学位論文

Towards the interpretation of complex visual hallucinations in terms of self-reorganization of neural networks.

(ネットワーク自己再組織化による視覚性幻覚の数理的解釈)

藤堂 真登

北海道大学大学院理学院数学専攻

2020 年 6 月

Contents

Chapter 1

Introduction

A human visual perception, which we usually achieve with no effort, is the result of unconscious, tremendous information processing in the brain. In neuroscience history, understanding of "abnormal" systems promotes that of "normal" systems. The case of patient H.M., whose medial temporal lobe was removed, significantly improved the understanding of the memory system [1]. In this regard, a mysterious visual phenomenon, visual hallucinations, might provide valuable insight into the normal vision.

Patients suffering from dementia with Lewy body (DLB) often see complex visual hallucinations (CVH). Many pathological, clinical, and neuroimaging studies have provided a wealth of insights, and several hypotheses converge on visual perception and attention deficit [2, 3, 4]. However, the limitations of these approaches have prevented elucidation of the neural basis, so that they fail to specify testable details of how normal vision develops into CVH. It is time to try to embody such a model [5]. A mathematical model that appropriately incorporates neuroscience findings is expected to provide testable details.

Neuroimaging studies do not necessarily provide consistent evidence [6], but hypometabolism and hypoperfusion in the occipital region, which corresponds to the lower visual cortex, are compatible with CVH [7, 8]. Limited imaging studies during CVH show activity in the higher visual cortex [9, 10]. Recent studies have shown that patients with DLB experience CVH through interaction with the visual environment [11, 12]. This finding suggests that top-down information, which reflects internal context, may be used to eliminate the ambiguity of visual environment [12, 13, 14, 15]. In other words, top-down compensation mechanisms, as well as bottom-up deficits, may be associated with CVH.

We investigate one possible scenario of CVH that top-down information is being used to compensate for the lack of bottom-up information. As a simplified model of the hallucinatory situation, we assume a bottom-up and a top-down network. The bottom-up network corresponds to the early visual areas of the occipital cortex and reflects the external visual stimuli. The top-down network corresponds to the frontal cortex and reflects contextual indices concerning visual objects [16]. In the brain, the structures corresponding to our bottom-up and top-down subsystems project information to the inferior temporal cortex (IT), which includes neurons sensitive to complex visual objects such as faces or objects. We are interested in the effects of selective loss of the synapses from the bottom-up network, and in how the plasticity of synapses (including those of the top-down network) might change the neural activity. It may explain the hallucinatory situation, especially if we assume that certain IT neurons show increased activity by the top-down input during CVH and the pre-reorganization representation of neurons remains after reorganization.

In terms of self-organization, theorists often assume constraints for computational accounts of brain organization, function, and a goal-directed behavior [17, 18]. In particular, strategies are employed to optimize specific quantities, such as information and prediction errors. This study proposes a new learning rule according to the information maximization with the stochastic binary unit and shows that it can be understood as one of the generalized Hebbian rules, which are the mathematical formulations closest to the phenomenology of biological synaptic plasticity.

In Chapter 2, background knowledge in neuroscience is given. This includes visual processing in the cerebral cortex, CVH mainly in DLB patients, the mathematical formulation of synaptic plasticity associated with the Hebbian learning, related computational works of CVH, and findings of phantom perception due to the network reorganization. These findings lead to a scenario of the development of CVH, and support the validity and uniqueness of our working hypothesis and computational model. In Chapter 3, we introduce a mathematical model of neural networks and describe our method of investigating the neural mechanism of CVH within the framework of computation theory. In Chapter 4, we present the computation results. We explain how the proposed learning rule can be understood as one of the Hebbian rules and what properties it has after learning. Changes in neural activity through the reorganization process are understood as state transitions. We conclude that one of the specific change corresponds to the proposed scenario of CVH. Chapter 5 contains a discussion of the validation of our hypothesis and computational model, including future works.

Chapter 2

Background

In search of constructing a working hypothesis on the mechanism of complex visual hallucinations (CVH), this chapter describes related physiological, medical, and computational studies in neuroscience. We provide an overview of visual information processing in the human brain, which is divided into two aspects of the processing system in Section 2.1. We explain clinical and neurophysiological evidence regarding CVH in dementia with Lewy body (DLB) in Section 2.2. We describe synaptic plasticity, especially conventional mathematical formulation in Section 2.3. Finally, we overview the related computational models of CVH and explain our working hypothesis in Section 2.4.

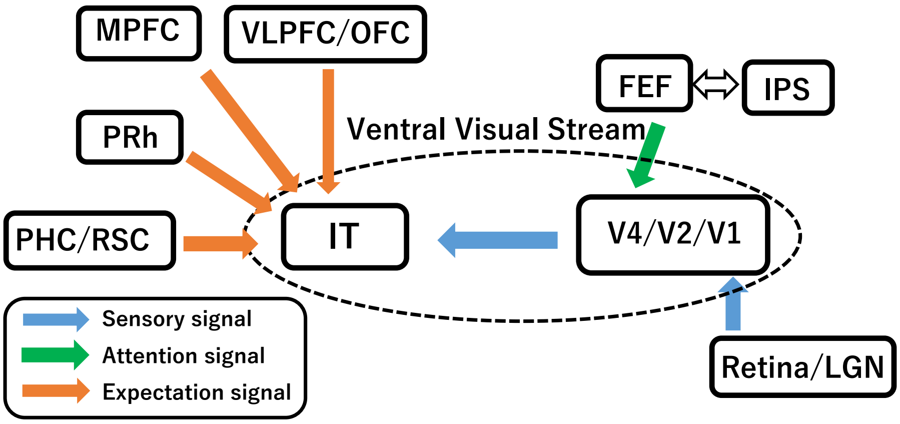

Figure 2.1: **Brain regions related to visual processing centered at IT cortex.** Each arrow represents a different effect on visual processing: external visual information (in blue), attention of the task-relevant object (in green), or expectation of the identity of the visual object (in orange). Abbreviations: IT, inferior temporal cortex; V4/V2/V1, visual cortices; RSC, retrosplenial cortex; PHC, parahippocampal cortex; PRh, perirhinal cortex; PFC (VLPFC/OFC), prefrontal cortex (ventrolateral PFC/ orbitofrontal cortex); MPFC, medial prefrontal cortex; FEF, frontal eye field; IPS, intraparietal sulcus; LGN, lateral geniculate nucleus.

2.1 Two visual processing

Visual information from the external world is processed in extensive areas, from the retina to the visual associative cortices in humans. A vast number of systematic researches have been conducted and some basic mechanisms have been established. In the following subsections, we divide two distinct visual processing, ”bottom-up” and ”top-down”, and describe their basic characteristics. Fig. 2.1 summarizes the brain regions and their connections that we will discuss.

2.1.1 Bottom-up process

The hierarchical parallel processing [19] could explain the bottom-up process for visual information. Parallelism refers to the processing of different properties of visual information via several independent visual pathways. In visual processing, there are two well-known pathways: dorsal visual stream and ventral visual stream [20]. The former, which is composed of occipotoparietal cortex areas, is involved in the detection of visual spatial and motor information. The latter, which is composed of occipototemporal cortex areas, is involved in the detection of the visual features such as shapes or colors.

Hierarchy refers to serial processing across the cortices in each pathway. As the process progresses, the size of the receptive field becomes larger, and it begins to use higher levels of visual information. For example, each neuron in the primary visual cortex (V1), which is the lowest domain of both streams, responds to the specific orientation within a tiny region of a particular location in the retina. On the other hand, each neuron in the inferior temporal cortex (IT), which is the higher domain of the ventral stream, responds to complex shapes or objects within a large foveal region in the retina. The following experimental procedure determines these specific responses of neurons .

We here take IT neurons as an example [21], which might be associated with CVH. Single IT neurons in the monkey show spike activities above baseline at 100 msec after the presentation of visual stimuli, then peak at several tens msec. Therefore, by setting a time window, several 10 to 100 msec around 100 msec after the presentation, and counting the number of spikes within this window, the response of the neuron to visual stimuli can be measured. Repetition of the same stimuli provides the average number of spikes per second, i.e., the average firing rate. Although various extensions, such as the population decoding, might be considered in the experimental studies, the primary characteristic of the neural response is the average firing rate thus obtained.

IT neurons in the monkey are selective to complex visual images, including inanimate, animate objects, human faces, or body parts [22]. However, it is not clear what these neurons actually respond to, and there are two proposals [23]. One is a holistic-based representation that responds to a whole face [24], and the other is a parts-based representation that responds to features contained in the face rather than the face itself [25, 22]. IT neurons with similar selectivities tend to be close together and form spatial clusters such as columns (0.5 mm) and patches (5 mm) [26, 27]. These clusters can also be detected by imaging in humans and monkeys because MRI has a spatial resolution of about 1 mm. Indeed, monkeys have similar category representations with humans [28]. These brain activities in IT are considered to follow a purely hierarchical bottom-up process in the ventral stream since the top-down effects described below are at a minimal level.

2.1.2 Top-down process

Attentions for visual perception are typical examples of the top-down effects for which behavior, human imaging, and single cell studies have been examined [29]. It is a voluntary process that preferentially treats a task-relevant stimulus over irrelevant ones. One example shows the attentional effect in the visual cortex of the behavioral monkey [30]. Firing rate for the presentation of multiple stimuli within the receptive field keeps the relation of the linear summation of each firing rate at a single presentation, and the balance is biased toward the attended stimulus. In other words, attention alters the contribution of weight to the firing rate. Several attentional effects on the firing rate in the visual cortex have been considered [31, 32], but all of these effects only facilitate neural response to visual input, rather than driving perception itself.

Attention signals to visual areas are projected from the frontal and parietal cortices, especially lateral intraparietal cortex (LIP) and frontal eye field (FEF) [33]. Human neuroimaging studies also support this finding and report increased activity during the maintenance of attention in the absence of visual stimuli in FEF and the intraparietal sulcus (IPS) [34], which is the human homologue of the monkey LIP [35].

Expectations are similar but distinct top-down examples compared with attentions. Both effects facilitate visual perception, but expectations constraint interpretation by using prior information about the visual environment, whereas attentions prioritize visual stimuli related to the behavioral goal [36]. Limited studies examine the neural correlation at a single cell level, but several candidates are being considered. One candidate of expectations comes from a series of experiments examining long-term memory reported in the higher visual cortex. These studies directly demonstrated the importance of the role of the lateral prefrontal cortex (PFC) neurons to IT neurons during a paired association task in adult monkeys. The perirhinal cortex (PRh) is also necessary for associative learning in IT [37, 38]. On the other hand, several studies proposed that expectations occur as a suppressed effect on sensory neurons for the prediction stimuli [39]. For example, one study showed the increased activity of IT neurons in response to unpredicted stimuli than in response to predicted stimuli [40].

Human neuroimaging studies have inferred the source of other top-down signals of expectation. For example, the increased activation in the parahippocampal cortex (PHC) and the retrosplenial complex (RSC) was found, when visual objects associated with a particular context (e.g., a hardhat) were compared with ones not related to any specific contexts (e.g., a fly) [41]. Although posterior PHC corresponds to the parahippocampal place area, which is involved in visual-spatial information such as a particular landmark or scene [42], it has been proposed that PHC is more generally involved in contextual information, including non-spatial information [43]. The orbitofrontal cortex (OFC) sends an initial guess to IT based on the low spatial frequency visual information [44, 45]. Furthermore, the medial prefrontal cortex (MPFC) is also involved in the expectation toward the face [46].

2.2 Complex Visual Hallucinations

Disorders that experience visual hallucinations include neurodegenerative disorders such as dementia with Lewy body (DLB), Parkinson's disease (PD), and Parkinson's disease with dementia (PDD), eye disease, schizophrenia, epilepsy, migraine, or arousal disorders such as narcolepsy as well as drug inducing or sensory deprivation [47]. In this study, we mainly focus on the visual hallucinations of DLB, but also provide findings regarding other symptoms to complement the missing evidence in DLB. In particular with PDD and DLB, they might have the same neural basis of CVH because Lewy pathology and characteristics of CVH are common [48].

2.2.1 Phenomenological findings

Visual hallucinations are defined as ”involuntary images that are experienced as real during the waking state but for which there is no objective reality” [2]. Complex visual hallucinations (CVH) include people, animals, and objects, on the other hand, simple hallucinations include dots, lines, and flashes [2]. The prevalence of CVH within each group is approximately 70 % in DLB [49], 50 % in PDD [49], 10 % in PD [49], 30 % in shizophrenia [50], and 15 % in eye disease [51]. Simple hallucinations are associated with eye disease [52] or migraine (known as a migraine aura).

In DLB and PDD, visual hallucinations are mostly complex, lasting for minutes rather than seconds or hours [48]. Patients commonly see a single colored object in the central visual field, superimposed onto a normal background scene. Hallucinated objects are static in more than 50 % cases, but moving objects or scenes are also experienced.

A wide range of cognitive deficits is known in DLB patients. Compared with Alzheimer’s disease, visual perception and visual attention deficits are highlighted, whereas memory deficits are preserved [53]. Indeed, these two deficits are also associated with CVH in DLB [54, 55, 56]. Also, PD patients with CVH have more significant deficits in object and face identification associated with higher visual regions [15, 57].

CVH in DLB patients may not be discrete symptoms and be associated with symptoms such as illusions or minor hallucinations. Patients with DLB tend to see pareidolias compared with AD or control people [11]. Seeing pareidolias is also associated with CVH [12]. Recent studies focus on minor hallucinations as a precursor to CVH in PD patients [58, 59], while another study suggests that these symptoms are independent [60].

2.2.2 Neuropathological findings: During hallucinations

Direct observations of brain activity during CVH in DLB patients have hitherto not been made; however, increased activity across the visual cortex during CVH has been recorded in brain imaging studies of PD [9, 10] and eye disease patients [61, 62]. In particular, the content of hallucinations (such as colors, faces, and objects) seems to be correlated with activity in the corresponding functional specializations of the visual cortex [61]. This increased activity may reflect the activity of neurons selective to complex visual features such as animals or humans. However, the only study in PD reported decreased activity in the fusiform gyri within IT and increased activity in other areas such as the frontal cortex [63]. There is limited direct evidence that the source of IT activity comes from outside IT during CVH [64], including the co-occurrence increased activity in the frontal region [10, 62].

2.2.3 Neuropathological findings: Associations with hallucinations

DLB and PD are neuropathologically characterized by Lewy bodies, which is the aggregation of α-synuclein. Lewy bodies are associated with neurodegeneration and might cause CVH. In limited studies of association with hallucinations mainly in PD and DLB, Lewy bodies have been observed in IT [65], parahippocampus [65], amygdala [65, 66, 67], frontal [57], temporal [67], parietal [67], and anterior cingulate cortex [57]. Structural imaging studies suggest a wide range of grey matter loss across cortical-subcortical regions in PD and PDD [68, 69, 70]. In these studies, however, a cognitive level between PD patients was not controlled, and brain atrophy can be related to cognitive differences rather than the presence of CVH [71]. In a neuroimaging study that controls the cognitive level in PD patients, CVH was associated with atrophy, mainly in occipital regions. In contrast, dementia was associated with the frontal cortex and medial temporal lobes [72].

Indeed, abnormalities other than atrophy in the visual cortex have been consistently reported. Hypometabolism and hypoperfusion in this area have also been reported with relation to CVH in DLB [7, 8] and PD [73, 74, 75]. Furthermore, fMRI studies during visual perception tasks have reported hypoactivation in this area of DLB [76] and PD patients [77, 78]. Despite no association between Lewy body deposition in the visual cortex and CVH [79], phosphorylation α-synuclein aggregates, which is smaller than Lewy body, have been located at presynapses; this abnormality may lead to the loss of postsynaptic dendrite spines in DLB patients [80]. White matter loss in the inferior longitudinal fasciculus or occipito-parietal regions might be associated with this pathological finding [81, 82].

CVH might be associated with neocortical cholinergic deficits, rather than a neuronal loss in the neocortex [83]. Neuronal loss by the aggregation of Lewy bodies, especially in the nucleus basalis of Meynert (NBM), leads to a decrease in acetylcholine (ACh) projections and the reduction of cortical choline acetyltransferase activity [84]. Decreased uptake of ACh in the occipital region has been involved in hypometabolism in DLB and PDD [85, 86]. Although there is a complex and not yet fully understood interplay between the CVH and the contribution of medical treatment, increasing ACh by acetylcholinesterase inhibitors has been reported to improve visual hallucinations [87]. NBM cholinergic input is known to modulate the visual cortical firing rate and improve visual discrimination [88, 89]. NBM signals seem to be a facilitator rather than an initiator [83].

In PD, early studies suggested that hallucinations were induced by the dopaminergic treatment, referred to as "levodopa psychosis" [90, 91], but some studies challenged this issue [92, 93, 47]. Indeed, hallucinations were known before the pre-levodopa era [94]. Nevertheless, the role of dopamine itself remains controversial, suggesting a role in promoting hallucinations rather than in isolated inducing [58].

2.3 Mathematical models for synaptic plasticity

Information transmission between neurons occurs via synapses. A spike generated from a presynaptic neuron changes the membrane potential of the postsynaptic neuron. This fluctuation difference can be regarded as a synaptic strength and leads to alter the average firing rates. A synaptic strength itself changes depending on the activities of pre- and post- synaptic neuron. In neuroscience, such synaptic change between neurons is a fundamental mechanism of learning for experience-dependent behavior. Current experimental and theoretical findings of synaptic plasticity are following Hebbian postulate, which is summarized as "cells that fire together wire together" [95].

A mathematical formulation of Hebbian plasticity can be divided into a rate-based or a spike timing-based learning rule [96]. The former is adopted in this paper because our targeting neurons are considered to represent information as firing rates (see Section 2.1). This section outlines several rate-based learning rules, including generalized Hebbian rules and metaplasticity. These formulations are useful to classify our proposed learning rule.

2.3.1 Rate-based local learning rule

We consider a simple neural model that consists of M input neurons, whose activity is described by an M-dimensional vector, $x \in \mathbb{R}^M$, and a single output neuron, whose membrane potential, $v(x) \in \mathbb{R}$ is defined by the following equation.

$$u(x) = \sum_{j=1}^{M} x_j w_j, \tag{2.1}$$

where $w_j \in \mathbb{R}$ is the synaptic weight from j th input neuron to the output neuron and $\theta \in \mathbb{R}$ denotes the bias. The first term, $u(x) \in \mathbb{R}$ represents an activity from input x.

One property of Hebbian plasticity is a "locality", which means that the change of synaptic weight w_j depends only on itself, the pre-synaptic activity x_j (j-th component of x) and the post-synaptic activity u. This is generally expressed as follows.

$$\tau \frac{d}{dt} w_j(t) = F(x_j, u(x), w_j), \ \ (j = 1, ..., M), \tag{2.2}$$

where F is an undetermined function. This equation implies that each synaptic weight w_j evolves from initial weights $w_j(0)$. This learning rule is called a rate-based learning rule because x_j and u are regarded as the averaged firing rate, but is should be cared to the correspondences with experiments due to their negative values.

2.3.2 Linear Hebbian rule

The other property of Hebbian plasticity is a "cooperativity", which implies that the synaptic weight is strengthened when pre- and postsynaptic neurons get active simultaneously. The simplest form is a linear Hebbian rule, which is represented by $F = x_j u$. When considering a finite set of input patterns $\mathcal{X} = \{x^k\}_{k=1}^K$, where K is the number of input patterns. We simply denote here $u(x^k), v(x^k)$ by u^k, v^k, respectively. Then, a linear Hebbian rule is described by

$$\tau \frac{d}{dt} w_j(t) = \sum_{k=1}^{K} x_j^k u^k, \ \ (j = 1, ..., M). \tag{2.3}$$

This equation can be rewritten as

$$\tau \frac{d}{dt} w(t) = Xu = XX^T w, \tag{2.4}$$

where $w = (w_1, ..., w_M)^T \in \mathbb{R}^M$, $u = (u^1, ..., u^K) \in \mathbb{R}^K$, and $X \in \mathbb{R}^{M \times K}$ is the matrix whose k th column is x^k. Since XX^T is a positive-semidefinite matrix, trajectories following this equation diverge exponentially. The instability due to positive feedback of the Hebbian rule is derived from this fact [97]. Theoretical and experimental approaches to homeostatic plasticity as a stabilizing mechanism have been actively pursued [98].

The natural way for stabilizing of learning is to introduce a constraint that keeps the overall synaptic weight constant. There are several variations on this approach [99, 100], but one systematic approach is taken by Miller and Mackey [97]. They showed that learning could converge or diverge depending on synaptic constraints. In the case of convergence, the weight strength is stabilized at the principal component of the correlation matrix XX^T.

2.3.3 Generalized Hebbian rules and metaplasticity

First, we introduce a generalized Hebbian rule which is a generalization of the linear Hebbian rule Eq. (2.4).

$$\tau \frac{d}{dt} w(t) = X g(u), \tag{2.5}$$

where $g(u) = (g_1(u), ..., g_K(u))^T$ is a vector-valued function. A function $g_k : \mathbb{R}^K \to \mathbb{R}$ is an undetermined function. When $g_k(u) = u^k$, this equation is equal to the linear Hebbian rule.

The other way has been proposed to satisfy both the stabilization of learning and the acquisition of stimulus selectivity. The Bienenstock-Cooper-Monroe (BCM) rule achieves these properties and is one of the generalized Hebbian rules with the following function [101, 102].

$$g_k(u) = p(x^k) u^k (u^k - \Psi(u)), \tag{2.6}$$

where $\Psi(u) = E[u^2] = \sum_k p(x^k)(u^k)^2$, and $p(x^k)$ is the probability that a certain pattern x^k is fed during learning. We consider that $p(x^k) > 0$ for any k and $\sum_k p(x^k) = 1$. $\Psi(u)$ is referred to as the sliding threshold, which is the post-synaptic activity dependent threshold at which synaptic weights increase or decrease. This effect is known experimentally as meta-plasticity [103]. After learning converges, an output neuron responds selectively to only one of the K input patterns.

Our proposed learning rule is one of the generalized Hebbian rules with a sliding threshold whose function g differs from the BCM rule (see Subsection 4.1). Our proposed learning rule realizes selective response to one or more input patterns, unlike the BCM rule. Different functional properties appear depending on the function g. A systematic study applying the generalized Hebbian rule to natural images reported that a receptive field such as a simple cell in V1 could be obtained with a wide range of nonlinear functions g, but not linear Hebbian rule [104].

2.4 Working Hypothesis

2.4.1 Previous computational studies related to hallucinations

Computational models of CVH are in their early stages, but provide insights for the generation of hallucinatory images or functional changes in the preservation or damage of networks [14]. Here, we describe four different approaches and specific examples. First, the effect of acetylcholine in DLB patients was studied using a recurrent neural network for associative memory [105, 106]. These studies have supported the hypothesis that the deficit of acetylcholine alters the landscape of attractor dynamics and evokes the internal templates of visual objects.

Second, there are several computational studies based on the hypothesis that CVH reflects the internal representation of the visual system. One computational study using a Boltzman machine, which is a generative model of the recurrent network, has supported the hypothesis that the loss of visual input and a homeostasis effect evoke the internal representation template [107]. Multilayer feedforward neural networks, including convolutional neural networks, are considered to be the most plausible models of the human visual system in terms of functional and physiological aspects [108]. One study generated psychedelic hallucinatory images by updating input images to make a specific layer of neural networks more active [109]. This approach makes it possible to reproduce the hallucinatory experience in human subjects [110].

Third, the most referenced computational framework in hallucinations follows the concept of Bayesian inference or predictive coding [111, 112, 113]. The Bayesian inference derives a posterior distribution from a likelihood and a prior distribution. In this context, the likelihood is assumed bottom-up information such as sensory evidence, and the prior distribution is assumed top-down information such as the expectation of cause. Hallucinations are caused by a lack of bottom-up information and/or a strong dependence on top-down information. A Bayesian modeling approach for estimating behavioral data from the hallucination-induced experiment shows that a parameter determining the balance between prior and likelihood is sufficient to distinguish between non-hallucinators and hallucinators, suggesting over-weighting to the subject's prior [114]. There is also an attempt to formulate the cholinergic effect as the precision of the sensory evidence [115, 111, 116].

Finally, computational hallucinations by conduction disturbance between functionally distinct networks were investigated [117, 118]. As these authors describe, the blockage of information from the early visual cortex to PFC generates predictions about object identity that is different from that of the external world, resulting in CVH. Our working hypothesis and model extends the concept of conduction disturbance.

2.4.2 Network self-reorganization as a working hypothesis

This study hypothesizes that consistently reported damage in the visual cortex is the core mechanism of CVH in DLB patients, as discussed in Section 2.2. However, why does specific increased activity occur in the higher visual cortex during CVH as a result? Perhaps there is a compensatory mechanism [119, 14] whereby top-down information is used to compensate for the lack of bottom-up information. To test this scenario, we assume synaptic plasticity as the compensation mechanism. Fig. 2.2 facilitates an intuitive understanding of our working hypothesis. This study treats an output network as IT, a bottom-up network as early visual cortices, and a top-down network as VLPFC/OFC, but the other brain area might also be considered as one of the candidate. The functional role of each network in visual processing is described in Section 2.1. We are interested in the effects of selective loss of the synapses from the bottom-up network, and in how the plasticity of synapses (including those of the top-down network) might change the neural activity.

Similar scenario of our working hypothesis has been proposed in a large body of hallucination literature [15, 77, 64, 12], but few details of the neural mechanism have been discussed. On the other hand, there is a discussion on phantom perceptions due to the network reorganization, suggesting potential neural mechanisms or representation, as described below.

2.4.3 Reorganization phenomena

Understanding the functional recovery of behavior by neural plasticity is an essential theme in neuroscience because of its benefit to disease [120]. Neural plasticity reflects a complicated mechanism on a micro to macro scale and includes the contribution of experience-dependent synaptic plasticity [121]. Neural plasticity can be viewed as adaptive when it relates to gain of function or as maladaptive when it relates to negative

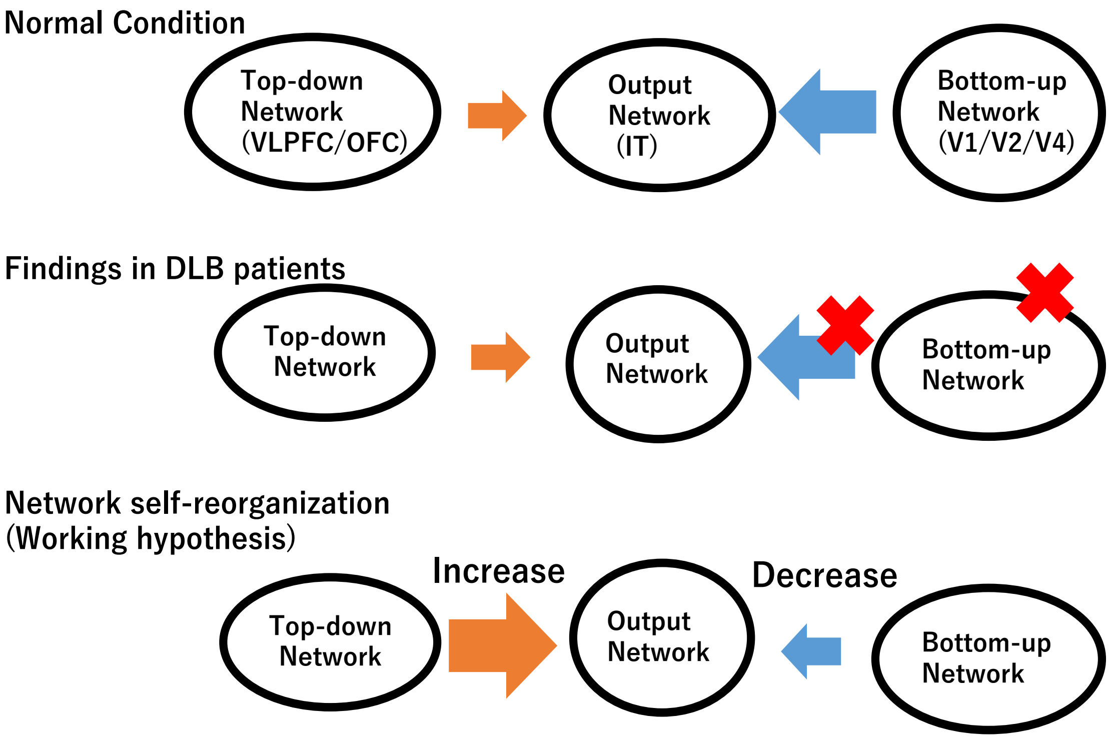

Figure 2.2: **Network self-reorganization hypothesis that explains CVH.** In this hypothesis, IT activities depend on bottom-up information, which reflects the external sensory information, in the normal condition. Then IT activities depend on top-down information due to the loss of bottom-up information and the subsequent self-reorganization process. Red cross marks represent damage in the corresponding parts. Blue and orange arrows represent the influence of bottom-up and top-down information. The thickness of an arrow represents the strength of these influences.

consequences of function [120].

One example of experience-dependent reorganization is an ocular dominance shift associated with V1 neurons after monocular deprivation [122]. Responses in V1 neurons to stimuli presented to the deprived eye are depressed despite responded before deprivation. The BCM rule was developed to explain this phenomenon [101, 123].

Another example is the phantom limb, in which amputees perceive their lost limbs [124, 125]. One famous scenario is the reorganization of the topographic map of the primary somatosensory cortex (S1) [126]. According to this scenario, hand amputation leads to the invasion of the face region adjacent to the hand region in S1, and face stimulation causes missing hand perception. However, nonfacial stimuli also produce lost arm perception [127, 128], suggesting a reorganization of some places through the afferent pathway leading to S1 [125].

The unique nature of this scenario is that the inputs that activate the neuron change with the reorganization, but the representation of the neuron remains [129, 130]. In fact, S1 activation by motor control [131] or microstimulation [132] have shown that representations persist over decades. The preservation of the representation after the reorganization is one of our assumptions in this study (see Section 3.4). Phantom perception in other modalities may also be a reorganization process associated with deafferentation [133].

Chapter 3

Method

In Section 3.1, we consider a self-organization principle as the maximization of mutual information. In Section 3.2, we formulate two types of self-reorganization process. In Section 3.3, we consider both bottom-up and top-down patterns with an assumption regarding conditioning effects such that top-down patterns co-occur with the corresponding bottom-up patterns during the learning process. In Section 3.4, we define a measure of CVH, which implies how much a bottom-up pattern is reconstructed from the output population activity when the corresponding top-down pattern is fed.

3.1 Information maximization with stochastic binary unit

We consider a simple neural model that consists of M input neurons, whose activity is described by an M-dimensional vector, $x \in \mathbb{R}^M$, and a single output neuron, for which the state is described by a binary value, $y \in \{0,1\} = \mathcal{Y}$. The membrane potential $v(x) \in \mathbb{R}$ is defined by the following equation.

$$v(x) = u(x) + \theta, \ \ u(x) = \sum_{j=1}^{M} x_j w_j, \tag{3.1}$$

where $w_j \in \mathbb{R}$ is the synaptic weight from j th input neuron to the output neuron and $\theta \in \mathbb{R}$ denotes the bias. The first term, $u(x) \in \mathbb{R}$ represents an activity from input x. The firing probability of the output neuron is defined by

$$p(y = 1|x) = f(v(x)), \tag{3.2}$$

where $f(v) = \frac{1}{1+exp(-v)}$ is a sigmoid function. Here, we consider a finite set of input patterns $\mathcal{X} = \{x^k\}_{k=1}^K$, where K is the number of input patterns. We simply denote here $u(x^k), v(x^k)$ by u^k, v^k, respectively. The probabilities of the occurence of input pattern x^k are denoted by $p(x^k)$. Mutual information between input and output patterns is defined by

$$I(\mathcal{X};\mathcal{Y}) = \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}, \tag{3.3}$$

where $p(x,y)$ denotes the joint probability distribution, and $p(y) = \sum_{x\in\mathcal{X}} p(x,y)$ denotes the marginal probability distribution. To derive the learning equation, we assume that the synaptic weights change to maximize the mutual information by the gradient method. Then, synaptic weights w_j evolve from initial weights $w_j(0)$ in the following way.

$$\tau \frac{dw_j(t)}{dt} = \frac{\partial I(\mathcal{X};\mathcal{Y})}{\partial w_j}, \quad (j = 1, ..., M) \tag{3.4}$$

where $\tau > 0$ is a time constant. This learning equation can be compared with conventional synaptic plasticity rule by calculating the gradient of mutual information (see Section 4.1). In numerical calculations of Eq. (3.4), we used the following discrete dynamical system.

$$w_j^{t+1} = w_j^t + \lambda \frac{\partial I(\mathcal{X};\mathcal{Y})}{\partial w_j}, \tag{3.5}$$

where $\lambda > 0$ is a learning rate.

For clarity, we define the preference for input patterns that gives rise to the definition of pattern selectivity for input patterns.

Definition 1. *For input pattern x^k, the output neuron is said to possess a preference if $v^k > 0$ and to possess no preference if $v^k \leq 0$. In the preference situation, the value "1" is allocated to the output neuron; otherwise, "0" is allocated. Then, the pattern selectivity for input patterns $\mathcal{X}$ is defined by the sequence of elements in $\{0, 1\}$.*

For example, if the output neuron has preference only for input pattern x^1, the pattern selectivity is expressed as "$10 \cdots 0$".

3.2 Self-reorganization process

We consider the situation that the network was damaged in some parts of synapses between input and output neurons, and then the network learning starts again from those damaged conditions. In the present study, this situation is called self-reorganization (SRO). The procedure for the computational study is divided into three phases: "self-organization" (SO) phase, "damage" phase, and "self-reorganization" (SRO) phase. At the SO phase, every synaptic weight $w_j(t)$ changes to follow the learning rule Eq. (3.4) from initial state $w_j(0)$ until a certain finite time T. At the damage phase after the SO phase, c connections among all are randomly selected and set to 0. This procedure is described by

$$w_j(T) \to 0, \qquad (\forall j \in \Omega), \tag{3.6}$$

where a set Ω consists of indices of selected connections. This manipulation is assumed to correspond to the synaptic loss in the physiological situation. At the SRO phase after the damage phase, synaptic weight $w_j(t)$ changes to follow the learning rule Eq. (3.4) until a certain finite time T'. Here, we consider two types of SRO cases depending on whether or not the damaged connections also change, referred to as "weakening" and "cutting", respectively. Thus, weakening and cutting are used as terms of the SRO phase in the present study. In particular, the latter procedure is described by

$$\tau \frac{dw_j(t)}{dt} = \frac{\partial I(\mathcal{X};\mathcal{Y})}{\partial w_j}, \qquad (\forall j \in \bar{\Omega},\ t > T), \tag{3.7}$$

where a set $\bar{\Omega}$ consists of indices of non-selected connections, that is $\bar{\Omega} = \{1, ..., M\} - \Omega$. Fig.3.1 shows the schematic diagrams in these four phases including two types for the SRO phase.

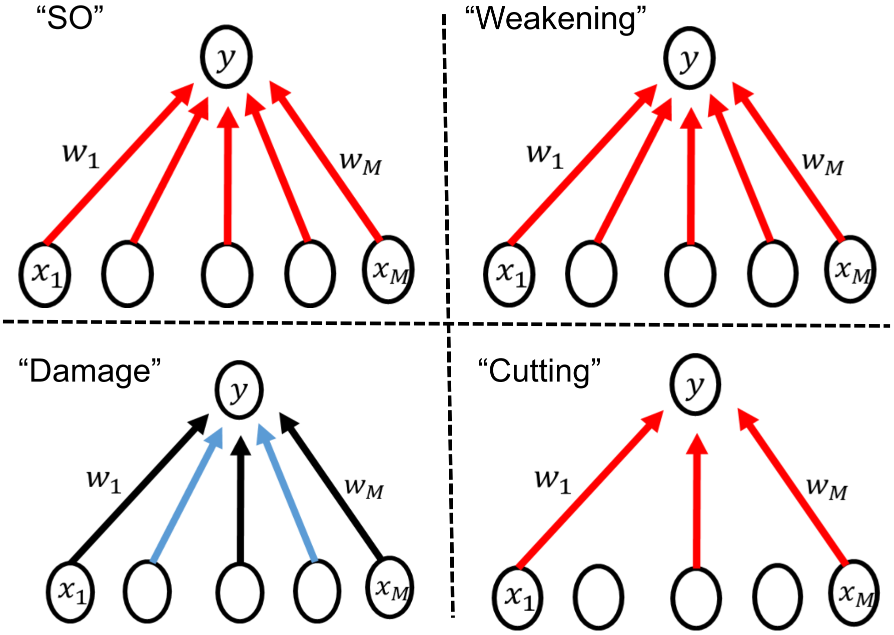

Figure 3.1: Four phases in the SRO procedure: "SO" phase, "damage" phase, "weakening" phase, and "cutting" phase. Red-colored connections between input and output neurons represent development to follow the learning rule. Blank connections represent non-development. Blue-colored connections denote damaged connections. Black-colored connections denote undamaged connections.

3.3 Bottom-up and top-down patterns

We consider two different networks, that is, bottom-up network α and top-down network β, as shown in Fig.3.2A. To consider the situation in which each network represents typical population activities, network α represents basis patterns $\{\alpha^l\}_{l=1}^L$, which have null elements $\alpha_j^l = 0$ for $j \in \{M_H+1, ..., 2M_H\}$. M_H is the number of input neurons in each network. Network β represents basis patterns $\{\beta^l\}_{l=1}^L$, which have null elements $\beta_j^l = 0$ for $j \in \{1, ..., M_H\}$. Here, L is the number of basis patterns. We assume such population activities of bottom-up network reflect information on visual objects such as people and animals, and ones of top-down network reflect information on the context of the visual scene to facilitate activation of the visual object. In anticipation of such effects, the following conditional effects were considered. Each bottom-up pattern appears alone, whereas each top-down pattern appears with only the corresponding bottom-up pattern during learning. Then, the input patterns are expressed by

$$\begin{cases} x^{2l-1} = \alpha^l + \beta^l & (l = 1, ..., L) \\ x^{2l} = \alpha^l & (l = 1, ..., L). \end{cases} \tag{3.8}$$

The example of these basis patterns and input patterns are expressed as image patterns in Fig.2.1B. These image patterns are treated as vectors.

To guarantee statistical robustness and symmetry, in addition to these image patterns, normalized patterns generated by a Gaussian distribution were taken into account in such a way that $\bar{\alpha}_j^l \sim N(0,1)$ for $j \in \{1, \cdots, M_H\}$, $\bar{\alpha}_j^l = 0$ for $j \in \{M_H+1, \cdots, 2M_H\}$ and $\alpha^l = \frac{\bar{\alpha}^l}{|\bar{\alpha}^l|}$, where $|\cdot|$ is a Euclidean norm and $N(0,1)$ denotes a Gaussian distribution with

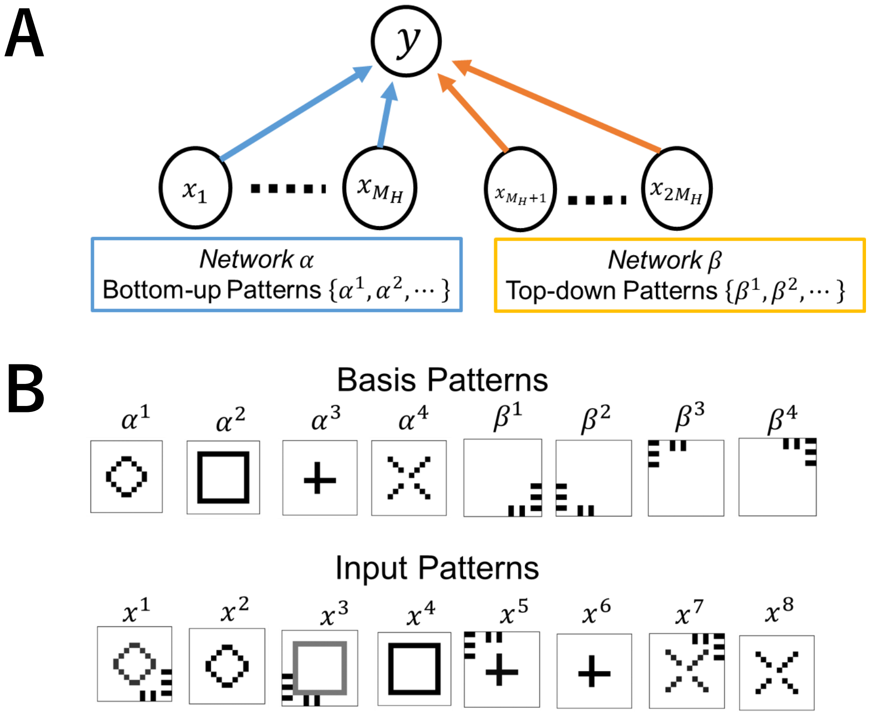
A
y
x_1
x_{M_H}
x_{M_H+1}
x_{2M_H}
Network α
Bottom-up Patterns $\{\alpha^1, \alpha^2, \cdots\}$
Network β
Top-down Patterns $\{\beta^1, \beta^2, \cdots\}$
B
Basis Patterns
α^1 α^2 α^3 α^4 β^1 β^2 β^3 β^4
Input Patterns
x^1 x^2 x^3 x^4 x^5 x^6 x^7 x^8

Figure 3.2: (A) Structure of network α, which is responsible for bottom-up patterns and network β, which is responsible for top-down patterns. (B) Image patterns with 14×14 pixels, which were used for the present study. The upper row represents basis patterns, where the left four patterns as bottom-up patterns, which represent the visual image, and the right four patterns as top-down patterns, which represent the indices of the context of the visual scene. The lower row indicates input patterns in the learning constructed by the visual images with and without contextual patterns. The central 10×10 pixels in each box were used for the representation of the activity of network α. The peripheral 96 pixels in each box and additional 4 zero entries were used for the representation of the activity of network β. Every white pixel has the value 0, and every black pixel has a positive constant value scaled to normalize the activity of each pattern. Here, patterns are orthogonalized with each other, because patterns are constructed to guarantee non-overlapping between any different patterns.

mean 0 and variance 1. Regarding the β^l, $\bar{\beta}^l_j \sim N(0,1)$ for $j \in \{M_H+1, \cdots, 2M_H\}$, $\bar{\beta}^l_j = 0$ for $j \in \{1,, \cdots, M_H\}$ and $\beta^l = \frac{\bar{\beta}^l}{|\bar{\beta}^l|}$.

3.4 Decoding the population activity

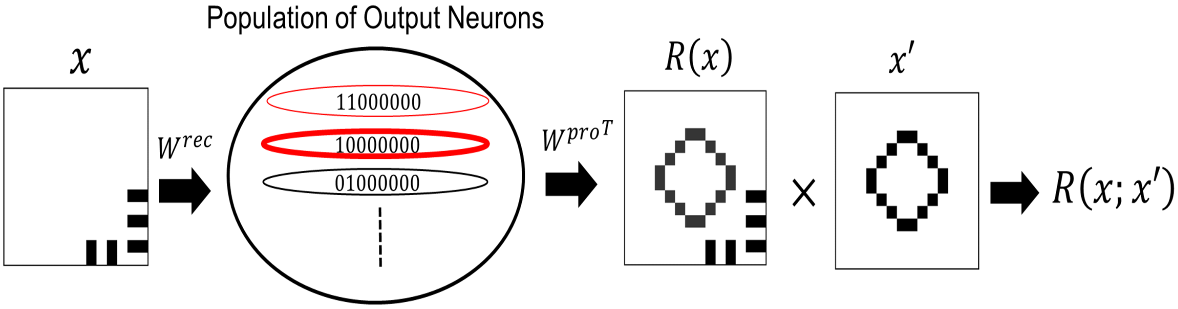

Figure 3.3: **A schematic diagram for the decoding method.** A pattern $R(x)$ is reconstructed from the activity of output population, which is composed of neurons specified by several types of pattern selectivity, when a pattern x is fed to input units. Several circles in the output population denote one example of the firing activity from each type of pattern selectivity. In this case, a thick red circle denotes high firing activity. A solid red circle denotes mild firing activity. A black circle denotes no firing activity. The agreement $R(x;x')$ is evaluated by the template matching between a reconstructed pattern $R(x)$ and a referenced pattern x'. Receptive field W^{rec} and projective field W^{pro} are introduced to reconstruct a pattern (details in the text).

In the present study, we distinguish between the "stimulus" that activates a certain neuron and the "stimulus" that the activity of the neuron represents. The former concept is referred to as receptive fields (RF) and the latter one as projective fields (PF) in this study. In order to deal with this situation computationally, the following procedure is considered and the overall schematic diagram is shown in Fig.3.3.

First, we consider the generation of a reconstructed pattern from the activity of the output population when pattern x is fed. Here, we introduce the concept of the receptive field (RF) $w_i^{rec} \in \mathbb{R}^M$ and the projective field (PF) $w_i^{pro} \in \mathbb{R}^M$ of output neuron i. Then, a reconstructed pattern $R(x) \in \mathbb{R}^M$ is defined by

$$R(x) = \sum_{i=1}^{N} f(v_i^{rec}(x)) w_i^{pro}, \tag{3.9}$$

where $v_i^{rec}(x) = \sum_{j=1}^{M} w_{ij}^{rec} x_j + \theta$, w_{ij}^{rec} is a synaptic weight from j th input neuron to i th output neuron.

To quantify the content of a reconstructed pattern, we employ a template matching procedure; that is, the agreement $R(x;x')$ between reconstructed pattern $R(x)$ and reference pattern x' is defined by

$$R(x;x') = R(x) \cdot x' = \sum_{i=1}^{N} r_i(x;x'), \tag{3.10}$$

where $R(x) \cdot x'$ is an inner product of $R(x)$ and x', and $r_i(x; x') = f(v_i^{rec}(x))u_i^{pro}(x')$ is the contribution of i th neuron to $R(x; x')$, where $u_i^{pro}(x') = \sum_{j=1}^{M} w_{ij}^{pro} x'_j$, w_{ij}^{pro} is a synaptic weight from i th output neuron to j th component of reconstructed pattern.

We examine how much single bottom-up pattern α^l is reconstructed when the corresponding top-down pattern β^l is fed. This measure is defined by $R(\beta^l; \alpha^l)$, referred to as the term "hallucination quality" (HQ). We are interested in how HQ changes in the SRO process. Regarding the RF, we consider the case in which the synaptic weights at that time are used after each set of four phases, SO, damage, weakening and cutting phase. However, regarding the PF, we compare two cases to determine whether synaptic weights at that time or synaptic weights after the SO phase are used after each set of four phases. In the former case, the neuron is considered to represent pattern selectivity at that point. In the latter case, the neuron represents pattern selectivity at pre-reorganization. The discrepancy between RF and PF might be a hallucination problem (see Section 4.1 or Subsection 2.4.3).

Chapter 4

Result

In Section 4.1, we discuss the fundamental mathematical properties of the proposed learning rule. In Section 4.2, we reformalize self-reorganization process. In Section 4.3, neurons that undergo a shift in activity from a bottom-up pattern to the corresponding top-down pattern after the occurrence of selective damage on the bottom-up network. Furthermore, in Section 4.4, this change can be explained in terms of state transitions. In Section 4.5, a single bottom-up pattern appears in the reconstructed pattern, which is generated by output population activity when the corresponding top-down pattern is fed with a specific projective field.

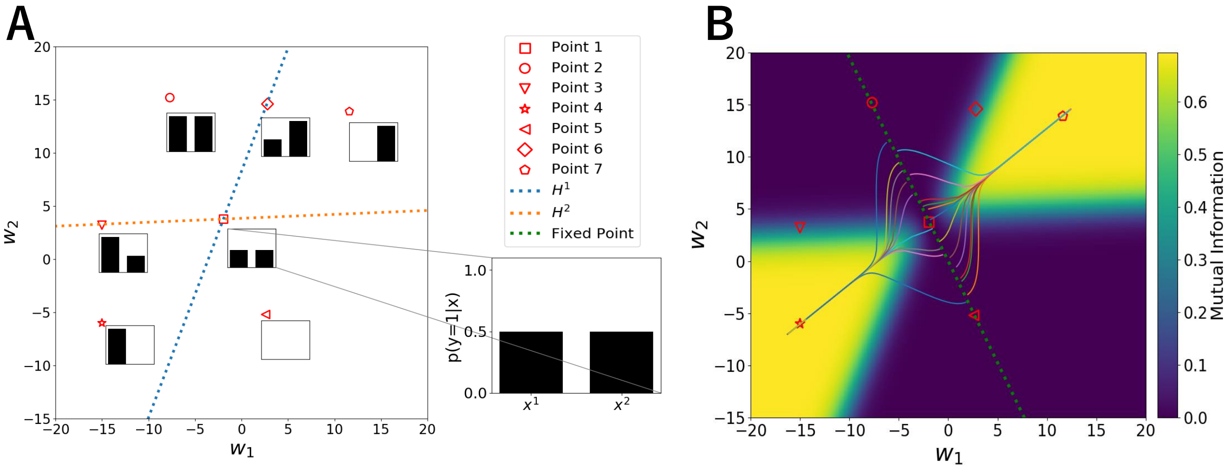

Figure 4.1: **Pattern selectivity and learning trajectories in two-dimensional weight space.** (A) Two hyperplanes H^1, H^2 (details in the text) are depicted for two input patterns x^1, x^2, respectively. As typical, this is shown in seven points, for which bar graphs demonstrate firing probabilities $p(y = 1|x^1)$ and $p(y = 1|x^2)$. (B) Quantity of mutual information is shown in a color bar. Dynamical trajectories starting from some initial conditions on weight space are drawn. Degenerate fixed point (details in the text) is depicted and same seven points as in (A) are illustrated.

4.1 Dynamics of synaptic weights

In understanding the relationship between synaptic weights and pattern selectivity of output neuron, hyperplane $H^k = \{w \in \mathbb{R}^M | v^k = 0\}$ for pattern x^k on weight space is an informative object. Fig.4.1A shows the firing probabilities $p(y = 1|x)$ at certain places of the two-dimensional weight space (inlet simply indicates a magnification of one figure) and two hyperplanes (yellow and blue lines) defined by a null membrane potential of the corresponding input pattern. There are four regions divided by two hyperplanes, each of which corresponds to one of the types of pattern selectivity, as the hyperplane is a boundary of the pattern selectivity by definition.

We can derive the following learning equation by calculating the gradient of mutual information Eq. (3.3).

$$\tau \frac{dw}{dt} = \sum_{k=1}^{K} x^k g_k(v) = Xg(v), \tag{4.1}$$

where $w = (w_1, ..., w_M)^T \in \mathbb{R}^M$, $X \in \mathbb{R}^{M\times K}$ is the matrix whose k th column is x^k, and $g(v) = (g_1(v), ..., g_K(v))^T$ is a vector-valued function. $g_k(v) = p(x^k)f'(v^k)(v^k - \Psi(v))$ is derived by using a formula $\log \frac{f(v)}{1-f(v)} = v$. $\Psi(v) = \log \frac{p(y=1)}{p(y=0)}$ is a meta-plasticity term that controls the overall firing activity such as the sliding threshold in the BCM rule [101] . In general, the conventional rate-based synaptic plasticity rule is expressed as the combination of the presynaptic neural activities x, the postsynaptic neural activity u, and the synaptic weights w (see Section 2.3 or [96]). This learning rule is therefore one of the rate-based synaptic plasticity rules.

Fig.4.1B shows the quantity of mutual information on weight space by providing the same two input patterns as in Fig.4.1A. Compared with the two figures, it is apparently seen that two regions of high mutual information (yellow regions) realize the pattern selectivity "10" or "01". Fig.4.1B also shows that every trajectory converges to either of these two regions. There is a degenerate fixed point of Eq. (4.1) (green line) that consists of an infinite set of fixed points. The reason for the degeneration is that Jacobian matrix at this fixed point has a zero eigenvalue. Convergence regions of trajectories are divided by this fixed point. Concerning this fixed point, the next proposition is proved.

Proposition 1. *If $K \leq M$ and input patterns are linearly independent, the fixed point of Eq. (4.1) is the set*

$$W_{FP} = \{w \in \mathbb{R}^M | X^T w = a\bar{1}, a \in \mathbb{R}, \bar{1} = (1, ..., 1)^T \in \mathbb{R}^K\} \tag{4.2}$$

, which is unstable.

This proof is given in Appendix A. In the case that $K > M$, W_{FP} is unstable and other fixed points appear depending on X.

However, this proposition does not address the pattern selectivity of an output neuron, which is established after learning. The next proposition is related to the number of types of pattern selectivity.

Proposition 2. *If $K \leq M$ and the input patterns are linearly independent, the number of convergence regions of Eq. (4.1) is $2^K - 2$. Each convergence region corresponds to one type of pattern selectivity.*

Even though this statement has not been proven yet, our numerical results strongly suggest that this proposition is true. Although the possible number of types of pattern selectivity is 2^K when $K \leq M$ and input patterns are linearly independent, two types of pattern selectivity "$00\cdots0$" and "$11\cdots1$" will not appear after learning. For example, there are no convergence regions related to "00" and "11" in Fig.4.1B.

4.2 Dynamical differences between weakening and cutting

Because the dynamics of w is not useful for considering the learning dynamics after the damages of connections, we rewrite Eq. (4.1) by multiplying X^T on both sides.

$$\tau\frac{dv}{dt} = X^T X g(v), \tag{4.3}$$

where $v = (v^1, ..., v^K)^T \in \mathbb{R}^K$ and (k, l)-element of $X^T X$ is $x^k \cdot x^l$. At the damage phase, v^k is also changed by Eq. (3.6) in the following way.

$$\bar{v}^k = \sum_{j\in\bar{\Omega}} x^k_j w_j + \theta, \tag{4.4}$$

where $\bar{v}^k$ implies v^k after damage. At the weakening phase, v changes to follow Eq. (4.3) from the point Eq. (4.4). However, at the cutting phase, it is necessary to modify Eq. (4.3) because only undamaged connections change. If we consider $\{j_l\}_{l=1}^{M-c} = \bar{\Omega}$, which is the set of indices of undamaged connections and $\bar{x}^k = (x^k_{j_1}, \cdots, x^k_{j_{M-c}})$, which is the k th input pattern affected by cutting connections, then v changes to follow the equation

$$\tau\frac{dv}{dt} = \bar{X}^T \bar{X} g(v), \tag{4.5}$$

where $\bar{X} \in \mathbb{R}^{(M-c)\times K}$ is the matrix whose k th column is $\bar{x}^k$. This implies that not only the initial points of trajectory but also the vector field itself changes at the cutting phase.

Concerning the effects of damage for $\bar{v}^k$ and $\bar{x}^k \cdot \bar{x}^l$, the next proposition is proved.

Proposition 3. *For every $c < M$, the expected value of $\bar{v}^k$ is*

$$E[\bar{v}^k] = (1 - \frac{c}{M})(v^k - \theta) + \theta. \tag{4.6}$$

For every $c < M$, the expected value of $\bar{x}^l \cdot \bar{x}^l$ is

$$E[\bar{x}^k \cdot \bar{x}^l] = (1 - \frac{c}{M})x^k \cdot x^l. \tag{4.7}$$

The proof is given in Appendix A. Therefore, the degree of damage linearly depends on the number of damaged connections c.

4.3 Over-compensation through SRO

Here, we consider the case of selective damage to the bottom-up network α for 100 output neurons which have a specific type of pattern selectivity, "11000000", "10000000", and "01000000" after the SO phase. We do not consider symmetrical types of pattern selectivity, such as "00110000" for "11000000" because they show similar change in the following results. Other types of pattern selectivity are not considered in the present study. To get an output neuron with a specific type of pattern selectivity after the SO, we set specific initial weights before the SO phase described in Appendix B. Parameters used for numerical simulation are described in Appendix B.

Fig.4.2A shows the change rate of pattern selectivity with respect to a parameter c, which indicates the number of the damaged connections. After the cutting phase, neurons, whose pattern selectivity changes from "11000000" to "10000000" increases with c. However, after the weakening phase, there is no such change. after the cutting phase, neurons, whose pattern selectivity changes from "01000000" to "00000000" appear for some large c. Only in this case, we need to discuss the delicate characteristics latent in the learning rule (see Appendix B).

Fig.4.2B shows the changes in activity u from each basis pattern $\{\alpha^l\}_{l=1}^4$ and $\{\beta^l\}_{l=1}^4$ with respect to c. After the damage phase, the activity $u(\alpha^l)$, which indicate the activity from bottom-up pattern α^l seems to change linearly with respect to c (top three figures in Fig.4.2B). Proposition 3 tells us about this change. After both SRO phases, neurons specified by "11000000" have a tendency that $u(\alpha^1)$ decreases and $u(\beta^1)$ increases with c, each of which indicates the activity from bottom-up pattern α^1 and top-down pattern β^1, respectively (two left figures from the bottom in Fig.4.2B). Moreover, significant discontinuous changes appear only after the cutting phase, caused by the neurons whose pattern selectivity changes from "11000000" to "10000000".

Fig.4.2C shows typical synaptic weights of output neurons after each phase. In particular, a neuron whose pattern selectivity does not change from "11000000" after the cutting phase has positively valued synaptic weights to bottom-up pattern α^1 (Fig.4.2C-i). A neuron whose pattern selectivity changes to "10000000" from "11000000" during the cutting phase has negatively valued synaptic weights to α^1 and larger positive valued synaptic weights to top-down pattern β^1 than ones after the SO phase (Fig.4.2C-ii). Because image patterns have only non-negative values, these results provide an intuitive explanation for activity u, which is calculated by the inner product between image pattern and synaptic weights.

To sum up, when the damage is small to some extent, the compensation mechanism works to recover the activation from the bottom-up pattern. However, when the damage is sufficiently large, the compensation mechanism works to reduce the activation from a bottom-up pattern and enhance the activation from the corresponding top-down pattern.

4.4 Over-compensation from the viewpoint of state transition

The above computation results can be understood by the framework discussed in Section 4.2. Here, we use normalized Gaussian input patterns. The procedure to obtain the following results is described in Appendix B, in which the most important point is that convergence property after SRO is characterized only by two variables (v^1, v^2), representing membrane potentials for input pattern x^1 and x^2. We numerically calculated SRO processes with the same parameters in Section 4.3.

Fig.4.3A shows the convergence regions for each pattern selectivity on phase space (v^1, v^2). These figures can be better understood by comparing with the change rate of pattern selectivity shown in Fig.4.3B. For example, when focusing on the SRO that occurred in neurons with pattern selectivity of "11000000" after damage to network α

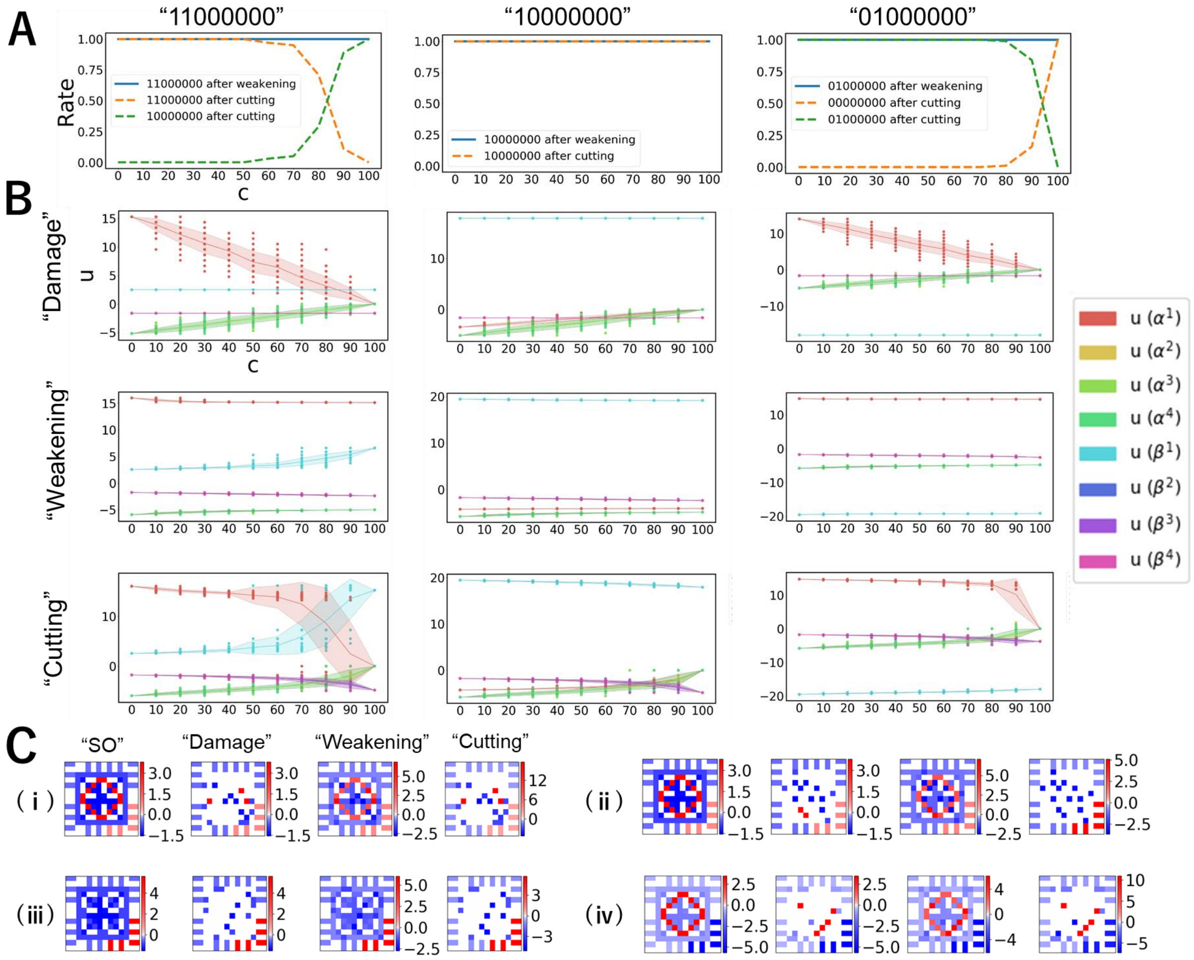

Figure 4.2: **Effect of SRO after damage to network α for specific type of pattern selectivity.** (A) For each c, the rate of 100 output neurons for which pattern selectivity has changed after each phase. Each column corresponds to "11000000","10000000", and "01000000" after the SO phase from left to right. (B) Each dot represents activity u from each pattern. Each line and shading region represents the mean and standard deviation of u respectively. Each column corresponds to the same pattern selectivity as in (A). Each row corresponds to three phases, "damage", "weakening", and" cutting" from top to bottom. (C) Typical synaptic weights of a specific output neuron at four phases when $c = 80$. (i) Synaptic weights of a neuron whose pattern selectivity remains "11000000" after the SO, weakening and cutting phase. (ii) Synaptic weights of a neuron whose pattern selectivity remains "11000000" after the SO and weakening phase, but changes to "10000000" after the cutting phase. (iii) Synaptic weights of a neuron whose pattern selectivity remains "10000000" after the SO, weakening and cutting phase. (iv) Synaptic weights of a neuron whose pattern selectivity remains "01000000" after the SO, weakening, and cutting phase.

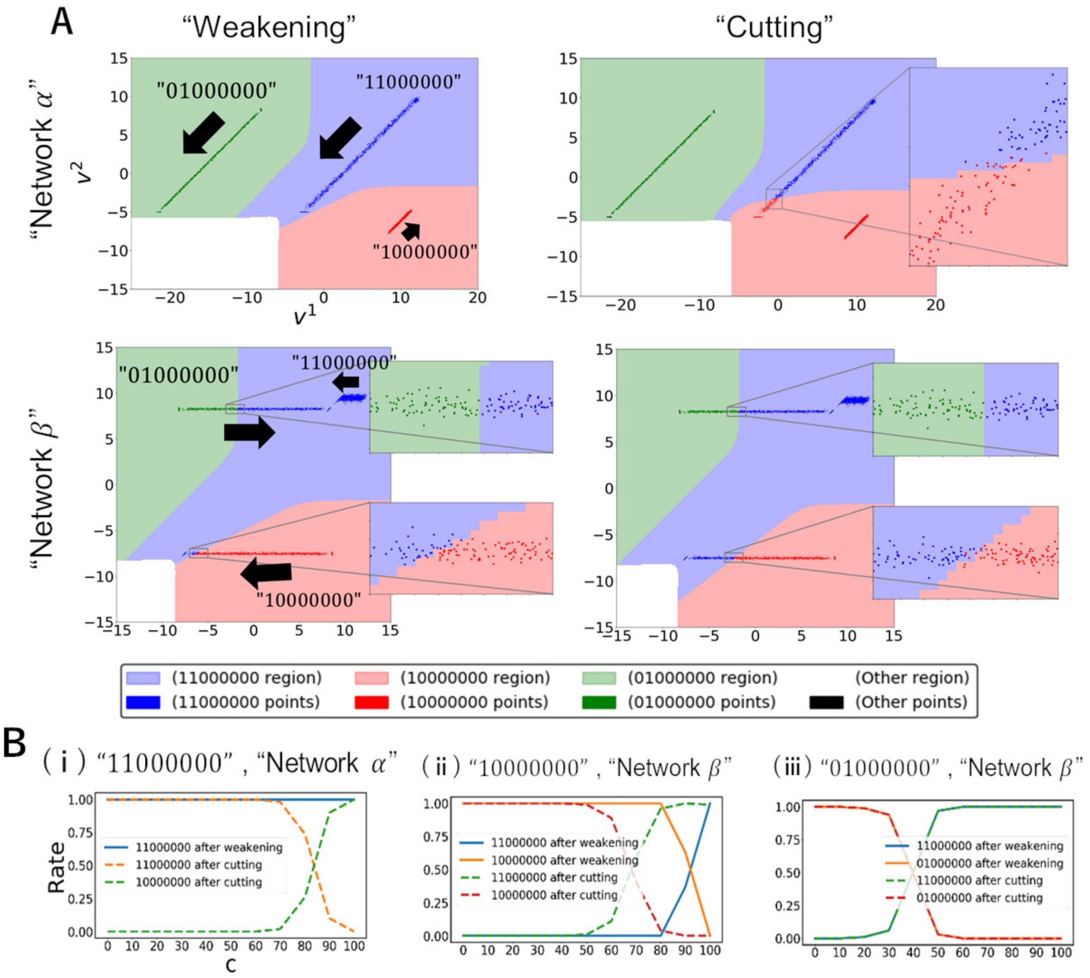

Figure 4.3: **Convergence regions on phase space constructed by membrane potential and the change rate for the specificity of pattern selectivity.** (A) On phase space (v^1, v^2), the convergence regions for the four types of pattern selectivity, ”11000000”, ”10000000”, ”01000000”, and others are shown. Values (v^1, v^2) for every neuron after the damage phase are shown as dots which are classified by color based on the pattern selectivity after SRO. The black arrows indicate the direction in which each dot moves after network damage. The left column indicates the case of “weakening”, and the right column shows the case of “cutting”. The upper row indicates the damage to network α, and the lower row demonstrates the damage to network β. (B) The change rate of pattern selectivity of some specific 100 output neurons for each c. (i) For neurons whose pattern selectivity is ”11000000” after damage to network α. (ii) For neurons whose pattern selectivity is ”10000000” after damage to network β. (iii) For neurons whose pattern selectivity is ”01000000” after damage to network β.

(dark blue dots in the upper two figures in Fig.4.3A), the states of the neuron go down from the upper right to lower left depending on parameter c. At the weakening phase, no dots cross the boundary of convergence region of pattern selectivity; thus, specificity via SRO never changes for all values of parameter c (Fig.4.3B- i). However, at the cutting phase, some dots cross the boundary (invasion from light blue to light red regions), which gives rise to change of pattern selectivity to "10000000" through the SRO processes. This is because convergence regions are altered in comparison with weakening phase due to the change in vector fields. The other cases are explained in the same way except for one case of "01000000", cutting phase, and damage to network α (black dots in the upper right in Fig.4.3A).

It is found that the change rate of pattern selectivity of neurons specified by "11000000" after damage to network α (Fig.4.3B- i) has the same tendency as those of the image patterns (left figure of Fig.4.2A), as the vector fields change in the same way with respect to c. We also found the same tendency to change in activity u from each pattern (data not shown).

From these observations, the mechanism of over-compensation can be understood in the following way. When the damage is small, v tends to fall inside the original convergence region, returning to the original state. Therefore, the original activation is recovered through SRO. However, when the damage is sufficiently large, on the other hand, v tends to fall outside the original convergence region, which brings about the deterioration of the original activity due to another pattern selectivity, thereby increasing the other activity through SRO.

4.5 Reconstructed patterns and contributions to change of HQ

First, we consider the reconstructed patterns from the population activity of 1200 output neurons, which are composed of 100 output neurons for each of the 12 types of pattern selectivity, "11000000", "10000000", "01000000", and other 8 symmetrical types of pattern selectivity, in the case of damage to network α. The same image patterns and parameters are used for the numerical simulation as in Section 4.3. Fig.4.4A shows an input pattern and four reconstructed patterns after each phase. In the first case of PF, in which synaptic weights after the SO phase are used as a projective field (PF), a bottom-up pattern α^l corresponding to the top-down pattern β^l is reconstructed after both SRO phases when $c = 90$ but is not reconstructed when $c = 30$. Thus, if the post-reorganization activity from the top-down pattern is read out as pre-reorganized activity with this specific PF, then the corresponding bottom-up pattern appears in the reconstructed pattern. In the second case of PF, in which synaptic weights at that time are used as PF, corresponding bottom-up patterns appear only after the weakening phase when $c = 90$.

Next, we consider hallucination quality (HQ), $R(\beta^l; \alpha^l)$, which is a measure of how much the bottom-up pattern α^l appears in the reconstruction pattern when the corresponding top-down pattern β^l is fed in the SRO process. Fig.4.4B shows a change in one of HQ, $R(\beta^1; \alpha^1)$ depending on c after each set of four phases (black lines of four different line styles). As expected, in the first case of PF, HQ increases with c after two SRO phases, whereas in the second case of PF, HQ increases only after the weakening phase. The difference in HQ after the cutting phase for different PFs is determined by whether there are null connections to network α or not. Interestingly, HQ after the damage phase never changes in the first case of PF as the activity from β^l, $u(\beta^l)$ does not change after damage to network α, and synaptic weights after the SO phase are used as PF after the damage phase.

It is possible to determine which types of pattern selectivity are responsible for HQ because it can be calculated by a linear sum of the contribution $r_i(\beta^l; \alpha^l)$ (see definition in the Section 3.4). Fig.4.4B also shows the amount of the contributions $r_i(\beta^1; \alpha^1)$ in neurons with a specific type of pattern selectivity. Only the change from "11000000" (red bars) contributes to the change of HQ with both PF cases. Therefore, as discussed in Section 4.3, changes in activities from a bottom-up pattern α^l to the corresponding top-down pattern β^l among neurons with pattern selectivity of "11000000" contribute to an increase in HQ.

One might assume that these results depend on the activation function of the output population, which is $f(v)$ in this case, or the variety of output neurons. In the first case of PF, it can be shown that the presence of the neuron whose activity shifts from the bottom-up pattern to the corresponding top-down pattern, such as the neuron specified by "11000000" contributes to the increase in HQ if the activation function is a strictly monotonically increasing function.

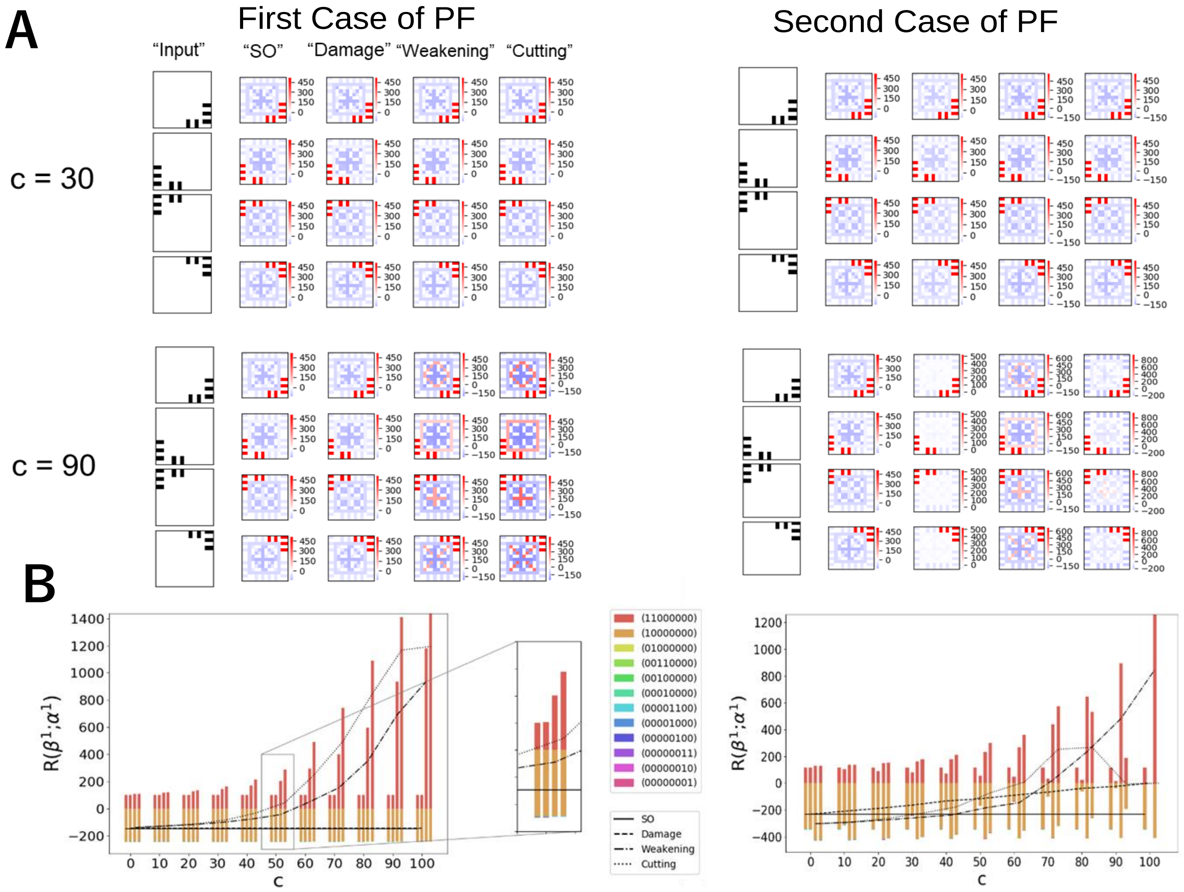

Figure 4.4: **Decoding the population activity from 12 types of output neurons when a top-down pattern is fed.** (A) Input pattern and reconstructed patterns after each set of four phases ”SO”, ”damage”, ”weakening”, and ”cutting”. The left column indicates the first case of the projective field (PF), in which synaptic weights after the SO phase are used as PF. The right column indicates the second case of PF, in which synaptic weights at that time are used. The upper row indicates the case in which $c = 30$. The lower row indicates the case in which $c = 90$. (B) One of ”hallucination quality” (HQ), $R(\beta^1; \alpha^1)$ depending on c, which is a measure of how much a bottom-up pattern α^1 is reconstructed. The black four lines represent HQ after each set of four phases. Each bar graph shows the amount of contribution $r_i(\beta^1; \alpha^1)$ in neurons with a specific type of pattern selectivity. The four bars on each c correspond to the four phases from left to right. The left and right column indicate the first case and the second case of PF, respectively. The sum of each color bar matches each $R(\beta^1; \alpha^1)$ (each black line), as $R(\beta^1; \alpha^1)$ is a linear sum of $r_i(\beta^1; \alpha^1)$.

Chapter 5

Discussion

5.1 Correspondence with clinical and neuroimaging findings

We used a computational model to examine the possibility that CVH in DLB occur when top-down information is used to compensate for the loss of bottom-up information. We examined how synaptic plasticity, including the plasticity of synapses in the top-down network, enables self-reorganization. We showed that the percentage of neurons which undergo a change in activity from the bottom-up pattern to the corresponding top-down pattern increased with the extent of synaptic loss through self-reorganization. Before reorganization, these neurons show sufficient increased activity for a particular bottom-up pattern and mild increased activity for the corresponding top-down pattern. After reorganization, in contrast, these neurons show decreased activity for the bottom-up pattern and sufficient increased activity for the top-down pattern. In general, a significant discontinuous shift in activity was observed, depending on the input patterns, the extent of damage, and the types of pattern selectivity, which can be understood in terms of state transitions.

This partly explains the results of neuroimaging during CVH in PD [9, 10] and eye disease [61, 62] by the occurrence of specific IT population activity since our computational results imply that, after reorganization, specific IT neurons show significantly increased activity triggered by the top-down input without bottom-up input. There is limited direct evidence that the source of IT activity comes from outside IT during CVH [64], including the co-occurrence increased activity in the frontal region [10, 62]. As indirect evidence, DLB patients tend to see CVH on the meaningless external environment, which is called pareidolias [11, 12]. This illusion might reflect the top-down facilitation with contextual indices of the background scene [16, 12]. Consistent with the decreased activity for the bottom-up pattern after reorganization, there are reports for decreased activity in the higher visual cortex for complex visual objects in DLB [76] and clinical assessment using silhouette visual images in PD [15].

This shift in activity before and after reorganization could give rise to CVH: the representation of the bottom-up pattern by IT neurons before reorganization is retained after reorganization, and a sensation of the bottom-up pattern is induced by IT neural activity evoked by the top-down pattern. In the present study, projective fields are characterized by the kind of pattern the neurons represent, rather than by their effects on downstream neurons as in the original computational study [134]. The discrepancy between the receptive and projective fields associated with reorganization has been discussed previously in the context of phantom limb sensations [130].

Assuming that the bottom-up pattern reflects external visual input and the top-down pattern reflects the internal representation of visual objects (e.g., expectation), our approach seems consistent with the misattribution model, which states that hallucinations are the misattribution of mental visual imagery to external visual input [15, 135]. Several studies support such a phenomenological mechanism in PD patients [15, 136]. The formation of mental imagery in the visual cortex has considerable overlap with the activity during perception [137]. However, because the electrophysiological properties of mental imagery at the cellular level are still unknown, it is difficult to determine whether our results reflect perception or mental imagery.

In DLB, α-synuclein aggregates have been located at presynapses; this abnormality may lead to the loss of postsynaptic dendrite spines [80]. White matter loss in the inferior longitudinal fasciculus or occipito-parietal regions might also be associated with this pathological finding [81, 82]. Although there is no evidence of synaptic plasticity as a compensatory mechanism in DLB per se, it does appear to play this role in many neurological conditions [138]. As indirect evidence of this, hyperconnectivity between the visual cortex and hippocampus or amygdala has been reported in patients with schizophrenia [139, 140], though not yet in those with DLB. From a theoretical viewpoint, as shown in the present study, synaptic damage can destabilize the dynamics of synaptic plasticity and prevent convergence to an original stable state.

It has been suggested that CVH are caused by a combined dysfunction of multiple distributed visual systems, particularly in DLB [2, 3, 4]. In the present study, we did not incorporate clinical or imaging findings, such as changes in attentional control or executive function [135, 64]. This is because it is not clear how these functions work at the cellular level during CVH. At the cellular level, one possibility is that the effect of gate modification on the firing rate of acetylcholine might influence hallucinations [141]. To address multiple dysfunction in the computational study, one obvious line of investigation would be to examine the interaction between functional modules and the effect of lesions [118]. The present study could serve as the basic framework for such a extension.

These experimental findings were referred not only from CVH in patients with DLB but also with other diseases or experience of hallucinations in other sensory modalities. If there is a common neural mechanism for hallucinations, our computational results provide new insights into data that have hitherto not been experimentally validated on patients with DLB. Furthermore, understanding the cognitive and neural mechanisms underlying the experience of hallucinations not only has considerable clinical significance but also provides insight into the veridical perception [142, 50]. To address this challenging problem, we expect that a computational approach is important and useful for generating hypotheses and getting an intuition of neural activities [14].

5.2 Computational implementation of CVH

To determine the fundamental mathematical structure of the mechanism of CVH, we adopted a rather abstract computational model in the present study. There is room for improvement in the treatment of the top-down mechanism and network architecture.

We have simply considered the top-down mechanism to be typical population activities, with top-down patterns representing the context of a visual scene to facilitate activation of the visual object [16, 36]. As a computational assumption, the top-down patterns co-occur with the corresponding bottom-up patterns during the learning process, but, as far as we know, related evidence is unclear.

A completely different approach for top-down architecture is Bayesian estimation [143]. For example, a Bayesian computational study of the false perception of DLB patients was found in the literature [116]. While such an approach has considerable explanatory power, especially with regard to functions such as expectations and anticipation, it is unclear that it can be tested experimentally [144].

Our network architecture could be extended in a variety of ways; however, the observational results would not necessarily be the same. For example, our results depended considerably on the characteristics of the convergence region of the learning rule (proposition 2). Therefore, the activity shift from a specific bottom-up pattern to the corresponding top-down pattern might not be obtained in the same setting with other local learning rules, such as the linear Hebbian rule with constraints [97] or the BCM rule [101].

Although computational models for synaptic plasticity have been studied for a long time [99], no single mechanism for the specific functions of the neural system has gained universal acceptance [104, 145]. For example, to maintain a cell assembly within a reccurent network, it is necassary to incorporate multiple factors as a synaptic plasticity [146, 147, 148]. Assuming CVH as the destabilized spontaneous activity within the IT population, such mechanisms might contribute to the generation. Several future directions of research are therefore possible, such as developing more physiologically realistic models, or studying exactly how plasticity affects self-reorganization. We expect other mathematical structures useful for explaining cognitive abnormality in mental disorder, not only CVH.

Appendix A

Proof of Proposition 1. At first, we can derive two formula by using the property of the sigmoid function. Here, we use the term $E[\cdot]$ as the mean of the distribution $p(x^k)$. For every $a \in \mathbb{R}$,

$$\Psi(a, ..., a) = a, \tag{5.1}$$

$$\frac{\partial \Psi}{\partial w_j}(a, ..., a) = E[x_j], \tag{5.2}$$

where $\Psi(v^1, \cdots, v^K) = \log \frac{\sum_k p(x^k) f(v^k)}{\sum_k p(x^k)(1 - f(v^k))}$.

Note that $Ker(X)$ is only the origin of $\mathbb{R}^K$ since input patterns $\{x^k\}_{k=1}^K$ are linearly independent. Then, we obtain the following relation about fixed point by using Eq. (5.1).

$$\begin{aligned} & g_k(v) = 0 \ \ (\forall k) \\ \iff & v^k = \Psi(v^1, ..., v^K) \ \ (\forall k) \\ \iff & v^1 = \cdots = v^K = \Psi(a, ..., a) = a \ \ (\forall a \in \mathbb{R}). \end{aligned} \tag{5.3}$$

Then, fixed point is the set $W_{FP} = \{w \in \mathbb{R}^M | X^T w + \Theta = a\bar{1}, a \in \mathbb{R}\}$.

Next, we will demonstrate the stability of this fixed point. Jacobi matrix $DF(w)$ of Eq. (4.1) is caluculated

$$DF(w) = XG, \tag{5.4}$$

where (j, k)-element of matrix G is $\frac{\partial g_k(v)}{\partial w_j}$. Using Eq. (5.2), this element at fixed point $w^* \in W_{FP}$ is

$$\frac{\partial g_k}{\partial w_j}(a, ..., a) = p(x^k) f'(a)(x_j^k - E[x_j]). \tag{5.5}$$

Then, Jacobi matrix at fixed point $w^* \in W_{FP}$ is

$$\begin{aligned} DF(w^*) &= f'(a) \begin{pmatrix} p(x^1)x_1^1 & p(x^2)x_1^2 & \cdots & p(x^K)x_1^K \\ p(x^1)x_2^1 & p(x^2)x_2^2 & \cdots & p(x^K)x_2^K \\ \vdots & \vdots & \ddots & \vdots \\ p(x^1)x_M^1 & p(x^2)x_M^2 & \cdots & p(x^K)x_M^K \end{pmatrix} \begin{pmatrix} x_1^1 - E[x_1] & x_2^1 - E[x_2] & \cdots & x_M^1 - E[x_M] \\ x_1^2 - E[x_1] & x_2^2 - E[x_2] & \cdots & x_M^2 - E[x_M] \\ \vdots & \vdots & \ddots & \vdots \\ x_1^K - E[x_1] & x_2^K - E[x_2] & \cdots & x_M^K - E[x_M] \end{pmatrix} \\ &= f'(a)E[(x - E[x])(x - E[x])^T] \end{aligned} \tag{5.6}$$

where $E[(x - E[x])(x - E[x])^T]$ is the covariance matrix of input patterns, which contains (i, j)-element as $E[(x_i x_j) - E[x_i]E[x_j]]$. Therefore, the Jacobi matrix is positive semi-definite at a fixed point.

□

Proof of Proposition 3. Consider the real values $\{t_j\}_{j=1}^M$, and the summation $T = \sum_j^M t_j$. Then, take c values $(t_{j_1}, t_{j_2}, ..., t_{j_c})$ randomly from these values and consider the summation

$T_c = \sum_l^c t_{j_l}$. Then if we represent the expected value of T_c as $E[T_c]$, we can show $E[T_c] = \frac{c}{M}T$.

At first, we show the following equation.

$$\sum_{j_1=1}^{M} \sum_{j_2=1, j_2 \neq j_1}^{M} \cdots \sum_{j_c=1, j_c \neq j_1, ..., j_{c-1}}^{M} t_{j_1} + t_{j_2} + \cdots + t_{j_c} = c_{M-1}\mathrm{P}_{c-1}T, \tag{5.7}$$

where ${}_M\mathrm{P}_c = \frac{M!}{(M-c)!}$ and $M!$ is the factorial of M. It can be shown by induction. When $c = 1$, $\sum_{j_1} t_{j_1} = T$. When $c + 1$,

$$\begin{aligned}
&\sum_{j_1=1}^{M} \sum_{j_2=1, j_2 \neq j_1}^{M} \cdots \sum_{j_{c+1}=1, j_c \neq j_1, ..., j_c}^{M} t_{j_1} + t_{j_2} + \cdots + t_{j_c} + t_{j_{c+1}} \\
&= \sum_{j_1=1}^{M} \sum_{j_2=1, j_2 \neq j_1}^{M} \cdots \sum_{j_c=1, j_c \neq j_1, ..., j_{c-1}}^{M} (M-c)(t_{j_1} + t_{j_2} + \cdots + t_{j_c}) + T - (t_{j_1} + t_{j_2} + \cdots + t_{j_c}) \\
&= (M-c-1) \sum_{j_1=1}^{M} \sum_{j_2=1, j_2 \neq j_1}^{M} \cdots \sum_{j_c=1, j_c \neq j_1, ..., j_{c-1}}^{M} (t_{j_1} + t_{j_2} + \cdots + t_{j_c}) + \sum_{j_1=1}^{M} \sum_{j_2=1, j_2 \neq j_1}^{M} \cdots \sum_{j_c=1, j_c \neq j_1, ..., j_{c-1}}^{M} T \\
&= (M-c-1)c_{M-1}\mathrm{P}_{c-1}T + {}_M\mathrm{P}_c T \\
&= (c+1)(M-c)_{M-1}\mathrm{P}_{c-1}T \\
&= (c+1)_{M-1}\mathrm{P}_c T.
\end{aligned}$$

Therefore, Eq. (5.7) holds. Using Eq. (5.7) and joint probability $p(t_{j_1}, ..., t_{j_c}) = \frac{1}{{}_M\mathrm{P}_c}$,

$$\begin{aligned}
E[T_c] &= \frac{1}{{}_M\mathrm{P}_c} \sum_{j_1=1}^{M} \sum_{j_2=1, j_2 \neq j_1}^{M} \cdots \sum_{j_c=1, j_c \neq j_1, ..., j_{c-1}}^{M} t_{j_1} + t_{j_2} + \cdots + t_{j_c} \\
&= \frac{c_{M-1}\mathrm{P}_{c-1}T}{{}_M\mathrm{P}_c} \\
&= \frac{c}{M}T.
\end{aligned} \tag{5.8}$$

Regarding v^k, assume $t_j = w_j x_j^k$. Regarding $x^k \cdot x^l$, assume $t_j = x_j^k x_j^l$. □

Appendix B

Parameters for numerical simulations

We set the synaptic weights to target initial values $v_{tar} \in \mathbb{R}^K$ to get an output neuron that has a specific type of pattern selectivity after the SO phase.

$$w(0) = X(X^TX)^{-1}v_{tar}, \tag{5.9}$$

where $(X^TX)^{-1}$ is the inverse matrix of X^TX. For each c and each pattern selectivity, 100 output neurons are generated by the setting following initial points, $v_{tar} = (0, 0, \theta, \cdots, \theta)$ for "11000000", $v_{tar} = (0, \theta, \cdots, \theta)$ for "10000000", $v_{tar} = (\theta, 5, \theta, \cdots, \theta)$ for "01000000" from Eq. (5.9). For symmetrical types of pattern selectivity, symmetrical initial target values are used. For example, considering "00110000" which can be considered as a symmetrical type of pattern selectivity with "11000000", $v_{tar} = (\theta, \theta, 0, 0, \theta, \cdots, \theta)$ which can be considered to be symmetrical values in which $v_{tar} = (0, 0, \theta, \cdots, \theta)$ are used. Other parameters are set to $M_H = 100, p(x^k) = 1/K, \lambda = 20.0, T = 500$, and $T' = 1000$.

Procedure for visualizing convergence regions

We consider the dynamics of membrane potentials discussed in Subsection 3.2. Although the actual dimension of v is $2L$, two dimensions are sufficient to characterize the SRO process for some pattern selectivity. Here, normalized Gaussian input patterns are considered. If M_H is sufficiently large, $\alpha^k \cdot \alpha^l \approx 0$ $(k \neq l)$ and $\beta^k \cdot \beta^l \approx 0$ $(k \neq l)$. Following this property, we consider the specific coefficient case.

$$X^TX = \begin{pmatrix} A & O & \cdots & O \\ O & A & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & A \end{pmatrix}, A = \begin{pmatrix} a_1 & a_3 \\ a_3 & a_2 \end{pmatrix}, \tag{5.10}$$

where every element of $O \in \mathbb{R}^{2\times 2}$ is zero and $a_1, a_2, a_3 \in \mathbb{R}$. Then, we consider the dynamics of v with coefficients Eq. (5.10) from initial points $v^1(0) = v_0^1, v^2(0) = v_0^2, v^3(0) = \cdots = v^{2L}(0) = v_0^3$. We can visualize convergence regions such as Fig.4.3A by classifying the initial points (v_1, v_2) based on the pattern selectivity after convergence with appropriate parameters such as v_0^3, a_1, a_2, a_3. In the case of weakening, it is sufficient to fix $a_1 = 2, a_2 = 1, a_3 = 1$ and adjust only v_0^3. However, in the case of cutting, since the value of the coefficient differs at each point, we adopt them near the boundary of the convergence regions.

"01000000" at the cutting phase

Note that the same coefficients Eq. (5.10) are used for image patterns as they have no overlap each other and are normalized. When network α is damaged and c is near to M_H, we can observe two SRO cases for a neuron with "01000000", which converges to one pattern except "00000000" or which converges to "00000000" after sufficient time T'. The former case is simply a problem of convergence time. The latter case is a problem of the relation between the zero coefficient, the order of v, and the role of Ψ.

In particular, when $c = M_H$, $v^{2l} = \theta$ and $v^{2l-1} < v^{2l}$ for each l since $u(\alpha^l) = 0$ and $u(\beta^l) < 0$ after the SO phase (see for example, figure after the damage phase with $c = 0$ of Fig.4.2B). Then, the order $v^{2l-1} < \Psi < v^{2l}$ is given. At first glance, v^{2l} seems to increase and v^{2l-1} seems to decrease owing to $g_{2l} > 0$ and $g_{2l-1} < 0$ at these points, but the coefficient $a_3 = 0$ makes $v^{2l}(t)$ remain at this point; thus, v^{2l-1} continues to decrease and v^{2l} remains at θ, such that pattern selectivity remains "00000000" indefinitely.

Acknowledgments

First of all, I would like to express my most significant appreciation to my supervisor Ichiro Tsuda at Chubu University for helpful advice and many opportunities. His financial support has also created an environment where I can focus on my research. These supports include MEXT KAKENHI Grant Number JP15H05878 (Non-linear Neuro-oscillology), Grant-in-Aid for Scientific Research (B) Grant Number JP17H01794, and JST CREST Grant Number JPMJCR17A4, Japan.

I would like to thank my supervisor Takao Namiki at Hokkaido University for helpful suggestions and supports in critical moments.

I would like to thank Mr. Watanabe at Chubu University for not only research discussion but also sharing issues in daily life. I would like to thank Dr. Seo Jong-Hyeon at Chubu University and Dr. Satoru Tadokoro at Hokkaido University for the helpful review.

I would like to thank Chubu University, especially people at Chubu University Academy of Emerging Sciences, for giving me a great environment of research. I would also express my appreciation to all the professors and staff of the department of mathematics at Hokkaido University. Especially, I would like to thank Prof. Hideo Kubo, who gives me opportunities for presentations in RIMS in Kyoto every year.

Finally, I would like to express my appreciation to my parents, who support me financially and mentally.

Bibliography

[1] William Beecher Scoville and Brenda Milner. Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, 20(1):11, 1957.

[2] Daniel Collerton, Elaine Perry, and Ian McKeith. Why people see things that are not there: A novel perception and attention deficit model for recurrent complex visual hallucinations. *Behavioral and Brain Sciences*, 28(6):737–757, 2005.

[3] Nico J Diederich, Christopher G Goetz, and Glenn T Stebbins. Repeated visual hallucinations in parkinson's disease as disturbed external/internal perceptions: focused review and a new integrative model. *Movement disorders: official journal of the Movement Disorder Society*, 20(2):130–140, 2005.

[4] James M Shine, Glenda M Halliday, Sharon L Naismith, and Simon JG Lewis. Visual misperceptions and hallucinations in parkinson's disease: dysfunction of attentional control networks? *Movement Disorders*, 26(12):2154–2159, 2011.

[5] Daniel Collerton, Urs P Mosimann, and Elaine K Perry. *The neuroscience of visual hallucinations*. Wiley Online Library, 2015.

[6] Stefania Pezzoli, Annachiara Cagnin, Oliver Bandmann, and Annalena Venneri. Structural and functional neuroimaging of visual hallucinations in lewy body disease: a systematic literature review. *Brain sciences*, 7(7):84, 2017.

[7] Michael J Firbank, Jim Lloyd, and John T O'Brien. The relationship between hallucinations and fdg-pet in dementia with lewy bodies. *Brain imaging and behavior*, 10(3):636–639, 2016.

[8] Camille Heitz, Vincent Noblet, Benjamin Cretin, Nathalie Philippi, Laurent Kremer, Mélanie Stackfleth, Fabrice Hubele, Jean Paul Armspach, Izzie Namer, and Frédéric Blanc. Neural correlates of visual hallucinations in dementia with lewy bodies. *Alzheimer's research & therapy*, 7(1):6, 2015.

[9] Kathy Dujardin, David Roman, Guillaume Baille, Delphine Pins, Stéphanie Lefebvre, Christine Delmaire, Luc Defebvre, and Renaud Jardri. What can we learn from fmri capture of visual hallucinations in parkinson's disease? *Brain imaging and behavior*, pages 1–7, 2019.

[10] Hiroshi Kataoka, Yoshiko Furiya, Masami Morikawa, Satoshi Ueno, and Makoto Inoue. Increased temporal blood flow associated with visual hallucinations in parkinson's disease with dementia. *Movement disorders: official journal of the Movement Disorder Society*, 23(3):464–465, 2008.

[11] Makoto Uchiyama, Yoshiyuki Nishio, Kayoko Yokoi, Kazumi Hirayama, Toru Imamura, Tatsuo Shimomura, and Etsuro Mori. Pareidolias: complex visual illusions in dementia with lewy bodies. *Brain*, 135(8):2458–2469, 2012.

[12] Kayoko Yokoi, Yoshiyuki Nishio, Makoto Uchiyama, Tatsuo Shimomura, Osamu Iizuka, and Etsuro Mori. Hallucinators find meaning in noises: pareidolic illusions in dementia with lewy bodies. *Neuropsychologia*, 56:245–254, 2014.

[13] Alan Robert Bowman, Vicki Bruce, Christopher J Colbourn, and Daniel Collerton. Compensatory shifts in visual perception are associated with hallucinations in lewy body disorders. *Cognitive Research: Principles and Implications*, 2(1):26, 2017.

[14] D. Collerton and et al. How can we see things that are not there?: Current insights into complex visual hallucinations. *Journal of Consciousness Studies*, 23(7-8):195–227, 2016.

[15] James Barnes, Laura Boubert, J Harris, A Lee, and Anthony S David. Reality monitoring and visual hallucinations in parkinson's disease. *Neuropsychologia*, 41(5):565–574, 2003.

[16] Moshe Bar. Visual objects in context. *Nature Reviews Neuroscience*, 5(8):617, 2004.

[17] Ichiro Tsuda, Yutaka Yamaguti, and Hiroshi Watanabe. Self-organization with constraints—a mathematical model for functional differentiation. *Entropy*, 18(3):74, 2016.

[18] Karl Friston and Stefan Kiebel. Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521):1211–1221, 2009.

[19] Daniel J Felleman and DC Essen Van. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 1(1):1–47, 1991.

[20] Mortimer Mishkin, Leslie G Ungerleider, and Kathleen A Macko. Object vision and spatial vision: two cortical pathways. *Trends in neurosciences*, 6:414–417, 1983.

[21] James J DiCarlo, Davide Zoccolan, and Nicole C Rust. How does the brain solve visual object recognition? *Neuron*, 73(3):415–434, 2012.

[22] Roozbeh Kiani, Hossein Esteky, Koorosh Mirpour, and Keiji Tanaka. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of neurophysiology*, 97(6):4296–4309, 2007.

[23] Sidney R Lehky and Keiji Tanaka. Neural representation for object recognition in inferotemporal cortex. *Current Opinion in Neurobiology*, 37:23–35, 2016.

[24] Doris Y Tsao and Margaret S Livingstone. Mechanisms of face perception. *Annu. Rev. Neurosci.*, 31:411–437, 2008.

[25] Keiji Tanaka. Inferotemporal cortex and object vision. *Annual review of neuroscience*, 19(1):109–139, 1996.

[26] Kazushige Tsunoda, Yukako Yamane, Makoto Nishizaki, and Manabu Tanifuji. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature neuroscience*, 4(8):832, 2001.

[27] Takayuki Sato, Go Uchida, Mark D Lescroart, Jun Kitazono, Masato Okada, and Manabu Tanifuji. Object representation in inferior temporal cortex is organized hierarchically in a mosaic-like structure. *Journal of Neuroscience*, 33(42):16642–16656, 2013.

[28] Nikolaus Kriegeskorte, Marieke Mur, Douglas A Ruff, Roozbeh Kiani, Jerzy Bodurka, Hossein Esteky, Keiji Tanaka, and Peter A Bandettini. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6):1126–1141, 2008.

[29] Robert Desimone and John Duncan. Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18(1):193–222, 1995.

[30] John H Reynolds, Leonardo Chelazzi, and Robert Desimone. Competitive mechanisms subserve attention in macaque areas v2 and v4. *Journal of Neuroscience*, 19(5):1736–1753, 1999.

[31] Geoffrey M Boynton. A framework for describing the effects of attention on visual responses. *Vision research*, 49(10):1129–1143, 2009.

[32] John H Reynolds and David J Heeger. The normalization model of attention. *Neuron*, 61(2):168–185, 2009.

[33] Timothy J Buschman and Earl K Miller. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *science*, 315(5820):1860–1862, 2007.

[34] Maurizio Corbetta and Gordon L Shulman. Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, 3(3):201, 2002.

[35] Rik Vandenberghe and Céline R Gillebert. Parcellation of parietal cortex: convergence between lesion-symptom mapping and mapping of the intact functioning brain. *Behavioural brain research*, 199(2):171–182, 2009.

[36] Christopher Summerfield and Tobias Egner. Expectation (and attention) in visual cognition. *Trends in cognitive sciences*, 13(9):403–409, 2009.

[37] S-I Higuchi and Yasushi Miyashita. Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. *Proceedings of the National Academy of Sciences*, 93(2):739–743, 1996.

[38] MJ Buckley and D Gaffan. Perirhinal cortex ablation impairs configural learning and paired–associate learning equally. *Neuropsychologia*, 36(6):535–546, 1998.

[39] Floris P de Lange, Micha Heilbron, and Peter Kok. How do expectations shape perception? *Trends in cognitive sciences*, 22(9):764–779, 2018.

[40] Travis Meyer and Carl R Olson. Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences*, 108(48):19401–19406, 2011.

[41] Moshe Bar and Elissa Aminoff. Cortical analysis of visual context. *Neuron*, 38(2):347–358, 2003.

[42] Russell A Epstein. Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in cognitive sciences*, 12(10):388–396, 2008.

[43] Elissa M Aminoff, Kestutis Kveraga, and Moshe Bar. The role of the parahippocampal cortex in cognition. *Trends in cognitive sciences*, 17(8):379–390, 2013.

[44] Moshe Bar, Karim S Kassam, Avniel Singh Ghuman, Jasmine Boshyan, Annette M Schmid, Anders M Dale, Matti S Hämäläinen, Ksenija Marinkovic, Daniel L Schacter, Bruce R Rosen, et al. Top-down facilitation of visual recognition. *Proceedings of the national academy of sciences*, 103(2):449–454, 2006.

[45] Maximilien Chaumon, Kestutis Kveraga, Lisa Feldman Barrett, and Moshe Bar. Visual predictions in the orbitofrontal cortex rely on associative content. *Cerebral cortex*, 24(11):2899–2907, 2013.

[46] Christopher Summerfield, Tobias Egner, Matthew Greene, Etienne Koechlin, Jennifer Mangels, and Joy Hirsch. Predictive codes for forthcoming perception in the frontal cortex. *Science*, 314(5803):1311–1314, 2006.

[47] Marco Onofrj, Astrid Thomas, Giovanni Martinotti, Francesca Anzellotti, Massimo Di Giannantonio, Fausta Ciccocioppo, and Laura Bonanni. The clinical associations of visual hallucinations. *The Neuroscience of Visual Hallucinations*, 5:91–117, 2015.

[48] Urs P Mosimann, Elise N Rowan, Cassie E Partington, Daniel Collerton, Elizabeth Littlewood, John T O'Brien, David J Burn, and Ian G McKeith. Characteristics of visual hallucinations in parkinson disease dementia and dementia with lewy bodies. *The American journal of geriatric psychiatry*, 14(2):153–160, 2006.

[49] Dag Aarsland, Clive Ballard, Jan P Larsen, and Ian McKeith. A comparative study of psychiatric symptoms in dementia with lewy bodies and parkinson's disease with and without dementia. *International journal of geriatric psychiatry*, 16(5):528–536, 2001.

[50] Flavie Waters, Daniel Collerton, Dominic H Ffytche, Renaud Jardri, Delphine Pins, Robert Dudley, Jan Dirk Blom, Urs Peter Mosimann, Frank Eperjesi, Stephen Ford, et al. Visual hallucinations in the psychosis spectrum and comparative information from neurodegenerative disorders and eye disease. *Schizophrenia bulletin*, 40(Suppl_4):S233–S245, 2014.

[51] G Jayakrishna Menon, Imran Rahman, Sharmila J Menon, and Gordon N Dutton. Complex visual hallucinations in the visually impaired: the charles bonnet syndrome. *Survey of ophthalmology*, 48(1):58–72, 2003.

[52] AM Santhouse, RJ Howard, and DH Ffytche. Visual hallucinatory syndromes and the anatomy of the visual brain. *Brain*, 123(10):2055–2064, 2000.

[53] Daniel Collerton, David Burn, Ian McKeith, and John O'Brien. Systematic review and meta-analysis show that dementia with lewy bodies is a visual-perceptual and attentional-executive dementia. *Dementia and geriatric cognitive disorders*, 16(4):229–237, 2003.

[54] Etsuro Mori, Tatsuo Shimomura, Misato Fujimori, Nobutsugu Hirono, Toru Imamura, Mamoru Hashimoto, Satoshi Tanimukai, Hiroaki Kazui, and Tokiji Hanihara. Visuoperceptual impairment in dementia with lewy bodies. *Archives of Neurology*, 57(4):489–493, 2000.

[55] Urs Peter Mosimann, George Mather, KA Wesnes, JT O'brien, DJ Burn, and IG McKeith. Visual perception in parkinson disease dementia and dementia with lewy bodies. *Neurology*, 63(11):2091–2096, 2004.

[56] Annachiara Cagnin, Francesca Gnoato, Nela Jelcic, Silvia Favaretto, Giulia Zarantonello, Mario Ermani, and Mauro Dam. Clinical and cognitive correlates of visual hallucinations in dementia with lewy bodies. *J Neurol Neurosurg Psychiatry*, 84(5):505–510, 2013.

[57] David A Gallagher, Laura Parkkinen, Sean S O'Sullivan, Alexander Spratt, Ameet Shah, Clare C Davey, Fion D Bremner, Tamas Revesz, David R Williams, Andrew J Lees, et al. Testing an aetiological model of visual hallucinations in parkinson's disease. *Brain*, 134(11):3299–3309, 2011.

[58] Dominic ffytche, Byron Creese, Marios Politis, K Ray Chaudhuri, Daniel Weintraub, Clive Ballard, Dag Aarsland, et al. The psychosis spectrum in parkinson disease. *Nature Reviews Neurology*, 13(2):81, 2017.

[59] Nico J Diederich, Glenn Stebbins, Christine Schiltz, and Christopher G Goetz. Are patients with parkinson's disease blind to blindsight? *Brain*, 137(6):1838–1849, 2014.

[60] Prabitha Urwyler, Tobias Nef, Alison Killen, Daniel Collerton, Alan Thomas, David Burn, Ian McKeith, and Urs Peter Mosimann. Visual complaints and visual hallucinations in parkinson's disease. *Parkinsonism & related disorders*, 20(3):318–322, 2014.

[61] DH ffytche, RJ Howard, MJ Brammer, A David, P Woodruff, S Williams, et al. The anatomy of conscious vision: an fmri study of visual hallucinations. *Nature neuroscience*, 1(8):738, 1998.

[62] Anne Marthe Meppelink, Bauke M de Jong, Johannes H van der Hoeven, and Teus van Laar. Lasting visual hallucinations in visual deprivation; fmri correlates and the influence of rtms. *Journal of Neurology, Neurosurgery & Psychiatry*, 81(11):1295–1296, 2010.

[63] Christopher G Goetz, Christina L Vaughan, Jennifer G Goldman, and Glenn T Stebbins. I finally see what you see: Parkinson's disease visual hallucinations captured with functional neuroimaging. *Movement Disorders*, 29(1):115–117, 2014.

[64] Anne Marthe Meppelink. Imaging in visual hallucinations. *The Neuroscience of Visual Hallucinations*, pages 151–166, 2014.

[65] AJ Harding, GA Broe, and GM Halliday. Visual hallucinations in lewy body disease relate to lewy bodies in the temporal lobe. *Brain*, 125(2):391–403, 2002.

[66] ME Kalaitzakis, LM Christian, LB Moran, MB Graeber, RKB Pearce, and SM Gentleman. Dementia and visual hallucinations associated with limbic pathology in parkinson's disease. *Parkinsonism & related disorders*, 15(3):196–204, 2009.

[67] Spiridon Papapetropoulos, Donald S McCorquodale, Jocely Gonzalez, Lucie Jean-Gilles, and Deborah C Mash. Cortical and amygdalar lewy body burden in parkinson's disease patients with visual hallucinations. *Parkinsonism & related disorders*, 12(4):253–256, 2006.

[68] Naroa Ibarretxe-Bilbao, Blanca Ramirez-Ruiz, Carme Junque, Maria Jose Marti, Francesc Valldeoriola, Nuria Bargallo, Silvia Juanes, and Eduardo Tolosa. Differential progression of brain atrophy in parkinson's disease with and without visual hallucinations. *Journal of Neurology, Neurosurgery & Psychiatry*, 81(6):650–657, 2010.

[69] B Ramírez-Ruiz, M-J Martí, E Tolosa, M Gimenez, N Bargallo, F Valldeoriola, and C Junque. Cerebral atrophy in parkinson's disease patients with visual hallucinations. *European journal of neurology*, 14(7):750–756, 2007.

[70] Soojeong Shin, Ji Eun Lee, Jin Yong Hong, Mun-Kyung Sunwoo, Young Ho Sohn, and Phil Hyu Lee. Neuroanatomical substrates of visual hallucinations in patients with non-demented parkinson's disease. *J Neurol Neurosurg Psychiatry*, 83(12):1155–1161, 2012.

[71] James Gratwicke, Marjan Jahanshahi, and Thomas Foltynie. Parkinson's disease dementia: a neural networks perspective. *Brain*, 138(6):1454–1476, 2015.

[72] Jennifer G Goldman, Glenn T Stebbins, Vy Dinh, Bryan Bernard, Doug Merkitch, Leyla deToledo Morrell, and Christopher G Goetz. Visuoperceptive region atrophy independent of cognitive status in patients with parkinson's disease with hallucinations. *Brain*, 137(3):849–859, 2014.

[73] N Oishi, F Udaka, M Kameyama, N Sawamoto, K Hashikawa, and H Fukuyama. Regional cerebral blood flow in parkinson disease with nonpsychotic visual hallucinations. *Neurology*, 65(11):1708–1715, 2005.

[74] Hideaki Matsui, Kazuto Nishinaka, Masaya Oda, Narihiro Hara, Kenichi Komatsu, Tamotsu Kubori, and Fukashi Udaka. Hypoperfusion of the visual pathway in parkinsonian patients with visual hallucinations. *Movement disorders: official journal of the Movement Disorder Society*, 21(12):2140–2144, 2006.

[75] Henning Boecker, Andres O Ceballos-Baumann, Dominik Volk, Bastian Conrad, Hans Forstl, and Peter Haussermann. Metabolic alterations in patients with parkinson disease and visual hallucinations. *Archives of Neurology*, 64(7):984–988, 2007.

[76] John-Paul Taylor, Michael J Firbank, Jiabao He, Nicola Barnett, Sarah Pearce, Anthea Livingstone, Quoc Vuong, Ian G McKeith, and John T O'Brien. Visual cortex in dementia with lewy bodies: magnetic resonance imaging study. *The British Journal of Psychiatry*, 200(6):491–498, 2012.

[77] GT Stebbins, CG Goetz, MC Carrillo, KJ Bangen, DA Turner, GH Glover, and JDE Gabrieli. Altered cortical visual processing in pd with hallucinations: an fmri study. *Neurology*, 63(8):1409–1416, 2004.

[78] Anne Marthe Meppelink, Bauke M de Jong, Remco Renken, Klaus L Leenders, Frans W Cornelissen, and Teus van Laar. Impaired visual processing preceding image recognition in parkinson's disease patients with visual hallucinations. *Brain*, 132(11):2980–2993, 2009.

[79] Ryoko Yamamoto, Eizo Iseki, Norio Murayama, Michiko Minegishi, Wami Marui, Takashi Togo, Omi Katsuse, Masanori Kato, Takeshi Iwatsubo, Kenji Kosaka, et al. Investigation of lewy pathology in the visual pathway of brains of dementia with lewy bodies. *Journal of the neurological sciences*, 246(1-2):95–101, 2006.

[80] Walter J Schulz-Schaeffer. The synaptic pathology of α-synuclein aggregation in dementia with lewy bodies, parkinson's disease and parkinson's disease dementia. *Acta neuropathologica*, 120(2):131–143, 2010.

[81] K Kantarci, R Avula, ML Senjem, AR Samikoglu, B Zhang, SD Weigand, SA Przybelski, HA Edmonson, P Vemuri, David S Knopman, et al. Dementia with lewy bodies and alzheimer disease: neurodegenerative patterns characterized by dti. *Neurology*, 74(22):1814–1821, 2010.

[82] Zuzana Nedelska, Christopher G Schwarz, Bradley F Boeve, Val J Lowe, Robert I Reid, Scott A Przybelski, Timothy G Lesnick, Jeffrey L Gunter, Matthew L Senjem, Tanis J Ferman, et al. White matter integrity in dementia with lewy bodies: a voxel-based analysis of diffusion tensor imaging. *Neurobiology of aging*, 36(6):2010–2017, 2015.

[83] EK Perry and RH Perry. Acetylcholine and hallucinations-disease-related compared to drug-induced alterations in human consciousness. *Brain and cognition*, 28(3):240–258, 1995.

[84] CF Lippa, TW Smith, and E Perry. Dementia with lewy bodies: choline acetyltransferase parallels nucleus basalis pathology. *Journal of Neural Transmission*, 106(5-6):525–535, 1999.

[85] Johannes C Klein, Carsten Eggers, Elke Kalbe, S Weisenbach, Carina Hohmann, Stefan Vollmar, S Baudrexel, NJ Diederich, Wolf-Dieter Heiss, and Rüdiger Hilker. Neurotransmitter changes in dementia with lewy bodies and parkinson disease dementia in vivo. *Neurology*, 74(11):885–892, 2010.

[86] H Shimada, S Hirano, H Shinotoh, A Aotsuka, K Sato, N Tanaka, T Ota, M Asahina, K Fukushi, S Kuwabara, et al. Mapping of brain acetylcholinesterase alterations in lewy body disease by pet. *Neurology*, 73(4):273–278, 2009.

[87] Peter Swann and John T O'Brien. Management of visual hallucinations in dementia and parkinson's disease. *International psychogeriatrics*, pages 1–22, 2018.

[88] Michael Goard and Yang Dan. Basal forebrain activation enhances cortical coding of natural scenes. *Nature neuroscience*, 12(11):1444, 2009.

[89] Shogo Soma, Satoshi Shimegi, Naofumi Suematsu, Hiroshi Tamura, and Hiromichi Sato. Modulation-specific and laminar-dependent effects of acetylcholine on visual responses in the rat primary visual cortex. *PLoS One*, 8(7):e68430, 2013.

[90] Charlene Moskovitz, Hamilton Moses, and Harold L Klawans. Levodopa-induced psychosis: a kindling phenomenon. *Am J Psychiatry*, 135(6):669–675, 1978.

[91] Christopher G Goetz, Caroline M Tanner, and Harold L Klawans. Pharmacology of hallucinations induced by long-term drug therapy. *The American journal of psychiatry*, 1982.

[92] S Holroyd, L Currie, and GF Wooten. Prospective study of hallucinations and delusions in parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 70(6):734–738, 2001.

[93] David R Williams and Andrew J Lees. Visual hallucinations in the diagnosis of idiopathic parkinson's disease: a retrospective autopsy study. *The Lancet Neurology*, 4(10):605–610, 2005.

[94] Gilles Fénelon, Christopher G Goetz, and Axel Karenberg. Hallucinations in parkinson disease in the prelevodopa era. *Neurology*, 66(1):93–98, 2006.

[95] Donald Olding Hebb. *The organization of behavior: A neuropsychological theory.* Psychology Press, 1949.

[96] Wulfram Gerstner and Werner M Kistler. Mathematical formulations of hebbian learning. *Biological cybernetics*, 87(5-6):404–415, 2002.

[97] Kenneth D Miller and David JC MacKay. The role of constraints in hebbian learning. *Neural Computation*, 6(1):100–126, 1994.

[98] Tara Keck, Taro Toyoizumi, Lu Chen, Brent Doiron, Daniel E Feldman, Kevin Fox, Wulfram Gerstner, Philip G Haydon, Mark Hübener, Hey-Kyoung Lee, et al. Integrating hebbian and homeostatic plasticity: the current state of the field and future research directions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1715):20160158, 2017.

[99] Chr Von der Malsburg. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14(2):85–100, 1973.

[100] Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of mathematical biology*, 15(3):267–273, 1982.

[101] Elie L Bienenstock, Leon N Cooper, and Paul W Munro. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2(1):32–48, 1982.

[102] Nathan Intrator and Leon N Cooper. Objective function formulation of the bcm theory of visual cortical plasticity: Statistical connections, stability conditions. *Neural Networks*, 5(1):3–17, 1992.

[103] Wickliffe C Abraham. Metaplasticity: tuning synapses and networks for plasticity. *Nature Reviews Neuroscience*, 9(5):387, 2008.

[104] Carlos SN Brito and Wulfram Gerstner. Nonlinear hebbian learning as a unifying principle in receptive field formation. *PLoS computational biology*, 12(9):e1005070, 2016.

[105] Takashi Kanamaru, Hiroshi Fujii, and Kazuyuki Aihara. Deformation of attractor landscape via cholinergic presynaptic modulations: a computational study using a phase neuron model. *PLoS One*, 8(1):e53854, 2013.

[106] Hiromichi Tsukada, Yutaka Yamaguti, and Ichiro Tsuda. Transitory memory retrieval in a biologically plausible neural network model. *Cognitive neurodynamics*, 7(5):409–416, 2013.

[107] David P Reichert, Peggy Series, and Amos J Storkey. Charles bonnet syndrome: evidence for a generative model in the cortex? *PLoS computational biology*, 9(7):e1003134, 2013.

[108] Nikolaus Kriegeskorte. Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual review of vision science*, 1:417–446, 2015.

[109] Alexander Mordvintsev, Christopher Olah, and Mike Tyka. Going deeper into neural networks. google research blog, 2015.

[110] Keisuke Suzuki, Warrick Roseboom, David J Schwartzman, and Anil K Seth. A deep-dream virtual reality platform for studying altered perceptual phenomenology. *Scientific reports*, 7(1):1–11, 2017.

[111] Karl J Friston. Hallucinations and perceptual inference. *Behavioral and Brain Sciences*, 28(6):764–766, 2005.

[112] Dirk De Ridder, Sven Vanneste, and Walter Freeman. The bayesian brain: phantom percepts resolve sensory uncertainty. *Neuroscience & Biobehavioral Reviews*, 44:4–15, 2014.

[113] Philip R Corlett, Guillermo Horga, Paul C Fletcher, Ben Alderson-Day, Katharina Schmack, and Albert R Powers III. Hallucinations and strong priors. *Trends in cognitive sciences*, 23(2):114–127, 2019.

[114] Albert R Powers, Christoph Mathys, and PR Corlett. Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science*, 357(6351):596–600, 2017.

[115] J Yu Angela and Peter Dayan. Acetylcholine in cortical inference. *Neural Networks*, 15(4-6):719–730, 2002.

[116] Thomas Parr, David A Benrimoh, Peter Vincent, and Karl J Friston. Precision and false perceptual inference. *Frontiers in integrative neuroscience*, 12, 2018.

[117] Hiroshi Fujii, Hiromichi Tsukada, Ichiro Tsuda, and Kazuyuki Aihara. Visual hallucinations in dementia with lewy bodies (i): a hodological view. In *Advances in Cognitive Neurodynamics (IV)*, pages 441–445. Springer, 2015.

[118] Hiromichi Tsukada, Hiroshi Fujii, Kazuyuki Aihara, and Ichiro Tsuda. Computational model of visual hallucination in dementia with lewy bodies. *Neural networks*, 62:73–82, 2015.

[119] J Ropero Peláez. Towards a neural network based therapy for hallucinatory disorders. *Neural networks*, 13(8-9):1047–1061, 2000.

[120] Steven C Cramer, Mriganka Sur, Bruce H Dobkin, Charles O'brien, Terence D Sanger, John Q Trojanowski, Judith M Rumsey, Ramona Hicks, Judy Cameron, Daofen Chen, et al. Harnessing neuroplasticity for clinical applications. *Brain*, 134(6):1591–1609, 2011.

[121] Karunesh Ganguly and Mu-ming Poo. Activity-dependent neural plasticity from bench to bedside. *Neuron*, 80(3):729–741, 2013.

[122] Torsten N Wiesel and David H Hubel. Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of neurophysiology*, 26(6):1003–1017, 1963.

[123] Leon N Cooper and Mark F Bear. The bcm theory of synapse modification at 30: interaction of theory with experiment. *Nature Reviews Neuroscience*, 13(11):798, 2012.

[124] Herta Flor, Lone Nikolajsen, and Troels Staehelin Jensen. Phantom limb pain: a case of maladaptive cns plasticity? *Nature reviews neuroscience*, 7(11):873–881, 2006.

[125] Tamar R Makin and Sliman J Bensmaia. Stability of sensory topographies in adult cortex. *Trends in cognitive sciences*, 21(3):195–204, 2017.

[126] Vilayanur S Ramachandran, Diane Rogers-Ramachandran, Marni Stewart, and Tim P Pons. Perceptual correlates of massive cortical reorganization. *SCIENCE-NEW YORK THEN WASHINGTON-*, 258:1159–1159, 1992.

[127] Stefan Knecht, Henning Henningsen, Thomas Elbert, Herta Flor, C Höhling, Christo Pantev, and Edward Taub. Reorganizational and perceptional changes after amputation. *Brain*, 119(4):1213–1219, 1996.

[128] SM Grüsser, W Mühlnickel, M Schaefer, K Villringer, C Christmann, C Koeppe, and H Flor. Remote activation of referred phantom sensation and cortical reorganization in human upper extremity amputees. *Experimental brain research*, 154(1):97–102, 2004.

[129] Gernot S Doetsch. Progressive changes in cutaneous trigger zones for sensation referred to a phantom hand: a case report and review with implications for cortical reorganization. *Somatosensory & motor research*, 14(1):6–16, 1997.

[130] Karen D Davis, Zelma HT Kiss, Lei Luo, Ronald R Tasker, Andres M Lozano, and Jonathan O Dostrovsky. Phantom sensations generated by thalamic microstimulation. *Nature*, 391(6665):385, 1998.

[131] Tamar R Makin, Jan Scholz, Nicola Filippini, David Henderson Slater, Irene Tracey, and Heidi Johansen-Berg. Phantom pain is associated with preserved structure and function in the former hand area. *Nature communications*, 4(1):1–8, 2013.

[132] Sharlene N Flesher, Jennifer L Collinger, Stephen T Foldes, Jeffrey M Weiss, John E Downey, Elizabeth C Tyler-Kabara, Sliman J Bensmaia, Andrew B Schwartz, Michael L Boninger, and Robert A Gaunt. Intracortical microstimulation of human somatosensory cortex. *Science translational medicine*, 8(361):361ra141–361ra141, 2016.

[133] Anusha Mohan and Sven Vanneste. Adaptive and maladaptive neural compensatory consequences of sensory deprivation—from a phantom percept perspective. *Progress in neurobiology*, 153:1–17, 2017.

[134] Sidney R Lehky and Terrence J Sejnowski. Network model of shape-from-shading: Neural function arises from both receptive and projective fields. *Nature*, 333(6172):452, 1988.

[135] Jim Barnes. Neuropsychological approaches to understanding visual hallucinations. *The Neuroscience of Visual Hallucinations*, pages 193–215, 2014.

[136] James M Shine, Rebecca Keogh, Claire O'Callaghan, Alana J Muller, Simon JG Lewis, and Joel Pearson. Imagine that: elevated sensory strength of mental imagery in individuals with parkinson's disease and visual hallucinations. *Proceedings of the Royal Society B: Biological Sciences*, 282(1798):20142047, 2015.

[137] Joel Pearson, Thomas Naselaris, Emily A Holmes, and Stephen M Kosslyn. Mental imagery: functional mechanisms and clinical applications. *Trends in cognitive sciences*, 19(10):590–602, 2015.

[138] H Duffau. Brain plasticity and tumors. In *Advances and technical standards in neurosurgery*, pages 3–33. Springer, 2008.

[139] Judith M Ford, Vanessa A Palzes, Brian J Roach, Steven G Potkin, Theo GM van Erp, Jessica A Turner, Bryon A Mueller, Vincent D Calhoun, Jim Voyvodic, Aysenil Belger, et al. Visual hallucinations are associated with hyperconnectivity between the amygdala and visual cortex in people with a diagnosis of schizophrenia. *Schizophrenia bulletin*, 41(1):223–232, 2014.

[140] A Amad, A Cachia, P Gorwood, D Pins, C Delmaire, B Rolland, M Mondino, P Thomas, and R Jardri. The multimodal connectivity of the hippocampal complex in auditory and visual hallucinations. *Molecular psychiatry*, 19(2):184, 2014.

[141] Alexander Thiele and Mark A Bellgrove. Neuromodulation of attention. *Neuron*, 97(4):769–785, 2018.

[142] Leor Zmigrod, Jane R Garrison, Joseph Carr, and Jon S Simons. The neural mechanisms of hallucinations: a quantitative meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, 69:113–123, 2016.

[143] Alan Yuille and Daniel Kersten. Vision as bayesian inference: analysis by synthesis? *Trends in cognitive sciences*, 10(7):301–308, 2006.

[144] Naoki Kogo and Chris Trengove. Is predictive coding theory articulated enough to be testable? *Frontiers in computational neuroscience*, 9:111, 2015.

[145] Friedemann Zenke, Wulfram Gerstner, and Surya Ganguli. The temporal paradox of hebbian learning and homeostatic plasticity. *Current opinion in neurobiology*, 43:166–176, 2017.

[146] Ashok Litwin-Kumar and Brent Doiron. Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature communications*, 5:5319, 2014.

[147] Friedemann Zenke, Everton J Agnes, and Wulfram Gerstner. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature communications*, 6:6922, 2015.

[148] James Humble, Kazuhiro Hiratsuka, Haruo Kasai, and Taro Toyoizumi. Intrinsic spine dynamics are critical for recurrent network learning in models with and without autism spectrum disorder. *bioRxiv*, page 525980, 2019.