic strength*, *synaptic efficacy* or *synaptic weight* (Dayan and Abbott 2001). Those factors are, for instance, the number of available neurotransmitters, the probability of transmitter release, or the number of postsynaptic ion channels with their receptors so-called *postsynaptic density* (PSD) (Markram et al. 1998; Senn et al. 2001).

Each neuron contains and releases its specific type of neurotransmitters at every axonal terminal, commonly known as *Dale's principle* (Whittaker 1983). This principle allows for a two-type classification of neurons. Thus, neurons that contain neurotransmitters that trigger a posivite response of the postsynaptic cell are called *excitatory neurons*, whereas neurons that contain neurotransmitter leading to a negative response of the postsynaptic cell are called *inhibitory neurons* (Markram et al. 2015). Along this line *excitatory* and *inhibitory synapses* are distinguished. At inhibitory synapses, the opening or closing of ion channels hyperpolarize (decrease) the postsynaptic potential towards its resting potential evoking an *inhibitory postsynaptic potential* (IPSP). GABAergic (gamma-aminobutyric acid) ion channels are common for those inhibitory synapses (Kubota et al. 2016). In contrast, the types of neurotransmitters involved in excitatory synapses trigger a depolarization (increase) of the postsynaptic membran potential, so-called *excitatory postsynaptic potential* (EPSP). Glutamatergic channels that bound specific types of receptors (e.g. α-amino-3-hydroxy-5-mehtyl-4-isoxazolepropionic acid (AMPA) or *N*-methyl D-aspartate (NMDA) receptors) are common for excitatory synapses (Chua et al. 2010). The regular form of synaptic transmission at excitatory synapses is mediated by AMPA receptors. An opening of AMPA-gated ion channels triggered by glutamate binding leads to an influx of positively charged sodium (Na^+) ions into the postsynaptic neuron. Hence, this influx evokes a depolarization of the postsynaptic cell. In the same way, glutamate also operates on NMDA-gated ion channels. However, NMDA-gated ion channels are addition-

ally blocked under resting membrane conditions by magnesium (Mg^{2+}) ions. Hence, an opening of NMDA-gated ion channels further require a sufficient strong level of depolarization that cause the Mg^{2+} ions to diffuse out of the ion channel. Thus, the opening of NMDA-gated ion channels requires the binding of neurotransmitters as well as a sufficiently strong level of depolarization. Such a required level of depolarization typically comes from back-propagating action potentials in the postsynaptic cell. Due to the dependency on pre- and postsynaptic quantities on the opening of NMDA-gated ion channels, NMDA receptors act as a molecular switch, a real *molecular coincidence detector* for pre- and postsynaptic firing (Korte and Schmitz 2016). In addition to the Na^+ influx into the postsynaptic cell from the AMPA-gated ion channels, NMDA-gated ion channels lead to a Ca^{2+} influx. Such Ca^{2+} influx has important relevance in further signal cascades in the postsynaptic cell, influencing, for instance, the AMPA receptor activation through kinases and phosphatases (detailed pathways are shown in Fig. 2.2 B; Lisman et al. 2002; Korte and Schmitz 2016).

In the broader context of behavior such discrete events of transmitting single spikes across neurons are not sufficient for explaining mental functions. Due to the homogeneous shape of each single action potential, it is known that the information is not stored within the transmission of individual action potentials between neurons. It is rather assumed that the information is encoded by the number of emitted spikes or precise temporal signaling patterns of neurons. This particular characteristic of the brain and local circuits is abstracted by the term *neural coding* (Richmond et al. 1987; Shinomoto et al. 2009). Therefore, the frequency at which a neuron emits single action potentials, i.e. number of spikes per second $[\frac{\text{spikes}}{\text{s}}]$ or Hertz [Hz], is defined as the *neuronal firing rate* or *neuronal activity*.

2.1.3 Cell Assemblies as Functional Units of the Brain

Beyond the scope of encoding and transmission of information between single neurons, Lorente de Nó and Hebb attributed higher cognitive brain functions to the encoding and transmission of information on the level of interconnected neurons forming networks of neurons. Lorente de Nó (1934) observed that specific patterns of neuronal activity can last for up to half a second after the removal of the stimulus which originally evoked the activity. This prolonged neuronal activity is called *functional reverberation* (Lorente de Nó 1934; Lorente de Nó 1949). Hebb (1949) explained this phenomenon with specific groups of neurons, termed *cell assemblies* (CAs), the functional units of the brain. He proposed that neurons are recurrently connected with each other building strongly. interconnected groups of neurons Therefore, Hebb speculated that the connections between neurons can be altered in their strength due to an activity-dependent learning rule for *synaptic plasticity*. This rule describes a strengthening of the synapses, if the pre- and postsynaptic neurons

show causally correlated coincident activity. Whereas a synapse gets weakened in its strength, if pre- and postsynaptic activity show noncoincident activity. Thus, this form of Hebbian synaptic plasticity links neurons into CAs by strong synaptic connections, if they exhibit causally correlated neuronal activity (Korte and Schmitz 2016). Decades later, Shatz (1992) condenses the underlying principle for synaptic plasticity dependent on the correlated activity in pre- and postsynaptic neurons towards: "What fires togehter, wires together". Hence, strongly interconnected neurons within cortical circuits allow for the self-sustaining reverberating activity observed by Lorente de Nó (Yuste 2015; Sejnowski 1999). Hebb (1949) attributed to CAs the ability to encode input stimuli by their ability to *learn* patterns of activity; and therefore, to store the respective memory into the strength of their interconnected neurons. Herewith, Hebbian CAs are thought to represent the neuronal correlate of memory, as their specific activity pattern can get reactivated even if the original learning stimulus is only partially applied. This phenomena is called *pattern completion* Holtmaat and Caroni 2016. Hence, Hebbian CAs represent the fundamental building blocks of the brain's memory system. Furthermore, Hebb suggested that the activation of CAs can be transmitted towards other CAs of the cortex and activate them, if they are properly connected. This principle of *sequence of activations* between distinct CAs is associated to the process underlying thinking. Hence, it is responsible for the encoding of specific behaviors (Yuste 2015; Sejnowski 1999; Holtmaat and Caroni 2016).

Hereby, Hebb established a theory that connects the psychological concept of behavior, in particular learning and memory, with the underlying neuronal mechanisms of synaptic plasticity. With this theory, the description of higher cognitive brain functions is extended from a cellular approach, described by the neuron doctrine, towards a precise mapping of neuronal circuits and their circuit-specific communication, summarized by the *neuronal circuit doctrine* (Kandel et al. 2013). Thus, neuronal networks are thought to be responsible for cognitive capabilities (Kandel et al. 2013; Palm et al. 2014; Pulvermüller et al. 2014; Trappenberg 2010).

2.1.4 Learning and Memory Underlie Synaptic Plasticity

The Hebbian principle of relating the psychological fact of a memory to specific traces in the neural system (Josselyn et al. 2015) is summarized in the *Synaptic-Plasticity-and-Memory Hypothesis* (Dayan and Abbott 2001; Eichenbaum 2012; Martin et al. 2000). This hypothesis states that information is dynamically and interactively stored in stimulus-specific neuronal networks by particular patterns of synaptic strength called *memory traces* or *engrams* (Josselyn et al. 2015). Here, the basic dynamic mechanism underlying learning and memory is determined by activity-dependent synaptic plasticity experimentally

observed in many different brain regions, including hippocampus, neocortex and cere-bellum (Dayan and Abbott 2001). However, the memory trace can be stored in different forms. Hence, either the engram is consolidated for a lifetime, and thus, being stable over time; or can get destabilized by memory recall, termed *long-term memory* (LTM) (Korte and Schmitz 2016). Here, the underlying form of plasticity leading to LTM, called *long-term plasticity* (Wood et al. 2012), that last more than 20-30[min], is opposed to *short-term plasticity* (Zucker and Regehr 2002), that lasts only hundreds of milliseconds. Hence, short-term plasticity is considered to be responsible for *working memory* rather than for LTM.

There are different forms of synaptic plasticity acting on different time scales. They are based on different molecular and biophysical properties, such as molecular coincidence detectors or dynamics of electrical signaling mechanisms in dendrites and spines (Korte and Schmitz 2016). However, a neuron, neuronal type, or brain region is not restricted to a specific form of synaptic plasticity. Moreover, the synapses of a single neuron underlie different plasticity mechanisms. The induction of the concrete synaptic plasticity mecha-nism is mainly dependent on the specific zone of a neuron, i.e. , synapse's localization on a dendritic tree in relation to the soma (Korte and Schmitz 2016). In the following, we will introduce important experimentally observed forms of synaptic plasticity with their un-derlying molecular and electrical mechanisms as well as their functional role for synaptic plasticity.

Long-term Potentiation

The first experimental confirmation that synapses can change their strength, i.e., that they are plastic, was observed by Bliss and Lømo (1973) under in-vivo conditions at a hip-pocampal pathway. Bliss and Lømo could arouse a long-lasting strengthening in the elec-trophysiological response properties of artificially stimulated neurons. This increase in synaptic strength that lasts for at least one hour is now called *long-term potentiation* (LTP) (Bliss and Lømo 1973). The changes in synaptic strength leading to LTP can operate on different time scales. Changes that occur within 1-3 h are called *early LTP* (E-LTP) and are independent of protein-synthesis, and thus, gene expression. Whereas changes that last longer than 3 h depend on changes in gene expression, thus, in changes on transcrip-tion and translation of plasticity-related proteins. Those changes are defined as *late LTP* (L-LTP) (Bliss and Collingridge 1993; Frey 1997; Kandel 2001).

For the induction of LTP at a synapse, the respective PSP has to surpass a certain thresh-old resulting into sufficient Ca^{2+} influx via NMDA-gated ion channels (Fig. 2.3 left). Such an increase in postsynaptic cytosolic calcium concentration causes an increase of AMPA receptor gated ion channel exocytosis into the postsynaptic membrane. This process is
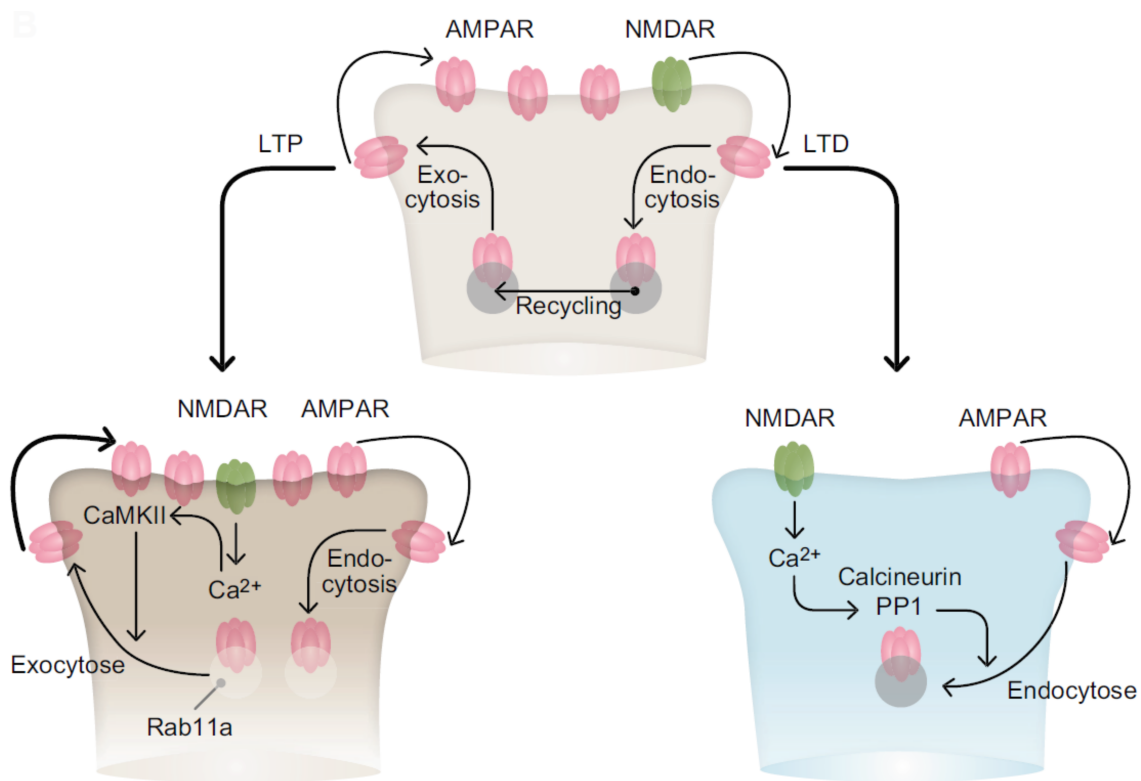
Figure 2.3: Basic cellular processes resulting in LTP or LTD. The basic cellular processes that occur at the postsynaptic cell under baseline conditions (top), LTP (left), or LTD (right). Details see main text (Figure taken from Korte and Schmitz 2016).

stabilized by Ca^{2+}-mediated processes including protein kinases (e.g., CaMKII). Furthermore, the recycling of recently used AMPA receptor gated ion channels towards reusable ion channels within the postsynaptic cell and their fusion back to the postsynaptic membrane is mediated by the cytosolic G protein Rab11a (Huganir and Nicoll 2013). The increased number of AMPA receptor gated ion channels within the postsynaptic cell membrane causes a stronger depolarization of the PSP as compared to baseline level. This, in turn, leads to Ca^{2+}-ion influx that causes the postsynaptic cell membrane to incorporate more AMPA receptor gated ion channels. Thus, LTP process ultimately describes a positive-feedback loop (Abbott and Nelson 2000).

LTP exhibits three key features. The first feature is *input specificity* (Engert and Bonhoeffer 1997), meaning that the synapses that undergo LTP have to be specifically activated or, at least, they have to be located in close vicinity of activated neurons. This constrain excludes inactive synapses for LTP (Engert and Bonhoeffer 1997; Wigström et al. 1986). Furthermore, as previously stated, to induce LTP, the PSP has to surpass an intensity threshold that is required to open NMDA-gated ion channels. This required level of depolarization can not be caused by individual presynaptic action potentials. Hence, the

second LTP feature is based on *cooperativity* of afferent fibers (Kitajima and Hara 1991). If a stimulus is applied in high-frequencies onto a neuron, the respective evoked postsynaptic depolarization is constructed by two mechanisms. On the one hand, by the collective depolarizations originating from different synapses within a close spacial vicinity (heterosynaptic information), and on the other hand, by the temporally integrated depolarizations at single synapses within a short time-window (homosynaptic information) (Bloodgood et al. 2009; Govindarajan et al. 2006). These two key features contribute to the third feature of LTP, called *associativity*. This mechanism describes the ability of weak synapses to get potentiated, if they concur with strong synapses in close spacial vicinity. Here, the depolarization of such strong synapses supports the particular postsynaptic potential to further increase, and by this, helps to reach a sufficient level of depolarization for LTP to occur (Kitajima and Hara 1991).

Long-term Depression

A counterpart to LTP is the process named *long-term depression* (LTD). This process was first described by Dudek and Bear (1992) and Mulkey and Malenka (1992). LTD establishes necessary differences in strength between synapses and further prevents excessive synaptic activity caused by LTP (Korte and Schmitz 2016). Similar to LTP also LTD is a self-enforcing positive-feedback loop. In contrast to LTP, low postsynaptic cytosolic Ca^{2+}-concentration triggering the endocytosis of AMPA receptor gated ion channels describes a positive-feedback loop that destabilizes neuronal activity.

Since LTP is induced by strong stimulation of the synapse causing the PSP to surpass a certain threshold, LTD is already induced at a lower level of PSP aroused by low-frequency stimulations leading to a moderate Ca^{2+} influx into the postsynaptic cell through NMDA-gated ion channels (Mulkey and Malenka 1992). This modest calcium influx activates postsynaptic cytosolic phosphatases such as calcineurin (PP2B), or protein phosphatases 1 (PP1) that guide the endocytosis of AMPA receptor gated ion channels into the postsynaptic cell. Subsequently, these ion channels are stored in endosomes or are further degraded (Lee et al. 2000; Song and Huganir 2002).

Another mechanism that mediates LTD is based on the activation of metabotropic glutamate receptors (mGluR) (Bashir et al. 1993; Huber et al. 2000); however, the underlying mechanisms are not completely understood. So far, it is confirmed that mGluR bind to G proteins, that results in a decreased frequency of the PSP, rather than to ion channels, that results into a decreased amplitude of the PSP. Thus, this form of LTD seems to be dependent on presynaptic mechanisms (Atwood et al. 2014).

It is confirmed that both forms of LTD mechanisms rely on different signaling mechanisms. However, the NMDA gated form of LTD is only reported in juvenile brain states (Korte and Schmitz 2016), thus, attributed to refining the young neuronal network by pruning less important synapses during late phases of the development. Whereas, the mGluR gated form of LTD is still observed in adult brain states (Huber et al. 2000), hence it is attributed to the balance of synaptic weights in the adult hippocampus (Korte and Schmitz 2016).

Spike-timing Dependent Plasticity

As the impact of LTP and LTD are described by dynamic processes such as the amount of transmitters and level of postsynaptic depolarization, *spike-timing dependent plasticity* (STDP) confirms the Hebbian postulated dependency on causally correlated neuronal activity of discrete pre- and postsynaptic firing (Korte and Schmitz 2016). The order and precise spike-timing of pre- and postsynaptic neurons is crucial for STDP and related to the opening of NMDA gated ion channels. Furthermore, STDP is related to LTP and LTD as it governs the induction of synaptic plasticity and defines its sign and magnitude (Feldman 2012; Korte and Schmitz 2016).

Markram et al. (1997) experimentally observed STDP in different brain areas by pairing experiments with whole-cell voltage recordings, where pre- and postsynaptic spikes where precisely triggered in time (Markram et al. 1997; Feldman 2012). With these experiments Markram et al. (1997) suggested that AP initated at the soma are back-propagated into the dendrite and, by this, serve to modify synaptic strength. Such an AP that runs backwards from the soma to the dendritic compartements is termed *back-propagating action potential* (bAP). In detail, if a presynaptic spike preceds the postsynaptic bAP within a narrow time window of 5-15 ms, the intertwined synapse get strengthened leading to LTP. Such a stimulation paradigm mimics the presynaptic contributes to the firing of the postsynaptic neuron, and consequently both spikes become causally correlated. In contrast, when a postsynaptic spike shortly precedes a presynaptic spike leads to LTD. This particular form of synaptic plasticity, dependent on causal (inducing LTP) and noncausal (inducing LTD) relation of pre- and postsynaptic activity is called *Hebbian STDP* (Abbott and Nelson 2000; Markram et al. 1997; Feldman 2012; Korte and Schmitz 2016).

The underlying biophysical mechanism yielding Hebbian STDP is similar to that contributing to LTP, as it depends on the activity-dependent increase in cytosolic Ca^{2+} concentration (Malinow and Malenka 2002; Huganir and Nicoll 2013; Korte and Schmitz 2016). Here, the biophysical mechanisms determined by bAPs contributes to the postsynaptic level of Ca^{2+} concentration within the postsynaptic nerve cell. The bAP opens

VSCCs in the postsynaptic dendrites and, by this, contributes to a higher level of depolarization. A sufficiently high concentration of postsynaptic Ca^{2+}-ions can be received in two ways: (i) by the opening of NMDA gated ion channels and VSCC enhanced by bAPs and/or (ii) by the activation of mGluR mediated by another coincidence detector named *phospholipase C* (PLC) resulting in a release of cytosolic Ca^{2+}-ions from cellular stores in combination with an bAP-dependent opening of VSCC (Koester and Sakmann 1998; Korte and Schmitz 2016; Hashimotodani et al. 2005).

Thus, the dependency of precise pre- and postsynaptic firing reflects the influence on the coincidence detectors such as NMDA receptors and PLC, as well as on the electrical dynamics gated by AMPA receptors and bAP, on the overall postsynaptic cytosolic Ca^{2+} concentration inducing plasticity (Korte and Schmitz 2016). However, bAPs, that function as postsynaptic spikes, show one major difference to APs, that function as presynaptic spikes. The difference is that the amplitude of depolarization caused by bAP decreases with distance to the soma and disappears completely in distal parts of the dendrites (Spruston 2008). Therefore, the influence of STDP is highly dependent on the localization of synapses on the dendritic tree, i.e., distance of synapses to the postsynaptic soma and seems to affect mostly proximal synapses (Froemke et al. 2005).

In addition to the cellular importance of STDP on single synapses, it might also be influential on the competition between convergent inputs (Abbott and Nelson 2000) and establishing temporal sequences (Rao and Sejnowski 2001), by this, being instrumental for sensory perception and to the development of sensory systems (Song and Abbott 2001; Korte and Schmitz 2016). Furthermore, the opposite case, so-called *Anti-Hebbian STDP*, has been observed by Feldman (2012). In this case, within a narrow time window, a preceding presynaptic (postsynaptic) spike before a postsynaptic (presynaptic) spike leads to LTD (LTP) (Feldman 2012).

Homeostatic Plasticity

As already mentioned, synaptic plasticity, through LTP and LTD, has strong destabilizing effects. Such effects can drive plastic synapses into diverging synaptic weight dynamics by evoking a positive-feedback loop. Cannon (1932) proposed a mechanism, termed *homeostasis*, that counterbalances such diverging dynamics in complex physiological systems and, by this, enforces stability. Here, homeostasis is thought to maintain the consistency of the internal state, in particular stabilize neuronal activity by an interplay of additional mechanisms (Cannon 1932; Turrigiano and Nelson 2004). The coexistence of stability and

adaptability in the network structure of a neuronal systems is mandatory to appropriately react on changes in the environment, and therefore, for the generation of behavior (Turrigiano 2012).

Although the functional role of homeostatic mechanisms serving as a regulatory mechanism is generally accepted, the underlying molecular dynamics and pathways are under intense investigation and still puzzling (Turrigiano 2012). However, different mechanisms that could generate stability in neuronal network activity, pooled by the term *homeostatic synaptic plasticity*, have been asserted. Those mechanisms are activity-dependent regulation of intrinsic neuronal firing (Marder and Prinz 2003; Zhang and Linden 2003), homeostatic regulation of intrinsic excitability (Marder and Goaillard 2006; Turrigiano 2011), adjustemts in the synaptic weights (Turrigiano and Nelson 2004) or the amount of excitatory synapses of a neuron (Kirov et al. 1999; Wierenga et al. 2006), balance of excitation and inhibition within neuronal networks (Maffei et al. 2004; Gonzalez-Islas and Wenner 2006), and metaplastic mechanisms regulating the induction of LTP and LTD (Bienenstock et al. 1982).

Turrigiano et al. (1998) observed one such regulatory mechanism, that prevents synaptic plasticity to induce hyper- or hypoactive neurons, by regulating the overall excitability of a neural circuit. This mechanism is called *synaptic scaling* (Turrigiano et al. 1998) and is the best understood type of homeostatic synaptic plasticity (Turrigiano 2012). The regulatory process selectively causes a cell to autonomously *up-scale* or *down-scale* the PSD depending on a reduced or elevated network activity, respectively. Here, synaptic scaling regulates the trafficking of AMPA receptor gated ion channels dependent on the postsynaptic cytosolic Ca^{2+} concentration in the opposite direction as, that described for LTP and LTD. Up scaling is induced by, for instance, a drop in postsynaptic calcium influx. The resulting biophysical mechanisms, yet not well understood, enhance the accumulation of postsynaptic AMPA receptors counteracting the reduced Ca^{2+} influx and maintaining a certain level of cytosolic Ca^{2+} concentration. This process results in a strengthening of the synaptic weight. Down scaling occurs in response to an enhanced calcium concentration in the postsynaptic cell and causes biophysical mechanisms that reduce the AMPA receptor accumulation in the postsynaptic membrane weakening the specific synaptic strength. This form of homeostatic synaptic plasticity promotes stablity in the network activity by strengthening or weakening the synaptic weight in the right direction to stabilize firing (Turrigiano et al. 1998; Turrigiano 2012; Abbott and Nelson 2000).

Inhibitory Plasticity

Heretofore, we have only considered activity-dependent synaptic plasticity acting on excitatory synapses, responsible for learning and memory. Although this form of synaptic plasticity constitutes the main focus for the study of synaptic plasticity, the field of *inhibitory synaptic plasticity* has recently gained more attention (Vogels et al. 2013). As excitatory synaptic plasticity, long-term activity-dependent inhibitory synaptic plasticity also provides refinement and sensitivity to neuronal circuits as well as experience-dependent learning and memory (Woodin and Maffei 2011; Vogels et al. 2013; Castillo et al. 2011). Therefore, inhibitory synaptic plasticity at GABAergic synapses regulates neuronal excitability and the impact of excitatory synapses (Woodin and Maffei 2011; Rozov et al. 2017; Castillo et al. 2011).

There are plenty of observed biophysical expression mechanisms mediated by changes in the GABAergic circuitry that guide inhibitory synaptic plasticity in two ways: is by changing properties in the presynaptic GABA release, or the postsynaptic number, sensitivity, and/or , responsiveness of $GABA_A$ receptors (Gaiarsa et al. 2002; McBain 2009). Thus, this form of plasticity depends on pre- and postsynaptic characteristics. For a review on the underlying biophysical mechanisms involved in the different forms of inhibitory synaptic plasticity see Castillo et al. (2011) and/or Rozov et al. (2017). Withal it is reported that excessive neuronal activity may trigger plasticity at GABAergic synapses (Castillo et al. 2011). Furthermore, there are some properties inhibitory and excitatory synaptic plasticity have in common. One joint characteristic is the associativity of inhibitory synaptic plasticity. This characteristic only modifies synapses with coincident activities of the enclosing pre- and postsynaptic neurons (Castillo et al. 2011). Another common characteristic is their homeostatic regulation that has emerged as an important mechanism for the control of the excitability and stability of neuronal circuits (Kilman et al. 2002; Woodin and Maffei 2011). Although, the induction of inhibitory synaptic plasticity slightly differs from that for excitatory synapses, both plasticity mechanisms interact with each other and can occur simultaneously (Castillo et al. 2011).

2.1.5 Schemata: The Building Blocks of Cognition

In the previous section we presented how a specific piece of knowledge (i.e. memory) is decoded in the brain by a specific group of neurons (i.e. Hebbian CAs) that are strongly interconnected due to activity-dependent synaptic plasticity processes. Those *memory representation*s (MRs) can be recalled, i.e. activated, by recall stimuli similar to the initial learning stimulus. However, it is well known that the encoding and recall of individual isolated memories is not sufficient to describe behavior, in particular cognitive abilities

(McKenzie and Eichenbaum 2011; Preston and Eichenbaum 2013). In this Section, we will outline the psychological view on the brain's main principles that allow for complex behavior.

James (1890) described the brain's ability to classify memories that helps to integrate new memories to the existing knowledge. Moreover, this ability leads to an organized mental representation of memories. Therefore, he suggested the *discriminaton* of memories that enables to distinguish different individual memories from another. Furthermore, James proposed that an *association* of memories decodes the coherence of several qualities of the individual memories. In the early 20th century, cognitive psychologists stated that a response, i.e. behavior, to a specific environmental stimulation arises from a comparison of that particular stimulus to the prior knowledge. Such phenomena can be represented by a cognitive structure that links and relates multiple representations of a phenomenon, termed *schema* (Head and Holmes 1911; Piaget 1926; Bartlett 1932). This schema is activated by new information and consequently, alters the individual's interpretation on the current situation (Piaget 1926). Bartlett (1932) specified that any incoming stimulus to an individual has to directly activate a specific portion of the *organized setting of past responses*, i.e. schema, which is most relevant in the current situation. He further examined that schemata are constantly developing and demonstrated their implication on memory retrieval processes (Ghosh and Gilboa 2014; McKenzie and Eichenbaum 2011; Preston and Eichenbaum 2013).

More recent advocates of relating schemas with behavior are, for instance, Rumelhart (1980) and Cooper et al. (1995) who acknowledged schemata being responsible for sequences of action. Rumelhart (1980) viewed schemata as knowledge structures with inherent meta knowledge about its utility and usage. Furthermore, Cooper et al. (1995) proposed that each schema has its inherent aim and tools to achieve a goal (Rumelhart 1980; Cooper et al. 1995; Ghosh and Gilboa 2014). Today such schemata, that serves to guide context-specific behavior by their inherent characteristic of embedded response options, are called *action schemata* (Rumelhart and Ortony 1976; Goodman 1980; Humphreys and Forde 1998; Cooper et al. 1995). In contrast, there is also the category of schemata that only account for knowledge structures and do not account for behavior (Preston and Eichenbaum 2013).

As briefly outlined, there exists several ambiguities for the definition and description of a *schema*. However, Ghosh and Gilboa (2014) reviewed the psychological and educational literature on memory neuroscience and proposed a general approach to describe a schema as a *cortical memory network* that allow for the encoding (Head and Holmes 1911; Piaget 1926; Bartlett 1932; Carmichael et al. 1932; Bransford and Johnson 1972; Rumelhart 1980;

Anderson 1984; Tse et al. 2007; Preston and Eichenbaum 2013; Kesteren et al. 2013) and retrieval (Bartlett 1932; Anderson and Pichert 1978; Rumelhart 1980; Anderson 1984; Arkes and Freedman 1984) of memory as well as the elaboration of context-specific patterns of behavior (Head and Holmes 1911; Bartlett 1932; Rumelhart 1980; Cooper et al. 1995; Shea et al. 2008; Kumaran et al. 2009).

Furthermore, Ghosh and Gilboa identified four necessary schema features that are central to the functionality of a schema. First, a schema constitutes an *associative network structure* that is composed of units (i.e. elements, events, variables, schema nodes) and their interrelationships (Fig. 2.4 gray networks). Please note that, the links (straight lines) between units (circles) are often considered to be more crucial for behavior than the units (Rumelhart 1980). Second, one schema is *based on multiple episodes* of one phenomenon and, by this, decodes their similarities (Fig. 2.4 colored networks). This feature implies the third feature, the *lack of unit details* (Fig. 2.4 gray networks with same thickness for all lines). Herein, single units of a schema represent distributions of real events rather than their clear definition; thereby, schemata is able to tolerate deviations from standard or reality. This feature further accounts for the generality and flexibility of schemata to organize new information. And last, schemata are *adaptive*, accounting for learning of specific behavior. Thus, schemata can be altered to be consistent with the external reality by two processes called *assimilation* and *accommodation*, cued by recalling previously stored information and learning new information (Fig. 2.4 green network and dotted green lines in the adapted schema) (Ghosh and Gilboa 2014; Preston and Eichenbaum 2013). These two processes of assimilation and accommodation were introduced and defined in the early 20th century. Piaget (1929) proposed that new memories are *assimilated* into existing schema in either two ways: If the new memory is consistent with an activated schema by a stimulus, the particular new memory gets incorporated into the specific activated structure of prior knowledge without challenging the existing relationships within the schema. Whereas, if the new memory is inconsistent with already existing schema, this memory undergoes modification until it fits the schema and subsequently gets assimilated to it (Piaget 1929; Preston and Eichenbaum 2013). In addition, Bartlett (1932) proposed that an existing schema undergoes modifications to *accommodate* new, conflicting memory. Moreover, new information that conflicts with existing schemata could also lead to the development of new schema (Preston and Eichenbaum 2013).

Ghosh and Gilboa (2014) further described four features to which schemata are sensitive. First, schemata that contain event-based information are likely to exhibit *chronological relationships* that can guide behavior. Second, schemata that store more complex information exhibit a *hierarchical organization* of memory that is enabled by specific connectivity among the units belonging to a schema. This hierarchical organization further allows
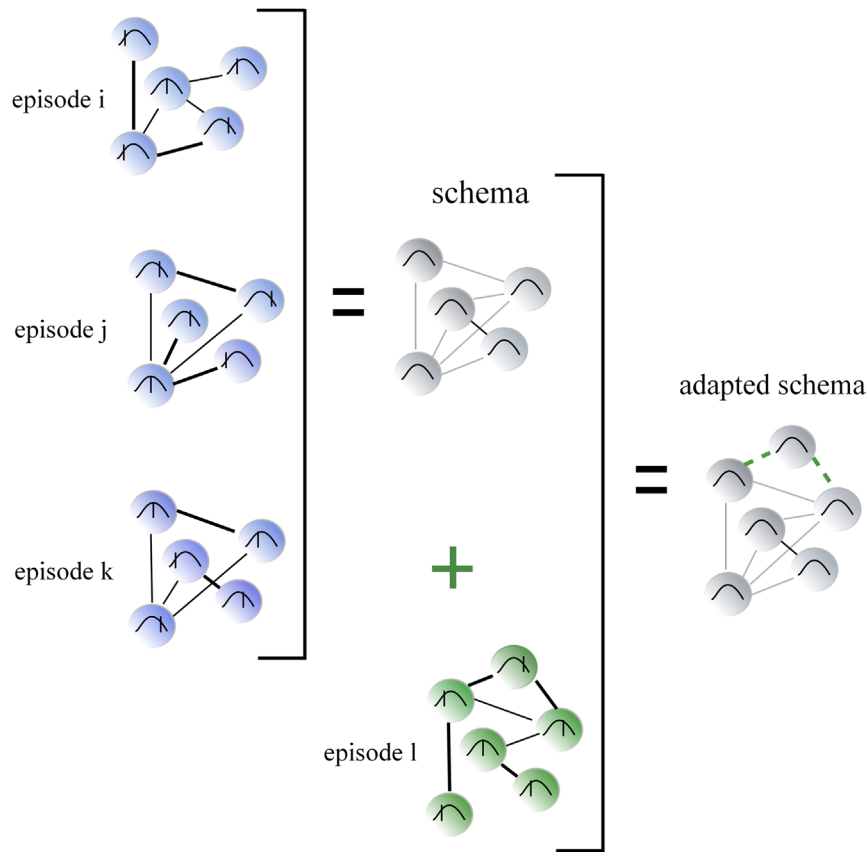
Figure 2.4: Necessary features of schema structure. Gray networks in the figure represent the schema as a latent neurocognitive structure of strongly interconnected nodes that could potentially be re-activated together. Colorful networks are either novel episodes or specific instantiations of the schema during a particular context of experience. The schema's *associative network structure (feature 1)* is depicted through circles, which represent schema units, and lines connecting those circles, which represent their associations. Differences in line connections and thickness indicate variability across episodes. The schema's *basis on multiple events (feature 2)* is illustrated through episodes i-k. Each episode differs in specificity, but all conform to the same general structure, which can be extracted as the schema. The schema's *lack of unit detail (feature 3)* is indicated by the normal distribution curve within each schema unit, which has the potential to take different values. For specific episodes or schema instantiations i-k, each unit takes a particular value on that curve. Lastly, the schema's *adaptability (feature 4)* is indicatd by the inclusion of new information from episode l as green dotted lines in the adapted schema (Figure and Caption taken from Ghosh and Gilboa 2014).

for the formation of sub-schemata; and thus, for top-down and bottom-up activation of cognitive structures. Closely linked to the hierarchical organization is the third feature, *cross-connectivity*, where different schemata share same overlapping units allowing the schemata to communicate or compete with each other. Lastly, schemata can have *embedded response options*, and by this, account for behavior.

There is experimental evidence for the existence of schemata and their specific features. Tse et al. (2007) underpinned the adaptability and development of schemata by paired-associative experiments with rats. In those experiments, the animals initially learned a specific set of food-location associations. After the learning phase, the rats were able to incorporate new associations within the same learning environment more rapidly than in a different environment. Thus, Tse et al. (2007) showed that new information can get assimilated to an existing schema that fits to the particular memory. Furthermore, Bunsey and Eichenbaum (1996) and Zeithamova et al. (2012) did experiments on the *transitive inference paradigm* where animals, which have learned multiple stimulus associations, such as A is associated with B and B is associated with C, are able to infer that A is associated with C. This ability to express knowledge of the relation between indirectly linked elements underpins the existence of an integrated organization of memory, hence, schemas (Preston and Eichenbaum 2013).

In summary, it is theoretically proposed and experimentally underpinned that behavior arises from the formation and interaction of the above described schemata. However, their formation is still not well understood. At present, there are several models describing the process of assimilating new information into pre-existing schemata. However, this process dependents on the communication of the cortex with other brain areas, such as the hippocampus or the brain area responsible for working memory. Hereby, the activity of the hippocampus is crucial for the maintenance and consolidation of most recent information (Scoville and Milner 1957; Smith and Squire 2009; McKenzie and Eichenbaum 2011). Whilst, the activation of cortical areas are most relevant for maintaining and recalling more remote memories (Haist et al. 2001; Douville et al. 2005; Bayley et al. 2006; McKenzie and Eichenbaum 2011). Different models that explain the communications of such brain areas and their different roles in memory consolidation are reviewed in McKenzie and Eichenbaum (2011) and called *cortical linkage*, *semantic transformation* and *schema modification*.

2.2 Computational Models of Learning and Memory

The field of *Computational Neurosciene* has revealed a vast number of mathematical models describing learning and memory acting on different levels of biophysical complexity, ranging from detailed molecule dynamics to collective network dynamics. These different levels of biological accuracy arise from the focus of study and its underlying scientific questions. As outlined in the previous Sections, within the context of behavior, detailed molecule or transmitter dynamics, or higher-level dynamics of discrete events, such as single APs, are not sufficient to understand the generation of cognitive brain functions.

In contrast, the number of emitted action potentials or precise temporal signaling patterns of neurons are thought to be crucial for transmitting information responsible for behavior (Richmond et al. 1987; Shinomoto et al. 2009). In order to address our specific scientific question on the principles underlying synaptic plasticity allowing for diverse mental organizations of memories, we focus on rate-coded nested in a recurrent neuronal network with its inherent modular organization for transforming sensory inputs, which is required for cognitive brain functions (Trappenberg 2010; Kandel et al. 2013; Palm et al. 2014; Pulvermüller et al. 2014). Thus, in this Section, we will outline some experimentally inspired approaches to mathematically model the biological dynamics of activity-dependent synaptic plasticity in neuronal networks of point neurons.

2.2.1 Neuron and Network Model

The state of an individual neuron i is generated by all incoming activities F_i from the interconnected neurons j pointed onto that particular neuron via either the synapse specific excitatory $\omega_{i,j}$ or inhibitory synaptic weight $\omega_{i,j}^-$ (Dayan and Abbott 2001). These afferent input stimulations are integrated towards a combined input current ϕ_i that determines a specific intrinsic *membrane potential* $u_i = f_u(\phi_i)$, defined as the difference between the electric potential within and outside a nerve cell. The membrane potential, in turn, defines the neuron's output rate, the neuronal firing rate or neuronal activity $F_i = f_F(u_i)$. Such a transformation of the membrane potential into a neuronal firing rate can be modeled by different activation functions f_F, such as a linear, step, threshold-linear, sigmoidal or radial-basis mapping (Trappenberg 2010).

In network theory (Papo et al. 2015), the brain or specific brain areas are interpreted as a network of nerve cells forming a massively parallel system by their links, via synapses, composed out of feed-forward and feedback connections in between the neurons. Such a network is called *recurrent neuronal network* and can change its connectivity p_{con}. In general, the connections between excitatory neurons are plastic, according to activity-dependent synaptic plasticity, while the inhibitory synapses are kept constant (Papo et al. 2015).

2.2.2 Models of Activity-dependent Synaptic Plasticity

Mathematical models of synaptic plasticity describe how specific factors drive the changes in efficacy of synapses ($\omega_{i,j}$), often formulated as *ordinary differential equation* (ODE). Those models aim to describe how the biophysical quantities, such as neuronal firing rates, precise spike timing or specific Ca^{2+}-concentration, influence changes in synaptic strength (Clopath 2015). Here, we focus on mathematical models of long-term plasticity rules

based on pre- (F_j) and postsynaptic neuronal firing rates (F_i) of the following form $\frac{\mathrm{d}\omega_{i,j}}{\mathrm{d}t} = \dot{\omega}_{i,j} = f(F_j, F_i, \omega_{i,j})$.

In line with Hebb's principle – "Who fires together wires together" (Hebb 1949; Shatz 1996), these models change the synaptic strength $\omega_{i,j}$ proportional to the correlation of the pre- (F_j) and postsynaptic firing rates (F_i) of the enclosing neurons $\dot{\omega}_{i,j} = \mu F_j F_i$, with μ being the learning rate. Since the neuronal firing rates only take positive values, such a formulation of synaptic plasticity exclusively increments the synaptic weight every time the learning rule is updated. Thus, Hebb's principle accounts for LTP without bound and does not include any processes leading to LTD (Dayan and Abbott 2001; Gerstner and Kistler 2002; Choe 2015).

This shortcoming lead to the formulation of covariance rules of the form of $\dot{\omega}_{i,j} = \mu(F_i - \theta_i)(F_j - \theta_j)$ (Sejnowski 1977). Here, the synaptic weights change relative to a mean, alternatively baseline or threshold value, for the pre- θ_j and/or postsynaptic firing θ_i. This model enables LTD, if neuronal activity falls below the threshold value (Stent 1973; Choe 2015). However such a regulation of the synaptic dynamics by assuming thresholds in the firing rates does not prevent unbounded synaptic weight dynamics (Dayan and Abbott 2001).

Further approaches to overcome this problem and simultaneously accounting for LTP and LTD are made by introducing either adaptive thresholds for the firing rates, implementing hard or soft bounds for the synaptic weights, or adding a synaptic weight normalization (Choe 2015; Clopath 2015; Gerstner et al. 2014; Dayan and Abbott 2001). The experimentally motivated BCM-rule (Bienenstock et al. 1982) introduces a nonlinear postsynaptic-activity dependent LTD in form of $\dot{\omega}_{i,j} = \mu F_j F_i(F_i - \theta_i)$. Here, θ_i is a sliding threshold modeled by threshold adaptation rule depending on the average output rate $\langle F_i \rangle$ and acting as a homeostatic mechanism (Dayan and Abbott 2001; Gerstner and Kistler 2002; Clopath 2015; Choe 2015). Moreover, Oja's model (Oja 1982) $\dot{\omega}_{i,j} = \mu F_i(F_j - F_i \omega_{i,j})$ with a multiplicative synaptic weight normalization further implies competition between the synapses connected onto the same postsynaptic neuron i and performs principle component analysis (Amari 1977; Oja 1982; Gerstner et al. 2014; Clopath 2015; Choe 2015). The weight normalization in Oja's model is based on the constant sum of squared weights ($\sum_j \omega_{i,j}^2 = 1$) (Miller and MacKay 1994), hence, describes a non-local principle.

Tetzlaff et al. (2011) combined both approaches, i.e., homeostatic plasticity and synaptic weight dependent regulation of the synaptic weight change, with the Hebbian formulation of synaptic plasticity resulting in $\dot{\omega}_{i,j} = \mu F_j F_i - \gamma(F_i - \mathrm{F^T})\omega_{i,j}^n$ ($n \in \mathbb{N}$), called *learning rule on Synaptic Plasticity and Synaptic Scaling* (SPaSS), with γ being the learning rate of the homeostatic mechanism. Here, the homeostasis describes the experimentally

observed form of synaptic scaling by Turrigiano et al. (1998), where the synaptic weights are scaled up or down when the postsynaptic activity is below or above the target firing rate F^T (Turrigiano et al. 1998; Turrigiano and Nelson 2004; Tetzlaff et al. 2011). This particular mathematical implementation of synaptic scaling is also able to stabilize other forms of synaptic plasticity rules, such as the BCM-rule, for at least a quadratic weight dependency $n \geq 2$ (Tetzlaff et al. 2011).

2.2.3 Models Describing Hebbian Cell Assemblies in Neuronal Networks

The above described SPaSS learning rule in a firing rate formalism reliably models the self-organized memory allocation (Auth et al. 2018), formation and consolidation of memories in recurrent neuronal networks (Tetzlaff et al. 2013; Tetzlaff et al. 2015). However, there are further approaches to form and retrieve memories, those models are based on more complex and computational expensive, as well as analytically not feasible, spiking neural network models (Litwin-Kumar and Doiron 2014; Zenke et al. 2015). For example, the model proposed by Litwin-Kumar and Doiron (2014) that combines a multitude of different plasticity mechanisms, such as spike-timing plasticity, synaptic normalization, as well as inhibitory spike-timing plasticity, to allow for the formation and consolidation of Hebbian *cell assemblies* (CAs) (Litwin-Kumar and Doiron 2014). Another model proposed by Zenke et al. (2015) comprises short-term plasticity, LTP and LTD, homeostatic plasticity, as well as, transmitter-induced plasticity and consolidation, to allow for the formation and retrieval of memories (Zenke et al. 2015).

Thus, due to the high abstraction level of the SPaSS learning rule (Tetzlaff et al. 2011; Tetzlaff et al. 2013) describing the synaptic dynamics in a firing rate formulation, this particular plasticity paradigm will be the starting point for the investigation in this thesis to examine the basic principles of synaptic plasticity responsible for the formation of diverse memory interactions.

3

Mathematical Formalism on the Interaction of Memories

In the previous Chapter (Sec. 2.1.5, p. 18), we have introduced the brain's principle to organize individual memories into a web of knowledge a so-called *schema* (Fig. 2.4, p. 21), which structures knowledge and allows for complex behavior. For modeling the dynamics of a schema, in Sec. 3.1, we set about to introduce the used recurrent neuronal network and neuron model (Fig. 3.1 A). Using this recurrent neuronal network model, in Sec. 3.2, we describe the dynamics of $n_{pop} \in \mathbb{N}$ individual and interconnected neuronal populations (Fig. 3.1 B). Here, we make three major assumptions for the population model: (i) all neuronal populations contain the same number of neurons, (ii) all neurons of a neuronal population receive input from the same subset of neurons of the input layer, (iii) each population-specific subset of the input layer contains as many neurons as the populations of the neuronal network. Two such neuronal populations ($n_{pop} = 2$) construct the smallest motif within the structure of a schema being its primary building block. Hence, using this generic neuronal population model, we restrict our numerical simulations and analysis on two anatomically interconnected memories ($n_{pop} = 2$). Hereby, our main objective is to classify all different forms of so-called *functional organization*s (FOs) of two such connected memories. We define those different forms of FOs based on the effect of a recalled memory to either excite or inhibit the interconnected memory, in accordance with the definitions propsed by Byrne and Huyck (2010):
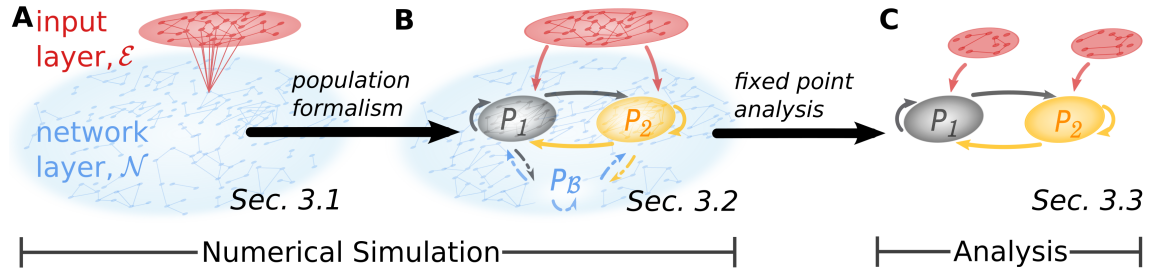
Figure 3.1: Procedure to analytically investigate the input-dependent functional organization of interconnected memories. (A) The used recurrent neuronal network (blue section) consists of recurrently connected neurons (blue circles). Here the individual neurons of the network are interconnected via plastic excitatory synapses (straight blue lines) and constant inhibitory synapses (not shown). Each neuron of the network receives neuron-specific external inputs from the input layer (red section) via constant excitatory synapses (red lines; example for the input stimulation of one neuron). **(B)** Description of $n_{pop} = 2$ individual neuronal populations nested in the recurrent neuronal network. All remaining neurons that are not part of a single neuronal population (p_1: black; p_2: yellow) of the network are combined to a background population ($p_\mathcal{B}$: blue). These background neurons serve as control neurons. All neurons belonging to the same neuronal population receive a population-specific input stimulation. **(C)** Analytical description of the neuronal populations at equilibrium by means of fixed point analysis. Here, the input stimulation onto each neuronal population originating from the background neurons ($p_\mathcal{B}$) is combined with the input layer.

Definition 1 *On the order of two interconnected memories, we define the FO of those as an*

- *association (asc), if both memories excite each other,*

- *sequence (seq), if one memory excites and the other memory inhibits its interconnected memory, and*

- *discrimination (disc), if both memories inhibit each other*

due to an appropriate recall stimulus.

For an evaluation of the dynamic FOs of memories underlying activity-dependent synaptic plasticity, we continue to explain how single neuronal populations decode specific environmental input stimuli. Here, the corresponding numerical analysis on the FOs of memories is computationally expensive and, furthermore, a corresponding synaptic-weight dependent analysis is not feasible. Therefore, by means of fixed point analysis, in Sec. 3.3 we provide an abstract low-dimensional population model at equilibrium state. With this reduced population model, the learning outcome (i.e. encoded by the synaptic weights) for a given input stimulation can be analytically calculated. Thus, this approach enables an analytic investigation on the FO of two interconnected memories on the respective synaptic weight space ($n_{pop} = 2$, Fig. 3.1 C). Therefore, in the Sections from Sec. 3.1 to

Sec. 3.3, we guide the reader step-by-step from the initial high dimensional neuronal network description considering individual neuronal activity and synaptic weight dynamics to the abstract low-dimensional description of interacting memories in a population model at equilibrium. Readers who are rather interested in the results on the underlying synaptic plasticity mechanisms can quickly skim the Sections from Sec. 3.1 to Sec. 3.3 and proceed in more detail with Sec. 3.4. For an evaluation on the interaction of two anatomically interconnected neuronal populations, in Sec. 3.4, we provide definitions of *memory representation* (MR) and the *functional organization* (FO) dependent on the coupling strength (i.e. synaptic weights) of the individual populations. Finally, we validate our derived method for the analysis of interconnected neuronal populations at equilibrium. For this, we compare the analytic results with those originating from the full network simulations for an exemplary synaptic plasticity rule.

The presented method and the corresponding conclusions in this Chapter have been published in the following article:

J. Herpich and C. Tetzlaff (2018). "Principles Underlying the Input-Dependent Formation and Organization of Memories". In: *bioRxiv*. A similar manuscript is currently under revision in Network Neuroscience.

3.1 Recurrent Neuronal Network Model

We consider a recurrent neuronal network model consisting of a set $\mathcal{N}$ of n rate-coded point neurons ($\mathcal{N} := \mathbb{N}_n = \{1,\ldots,n\}$, Fig. 3.1 A, blue dots). All neurons of the network layer $\mathcal{N}$ are excitatory and inhibitory connected with global (i.e. network wide) connection probabilities p_{con}^+ and p_{con}^-, respectively. Note, in our analysis, we assume $p_{con}^+ = p_{con}^- =: p_{con}$. By this, the *state of the network* is defined by:

- the activities $F_i \in \mathbf{F} := \mathrm{M}(n \times 1|\mathbb{R}_+)$ of all neurons i of the network ($i \in \mathcal{N}$),

- the excitatory synaptic weights $\omega_{i,j} \in \mathbf{W} := \mathrm{M}(n \times n|\mathbb{R}_+)$ connecting the presynaptic neuron j with the postsynaptic neuron i, as well as,

- all inhibitory synaptic weights $\omega_{i,j}^- \in \mathbf{W}^- := \mathrm{M}(n \times n|\mathbb{R}_+)$.

Note that the excitatory connections are plastic while the inhibitory connections are constant and globally set to $\omega_{i,j}^- := \theta \in \mathbb{R}_+$ throughout this thesis. Only in Sec. 4.4.2 we introduce inhibitory synaptic plasticity.

3.1.1 Environmental Input Stimulation

In addition to the network input, each neuron $i \in \mathcal{N}$ receives input from $n_i^{\text{ex}} \in \mathbb{N}$ different neurons (Fig. 3.1 A, red dots) of a neuron-i-specific subset $\mathcal{E}_i$ of the external input layer $\mathcal{E}$ (Fig. 3.1 A, red section; red lines indicate the neuron-i-specific input for one particular neuron i). The activities F_k^{ex} of all input neurons $k \in \mathcal{E}_i$, that are interconnected to neuron i of the recurrent network, are summarized by $\mathbf{F}_i^{\text{ex}} := \mathrm{M}(n_i^{\text{ex}} \times 1 | \mathbb{R}_+)$. Hence, the whole input stimulation of the neuronal network layer with n neurons is given by $\mathbf{F}^{\text{ex}} := (\mathbf{F}_1^{\text{ex}}, \ldots, \mathbf{F}_n^{\text{ex}}) = \mathrm{M}(n_i^{\text{ex}} \times n | \mathbb{R}_+)$. The specific input activities $F_k^{\text{ex}} \in \mathbf{F}^{\text{ex}}$ of these input neurons k are transmitted via excitatory synapses to the postsynaptic neuron i with global constant strength $\omega^{\text{ex}} := \omega_{i,k}$, $\omega_{i,k} \in \mathbf{W}_i^{\text{ex}} := \mathrm{M}(1 \times n_i^{\text{ex}} | \mathbb{R}_+)$.

3.1.2 Neuron Model

Each rate-coded point-neuron $i \in \mathcal{N}$ of the network integrates all incoming neuronal activities $F_j \in \mathbf{F}$ via the interconnected excitatory ($\omega_{i,j} \in \mathbf{W}$; Fig. 3.1 A, blue connections) and inhibitory ($\omega_{i,j}^- \in \mathbf{W}^-$; not shown) synapses, as well as, all incoming neuronal activities $F_k^{\text{ex}} \in \mathbf{F}_i^{\text{ex}}$ via the interconnected neuron-i-specific external synapses ($\omega_{i,k} \in \mathbf{W}_i^{\text{ex}}$; Fig. 3.1 A, red lines) to its neuron-specific *input current* (ϕ_i):

$$\phi_i(\mathbf{F}_i^{\text{ex}}, \mathbf{F}, \mathbf{W}, \mathbf{W}^-, \mathbf{W}_i^{\text{ex}}) = \underbrace{\sum_{j \in \mathcal{N}_i^+} \omega_{i,j} F_j}_{\text{network excitation}} - \underbrace{\sum_{j \in \mathcal{N}_i^-} \omega_{i,j}^- F_j}_{\text{network inhibition}} + \underbrace{\sum_{k \in \mathcal{E}_i} \omega^{\text{ex}} F_k^{\text{ex}}}_{\text{external excitation}}, \qquad [\phi] = 1\,\text{pA}.$$

(3.1)

Here, $\mathcal{N}_i^+$, $\mathcal{N}_i^- \in \mathcal{N}$ are the sets of indices for the excitatory and inhibitory presynaptic neurons j connected to the postsynaptic neuron i and $\mathcal{E}_i$ is the neuron-i-specific input layer, defined above. Although Dale's principle states that any neuron can have either excitatory or inhibitory connections, but not both at the same time, we merge both distinct neuronal sets for the excitatory and inhibitory neurons towards one neuronal population generalized by $\mathcal{N}$. This approach does not violates Dale's principle, it is just one technical simplification of the used model. Hereby, all excitatory neurons are interconnected to the inhibitory population, which, in turn, inhibits the excitatory population. Due to the constant inhibitory synaptic weights (θ) and constant external excitatory synaptic weights (ω^{ex}) the neuron specific input is simplified to:

$$\phi_i(\mathbf{F}_i^{\text{ex}}, \mathbf{F}, \mathbf{W}) = \sum_{j \in \mathcal{N}} (\omega_{i,j} - \theta)\, F_j + \omega^{\text{ex}} \left(\sum_{k \in \mathcal{E}_i} F_k^{\text{ex}} \right). \tag{3.2}$$

This neuron i specific current drives the *leaky membrane potential* (u_i) of the respective neuron and is described by the following *ordinary differential equation* (ODE):

$$\tau \dot{u}_i = -u_i + R\phi_i(\mathbf{F}_i^{ex}, \mathbf{F}, \mathbf{W}), \qquad \tau, R \in \mathbb{R}_+, \qquad [u] = 1\,\text{mV}, \qquad (3.3)$$

with τ ($[\tau] = 1\,\text{s}$) being the time constant of the membrane potential set to $\tau = 1\,\text{s}$ and $R = 0.1\,\text{n}\Omega$, ($[R] = 1\,\text{n}\Omega$) being the the membrane resistance. The neuron-specific membrane potential u_i is non-linearly mapped to a *neural firing rate* (F_i) by a sigmoidal transfer function:

$$F_i = \frac{F^{max}}{1 + \exp\left(\beta[\epsilon - u_i(\mathbf{F}_i^{ex}, \mathbf{F}, \mathbf{W})]\right)}, \qquad F^{max}, \beta, \epsilon \in \mathbb{R}_+, \qquad [F] = 1\,\text{Hz}, \qquad (3.4)$$

with global parameters of the neuronal network of the *maximal firing rate* $F^{max} = 100\,\text{Hz}$, the steepness β ($[\beta] = 1\,{}^1\!/\text{mV}$) and the inflexion point ϵ ($[\epsilon] = 1\,\text{mV}$) of the sigmoid. Thus, the firing rate of each neuron i varies in between $0\,\text{Hz}$ and F^{max} ($\mathbf{F} = M(n \times 1|[0, F^{max}])$). This sigmoidal transfer function approximates the characteristics of the neuronal activity: The neuronal acticity function only yields positive values (Gerstner et al. 2014) and convergences to a maximal firing rate F^{max} evoked by the refractory period for signal transmission in form of *action potential*s (APs). By combining Eq. (3.3) and Eq. (3.4), we simplify the description of the neuronal dynamics to one ODE for the neuronal firing rate:

$$\tau \dot{F}_i = f_F(\mathbf{F}_i^{ex}, \mathbf{F}, \mathbf{W}) := \frac{(F^{max} - F_i)F_i}{F^{max}} \left[\log\left(\frac{F^{max}}{F_i} - 1\right) + \beta[R\phi_i(\mathbf{F}_i^{ex}, \mathbf{F}, \mathbf{W}) - \epsilon]\right]. \quad (3.5)$$

As the dynamics of the neuronal firing rate (F_i) of all neurons i of the network underlie the same global rule (Eq. 3.5), we summarize the whole activity dynamics of the neuronal network to:

$$f_{\mathbf{F}^n}(\mathbf{F}^{ex}, \mathbf{F}, \mathbf{W}) := \begin{pmatrix} f_F(\mathbf{F}_1^{ex}, \mathbf{F}, \mathbf{W}) \\ \vdots \\ f_F(\mathbf{F}_n^{ex}, \mathbf{F}, \mathbf{W}) \end{pmatrix} \qquad (3.6)$$

with $\mathbf{F}^{ex} := (\mathbf{F}_1^{ex}, \ldots, \mathbf{F}_n^{ex})$.

3.1.3 Excitatory Synaptic Plasticity

In contrast to all inhibitory synaptic weights ($\omega_{i,j}^- \in \mathbf{W}^-$), all excitatory synaptic weights ($\omega_{i,j} \in \mathbf{W}$) in the recurrent neuronal network connecting the presynaptic neuron j with the postsynaptic neuron i undergo synaptic plasticity. There are multiple plasticity mechanisms and modeling approaches describing learning, introduced in Sec. 2.2.2. We will focus on their application in the next Chapter (Chapter 4) of this thesis. So far we only

specify that we consider activity-dependent learning rules for modeling synaptic plasticity only considering local state variables, thus the pre- and postsynaptic activities (F_j, F_i) and the synaptic weight ($\omega_{i,j}$) itself. Thus, we express the plastic excitatory synaptic weight by the following initial value problem:

$$f_\omega(F_j, F_i, \omega_{i,j}) := \tau_\omega \dot\omega_{i,j} \qquad \text{with} \qquad \omega_{i,j}(t=0) \in \mathbb{R}_+, \qquad [\omega] = 1\,\text{pC}, \qquad (3.7)$$

with τ_ω being the time constant for the synaptic weight dynamics. Again, as the dynamics of the excitatory synaptic weights ($\omega_{i,j}$) for all synapses of the network underlie the same global rule (Eq. 3.7), we summarize the whole excitatory synaptic weight dynamics of the neuronal network to:

$$f_{\mathbf{W}^n}(\mathbf{F}, \mathbf{W}) := \begin{pmatrix} f_\omega(F_1, F_1, \omega_{1,1}) & \dots & f_\omega(F_n, F_1, \omega_{1,n}) \\ \vdots & \ddots & \vdots \\ f_\omega(F_1, F_n, \omega_{n,1}) & \dots & f_\omega(F_n, F_n, \omega_{n,n}) \end{pmatrix}. \qquad (3.8)$$

3.2 Interaction of Memories in a Neuronal Population Model

In the following Section, we will derive a model of interacting neuronal populations (Fig. 3.1 B and Fig. 3.2 A for a more detailed description) nested in the recurrent neuronal network as introduced in the previous Section (Sec. 3.1). For this, we transfer the high dimensional recurrent network model defined by its single neuronal activity- and synaptic weight-dynamics to a low dimensional *neuronal population model* described by the dynamics of the average neuronal population's activity, as well as, the intra- and inter-population's synaptic weights.

To describe the dynamics of $n_{\text{pop}} \in \mathbb{N}$ interacting neuronal populations within the recurrent network $\mathcal{N}$, we will consider a set $\mathcal{P}$ of n_{pop} distinct subsets of neurons $\mathcal{P} := \{\mathcal{P}_{p_1}, \dots, \mathcal{P}_{p_{n_{\text{pop}}}}\}$. Note we refer to the population r as a whole by p_r whilst we refer to all neurons of the population r by $\mathcal{P}_{p_r} \subset \mathcal{N}$.

The specific subset of neurons $\mathcal{P}_{p_r} \subset \mathcal{N}$ belonging to the respective neuronal population r is selected by the population-r-specific subset of the external input layer $\mathcal{E}_r \subset \mathcal{E}$. Here, all neurons i, j of the specific neuronal population r receive input from the same input neurons k of the input layer $\mathcal{E}$. Thus, it holds $\mathcal{E}_i = \mathcal{E}_j =: \mathcal{E}_r$ for the population-r-specific input set. Consequently, each neuron $k \in \mathcal{E}_r$ of the population-r-specific input layer has an excitatory connection towards all neurons $i \in \mathcal{P}_{p_r}$ of the neuronal population r with constant synaptic weight strength ω^{ex}.
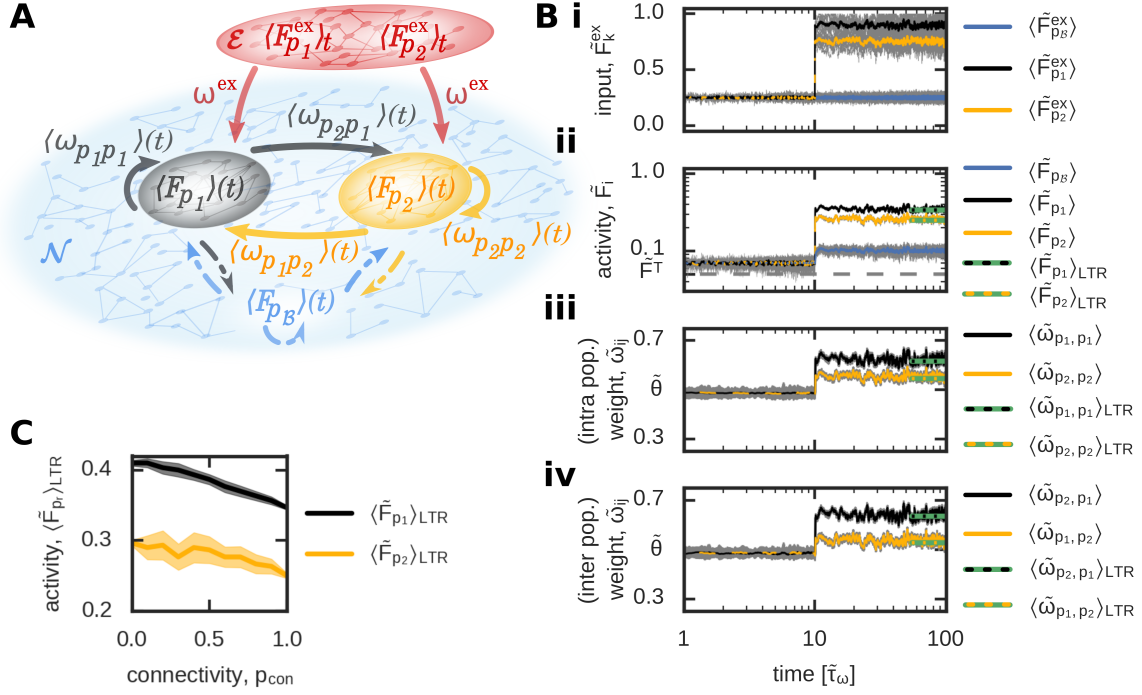
Figure 3.2: The interaction of two neuronal populations in a recurrent neuronal network. (A) In a recurrent network $\mathcal{N}$ (blue section), two neuronal populations ($\mathcal{P}_{p_1}$: black; $\mathcal{P}_{p_2}$: yellow) receive population-specific external inputs of average amplitudes $\langle F_{p_1}^{ex} \rangle_t$ and $\langle F_{p_2}^{ex} \rangle_t$, respectively. All remaining neurons that are not part of a specific neuronal population ($\mathcal{P}_{p_1}$: black; $\mathcal{P}_{p_2}$: yellow) of the network are combined to a background population ($\mathcal{P}_{p_\mathcal{B}}$: blue). These background neurons serve as control neurons receiving noisy external inputs $\langle F_{p_\mathcal{B}}^{ex} \rangle_t$. Each population $r \in \{1, 2, \mathcal{B}\}$ is described by its mean activity $\langle F_{p_r} \rangle$ and its mean intra-population-synaptic weight $\langle \omega_{p_r p_r} \rangle$ which changes over time. Furthermore, the mean inter-population-synaptic weights connecting two populations r, s are given by $\langle \omega_{p_r p_s} \rangle$ ($s \in \{1, 2, \mathcal{B}\} \backslash r$). **(B)** Numerical simulation of the complete $n + n^2$-dimensional network dynamics (gray lines, Sec. 3.1) for three interconnected neuronal populations 1, 2 and $\mathcal{B}$ in a normalized network model (see Appendix) for full connectivity ($p_{con} = 1$) and a population-specific input of $\tilde{\mathbf{F}}^{ex} = (0.9, 0.75, 0.25)^{\mathsf{T}}$ (see Sec. 3.2.1 for details on the input stimulation). Colored lines indicate the mean activity and synaptic weight dynamics of the different neuronal populations 1 (black), 2 (yellow), $\mathcal{B}$ (blue) (Sec. 3.2.2). The long-term-representation (LTR) of the applied population-specific input stimulation is evaluated for the second half of stimulation time $t \in [55, 100]\tilde{\tau}_\omega$ (indicated by green bars, Sec. 3.3.1). For the applied learning rule see Sec. 4.2.2 (SPaSS-learning rule) and for the input paradigm see Sec. 3.2.1. **(C)** Same input paradigm as in (B) for different network connectivities $p_{con} \in [0, 1]$. Solid line: LTR of the neuronal population's activities; Colored area: variance of the LTR of single neuronal activities of the respective neuronal population. Used parameters for (B) and (C) see Tab. 3.1; (adapted from Herpich and Tetzlaff 2018).

In our model of n_{pop} different interacting neuronal populations $r \in \{1, \dots, n_{\text{pop}}\}$, we make two assumptions to simplify the network structure and connectivity.

Assumption 1 *All neuronal populations consist of the same amount of neurons $n_{\mathcal{P}} \in \mathbb{N}$ and the neuronal sets of each two neuronal populations r and s do not overlap ($\mathcal{P}_{p_r} \cap \mathcal{P}_{p_s} = \emptyset$).*

Assumption 2 *We set the number n_i^{ex} of afferent input neurons projecting to a neuronal population r equal to the size of the neuronal population $n_i^{ex} = n_{\mathcal{P}} =: n_{\mathcal{P}}^{ex}$ to consider the same order of magnitude for the input population as for the neuronal populations within the network.*

The remaining neurons of the network, which are not part of any neuronal population $p_r \in \mathcal{P}$, are unified to the neuronal set $\mathcal{P}_{p_{\mathcal{B}}} := \mathcal{N} \backslash \mathcal{P}$ and defined as *background neurons*. We refer to this neuronal population by $p_{\mathcal{B}}$ with size $n_{\mathcal{B}} := |\mathcal{P}_{p_{\mathcal{B}}}| = n - n_{pop}n_{\mathcal{P}}$. With this approch, we reduce the high-dimensional neuronal network model to the interaction of $n_{\text{pop}} + 1$ different neuronal populations nested in a recurrent neuronal network (see Fig. 3.2 A for $n_{\text{pop}} = 2$ interconnected neuronal populations).

3.2.1 Environmental Stimulation of the Neuronal Populations

Each neuronal simulation starts with a tuning phase (Fig. 3.2 Bi), where all neurons of the network receive noisy input $F_k^{\text{ex}} \sim \mathcal{N}(0.25, 0.02)$ from $n_{\mathcal{P}}^{\text{ex}}$ different neurons k of the input layer $\mathcal{E}$. Along with the onset of the environmental input stimulation ($t = 10\,\tau_\omega$), all neurons $i \in \mathcal{P}_{p_r}$ belonging to a particular neuronal population $r \in \{1, \ldots, n_{\text{pop}}\}$ receive their neuronal population specific input stimulation:

Assumption 3 *As the input stimulation of a neuronal population r represents a specific environmental piece of information, we assume that the firing rate of all population-r-specific input neurons $k \in \mathcal{E}_r$ fluctuate around a mean value over time, described by $\langle F_{p_r}^{ex} \rangle_t$.*

With this mean population-r-specific input paradigm $\langle F_{p_r}^{\text{ex}} \rangle_t$ over time, the activities of all neurons k of the population-r-specific external input layer $\mathcal{E}_r$ are modeled by an Ornstein-Uhlenbeck process, respectively:

$$\dot{F}_k^{\text{ex}} = \underbrace{\delta(\langle F_{p_r}^{\text{ex}} \rangle_t - F_k^{\text{ex}})}_{\text{drift}} + \underbrace{\sigma\zeta}_{\text{diffusion}}, \quad \zeta \sim \mathcal{N}(0,1) \quad \text{with} \quad F_k^{\text{ex}}(t=0) = \langle F_{p_r}^{\text{ex}} \rangle_t \quad (3.9)$$

with the drift term scaled by $\delta = 0.025$ and with a normal distributed diffusion term with constant $\sigma = 0.0125$. All neurons k of the external input layer that are connected to the background neurons $i \in P_{p_{\mathcal{B}}}$ continue to fire at the noise level $F_k^{\text{ex}} \sim \mathcal{N}(0.25, 0.02)$ similar to the tuning phase.

Table 3.1: Used parameters for the numerical simulations in Fig. 3.2 - Fig. 3.4 and Fig. 3.6.

Network model		Neuron model		Synaptic plasticity & Other	
parameter	value	parameter	value	parameter	value
n	100 neurons	τ	1 s	τ_ω	60s
$n_\mathcal{P}$	10 neurons	R	$0.1\text{n}\Omega$	F^T	5 Hz
$n_\mathcal{B}$	$n - 2n_\mathcal{P}$	F^max	100 Hz	ω^max	97.33 pC
$n_\mathcal{P}^\text{ex}$	$n_\mathcal{P}$	n_ϵ	20 neurons	δ	2.5 Hz
$\tilde{\omega}^\text{ex}$	$1[\omega^\text{max}]$	β	$0.00035\frac{1}{\text{mV}}$	σ	1.25 Hz
$\tilde{\theta}$	$0.5[\omega^\text{max}]$			Δt	0.01s

Thus, given $n_\text{pop} + 1$ different neuronal populations $\mathcal{P} = \{\mathcal{P}_{p_1}, \ldots, \mathcal{P}_{p_{n_\text{pop}}}, \mathcal{P}_{p_\mathcal{B}}\}$, the parameter space of the input is defined by the different mean neuronal population-r-specific inputs over time:

$$\mathbf{F}^\text{ex} := \left(\langle F_{p_1}^\text{ex}\rangle_t, \ldots, \langle F_{p_{n_\text{pop}}}^\text{ex}\rangle_t, \langle F_{p_\mathcal{B}}^\text{ex}\rangle_t\right)^\mathsf{T}. \tag{3.10}$$

One exemplary simulation for a fully connected network ($p_{con}^+ = p_{con}^- = p_{con} = 1$) with $2 + 1$ interacting neuronal populations $\mathcal{P} = \{\mathcal{P}_{p_1}, \mathcal{P}_{p_2}, \mathcal{P}_{p_\mathcal{B}}\}$ is shown in Fig. 3.2 B (gray curves). For the simulation we used the normalized neuronal network model (see Appendix). Here, we drive the neuronal network with an input stimulation (Fig. 3.2 Bi) specified by the population-specific input paradigm

$$\tilde{\mathbf{F}}^\text{ex} = \left(\langle \tilde{F}_{p_1}^\text{ex}\rangle_t, \langle \tilde{F}_{p_2}^\text{ex}\rangle_t, \langle \tilde{F}_{p_\mathcal{B}}^\text{ex}\rangle_t\right)^\mathsf{T} = (0.9, 0.75, 0.25)^\mathsf{T}.$$

For the simulation, we numerically solve the *ordinary differential equation*s (ODEs) of the normalized model for all individual neuronal activity (Fig. 3.2 Bii) and synaptic weight dynamics (Fig. 3.2 Biii,iv) with the *Euler method* ($\Delta t = 0.01$s). Note, for reasons of clarity in the upcoming Sections, we only plotted the intra- (Fig. 3.2 Biii) and inter-population synaptic weights (Fig. 3.2 Biv) of both neuronal populations 1 and 2 and excluded the intra- and inter-population synaptic weights related with the population on the background neurons (grey lines indicate the single neuronal and synaptic dynamics). All used parameters for the numercial simulation are collected in Tab. 3.1.

3.2.2 Dynamics of the Neuronal Populations

To describe the time evolution of the different neuronal populations r and s within the recurrent neuronal network for a specific input stimulation $\mathbf{F}^{\text{ex}}$, we calculate the time-dependent mean state variables of this system of interconnected neuronal populations. Note, the mean state variables are taken over all neuronal activity and synaptic weight variables of one population not over time:

- the mean neuronal population activities $\mathbf{F}_{\mathcal{P}}(t)$ given by:

$$\mathbf{F}_{\mathcal{P}}(t) := \begin{pmatrix} \langle F_{p_1} \rangle(t) \\ \vdots \\ \langle F_{p_{n_{\text{pop}}}} \rangle(t) \\ \langle F_{p_{\mathcal{B}}} \rangle(t) \end{pmatrix} \quad \text{with } \langle F_{p_r} \rangle(t) := \tfrac{1}{n_{\mathcal{P}}} \sum_{i \in \mathcal{P}_{p_r}} F_i(t),$$

- the mean intra-population synaptic weights $\mathbf{W}_{\text{intra}}(t)$ given by:

$$\mathbf{W}_{\text{intra}}(t) := \begin{pmatrix} \langle \omega_{p_1, p_1} \rangle(t) \\ \vdots \\ \langle \omega_{p_{n_{\text{pop}}}, p_{n_{\text{pop}}}} \rangle(t) \\ \langle \omega_{p_{\mathcal{B}}, p_{\mathcal{B}}} \rangle(t) \end{pmatrix} \quad \text{with } \langle \omega_{p_r, p_r} \rangle(t) := n_{\mathcal{P}}^{-2} \sum_{j \in \mathcal{P}_{p_r}} \sum_{i \in \mathcal{P}_{p_r}} \omega_{i,j}(t), \text{ and}$$

- the mean inter-population synaptic weight $\mathbf{W}_{\text{inter}}(t)$ given by:

$$\mathbf{W}_{\text{inter}}(t) := \begin{pmatrix} - & \langle \omega_{p_1, p_2} \rangle(t) & \cdots & \langle \omega_{p_1, p_{n_{\text{pop}}}} \rangle(t) & \langle \omega_{p_1, p_{\mathcal{B}}} \rangle(t) \\ \langle \omega_{p_2, p_1} \rangle(t) & - & \cdots & \langle \omega_{p_2, p_{n_{\text{pop}}}} \rangle(t) & \langle \omega_{p_2, p_{\mathcal{B}}} \rangle(t) \\ \vdots & & \ddots & \vdots & \vdots \\ \langle \omega_{p_{n_{\text{pop}}}, p_1} \rangle(t) & & \cdots & - & \langle \omega_{p_{n_{\text{pop}}}, p_{\mathcal{B}}} \rangle(t) \\ \langle \omega_{p_{\mathcal{B}}, p_1} \rangle(t) & & \cdots & \langle \omega_{p_{\mathcal{B}}, p_{n_{\text{pop}}}} \rangle(t) & - \end{pmatrix}$$

with $\langle \omega_{p_r, p_s} \rangle(t) := n_{\mathcal{P}}^{-2} \sum_{j \in \mathcal{P}_{p_s}} \sum_{i \in \mathcal{P}_{p_r}} \omega_{i,j}(t)$.

These mean state variables correspond to the colored curves ($\mathcal{P}_{p_1}$: black, $\mathcal{P}_{p_2}$: yellow, $\mathcal{P}_{p_{\mathcal{B}}}$: blue) in the example shown in Fig. 3.2 B.

3.2.3 Dimensionality of the Neuronal Population Model

With this description of the mean state variables of $n_{\text{pop}} + 1$ interacting neuronal populations, we reduce the state space of the neuronal network model from $(n + n^2)$-dimensions with n dimensions for the neuronal activity space $\mathbf{F}$ and n^2 dimensions for the neuronal excitatory synaptic weight space $\mathbf{W}$ to $((n_{\text{pop}} + 1) + (n_{\text{pop}} + 1)^2)$-dimensions for the state space of the mean neuronal population model. Here, the mean neuronal population's activity space $\mathbf{F}_{\mathcal{P}}$, as well as the mean intra-population excitatory synaptic weights $\mathbf{W}_{\text{intra}}$

space each consists of $n_{\text{pop}}+1$-dimensions. The mean inter-population excitatory synaptic weight space $\mathbf{W}_{\text{inter}}$ consists of $(n_{\text{pop}}^2 + n_{\text{pop}})$-dimensions. These mean state variables for the neuronal population's activities and synaptic weights are depicted in the specific example (Fig. 3.2 B) for the interaction of 2+1 different neuronal populations by the colored curves (black: neuronal population 1, yellow: neuronal population 2, blue: background neurons $\mathcal{B}$).

3.2.4 Long-term Representation of Stimulation

As the neuronal populations are designed to represent specific environmental stimuli, we consider the *long-term representation* (LTR) of the stimulus in the population-specific state variables to analyze how a specific input paradigm is encoded within the neuronal network. The LTR is determined within a time window t_w of the input stimulation when the system's state variables have reached stable representations. Here, all neuronal and synaptic state variables fluctuate around mean values. These mean values determine the LTR of the specific input stimulation. In our neuronal network simulations (exemplary simulation shown in Fig. 3.2 B), we extract the LTR as the system's mean state variables over the specific time window of $t_w = [55, 100]\tilde{\tau}_\omega$, for the second half of stimulation time (indicated by green bars in Fig. 3.2 B). Here, we indicate the LTR of the system's state variables in the population model formalism by $\langle F_{p_r} \rangle_{\text{LTR}}$ for the neuronal population's mean activity and $\langle \omega_{p_r, p_s} \rangle_{\text{LTR}}$ for the mean excitatory synaptic weights connecting neuronal population s towards neuronal population r.

3.3 Interaction of Memories in a Population Model at Equilibrium

In Fig. 3.2 B, we have exemplary shown how a specific input stimulation is mentally encoded by the synaptic weights in the population model formalism. These synaptic weights, at stable state, fluctuate around mean values, which, are defined as the *long-term representation* (LTR) for the respective input stimulation. In this Section, by means of fixed point analysis, we derive a method to analytically determine the state variables for the LTR for a given input stimulation (Fig. 3.1 C) without simulating the whole neuronal activity and synaptic weight dynamics.

The here derived *population model at equilibrium* holds for specific neuronal systems with the following inherent condition: When the environmental input stimulation $\mathbf{F}^{\text{ex}}$ is constantly applied onto the network over time, specific learning rules f_ω lead the system into an *equilibrium state* $\mathfrak{E}(\mathbf{F}^{\text{ex}}, f_\omega)$, indicating the LTR of the respective input stimulation.

Condition 1 *Given a specific external input stimulation $\mathbf{F}^{ex}$ and a specific excitatory synaptic learning rule f_ω driving the dynamics of a neuronal population model, the system evolves into a equilibrium state $\mathfrak{E}(\mathbf{F}^{ex}, f_\omega)$ if and only if $f_{\mathbf{F}^n}(\mathbf{F}^{ex}, \mathbf{F}^*, \mathbf{W}^*) = \mathbf{0}$ and $f_{\mathbf{W}^n}(\mathbf{F}^*, \mathbf{W}^*) = \mathbf{0}$ being the zero matrix. Whereby, $\mathbf{F}^* := M(n \times 1 | [0, F^{max}])$ pools the fixed activities F_i^* of all neurons $i \in \mathcal{N}$ of the network and $\mathbf{W}^* := M(n \times n | \mathbb{R}_+)$ summarizes all fixed excitatory synaptic weights $\omega_{i,j}^*$ of the neuronal network. Throughout this thesis, we refer to the stable state by the following formalism:*

$$\mathfrak{E}(\mathbf{F}^{ex}, f_\omega) := \{\mathbf{F}^*(\mathbf{F}^{ex}, f_\omega), \ \mathbf{W}^*(\mathbf{F}^{ex}, f_\omega)\}. \tag{3.11}$$

In this thesis equilibrium *and* fixed point *are synonyms.*

Note, the state variables representing the numerically obtained *long-term representation* (LTR) for a given input stimulation are described by the notation of $\langle \circ \rangle_{\text{LTR}}$. Whereas the state variables representing the analytically determined system's equilibrium state is described by the notation of $\circ^*$. To clearly distinct numerically and analytically calculated LTRs, in the following, we consequently make use of $\langle F_i \rangle_{\text{LTR}}$, $\langle \omega_{i,j} \rangle_{\text{LTR}}$ for the simulation and F_i^*, $\omega_{i,j}^*$ for the analysis. However, both notations describe the stable state of the system for a given input paradigm.

3.3.1 Populations at Equilibrium

For all neuronal populations $r, s \in \{1, \ldots, n_{\text{pop}}\}$ of the network, we describe the average neuronal population's state variables at equilibrium as:

- the *fixed average neuronal population's activities*
  $\mathfrak{F} := (\mathfrak{F}_{p_1}, \ldots, \mathfrak{F}_{p_\mathcal{B}})^\mathsf{T}$ with $\mathfrak{F}_{p_r} := \frac{1}{n_\mathcal{P}} \sum_{i \in \mathcal{P}_{p_r}} F_i^*$,

- the *fixed average intra-population's excitatory synaptic weights*
  $\mathfrak{w}_{\text{intra}} := (\mathfrak{w}_{p_1,p_1}, \ldots, \mathfrak{w}_{p_\mathcal{B},p_\mathcal{B}})^\mathsf{T}$ with $\mathfrak{w}_{p_r,p_r} := n_\mathcal{P}^{-2} \sum_{i \in \mathcal{P}_{p_r}} \sum_{j \in \mathcal{P}_{p_r}} \omega_{i,j}^*$, and

- the *fixed average inter-population's excitatory synaptic weights*
  $\mathfrak{w}_{\text{inter}} := \left(\mathfrak{w}_{p_1,p_2}, \ldots, \mathfrak{w}_{p_{n_{\text{pop}}},p_\mathcal{B}}\right)^\mathsf{T}$ with $\mathfrak{w}_{p_r,p_s} := n_\mathcal{P}^{-2} \sum_{i \in \mathcal{P}_{p_r}} \sum_{j \in \mathcal{P}_{p_s}} \omega_{i,j}^*$.

Using these state variables in the population formalism of $n_{\text{pop}} + 1$-dimensions, we reduce the equilibrium state, previously defined on the individual fixed neuronal activities $\mathbf{F}^*(\mathbf{F}^{ex}, f_\omega)$ and fixed synaptic weights $\mathbf{W}^*(\mathbf{F}^{ex}, f_\omega)$ (compare Eq. (3.11) with the following Eq. (3.12)):

$$\mathfrak{E}(\mathbf{F}^{ex}, f_\omega) = \{\mathfrak{F}(\mathbf{F}^*(\mathbf{F}^{ex}, f_\omega)), \ \mathfrak{w}_{\text{intra}}(\mathbf{W}^*(\mathbf{F}^{ex}, f_\omega)), \ \mathfrak{w}_{\text{inter}}(\mathbf{W}^*(\mathbf{F}^{ex}, f_\omega))\}. \tag{3.12}$$

To enable a fixed point analysis of the population model to determine its LTR for a given input stimulation, the applied learning rule f_ω has to have a specific property:

Condition 2 *The fixed excitatory synaptic weights $\omega_{i,j}^*$ can explicitly be calculated only from the fixed pre- F_j^* and postsynaptic neuronal activities F_i^* and by a function $f_\omega^*(F_j^*, F_i^*)$. We refer to this function as the* fixed point function *of the synaptic learning rule f_ω.*

For the exemplary network simulation in Fig. 3.2 B, we note that for all neurons $i \in \mathcal{P}_{p_r}$ of a given neuronal population r the LTR of the single neuronal activities (gray lines) fluctuate around the LTR of the average neuronal population's activity (colored lines) with small variance. In Fig. 3.2 C, we plotted the average population's activities (solid lines) and the variance of the respective single neuronal activities for different network connectivities p_{con}. Thus, we see that for the whole parameter space of $p_{con} \in [0, 1]$ the variance of the individual neuronal activities compared to the population average is very small. Given such small deviations of the single neuronal dynamics, we can make a further assumption on the equilibrium state $\mathfrak{E}(\mathbf{F}^{ex}, f_\omega)$ of the population model:

Assumption 4 *Due to the same environmental input stimulation $\langle F_{p_r}^{ex} \rangle_t$ onto all neurons $i \in \mathcal{P}_{p_r}$ of a neuronal population r, we assume that the variance of each fixed neuronal activity to the fixed average population's activity is negligibly small. Thus, it holds*

$$F_i^* = \mathfrak{F}_{p_r} + \sigma_{F_i}^2 \qquad \text{with variance} \qquad \sigma_{F_i}^2 \approx 0. \tag{3.13}$$

Corollary 1 *The fixed activity of each neuron i of a neuronal population r approaches the fixed average activity of the respective neuronal population*

$$F_i^* \approx \mathfrak{F}_{p_r}. \tag{3.14}$$

At this point, we are able to analytically calculate the LTR for the population model, independent on the individual fixed neuronal activities $\mathbf{F}^*(\mathbf{F}^{ex}, f_\omega)$ and fixed synaptic weights $\mathbf{W}^*(\mathbf{F}^{ex}, f_\omega)$ (compare Eq. (3.12) with the following Eq. (3.15)):

Theorem 2 *Let $f_\omega(F_j, F_i, \omega_{i,j})$ be the learning rule of the neuronal system and $f_\omega^*(F_j^*, F_i^*)$ its fixed point function. The equilibrium state of the neuronal population model is clearly defined by the input paradigm $\mathbf{F}^{ex}$, the fixed average neuronal population's activities $\mathfrak{F}$ and the learning rule on the excitatory synaptic weights f_ω with its fixed point function f_ω^*:*

$$\mathfrak{E}(\mathbf{F}^{ex}, f_\omega, f_\omega^*) = \{\mathfrak{F}(\mathbf{F}^{ex}, f_\omega), \ \mathfrak{w}_{intra}(\mathfrak{F}(\mathbf{F}^{ex}, f_\omega), f_\omega^*), \ \mathfrak{w}_{inter}(\mathfrak{F}(\mathbf{F}^{ex}, f_\omega), f_\omega^*)\}. \tag{3.15}$$

Proof. The fixed average synaptic weights of the population at equilibrium are given by:

$$\mathfrak{w}_{p_r,p_s} = \mathrm{n}_{\mathcal{P}}^{-2} \sum_{i\in\mathcal{P}_{p_r}} \sum_{j\in\mathcal{P}_{p_s}} f_\omega^*(F_j^*, F_i^*).$$

The assumption that all neurons i of a neuronal population $\mathcal{P}_{p_r}$ approximate the fixed average neuronal population's activity $F_i^* \approx \mathfrak{F}_{p_r}$ (Eq. 3.14) yields an expression of the fixed average synaptic weight independent of the single neuronal dynamics:

$$\mathfrak{w}_{p_r,p_s} = \mathrm{n}_{\mathcal{P}}^{-2} \sum_{i\in\mathcal{P}_{p_r}} \sum_{j\in\mathcal{P}_{p_s}} f_\omega^*(\mathfrak{F}_{p_s}, \mathfrak{F}_{p_r}) = f_\omega^*(\mathfrak{F}_{p_s}, \mathfrak{F}_{p_r}).$$

Thus, given the fixed neuronal population's activities at equilibrium $\mathfrak{F} = \left(\mathfrak{F}_{p_1}, \ldots, \mathfrak{F}_{p_{n_{\mathrm{pop}}}} \mathfrak{F}_{p_{\mathcal{B}}}\right)^{\mathsf{T}}$, the respective fixed neuronal population's intra- and inter-synaptic weights are given by:

$$\mathfrak{w}_{p_r,p_r} = f_{\omega_{i,j}}^*(\mathfrak{F}_{p_r}, \mathfrak{F}_{p_r}) \qquad \text{and} \qquad \mathfrak{w}_{p_r,p_s} = f_{\omega_{i,j}}^*(\mathfrak{F}_{p_s}, \mathfrak{F}_{p_r}).$$

$\square$

3.3.2 Calculation of the Fixed Neuronal Population's Activities.

Up to this point, for any input stimulation $\mathbf{F}^{\mathrm{ex}}$ and synaptic learning rule f_ω, we still have to observe the corresponding average neuronal population's activities $\mathfrak{F} = \left(\mathfrak{F}_{p_1}, \ldots, \mathfrak{F}_{p_{n_{\mathrm{pop}}}}, \mathfrak{F}_{p_{\mathcal{B}}}\right)^{\mathsf{T}}$ at stable state to calculate the equilibrium state of the population model (Eq. 3.15). In the following, we derive a method that combines the external input stimulation $\mathbf{F}^{\mathrm{ex}}$ with the input originating from the background population $p_{\mathcal{B}}$ towards a new system's input paradigm $\mathfrak{I} = \left(\mathfrak{I}_{p_1}, \ldots, \mathfrak{I}_{p_{n_{\mathrm{pop}}}}\right)^{\mathsf{T}}$. On the new $\mathfrak{I}$-input-space, we are then able to numerically determine the respective average neuronal population's activities $\mathfrak{F} = \left(\mathfrak{F}_{p_1}, \ldots, \mathfrak{F}_{p_{n_{\mathrm{pop}}}}\right)^{\mathsf{T}}$ at equilibrium, without simulation the whole system's dynamics over time. For this, we set about to specify $\mathfrak{Q}_{p_r}$ the *average fixed input onto a neuronal population r at equilibrium* in a population formalism.

Corollary 3 *The* average fixed input onto a neuronal population r at equilibrium *is given by:*

$$\mathfrak{Q}_{p_r}(\mathbf{F}^{ex}, \mathfrak{F}, f_\omega^*) := n_\mathcal{P} \omega^{ex} \langle F_{p_r}^{ex} \rangle_t \tag{3.16}$$
$$+ p_{con} \left(n_\mathcal{B}(f_\omega^*(\mathfrak{F}_{p_\mathcal{B}}, \mathfrak{F}_{p_r}) - \theta)\mathfrak{F}_{p_\mathcal{B}} + n_\mathcal{P} \left(\sum_s (f_\omega^*(\mathfrak{F}_{p_s}, \mathfrak{F}_{p_r}) - \theta)\mathfrak{F}_{p_s} \right) \right)$$

only dependent on the environmental input stimulation $\mathbf{F}^{ex}$*, the fixed average neuronal population's activities* $\mathfrak{F}$ *and the fixed point function of the synaptic learning rule* f_ω^*.

Proof. The average input onto a neuronal population r at equilibrium is given by:

$$\mathfrak{Q}_{p_r} := n_\mathcal{P}^{-1} \sum_{i \in \mathcal{P}_{p_r}} \phi_i(\mathbf{F}_i^{ex}, \mathbf{F}^*, \mathbf{W}^*) \overset{(Eq.\ 3.2)}{=} n_\mathcal{P}^{-1} \sum_{i \in \mathcal{P}_{p_r}} \left[\omega^{ex} \left(\sum_{k \in \mathcal{E}_i} F_k^{ex} \right) + \sum_{j \in \mathcal{N}} \left(\omega_{i,j}^* - \theta \right) F_j^* \right].$$

When we pool the presynaptic neuron's affiliation to its respective neuronal population $s \in \{1, \ldots, n_{\text{pop}}\}$, background neurons $\mathcal{B}$ and the population-r-specific external input population $\mathcal{E}_r$, this equation becomes

$$\mathfrak{Q}_{p_r} = n_\mathcal{P}^{-1} \sum_{i \in \mathcal{P}_{p_r}} \left[\omega^{ex} \left(\sum_{k \in \mathcal{E}_r} \langle F_{p_r}^{ex} \rangle_t \right) + \sum_{j \in \mathcal{P}_\mathcal{B}} \left(\omega_{i,j}^* - \theta \right) F_j^* + \sum_s \left(\sum_{j \in \mathcal{P}_{p_s}} \left(\omega_{i,j}^* - \theta \right) F_j^* \right) \right]$$
$$= n_\mathcal{P} \omega^{ex} \langle F_{p_r}^{ex} \rangle_t + p_{con} \left(n_\mathcal{B} \left(\mathfrak{w}_{p_r, p_\mathcal{B}} - \theta \right) \mathfrak{F}_{p_\mathcal{B}} + n_\mathcal{P} \left(\sum_s \left(\mathfrak{w}_{p_r, p_s} - \theta \right) \mathfrak{F}_{p_s} \right) \right)$$

independent of individual neuronal inputs. Furthermore, with Theorem 2 the average input onto neuronal population r at equilibrium is only dependent on the neuronal population's activities and the environmental population specific input stimulation:

$$\mathfrak{Q}_{p_r} = n_\mathcal{P} \omega^{ex} \langle F_{p_r}^{ex} \rangle_t + p_{con} \left(n_\mathcal{B}(f_\omega^*(\mathfrak{F}_{p_\mathcal{B}}, \mathfrak{F}_{p_r}) - \theta)\mathfrak{F}_{p_\mathcal{B}} + n_\mathcal{P} \left(\sum_s (f_\omega^*(\mathfrak{F}_{p_s}, \mathfrak{F}_{p_r}) - \theta)\mathfrak{F}_{p_s} \right) \right).$$

$\square$

As briefly outlined, we now combine the average input onto each neuronal population p_r originating from the background population $p_\mathcal{B}$ at equilibrium with the constant average input originating from the external input layer $\mathcal{E}_r$. In doing so and relative to the size $n_\mathcal{P}$ of the neuronal population-r-specific external input layer $\mathcal{E}_r$, the combined input parameter $\mathfrak{I}_{p_r}$, of the population model is given by:

$$\mathfrak{I}_{p_r} := \omega^{ex} \langle F_{p_r}^{ex} \rangle_t + p_{con} \frac{n_\mathcal{B}}{n_\mathcal{P}} (f_\omega^*(\mathfrak{F}_{p_\mathcal{B}}, \mathfrak{F}_{p_r}) - \theta)\mathfrak{F}_{p_\mathcal{B}}. \tag{3.17}$$
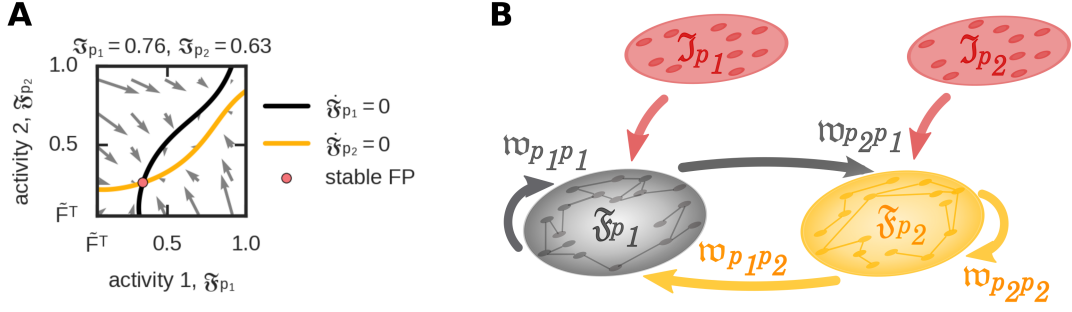
Figure 3.3: Abstraction to a low-dimensional population model of two interacting memories at equilibrium. (A) Calculation of the fixed average neuronal population's activities at equilibrium for the exemplary network simulation (Fig. 3.2 B; for used parameters see see Tab. 3.1). Here, the intersection of the nullclines describes the average neuronal population activities at equilibrium (black: $\dot{\mathfrak{F}}_1 = 0$, yellow: $\dot{\mathfrak{F}}_2 = 0$). Grey arrows indicate the flow field for the activities. **(B)** The abstraction of the neuronal network model yields a low-dimensional population model described by the mean long-term activities ($\mathfrak{F}_p$). These in turn clearly define the corresponding mean long-term synaptic weights ($\mathfrak{w}_{p_r, p_s}$). Here, the input (red) combines inputs from the former described background neurons and external inputs given in the neuronal population model Fig. 3.2 A (adapted from Herpich and Tetzlaff 2018).

With this approach, we combine the input parameter set for the population model formalism from $\mathbf{F}^{ex}$ (Eq. 3.10) with the input originating form the background population $p_{\mathcal{B}}$ at equilibrium towards the new input parameter set at equilibrium $\mathfrak{I} := \left(\mathfrak{I}_{p_1}, \ldots, \mathfrak{I}_{p_{n_{pop}}} \right)^{\mathsf{T}}$. This formalism further reduces the equation on the average fixed input onto a neuronal population r at equilibrium (Eq. 3.16) towards:

$$\mathfrak{Q}_{p_r}(\mathfrak{I}_{p_r}, \mathfrak{F}, f_\omega^*) = \mathrm{n}_{\mathcal{P}} \left[\mathfrak{I}_{p_r} + p_{con} \left(\sum_s \left(f_\omega^*(\mathfrak{F}_{p_s}, \mathfrak{F}_{p_r}) - \theta \right) \mathfrak{F}_{p_s} \right) \right]. \tag{3.18}$$

Furthermore, the equilibrium state of the population model is now clearly defined by the input stimulation $\mathfrak{I}$ and the fixed point function f_ω^* for the excitatory synaptic weights:

Theorem 4 *Let $f_\omega(F_j, F_i, \omega_{i,j})$ be the learning rule of the neuronal system and $f_\omega^*(F_j^*, F_i^*)$ its fixed point function. The equilibrium state of the neuronal population model is clearly defined by the input paradigm $\mathfrak{I}$, the fixed average neuronal population's activities $\mathfrak{F}$ and the fixed point function f_ω^* of the learning rule for the excitatory synaptic weights:*

$$\mathfrak{E}(\mathfrak{I}, f_\omega^*) = \{ \mathfrak{F}(\mathfrak{I}, f_\omega^*), \ \mathfrak{w}_{intra}(\mathfrak{F}(\mathfrak{I}, f_\omega^*), f_\omega^*), \ \mathfrak{w}_{inter}(\mathfrak{F}(\mathfrak{I}, f_\omega^*), f_\omega^*) \}. \tag{3.19}$$

Proof. This plain $\mathfrak{F}$-activity and f_ω^*-learning rule dependent expression of the average input current $\mathfrak{Q}_{p_r}$ onto a neuronal population r at stable state (Eq. 3.18), allows us to numer-
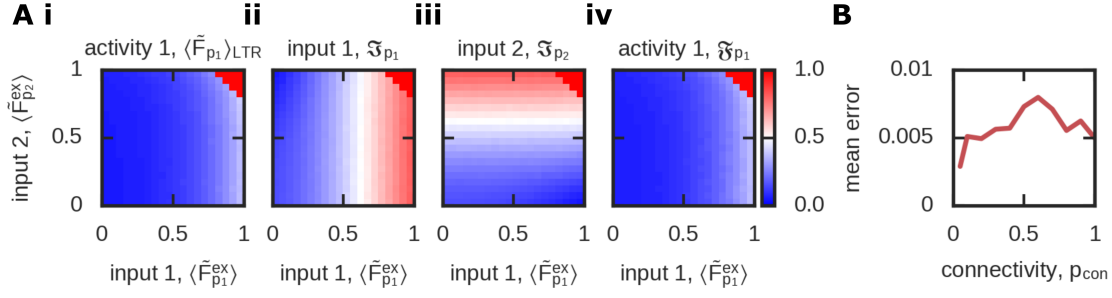
Figure 3.4: Full-network simulation and analysis of two interconnected memories 1 and 2.
(Ai) LTR of the average neuronal population 1 activity within the applied $\langle \tilde{F}^{ex}_{p_1} \rangle - \langle \tilde{F}^{ex}_{p_2} \rangle$−input space. **(Aii,iii)** Conversion of the input space of the neuronal simulation towards the population model at equilibrium. **Aiv** Analytically calculated fixed average activity of population 1 with the given input parameters $\mathfrak{I} = (\mathfrak{I}_{p_1}, \mathfrak{I}_{p_2})^{\mathsf{T}}$. **(B)** Mean error of the numerically simulated and analytically calculated mean activity of population 1 over the whole normalized $\langle \tilde{F}^{ex}_{p_1} \rangle - \langle \tilde{F}^{ex}_{p_2} \rangle$−input space. For used parameters see Tab. 3.1.

ically determine the population's activities at equilibrium $\mathfrak{F}$ for a specific input paradigm $\mathfrak{I}$. By means of the *ordinary differential equation* (ODE) for the neuronal activity (Eq. 3.5), we specify the nullcline for the neuronal population r's average activity:

$$\dot{\mathfrak{F}}_{p_r} := \frac{(F^{max} - \mathfrak{F}_{p_r})\mathfrak{F}_{p_r}}{F^{max}} \left(\log \left(\frac{F^{max}}{\mathfrak{F}_{p_r}} - 1 \right) + \beta(R\mathfrak{Q}_{p_r}(\mathfrak{I}_{p_r}, \mathfrak{F}, f^*_\omega) - \epsilon) \right) \overset{!}{=} 0.$$

And, second, we calculate the intersection of the n_{pop} different nullclines:

$$\mathfrak{F}(\mathfrak{I}, f^*_\omega) = \begin{pmatrix} \mathfrak{F}_{p_1}(\mathfrak{I}_{p_1}, \mathfrak{F}, f^*_\omega) \\ \vdots \\ \mathfrak{F}_{p_{n_{pop}}}(\mathfrak{I}_{p_{n_{pop}}}, \mathfrak{F}, f^*_\omega) \end{pmatrix} \quad \text{with} \quad \dot{\mathfrak{F}}_{p_r}(\mathfrak{I}_{p_r}, \mathfrak{F}, f^*_\omega) \overset{!}{=} 0 \ \text{ for all } r \in \{1, \ldots, n_{pop}\}$$

that determines the system's equilibrium state (see Fig. 3.3 A; red circle indicates the intersection of both nullclines (black and yellow curve)). □

3.3.3 The Population Model Reliably Predicts the Equilibrium State

In the following, we validate our derived population model at equilibrium for the example of two interacting populations (Fig. 3.3 B) to directly calculate the neuronal population's fixed average activities. We compare the results for the LTR from the neuronal network simulation (Fig. 3.1 B, Fig. 3.2 B) for the whole normalized $\langle \tilde{F}^{ex}_{p_1} \rangle - \langle \tilde{F}^{ex}_{p_2} \rangle$−input space (see Appendix) with the predicted fixed population's activities at equilibrium state (Fig. 3.1 C, Fig. 3.3 A).

Thus, we simulate the whole activity- and synaptic weight-dynamics for two interconnected neuronal populations 1 and 2 nested in a fully connected network ($p_{con} = 1$) and extract the LTR of the respective neuronal population's state variables for the whole normalized $\langle \tilde{F}^{ex}_{p_1} \rangle - \langle \tilde{F}^{ex}_{p_2} \rangle$−input space. To make a comparison of the numerical simulation and analytic calculation feasible, we need to calculate $\mathfrak{I}_{p_r}$ for the simulations. For this, we adapt the parameters of Eq. (3.17) for the combined input onto a neuronal population and replace the state variables at fixed point $\mathfrak{F}_{\mathcal{B}}$ and $\mathfrak{w}_{p_r,p_{\mathcal{B}}}$ with the obtained LTR of the numerical simulation $\langle \tilde{F}_{\mathcal{B}} \rangle_{\mathrm{LTR}}$ and $\langle \tilde{\omega}_{p_r,p_{\mathcal{B}}} \rangle_{\mathrm{LTR}}$ at stable state, respectively:

$$\mathfrak{I}_{p_r} := \omega^{ex} \langle F^{ex}_{p_r} \rangle_t + p_{con} \frac{n_{\mathcal{B}}}{n_{\mathcal{P}}} (\langle \omega_{p_r,p_{\mathcal{B}}} \rangle_{\mathrm{LTR}} - \theta) \langle F_{\mathcal{B}} \rangle_{\mathrm{LTR}}. \tag{3.20}$$

Applying this function Eq. (3.20) for the exemplary neuronal network simulation in Fig. 3.2 B where we use a normalized population-specific input of $\tilde{\mathbf{F}}^{ex} = \left(\langle \tilde{F}^{ex}_{p_1} \rangle_t, \langle \tilde{F}^{ex}_{p_2} \rangle_t, \langle \tilde{F}^{ex}_{p_{\mathcal{B}}} \rangle_t \right)^{\mathsf{T}} = (0.9, 0.75, 0.25)^{\mathsf{T}}$ corresponds to the new input stimulation of $\mathfrak{I} = (\mathfrak{I}_{p_1}, \mathfrak{I}_{p_2})^{\mathsf{T}} = (0.76, 0.63)^{\mathsf{T}}$ for the population model at equilibrium, shown in Fig. 3.3 A.

In Fig. 3.4 Ai the LTR for the activity $\langle \tilde{F}_{p_1} \rangle_{\mathrm{LTR}}$ of neuronal population 1 within the whole normalized $\langle \tilde{F}^{ex}_{p_1} \rangle - \langle \tilde{F}^{ex}_{p_2} \rangle$−input space is shown. Using the state variables for the LTR of the neuronal population's activities ($\langle \tilde{F}_{p_1} \rangle_{\mathrm{LTR}}, \langle \tilde{F}_{p_2} \rangle_{\mathrm{LTR}}, \langle \tilde{F}_{p_{\mathcal{B}}} \rangle_{\mathrm{LTR}}$) and synaptic weights ($\langle \tilde{\omega}_{p_{\mathcal{B}} p_r} \rangle_{\mathrm{LTR}}$) and by means of Eq. (3.20), we calculated the converted input parameter $\mathfrak{I}_{p_1}$ and $\mathfrak{I}_{p_2}$ for the population specific input stimulation at equilibrium (Fig. 3.4 Aii,iii), respectively. Using these input parameters $\mathfrak{I}$ and Corollary 4, we analytically calculated the fixed average activity of population 1 at equilibrium state $\mathfrak{F}_{p_1}$ as described (Fig. 3.4 Aiv). To compare the numerical (Fig. 3.4 Ai) with the analytical results for the average activity of neuronal population 1 (Fig. 3.4 Aiv) for the whole normalized $\langle \tilde{F}^{ex}_{p_1} \rangle - \langle \tilde{F}^{ex}_{p_2} \rangle$−input space yields a mean error defined by

$$\text{mean error} = \left\langle \sum_{i \in \langle \tilde{F}^{ex}_{p_1} \rangle} \sum_{j \in \langle \tilde{F}^{ex}_{p_2} \rangle} \frac{1}{2} \left(|\langle \tilde{F}_{p_1} \rangle_{\mathrm{LTR}} - \mathfrak{F}_{p_1}| + |\langle \tilde{F}_{p_2} \rangle_{\mathrm{LTR}} - \mathfrak{F}_{p_2}| \right) \right\rangle, \tag{3.21}$$

of *mean error* $= 0.005 \, \mathrm{F}^{max} = 0.5 \, \mathrm{Hz}$. In Fig. 3.4 B, we did the same analysis as in Fig. 3.4 A for different network connectivities $p_{con} \in [0, 1]$. The mean error globally stays at low values with *mean error* $< 0.01 \, \mathrm{F}^{max} = 1 \, \mathrm{Hz}$. Thus, we conclude that our analytic model for the direct calculation of the system's equilibrium state matches the numerical results very well and, which, supports our population model at equilibrium, derived in this Section.

3.3.4 Dimensionality of the Population Model at Equilibrium

With this description of the fixed average state variables of n_{pop} interacting populations, we reduce the state space of the neuronal network model towards n_{pop}-dimensions, clearly defined by the n_{pop} different fixed population's activities $\mathfrak{F}$. In Fig. 3.3 A, the vector field and the corresponding calculation of the fixed average neuronal population's activities of the corresponding simulation in Fig. 3.2 B is shown. Here, the former population specific input parameter $\tilde{\mathbf{F}}^{\text{ex}} = (0.9, 0.75, 0.25)^{\mathsf{T}}$ is converted towards the respective input parameter $\mathfrak{I} = (0.76, 0.63)^{\mathsf{T}}$ of the population model at equilibrium according to (Eq. 3.20). The stable state of the neuronal population's activities is determined by the intersection of both nullclines $\dot{\mathfrak{F}}_{p_1} = 0$ and $\dot{\mathfrak{F}}_{p_2} = 0$ emphasized by the red circle.

3.4 Definition of Functional Organization of Memories

To classify the system's ability to build up different functionally organized memories based on synaptic plasticity, as they are proposed to appear for the formation of schemata (Sec. 2.1.5, Fig. 2.4, Def. 1), we consider the *long-term representation* (LTR) of the neuronal population model representing the respective stimuli. For learning rules f_ω, that lead the system into an equilibrium state, the LTR of the environmental stimuli are given by the state variables $\mathfrak{F}$, $\mathbf{w}_{\text{intra}}$ and $\mathbf{w}_{\text{inter}}$ of this equilibrium state (see previous Section Sec. 3.3.2). The smallest functional entity of a set $\mathcal{P} = \{1, \ldots, n_{\text{pop}}\}$ of n_{pop} interacting memories in the cognitive structure of schemata is the *functional organization* (FO) of two such interconnected memories r and s. Thus, in the following we focus on the constraints for *memory representation* (MR) and different forms of FO of two interconnected neuronal populations r and s. Note, for general validity, we define MR and the FO of neuronal populations in the general notation of the LTR resulting from the numerical simulation ($\langle \circ \rangle_{\text{LTR}}$) that also holds for the analytic population model at equilibrium.

3.4.1 Memory Representation

To determine whether a neuronal population is a proper memory representation of a certain input, we introduce a binary classification of the average recurrent intra-population excitatory synaptic weight in relation to the constant average inhibition inherent of the neuronal system in the LTR. Therefore, we classify whether the mean intra-population synaptic weight of neuronal population r $\langle \omega_{p_r, p_r} \rangle_{\text{LTR}}$ is weaker or stronger than the average inhibitory synaptic weight θ. As long as the recurrent excitatory synaptic weight of a neuronal population r is weaker than the constant mean inhibitory synaptic weight θ

A Memory Representation (MR)   **B** Functional Organization (FO)
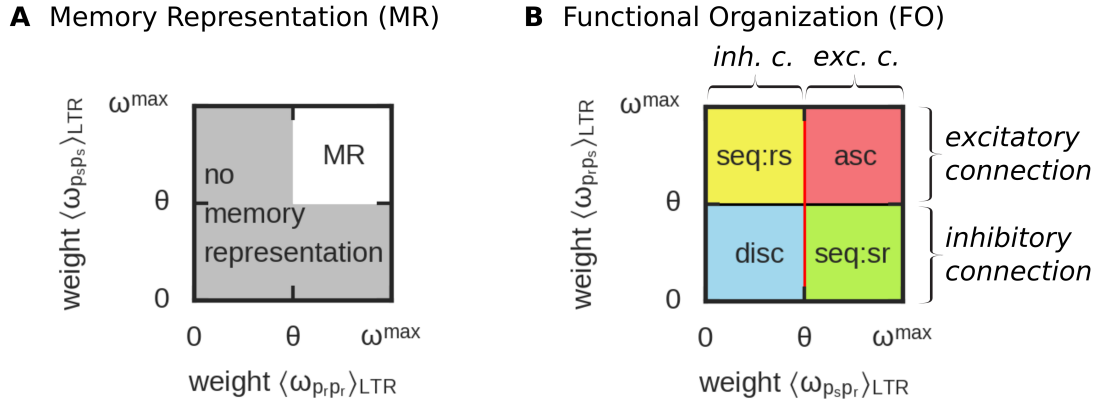
Figure 3.5: Definition of memory representation and functional organizations of two interconnected neuronal populations on the respective synaptic weight space. (A) Definition of the MR for two interconnected neuronal populations dependent on the binary relation of the average recurrent synaptic weights ($\langle \omega_{p_r,p_r} \rangle_{\text{LTR}}$) to the inhibitory synaptic weight strength (θ) inherent to the network. **(B)** Definition of different forms of FOs, such as association (asc), sequence (seq) and discrimination (disc), based on the average excitatory synaptic weights ($\langle \omega_{p_s,p_r} \rangle_{\text{LTR}}$) connecting neuronal population r towards neuronal population s, again, dependent on their binary relation to the constant inhibition of the network (θ). Details see main text (adapted from Herpich and Tetzlaff 2018).

(Fig. 3.5 A, gray; Tab. 3.2 first column above horizontal line), an external input provided to the population will cause the activity of this particular population to decrease. Thus, in those cases we refer this neuronal population to be in a *no-memory state* (nm), that consequently, does not serve as MR. By contrast, a neuronal population is said to encode a memory trace, if the recurrent excitatory synaptic weight gets stronger than the constant inhibitory synaptic weight strength (Fig. 3.5 A, white space; Tab. 3.2 first column below horizontal line). Here, when recalling the memory by an external input cue, the neuronal population will respond with an increased activity level and, therefore, serves as a MR.

Definition 2 *Let the neuronal population r encode an environmental stimulus. This neuronal population is said to be a memory representation (MR), if the LTR of its average recurrent excitatory synaptic weight ($\langle \omega_{p_r,p_r} \rangle_{\text{LTR}}$) is above the neural system's inherent level of inhibition (θ):*

$$\langle \omega_{p_r,p_r} \rangle_{\text{LTR}} > \theta. \tag{3.22}$$

The condition implied by Def. 2 corresponds to the white synaptic weight space in Fig. 3.5 A for two interconnected populations r, s.

3.4.2 Functional Organization of Memories

In the following, we define the long-term relation $\langle \omega_{p_s,p_r} \rangle_{\text{LTR}}$ connecting neuronal population r with neuronal population s. We do this in a similar manner as for the MR based on the relation of excitation and inhibition between the corresponding neuronal populations (Fig. 3.5 B; Tab. 3.2 column 2 & 3). If the LTR of the average excitatory synaptic weight $\langle \omega_{p_s,p_r} \rangle_{\text{LTR}}$ is larger than the average inhibitory synaptic strength θ, an increased level of activity in the former population r triggers an increased activation in the latter population s. In this case, we refer to $\langle \omega_{p_s,p_r} \rangle_{\text{LTR}}$ being an *excitatory connection*. In the opposite case ($\langle \omega_{p_s,p_r} \rangle_{\text{LTR}} < \theta$), we refer to an *inhibitory connection*. Note that the LTRs of the inter memory connections can be different for both directions $\langle \omega_{p_s,p_r} \rangle_{\text{LTR}}$ and $\langle \omega_{p_r,p_s} \rangle_{\text{LTR}}$. Thus, there are four different combinations for the binary states of the two connections possible, defining the range of four different forms of FO. These FOs are *discrimination* (disc, blue), *sequence* (seq:*sr* green and seq:*rs* yellow) and *association* (asc, red). The definition of sequences introduces a *sequential order* of both two interconnected memories regarding the direction of the excitatory conduction of an external input stimulus.

Definition 3 *Let two neuronal populations r and s encode memory items ($\langle \omega_{p_r,p_s} \rangle_{LTR} > \theta$, $\langle \omega_{p_s,p_r} \rangle_{LTR} > \theta$). The different forms of FO of two interconnected memory items are defined based on the binary conditions of either excitatory or inhibitory inter-memory connection ($\langle \omega_{p_r,p_s} \rangle_{LTR} \gtrless \theta$, $\langle \omega_{p_s,p_r} \rangle_{LTR} \gtrless \theta$) of the two memories:*

disc *Both connections $\langle \omega_{p_r,p_s} \rangle_{LTR}$ and $\langle \omega_{p_s,p_r} \rangle_{LTR}$ are inhibitory (Fig. 3.5 B, blue synaptic weight space).*

seq *One connection is inhibitory while the reverse connection is excitatory. Regarding the sequential order, we define a sequence from memory s to memory r (seq:rs) as the sequence with an excitatory conduction of a recall stimulus from memory s towards memory r (Fig. 3.5 B, green and yellow synaptic weight space).*

asc *Both connections $\langle \omega_{p_r,p_s} \rangle_{LTR}$ and $\langle \omega_{p_s,p_r} \rangle_{LTR}$ are excitatory (Fig. 3.5 B, red synaptic weight space).*

3.4.3 Comparison of the Simulated with the Analytically Predicted Functional Organization of Memories

From the full-network simulation of two interconnected neuronal populations 1 and 2 (Fig. 3.4 A), we extract the corresponding LTRs of the intra- and inter-population synaptic weights in Fig. 3.6 A,B. In panel Aiii and Panel Biii, we plotted the resulting classification

Table 3.2: Synaptic weight-dependent definition of memory representation (MR) and different forms of functional organization of two interconnected neuronal populations $r, s \in \{1, 2\}$**.** The different FOs are defined based on the LTR of the average intra population's excitatory synaptic weights $\langle \omega_{p_r p_r} \rangle_{\mathrm{LTR}}$ and inter population's synaptic weight $\langle \omega_{p_r p_s} \rangle_{\mathrm{LTR}}$ in relation to the average inhibitory synaptic weights θ (adapted from Herpich and Tetzlaff 2018).

MR	Functional Organization				
$\langle \omega_{p_r,p_r} \rangle_{\mathrm{LTR}}$	$\langle \omega_{p_r,p_s} \rangle_{\mathrm{LTR}}$	$\langle \omega_{p_s,p_r} \rangle_{\mathrm{LTR}}$	definition	abbreviation	color
$< \theta$	-	-	none	nm	gray
$> \theta$	$< \theta$	$< \theta$	discrimination	disc	blue
	$< \theta$	$> \theta$	sequence sr	seq:sr	green
	$> \theta$	$< \theta$	sequence rs	seq:rs	yellow
	$> \theta$	$> \theta$	association	asc	red

in accordance with the definitions on MR and FO. These results are further mapped onto the applied $\langle \tilde{F}^{\mathrm{ex}}_{p_1} \rangle_{\mathrm{LTR}} - \langle \tilde{F}^{\mathrm{ex}}_{p_2} \rangle_{\mathrm{LTR}} -$ input space (Fig. 3.6 Ci). For comparison with the analytically predicted FOs (Fig. 3.6 D), we map this space to the $\mathfrak{I}_{p_1} - \mathfrak{I}_{p_2} -$input space of the population model at equilibrium (Fig. 3.6 Cii) by means of Eq. (3.20). When we compare the FO resulting of the numerical simulation (Fig. 3.6 Cii) and analytically predicted forms of FOs (Fig. 3.6 D), we find a good match which again highlight the applicability of our analytic approach to calculate the system's state variables at equilibrium in a population-model formalism.

3.5 Conclusion

In this Chapter, we have derived a mathematical framework to investigate the ability of adaptive neural networks to form different *functional organization*s (FOs) of interconnected *memory representation*s (MRs) in a dynamic and input-dependent manner. Those structures are likely to occur in schemata, the networks of knowledge in the human brain responsible for basic cognitive functions. Multiple studies that precede this thesis only consider the FO of sequences, such as temporal or motor sequences or dynamic phase transitions (Griniasty et al. 1993; Abbott and Blum 1996; Leibold and Kempter 2006; Herrera-Aguilar et al. 2012; Tully et al. 2016; Chenkov et al. 2017). In contrast, in this thesis, we consider all possible combinations of two interconnected MRs, defined as *association* (asc),
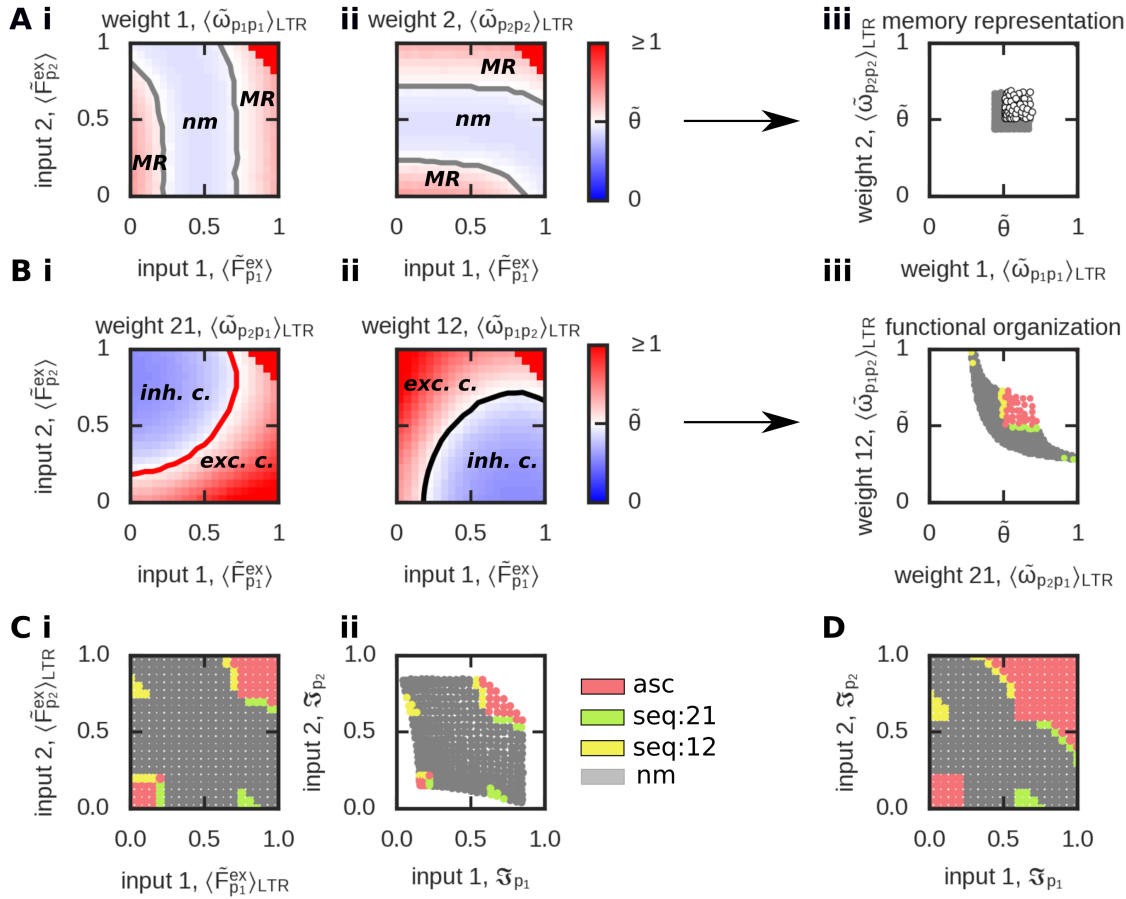
Figure 3.6: Complete analysis for the population model nested in a neuronal network (A-C) and analyzed by the mathematical framework at equilibrium (D). (A) LTRs for the intra-population synaptic weights within the applied $\langle F^{ex}_{p_1} \rangle_{LTR} - \langle F^{ex}_{p_2} \rangle_{LTR}$−input space classified accoring to the contraints on MRs in the synaptic weight space. **(B)** Same as in (A) for the synaptic weight connecting both neuronal population with a classification on the constraints of FO in (Biii). **(C)** Mapping of the result in (Aiii, Biii) onto the the applied $\langle F^{ex}_{p_1} \rangle_{LTR} - \langle F^{ex}_{p_2} \rangle_{LTR}$−input space (Ci) with transfer to the input space at equilibrium (Cii) by means of (Eq. 3.20). **(D)** Analytically predicted MRs and FOs resulting of the population model at equilibrium. For used parameters see Tab. 3.1 (adapted from Herpich and Tetzlaff 2018).

sequence (seq) and *discrimination* (disc). Those structures represent the primary building blocks of interconnected memories present in schemata and we summarize these different forms of relations by the term FOs.

An analysis of those FOs of interconnected MRs embedded in a recurrent neuronal network is made feasible by introducing a population formalism. This approach enables us to reduce the high-dimensional neuronal network problem to a low dimensional problem using average synaptic weights and activities of the neuronal populations involved (Fig. 3.2). With this level of description, we are able to determine analytically whether a

neuronal population will either excite or inhibit an interconnected neuronal population at a given input. Since two interconnected memories can either excite or inhibit each other reciprocally or, alternatively, form an excitatory connection in one direction while the reverse connection is an inhibitory connection, this classification defines the different form of FOs of interconnected MRs. Finally, by considering the long-term equilibrium states of the excitatory synaptic weights of the population model, we further reduce the system in its complexity. On this reduced model of interconnected neuronal populations at equilibrium, we are able to map the resulting *long-term representation*s (LTRs) onto the two-dimensional activity-space of two interconnected neural populations (Fig. 3.3). This is a sufficient description to solve the complex problem of input-dependent and plastic memory interactions in a recurrent neuronal network (Fig. 3.4, Fig. 3.6).

Next, we will use this generic analytic framework to investigate the effect of diverse activity-dependent synaptic learning rules on the formation and interaction between MRs in the next Chapter.

4

Implication of Memory Interactions on Synaptic Plasticity

In this Section we analyze the ability of different synaptic plasticity mechanisms to reliably describe the formation and storage of memories and their dynamical organization into a web of knowledge, which is essential for the generation of complex behaviors. The primary building blocks of such a web of knowledge are the different forms of *functional organization* (FO) of two interconnected memories (Sec. 3.4, p. 45). Therefore, here we consider such interaction of two neuronal populations $r = 1$ and $s = 2$ in the derived population model at equilibrium (Sec. 3.3, p. 37, Fig. 3.3). In the previous Chapter, we referred to each individual populations r and s of the neuronal network with the specific notation of p_r and p_s, respectively. In this Chapter, for better readability, we drop the p and directly refer to each individual population by $r, s \in \{1, 2\}$.

4.1 Correlation-based Plasticity

The underlying neuronal correlate of learning for all plastic excitatory synapses ($\omega_{i,j}$) within the recurrent layer $\mathcal{N}$ of rate coded neurons is proposed to be Hebbian like (Hebb 1949), as introduced in Sec. 2.2.2:

$$f_\omega(F_j, F_i, \omega_{i,j}) = \tau_\omega \dot{\omega}_{i,j} = C(F_i, F_j), \tag{4.1}$$

with neural activities of the pre- (F_j) and postsynaptic (F_i) neuron and τ_ω being the time scale of synaptic plasticity. To enable an analysis of different learning rules on synaptic plasticity, we make several assumptions on the specific formulation of Hebbian like plasticity at stable state ($\circ^*$):

Assumption 5 *The correlation-based plasticity mechanism C at equilibrium state is assumed to be:*

- *positive (abbr. C:p); for all $x^*, y^* \in \mathbb{R}_+$:*

$$C(x^*, y^*) \in \mathbb{R}_+, \tag{4.2}$$

- *monotonous (abbr. C:m); for all $x_1^* > x_2^* \in \mathbb{R}_+$ and $y_1^* > y_2^* \in \mathbb{R}_+$:*

$$C(x_1^*, y^*) > C(x_2^*, y^*) \quad and \quad C(x^*, y_1^*) > C(x^*, y_2^*), \tag{4.3}$$

- *symmetric (abbr. C:s); for all $x^*, y^* \in \mathbb{R}_+$:*

$$C(x^*, y^*) = C(y^*, x^*), \tag{4.4}$$

- *independent (abbr. C:i); for all $x^*, y^* \in \mathbb{R}_+$:*

$$C(x^*, y^*) = \hat{C}(x^*)\hat{C}(y^*) \tag{4.5}$$

with a function $\hat{C} : \mathbb{R}_+ \to \mathbb{R}_+$.

Example 1 *The prototype of the Hebbian learning mechanism:*

$$C : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+, \ (F_j, F_i) \mapsto F_j F_i \tag{4.6}$$

describes the AND condition for joint activity of the pre- and postsynaptic neurons (Dayan and Abbott 2001; Gerstner and Kistler 2002; Choe 2015). Furthermore, this prototype fulfills each single assumption (C:p, C:m, C:s, C:i) introduced above.

The mathematical formulation of plain correlation-based Hebbian learning rules (generic formulation (Eq. 4.1), specific example (Eq. 4.6)) describe a positive feedback loop and, thus, leads to diverging synaptic weight dynamics. Hence, plain Hebbian learning rules are insufficient to describe the stable *long-term representation* (LTR) of environmental input stimulations (Dayan and Abbott 2001; Gerstner and Kistler 2002; Choe 2015).

4.2 Correlation-based Plasticity and Homeostasis

Previous studies counteract on the divergence of synaptic weights by introducing a stabilizing homeostatic mechanism H in addition to the Hebbian learning mechanism C. Consequently, these rules are suitable models for synaptic plasticity in a biological realistic manner (Bienenstock et al. 1982; Oja 1982; Dayan and Abbott 2001; Gerstner and Kistler 2002; Tetzlaff et al. 2013; Tetzlaff et al. 2015). Thereby, a subset of these different synaptic learning rules introduce a direct weight-dependency of the homeostatic dynamics that is driven by the postsynaptic neuronal activity (Oja 1982; Tetzlaff et al. 2013).

Definition 4 *The combination of the Hebbian learning C rule with a postsynaptic activity-dependent homeostatic mechanism H proportional to the synaptic weight:*

$$f_\omega(F_j, F_i, \omega_{i,j}) = \tau_\omega \dot{\omega}_{i,j} = C(F_j, F_i) - \gamma H(F_i)\omega_{i,j}^n, \ n \in \mathbb{N}, \tag{4.7}$$

with γ being the time scale of the homeostatic mechanism and $n \in \mathbb{N}$ the exponential dependency on the synaptic weight. This is a learning rule leading to stable synaptic weight dynamics (Tetzlaff et al. 2013; Tetzlaff et al. 2015).

For the analysis on the reduced population model at equilibrium, we make one assumption on the homeostatic plasticity mechanisms at stable state.

Assumption 6 *The postsynaptic activity-dependent synaptic scaling mechanism H is assumed to be monotonous at equilibrium (H:m); for all $x^* > y^* \in \mathbb{R}_+$ it holds*

$$H(x^*) > H(y^*). \tag{4.8}$$

Such learning rules on correlation-based and homeostatic synaptic plasticity are biologically motivated. Some examples based on a Hebbian learning mechanism for the correlation-based plasticity mechanism as stated in Eq. (4.6), are given in the following examples.

Example 2 *Oja's rule (Oja 1982), defined by a quadratic postsynaptic activity-dependent homeostatic mechanism:*

$$H : \mathbb{R}_+ \to \mathbb{R}_+, \ F_i \mapsto F_i^2,$$

depends linearly on the synaptic weight ($n = 1$) yielding the specific learning rule:

$$\tau_\omega \dot{\omega}_{i,j} = F_j F_i - \gamma F_i^2 \omega_{i,j}. \tag{4.9}$$

Example 3 *The learning rule on Synaptic Plasticity and Synaptic Scaling (SPaSS) (Tetzlaff et al. 2011; Tetzlaff et al. 2012) with a biologically motivated homeostatic mechanism, called synaptic scaling (Turrigiano et al. 1998; Turrigiano and Nelson 2004), depends on the postsynaptic activity in relation to a specific* target firing rate $F^T \in \mathbb{R}_+$:

$$H : \mathbb{R}_+ \to \mathbb{R}, \ F_i \mapsto F_i - F^T.$$

Moreover, the synaptic scaling mechanism is dependent on the synaptic weight in second power ($n = 2$). Hence, the synaptic learning rule is given by the following mathematical formulation:

$$\tau_\omega \dot{\omega}_{i,j} = F_j F_i - \gamma \left(F_i - F^T \right) \omega_{i,j}^2. \tag{4.10}$$

Both examples (Example 2 & Example 3) hold Assumption 6 at equilibrium. Please note that previous studies indicate that the SPaSS-learning rule with its combination of the specific plasticity mechanisms of Hebbian learning and synaptic scaling normalized to the synaptic weight yield the reliable formation of individual *memory representation*s (MRs) (Tetzlaff et al. 2013; Tetzlaff et al. 2015). Thus, the SPaSS-learning rule seems to be a suitable learning rule to analyze its dynamics in the interaction of distinct MRs and, consequently, its abilities to form different forms of *functional organization*s (FOs) of interconnected memories. For an analysis of such learning rules f_ω on the introduced *population model at equilibrium* (Sec. 3.3, p. 37), we have to specify the corresponding fixed point function f_ω^*:

Corollary 5 *Let f_ω be a synaptic learning rule on correlation-based plasticity in combination with a synaptic-weight dependent homeostatic mechanism. The respective fixed point function is given by:*

$$f_\omega^*(F_j^*, F_i^*) = \omega_{i,j}^*(F_j^*, F_i^*) = \sqrt[n]{\frac{C(F_j^*, F_i^*)}{\gamma H(F_i^*)}}. \tag{4.11}$$

Proof. Due to the weight dependency of the homeostatic mechanism, the fixed point function f_ω^* of the applied learning rule (Eq. 4.7) can be calculated by setting the *ordinary differential equation* (ODE) of the learning rule equal to zero and solve it to $\omega_{i,j}$. $\qquad\square$

The existence of such a fixed point function f_ω^* for the plastic synaptic weights in combination with Theorem 4 (p. 42) allows us to express the system's equilibrium state $\mathfrak{E}(\mathfrak{I}, f_\omega^*)$ for a given input paradigm $\mathfrak{I} = (\mathfrak{I}_1, \mathfrak{I}_2)^\mathsf{T}$ only dependent on the neuronal population's fixed activities $\mathfrak{F} = (\mathfrak{F}_1, \mathfrak{F}_2)^\mathsf{T}$. By this, we analyze the conditions for MR and the different forms of FO on the neuronal population's $\mathfrak{F}_1 - \mathfrak{F}_2-$activity space at fixed point.

4.2.1 Memory Representation and Functional Organization in Activity-Space

In the following, we analyze learning rules on correlation-based synaptic plasticity in combination with a homeostatic mechanism in regard to their ability to build up *memory representation* (MR) of given stimuli $\mathfrak{I}$ and, simultaneously, form the different types of *functional organization* (FO) of two interconnected neuronal populations. As we have seen, synaptic learning rules based on correlation-based plasticity in combination with a weight dependent homeostatic mechanism exhibit a fixed point function f_ω^* (Eq. 4.11) on the synaptic weights. Thus, for our analysis, we consider the system's equilibrium state $\mathfrak{E}(\mathfrak{I}, f_\omega^*)$ indicating the *long-term representation* (LTR) for a given input.

Necessary Conditions on the Different Forms of Functional Organization

Using the fixed point function f_ω^* (Eq. 4.11), we can express the conditions for MR for each neuronal population $r \in \{1, 2\}$ and the FO of two interconnected memories by the different activity-dependent plasticity mechanisms C and H. The condition on MR (Eq. 3.22, Fig. 3.5 A, p. 46) yields for each neuronal population r the following expression on the involved plasticity mechanisms:

$$\mathfrak{w}_{r,r} \overset{\text{MR}}{>} \theta \qquad \Leftrightarrow \qquad \sqrt[n]{\frac{C(\mathfrak{F}_r, \mathfrak{F}_r)}{\gamma H(\mathfrak{F}_r)}} \overset{\text{MR}}{>} \theta. \tag{4.12}$$

In the same way, the long-term average synaptic weights connecting neuronal population s with neuronal population r can be expressed by their activity-dependent plasticity mechanisms C and H. Here, we consider the binary states of such connections being an *excitatory* ($>$) or *inhibitory connection* ($<$; Fig. 3.5 B, p. 46):

$$\mathfrak{w}_{r,s} \overset{\text{exc}}{\underset{\text{inh}}{\gtrless}} \theta \qquad \Leftrightarrow \qquad \sqrt[n]{\frac{C(\mathfrak{F}_s, \mathfrak{F}_r)}{\gamma H(\mathfrak{F}_r)}} \overset{\text{exc}}{\underset{\text{inh}}{\gtrless}} \theta. \tag{4.13}$$

Theorem 6 *Let θ be the constant inhibitory synaptic weight level within the neuronal network. The constraints on the MR for the interconnected neuronal population 1 and 2 and their different forms of functional organizations (FOs) (FO $\in \{asc, seq{:}21, seq{:}12, disc\}$; Tab. 3.2) can be combined into two ordered sets on the plasticity mechanisms C and H of the applied learning rule:*

$$\left[\gamma H(\mathfrak{F}_1)\theta^n \begin{cases} \overset{MR}{<} C(\mathfrak{F}_1, \mathfrak{F}_1) \\ \underset{exc}{\overset{inh}{\gtrless}} C(\mathfrak{F}_2, \mathfrak{F}_1) \end{cases} ; \; \gamma H(\mathfrak{F}_2)\theta^n \begin{cases} \overset{MR}{<} C(\mathfrak{F}_2, \mathfrak{F}_2) \\ \underset{exc}{\overset{inh}{\gtrless}} C(\mathfrak{F}_1, \mathfrak{F}_2) \end{cases} \right] \mapsto FO. \tag{4.14}$$

Proof. The constraints on the memory representation (MR1 & MR2) are given by:

$$\mathfrak{w}_{11} = \sqrt[n]{\frac{C(\mathfrak{F}_1, \mathfrak{F}_1)}{\gamma H(\mathfrak{F}_1)}} > \theta \qquad \Leftrightarrow \qquad C(\mathfrak{F}_1, \mathfrak{F}_1) > \gamma H(\mathfrak{F}_1)\theta^n \qquad \text{(MR1)}$$

and

$$\mathfrak{w}_{22} = \sqrt[n]{\frac{C(\mathfrak{F}_2, \mathfrak{F}_2)}{\gamma H(\mathfrak{F}_2)}} > \theta \qquad \Leftrightarrow \qquad C(\mathfrak{F}_2, \mathfrak{F}_2) > \gamma H(\mathfrak{F}_2)\theta^n. \qquad \text{(MR2)}$$

Whereas the constraints on the excitatory ($>$) or inhibitory ($<$) connections in between both memories (R12 & R21) are given by:

$$\mathfrak{w}_{12} = \sqrt[n]{\frac{C(\mathfrak{F}_2, \mathfrak{F}_1)}{\gamma H(\mathfrak{F}_1)}} \overset{\text{exc}}{\underset{\text{inh}}{\gtrless}} \theta \qquad \Leftrightarrow \qquad C(\mathfrak{F}_2, \mathfrak{F}_1) \overset{\text{exc}}{\underset{\text{inh}}{\gtrless}} \gamma H(\mathfrak{F}_1)\theta^n \qquad \text{(R12)}$$

and

$$\mathfrak{w}_{21} = \sqrt[n]{\frac{C(\mathfrak{F}_1, \mathfrak{F}_2)}{\gamma H(\mathfrak{F}_2)}} \overset{\text{exc}}{\underset{\text{inh}}{\gtrless}} \theta \qquad \Leftrightarrow \qquad C(\mathfrak{F}_1, \mathfrak{F}_2) \overset{\text{exc}}{\underset{\text{inh}}{\gtrless}} \gamma H(\mathfrak{F}_2)\theta^n. \qquad \text{(R21)}$$

Due to the postsynaptic activity-dependent scaling term H, the conditions (MR1) and (R12) yield an estimation on the homeostatic mechanism $\gamma H(\mathfrak{F}_1)\theta^n$. This estimation comprises the first ordered set of the plasticity mechanisms in (Eq. 4.14). In the same way, the conditions (MR2) and (R21) yield the second ordered set of the plasticity mechanisms. $\square$

Sufficient Conditions on the Different Forms of Functional Organization

The conditions for the formation of an *association* (asc) of two interconnected populations result from the condition on the MR of the population with the lower level of activity and the condition for an excitatory relation in the direction from the population with the lower activity level towards the population with the higher activity level:

Corollary 7 *Let $\mathfrak{F}_{min} := \min(\mathfrak{F}_1, \mathfrak{F}_2)$ be the lower activity level and $\mathfrak{F}_{max} := \max(\mathfrak{F}_1, \mathfrak{F}_2)$ be the higher activity level of two interconnected neuronal populations 1 and 2. An association (asc) of both neuronal populations is defined by the sufficient condition:*

$$\left[\gamma H(\mathfrak{F}_{min})\theta^n \overset{MR}{<} C(\mathfrak{F}_{min}, \mathfrak{F}_{min}); \ \gamma H(\mathfrak{F}_{max})\theta^n \overset{exc}{<} C(\mathfrak{F}_{min}, \mathfrak{F}_{max}) \right] \mapsto asc. \qquad (4.15)$$

Proof. The specific conditions for an asc are given by the bidirectional excitatory connections of both memories:

$$\left[\gamma H(\mathfrak{F}_1)\theta^n \begin{cases} \overset{MR}{<} C(\mathfrak{F}_1, \mathfrak{F}_1) \\ \overset{exc}{<} C(\mathfrak{F}_2, \mathfrak{F}_1) \end{cases} ; \ \gamma H(\mathfrak{F}_2)\theta^n \begin{cases} \overset{MR}{<} C(\mathfrak{F}_2, \mathfrak{F}_2) \\ \overset{exc}{<} C(\mathfrak{F}_1, \mathfrak{F}_2) \end{cases} \right] \mapsto \text{asc.}$$

Case 1: $\mathfrak{F}_1 < \mathfrak{F}_2$. According to the premises on the MR of the population 1 with the lower level of activity, $\gamma H(\mathfrak{F}_1)\theta^n \overset{\text{MR}}{<} C(\mathfrak{F}_1, \mathfrak{F}_1)$. Due to the order-preserving property of the Hebbian plasticity mechanism (C:m, Eq. 4.3), this MR implies the condition on the excitatory relation for the connection from memory 2 to memory 1 (R12): $\gamma H(\mathfrak{F}_1)\theta^n \overset{\text{MR}}{<} C(\mathfrak{F}_1, \mathfrak{F}_1) \overset{\text{C:m}}{<} C(\mathfrak{F}_2, \mathfrak{F}_1)$. This implication reduces both estimations on the homeostatic mechanism dependent on the fixed average activity of neuronal population 1 to one estimation:

$$\left[\gamma H(\mathfrak{F}_1)\theta^n \overset{\text{MR}}{<} C(\mathfrak{F}_1, \mathfrak{F}_1) \; ; \; \gamma H(\mathfrak{F}_2)\theta^n \begin{cases} \overset{\text{MR}}{<} C(\mathfrak{F}_2, \mathfrak{F}_2) \\ \overset{\text{exc}}{<} C(\mathfrak{F}_1, \mathfrak{F}_2) \end{cases} \right] \mapsto \text{asc.}$$

Equivalent, the second premises on the excitatory relation in the direction from the population 1 with the lower activity level towards the population 2 with the higher activity level (R21) $\gamma H(\mathfrak{F}_2)\theta^n \overset{\text{exc}}{<} C(\mathfrak{F}_1, \mathfrak{F}_2)$ implies the condition on the MR of population 2 with the higher activity level: $\gamma H(\mathfrak{F}_2)\theta^n \overset{\text{exc}}{<} C(\mathfrak{F}_1, \mathfrak{F}_2) \overset{\text{C:m}}{<} C(\mathfrak{F}_2, \mathfrak{F}_2)$. Hence, the second ordered set of plasticity mechanisms defining the asc is reduced towards one estimation:

$$\left[\gamma H(\mathfrak{F}_1)\theta^n \overset{\text{MR}}{<} C(\mathfrak{F}_1, \mathfrak{F}_1) \; ; \; \gamma H(\mathfrak{F}_2)\theta^n \overset{\text{exc}}{<} C(\mathfrak{F}_1, \mathfrak{F}_2) \right] \mapsto \text{asc.}$$

Case 2: $\mathfrak{F}_1 > \mathfrak{F}_2$. Equivalent to case 1, we get

$$\left[\gamma H(\mathfrak{F}_1)\theta^n \overset{\text{exc}}{<} C(\mathfrak{F}_2, \mathfrak{F}_1) \; ; \; \gamma H(\mathfrak{F}_2)\theta^n \overset{\text{MR}}{<} C(\mathfrak{F}_2, \mathfrak{F}_2) \right] \mapsto \text{asc.}$$

Case 3: $\mathfrak{F}_1 = \mathfrak{F}_2$. This special case reduces the problem to consider the condition for MR for both populations

$$\left[\gamma H(\mathfrak{F}_1)\theta^n \overset{\text{MR}}{<} C(\mathfrak{F}_1, \mathfrak{F}_1) \; ; \; \gamma H(\mathfrak{F}_2)\theta^n \overset{\text{MR}}{<} C(\mathfrak{F}_2, \mathfrak{F}_2) \right] \mapsto \text{asc,}$$

as these conditions imply an excitatory connection in both directions (R21 and R12), respectively. This expression is in accordance with (Eq. 4.15). $\qquad\square$

Corollary 8 *The sufficient conditions for the formation of a sequence from memory s to memory r (seq:rs) are given by:*

$$\left[\gamma H(\mathfrak{F}_r)\theta^n \begin{cases} \overset{MR}{<} C(\mathfrak{F}_r, \mathfrak{F}_r) \\ \overset{exc}{<} C(\mathfrak{F}_s, \mathfrak{F}_r) \overset{inh}{<} \gamma H(\mathfrak{F}_s)\theta^n \overset{MR}{<} C(\mathfrak{F}_s, \mathfrak{F}_s) \end{cases} \right] \mapsto seq{:}rs \qquad (4.16)$$

Proof. The two ordered sets on the plasticity mechanisms for the conditions of an seq:rs are given by

$$\left[\gamma H(\mathfrak{F}_r)\theta^n \begin{cases} \overset{\text{MR}}{<} C(\mathfrak{F}_r, \mathfrak{F}_r) \\ \overset{\text{exc}}{<} C(\mathfrak{F}_s, \mathfrak{F}_r) \end{cases} ; \ C(\mathfrak{F}_r, \mathfrak{F}_s) \overset{\text{inh}}{<} \gamma H(\mathfrak{F}_s)\theta^n \overset{\text{MR}}{<} C(\mathfrak{F}_s, \mathfrak{F}_s) \right] \mapsto \text{seq:}rs.$$

Due to the symmetric characteristic of the Hebbian learning mechanism, it holds $C(\mathfrak{F}_s, \mathfrak{F}_r) \overset{\text{C:s}}{=} C(\mathfrak{F}_r, \mathfrak{F}_s)$. Hence, both ordered sets of plasticity mechanisms can be combined towards one ordered set of the following form:

$$\left[\gamma H(\mathfrak{F}_r)\theta^n \begin{cases} \overset{\text{MR}}{<} C(\mathfrak{F}_r, \mathfrak{F}_r) \\ \overset{\text{exc}}{<} C(\mathfrak{F}_s, \mathfrak{F}_r) \overset{\text{C:s}}{=} C(\mathfrak{F}_r, \mathfrak{F}_s) \overset{\text{inh}}{<} \gamma H(\mathfrak{F}_s)\theta^n \overset{\text{MR}}{<} C(\mathfrak{F}_s, \mathfrak{F}_s) \end{cases} \right] \mapsto \text{seq:}rs$$

$\square$

Corollary 9 *All learning rules based on correlation-based Hebbian plasticity in combination with postsynaptic activity-dependent homeostatic mechanism can not describe the functional organization of discrimination of two interconnected memories.*

Proof. The two ordered sets of the plasticity mechanisms describing the *discrimination* (disc) of two interconnected memories are defined by the bidirectional inhibitory connections of both memories:

$$\left[C(\mathfrak{F}_1, \mathfrak{F}_1) \overset{\text{MR}}{>} \gamma H(\mathfrak{F}_1)\theta^n \overset{\text{inh}}{>} C(\mathfrak{F}_2, \mathfrak{F}_1) ; \ C(\mathfrak{F}_2, \mathfrak{F}_2) \overset{\text{MR}}{>} \gamma H(\mathfrak{F}_2)\theta^n \overset{\text{inh}}{>} C(\mathfrak{F}_1, \mathfrak{F}_2) \right] \mapsto \text{disc.}$$

When we neglect the condition on the postsynaptic activity-dependent scaling condition, this set of orders is reduced to:

$$\left[C(\mathfrak{F}_1, \mathfrak{F}_1) \overset{\text{MR \& inh}}{>} C(\mathfrak{F}_2, \mathfrak{F}_1) ; \ C(\mathfrak{F}_2, \mathfrak{F}_2) \overset{\text{MR \& inh}}{>} C(\mathfrak{F}_1, \mathfrak{F}_2) \right] \mapsto \text{disc}$$

and, using the monotonous characteristic of the Hebbian learning mechanism (C:m), expressed to a contradiction:

$$\left[\mathfrak{F}_1 > \mathfrak{F}_2 ; \ \mathfrak{F}_1 < \mathfrak{F}_2 \right] \mapsto \text{disc} \qquad \Rightarrow\Leftarrow; \text{contradiction.}$$

$\square$

Thus, we conclude that such learning rules on synaptic plasticity in combination with a homeostatic mechanism are only able to reliably form associated memories or memories

that are connected in a sequential order. In the next Section, we validate this conclusion given the example of the SPaSS-learning rule.

4.2.2 Example: Learning Rule on Synaptic Plasticity and Synaptic Scaling

As already stated, the *learning rule on Synaptic Plasticity and Synaptic Scaling* (SPaSS) reliably builds up *memory representation*s (MRs) of environmental stimuli. Thus, here we analyze this learning rule in regard to its abilities to build up different *functional organization*s (FOs) of two interconnected neuronal populations $r, s \in \{1, 2\}$ dependent on the environmental and sensory input stimulation in a fully connected neuronal network ($p_{con} = 1$). To further reduce complexity of the parameter space, we do the analysis on the state-space for the respective normalized neuronal network model (see Appendix for details):

$$\textbf{input} \quad \tilde{\mathbf{F}}^{\text{ex}} := \left(\langle \tilde{F}_1^{\text{ex}} \rangle_t, \langle \tilde{F}_2^{\text{ex}} \rangle_t, \langle \tilde{F}_{\mathcal{B}}^{\text{ex}} \rangle_t \right)^{\mathsf{T}} \tag{4.17}$$

$$\textbf{activity} \quad \tau \dot{\tilde{F}}_i = (1 - \tilde{F}_i) \tilde{F}_i \left[\log \left(\tilde{F}_i^{-1} - 1 \right) + \beta (\mathrm{R} \phi_i - \epsilon) \right], \tag{4.18}$$

$$\text{with} \quad \phi_i = \mathrm{F}^{\text{max}} \omega^{\text{max}} \left(\sum_{r \in \{\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_{\mathcal{B}}\}} \sum_{j \in \mathcal{P}_r} (\tilde{\omega}_{i,j} - \tilde{\theta}) \tilde{F}_j + \tilde{\omega}^{\text{ex}} \sum_{k \in \mathcal{E}_r} \tilde{F}_k^{\text{ex}} \right), \tag{4.19}$$

$$\textbf{syn. weight} \quad \tilde{\tau}_\omega \dot{\tilde{\omega}}_{i,j} = \tilde{F}_j \tilde{F}_i - \tilde{\gamma} (\tilde{F}_i - \tilde{\mathbf{F}}^{\mathsf{T}}) \tilde{\omega}_{i,j}^2, \tag{4.20}$$

$$\text{with} \quad \tilde{\tau}_\omega = \frac{\tau_\omega}{\left(\mathrm{F}^{\text{max}} \sqrt{\mathrm{F}^{\text{max}} \gamma (1 - \tilde{\mathrm{F}}^{\mathsf{T}})} \right)}, \quad \tilde{\gamma} = \frac{1}{1 - \tilde{\mathrm{F}}^{\mathsf{T}}}. \tag{4.21}$$

By this, we map the whole activity and synaptic weight dynamics to a state space normalized to the maximal firing rate and maximal synaptic weight, respectively. Furthermore, the normalized time scale $\tilde{\gamma}$ for the synaptic scaling mechanism is express by the normalized target firing rate $\tilde{\mathrm{F}}^{\mathsf{T}}$, thus, reduces the number of free parameters. Note, $\tilde{x}$ indicates the normalized variable x.

The results presented in this Section (Sec. 4.2.2) have been published in the following article:

J. Herpich and C. Tetzlaff (2018). "Principles Underlying the Input-Dependent Formation and Organization of Memories". In: *bioRxiv*. A similar manuscript is currently under revision in Network Neuroscience.

Activity-dependent Constraints of Memory Representation and Functional Organization

In this Section, we map the general synaptic weight-dependent conditions for *memory representation* (MR) and *functional organization* (FO) of two interconnected memories (Theorem 6, p. 55) onto the respective activity-dependent conditions underlying the specific SPaSS-learning rule. Therefore, we map the conditions on MR, as well as the condition for an excitatory or inhibitory relation between two interconnected neuronal populations, defining the different forms of FOs to the average population activities (Fig. 4.1). For this, we specify the activity-dependent fixed point function of the applied normalized SPaSS-learning rule (Eq. 4.20) for the excitatory synaptic weights:

$$f_\omega^* = \omega_{i,j}^* = \sqrt[n]{\frac{C(\tilde{F}_j^*, \tilde{F}_i^*)}{\tilde{\gamma} H(\tilde{F}_i^*)}} = \sqrt[2]{\frac{\tilde{F}_j^* \tilde{F}_i^* (1 - \tilde{F}^{\mathrm{T}})}{\tilde{F}_i^* - \tilde{F}^{\mathrm{T}}}}. \tag{4.22}$$

This SPaSS-learning rule specific fixed point function with its specific plasticity-mechanism dependent condition on MR (Eq. 4.12) yields two distinct activity-regimes for the neuronal population $r \in \{1, 2\}$ to become a MR (Herpich and Tetzlaff 2018):

$$\mathfrak{w}_{r,r} \overset{\mathrm{MR}}{>} \tilde{\theta} \quad \overset{\text{Theorem 2}}{\Leftrightarrow} \quad \sqrt[2]{\frac{\mathfrak{F}_r \mathfrak{F}_r (1 - \tilde{F}^{\mathrm{T}})}{\mathfrak{F}_r - \tilde{F}^{\mathrm{T}}}} > \tilde{\theta} \quad \Leftrightarrow \quad \mathfrak{F}_r \gtrless \frac{\tilde{\theta}^2 \pm \tilde{\theta}\sqrt{D}}{2(1 - \tilde{F}^{\mathrm{T}})} \tag{4.23}$$

with $D = \tilde{\theta}^2 - 4\tilde{F}^{\mathrm{T}}(1 - \tilde{F}^{\mathrm{T}})$. Here, we see that the quadratic synaptic weight dependency ($n = 2$) yield two distinct open intervals $\mathfrak{F}^{\mathrm{low}}$, $\mathfrak{F}^{\mathrm{high}}$ (Fig. 4.1 Ai, white regimes) for the activity of population r. This corresponds to a sufficiently strong recurrent excitability conditioned on the synaptic weight-dependent definition for MR:

$$\mathfrak{F}^{\mathrm{low}} := \left(\tilde{F}^{\mathrm{T}}, \overline{\mathfrak{F}}\right) \qquad \text{with } \overline{\mathfrak{F}} := \frac{\tilde{\theta}^2 - \tilde{\theta}\sqrt{D}}{2(1 - \tilde{F}^{\mathrm{T}})}, \tag{4.24}$$

and

$$\mathfrak{F}^{\mathrm{high}} := \left(\underline{\mathfrak{F}}, 1\right) \qquad \text{with } \underline{\mathfrak{F}} := \frac{\tilde{\theta}^2 + \tilde{\theta}\sqrt{D}}{2(1 - \tilde{F}^{\mathrm{T}})}. \tag{4.25}$$

Note the open interval $\mathfrak{F}^{\mathrm{low}}$ for the activity of a population reveals a lower bound defined by the target activity $\tilde{F}^{\mathrm{T}}$ of the synaptic scaling mechanism, since all activities below this particular target firing rate lead to unbounded growth in the synaptic weights (Eq. 4.20). Consequently, a neuronal population at equilibrium can not reach activity levels below this target firing rate (Fig. 4.1 Ai, hatched regime; Eq. 4.22). Since the two distinct activity regimes $\mathfrak{F}^{\mathrm{low}}$ and $\mathfrak{F}^{\mathrm{high}}$ do not overlap, the enclosed activity regime $\overline{\mathfrak{F}} < \mathfrak{F}_r < \underline{\mathfrak{F}}$

corresponds to the *no-memory state* (nm) (nm $:= [\overline{\mathfrak{F}}, \underline{\mathfrak{F}}]$, Fig. 4.1 Ai, grey regime) with size $|\text{nm}| = \tilde{\theta}\sqrt{D}/(1 - \tilde{F}^T)$, where no proper memory representation can be formed (equivalent to $\mathfrak{w}_{rr} \leq \tilde{\theta}$). For the definition of different forms of FOs of two interconnected neuronal populations 1 and 2, the restricted $\mathfrak{F}_1 - \mathfrak{F}_2 -$activity phase space, enabling the formation of two memory representations, is constructed by four distinct regimes (Fig. 4.1 Aii; activity space that is not grayed out). These four distinct regimes emerge from all possible combinations of $\mathfrak{F}^{\text{low}}$ and $\mathfrak{F}^{\text{high}}$ in both dimensions of $\mathfrak{F}_1$ and $\mathfrak{F}_2$ for each neuronal population, respectively. Furthermore, these activity regimes are separated by the no memory phase (nm, Fig. 4.1 Aii, gray regimes) (Herpich and Tetzlaff 2018).

To describe the different forms of FOs of two interconnected populations, we first specify the activity-dependent conditions for an excitatory and inhibitory connection in between two populations (Eq. 4.13) by means of the fixed point function (Eq. 4.22) of the applied SPaSS-learning rule. In doing so, the condition for an inhibitory relation $\mathfrak{w}_{rs} < \tilde{\theta}$ becomes

$$\mathfrak{F}_s < \frac{\tilde{\theta}^2}{1 - \tilde{F}^T}\left(1 - \frac{\tilde{F}^T}{\mathfrak{F}_r}\right) \tag{4.26}$$

and the condition for an excitatory relation $\mathfrak{w}_{rs} > \tilde{\theta}$ becomes

$$\mathfrak{F}_s > \frac{\tilde{\theta}^2}{1 - \tilde{F}^T}\left(1 - \frac{\tilde{F}^T}{\mathfrak{F}_r}\right). \tag{4.27}$$

To clearly differentiate an excitatory-dominated connection from an inhibitory-dominated connection, we define a the marginal case that separates the two distinct cases, given by:

$$S_{rs} := \mathfrak{F}_s = \frac{\tilde{\theta}^2}{1 - \tilde{F}^T}\left(1 - \frac{\tilde{F}^T}{\mathfrak{F}_r}\right). \tag{4.28}$$

Consequently, for two interconnected neuronal populations with their inherent directed connections (R21 & R12), we get two such separatrixes, respectively:

$$S_{21} := \mathfrak{F}_1 = \frac{\tilde{\theta}^2}{1 - \tilde{F}^T}\left(1 - \frac{\tilde{F}^T}{\mathfrak{F}_2}\right), \qquad S_{12} := \mathfrak{F}_2 = \frac{\tilde{\theta}^2}{1 - \tilde{F}^T}\left(1 - \frac{\tilde{F}^T}{\mathfrak{F}_1}\right). \tag{4.29}$$

Whereby, S_{21} maps the synaptic weight dependent condition $\mathfrak{w}_{21} = \tilde{\theta}$ (Fig. 3.5 B, red vertical line) into the $\mathfrak{F}_1 - \mathfrak{F}_2 -$activity-space (Fig. 4.1 Aii, red curve) while S_{12} maps the synaptic weight dependent condition $\mathfrak{w}_{12} = \tilde{\theta}$ (Fig. 3.5 B, black horizontal line) onto the $\mathfrak{F}_1 - \mathfrak{F}_2 -$activity-space (Fig. 4.1 Aii, black curve).

By means of those two separatrixes (Eq. 4.29), the activity-dependent conditions, that specify the different forms of FOs, can be obtained by the specific combination of excitatory

and inhibitory relations for the two directed connections in between both populations:

Association. When both populations have activities above the respective separatrix $\mathfrak{F}_1 > S_{21}$ and $\mathfrak{F}_2 > S_{12}$, each corresponding to an excitatory relation, both memories are organized in an associational entity (Fig. 4.1 Aii, red phase).

Sequence. When the activity level of one population is above the respective separatrix (i.e. excitatory relation) and the activity level of the other population is below the respective separatrix (i.e. inhibitory relation), the system is in a sequence state. Hereby, the sequence with specific direction from memory 1 to memory 2 is formed when the activity $\mathfrak{F}_1$ is above the corresponding separatrix S_{21} while the activity $\mathfrak{F}_2$ stays below separatrix S_{12} (Fig. 4.1 Aii, green, seq:21) and vice versa for a sequence from memory 2 to memory 1 (yellow, seq:12).

Discrimination. When both populations have activities below the respective separatrix $\mathfrak{F}_1 < S_{21}$ and $\mathfrak{F}_2 < S_{12}$, each corresponding to an inhibitory relation, both memories are organized into an discriminatory functional organization (Fig. 4.1 A ii, blue regime). Note that the contradiction for the constraints on MRs and the functional discriminated memories already proofed in Corollary 9 is graphically shown (Herpich and Tetzlaff 2018).

In the following, we refer to this synaptic learning rule-specific mapping of the synaptic weight-dependent conditions for the different forms of FOs of two interconnected populations to the respective categorized $\mathfrak{F}_1 - \mathfrak{F}_2-$activity phase (Fig. 4.1,Aii) by the term characteristic activity phase space (χAPS).

Functional Organizations in Activity-space

To avoid the computational expensive calculation of the LTR resulting from the whole neuronal network simulation, we directly make use of the population model at equilibrium as described in Sec. 3.3.2. Thereby, to examine the resulting FO for a given input paradigm $\mathfrak{I} = (\mathfrak{I}_1, \mathfrak{I}_2)^{\mathsf{T}}$ (Eq. 3.20), we only need to calculate the synaptic learning rule-dependent average population's activities $\mathfrak{F}(\mathfrak{I}, f_\omega^*) = (\mathfrak{F}_1(\mathfrak{I}_1, \mathfrak{F}, f_\omega^*), \mathfrak{F}_2(\mathfrak{I}_2, \mathfrak{F}, f_\omega^*))^{\mathsf{T}}$ in the fixed point (Theorem 4, p. 42). By this, we reduce the high-dimensional mathematical problem in the neuronal activity- and synaptic-weight-space of the neuronal network model towards a two-dimensional problem in the population's $\mathfrak{F}_1 - \mathfrak{F}_2-$activity space in the population model at equilibrium (Herpich and Tetzlaff 2018).

One exemplary calculation for a specific input stimulation of $\mathfrak{I} = (0.95, 0.4)^{\mathsf{T}}$ is shown in Fig. 4.1 Bi. Here, the intersection (green dot) of the black thick curve (i.e. fixed average activity of population 1) and yellow thick curve (i.e. fixed average activity of population 2)

Figure 4.1: Memory representation and functional organizations of two interconnected neuronal populations in activity-space. (Ai) Mapping of the synaptic weight-dependent condition on MR for population r onto its respective activity-dependent condition (blue curve in white area). **(Aii)** Conditions for the different FOs of two interconnected memories mapped onto the $\mathfrak{F}_1 - \mathfrak{F}_2-$activity space of the respective populations 1 and 2. **(B i)** Given this reduced 2-dimensional $\mathfrak{F}_1 - \mathfrak{F}_2-$activity space one can calculate the equilibrium state of the population activities $\mathfrak{F}_1, \mathfrak{F}_2$. The Equilibrium is determined by the intersection of both fixed point functions given an specific input stimulation $\mathfrak{I}_1, \mathfrak{I}_2$ and further compare the resulting fixed point with the constraints on the FO of interconnected MRs (Aii), we obtain the respective FO. **(Bii)** Same as in (Bi) for the whole $\mathfrak{I}_1 - \mathfrak{I}_2-$input space with $\mathfrak{I}_r \in [0,1]\mathrm{F}^{max}$, respectively. Used parameters: $\tilde{\theta} = 0.5$, $\tilde{\mathrm{F}}^{\mathrm{T}} = 0.05$, $n_\epsilon = 20$ (adapted from Herpich and Tetzlaff 2018).

indicates the system's stable state and, by this, the respective FO in relation to the characteristic activity phase space (χAPS). To determine the FO (Fig. 4.1 Bi), the stable state (green dot) has to be compared to the activity regimes leading to MRs (white activity phase), as well as the separatrixes S_{21} (thin red curve) and S_{12} (thin black curve), separating the activity regimes, leading to an excitatory or inhibitory connection linking both populations, as described above. For this particular example, the specific input stimulation $\mathfrak{I} = (0.95, 0.4)^\mathsf{T}$ leads to the formation of a sequence 21, with an excitatory connection from memory 1 to memory 2 (Herpich and Tetzlaff 2018).

For a comprehensive analysis, we do the above described procedure for the whole $\mathfrak{I}_1 - \mathfrak{I}_2-$input space (Fig. 4.1 Bii). The results, mapped onto the applied input space, are shown in Fig. 3.6 D (p. 49). With this analysis, we validate the hypothesis that synaptic plasticity in combination with a postsynaptic activity-dependent synaptic scaling mechanism leads to the formation of associations and sequences in both directions but lacks the formation of discrimination. This shortcoming is already graphically underpinned by the characteristic $\mathfrak{F}_1 - \mathfrak{F}_2$-activity phase space ($\chi$APS, Fig. 4.1 Aii), where the activity phase, leading to the discrimination of interconnected memories, is completely overlaid by the activity space resulting in insufficient strong representations of the given environmental input stimulation (nm-state, gray activity space). In addition to functionally connected

MRs, there is a broad input regime that leads to no proper MRs in either one or both populations (Fig. 4.1 B ii, gray dots) (Herpich and Tetzlaff 2018).

Thus, we see that the synaptic learning rule-specific characteristic activity phase space (χAPS) (Fig. 4.1 Aii), provides sufficient insights to qualitatively assess the system's ability to build up different FOs of interconnected memories. Consequently, an evaluation of the system's fixed point is not required anymore, but, nevertheless, contributes to a full analysis (Herpich and Tetzlaff 2018).

Synaptic Plasticity Dominated Activity-input Mapping

A closer look onto the input space for the numerical simulations (Fig. 3.6) and analytical calculations (Fig. 4.1) indicates that for very low external input stimulations $\langle \tilde{F}_r^{\text{ex}} \rangle_{\text{LTR}} \in [0, 0.15]$ ($r \in \{1, 2\}$ for the numerical simulation (Fig. 3.6 C) and $\mathfrak{I}_r \in [0, 0.2]$ for the reduced population model (Fig. 3.6 D)) the system already associates both neuronal populations. Such an intrinsic association of the population model is in contradiction to the *synaptic-plasticity-and-memory* (SPM) hypothesis (Hebb 1949; Martin et al. 2000) since higher rates of neuronal activity should cause stronger correlation-based plasticity mechanisms, responsible for learning (Herpich and Tetzlaff 2018).

This inadequacy of the here investigated SPaSS-learning rule results from the quadratic weight-dependency of synaptic scaling that yields a so-called *up-scaling* effect of the synaptic dynamics. This effect can be seen in Fig. 4.1 Ai, where the fixed point function for the average recurrent population's synaptic weight (Eq. 4.22) shows a global minimum. By means of curve sketching, we determine this local minumum for the specific activity level of $\mathfrak{F}_r = \mathfrak{F}^{\text{min}} = 2\tilde{F}^{\text{T}}$. Consequently, all activities below $\mathfrak{F}^{\text{min}}$ (i.e. low average population activities, $\mathfrak{F}_r \in (\tilde{F}^{\text{T}}, \mathfrak{F}^{\text{min}})$) are dominated by synaptic scaling and yield higher recurrent synaptic weight values than for $\mathfrak{F}_r = \mathfrak{F}^{\text{min}}$. The same effect holds for the average synaptic weights connecting the different populations (not shown graphically), that finally leads to such an intrinsic association of both populations. Note, this effect of up-scaling has already made an impact in the activity-dependent constraint on MR where we obtained two distinct regimes $\mathfrak{F}^{\text{low}}$ (Eq. 4.24) and $\mathfrak{F}^{\text{high}}$ (Eq. 4.25) leading to MR (Herpich and Tetzlaff 2018).

Synaptic scaling is experimentally observed (Turrigiano and Nelson 2004) and theoretically implemented (Tetzlaff et al. 2013) to stabilize and regulate the synaptic weight dynamics, acting as a homeostatic mechanism. However, the mathematical implementation by a quadratic synaptic weight behavior lead to biologically and theoretically implausible up-scaling effects. In principal, an increase in the population's activity should result in an increase of the respective synaptic weights. In other words, all up-scaling effects
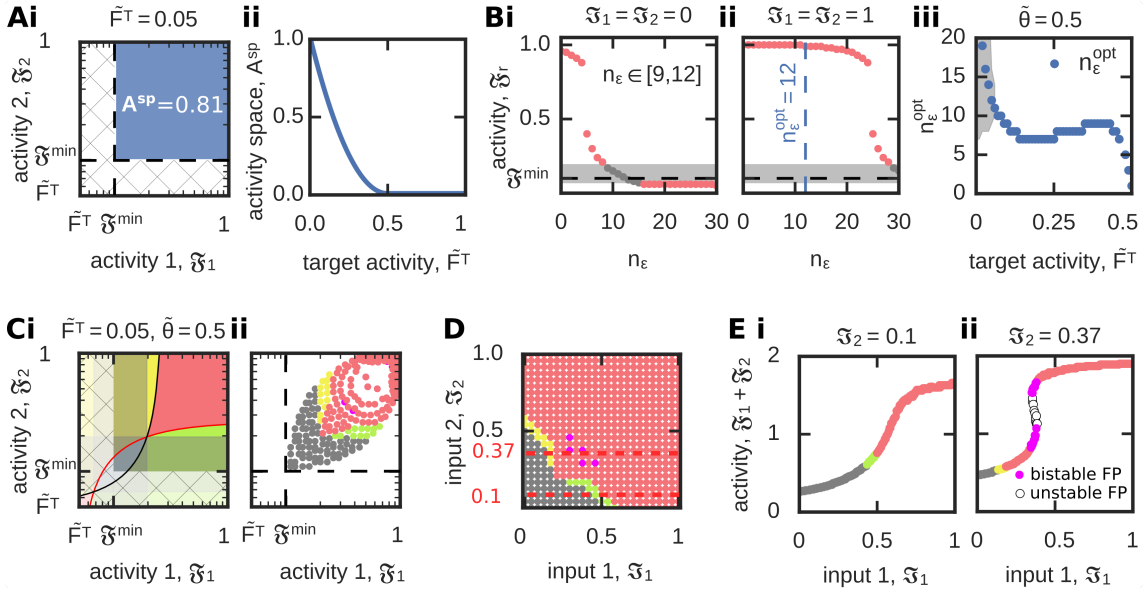
Figure 4.2: Synaptic plasticity dominated functional organization of two interconnected memories. (A, B) The regime A^{SP}, in which synaptic plasticity dominates the synaptic dynamics, depends on the target firing rate $\tilde{F}^T$ of the SPaSS-learning rule (A) and inflexion point $\tilde{\epsilon} = n_\epsilon \tilde{u}^{max}$ of the neuronal activity function (B). **(A)** The area of the $\mathfrak{F}_1 - \mathfrak{F}_2-$activity phase space (A i, blue space) leading to synaptic plasticity dominated FOs decreases with increasing target firing rate $\tilde{F}^T$ (A ii). **(B)** The inflexion point (measured in n_ϵ) determines the activity-input mapping such that for the same input different activities and, thus, different FOs are realized. If the inflexion point equals n_ϵ^{opt}, $\mathfrak{F}_r \approx \mathfrak{F}^{min}$. **(B i)**: $\mathfrak{I}_1 = \mathfrak{I}_2 = 0$. **(B ii)**: $\mathfrak{I}_1 = \mathfrak{I}_2 = 1$. **(B iii)** The value of n_ϵ^{opt} (blue) depends on the target firing rate $\tilde{F}^T$. The grey area specifies all n_ϵ that yield to the no-memory state. **(C - E)** One example of synaptic plasticity dominated formation of FOs. Used parameters: $\tilde{F}^T = 0.05$, $\tilde{\theta} = 0.5$, $n_\epsilon = 12$. **(C)** Although the system implies regimes of scaling-dominated synaptic dynamics (hatched area; i), the activity-input mapping excludes that the system can reach these by external inputs (ii). **(D)** The resulting $\mathfrak{I}_1 - \mathfrak{I}_2-$input phase space, color-coded according to the respective FOs, shows that associations can only be formed for stronger inputs (compare to Fig. 3.6 Ciii). **(E)** The sum of both population activities ($\mathfrak{F}_1 + \mathfrak{F}_2$) for a fixed input $\mathfrak{I}_2$ shows for some cases the existence of two equilibrium states encoding associations (pink). **(E i)**: $\mathfrak{I}_2 = 0.1$. **(E ii)**: $\mathfrak{I}_2 = 0.37$ (adapted from Herpich and Tetzlaff 2018).

should be avoided by the computational model. Hereinafter, the dynamics of the population model should operate in the remaining activity space of $\mathfrak{F}_r \in [\mathfrak{F}^{min}, 1]$. Consequently, the correlation-based synaptic plasticity dominated and biologically plausible activity regime (A^{SP}) of two interconnected populations is given by the

$$[\mathfrak{F}^{min}, 1] \times [\mathfrak{F}^{min}, 1] - \text{activity space} \tag{4.30}$$

(Fig. 4.2 Ai, blue space, $A^{SP} = (1 - 2\tilde{F}^T)^2$). The size of this synaptic plasticity dominated

activity phase space (A^{sp}) depends on the synaptic scaling specific target firing rate $\tilde{\mathrm{F}}^{\mathrm{T}}$ (Fig. 4.2 Aii) and vanishes for target activities larger than 0.5. Thus, we restrict the parameter space for the target activity level to $\tilde{\mathrm{F}}^{\mathrm{T}} \in (0, 0.5)$ (Herpich and Tetzlaff 2018).

In order to avoid long-term neuronal activities being within the scaling-dominated activity regime $\mathfrak{F}_r \in (\tilde{\mathrm{F}}^{\mathrm{T}}, \mathfrak{F}^{\mathrm{min}})$, we investigate the characteristics underlying the mapping of the input stimulation ($\mathfrak{I}$) onto the neuronal population's activities ($\mathfrak{F}$). However, the sigmoidal transfer function, that maps the neuronal membrane potential to a neuronal activity (Eq. 3.4), is mainly determined by the global parameter of the inflexion point ϵ. This parameter represents the critical membrane potential of a neuron corresponding to the strongest increase its neuronal activity. Thus, in the following, we derive an appropriate choice for this membrane potential-related parameter ϵ on the more detailed level of single neuron dynamics. Therefore, we consider one presynaptic neuron j connected via a synapse with maximal strength ($\omega_{ij} = \omega^{\mathrm{max}}$) a postsynaptic neuron i. Using this two-neuron set up, we define the maximal evokable membrane potential ($\mathrm{u}^{\mathrm{max}}$) within the postsynaptic neuron i by assuming maximal pre- and postsynaptic firing rates ($F_j = F_i = \mathrm{F}^{\mathrm{max}}$). By means of the ODE for the leaky membrane potential (Eq. 3.3), we can calculate the maximal network ($\sum_{k \in \mathcal{E}_i} F_k^{\mathrm{ex}} = 0$) evoked membrane potential, given by $\mathrm{u}^{\mathrm{max}} := R\mathrm{F}^{\mathrm{max}}(\omega^{\mathrm{max}} - \theta)$. With this neuron specific quality of $\mathrm{u}^{\mathrm{max}}$, we can interpret the inflexion point ϵ for the activity function of a particular neuron in units of n_ϵ. The parameter n_ϵ represents the number of maximally activated ($F_j = \mathrm{F}^{\mathrm{max}}$) and maximally interconnected ($\omega_{ij} = \omega^{\mathrm{max}}$) presynaptic neurons j. With this approach, we express the inflexion point as a linear function of the number of interconnected neurons $\epsilon = \mathrm{n}_\epsilon \mathrm{u}^{\mathrm{max}}$. Please note, the system analyzed in Fig. 4.1 has the specific parameter of $\mathrm{n}_\epsilon = 20$ defining the inflexion point and $\tilde{\mathrm{F}}^{\mathrm{T}} = 0.05$ defining the target activity level for the synaptic scaling mechanism (Herpich and Tetzlaff 2018).

Using the interpretation of the global parameter for the inflexion point ϵ, we now analyze the intrinsic activity dynamics (i.e. no external inputs, $\mathfrak{I}_1 = \mathfrak{I}_2 = 0$; i.e. intrinsic state) of one population r described by the population model at equilibrium (Fig. 4.2 B i). By changing the specific parameter n_ϵ determining the inflexion point of the neurons, we calculate the resulting intrinsic population's activity $\mathfrak{F}_r$ as well as the corresponding FO with the interconnected population, indicated by the colored dots. The results show that the two interconnected populations either form an association (Fig. 4.2 Bi, red dot: association) or they are in the no memory state (Fig. 4.2 Bi, gray dot: nm-state), intrinsically. The association occurs in two distinct regimes: In the regime of $\mathrm{n}_\epsilon < 9$, the neurons are too easy to excite underpinned by high intrinsic activity levels above the global minimum $\mathfrak{F}^{\mathrm{min}}$, partially reaching the maximum activity level. For $\mathrm{n}_\epsilon > 16$, the intrinsic activity level of the neuronal population is suppressed to levels below $\mathfrak{F}^{\mathrm{min}}$. Hence, a scaling-induced FO

of an association is formed. Moreover, there is a parameter regime $9 \leq n_\epsilon \leq 16$, where the system intrinsically is in the nm-state. In particular, for $9 \leq n_\epsilon \leq 12$ the intrinsic activity level is within the biological realistic activity regime A^{sp} above $\mathfrak{F}^{\mathrm{low}}$. This parameter regime prevents an input-*in*dependent association of two interconnected neuronal populations. Hence, to drive the computational model towards this desired activity regime $9 \leq n_\epsilon \leq 16$, we assign the parameter regime of $9 \leq n_\epsilon \leq 12$ leading to biological realistic dynamics. In particular, activity levels nearby the local minimum $\mathfrak{F}^{\mathrm{min}}$ are generated by the specific parameter value of $n_\epsilon^{\mathrm{opt}} = 12$ (Herpich and Tetzlaff 2018).

For a comprehensive model of interacting populations, it is not only important to consider a biological realistic intrinsic state, but also to make sure that the model fullfils all desired requirements. In our specific case, it is desired that the model is able to reach each single distinct $\mathfrak{F}_1 - \mathfrak{F}_2-$activity phase spaces corresponding to a specific FO, such as association (red), discrimination (blue) and sequence (green and yellow). In this sense, we do the same analysis as for the intrinsic state (Fig. 4.2 B i), now considering maximal external input stimulation $\mathfrak{I}_1 = \mathfrak{I}_2 = 1$ (Fig. 4.2 B ii). By this, we specify the parameter range for n_ϵ that enables the neuronal population to exhibit a sufficient strong activity level that, at least, corresponds to an association of two interconnected populations. For the restricted parameter range of $n_\epsilon \in [0, 12]$, a population r reaches its maximal firing rate of $\mathfrak{F}_r = 1$; therefore it allows a broad range of the desired activity phase space ($\mathfrak{F}_r \in [\mathfrak{F}^{\mathrm{min}} - 1]$) to be covered with changing input stimulation $\mathfrak{I}_r \in [0, 1]$. In contrast, for the parameter range $n_\epsilon > 24$ a single population's activity stays far from the maximal firing rate, in particular, for $n_\epsilon > 27$ such populations are not able to build up proper MRs, although they are maximally stimulated. Thus, combining both analysis (Fig. 4.2 Bi,ii) an optimal choice for the parameter $n_\epsilon^{\mathrm{opt}}$ defining the inflexion point of the neuronal activity function is $n_\epsilon^{\mathrm{opt}} = 12$ (Herpich and Tetzlaff 2018).

However, this analysis on the population's intrinsic and maximal evoked activity is done for the specific target firing rate of $\tilde{\mathrm{F}}^{\mathrm{T}} = 0.05$ of the synaptic scaling mechanism. As we can see in Fig. 4.2 A, an optimal choice of $n_\epsilon^{\mathrm{opt}}$ also depends on the specific choice for the target firing rate $\tilde{\mathrm{F}}^{\mathrm{T}}$. Thus, in Fig. 4.2 Biii, we provide the analysis shown in Fig. 4.2 Bi,ii for different values of the target activities $\tilde{\mathrm{F}}^{\mathrm{T}}$. The gray shaded area corresponds to the parameter that yields an intrinsic nm-state of the populations (Herpich and Tetzlaff 2018).

In the following (Fig. 4.2 C-E), we will consider $n_\epsilon = 12$ and $\tilde{\mathrm{F}}^{\mathrm{T}} = 0.05$. These parameter values ensure that the population system is intrinsically in the nm-state and, in addition, that for each single input, the resulting FO of interconnected populations is due to a dominating correlation-based synaptic plasticity mechanism. For clarity, the activity regime yielding scaling-dominated learning (hatched area in Fig. 4.2 C i) is theoretically possible; however, the adapted activity-input mapping now assures that this regime cannot be

reached for given external inputs (Fig. 4.2 C ii and D). An analysis of this adapted system (Fig. 4.2 C ii and D) shows that low inputs $\mathfrak{I}_1$, $\mathfrak{I}_2$ lead to a nm-state (gray), while in a small input regime sequences are formed (yellow and green). For sequences, the excitatory connection between both populations is from the population receiving a stronger input to the population receiving the weaker input stimulation. If both inputs are strong, an association between the MRs is being build (red). Note that there is a small bimodal regime with two long-term equilibrium states both being an association (pink; see two exemplary cross sections in Fig. 4.2 E) (Herpich and Tetzlaff 2018).

Monotonous Activity Function Preserves the Topology of Functional Organization

The specific topology of the already described characteristic $\mathfrak{F}_1 - \mathfrak{F}_2-$activity phase ($\chi$APS, Fig. 4.2 Ci), separating the distinct activity spaces leading to MR, as well as different FOs of two interconnected populations, is preserved within the $\mathfrak{I}_1 - \mathfrak{I}_2-$input-space (Fig. 4.2 D). This is due to the monotonously increasing sigmoidal-shaped transfer function of the neuronal membrane potential to a neuronal firing rate (Eq. 3.4). In Fig. 4.2 E, we exemplary show the $\mathfrak{F} - \mathfrak{I}-$curve, where we sum the neuronal activities of both neuronal populations ($\mathfrak{F}_1 + \mathfrak{F}_2$). Here, the environmental input stimulation $\mathfrak{I}_1$ on neuronal population 1 changes while the input stimulation on neuronal population 2 is kept constant to $\mathfrak{I}_2 = 0.1$ (Fig. 4.2 Ei) and $\mathfrak{I}_2 = 0.37$ (Fig. 4.2 Eii), corresponding to the horizontal dashed red lines in Fig. 4.2 D, respectively. In contrast to the monotonous increase of the sum of both neuronal firing rates in Fig. 4.2 Ei, the sum bifurcates in two distinct fixed point regimes in Fig. 4.2 Eii for a stronger environmental input stimulation on neuronal population 2. However, this does not change the topology of the FOs within the $\mathfrak{I}_1 - \mathfrak{I}_2-$input space.

Consequently, input frequencies $\mathfrak{I}$, that are slightly higher than the input frequencies leading to the no nm-state (Fig. 4.2 D, gray), drive the system into a sequential entity (green and yellow). For even stronger environmental input frequencies, the system gets organized into an associational entity.

Parameter-dependency of Functional Organizations

In the previous Sections, we have analyzed one specific set of global parameters of the population model. These are $\tilde{F}^T = 0.05$ for the target activity of the scaling mechanism and $\tilde{\theta} = 0.5$ for the average level of inhibition inherent of the neuronal network. In the following, we will analyze which kind of FOs can be formed with different parameters acting on the synaptic weights (Herpich and Tetzlaff 2018).
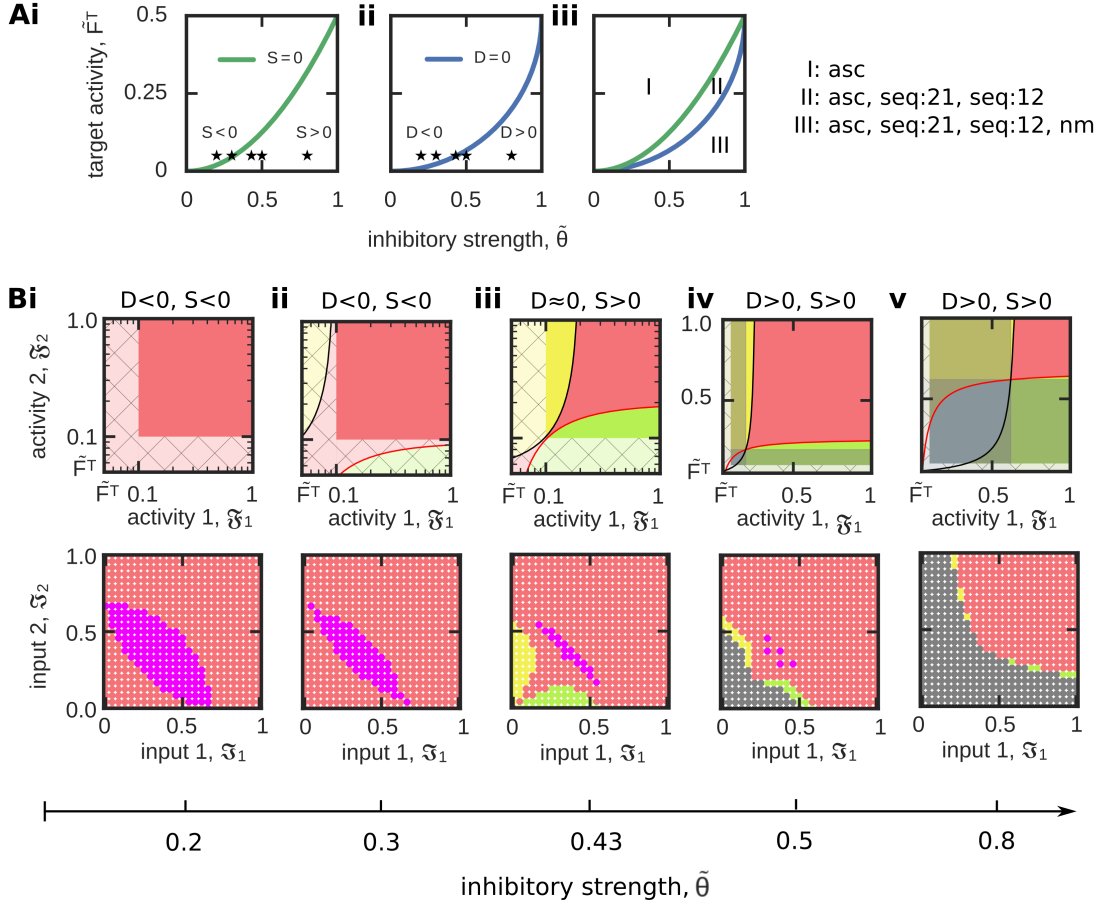
Figure 4.3: Quantification of the system ability to form different FOs dependent on the parameters $\tilde{F}^T$ **and** $\tilde{\theta}$**. (A)** The measure S (i; green) indicates whether sequences can be formed (S > 0), while the measure D (ii; blue) specifies the existence of no-memory states (D > 0). (iii) Both measures together separate the $\tilde{F}^T - \tilde{\theta}-$parameter phase space into three distinct regimes. Please see main text for details. **(B)** For a constant target activity ($\tilde{F}^T = 0.05$), we show several examples of resulting functional organizations for different values of inhibition $\tilde{\theta}$ (asterisks in (A)) in activity- (top row) and input-space (bottom) (adapted from Herpich and Tetzlaff 2018).

In general, the *no-memory state* (nm) implies that neuronal populations can exist which do not encode information (or have "forgotten" this information); hence, this state has a strong influence on the overall system properties. As already derived, the size of this state is given by $|nm| = \tilde{\theta}\sqrt{D}/(1 - \tilde{F}^T)$ with $D = \tilde{\theta}^2 - 4\tilde{F}^T(1 - \tilde{F}^T)$. As $\tilde{\theta}$ is larger than zero, the discriminant D, determined by $\tilde{F}^T$ and $\tilde{\theta}$, defines whether the nm-state can exist in a given system (Fig. 4.3). Thus, the $\tilde{F}^T$-$\tilde{\theta}$-dependency of D (Fig. 4.3 Aii) determines the potential of the system to form sufficient MRs of interconnected populations. Thus, the $\tilde{F}^T$-$\tilde{\theta}$-parameter-space corresponding to D < 0 will globally exhibit sufficiently strong MRs of both populations, thus excludes the nm-state. Whereas, in the remaining

$\tilde{F}^T$-$\tilde{\theta}$-parameter-space the nm-state can still occur dependent on the input stimulation. One exemplary limit case for this categorization is given in Fig. 4.3 Biii for $\tilde{\theta} = 0.5$ (Herpich and Tetzlaff 2018).

In addition, the relation of the activity levels to the separatrices S_{21} and S_{12} (Eq. 4.29) define which kind of FO is present. However, the separatrices have to reach the synaptic plasticity dominated activity-regime (A^{sp}; $\mathfrak{F}_r \in (\mathfrak{F}^{min}, 1)$) to enable the formation of sequences (*seq:12*: $\mathfrak{F}_1 < S_{21}$, $\mathfrak{F}_2 > S_{12}$; *seq:21*: $\mathfrak{F}_1 > S_{21}$, $\mathfrak{F}_2 < S_{12}$) and discrimination (*disc*: $\mathfrak{F}_1 < S_{21}$, $\mathfrak{F}_2 < S_{12}$). This necessary condition can be obtained by comparing the χAPSs for two different sets of parameters: In Fig. 4.3 Bii, no sequence is made up in the biologic realistic activity regime A^{sp}, whereas in Fig. 4.3 Biii the separatrices reach the restricted activity regime A^{sp}. S_{12} (S_{21}) monotonously increase with $\mathfrak{F}_1$ ($\mathfrak{F}_2$), the maximum difference S between the lower activity level ($\mathfrak{F}^{min}$) of the synaptic plasticity dominated dynamics and the separatrix is given for $\mathfrak{F}_1 = 1$ ($\mathfrak{F}_2 = 1$) such that

$$ S = S_{12}[\mathfrak{F}_1 = 1] - \mathfrak{F}^{min} = \tilde{\theta}^2 - 2\,\tilde{F}^T. \tag{4.31} $$

Thus, as soon as the measure S (Eq. 4.31) takes on positive values, the separatrices reaches A^{sp} (Herpich and Tetzlaff 2018).

The overall potential of the system to form diverse FOs is determined by combining both categorizations D and S of the $\tilde{F}^T$-$\tilde{\theta}$-parameter-space (Fig. 4.3 Aiii). This leads to three functionally different system configurations: For $D < 0$, $S < 0$, the system can only form associations (regime I in Fig. 4.3 Aiii, Bi,ii). If $D \leq 0$, $S > 0$, the system can form either associations or sequences (*seq:12* as well as *seq:21*; regime II in Fig. 4.3 Aiii, Biii). And if $D > 0$, $S > 0$, associations, sequences, and the nm- state can be formed and reached by the system (regime III in Fig. 4.3 Aiii, Biv,v). Thus, this analysis shows that with larger average inhibitory weight $\tilde{\theta}$ and smaller target activity level $\tilde{F}^T$ the system receives a larger repertoire of FOs. Once more, this analysis also shows that the FO of discrimination cannot be formed in a long-term manner. Although for large values of inhibition both separatrices intersect (see, for instance, Fig. 4.3 Bv), both activity levels could be simultaneously below their corresponding separatrix (blue), the resulting area of discrimination cannot be reached by any inputs $\mathfrak{I}$ since in all these cases the neuronal populations cannot serve as MRs (gray activity phase) (Herpich and Tetzlaff 2018).

4.2.3 The Discrimination of Memories Challenges Synaptic Plasticity

In the first part of this Section (Sec. 4.2.1), we have analyzed a general form of synaptic learning rules pooling all rules, that are based on correlation-based synaptic plasticity in combination with any activity-dependent homeostatic mechanism, by means of the derived population model at equilibrium (Chapter 3). By this, we focused on the ability of such generic formulations of learning rules to form all possible forms of FOs of interconnected MRs. From our theoretical considerations we conclude that such synaptic learning rules are able to form the FO of association and sequences of interconnected memories. Furthermore, we have proofed that these learning rules are not sufficient to distinguish the two different processes of memory formation their discrimination (Herpich and Tetzlaff 2018).

This shortcoming is due to the correlation-based formulation of synaptic plasticity in combination with a postsynaptic activity-dependent homeostatic mechanism. This combination of plasticity mechanisms mathematically couples the condition of MR with the condition of FO of interconnected memories (Theorem 6). In particular, the monotonous characteristic of the correlation-based plasticity mechanism excludes the formation of discriminated memories. Here, even the correlation-independent dynamics caused by the postsynaptic activity dependent homeostatic mechanism are not sufficient to decouple these conditions. However, such a homeostatic mechanism enables the formation of sequences. Thus, we can assign an additional functional role on the level of memory interaction of enabling sequential orders in between MRs to the homeostatic mechanism besides their inherent functions on the neuronal level of synaptic stabilization (Tetzlaff et al. 2011; Zenke et al. 2013; Zenke et al. 2017) and homeostatic regulation (Abbott and Nelson 2000; Turrigiano and Nelson 2004) of neuronal activities (Herpich and Tetzlaff 2018).

With this general approach to investigate the abilities of synaptic learning rules to form different FOs of memories, we can generalize that all synaptic learning rules, which are composed of correlation-based plasticity and a postsynaptic activity-dependent homeostatic term, for instance, Oja's rule (Oja 1982) or BCM rule (Bienenstock et al. 1982) are only able to form associations and sequences of memories (Herpich and Tetzlaff 2018).

In the second part (Sec. 4.2.2), we have underpinned these theoretical predictions with an extensive analysis of the specific *learning rule on Synaptic Plasticity and Synaptic Scaling* (SPaSS)-learning rule (Tetzlaff et al. 2011; Tetzlaff et al. 2012). For such learning rules it was already shown that they are sufficient to form individual MRs (Tetzlaff et al. 2013; Tetzlaff et al. 2015). Our analysis, confirmed that a neuronal system with correlation-based synaptic plasticity in combination with postsynaptic activity-dependent synaptic scaling mechanism is able to form associations and sequences but is not able to form two

inhibitory relations in between two memory representations which are weaker than the average inhibition, i.e. discrimination, as predicted in the first part.

In addition to the association and sequential order of MRs, a discrimination of MRs is a functionally very important ability for a neuronal system (James 1890), in particular for the construction of schemata (Preston and Eichenbaum 2013). This feature of discriminating environmental input stimulations is, for instance, necessary for the development of new schemata, if new information of the environment conflicts with the prior knowledge that is stored in the existing schemata (Preston and Eichenbaum 2013). Furthermore, an anatomical connection of memories, as it is the case in our model, does not necessarily imply a functional connection. Therefore, we are challenged to decouple the condition for MRs and FOs of anatomically connected neuronal populations to overcome the lack of discriminatory FOs of memories (Herpich and Tetzlaff 2018).

Thus, in the next Section (Sec. 4.3), we will introduce an additional plasticity mechanism on the underlying synaptic dynamics of the network and derive its specific properties that, besides the formation of associations and sequences, allow for the discrimination of MRs.

4.3 Correlation-and-Causality-based Plasticity and Homeostasis

Our previous analysis on learning rules based on the correlation of pre- and postsynaptic activities combined with a homeostatic mechanism has shown that those learning rules do not reliably describe the formation of all different forms of *functional organization* (FO) of interconnected memories. In particular, the formation of discriminated memories is excluded by this category of synaptic learning rules, as the conditions on *memory representation* (MR) and FO are inevitable coupled via the postsynaptic-activity dependent homeostatic mechanism (Theorem 6, p. 55).

The generic approach to mathematically formulate the experimentally observed *long-term potentiation* (LTP) (Bliss and Lømo 1973; Bliss and Collingridge 1993) by the plain correlation of pre- and postsynaptic activity levels, i.e. $C(F_j, F_i) = F_j F_i$ (Dayan and Abbott 2001; Gerstner and Kistler 2002; Choe 2015), does not account for the plasticity mechanism of *long-term depression* (LTD) (Dudek and Bear 1992), also experimentally observed. The mechanisms of synaptic plasticity guiding LTP (LTD), i.e. an increase (decrease) of AMPA and NMDA-receptors within the postsynaptic membrane (i.e. change in *postsynaptic density* (PSD)) causing an enhanced (decreased) excitability of the postsynaptic neuron, have already been described in Sec. 2.1.4. However, both mechanisms are mediated by specific levels of pre- and postsynaptic cytosolic Ca^{2+}-concentrations guiding the presynaptic release of neurotransmitters (proportional to the presynaptic firing rate) and the size of the postsynaptic density (proportional to the pre- and postsynaptic firing rate), respectively. There is experimental evidence (Mulkey and Malenka 1992) showing that the different forms of synaptic plasticity are initiated by distinct activity levels of pre- (F_j) and postsynaptic firing (F_i) affecting the respective Ca^{2+}-concentrations. In particular, it was shown that high neuronal activities cause LTP, whereas, low neuronal activities induce the process of LTD.

Furthermore, the mathematical model describing correlation-based synaptic plasticity $C(F_j, F_i) = F_j F_i$, (Dayan and Abbott 2001; Gerstner and Kistler 2002; Choe 2015) fails to describe specific cases, explained in the following. When there is a high presynaptic transmitter release (proportional to presynaptic firing) but a small PSD, the impact of the presynaptic neurotransmitters is too small to evoke a sufficient strong depolarization of the postsynaptic cell. This is not captured in the mathematical model that only depends on the plain correlation of pre- and postsynaptic firing rates. The same restriction on plasticity holds in the opposite case for a low amount of presynaptic transmitter release combined with a high postsynaptic density. In both cases, the neurons may be anatomically but not causally connected. Thus, modeling synaptic plasticity by a dominating correlation-based mechanism of pre- and postsynaptic firing can not describe compre-

hensively the complexity of synaptic plasticity mechanisms of the enclosing synapses. In particular, the experimentally observed distinction in different levels of pre- and postsynaptic activity being responsible for the initiation of opposing synaptic plasticity effects, such as LTP and LTD, is not present in the used correlation-based synaptic plasticity mechanism of rate coded neurons. When we further take into account the mechanisms of *spike-timing dependent plasticity* (STDP) (Markram et al. 1997; Feldman 2012), we recognize that the mathematical formulation of correlation-based plasticity of rate coded neurons ($C(F_j, F_i) = F_j F_i$) does not account for such STDP mechanism. This level of description fails to describe that correlated spike-times of pre- and postsynaptic neurons have to be within a given time window to successfully induce LTP (Dayan and Abbott 2001), whereas uncorrelated spike events either in the pre- or postsynaptic neuron would trigger LTD. In other words, the generic approach to describe synaptic plasticity based on the correlation of pre- and postsynaptic firing levels, previously described, does not take into account the failure of presynaptic firing to evoke the respective postsynaptic firing (*non-causality*) for the enclosed synapse. Thus, when synaptic plasticity is mainly defined by the correlation of pre- and postsynaptic firing, it only describes LTP. However, there is no measure for connected but causally uncorrelated neurons, in particular a measure for non-correlated pre- and postsynaptic firing that would result in a decrease of the synaptic efficacy (i.e. LTD). Nevertheless, such a mechanism is mandatory for learning and memory, because these processes require both activity-dependent synaptic potentiation and depression (Song et al. 2000).

To overcome this inability of the mathematical formulation for synaptic plasticity, in this Section we will introduce another plasticity mechanism that shall account for such LTD in addition to the previously analyzed learning rule Eq. (4.7) in Sec. 4.2.1. Thus, the resulting *augmented model for synaptic plasticity* shall be able to decouple the two distinct processes of memory formation and their FO of interconnected memories also enabling *discrimination* (disc). To investigate the emerging new properties of this augmented model, here we analyze if and under which specific conditions the combination of the different plasticity mechanisms involved are able to describe all different forms of FO.

Definition 5 *Synaptic learning rules that are based on the combination of correlation-based Hebbian learning $C(F_i, F_j)$ and postsynaptic activity-dependent synaptic scaling $H(F_j)$ and further augmented by a mechanism $R(F_i, F_j)$ being a measure for the non-causality of the pre- and postsynaptic neuronal activities:*

$$f_\omega(F_j, F_i, \omega_{i,j}) = \tau_\omega \dot{\omega}_{i,j} = C(F_j, F_i) - \nu R(F_j, F_i) - \gamma H(F_i)\omega_{i,j}^n \qquad (4.32)$$

with ν being the time scale of the causality-based plasticity mechanism R, are summarized as learn-

ing rules on Correlation-and-Causality-based Synaptic Plasticity in combination with a Homeostatic mechanism.

Corollary 10 *Let f_ω be a synaptic learning rule on correlation-and-causality-based plasticity in combination with a synaptic-weight dependent homeostatic mechanism. The respective fixed point function is given by:*

$$f_\omega^*(F_j^*, F_i^*) = \omega_{i,j}^*(F_j^*, F_i^*) = \sqrt[n]{\frac{C(F_j^*, F_i^*) - \nu R(F_j^*, F_i^*)}{\gamma H(F_i^*)}}. \tag{4.33}$$

Proof. Equivalent to Eq. 4.11 (p. 54). □

4.3.1 Memory Representation and Functional Organization in Activity-space

In the following, we show that Correlation-and-Causality-based Synaptic Plasticity decouples the two processed of MR and the FO of interconnected memories. In particular, this learning rule enables the discrimination of memories while the formation of sequences and associations are still possible events. Also here, by means of the fixed point function on the excitatory synaptic weights (Eq. 4.33), we do the analysis on the reduced population model at equilibrium state introduced in Chapter 3.

For our analysis, we first make several assumptions on the here introduced causality-based plasticity mechanism $R(F_i, F_j)$. In addition to the measure C of pre- and post-synaptic correlation, the causality-based plasticity mechanism is introduced to describe the non-causality (i.e. independence) on the pre- and postsynaptic firing, as described above. Thus, we do not assume R being independent as the correlation-based plasticity mechanism C (C:i, Eq. 4.5, p. 52).

Moreover, when the pre- and postsynaptic neurons fire at the same activity level, we expect them to have an optimal configuration for the transmission of information. In other words, we assume interconnected neurons with the same activity level to be causally connected at a maximum value. In addition, based on the underlying pre- and postsynaptic requirements to maintain an information flow at one synapse, we suppose the causality-based plasticity mechanism to be symmetric.

Assumption 7 *At equilibrium state, we assume the causality-based plasticity mechanism R to be:*

- *minimal for fixed equal pre- and postsynaptic firing rates (R:0), as we assume that interconnected neurons with similar firing rates bring along the highest level of causality; Thus, it holds:*

$$R(x^*, x^*) = \min_{\forall x^*, y^* \in \mathbb{R}_+} R(x^*, y^*) =: R_{min}. \tag{4.34}$$

- *symmetric (R:s); For all $x^*, y^* \in \mathbb{R}_+$ it holds:*

$$R(x^*, y^*) = R(y^*, x^*). \tag{4.35}$$

Necessary Conditions on the Different Forms of Functional Organization

By means of the fixed point function f_ω^* (Eq. 4.33), we can express the synaptic plasticity dependent conditions for the MR of two interconnected memories and their FO equivalent to the analysis in Sec. 4.2.1 of the previous plasticity model.

Theorem 11 *Let θ be the constant inhibitory synaptic weight level within the network. The constraints on the MR for the interconnected neuronal population 1 and 2 and their different forms of FO can be combined into two ordered sets on the involved plasticity mechanisms C, R, H of the applied learning rule:*

$$\begin{bmatrix} \gamma H(\mathfrak{F}_1)\theta^n \begin{cases} \overset{(MR)}{<} C(\mathfrak{F}_1, \mathfrak{F}_1) - \nu R_{min} \\ \underset{exc}{\overset{inh}{\gtrless}} C(\mathfrak{F}_2, \mathfrak{F}_1) - \nu R(\mathfrak{F}_2, \mathfrak{F}_1) \end{cases} \\ \\ \gamma H(\mathfrak{F}_2)\theta^n \begin{cases} \overset{(MR)}{<} C(\mathfrak{F}_2, \mathfrak{F}_2) - \nu R_{min} \\ \underset{exc}{\overset{inh}{\gtrless}} C(\mathfrak{F}_1, \mathfrak{F}_2) - \nu R(\mathfrak{F}_1, \mathfrak{F}_2) \end{cases} \end{bmatrix} \mapsto FO = \{asc, seq\text{:}21, seq\text{:}12, disc\}. \tag{4.36}$$

Proof. Equivalent to (Eq. 4.14). $\qquad\qquad\qquad\square$

Theorem 12 *The necessary condition for the discrimination to occur is given by:*

$$\nu R(\mathfrak{F}_1, \mathfrak{F}_2) > \min\left(\hat{C}(\mathfrak{F}_1), \hat{C}(\mathfrak{F}_2)\right) \left|\hat{C}(\mathfrak{F}_1) - \hat{C}(\mathfrak{F}_2)\right| + R_{min}. \tag{4.37}$$

Proof. The two ordered sets of the plasticity mechanisms (Eq. 4.36) describing the *discrimination* (disc) of two interconnected memories is defined by the bidirectional inhibitory inter-memory relation of neuronal populations:

$$\begin{bmatrix} C(\mathfrak{F}_1, \mathfrak{F}_1) - \nu R_{\min} > \gamma H(\mathfrak{F}_1)\theta^n > C(\mathfrak{F}_2, \mathfrak{F}_1) - \nu R(\mathfrak{F}_2, \mathfrak{F}_1) \\ C(\mathfrak{F}_2, \mathfrak{F}_2) - \nu R_{\min} > \gamma H(\mathfrak{F}_2)\theta^n > C(\mathfrak{F}_1, \mathfrak{F}_2) - \nu R(\mathfrak{F}_1, \mathfrak{F}_2) \end{bmatrix} \mapsto disc.$$

When we neglect the condition on the postsynaptic activity-dependent scaling mechanism, the chain of relations of the plasticity mechanism describing the discrimination of two memories is reduced to the following necessary condition:

$$\begin{bmatrix} C(\mathfrak{F}_1, \mathfrak{F}_1) - \nu R_{\min} > C(\mathfrak{F}_2, \mathfrak{F}_1) - \nu R(\mathfrak{F}_2, \mathfrak{F}_1) \\ C(\mathfrak{F}_2, \mathfrak{F}_2) - \nu R_{\min} > C(\mathfrak{F}_1, \mathfrak{F}_2) - \nu R(\mathfrak{F}_1, \mathfrak{F}_2) \end{bmatrix} \mapsto disc.$$

Due to the symmetric property of the Correlation- and Causality-based plasticity mechanism (C:s, R:s) both sets of equations can be combined towards one set of equations

$$\begin{bmatrix} \nu R(\mathfrak{F}_1, \mathfrak{F}_2) > \begin{cases} C(\mathfrak{F}_2, \mathfrak{F}_1) - C(\mathfrak{F}_1, \mathfrak{F}_1) + \nu R_{\min} \\ C(\mathfrak{F}_1, \mathfrak{F}_2) - C(\mathfrak{F}_2, \mathfrak{F}_2) + \nu R_{\min} \end{cases} \end{bmatrix} \mapsto disc.$$

Here, due to the monotonous characteristic of the Hebbian learning mechanism C (4.3) and a case differentiation for $\mathfrak{F}_1 > \mathfrak{F}_2$ and $\mathfrak{F}_1 < \mathfrak{F}_2$ we can split this conditions into two distinct cases:

$$\nu R(\mathfrak{F}_1, \mathfrak{F}_2) > \begin{cases} C(\mathfrak{F}_2, \mathfrak{F}_1) - C(\mathfrak{F}_1, \mathfrak{F}_1) + \nu R_{\min}, & \text{for } \mathfrak{F}_1 < \mathfrak{F}_2 \\ C(\mathfrak{F}_1, \mathfrak{F}_2) - C(\mathfrak{F}_2, \mathfrak{F}_2) + \nu R_{\min}, & \text{for } \mathfrak{F}_1 > \mathfrak{F}_2 \end{cases}$$

$$\overset{(4.5, 4.4)}{=} \begin{cases} \hat{C}(\mathfrak{F}_1)\left(\hat{C}(\mathfrak{F}_2) - \hat{C}(\mathfrak{F}_1)\right) + \nu R_{\min}, & \text{for } \mathfrak{F}_1 > \mathfrak{F}_2 \\ \hat{C}(\mathfrak{F}_2)\left(\hat{C}(\mathfrak{F}_1) - \hat{C}(\mathfrak{F}_2)\right) + \nu R_{\min}, & \text{for } \mathfrak{F}_1 > \mathfrak{F}_2 \end{cases}$$

$$= \min\left(\hat{C}(\mathfrak{F}_1), \hat{C}(\mathfrak{F}_2)\right) \left|\hat{C}(\mathfrak{F}_1) - \hat{C}(\mathfrak{F}_2)\right| + \nu R_{\min}.$$

$\square$

This indicates that such a learning rule, considering the causality of pre- and postsynaptic firing in addition to its correlation, enables a decoupling of the two distinct processes of MR and their formation of different forms of FOs allowing the formation of discriminated memories.

4.3.2 Example: Learning Rule on Synaptic Plasticity and Synaptic Scaling Augmented by Causality

In this Section, we propose a specific learning rule that considers the introduced causality-based plasticity mechanism and, furthermore, fullfils the requirements derived in Theorem 12. This specific example is an extension of the already investigated *learning rule on Synaptic Plasticity and Synaptic Scaling* (SPaSS) by a causality based plasticity mechanism of the following form:

$$R : (\mathbb{R}_+, \mathbb{R}_+) \to \mathbb{R}_+, \qquad (F_j F_i) \mapsto \sqrt{F_j F_i} \Delta F + R_{\min}, \qquad R_{\min} \in \mathbb{R} \qquad (4.38)$$

with $R_{\min} = 0$ and $\Delta F := |F_j - F_i|$ describing the difference in pre- and postsynaptic firing.

Thus, we refer to the here proposed learning rule by *augmented learning rule on Synaptic Plasticity and Synaptic Scaling by Causality* (aSPaSS). This aSPaSS learning rule given by

$$\tau_\omega \dot{\omega}_{i,j} = F_j F_i - \nu \sqrt{F_j F_i} \Delta F - \gamma \left(F_i - \mathrm{F}^\mathrm{T} \right) \omega_{i,j}^2 \qquad (4.39)$$

consists of the prototype of Hebbian learning mechanism C (Eq. 4.6), a causality-based plasticity mechanism R (Eq. 4.38) proportionally dependent to the difference in pre- and postsynaptic firing, and a quadratic ($n = 2$) synaptic weight dependency for the homeostatic mechanism H. For the analysis of this specific example we use its respective normalized model (see Appendix for more detail):

$$\textbf{input} \qquad \tilde{\mathbf{F}}^\mathrm{ex} := \left(\langle \tilde{F}_1^\mathrm{ex} \rangle_t, \langle \tilde{F}_2^\mathrm{ex} \rangle_t, \langle \tilde{F}_\mathcal{B}^\mathrm{ex} \rangle_t \right)^\mathsf{T} \qquad (4.40)$$

$$\textbf{activity} \qquad \tau \dot{\tilde{F}}_i = (1 - \tilde{F}_i)\tilde{F}_i \left[\log \left(\tilde{F}_i^{-1} - 1 \right) + \beta(\mathrm{R}\phi_i - \epsilon) \right], \qquad (4.41)$$

$$\textbf{with} \qquad \phi_i = \mathrm{F}^\mathrm{max} \omega^\mathrm{max} \left(\sum_{r \in \{\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_\mathcal{B}\}} \sum_{j \in \mathcal{P}_r} (\tilde{\omega}_{i,j} - \tilde{\theta})\tilde{F}_j + \tilde{\omega}^\mathrm{ex} \sum_{k \in \mathcal{E}_r} \tilde{F}_k^\mathrm{ex} \right), \qquad (4.42)$$

$$\textbf{syn. weight} \qquad \tilde{\tau}_\omega \dot{\tilde{\omega}}_{i,j} = \tilde{F}_j \tilde{F}_i - \tilde{\nu} \sqrt{\tilde{F}_j \tilde{F}_i} |\tilde{F}_j - \tilde{F}_i| - \tilde{\gamma}(\tilde{F}_i - \tilde{\mathrm{F}}^\mathrm{T})\tilde{\omega}_{i,j}^2, \qquad (4.43)$$

$$\textbf{with} \qquad \tilde{\tau}_\omega = \frac{\tau_\omega}{\left(\mathrm{F}^\mathrm{max} \sqrt{\mathrm{F}^\mathrm{max} \gamma (1 - \tilde{\mathrm{F}}^\mathrm{T})} \right)}, \quad \tilde{\nu} = \frac{\nu}{\mathrm{F}^\mathrm{max2}}, \quad \tilde{\gamma} = \frac{1}{1 - \tilde{\mathrm{F}}^\mathrm{T}}. \qquad (4.44)$$

Activity-dependent Constraints of Memory Representation and Functional Organization

In the following, we map the general synaptic-weight dependent conditions for MR and FO of two interconnected memories towards the respective conditions for the specific nor-

malized aSPaSS-learning rule in the $\mathfrak{F}_1 - \mathfrak{F}_2-$activity space. Therefore, we set about to calculate its specific normalized fixed point function:

$$f_\omega^* = \omega_{i,j}^* = \sqrt{\frac{C(\tilde{F}_j^*, \tilde{F}_i^*) - \tilde{\nu}R(\tilde{F}_j^*, \tilde{F}_i^*)}{\tilde{\gamma}H(\tilde{F}_i^*)}} = \sqrt{\frac{(\tilde{F}_j^*\tilde{F}_i^* - \tilde{\nu}\sqrt{\tilde{F}_j^*\tilde{F}_i^*}\Delta\tilde{F})(1 - \tilde{F}^\mathsf{T})}{\tilde{F}_i^* - \tilde{F}^\mathsf{T}}}. \quad (4.45)$$

As already derived, the *long-term representation* (LTR) of an environmental input stimulus $\mathfrak{I}_r$ is decoded by the average activity of the respective neuronal population ($\mathfrak{w}_{rr} = f_\omega^*(\mathfrak{F}_r, \mathfrak{F}_r)$) at equilibrium state. Therefore, the causality-based plasticity mechanism R vanishes for the condition on MR of a neuronal population r, as it holds $\Delta F = 0$ and $R_{\min} = 0$ for the applied learning rule. Thus, for the specific learning rule on aSPaSS (Eq. 4.39) the same $\mathfrak{F}_r$-activity dependent condition for MR holds as in the previous analyzed SPaSS-learning rule (Eq. 4.23, Fig. 4.1 Ai).

As derived in the previous Section, the causality-based plasticity mechanism of the specific aSPaSS-learning rule starts to make an impact for distinct activity levels of both neuronal populations $\mathfrak{F}_1 \neq \mathfrak{F}_2$ by construction. By this, the corresponding fixed point function for the excitatory synaptic weights (Eq. 4.45) can not be expressed by a closed form for either the fixed presynaptic $\tilde{F}_j^*$ nor the fixed postsynaptic firing rate $\tilde{F}_i^*$. In particular, we can not provide a closed form for the separatrixes $S_{21}(\mathfrak{F}_2)$, $S_{12}(\mathfrak{F}_1)$ dividing an excitatory from an inhibitory relation (R21, R12) of both populations within the $\mathfrak{F}_1 - \mathfrak{F}_2$-activity space, determining the characteristic activity phase space (χAPS). Thus, we can not express the synaptic weight-dependent conditions on different forms of FO of two interconnected memories 1 and 2 by conditions on their respective activities $\mathfrak{F}_1$ and $\mathfrak{F}_2$, as we did for the SPaSS-learning rule. Instead, we can do the analysis based on the respective synaptic weight dependent-conditions, explained in the following.

Functional Organization in Activity-space

Even though we can not specify the activity-dependent conditions for the different forms of FO of two interconnected neuronal population in the $\mathfrak{F}_1 - \mathfrak{F}_2-$activity-space analytically, we can do the analysis based on the respective synaptic weight dependent-conditions. Therefore, for each possible $\mathfrak{F} = (\mathfrak{F}_1, \mathfrak{F}_2)^\mathsf{T}$-activity tuple of the population's activities at equilibrium state, we can calculate the respective fixed average synaptic weights of the population model by means of the fixed point function for the excitatory synaptic weights $\mathfrak{w}_{r,r} = f_\omega^*(\mathfrak{F}_r, \mathfrak{F}_r)$ and $\mathfrak{w}_{s,r} = f_\omega^*(\mathfrak{F}_r, \mathfrak{F}_s)$ (Fig. 4.4 A). Finally, we classify these synaptic weights ($\mathfrak{w}_{\text{intra}}$, $\mathfrak{w}_{\text{inter}}$) in the population formalism at equilibrium according to the synaptic weight-dependent constraints on MR and FO of two interconnected populations
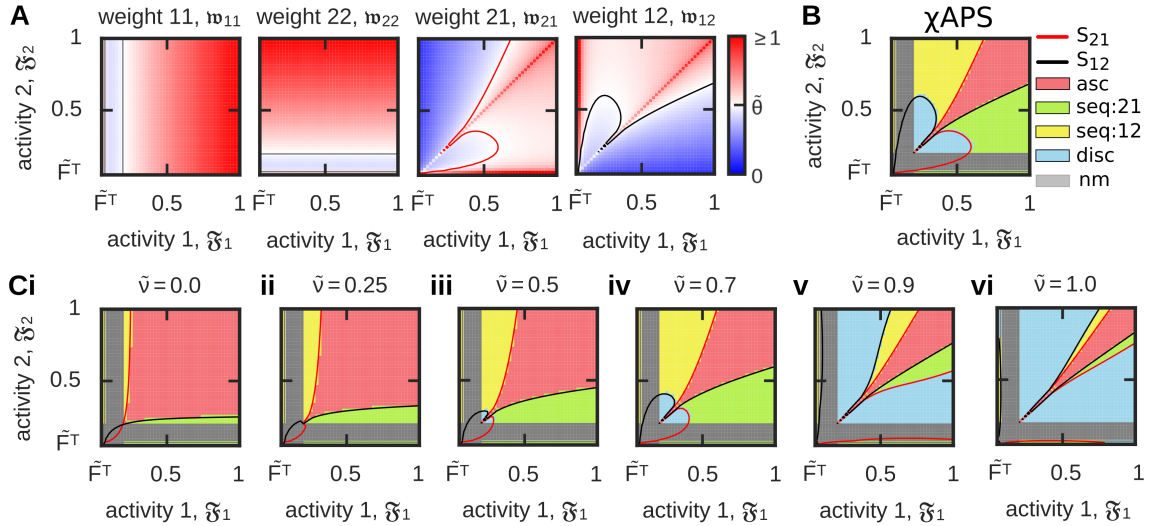
Figure 4.4: Analysis of the specific aSPaSS-learning rule on its abilities to describe all different forms of FO of two interconnected memories. (A) Calculation of the respective synaptic weights at equilibrium in the $\mathfrak{F}_1 - \mathfrak{F}_2-$activity space. Used parameters: $\tilde{F}^T = 0.05$, $\tilde{\theta} = 0.5$, $\tilde{\nu} = 0.8$. **(B)** Classification of the synaptic weights at equilibrium obtained in (A) according to the synaptic weight-dependent constraints on MR and FO yielding the characteristic activity space. **(C)** Characteristic activity phase spaces for different time scales $\tilde{\nu}$ on the causality-based plasticity mechanism.

(Fig. 3.5, p. 46). Hence, we obtain the $\mathfrak{F}_1 - \mathfrak{F}_2-$characteristic activity phase space (χAPS) (Fig. 4.4 B). Thus, we show that this specific aSPaSS-learning rule reveals the formation of all different forms of FO of two interconnected memories (Fig. 4.4 B). Note that for $\tilde{\nu} = 0$, the learning rule describes the former SPaSS-learning rule, which only exhibits association and sequences of memories (compare to Fig. 4.4 Ci, $\tilde{\nu} = 0$). With an increasing time-scale $\tilde{\nu}$ for the causality-based plasticity mechanism, the activity space leading to a discrimination of two interconnected memories (blue phase space) emerges and increases at the border of the *no-memory state* (nm)-state (grey phase space) in the $\mathfrak{F}^{high} - \mathfrak{F}^{high}-$activity space (Fig. 4.4 C).

4.3.3 Long-term Depression is Mandatory for the Functional Organization of Memories

In this Section, we have augmented the previously discussed synaptic learning rule describing LTP, by an additional plasticity mechanism describing LTD, based on a measure for causality of pre- and postsynaptic firing.

Besides the mathematical motivation for such an additional plasticity mechanism, outlined at the beginning of this Section, there is a multitude of studies indicating the ex-

istance of those additional factors influencing synaptic plasticity. For instance, neuro-modulatory transmitters, such as acetylcholine, noradrenaline, serotonin, and dopamine, can serves as third factor (Frémaux and Gerstner 2016; Gu 2002). The here presented mechanism may abstract those molecular mechanisms of such additional factors yielding, together with correlation-based and homeostatic plasticity, all different forms of FOs of interconnected memories. For instance, the mathematical implementation of spike-timing-dependent triggered LTD (Bi and Poo 1998; Rossum et al. 2000); in contrast to the here presented firing rate-dependent LTD (Bienenstock et al. 1982; Sjöström et al. 2001; Malenka and Bear 2004)) could comprehensively describe uncorrelated spike trains and, by this, decouple the condition for MR the discrimination condition. Thus, the STDP (Gerstner et al. 1996; Bi and Poo 1998; Markram et al. 2011) mechanism leading to LTP could be a measure for the probability that the pre- and postsynaptic neurons fire correlated spikes during a small time window (Dayan and Abbott 2001) described in the rate-model by the correlation-based plasticity mechanism. Whereas LTD triggered by uncorrelated spike pairs could be described in the here introduced causality-base plasticity mechanisms in the rate-model by the difference between the pre- and postsynaptic firing rates.

With these assumptions on the causality-based plasticity mechanism, we conclude that the difference in pre- and postsynaptic firing is a sufficient measure for the non-causality of neural firing rates. Furthermore, we have validated this basic principle for an exemplary learning rule. This example reliably decouples the formation of MR and FO of interconnected neuronal populations and enables the formation of all different forms of FOs, such as association, sequence and discrimination.

4.4 Local Characteristics for Different Plasticity Mechanisms

For generality, in our previous investigations on the ability of adaptive neuronal systems to form all different types of FO of interconnected memories, we considered a homogeneous recurrent neuronal network. In particular, we applied a globally constant parameter set for all individual neurons and synapses of the whole neuronal network. In the upcoming Section, we allow local differences for the individual parameter sets of the underlying neuronal and synaptic dynamics of the network.

The discrimination of two interconnected memories $r, s \in \{1, 2\}$ turned out to be the key problem of correlation-based synaptic plasticity mechanism $C(F_j, F_i)$ in combination with a synaptic weight and postsynaptic activity-dependent homeostatic mechanism $H(F_i)\omega_{i,j}^n$. Without any local differences in the homeostatic mechanism and inhibitory synaptic weight strength θ for all synapses of the network, we have seen in Theorem 6 (p. 55) that the constraints on MR and FO can be expressed by an estimation of the involved plasticity mechanism on the postsynaptic activity-dependent homoestatic mechanism H and global inhibitory synaptic weight strength θ. Thus, the underlying reason for this shortcoming of discriminated MRs is the strong coupling of the constraints for the conditions of MR and discrimination of interconnected memories. This conditions are described by the following, already derived (Corollary 9, p. 58), chain of relations:

$$
\begin{bmatrix}
C(\mathfrak{F}_1, \mathfrak{F}_1) \overset{\text{MR}}{>} \gamma H(\mathfrak{F}_1)\theta^n \overset{\text{R12: inh}}{>} C(\mathfrak{F}_2, \mathfrak{F}_1) \\
C(\mathfrak{F}_2, \mathfrak{F}_2) \overset{\text{MR}}{>} \gamma H(\mathfrak{F}_2)\theta^n \overset{\text{R21: inh}}{>} C(\mathfrak{F}_1, \mathfrak{F}_2)
\end{bmatrix} \mapsto \text{disc}
\tag{4.46}
$$

in combination with the monotonous characteristic of the correlation-based plasticity mechanism (C:m, Eq. 4.3, p. 52). Thus, in the following, we focus on strategies that decouple these chain of relations, and, hence, enable such adaptive systems to build up discriminated memory representations.

4.4.1 Local Characteristics of the Homeostatic Mechanism

In order to decouple the chain of relations for the discrimination of two interconnected memories (Eq. 4.46), we introduce local differences for the inter- and intra-neuronal population's dynamics of the postsynaptic activity-dependent homeostatic mechanism H in two different ways. First, we introduce a generic solution on the level of interacting plasticity mechanisms of correlation-based plasticity in combination with a homeostatic mechanism. Then, secondly, a SPaSS-learning rule specific solution is included.

Enhanced Correlation-based Plasticity within Neuronal Populations

To facilitate the discrimination of interconnected memories, here we introduce a damping constant (η) for the homeostatic mechanism only for the synapses within the neuronal populations $r \in \{1, 2\}$ (Fig. 4.5 A):

$$\tau_\omega \dot{\omega}_{i,j} = F_j F_i + \eta \gamma H(F_i) \omega_{i,j}^n, \qquad \forall i,j \in \mathcal{P}_{p_r}, \qquad \eta \in (0,1) \in \mathbb{R}. \tag{4.47}$$

This approach has the advantage that synapses within one neuronal population underlie a stronger influence of the correlation-based plasticity mechanism compared to the inter-population synapses connecting both neuronal populations. Such an assumption can be based on the homogeneous input stimulation onto all neurons of a network resulting in a causally correlation of the respective pre- and postsynaptic neurons.

At equilibrium, this approach decouples the previous coupled chain of relations for the plasticity mechanisms, describing a discrimination of interconnected memories:

$$\begin{bmatrix} C(\mathfrak{F}_1, \mathfrak{F}_1) \overset{(\text{MR})}{>} \eta\gamma H(\mathfrak{F}_1)\theta^n, \ \gamma H(\mathfrak{F}_1)\theta^n \overset{(\text{R12})}{>} C(\mathfrak{F}_2, \mathfrak{F}_1) \\ C(\mathfrak{F}_2, \mathfrak{F}_2) \overset{(\text{MR})}{>} \eta\gamma H(\mathfrak{F}_2)\theta^n, \ \gamma H(\mathfrak{F}_2)\theta^n \overset{(\text{R21})}{>} C(\mathfrak{F}_1, \mathfrak{F}_2) \end{bmatrix} \mapsto \text{disc.} \tag{4.48}$$

While the relations R12 and R21 of both neuronal populations are in an inhibitory state for both directions (Fig. 4.5 Ai, blue phase space), the average fixed synapses within the neuronal populations can already increase the population specific constraints on MRs (Fig. 4.5 Ai, decoupled grey phase space from blue phase space) and, by this, yield an activity phase space for discriminated memory representations. This corresponds to low environmental $\mathfrak{I}_1 - \mathfrak{I}_2-$input stimulations that lead to discriminated memories (Fig. 4.5 Bi, blue phase space). Furthermore, the characteristic $\mathfrak{F}_1 - \mathfrak{F}_2-$activity phase space ($\chi$APS), describing the system's ability to form the different types of FOs of two interconnected memories, follows the usual pattern as obtained for a globally homogeneous parameter set (compare with Fig. 4.2 D, p. 65). Here, due to the facilitation of MR, the input phase space leading to sequences is enlarged along with a shrinked $\mathfrak{I}_1 - \mathfrak{I}_2-$input phase space leading to the nm-state of the system.

Distinct Neuronal Target Activities

Given the SPaSS-learning rule, we consider local differences for the neuronal target activity level F^T (the synaptic scaling dependent parameter). Here, we assume a binary, distance-dependent target firing rate for the scaling mechanism (Fig. 4.5 B). Therefore,
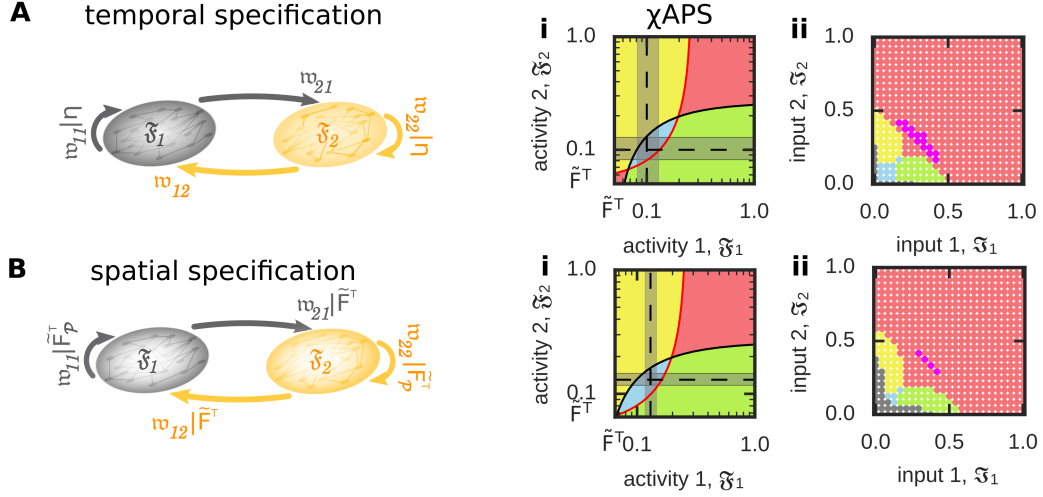
Figure 4.5: Different modeling approaches to ensure all possible forms of functional organizations of two interconnected populations. (A) Temporal specification, introducing a damping constant ζ for the homeostatic mechanisms of the learning rule for all synapses within a neuronal population leading to a stronger impact of the Hebbian correlation-based plasticity mechanism of the learning rule. **(B)** Spatial specification, introducing a higher target activity level for all synapses within one neuronal population ($\tilde{F}_{\mathcal{P}}^{\mathrm{T}}$) as compared to the target activity level of synapses connected to other neuronal population $\tilde{F}^{\mathrm{T}}$. Used parameters: (A) $\eta = 0.8$, (B) $\tilde{F}_{\mathcal{P}}^{\mathrm{T}} = 0.065$, (A & B) $\tilde{F}^{\mathrm{T}} = 0.05$, $\tilde{\theta} = 0.5$, $n_{\epsilon} = 12$.

for synapses building connections to neurons that are closely located to the presynaptic neuron (i.e. neurons that belong to one neuronal population), we assume them to have a higher target firing rate ($F_{\mathcal{P}}^{\mathrm{T}}$) than compared to far off neurons (i.e. neurons belonging to different populations, F^{T}). Thus, the learning rule for all neurons i, j belonging to the same neuronal population r changes according to:

$$\tau_{\omega}\dot{\omega}_{i,j} = F_i F_j - \gamma(F_i - F_{\mathcal{P}}^{\mathrm{T}})\omega_{i,j}^2, \qquad \forall i, j \in \mathcal{P}_{p_r}, \tag{4.49}$$

with $F_{\mathcal{P}}^{\mathrm{T}}$ denoting the target firing rate for all neurons belonging to a neuronal population $r \in \{1, 2\}$. Whereby, the learning rule for synapses connecting neurons of distinct populations still follows the former formulation of

$$\tau_{\omega}\dot{\omega}_{i,j} = F_i F_j - \gamma(F_i - F^{\mathrm{T}})\omega_{i,j}^2, \qquad \forall i \in \mathcal{P}_{p_r} \text{ and } \forall j \in \mathcal{P}_{p_s}. \tag{4.50}$$

Thus, when we assume a higher target firing rate for the synaptic scaling mechanism for all synapses within one neuronal population compared for synapses in between both neuronal populations ($\tilde{F}^{\mathrm{T}} < \tilde{F}_{\mathcal{P}}^{\mathrm{T}}$), the system is able to build up MRs of interconnected populations while the synaptic weights linking the populations in both directions still

may build up inhibitory relations. Hence, enabling the FO of discrimination as shown in Fig. 4.5 Bi, where the blue and grey activity phases are decoupled.

This approach, again, facilitates the Hebbian learning term at equilibrium for synapses within the neuronal populations compared the the synapses connecting both populations. Thus, the formation of a discriminatory FO in between two competing memories is enabled.

4.4.2 Local Inhibitory Synaptic Plasticity

Heretofore, we have only considered excitatory synaptic plasticity while the inhibition of the network has been set to a constant value of θ. However, the inhibitory synaptic plasticity (Castillo et al. 2011) is experimentally observed and described. Thus, in this Section, we introduce an additional synaptic learning rule acting on the weights of inhibitory synapses $\omega_{j,i}^-$.

This approach and the corresponding results have been published in the following article:

J. Herpich and C. Tetzlaff (2018). "Principles Underlying the Input-Dependent Formation and Organization of Memories". In: *bioRxiv*. A similar manuscript is currently under revision in Network Neuroscience.

Similar to excitatory synaptic plasticity, here we introduce a correlation-based inhibitory plasticity mechanism depending on the correlation of pre- and postsynaptic firing. Moreover, we augment the inhibitory synaptic plasticity rule by two additional constrains. The first constraint is given by a threshold in the pre- and postsynaptic firing rates, defining a minimum activity level θ_F, for inhibitory synaptic plasticity to occur

$$\dot{\omega}_{j,i}^- \propto F_i F_j \mathsf{H}(\textstyle\sum F - \theta_F), \quad \textstyle\sum F := F_j + F_i, \tag{4.51}$$

with $\omega_{j,i}^-$ being the strength of the inhibitory synapse connecting the presynaptic neuron i with the postsynaptic neuron j. $\mathsf{H}(x)$ ($\mathsf{H}(x \geq 0) = 1$, $\mathsf{H}(x < 0) = 0$) is the heaviside step function to monitor the here introduced constraint on the sum of pre- and postsynaptic activity levels ($\sum F$) exceeding the defined threshold level (θ_F). The second constraint for inhibitory synaptic plasticity to occur depends on the causality of pre- and postsynaptic firing rates:

$$\dot{\omega}_{j,i}^- \propto F_i F_j \mathsf{H}(\delta F - \Delta F), \quad \Delta F := |F_i - F_j|. \tag{4.52}$$

Here, the underlying principle is the same as for the excitatory synaptic plasticity in the augmented learning rule being proportional to the difference in pre- and postsynaptic

firing (ΔF, Sec. 4.3.2, p. 78). To allow a certain tolerance range for such a variation in the pre- and postsynaptic firing rates, we further introduce a threshold δF on the difference in pre- and postsynaptic firing. This is monitored by the heaviside step function $H(x)$. We combine these constraints Eq. (4.51) and Eq. (4.52) in a specific way in order to drive the inhibitory synaptic weights to converge either to

- an up-state (θ_u), if the sum of pre- and postsynaptic firing stays below its threshold ($\sum F < \theta_F$) **and/or** the difference in pre- and postsynaptic neuronal activities is larger than the respective tolerance range ($\Delta F > \delta F$), or

- a down-state (θ_d), if the sum of pre- and postsynaptic firing is larger than its threshold ($\sum F > \theta_F$) **and** the difference in the pre- and postsynaptic neuronal activities stays within its tolerance range ($\Delta F < \delta F$)

at equilibrium state. These constraints lead to the following learning rule for inhibitory synaptic plasticity:

$$\dot{\omega}_{j,i}^- = \rho F_i F_j ((\theta_u - \omega_{j,i}) H'[H(\Delta F - \delta F) + H(\theta_F - \sum F)] \qquad (4.53)$$
$$+ (\theta_d - \omega_{j,i}) H(\delta F - \Delta F) H(\sum F - \theta_F)),$$

with ρ being the learning rate for the inhibitory synaptic plasticity and H' an adjusted heaviside function ($H'(0) = 0$) to account for the and/or condition for the constraints driving the inhibitory synaptic weights towards the up-state (Herpich and Tetzlaff 2018).

An exemplary network simulation for this type of inhibitory synaptic plasticity together with plastic excitatory synapses governed by the already described SPaSS-learning rule is shown in Fig. 4.6 with a specific input stimulation of $\tilde{\mathbf{F}}^{ex} = (0.85, 0.7, 0.05)^\mathsf{T}$ of the normalized network model (see Appendix). For the stimulation phase ($t > 10\tilde{\tau}_\omega$), all neurons of the neuronal populations exhibit a sufficient strong activity level of $\sum F > \theta_F = 20\,\text{Hz} = 0.2\,F^{max}$ and the variance of the activities of all neurons within one population is smaller than the specific tolerance range of $\delta F = 5\,\text{Hz} = 0.05\,F^{max}$ (Fig. 4.6 B). Thus, the inhibitory synaptic weight level for all synapses within one population converge to the down-state ($\theta_d = 0.5\,\omega^{max}$, Fig. 4.6 C, red curves). Whereas for inter-population synapses the difference in pre- and postsynaptic firing increases the tolarance range δF. Hence, the respective inter-population inhibitory synaptic weights converge to the upstate $\theta_u = 0.8\,\omega^{max}$ (Fig. 4.6 D, red curves). Therefore, this specific example shows the dynamic formation of two discriminated memories. The synaptic weights within each population are stronger than the inhibitory synaptic strength (Fig. 4.6 C, black and yellow curves in relation to the red curves) and the synaptic weights connecting both populations stay below the inhibitory synaptic strength (Fig. 4.6 D, black and yellow curves in relation to the red curves).
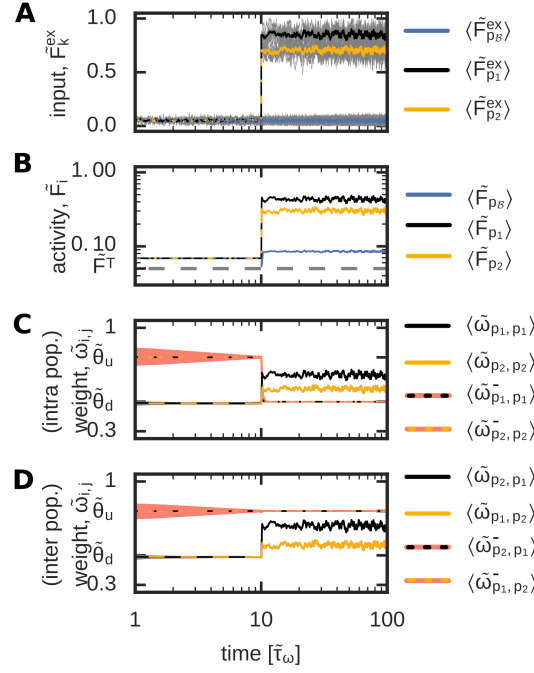
Figure 4.6: An exemplary inhibitory plasticity rule enables the dynamic formation of discriminated memories. The development of the input-driven dynamics of the complete neural network underlying SPaSS-learning rule for the excitatory synaptic weight in combination with plastic inhibitory synaptic weights (red lines). Stimulation protocol same as in Fig. 3.2 B for $\tilde{\mathbf{F}}^{\text{ex}} = (0.85, 0.7, 0.05)^{\mathsf{T}}$. Used parameters: $\rho = 1$, $\delta\tilde{F} = 0.05$, $\tilde{\theta}_F = 0.2$, $\tilde{\theta}_d = 0.5$, $\tilde{\theta}_u = 0.8$ (adapted from Herpich and Tetzlaff 2018).

Functional Organization in Activity-space

For our analysis of the system's abilities to build up different FOs with plastic excitatory and inhibitory synaptic weights, we apply the SPaSS-learning rule and assume two distinct inhibitory weight levels θ_u and θ_d at equilibrium. To decouple the condition on MR and FO, the inhibition within the neuronal population has to converge to the downstate (θ_d) while the inhibition for the inter-population synaptic weight has to converge to the upstate (θ_u) to spawn a discriminatory FO of two interconnected memories. For those cases, we introduce an inhibitory synaptic weight strength $(\theta_{\mathcal{P}} = \theta_d)$ within the neuronal populations. This synaptic weight strength is different to the inhibitory synaptic weight strength for all other connections $(\theta = \theta_u$, Fig. 4.7 A). In other words, the parameter θ is different for the discrimination condition as for the memory condition (which is now $\theta_{\mathcal{P}}$) (Herpich and Tetzlaff 2018).

To quantify the system's potential to form two MRs that are functionally discriminated, we calculate the size of the characteristic activity phase space (A^{disc}) leading to discriminated
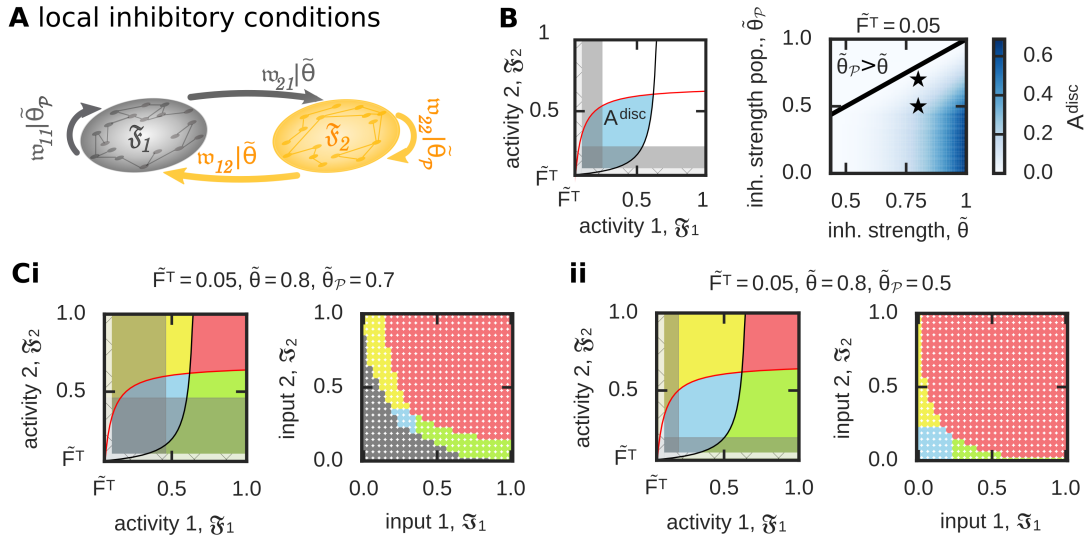
Figure 4.7: Considering different levels of inhibition level for connections within the neuronal populations compared to all others enables the formation of two discriminated memory representations. **(A)** We consider a different average inhibitory synaptic weight within the neuronal populations ($\tilde{\theta}_p$) compared to all others ($\tilde{\theta}$). **(B)** Left: To quantify the effect of different inhibition levels, we calculate the area of discrimination states (A^{disc}; blue) not being "covered" by the no-memory states (grey) in the $\mathfrak{F}_1$-$\mathfrak{F}_2$-activity space. Right: A^{disc} dependency on different relations between $\tilde{\theta}_{\mathcal{P}}$ and $\tilde{\theta}$. **(C)** Given a lower level of inhibition within the populations than otherwise provides the neural system the ability to form all FOs as indicated here by two examples (asterisked in (B)). (i) $\tilde{\theta} = 0.8$, $\tilde{\theta}_{\mathcal{P}} = 0.7$; (ii) $\tilde{\theta} = 0.8$, $\tilde{\theta}_{\mathcal{P}} = 0.5$ (adapted from Herpich and Tetzlaff 2018).

MRs (Fig. 4.7 B, left, occuring blue phase space). Apparently, as long as the inhibition within the populations takes on higher values than for the inter-population connections ($\theta_{\mathcal{P}} > \theta$, Fig. 4.7 B; right, parameter space above the solid black line), the system is not able to build up discriminated MRs. When, the inhibition within the populations takes on lower values than for the inter-population connections ($\theta_{\mathcal{P}} > \theta$Fig. 4.7 B; right, parameter space below the solid), the discrimination of two memories becomes possible within the $\mathfrak{F}_1 - \mathfrak{F}_2-$activity space. In addition to the discrimination of interconnected populations, all remaining forms of FOs, such as association and sequences, are maintained. Thus, all different types of FOs of interconnected populations can be obtained in the characteristic activity phase space. Two examples of characteristic activity phase space (χAPS) for different parameter sets are shown in Fig. 4.7 C (Herpich and Tetzlaff 2018).

Although, we predefined different levels of inhibition to enable an analysis of the population model at equilibrium, the self-organized neuronal network model can dynamically establish these different levels of inhibition at equilibrium by considering inhibitory synaptic plasticity (Eq. 4.53), as exemplary shown in Fig. 4.6 (Herpich and Tetzlaff 2018).

4.4.3 Local Characteristics for Plasticity Enable the Functional Organization of Memories

In general, our analysis indicates that a self-organized neural network can form all types of FOs for interconnected neuronal populations if the interaction of excitatory correlation-based plasticity and homeostasis are complemented by adaptive local constraints.

First, we have shown that local characteristics for the excitatory synaptic learning rule, in particular local constraints on the homeostatic mechanisms, enable the discrimination between memory representations while all other FOs of interconnected MRs are still enabled. The introduced constraints on the local characteristics are restricted to those neurons that receive specific environmental input stimulation. Since the collective stimulation of all neurons belonging to a population implies a higher level of causally correlated neural firing, such local mechanisms could serve as a measure for causally correlated pre- and postsynaptic firing of single neurons. Such local mechanisms serve in a way to facilitate the recurrency within those collectively stimulated neurons. This issue was already studied in the previous Section (Sec. 4.3).

Second, by introducing local variations in the level of inhibition, we have shown that these variations serve as a factor to enable a dynamic neuronal network to exhibit the whole range in functionally connected MRs. This is possible given the condition that the average inhibitory synaptic strength within a neuronal population has to be weaker than all other inhibitory synaptic weights of the remaining network. This is in contrast to the general understandig on the mode of operation for inhibition, which is considered to balance strong excitation within interconnected groups of neurons (Vogels et al. 2011; Litwin-Kumar and Doiron 2014). Nevertheless, such local differences in the balance of inhibition and excitation vanish when we broaden the view on the whole neuronal network and its whole balance of excitation and inhibition (Vreeswijk and Sompolinsky 1998; Denève and Machens 2016).

5

Discussion, Conclusions and Prospects

5.1 Schema and Synaptic Plasticity for Cognitive Strategies

Human behavior arises due to the brain's ability to exhibit a multitude of specific cognitive structures. For a review and comparison on the different cognitive structures see Ghosh and Gilboa (2014). Those structures are, for example, *narratives* (Mar 2004), *scripts* (Schank and Abelson 1977), *plans* (Schank and Abelson 1977), *concepts* and *categories* (Collins and Quillian 1969; Patterson et al. 2007), *event gists* (Thorndyke 1977; Nadel et al. 2000; Winocur and Moscovitch 2011), *statistical regularities* (Durrant et al. 2011; Posner and Keele 1968; Tobia et al. 2012) and *schemata* (Head and Holmes 1911; Piaget 1926; Bartlett 1932). These cognitive structures are differentiated due to their inherent features. Hereby, the concept of schema allows for cognitive structures and combines all features attributed to the specific knowledge structures listed above (Ghosh and Gilboa 2014). Thus, in this thesis, we focused on the concept of schemata. Evidently, a schema describes the brain's variety and diversity of ways to store knowledge and exhibit complex behavior in a generic way. The concept of schema was psychologically proposed (Piaget 1926; Bartlett 1932) and experimentally underpinned by biological and behavioral experiments (Tse et al. 2007; Tse et al. 2011; Wang et al. 2012; Preston and Eichenbaum 2013). Inherent properties of schemata include the behaviorally most important feature of elaborating context-specific behavior (Head and Holmes 1911; Bartlett 1932; Rumelhart 1980; Cooper et al. 1995; Shea et al. 2008; Kumaran et al. 2009) and the encoding and retrieval of memory (Head and Holmes 1911; Piaget 1926; Bartlett 1932; Carmichael et al. 1932; Bransford and Johnson 1972; Anderson and Pichert 1978; Rumelhart 1980; Anderson 1984; Arkes and Freedman 1984; Tse et al. 2007; Preston and Eichenbaum 2013; Kesteren et al. 2013). The *synaptic-plasticity-and-memory* (SPM) hypothesis (Martin et al. 2000; Hebb 1949; Dayan and Abbott 2001; Eichenbaum 2012) relates the psychological concept of memory to the neuronal correlate of synaptic plasticity, affecting the coupling strength of individual neurons. Thus, in this thesis, we evaluate the hypothesis that *complex cognitive strategies, dependent on the interaction and connection of numerous memories, are generated and performed based on the dynamics of synaptic plasticity*.

5.2 The Primary Building Blocks of a Schema

In order to evaluate the hypothesis, we restricted our investigation on a network of memories defining the structure of a schema. Since two related memories constitute the *primary building block* of a schema, heretofore, we studied the self-organization of two anatomically connected memories, memory 1 and memory 2 (indicated by $1 \sim 2$). By this, we distinguish different types for the *functional organization* (FO) of these primary building blocks, based on the ability of one memory to either excite or inhibit the interconnected memory (Byrne and Huyck 2010): They can either be functionally organized as an association ($1 \Leftrightarrow 2$, mutual excitation), sequence ($1 \Rightarrow 2$, sequential excitation) or discrimination ($1 - 2$, mutual inhibition; see Def. 1, p. 28). Studies by Preston and Eichenbaum (2013) showed that already two overlapping primary building blocks ($1 \sim 2$ and $2 \sim 3$), that have one memory in common (i.e., memory 2), can describe simple behavior such as the experimentally observed *transitive-inference paradigm*. In the study by Preston and Eichenbaum (2013), rats were trained to associate the memories for each primary building block: $1 \Leftrightarrow 2$ and $2 \Leftrightarrow 3$. During interference trials, they were successfully tested on their ability to infer the relationship between memory 1 and memory 3 ($1 \Leftrightarrow 3$). Thus, the simplified behavioral model of a schema at minuscule scale composed out of two overlapping primary building blocks ($1 \sim 2$ and $2 \sim 3$) can explain cognitive functions (Preston and Eichenbaum 2013; Ghosh and Gilboa 2014). These experimental results supports our approach to analyze two interconnected memories and their ability to form different types of FOs.

5.2.1 Synaptic Learning Rules Leading to Stable Dynamics Enable a Comprehensive Analysis of the Population Model

The different types of FOs have alrady been investigated by other theoretical studies (Griniasty et al. 1993; Abbott and Blum 1996; Leibold and Kempter 2006; Herrera-Aguilar et al. 2012; Tully et al. 2016; Chenkov et al. 2017; Litwin-Kumar and Doiron 2014; Byrne and Huyck 2010). However, they differ in their research questions compared to the one, analyzed in this thesis. For instance, studies by (Litwin-Kumar and Doiron 2014) investigate the association of memories whilst studies by (Chen 2017) consider the discrimination of memories. Further studies by (Griniasty et al. 1993; Abbott and Blum 1996; Leibold and Kempter 2006; Herrera-Aguilar et al. 2012; Tully et al. 2016; Chenkov et al. 2017) focus on the formation of sequences between memories. These studies often aim to generate temporally extended activity patterns, as needed for motor sequences (Abbott and Blum 1996) by the consecutive replay of memories.

All these different studies highlight the functional significance of the respective organization of interconnected memories but use or propose different modeling approaches, which typically give rise to the investigated FO. On the other hand, the work by Byrne and Huyck (2010) already demonstrates how a network with all of these FOs of interconnected memories can execute computational algorithms. However, they construct the network by predefined and suitably connected memories, thereby, neglecting their self-organization due to environmental input stimulation. Therefore, we here ask how one neuronal network with a given synaptic plasticity rule can at the same time explain the input-dependent formation of *memory representation*s (MRs) and give rise to all these different forms of FOs.

To investigate the self-organization of two interconnected memories, we developed a theoretical framework that determines the equilibrium state of two interconnected neuronal populations for given input stimulation (Chapter 3). This approach reduces the dimension of the adaptive neuronal network model (Sec. 3.1; Fig. 3.2, p. 33) to a low dimensional *population model at equilibrium* (Sec. 3.3; Fig. 3.3, p. 42) that enables analytic investigations on the input-dependent self-organization of two interconnected memories. Remarkably, the derived reduced model in this thesis matches the learning results of the corresponding full network simulation (Sec. 3.4.3;Fig. 3.6, p. 49). Using the reduced population model at equilibrium state, we mapped the definitions for the different types of FO of two interconnected memories on their ability to either excite or inhibit the interconnected memory to a synaptic weight-dependent definition for the population model (see Sec. 3.4, p. 45).

In Chapter 4, we analyzed the effect of synaptic plasticity mechanisms that determine on the self-organization of two interconnected memories into their FO. For this, we consider a general mathematical formulation for the different synaptic plasticity mechanisms. We show that synaptic learning rules that lead to stable synaptic weight dynamics, i.e., providing a fixed point function for the synaptic weights, enable a comprehensive analysis of the FOs. Using this fixed point function, the synaptic weight-based constraints for the different types of FO are mapped to the low-dimensional $\mathfrak{F}_1 - \mathfrak{F}_2-$activity space of the population model ($\mathfrak{F}_1$: average fixed activity of population 1; $\mathfrak{F}_2$: average fixed activity of population 2). We, refer to the corresponding classification of the $\mathfrak{F}_1 - \mathfrak{F}_2-$activity space in regard to the different types of FOs of two interconnected memories by the term characteristic activity phase space (χAPS).

In the following, we discuss the different activity-based synaptic plasticity mechanisms of rate-coded neurons we have analyzed in Chapter 4. Note, for each synaptic learning rule on their biological interpretation is already briefly discussed at the end of each corresponding Section in Chapter 4.

5.2.2 Correlation-based Plasticity Accounts For Associative Memory Structures

In Chapter 4, we set about analyzing synaptic learning rules that are based on Hebbian correlation-based learning in combination with a homeostatic mechanism on their ability to form diverse FOs of two interconnected memories. This class of learning rules comprises for example the *learning rule on Synaptic Plasticity and Synaptic Scaling* (SPaSS). It has been shown that the SPaSS-learning rule can account for formation of individual unconnected memories in neuronal networks (Tetzlaff et al. 2011; Tetzlaff et al. 2013; Auth et al. 2018). In this thesis, we extended these results by showing that the generic formulation of correlation-based Hebbian learning in combination with postsynaptic acticity-dependent homeostasis enables the formation of association and sequences of two memories (Fig. 4.3, p. 69). However, we also were able to show that the condition of MR is coupled with the condition for the different forms of FO via the plasticity mechanisms involved. Our analysis revealed that this strong interdependence on the plasticity mechanisms excludes the formation of a discrimination of interconnected memories. Furthermore, we have shown that sequences of memories only occur due to effects of the homeostatic plasticity mechanism. Since homeostatic plasticity is considered to balance neuronal and synaptic weight dynamics (Cannon 1932; Turrigiano and Nelson 2004), it is unlikely that a functional structure, such as sequences of memories only rely on this homeostatic plasticity mechanism. Hence, we conclude that learning rules comprising correlation-based and homeostatic plasticity mechanisms mainly account for associative memory structures.

5.2.3 Augmentation of Synaptic Plasticity Yield Diverse Functional Organizations

As discussed above, synaptic plasticity dependent on the dynamics of correlated activity account for associative memory structures. This is due to the fact the specific mathematical formulation of correlation-based synaptic plasticity, analyzed in this thesis, only accounts for *long-term potentiation* (LTP), i.e. strengthening of the synaptic weights. Thus, we continued to augment the model on activity-based plasticity mechanisms to reliably describe the diversity in the formation of functionally organized memories.

Synaptic Plasticity Based on LTP and LTD

At a biological synapse, the mechanism of LTP accounts for an enhanced exocytosis of ion-channels into the postsynaptic membrane (Fig. 2.3 bottom left, p. 13, Huganir and Nicoll 2013; Korte and Schmitz 2016) that enable signal transmission between the pre- and postsynaptic neurons. Hence, an increase of ion-channels in the postsynaptic cell cause

this postsynaptic neuron to be more sensitive to presynaptic signals. *Long-term depression* (LTD), on the other hand, accounts for an enhanced endocytosis of these ion channels into the postsynapic cell (Fig. 2.3 bottom right, p. 13, Korte and Schmitz 2016), thereby, decreasing the excitability of the postsynaptic cell. The latter mechanism is not considered by the mathematical formulation of correlation-based synaptic plasticity. Even though, this formulation is augmented by a stability-serving homeostatic process, this process does not imply LTD. Thus, as a first approach, we additionally introduced LTD, that take those plasticity mechanisms into account that yield a weakening of the synaptic weights (Sec. 4.3; p. 73).

Although, previous studies already engineered LTD in mathematical models of rate-coded neurons with synaptic plasticity (Sejnowski 1977; Bienenstock et al. 1982; Oja 1982), a mathematical formulation of biologically plausible activity-dependent learning rules were still missing. For example, models proposed by Sejnowski (1977) and Bienenstock et al. (1982) mathematically implement LTD by means of threshold values for the pre- and/or prostsynaptic firing. However, these thresholds yielding stable synaptic weight dynamics were technically derived (Sejnowski 1977; Bienenstock et al. 1982). As another example, the model propsed by (Oja 1982) does not make use of such arbitrary firing thresholds but still need to implement biologically questionable non-local synaptic weight normalization enabling stable synaptic weight dynamics. Our theoretical considerations led to the mathematical formulation of an additional plasticity mechanism accounting for LTD which is dependent on the difference in pre- and postsynaptic firing rates. Hence, this formulation differs from the previous models as it is independent of additional threshold values or biologically implausible mechanisms such as non-local synaptic weight normalization serving for stability. For stability, in our model, we used the experimentally observed homeostatic mechanism of synaptic scaling (Turrigiano et al. 1998; Turrigiano and Nelson 2004) that has aleady been mathematically implemented for synaptic plasticity rules by Tetzlaff et al. (2011). Furthermore, our proposed mechanism describing activity-dependent LTD, dependent on the difference in pre- and postsynaptic firing rates, enables the system of two interconnected memories to form a discrimination which we show in theory (Theorem 12, p. 76) and simulation (Sec. 4.3.2, p. 78). We validated our theoretical proposition that a LTD-term depending on the difference in the pre- and postsynaptic firing rates of rate-coded neurons is able to form discriminated memories. For this, we analyzed a specific learning rule based on the previously analyzed SPaSS-learning rule with an additional plasticity mechanism account for LTD (Sec. 4.3.2, p. 78). Evidently, this analysis showed that an increased impact of the LTD-dependent plasticity mechanism causes an enlarged activity space leading to the formation of discriminated memories. Strikingly, also the formation of memories arranged in a sequential order is facilitated by this additional plasticity mechanism (Fig. 4.4, p. 80).

Synaptic Plasticity with Local Characteristics

Heretofore, we only considered a homogeneous neuronal network model where all neuronal and synaptic parameters are unchanged for the whole neuronal network. As a second approach, we allowed local differences in the neuronal network dynamics enabling the formation of all FOs in two different ways: local differences in the homeostatic plasticity mechanism (Sec. 4.4.1; Fig. 4.5, p. 84) or inhibitory plasticity with local differences in the inhibitory synaptic weight dynamics (Sec. 4.4.2; Fig. 4.7, p. 88).

We have already discussed (Sec. 5.2.2) that it is biologically unplausible when the formation of specific FOs is caused by effects of homeostatic plasticity. However, the here introduced approach on local differences in the homeostatic mechanisms can alternatively be interpreted as a form of memory consolidation (McGaugh 2000). Memory consolidation describes the process that facilitates memories to become increasingly resistant to disruption or interference from competing factors over time (Walker and Stickgold 2004). Our general formulation, by lowering the time constant for the homeostatic plasticity mechanism γ (Fig. 4.5 A, p. 84) for all neurons belonging to one neuronal population implies a lower impact of the homeostatic mechanism. This in turn, implies all synapses within one neuronal population are less regulated and, hence, more autonomously stabilize their neuronal and synaptic dynamics. Thus, we interpreted this approach as a form of memory consolidation. Furthermore, this form of memory consolidation leads to the self-organized formation of all FOs of interconnected memories.

Moreover, we proposed an additional solution to enable the neuronal system to build up all different forms of FOs of two interconnected memories by inhibitory synaptic plasticity (Sec. 4.4.2, Fig. 4.7, p. 88). Although this category of synaptic plasticity is not well understood, we have shown that lower inhibitory synaptic weights within each population compared to the inhibitory synaptic weight level within the remaining network yields the reliable formation of all different forms of functionally connected memories. Yet inhibition has the functional role to balance strong excitation within interconnected groups of neurons (Vogels et al. 2011; Litwin-Kumar and Doiron 2014). The approach considering distinct inhibitory synaptic weight levels within the neuronal populations as compared to the remaining neurons of the network is in contrast to the general understandig on the function of inhibition. However, such local differences in the balance of inhibition and excitation become insignificant when we broaden the view on the whole neuronal network and its whole balance of excitation and inhibition (Vreeswijk and Sompolinsky 1998; Denève and Machens 2016).

5.2.4 Conclusion and Future Prospects

In summary, we developed a generic mathematical framework to analyze rate-based synaptic plasticity rules with respect to their ability to form different types of *functional organization* (FO). Note, the specific SPaSS learning rule is exemplary investigated in this thesis, but other formulations on rate-based synaptic plasticity rules can also be investigated and augmented by means of our derived mathematical framework.

At the moment, our analysis is restricted on the learning phase, i.e., the neuronal populations are continuously stimulated over time by environmental inputs. Thus, there is still an open question whether the FOs of memories can be maintained after the stimulation of the system is removed. Our analysis would predict that, as soon as the stimulation ends, the activities of the neuronal populations decline to their intrinsic states and consequently, change their respective FO. Thus, it might be possible that the here presented plasticity mechanisms have to be augmented by further processes stabilizing the memories as well as their FOs. One such mechanism that accounts for the stabilization of memories is consolidation (McGaugh 2000; Walker and Stickgold 2004), which is described above and experimentally observed.

Informed by our analytic framework, we were able to identify two strategies for augmenting the underlying synaptic weight dynamics to enable a system of two interconnected memories to form all types of FOs. Both augmentations in synaptic plasticity indeed enabled the formation of all FOs, thus, are qualitatively equivalent in the system of two interacting memories. However, it remains questionable, if both strategies are equally potent to generate functional motifs of more than two memories.

5.3 Comparison of Proposed Synaptic Learning Rules

Heretofore, we have restricted our investigations on the primary building blocks of schema. Thereby, we analyzed the ability of two interconnected memories to dynamically build diverse forms of functional connections (i.e. FO). In the following, we evaluate the different synaptic plasticity mechanism, analyzed in this thesis, on their ability to form diverse forms of schemata. Therefore, we augment a primary building block of schema by one additional memory, thus we consider three interconnected memories. In Fig. 5.1, we collect all 16 different equivalence classes of functional motifs that can be formed by three interconnected memories. (Fig. 5.1). To discuss the suitability of the different synaptic learning rules to form diverse structures of a schema, we ask if the formation of one specific FO between two memories already affects and determines the formation of the remaining FOs
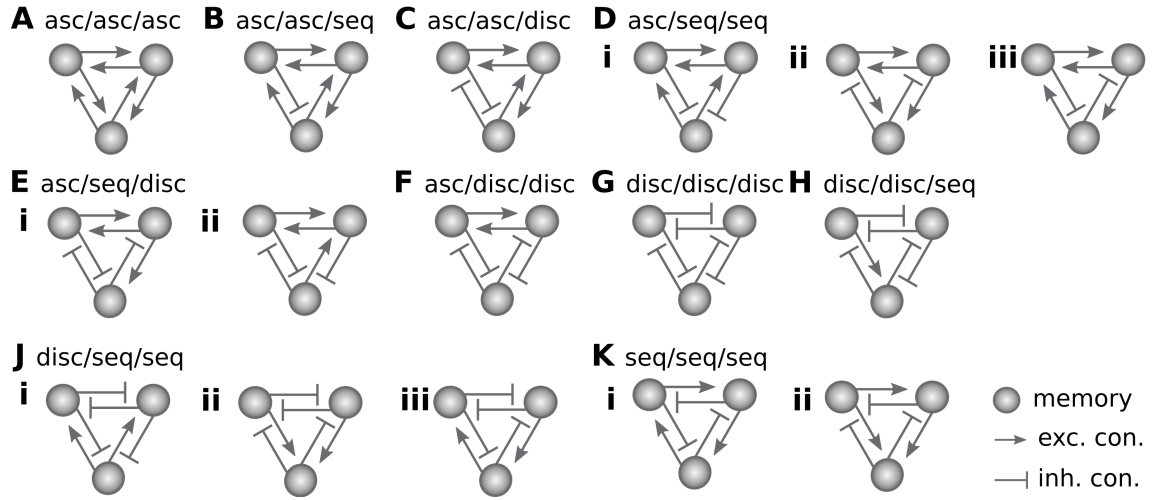
Figure 5.1: All equivalence classes for functional structures based on three interconnected memories. All equivalent combinatorics of functionally ordered memories are abstracted by their equivalence class, i.e. unordered memories.

in the structure of three interconnected memories. Thus, using this enlarged system of three interconnected memories, we briefly sketch and discuss the ability of the different learning rules to form diverse different *functional motifs* or *functional structures*. Therefore, we use the synaptic learning rule specific characteristic activity phase space (χAPS) for two interconnected memories, discussed in the following. For this, we exemplary determine one primary building block by restricting the activity of the respective memories and evaluate the influence of this restriction on the remaining two FOs within this structure of three interconnected memories. In doing this, we compare the two different synaptic plasticity mechanisms proposed in this thesis.

5.3.1 LTP and LTD Account for Diverse Schema Structures

Our results in Sec. 4.3 (p. 73) reveal that synaptic plasticity driven by the processes of LTP in combination with LTD yield an omnifarious characteristic activity phase space (χAPS) for the formation of all different forms of FO, such as association, sequence and discrimination (see Fig. 4.4, p. 80). In Fig. 5.2 we apply a specific *augmented learning rule on Synaptic Plasticity and Synaptic Scaling by Causality* (aSPaSS):

$$\tilde{\tau}_\omega \dot{\tilde{\omega}}_{i,j} = \tilde{F}_j \tilde{F}_i - \tilde{\nu} \left(\sqrt{\tilde{F}_j \tilde{F}_i} |\tilde{F}_j - \tilde{F}_i| + R_{\min} \right) - \frac{\tilde{F}_i - \tilde{F}^{\mathrm{T}}}{1 - \tilde{F}^{\mathrm{T}}} \tilde{\omega}_{i,j}^2 \tag{5.1}$$

defined by the set of parameter of $\tilde{\nu} = 0.9$, $R_{\min} = 0$ and $\tilde{F}^{\mathrm{T}} = 0.05$ as already used in Fig. 4.4 Cv (p. 80). Note, the analysis on the system's equilibrium state ($\tilde{\tau}_\omega \dot{\omega}_{i,j} \overset{!}{=} 0$) is

independent of the learning rate for synaptic plasticity. Thus, we do not have to specify the learning rate $\tilde{\tau}_\omega$ as it does not affect the resulting FO.

To sketch the ability of this exemplary learning rule to form multifarious functional interconnected memories, we exemplary limit the $\mathfrak{F}_1 - \mathfrak{F}_2 - \mathfrak{F}_3-$activity phase space for three interconnected memories and analyze the implications on the resulting functional structures. For example we start to restrict the average activity level of memory 1 to $\mathfrak{F}_1 \in [0.65, 0.8]$. This restricted activity level for memory 1 reduces the χAPS for the FOs with memory 2 (Fig. 5.2 Ai). This exemplary restriction in the activity for memory 1 still enables the formation of all possible FO, thus, it does not affect the diversity of possible FOs. The average activity of memory 2 determines the FO between memory 1 and memory 2 (1 ∼ 2). Equivalently, the restricted activity level for memory 1 reduces the χAPS for the FO with memory 3 (Fig. 5.2 Aiii) but does not affect the diversity of possible FOs. Evidently, the χAPS defining the FOs of memory 2 and memory 3 is unaffected by the restriction for the activity level of memory 1 (Fig. 5.2 Aii). When the average activity level of memory 2 is additionally restricted to operate in the activity regime of $\mathfrak{F}_2 \in [0.625, 0.825]$ (Fig. 5.2 B), this defines the FO of memory 1 and memory 2 to be an association (Fig. 5.2 Bi). Furthermore, the χAPS defining the FOs of memory 2 and memory 3 is reduced. However, they still can build up all forms of FOs (Fig. 5.2 Bi). Finally, the limitation of the average activity level for memory 3 to $\mathfrak{F}_3 \in [0.75, 1.0]$ determines the FO of memory 2 and memory 3 to either form an association or a sequence from memory 3 to memory 2 (seq:23, Fig. 5.2 Cii) and further the FO of memory 1 and memory 3 to either form an association or a sequence from memory 3 to memory 1 (seq:13, Fig. 5.2 Ciii). Thus, this exemplary restriction of the average activity levels for all three memories still accounts for different functional motifs. One such functional structure is composed out of association for each single primary building block, where all memories build up associations towards the interconnected memories (Fig. 5.2 Ei). The second equivalence class of functional motifs is that of two associations paired with on sequence (Fig. 5.2 Eii). This specific motif occurs for an association of memory 2 and memory 3 and a sequence from memory 3 to memory 1 (seq:13). This specific structure is equivalent that the structure determined of a sequence from memory 3 to memory 2 (seq:23) and an association of memory 1 and memory 3. The third functional motif is made up by an association and two sequences each building up an excitatory connection toward the memories that form an association (Fig. 5.2 Eiii). This particular functional structure is formed by a sequence from memory 3 to memory 2 and from memory 3 to memory 1, respectively.

Alternatively, if the average activity of memory 3 is restricted to $\mathfrak{F}_3 \in [0.3, 0.55]$ (Fig. 5.2 D) corresponds to a different set of functional motifs. However, the FO of memory 1 and memory 2 is unaffected of this changed activity and still forms an association. When
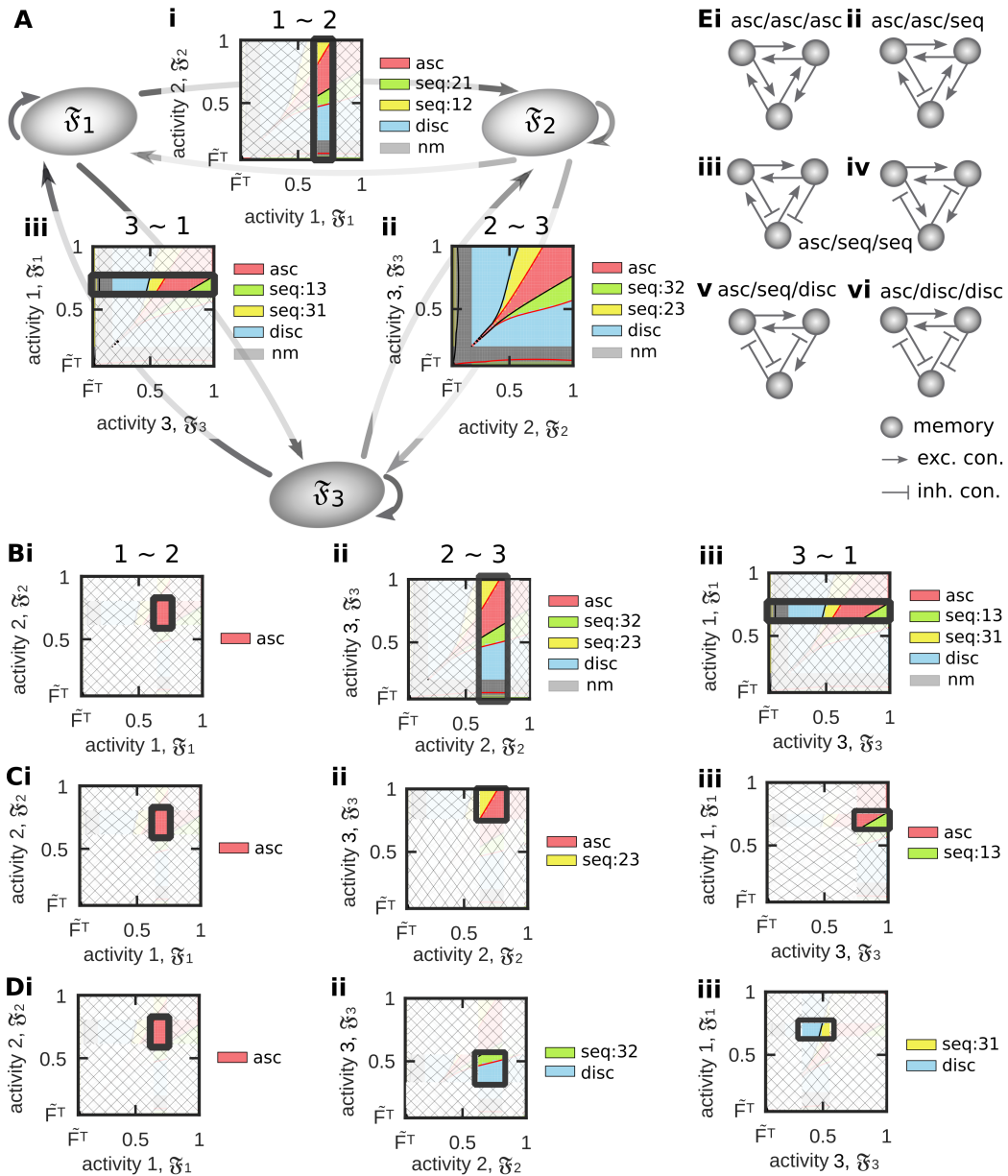
Figure 5.2: Motifs of three memories underlying the aSPaSS-learning rule. (A) Average activity of memory 1 limited to $\mathfrak{F}_1 \in [0.65, 0.8]$ restricts the χAPS for its relation to memory 2 (Ai) and memory 3 (Aiii). The χAPS defining the FOs of memory 2 and 3 is not restricted (Aii). **(B)** When the average activity of memory 2 is constrained to $\mathfrak{F}_2 \in [0.625, 0.825]$ (Bi,ii), yields an asc of memory 1 and 2 (Bi) and further restricts theχAPS defining the FOs of memory 2 and 3 (Bii). **(C)** When the average activity of memory 3 is constrained to $\mathfrak{F}_3 \in [0.75, 1.0]$ (Cii,iii), yields either an asc or seq:23 of memory 2 and 3 (Cii) and further restricts memory 1 and 3 to an asc or seq:13 (Ciii). **(D)** In contrast to (C), the average activity of memory 3 constrained to $\mathfrak{F}_3 \in [0.3, 0.55]$ (Dii,iii) yields either a seq:32 or disc of memory 2 and 3 (Dii) and further restricts memory 1 and 3 to a seq:31 or disc (Diii). **(E)** All equivalence classes of three functionally connected memories formed by the restrictions made in C and D.

memory 1 and memory 2 exemplary build up a sequence to memory 3, the system builds up a fourth functional motif with one association and two sequences that emerge from the association (Fig. 5.2 Eiii). Moreover, a fifth specific functional motif containing an association, a sequence and a discrimination can be formed. Here, the sequence builds up an excitatory relation from the memory belonging to an *association* (asc) towards the memory belonging to the discrimination (Fig. 5.2 Ev). Thus, the specific functional organization of a sequence from memory 2 to memory 3 combined with a discrimination between memory 1 and memory 3 belong to the same equivalence class as the specific functional organization of a sequence from memory 1 to memory 3 combined with a discrimination between memory 2 and memory 3. Finally, a sixth functional motif of one association and two discrimination can be formed (Fig. 5.2 Evi).

In summary, these two specific restrictions in the $\mathfrak{F}_1 - \mathfrak{F}_2 - \mathfrak{F}_3-$activity phase space of this small web of three interconnected memories (Fig. 5.2 C,D) in combination with the underlying dynamics for synaptic plasticity (aSPaSS) enables the formation of six out of nine possible functional motifs (all listed in Fig. 5.1 A-F) that contain at least one association between two memories (Fig. 5.2 E). Even though we have not provided specific examples for the formation of the three remaining functional motifs (Fig. 5.2 C, Diii, Ei), we strongly advocate their occurrence for specifically tuned restrictions in the average three-dimensional $\mathfrak{F}_1 - \mathfrak{F}_2 - \mathfrak{F}_3-$activity space of the memories involved.

Moreover, when we limit the χAPS for the interaction of memory 1 and memory 2 to form a *discrimination* (disc), we expect the same high diversity in the resulting functional structures of three interconnected memories. On top of this, adjusting the parameter $\tilde{\nu}$ for the time constant of LTD, which has a strong impact on the effective structure of the learning-rule specific χAPS, enlarges the scope of action for different forms of functional structures (for comparison see Fig. 4.4 C, p. 80).

5.3.2 Local Characteristics for Synaptic Plasticity Restricts the Diversity of Schema Structures

The results presented in Sec. 4.4 (p. 82) reveal that local characteristics in the plasticity mechanisms can form diverse functionally organized building blocks of schemata. In particular, we have shown that either local characteristics in the homeostatic mechanism (Sec. 4.4.1, p. 82) or local inhibitory synaptic plasticity (Sec. 4.4.2, p. 85) enables the formation of all FOs of two interconnected memories. Both approaches differ in the constraints on MR but do not affect the constraints on FOs. This becomes obvious from the mapping of the constraints for the different forms of FO onto the χAPS. For comparison see Fig. 4.5 Ai,Bi and Fig. 4.7 Ci,ii. Here, the χAPS exhibit qualitatively equivalent topologies

for the constraints on the different forms of FO. Thus, for the discussion of this approach it is sufficient to consider one example. To obtain the results shown in Fig. 5.3, we exemplary use the approach of local inhibitory synaptic conditions as already used in Fig. 4.7 Cii.

Again, we limit the $\mathfrak{F}_1 - \mathfrak{F}_2 - \mathfrak{F}_3 -$activity phase space for three interconnected memories in the same way as we already did in the previous sketch (Fig. 5.2) and deduce its implication on the resulting functional structures. We note that, strikingly, the variability of functional structures is drastically reduced to only three different equivalence classes of functional structures (Fig. 5.3 E). This is due to the nearly quadratically arranged activity spaces for the different FOs within the respective χAPS. In particular, when two memories build up an association, this excludes the formation of a discrimination for any of the remaining two FOs within this set of three interconnected memories. Vice versa, when two memories build up a disc, this excludes the formation of an association for any of the remaining two FOs.

5.3.3 Conclusion and Future Prospects

Regarding our brief sketches to form diverse motifs of three functionally connected memories, evidently, the method to describe synaptic plasticity by mechanisms including LTP and LTD has less inherent restrictions and yields more diverse structures than the method of local characteristics in synaptic plasticity mechanisms. Thus, synaptic plasticity described by LTP and LTD might be able to form more diverse different networks and, by this, generate more complex behavior. However, a more elaborated and comprehensive analysis on the implications of the different synaptic plasticity learning rules to form diverse functional structures of three interconnected memories is still to be done.

If a synaptic learning rule is able to form all 16 different equivalence classes of functional structures of three interconnected memories, theoretically, this learning rule is able to form diverse different networks of more than three interconnected memories without any restrictions. Thus, it yields the neuronal correlate for the generation of complex behavior.

5.4 Outlook: Processing with Memories

At the level of adaptive neuronal networks, we have extracted the essential neuronal principles for activity-dependent synaptic plasticity that enable the brain to form diverse *functional organization* (FO) of interconnected memories. We further discussed their ability to form diverse functional structures of three interconnected memories. We conclude that
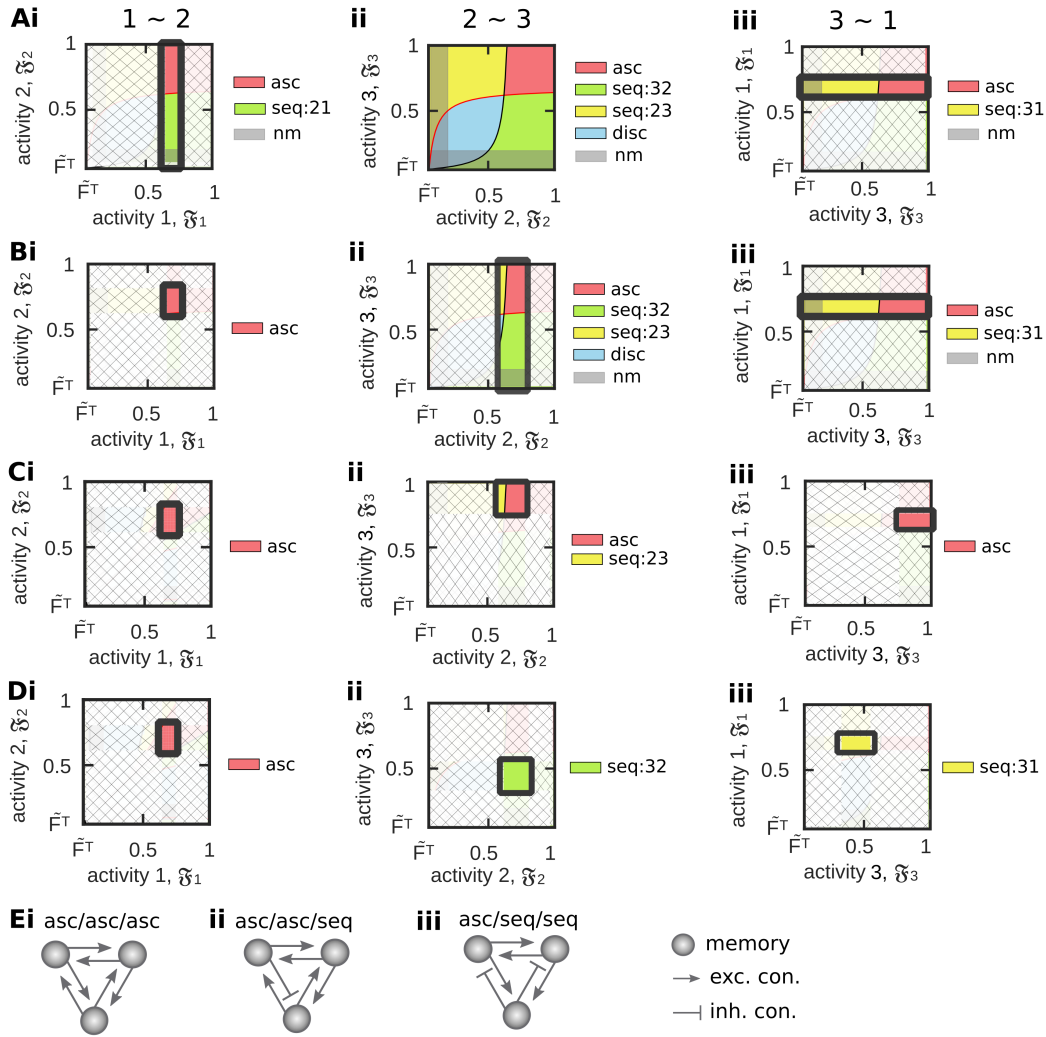
Figure 5.3: Motifs of three memories underlying the SPaSS-learning rule with local characteristics. (A) Average activity of memory 1 limited to $\mathfrak{F}_1 \in [0.65, 0.8]$ restricts the χAPS for its relation to memory 2 (Ai) and memory 3 (Aiii). The χAPS defining the FOs of memory 2 and 3 is not restricted (Aii). **(B)** When the average activity of memory 2 is constrained to $\mathfrak{F}_2 \in [0.625, 0.825]$ (Bi,ii), yields an asc of memory 1 and 2 (Bi) and further restricts theχAPS defining the FOs of memory 2 and 3 (Bii). **(C)** When the average activity of memory 3 is constrained to $\mathfrak{F}_3 \in [0.75, 1.0]$ (Cii,iii), yields an asc or seq:23 of memory 2 and 3 (Cii) and further restricts memory 1 and 3 to an asc (Ciii). **(D)** In contrast to (C), the average activity of memory 3 constrained to $\mathfrak{F}_3 \in [0.3, 0.55]$ (Dii,iii) yields either a seq:32 of memory 2 and 3 (Dii) and further restricts memory 1 and 3 to a seq:31 (Diii). **(E)** All equivalence classes of three functionally connected memories formed by the restrictions made in C and D.

activity-dependent synaptic plasticity that accounts for LTP and LTD might be key principles for the human brain's ability to create complex cognitive structures. Beyond the scope of human behavior, the possibility to form complex networks of memories in adaptive neuronal networks can be used for input-driven computation with memories.

Structured program theory (Böhm and Jacopini 1966) states that any system, which implements the three principle computational structures of sequence, selection, and iteration, is *complete* as it can execute any procedure which can be written as an algorithm (Böhm and Jacopini 1966; Byrne and Huyck 2010). The basic idea of structured program theory (Böhm and Jacopini 1966) is that programs, computers, etc. are usually composed of three functional types (i.e. , the basic computational structures) representing elementary operations mutually connected by oriented lines. Thus, the question arises, if and under which conditions the structured program theorem is also valid for adaptive neuronal networks. Byrne and Huyck (2010) already showed that five interconnected memories are required to construct the basic computational structures listed above. However, Byrne and Huyck (2010) predefined these memories as well as their required FOs to carry out the basic computational structures of sequence, selection, and iteration (Fig. 5.4 A). For instance, the computational structure of an iteration is composed out of five suitably connected memories I, II, III, A and B as depicted in Fig. 5.4 A (right panel). Here, as long as memory I and memory II are externally activated, their collective activity is sufficient to excite their excitatory interconnected memory A. The evoked excitation of memory A in turn maintains the activity of memory I and II due to their excitatory reverse connections. Thus, as long as no other stimulation is applied onto the system, the circular excitation, i.e. an iteration, of memory I and II with memory A is maintained. However, as soon as the external stimulation further excites memory III, the collective activity of memory II and III activates the excitatory interconnected memory B. This memory B has an inhibitory connection towards memory A. Hence the evoked activity of memory B suppresses the excitation of memory A and, by this, shuts down the iterative cycle described above (Byrne and Huyck 2010).

Our work strongly suggests that adaptive neuronal networks are able to learn such basic computational structures (Fig. 5.4 A), dependent on the external input stimulation, by suitably combining a specific set of functional structures (collected in Fig. 5.1). For instance, the iteration of memories can be composed out of three overlapping functional structures of three memories as depicted in Fig. 5.4 B. Note, memory II, A and B respectively belong to two functional structures that overlap. The first functional structure, containing memory I, II and A, is composed out of two associations and one discrimination (compare to Fig. 5.1 B). The second functional structure, containing memory II, A and B, is composed out of one association, secquence and discrimination (compare to Fig. 5.1 Ei). Finally, the third functional structure, containing memory II, III and B, is composed out of two secquences and one discrimination (compare to Fig. 5.1 Jii).

To test this hypothesis, one has to take into account the here derived input principles resulting either from the *two* population system forming diverse FOs (condensed in the
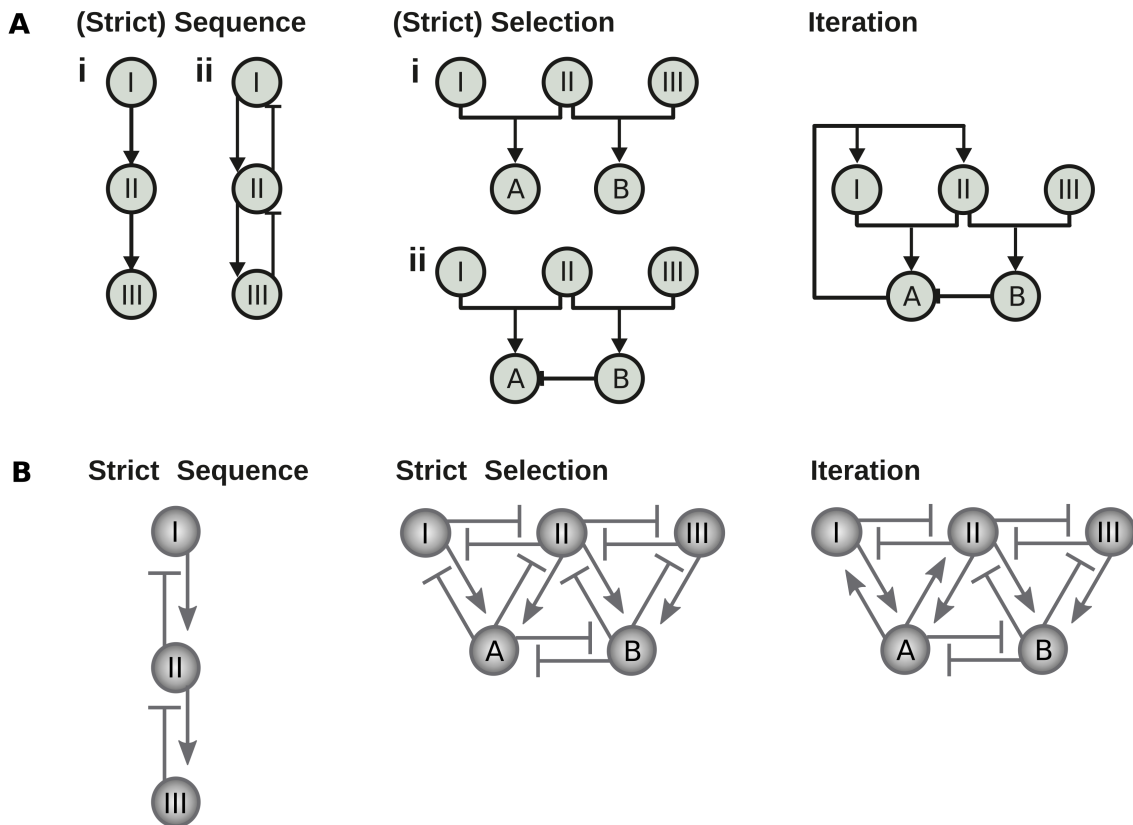
Figure 5.4: Composition of the basic computational structures out of a network of memories. **(A)** Schematic of the basic computational structures (sequence, selection, iteration), consisting of predefined and suitably connected memories (circles; arrows: excitatory connections; bars: inhibitory connections). **(B)** Construction of the basic computational tasks composed of the input driven basic functional structures of schemata.

respective synaptic learning rule dependent χAPS) or from the *three* population system forming different functional structures of schemata and transfer these principles to the wiring of up to *five* interconnected memories (Fig. 5.4 B). When all basic computational structures of sequence, selection and iteration can also be established in the population model or in the more detailed recurrent neuronal network model, one can conclude that adaptive neuronal networks are complete with respect to structured program theory. In particular, this indicates that a neuronal network can perform any algorithm by successively sequencing different basic computational tasks in a specific order. To test this, one can exemplary implement some task-based algorithms in the adaptive neuronal network which can be solved by different cognitive strategies. For instance, one such task is to write out as many of the permutations of the numbers 1, 2, 3, and 4 as possible. There are at least two strategies: (1) holding initial digits constant and changing digits on the right, or (2) rotating the preceding permutation (Leskow and Smock 1970). Therefore, one can start implementing these algorithms in the reduced population model and ex-

tract all required input patterns leading to the respective algorithm. Subsequently, one can implement those input patterns in a full network simulation and analyze the established structures. Thereby, it can be shown that, indeed, memory-based adaptive neural networks can compute programs and, thus, perform complex cognitive strategies.

Appendices

Normalized Neuronal Network Model

In the following, to reduce complexity, we normalize the neuronal the neuronal network model according to the *maximal firing rate* F^{max} and the *maximal synaptic weight* ω^{max}. The maximal firing rate is already defined as a system's parameter and set to $F^{\text{max}} = 100\,\text{Hz}$. Thus, we only have to calculate the maximal synaptic weight ω^{max}. Therefore, we set the pre- and postsynaptic neuronal activities to the maximal activity level of ($F_j = F_i = F^{\text{max}}$) and, by this, calculate the respective fixed synaptic weight of the applied learning rule ($\tau_\omega \dot{\omega}_{i,j}(F^{\text{max}}, F^{\text{max}}) \overset{!}{=} 0$) and define it as the maximal synaptic weight. For the specific learning rules of SPaSS and aSPaSS this maximal synaptic weight is given by:

$$0 = F^{\text{max}}F^{\text{max}} - \gamma \left(F^{\text{max}} - F^{\text{T}} \right)(\omega^{\text{max}})^2 \qquad \Rightarrow \qquad \omega^{\text{max}} := \frac{F^{\text{max}}}{\sqrt{\gamma(F^{\text{max}} - F^{\text{T}})}}. \quad \text{(A.1)}$$

By this, the normalized activities $\tilde{F}_i$ of all neurons $i \in \mathcal{N}$ are given by $\tilde{F}_i := F_i/F^{\text{max}}$ and the synaptic weights normalized to the maximal excitatory synaptic weight are given by $\tilde{\omega}_{i,j} := \omega_{i,j}/\omega^{\text{max}}$, $\tilde{\theta} := \theta/\omega^{\text{max}}$ and $\tilde{\omega}^{\text{ex}} := \omega^{\text{ex}}/\omega^{\text{max}}$. In a same procedure, we normalize the external input stimulation (Eq. 3.10) to

$$\tilde{\mathbf{F}}^{\text{ex}} := \left(\langle \tilde{F}_1^{\text{ex}} \rangle_t, \langle \tilde{F}_2^{\text{ex}} \rangle_t, \langle \tilde{F}_{\mathcal{B}}^{\text{ex}} \rangle_t \right)^{\mathsf{T}}. \qquad \text{(A.2)}$$

This normalized model maps the neuronal activity $\tilde{F}_i$ and the excitatory synaptic weight

$\tilde{\omega}_{i,j}$ to a state space of $[0,1] \in \mathbb{R}$, respectively:

$$\textbf{activity} \quad \tau\dot{\tilde{F}_i} = (1 - \tilde{F}_i)\tilde{F}_i \left[\log\left(\tilde{F}_i^{-1} - 1\right) + \beta(R\phi_i - \epsilon) \right], \quad \text{(A.3)}$$

$$\text{with} \quad \phi_i = \text{F}^{\text{max}}\omega^{\text{max}} \left(\sum_{r \in \{\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_\mathcal{B}\}} \sum_{j \in \mathcal{P}_r} (\tilde{\omega}_{i,j} - \tilde{\theta})\tilde{F}_j + \tilde{\omega}^{\text{ex}} \sum_{k \in \mathcal{E}_r} \tilde{F}_k^{\text{ex}} \right),$$

$$\text{(A.4)}$$

$$\textbf{synaptic weight} \quad \tilde{\tau}_\omega\dot{\tilde{\omega}}_{i,j} = \tilde{F}_j\tilde{F}_i - \tilde{\nu}\sqrt{\tilde{F}_j\tilde{F}_i}|\tilde{F}_j - \tilde{F}_i| - \tilde{\gamma}(\tilde{F}_i - \tilde{\text{F}}^{\text{T}})\tilde{\omega}_{i,j}^2, \quad \text{(A.5)}$$

$$\text{with} \quad \tilde{\tau}_\omega = \frac{\tau_\omega}{\left(\text{F}^{\text{max}}\sqrt{\text{F}^{\text{max}}\gamma(1 - \tilde{\text{F}}^{\text{T}})} \right)}, \quad \tilde{\nu} = \frac{\nu}{\text{F}^{\text{max}2}}, \quad \tilde{\gamma} = \frac{1}{1 - \tilde{\text{F}}^{\text{T}}}. \quad \text{(A.6)}$$

Note, the specific normalized model for the SPaSS-learning rule is defined by the normalized equation for the synaptic weight (Eq. A.5) with $\tilde{\nu} = 0$, excluding the causality-based plasticity mechanism.

Bibliography

Abbott, L. F. and K. I. Blum (1996). "Functional significance of long-term potentiation for sequence learning and prediction." In: *Cerebral Cortex* 6 (3), pp. 406–416.

Abbott, L. F. and S. B. Nelson (2000). "Synaptic plasticity: taming the beast". In: *Nature* 3, pp. 1178–1183.

Amari, S.-I. (1977). "Neural theory of association and concept-formation". In: *Biological Cybernetics* 26.3, pp. 175–185.

Anderson, R. C. (1984). "Role of the reader's schema in comprehension, learning, and memory". In: *Learning to read in American schools: Basal readers and content texts* 29, pp. 243–257.

Anderson, R. C. and J. W. Pichert (1978). "Recall of previously unrecallable information following a shift in perspective". In: *Journal of Verbal Learning and Verbal Behavior* 17.1, pp. 1–12.

Arkes, H. R. and M. R. Freedman (1984). "A demonstration of the costs and benefits of expertise in recognition memory". In: *Memory & Cognition* 12.1, pp. 84–89.

Atwood, B. K., D. M. Lovinger, and B. N. Mathur (2014). "Presynaptic long-term depression mediated by Gi/o-coupled receptors". In: *Trends in Neurosciences* 37.11, pp. 663–673.

Auth, J. M., T. Nachstedt, and C. Tetzlaff (2018). "The interplay of synaptic plasticity and scaling enables the self-organized allocation of multiple memory representations". In: *bioRxiv*.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Vol. 14. Cambridge University Press.

Bashir, Z. I., Z. A. Bortolotto, C. H. Davies, N. Berretta, A. J. Irving, A. J. Seal, J. M. Henley, D. E. Jane, J. C. Watkins, and G. L. Collingridge (1993). "Induction of LTP in the hippocampus needs synaptic activation of glutamate metabotropic receptors". In: *Nature* 363.6427, p. 347.

Bayley, P. J., R. O. Hopkins, and L. R. Squire (2006). "The fate of old memories after medial temporal lobe damage". In: *Journal of Neuroscience* 26.51, pp. 13311–13317.

Bennett, M. V. L. and R. S. Zukin (2004). "Electrical coupling and neuronal synchronization in the mammalian brain". In: *Neuron* 41.4, pp. 495–511.

Bi, G. Q. and M. M. Poo (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.

Bibliography

Bienenstock, E. L., L. N. Cooper, and P. W. Munro (1982). "Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex". In: *Journal of Neuroscience* 2.1, pp. 32–48.

Bliss, T. V. P. and G. L. Collingridge (1993). "A synaptic model of memory: long-term potentiation in the hippocampus". In: *Nature* 361.6407, p. 31.

Bliss, T. V. P. and T. Lømo (1973). "Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path". In: *Journal of Physiology* 232.2, pp. 331–356.

Bloodgood, B. L., A. J. Giessel, and B. L. Sabatini (2009). "Biphasic synaptic Ca influx arising from compartmentalized electrical signals in dendritic spines". In: *PLoS Biology* 7.9.

Böhm, C. and G. Jacopini (1966). "Flow diagrams, turing machines and languages with only two formation rules". In: *Communications of the ACM* 9.5, pp. 366–371.

Braitenberg, V. (1978). "Cell Assemblies in the Cerebral Cortex". In: *Theoretical approaches to complex systems*. Berlin: Springer Verlag, pp. 171–188.

Bransford, J. D. and M. K. Johnson (1972). "Contextual prerequisites for understanding: Some investigations of comprehension and recall". In: *Journal of Verbal Learning and Verbal Behavior* 11.6, pp. 717–726.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth.

Bunsey, M. and H. Eichenbaum (1996). "Conservation of hippocampal memory function in rats and humans". In: *Nature* 379.6562, pp. 255–257.

Buzsaki, G. (2010). "Neural syntax, cell assemblies, synapsembles, and readers". In: *Neuron* 68, pp. 362–385.

Byrne, E. and C. Huyck (2010). "Processing with Cell Assemblies". In: *Neurocomputing* 74, pp. 76–83.

Cannon, W. B. (1932). "The wisdom of the body". In: *NY Norton*.

Carmichael, L., H. P. Hogan, and A. A. Walter (1932). "An experimental study of the effect of language on the reproduction of visually perceived form." In: *Journal of experimental Psychology* 15.1, p. 73.

Castillo, P. E., C. Q. Chiu, and R. C. Carroll (2011). "Long-term plasticity at inhibitory synapses". In: *Current Opinion in Neurobiology* 21.2, pp. 328–338.

Chen, Y. (2017). "Mechanisms of winner-take-all and group selection in neuronal spiking networks". In: *Frontiers in Computational Neuroscience* 11, p. 20.

Chenkov, N., H. Sprekeler, and R. Kempter (2017). "Memory replay in balanced recurrent networks." In: *PLoS Computational Biology* 13 (1).

Choe, Y. (2015). "Hebbian Learning". In: *Encyclopedia of Computational Neuroscience*, pp. 1305–1309.

Chua, J. J. E., S. Kindler, J. Boyken, and R. Jahn (2010). "The architecture of an excitatory synapse". In: *Journal of Cell Science* 123.6, pp. 819–823.

Clopath, C. (2015). "Long Term Plasticity, Biophysical Models". In: *Encyclopedia of Computational Neuroscience*, pp. 1628–1640.

Collins, A. M. and M. R. Quillian (1969). "Retrieval time from semantic memory". In: *Journal of Verbal Learning and Verbal Behavior* 8.2, pp. 240–247.

Cooper, R., T. Shallice, and J. Farringdon (1995). "Symbolic and Continuous Processes in the Automatic Selection of Actions". In: *Hybrid Problems, Hybrid Solutions* 27, p. 27.

Dayan, P. and L. F. Abbott (2001). *Theoretical Neuroscience - Computational and Mathematical Modeling of Neural Systems*. Vol. 806. The MIT Press, Cambridge, Massachusetts.

Denève, S. and C. K. Machens (2016). "Efficient codes and balanced networks." In: *Nature Neuroscience* 19 (3), pp. 375–382.

Dierig, S. (1994). "Extending the neuron doctrine: Carl Ludwig Schleich (1859–1922) and his reflections on neuroglia at the inception of the neural-network concept in 1894". In: *Trends in Neurosciences* 17.11, pp. 449–452.

Douville, K., J. L. Woodard, M. Seidenberg, S. K. Miller, C. L. Leveroni, K. A. Nielson, M. Franczak, P. Antuono, and S. M. Rao (2005). "Medial temporal lobe activity for recognition of recent and remote famous names: an event-related fMRI study". In: *Neuropsychologia* 43.5, pp. 693–703.

Dudek, S. M. and M. F. Bear (1992). "Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade." In: *Proceedings of the National Academy of Sciences* 89.10, pp. 4363–4367.

Durrant, S. J., C. Taylor, S. Cairney, and P. A. Lewis (2011). "Sleep-dependent consolidation of statistical learning". In: *Neuropsychologia* 49.5, pp. 1322–1331.

Eichenbaum, H. (2012). *The Cognitive Neuroscience of Memory: An Introduction*. Oxford University Press.

Engert, F. and T. Bonhoeffer (1997). "Synapse specificity of long-term potentiation breaks down at short distances". In: *Nature* 388.6639, p. 279.

Exner, S. (1894). *Entwurf zu einer physiologischen Erklärung der psychischen Erscheinungen*. Vol. 1. Deuticke F.

Feldman, D. E. (2012). "The spike-timing dependence of plasticity". In: *Neuron* 75.4, pp. 556–571.

Flourens, P. (1842). *Recherches expérimentales sur les propriétés et les fonctions du système nerveux dans les animaux vertébrés*. Ballière.

Frémaux, N. and W. Gerstner (2016). "Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules". In: *Frontiers in Neural Circuits* 9, p. 85.

Frey, U. (1997). "Cellular mechanisms of long-term potentiation: Late maintenance". In: *Advances in Psychology*. Vol. 121. Elsevier, pp. 105–128.

Bibliography

Froemke, R. C., M. Poo, and Y. Dan (2005). "Spike-timing-dependent synaptic plasticity depends on dendritic location". In: *Nature* 434.7030, p. 221.

Gagné, R. M. (1965). *The Conditions of Learning*. Third. New York: Holt, Rinehart and Winston.

Gaiarsa, J.-L., O. Caillard, and Y. Ben-Ari (2002). "Long-term plasticity at GABAergic and glycinergic synapses: mechanisms and functional significance". In: *Trends in Neurosciences* 25.11, pp. 564–570.

Gall, F. J. (1818). *Anatomie et physiologie du systême nerveux en général, et du cerveau en particulier: avec des observations sur la possibilité de reconnoitre plusieurs dispositions intellectuelles et morales de l'homme et des animaux par la configuration de leurs têtes*. Vol. 3. F. Schoell.

Garagnani, M., T. Wennekers, and F. Pulvermüller (2009). "Recruitment and consolidation of cell assemblies for words by way of Hebbian learning and competition in a multi-layer neural network". In: *Cognitive Computing* 1, pp. 160–176.

Gerstner, W., R. Kempter, J. L. van Hemmen, and H. Wagner (1996). "A neuronal learning rule for sub-millisecond temporal coding." In: *Nature* 383 (6595), pp. 76–81.

Gerstner, W. and W. M. Kistler (2002). "Mathematical formulations of Hebbian learning". In: *Biological Cybernetics* 87.5-6, pp. 404–415.

Gerstner, W., W. M. Kistler, R. Naud, and L. Paninski (2014). *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press.

Ghosh, V. E. and A. Gilboa (2014). "What is a memory schema? A historical perspective on current neuroscience literature". In: *Neuropsychologia* 53, pp. 104–114.

Golgi, C. (1906). "The neuron doctrine: theory and facts". In: *Nobel Lectures: Physiology or Medicine* 1901–1921, pp. 190–217.

Gonzalez-Islas, C. and P. Wenner (2006). "Spontaneous network activity in the embryonic spinal cord regulates AMPAergic and GABAergic synaptic strength". In: *Neuron* 49.4, pp. 563–575.

Goodman, G. S. (1980). "Picture memory: How the action schema affects retention". In: *Cognitive Psychology* 12.4, pp. 473–495.

Govindarajan, A., R. J. Kelleher, and S. Tonegawa (2006). "A clustered plasticity model of long-term memory engrams". In: *Nature Reviews Neuroscience* 7.7, p. 575.

Griniasty, M., M. Tsodyks, and D. Amit (1993). "Conversion of temporal correlations between stimuli to spatial correlations between attractors". In: *Neural Computation* 5, pp. 1–17.

Gu, Q. (2002). "Neuromodulatory transmitter systems in the cortex and their role in cortical plasticity". In: *Neuroscience* 111.4, pp. 815–835.

Haist, F., J. B. Gore, and H. Mao (2001). "Consolidation of human memory over decades revealed by functional magnetic resonance imaging". In: *Nature Neuroscience* 4.11, p. 1139.

Hashimotodani, Y., T. Ohno-Shosaku, H. Tsubokawa, H. Ogata, K. Emoto, T. Maejima, K. Araishi, H.-S. Shin, and M. Kano (2005). "Phospholipase Cβ serves as a coincidence detector through its Ca 2+ dependency for triggering retrograde endocannabinoid signal". In: *Neuron* 45.2, pp. 257–268.

Hasselhorn, M. and A. Gold (2013). *Pädagogische Psychologie. Erfolgreiches Lernen und Lehren*. Kohlhammer, Stuttgart.

Head, H. and G. Holmes (1911). "Sensory disturbances from cerebral lesions". In: *Brain* 34.2-3, pp. 102–254.

Hebb, D. O. (1949). *The Organization of Behavior*. Wiley, New York.

Herpich, J. and C. Tetzlaff (2018). "Principles Underlying the Input-Dependent Formation and Organization of Memories". In: *bioRxiv*. A similar manuscript is currently under revision in Network Neuroscience.

Herrera-Aguilar, J. L., H. Larralde, and M. Aldana (2012). "Pattern recognition in neural networks with competing dynamics: coexistence of fixed-point and cyclic attractors." In: *PloS One* 7 (8).

Holtmaat, A. and P. Caroni (2016). "Functional and structural underpinnings of neuronal assembly formation in learning". In: *Nature Neuroscience* 19.12, p. 1553.

Huber, K. M., M. S. Kayser, and M. F. Bear (2000). "Role for rapid dendritic protein synthesis in hippocampal mGluR-dependent long-term depression". In: *Science* 288.5469, pp. 1254–1256.

Huganir, R. L. and R. A. Nicoll (2013). "AMPARs and synaptic plasticity: the last 25 years". In: *Neuron* 80.3, pp. 704–717.

Humphreys, G. W. and E. M. E. Forde (1998). "Disordered action schema and action disorganisation syndrome". In: *Cognitive Neuropsychology* 15.6, pp. 771–812.

James, W. (1890). *The Principles of Psychology*. Holt, New York.

Jessen, K. R. and R. Mirsky (1980). "Glial cells in the enteric nervous system contain glial fibrillary acidic protein". In: *Nature* 286.5774, p. 736.

Josselyn, S. A., S. Köhler, and P. W. Frankland (2015). "Finding the engram". In: *Nature Reviews Neuroscience* 16.9, p. 521.

Kandel, E. R., J. H. Schwartz, T. M. Jessel, S. A. Siegelbaum, and A. J. Hudspeth (2013). *Principles of Neural Science*. The McGraw-Hill Companies, Inc.

Kandel, E. R. (2001). "The molecular biology of memory storage: a dialogue between genes and synapses". In: *Science* 294.5544, pp. 1030–1038.

Kesteren, M. T. R. van, S. F. Beul, A. Takashima, R. N. Henson, D. J. Ruiter, and G. Fernández (2013). "Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: from congruent to incongruent". In: *Neuropsychologia* 51.12, pp. 2352–2359.

Bibliography

Kilman, V., M. C. W. Van Rossum, and G. G. Turrigiano (2002). "Activity deprivation reduces miniature IPSC amplitude by decreasing the number of postsynaptic GABA$_A$ receptors clustered at neocortical synapses". In: *Journal of Neuroscience* 22.4, pp. 1328–1337.

Kirov, S. A., K. E. Sorra, and K. M. Harris (1999). "Slices have more synapses than perfusion-fixed hippocampus from both young and mature rats". In: *Journal of Neuroscience* 19.8, pp. 2876–2886.

Kitajima, T. and K.-i. Hara (1991). "A model of the mechanism of cooperativity and associativity of long-term potentiation in the hippocampus: a fundamental mechanism of associative memory and learning". In: *Biological Cybernetics* 64.5, pp. 365–371.

Koester, H. J. and B. Sakmann (1998). "Calcium dynamics in single spines during coincident pre-and postsynaptic activity depend on relative timing of back-propagating action potentials and subthreshold excitatory postsynaptic potentials". In: *Proceedings of the National Academy of Sciences* 95.16, pp. 9596–9601.

Korte, M. and D. Schmitz (2016). "Cellular and system biology of memory: timing, molecules, and beyond". In: *Physiological reviews* 96.2, pp. 647–693.

Kubota, Y., F. Karube, M. Nomura, and Y. Kawaguchi (2016). "The diversity of cortical inhibitory synapses". In: *Frontiers in Neural Circuits* 10, p. 27.

Kumaran, D., J. J. Summerfield, D. Hassabis, and E. A. Maguire (2009). "Tracking the emergence of conceptual knowledge during human decision making". In: *Neuron* 63.6, pp. 889–901.

Lee, H.-K., M. Barbarosie, K. Kameyama, M. F. Bear, and R. L. Huganir (2000). "Regulation of distinct AMPA receptor phosphorylation sites during bidirectional synaptic plasticity". In: *Nature* 405.6789, p. 955.

Leibold, C. and R. Kempter (2006). "Memory capacity for sequences in a recurrent network with biological constraints." In: *Neural Computation* 18 (4), pp. 904–941.

Leskow, S. and C. D. Smock (1970). "Developmental Changes in Problem-Solving Strategies: Permutation". In: *Developmental Psychology* 2.3, p. 412.

Lieberman, D. A. (2004). *Learning and Memory: An integrative Approach*. Wadsworth, Belmont.

Lisman, J., H. Schulman, and H. Cline (2002). "The molecular basis of CaMKII function in synaptic and behavioural memory". In: *Nature Reviews Neuroscience* 3.3, p. 175.

Litwin-Kumar, A. and B. Doiron (2014). "Formation and maintenance of neuronal assemblies through synaptic plasticity". In: *Nature Communications*.

Lorente de Nó, R. (1949). "Cerebral cortex. Architecture, intracortical connections, motor projections". In: ed. by J. F. Fulton. Oxford University Press Inc., New York, pp. 288–330.

Lorente de Nó, R. (1934). "Studies on the structure of the cerebral cortex. II. Continuation of the study of the ammonic system." In: *Journal für Psychologie und Neurologie*.

Maffei, A., S. B. Nelson, and G. G. Turrigiano (2004). "Selective reconfiguration of layer 4 visual cortical circuitry by visual deprivation". In: *Nature Neuroscience* 7.12, p. 1353.

Malenka, R. C. and M. F. Bear (2004). "LTP and LTD: an embarrassment of riches." In: *Neuron* 44 (1), pp. 5–21.

Malinow, R. and R. C. Malenka (2002). "AMPA receptor trafficking and synaptic plasticity". In: *Annual Review of Neuroscience* 25.1, pp. 103–126.

Mar, R. A. (2004). "The neuropsychology of narrative: Story comprehension, story production and their interrelation". In: *Neuropsychologia* 42.10, pp. 1414–1434.

Marder, E. and J.-M. Goaillard (2006). "Variability, compensation and homeostasis in neuron and network function". In: *Nature Reviews Neuroscience* 7.7, p. 563.

Marder, E. and A. A. Prinz (2003). "Current compensation in neuronal homeostasis". In: *Neuron* 37.1, pp. 2–4.

Markram, H., W. Gerstner, and P. J. Sjöström (2011). "A history of spike-timing-dependent plasticity." In: *Frontiers in Synaptic Neuroscience* 3, p. 4.

Markram, H., J. Lübke, M. Frotscher, and B. Sakmann (1997). "Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs". In: *Science* 275.5297, pp. 213–215.

Markram, H., E. Muller, S. Ramaswamy, M. W. Reimann, M. Abdellah, C. A. Sanchez, A. Ailamaki, L. Alonso-Nanclares, N. Antille, S. Arsever, et al. (2015). "Reconstruction and simulation of neocortical microcircuitry". In: *Cell* 163.2, pp. 456–492.

Markram, H., Y. Wang, and M. Tsodyks (1998). "Differential signaling via the same axon of neocortical pyramidal neurons". In: *Proceedings of the National Academy of Sciences* 95.9, pp. 5323–5328.

Martin, S. J., P. D. Grimwood, and R. G. M. Morris (2000). "Synaptic Plasticity and Memory: An Evaluation of the Hypothesis". In: *Annual Review Neuroscience* 23, pp. 649–711.

McBain Chris J .and Kauer, J. A. (2009). "Presynaptic plasticity: targeted control of inhibitory networks". In: *Current Opinion in Neurobiology* 19.3, pp. 254–262.

McGaugh, J. L. (2000). "Memory–a century of consolidation". In: *Science* 287.5451, pp. 248–251.

McKenzie, S. and H. Eichenbaum (2011). "Consolidation and reconsolidation: two lives of memories?" In: *Neuron* 71.2, pp. 224–233.

Miller, K. D. and D. J. C. MacKay (1994). "The role of constraints in Hebbian learning". In: *Neural Computation* 6.1, pp. 100–126.

Mulkey, R. M. and R. C. Malenka (1992). "Mechanisms underlying induction of homosynaptic long-term depression in area CA1 of the hippocampus". In: *Neuron* 9.5, pp. 967–975.

Bibliography

Nadel, L., A. Samsonovich, L. Ryan, and M. Moscovitch (2000). "Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results". In: *Hippocampus* 10.4, pp. 352–368.

Oja, E. (1982). "Simplified neuron model as a principal component analyzer". In: *Journal of Mathematical Biology* 15.3, pp. 267–273.

Palm, G. (1981). "Towards a Theory of Cell Assemblies". In: *Biological Cybernetics* 39, pp. 181–194.

Palm, G., A. Knoblauch, F. Hauser, and A. Schütz (2014). "Cell assemblies in the cerebral cortex". In: *Biological Cybernetics* 108.5, pp. 559–572.

Papo, D., J. M. Buldú, and B. Stefano (2015). "Network Theory in Neuroscience". In: *Encyclopedia of Computational Neuroscience*, pp. 1852–1868.

Patterson, K., P. J. Nestor, and T. T. Rogers (2007). "Where do you know what you know? The representation of semantic knowledge in the human brain". In: *Nature Reviews Neuroscience* 8.12, p. 976.

Piaget, J. (1926). "The language and thought of the child". In: *London: Routledge & Kegan Paul Ltd*.

– (1929). "The child's concept of the world". In: *Londres, Routldge & Kegan Paul*.

Porter, K. R., A. Claude, and E. F. Fullam (1945). "A study of tissue culture cells by electron microscopy: methods and preliminary observations". In: *Journal of Experimental Medicine* 81.3, pp. 233–246.

Posner, M. I. and S. W. Keele (1968). "On the genesis of abstract ideas." In: *Journal of Experimental Psychology* 77.3p1, p. 353.

Preston, A. R. and H. Eichenbaum (2013). "Interplay of hippocampus and prefrontal cortex in memory". In: *Current Biology* 23.17, R764–R773.

Pulvermüller, F., M. Garagnani, and T. Wennekers (2014). "Thinking in circuits: toward neurobiological explanation in cognitive neuroscience". In: *Biological Cybernetics* 108:573-593.

Ramón y Cajal, S. (1906). "The structure and connections of neurons". In: *Nobel Lectures: Physiology or Medicine* 1901–1921, pp. 220–253.

Rao, R. P. N. and T. J. Sejnowski (2001). "Spike-timing-dependent Hebbian plasticity as temporal difference learning". In: *Neural Computation* 13.10, pp. 2221–2237.

Richmond, B. J., L. M. Optican, M. Podell, and H. Spitzer (1987). "Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis". In: *Journal of Neurophysiology* 57.1, pp. 162–178.

Richter, H. and S. Yang (2009). "Learning behavior in abstract memory schemes for dynamic optimization problems". In: *Soft Computing* 13, pp. 1163–1173.

Rossum, M. C. van, G. Q. Bi, and G. G. Turrigiano (2000). "Stable Hebbian learning from spike timing-dependent plasticity." In: *The Journal of Neuroscience* 20 (23), pp. 8812–8821.

Rozov, A. V., F. F. Valiullina, and A. P. Bolshakov (2017). "Mechanisms of long-term plasticity of hippocampal GABAergic synapses". In: *Biochemistry* 82.3, pp. 257–263.

Rumelhart, D. (1980). "Schemata: the building blocks of cognition." In: *Theoretical Issues in Reading and Comprehension.* Pp. 33–58.

Rumelhart, D. E. and A. Ortony (1976). "The representation of knowledge in memory". In: *Schooling and the acquisition of knowledge.* Hillsdale: Erlbaum, pp. 99–135.

Schank, R. C. and R. Abelson (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures.*

Schleich, C. L. (1894). *Schmerzlose Operationen, örtliche Betäubung mit indifferenten Flüssigkeiten, Psychophysik des natürlichen und künstlichen Schlafes.* J. Springer.

Scoville, W. B. and B. Milner (1957). "Loss of recent memory after bilateral hippocampal lesions". In: *Journal of Neurology, Neurosurgery, and Psychiatry* 20.1, p. 11.

Sejnowski, T. J. (1977). "Storing Covariance With Nonlinearly Interacting Neurons". In: *Journal of Theoretical Biology* 4.4, pp. 303–321.

Sejnowski, T. J. (1999). "The book of Hebb". In: *Neuron* 24.4, pp. 773–776.

Senn, W., H. Markram, and M. Tsodyks (2001). "An algorithm for modifying neurotransmitter release probability based on pre-and postsynaptic spike timing". In: *Neural Computation* 13.1, pp. 35–67.

Shatz, C. J. (1996). "Emergence of order in visual system development". In: *Proceedings of the National Academy of Sciences* 93.2, pp. 602–608.

Shatz, C. J. (1992). "The developing brain". In: *Scientific American* 267.3, pp. 60–67.

Shea, N., K. Krug, and P. N. Tobler (2008). "Conceptual representations in goal-directed decision making". In: *Cognitive, Affective, & Behavioral Neuroscience* 8.4, pp. 418–428.

Shepherd, G. M. (1991). *Foundations of the Neuron Doctrine.* Oxford University Press.

Shinomoto, S., H. Kim, T. Shimokawa, N. Matsuno, S. Funahashi, K. Shima, I. Fujita, H. Tamura, T. Doi, K. Kawano, et al. (2009). "Relating neuronal firing patterns to functional differentiation of cerebral cortex". In: *PLoS Computational Biology* 5.7.

Sjöström, P. J., G. G. Turrigiano, and S. B. Nelson (2001). "Rate, timing, and cooperativity jointly determine cortical synaptic plasticity." In: *Neuron* 32 (6), pp. 1149–1164.

Smith, C. N. and L. R. Squire (2009). "Medial temporal lobe activity during retrieval of semantic memory is related to the age of the memory". In: *Journal of Neuroscience* 29.4, pp. 930–938.

Song, I. and R. L. Huganir (2002). "Regulation of AMPA receptors during synaptic plasticity". In: *Trends in Neurosciences* 25.11, pp. 578–588.

Song, S. and L. F. Abbott (2001). "Cortical development and remapping through spike timing-dependent plasticity". In: *Neuron* 32.2, pp. 339–350.

Song, S., K. D. Miller, and L. F. Abbott (2000). "Competitive Hebbian learning through spike-timing-dependent synaptic plasticity". In: *Nature Neuroscience* 3.9, p. 919.

Bibliography

Spruston, N. (2008). "Pyramidal neurons: dendritic structure and synaptic integration". In: *Nature Reviews Neuroscience* 9.3, p. 206.

Stent, G. S. (1973). "A physiological mechanism for Hebb's postulate of learning". In: *Proceedings of the National Academy of Sciences* 70.4, pp. 997–1001.

Stuart, G., N. Spruston, B. Sakmann, and M. Häusser (1997). "Action potential initiation and backpropagation in neurons of the mammalian CNS". In: *Trends in Neurosciences* 20.3, pp. 125–131.

Tetzlaff, C., C. Kolodziejski, M. Timme, M. Tsodyks, and F. Wörgötter (2013). "Synaptic Scaling Enables Dynamically Distinct Short- and Long-Term Memory Formation". In: *PLoS Computational Biology* 9(10).

Tetzlaff, C., C. Kolodziejski, M. Timme, and F. Wörgötter (2011). "Synaptic scaling in combination with many generic plasticity mechanisms stabilizes circuit connectivity". In: *Frontiers in Computational Neuroscience* 5.47.

Tetzlaff, C., S. Dasgupta, T. Kulvicius, and F. Wörgötter (2015). "The Use of Hebbian Cell Assemblies for Nonlinear Computation." In: *Scientific Reports* 5, p. 12866.

Tetzlaff, C., C. Kolodziejski, M. Timme, and F. Wörgötter (2012). "Analysis of synaptic scaling in combination with hebbian plasticity in several simple networks". In: *Frontiers in Computational Neuroscience* 6.

Thorndyke, P. W. (1977). "Cognitive structures in comprehension and memory of narrative discourse". In: *Cognitive Psychology* 9.1, pp. 77–110.

Tobia, M. J., V. Iacovella, B. Davis, and U. Hasson (2012). "Neural systems mediating recognition of changes in statistical regularities". In: *NeuroImage* 63.3, pp. 1730–1742.

Trappenberg, T. P. (2010). *Fundamentals of Computational Neuroscience*. second. Oxford University Press Inc., New York.

Tse, D., R. F. Langston, M. Kakeyama, I. Bethus, P. A. Spooner, E. R. Wood, M. P. Witter, and R. G. M. Morris (2007). "Schemas and Memory Consolidation". In: *Science* 316.

Tse, D., T. Takeuchi, M. Kakeyama, Y. Kajii, H. Okuno, C. Tohyama, H. Bito, and R. G. M. Morris (2011). "Schema-Dependent Gene Activation and Memory Encoding in Neocortex". In: *Science* 333.891.

Tully, P. J., H. Lindén, M. H. Hennig, and A. Lansner (2016). "Spike-Based Bayesian-Hebbian Learning of Temporal Sequences." In: *PLoS Computational Biology* 12 (5).

Turrigiano, G. G., K. R. Leslie, N. S. Desai, L. C. Rutherford, and S. B. Nelson (1998). "Activity-dependent scaling of quantal amplitude in neocortical neurons". In: *Nature* 391, pp. 892–896.

Turrigiano, G. and S. B. Nelson (2004). "Homeostatic plasticity in the developing nervours system". In: *Nature Reviews Neuroscience* 5, pp. 97–107.

Turrigiano, G. (2011). "Too many cooks? Intrinsic and synaptic homeostatic mechanisms in cortical circuit refinement". In: *Annual Review of Neuroscience* 34, pp. 89–103.

– (2012). "Homeostatic synaptic plasticity: local and global mechanisms for stabilizing neuronal function". In: *Cold Spring Harbor Perspectives in Biology* 4.1.

Vogels, T. P., H. Sprekeler, F. Zenke, C. Clopath, and W. Gerstner (2011). "Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks." In: *Science* 334 (6062), pp. 1569–1573.

Vogels, T. P., R. C. Froemke, N. Doyon, M. Gilson, J. S. Haas, R. Liu, A. Maffei, P. Miller, C. Wierenga, M. A. Woodin, et al. (2013). "Inhibitory synaptic plasticity: spike timing-dependence and putative network function". In: *Frontiers in Neural Circuits* 7, p. 119.

Vreeswijk, C. van and H. Sompolinsky (1998). "Chaotic balanced state in a model of cortical circuits." In: *Neural Computation* 10 (6), pp. 1321–1371.

Waldeyer, W. (1891). "Über einige neuere Forschungen im Gebiete der Anatonmie des Centralnervensystems". In: *Deutsche medizinische Wochenzeitschrift* 17.

Walker, M. P. and R. Stickgold (2004). "Sleep-dependent learning and memory consolidation". In: *Neuron* 44.1, pp. 121–133.

Wang, S.-H., D. Tse, and R. G. M. Morris (2012). "Anterior cingulate cortex in schema assimilation and expression". In: *Learning & Memory* 19.8, pp. 315–318.

Whittaker, V. P. (1983). "What is Dale's principle?" In: *Dale's Principle and Communication Between Neurones*. Elsevier, pp. 1–5.

Wickelgren, W. A. (1999). "Webs, Cell Assemblies, and Chunking in Neural Nets: Introduction." In: *Canadian Journal of Experimental Psychology* 53.1, p. 118.

Wierenga, C. J., M. F. Walsh, and G. G. Turrigiano (2006). "Temporal regulation of the expression locus of homeostatic plasticity". In: *Journal of Neurophysiology* 96.4, pp. 2127–2133.

Wigström, H., B. Gustafsson, Y.-Y. Huang, and W. C. Abraham (1986). "Hippocampal long-term potentiation is induced by pairing single afferent volleys with intracellularly injected depolarizing current pulses". In: *Acta Physiologica Scandinavica* 126.2, pp. 317–319.

Winocur, G. and M. Moscovitch (2011). "Memory transformation and systems consolidation". In: *Journal of the International Neuropsychological Society* 17.5, pp. 766–780.

Wood, R., P. Baxter, and T. Belpaeme (2012). "A review of long-term memory in natural and synthetic systems". In: *Adaptive Behavior* 20.2, pp. 81–103.

Woodin, M. A. and A. Maffei (2011). *Inhibitory synaptic plasticity*. Springer Science & Business Media.

Yuste, R. (2015). "From the neuron doctrine to neural networks". In: *Nature Reviews Neuroscience* 16.8, p. 487.

Zeithamova, D., A. L. Dominick, and A. R. Preston (2012). "Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference". In: *Neuron* 75.1, pp. 168–179.

Bibliography

Zenke, F., E. J. Agnes, and W. Gerstner (2015). "Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks". In: *Nature Communication* 6, p. 6922.

Zenke, F., W. Gerstner, and S. Ganguli (2017). "The temporal paradox of Hebbian learning and homeostatic plasticity." In: *Current Opinion in Neurobiology* 43, pp. 166–176.

Zenke, F., G. Hennequin, and W. Gerstner (2013). "Synaptic plasticity in neural networks needs homeostasis with a fast rate detector." In: *PLoS Computational Biology* 9 (11).

Zhang, W. and D. J. Linden (2003). "The other side of the engram: experience-driven changes in neuronal intrinsic excitability". In: *Nature Reviews Neuroscience* 4.11, p. 885.

Zucker, R. S. and W. G. Regehr (2002). "Short-term synaptic plasticity". In: *Annual Review of Physiology* 64.1, pp. 355–405.

Acknowledgments

It is with immense gratitude that I acknowledge Dr. Christian Tetzlaff, Prof. Dr. Stefan Klumpp, and Prof. Dr. Robert Gütig for their supervision and support guiding my scientific work. Moreover, I want to thank Prof. Dr. Florentin Wörgötter for hiring me and providing useful comments and engagements for this thesis. Furthermore, I would like to thank all former and current colleagues from the Tetzlab as well as Wörgötter-lab for their beneficial scientific inputs and technical help. Here, special thanks to Timo Nachstedt, Johannes Auth, Simon Reich, Jan-Matthias Braun, Timo Lüddecke, Jannik Luboeinski, Michael Fauth, Steffen Krüppel and Moritz Becker. They have always been available for questions and discussions. Further credits go to Mayte Bonilla Quintana, Michael Fauth, David Kappel, Daniel Miner, and Jannik Luboeinski who read through parts of this thesis and provided valuable feedback. I also thank Ursula Hahn Wörgötter, Elke Zech, Antje Erdmann and Frauke Bergmann for their administrative support and Thomas Geiling for his IT support who can always be counted on.

Special thanks go to my family and friends. My parents, Ina-Maria and Fred, have ever gave me personal freedom to develop, guided me when necessary, and supported me in every possible way. I consider myself incredibly lucky to have such devoted parents. I dedicate this work for you! To the same extend I acknowledge my brother Jakob. He has ever been and will ever be a great inspiration for me and important person in my life. Special thanks go to Jan who inspires and supports me every day. Last but not least, I thank all friends who have accompanied me so far. All of you make life worth living.

Curriculum Vitæ

Personal Information

date of birth	13.12.1988 in Grevesmühlen (Germany)
email	herpich.juliane@gmail.com
address	Georg-August-Universität Göttingen
	III. Physikalisches Institut - Biophysik
	Friedrich-Hund-Platz 1
	37077 Göttingen

Education

2015 – 2018	**Doctorate**
	International Max Planck Research School on Physics of Biological and Complex Systems
	Georg-August-Universität Göttingen
	Thesis: *The Principles of Self-Organization of Memories in Neural Networks for Generating and Performing Cognitive Strategies*
2012 – 2015	**Master of Education in Mathematics and Biology**
	Georg-August-Universität Göttingen
	Thesis: *Interactions between Memories in a Mathematical Model Based on the Hebbian Cell Assembly Hypothesis*
2010 – 2012	**Bachelor of Science in Mathematics and Biology**
	Universität Bremen
	Thesis: *Computer-Aided Analysis of several Poplar Promoters*
2008	**Abitur**
	Carl-Friedrich-Gauß Gymnasium, Frankfurt (Oder)
	Advanced level subjects: Mathematics, Biology

Scholarship

2017 – 2018	Stipend of the International Max Planck Research School Göttingen

Journal Publication

J. Herpich and C. Tetzlaff (2018). "Principles Underlying the Input-Dependent Formation and Organization of Memories". In: *bioRxiv*. A similar manuscript is currently under revision in Network Neuroscience.