
Prior experience modulates top-down predictive processing in the ventral visual areas

for the acquisition of the academic degree

doctor philosophiae (Dr. phil.)

presented to the council of the Faculty of Social and Behavioural Sciences

Friedrich-Schiller-University Jena

written by

Chenglin Li

Born on the 5th of January 1993 in Anhui, China



**FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA**

Biological Psychology and Cognitive Neurosciences

First examiner: Prof. Dr. Gyula Kovács

Second examiner: Prof. Dr. Stefan R. Schweinberger

Oral examination date: 11.07.2022

Table of Contents

Summary.....	i
Zusammenfassung	v
1 Introduction.....	1
1.1 Predictive coding theories	2
1.1.1 The predictive coding framework.....	2
1.1.2 Predictive processing in cognitive neuroscience	4
1.1.2.1 The predictive coding model of Rao and Ballard	4
1.1.2.2 Empirical evidence for the predictive coding in cognitive neuroscience..	8
1.2 Fundamental mechanisms of neuronal phenomena in visual processing and related models	10
1.2.1 The neuronal phenomena of repetition and expectation.....	10
1.2.2 Bottom-up models of repetition suppression	16
1.2.3 Repetition suppression as explained by the predictive coding model.....	19
1.3 Critical moderators of the prediction-related neural responses.....	22
1.3.1 Prior experiences with the stimuli	23
1.3.2 Stimulus familiarity and expectation.....	24
1.3.3 Attention.....	26
1.3.4 Short-term memory	28
1.4 Objectives and questions of the present thesis	30
2 General methods.....	33
2.1 Functional magnetic resonance imaging	33
2.2 Regions-of-interest in the ventral visual system	36
3 Empirical contributions	45
3.1 study I: Repetition probability effects for Chinese characters and German words in the visual word form area	47
3.2 study II: Short-term training attenuates the repetition probability effect for non-face objects	61
3.3 study III: 3a: Visual short-term memory load modulates repetition related	

fMRI signal adaptation	85
3b: Active attention counteracts the effects of increased load on	
visual short-term memory on repetition suppression.....	95
4 Discussion.....	103
4.1 Summary of the current studies	103
4.2 The role of experience in repetition and expectation suppression.....	105
4.3 The relationship between experience and repetition probability effect	108
4.4 The modulation of top-down prediction for repetition suppression	113
4.5 Limitations of the present thesis.....	116
4.6 Open questions and future directions.....	118
5. Conclusion	121
References	122
List of figures.....	149
List of tables.....	151
List of Abbreviations.....	152
Author contributions	155
Acknowledgements	157
Statutory declaration	159
Curriculum vitae	161

Summary

One of the most classic and well-established phenomena in cognitive neuroscience is repetition suppression (RS), a significant reduction of neural activities for repeated presentations of a given stimulus compared to its first presentation (Desimone, 1996). Four models, relying on a primary role of bottom-up processing, have been put forward to explain the neural mechanism underlying RS (*Fatigue model*, *Sharpening model*, *Facilitation model*, *Synchronization model*). However, recent studies found that RS can be better explained by the influential and popular predictive coding (PC) model (Rao and Ballard, 1999). Summerfield and colleagues (2008) firstly provided empirical evidence for explaining RS from the perspective of PC. They found the magnitude of RS in the Fusiform Face Area is affected by the probability of repetitions of faces, known as P(rep) effect, revealing that perceptual expectations modulate repetition-related processes.

Within the framework of PC, the brain constantly attempts to match the incoming sensory information to prior expectations, thereby minimizing their mismatch (prediction error) (Kok and de Lange, 2015). According to this theory, prior experiences about the sensory inputs in our brain is necessary to optimally achieve cognitive processes. However, it remains unclear how prior experiences modulate top-down predictive processes. To address this issue, the present thesis concentrates on the modulation of prior experiences on prediction-related neural effects (i.e., RS and P(rep) effects) associated with visual stimuli (i.e., faces, words, and objects).

In Study I (Li and Kovács, 2022), we asked the question of whether the P(rep) effect for words appears in the Visual Word Form Area (VWFA). Previous studies have observed this effect for various sensory stimulus categories, such as faces, objects and Roman letters (e.g., Grotheer and Kovács, 2014; Mayrhauser et al., 2014; Summerfield et al., 2008). Especially, Grotheer and Kovács (2014) suggested that this P(rep) effect depends on prior experiences of participants with the stimuli. In our Study I, we tested the extent

Summary

to which prior experiences affect the P(rep) effect could be extended to more complex lexical stimuli (i.e. Chinese characters and German words) as well. We measured the blood oxygen level dependent (BOLD) signal in the VWFA of native Chinese and German participants and estimated the P(rep) effects for Chinese characters and German words. The results showed that the P(rep) effect is only manifest for words of a language with which participants had extensive prior experiences. This further supports the idea that predictive processes, measured by P(rep) effect, require extensive prior experiences with the stimuli.

Study II focused on the modulation of short-term experience on the P(rep) effect for non-face objects. As mentioned above, the presence of the P(rep) effect depends on extensive experiences with the stimuli. However, it remains unknown whether a relatively shorter-term experience with the stimuli could also modulate the P(rep) effect. Study II addressed this question by performing fMRI measurements before and after a ten-day perceptual learning (PL) training for non-face objects (cars). The results replicated the P(rep) effect for faces as well as for cars. More interestingly, the P(rep) effect can be temporarily abolished by the short-term perceptual learning experience produced by the training, which suggests predictive processes can be modulated by short-term learning experience as well.

While the first two studies mainly focused on the modulation of prior experiences on the P(rep) effect, the third study investigated how prior experiences modulate bottom-up sensory inputs in the brain. According to previous studies, two potential cognitive processes (short-term memory (STM) and attention) could contribute to such a modulation. To further test this hypothesis, Study 3a (Li, Kovács, and Trapp, 2021) included a classic stimulus repetition paradigm to measure RS for faces, together with either concurrent STM load or a control condition. The results showed that RS is significantly attenuated when visual STM is loaded with other visual information, which suggests an important role of visual STM for predictive processes. In addition, Study 3b explicitly manipulates attention by introducing a face inversion detection task. In contrast

Summary

to the results in Study 3a, when participants attend to faces, the RS effect in the fusiform face area appears again in the STM loaded condition, which suggests that active attention can effectively counteract the reduction of RS effect due to increased visual STM load. Together, the results of Study III demonstrate that STM and attention play an active role in predictive processes as indexed by an expectation-related RS effect.

These results lead to the following conclusions: i) predictive processes, as measured by the P(rep) effect, require extensive prior experiences with stimulus categories, but ii) these can also be modulated by short-term learning experience. Further, iii) STM and attention are two important modulators of prior experiences on predictive processes.

Summary

Zusammenfassung

Repetition Suppression (RS; auf dt. etwa „Wiederholungsunterdrückung“) ist ein etabliertes Phänomen in den kognitiven Neurowissenschaften. Es beschreibt die signifikante Abnahme neuronaler Aktivität nach wiederholter Präsentation eines Stimulus im Vergleich zur Aktivität während der erstmaligen Präsentation (Desimone, 1996). Bisher wurden vier Modelle, welche alle auf bottom-up (= reizgesteuert) Prozessen beruhen, zur Erklärung der zugrundeliegenden neuronalen Mechanismen herangezogen (Fatigue-Modell, Sharpening-Modell, Facilitation-Modell, Synchronization-Modell). Neuere Studien deuten allerdings darauf hin, dass RS noch besser durch die einflussreiche „Predictive Coding“-Theorie (PC; Rao und Ballard, 1999) erklärt werden kann. Summerfield und Kollegen konnten bereits 2008 erste empirische Beweise für die Bedeutsamkeit von PC für RS vorlegen. Sie fanden heraus, dass das Ausmaß von RS im Fusiformen Gesichtsareal durch die Auftretenswahrscheinlichkeit von sich wiederholenden Gesichtsstimuli beeinflusst wurde. Dieses Ergebnis ist heute als P(rep)-Effekt bekannt und zeigt eindeutig, dass wahrnehmungsbezogene Erwartungen einen Einfluss auf wiederholungsbasierte Prozesse haben.

Eine Grundannahme der PC-Theorie ist, dass das Gehirn fortwährend ankommende Sinnesinformationen mit vorgefassten Erwartungen an diese vergleicht, und dabei versucht, die Differenz aus eingehender Information und Erwartung zu minimieren (Kok und de Lange, 2015). Diese Differenz wird auch Prediction Error (= Vorhersagefehler) genannt. Die Theorie besagt weiterhin, dass frühere Erfahrungen mit Sinnesinformationen für ein optimales Ablaufen dieser Prozesse notwendig sind. Es ist jedoch noch ungeklärt, wie genau sich frühere Erfahrungen auf top-down (= kognitiv gesteuert) Vorhersageprozesse auswirken. Die vorliegende Arbeit knüpft an diese Problematik an und untersucht die Beeinflussung von PC-bezogenen neuronalen Prozessen (RS und P(rep)) durch frühere Erfahrungen anhand bestimmter visueller Stimuli (Gesichter, Wörter und Objekte).

Zusammenfassung

In Studie 1 (Li und Kovács, 2022) wurde untersucht, ob der P(rep)-Effekt für Wörter im visuellen Wortformareal (VWFA, visual word form area) zu finden ist. In vorherigen Studien konnte den Effekt für verschiedene Stimuluskategorien beobachtet werden, u.a. Gesichter, Objekte und lateinische Buchstaben (Grotheer und Kovács, 2014; Mayrhofer et al., 2014; Summerfield et al., 2008). Besonders Grotheer und Kovács (2014) hoben hervor, dass der P(rep)-Effekt von vorangegangenen Erfahrungen der Proband:innen mit den entsprechenden Stimuli abhängig ist. Studie 1 untersucht daher, ob die Beeinflussung des P(rep)-Effekts durch vorangegangene Erfahrungen auch auf komplexere Stimuli (chinesische Schriftzeichen und deutsche Wörter) übertragen werden kann. Dafür haben wir den BOLD-Kontrast (BOLD = blood oxygen level dependent) im VWFA gemessen, während muttersprachlich chinesische oder deutsche Proband:innen sowohl chinesische Schriftzeichen als auch deutsche Wörter sahen. Die Ergebnisse bestätigen, dass der P(rep)-Effekt nur auftritt, wenn umfangreiche Erfahrungen mit der jeweiligen Sprache vorhanden waren. Vorhersageprozesse, welche durch den P(rep)-Effekt sichtbar gemacht werden, sind demnach von vorherigen Erfahrungen mit den jeweiligen Stimuli abhängig.

Studie 2 stellte die Modulation des P(rep)-Effekts durch vergleichsweise kurze Lernerfahrungen in den Mittelpunkt. Wie bereits erwähnt, hängt das Auftreten des P(rep)-Effekts von vorheriger, langfristiger Erfahrung mit den jeweiligen Stimuli ab. Es bleibt jedoch unklar, inwiefern auch relativ kurze Erfahrungen den Effekt beeinflussen können. In Studie 2 wurden fMRT-Messungen, bei welchen die Proband:innen Gesichter und Autos sahen, vor und nach einem 10-tägigen Training durchgeführt. Gegenstand des Trainings waren Objektstimuli (Autos), für welche die Wahrnehmung geschult werden sollte. Der P(rep)-Effekt konnte sowohl für Gesichter als auch für Autos gefunden werden. Interessanterweise führte das Training zu einer vorübergehenden Aufhebung des Effekts aufgrund der veränderten Erfahrung mit den Stimuli, was darauf schließen lässt, dass auch kurzfristig gemachte Erfahrungen einen Einfluss auf Vorhersageprozesse haben können.

Während sich die ersten beiden Studien mit dem Vorhandensein des Einflusses von früheren Erfahrungen auf den P(rep)-Effekt auseinandersetzten, untersuchte die dritte

Zusammenfassung

Studie die dafür zugrundeliegenden Mechanismen. Zwei kognitive Prozesse (Kurzzeitgedächtnis und Aufmerksamkeit) waren aufgrund früherer Studien besonders im Fokus. Studie 3a (Li, Kovács und Trapp, 2021) basiert auf einem klassischen Paradigma mit Stimulus-Wiederholung, um RS für Gesichter zu messen. Gleichzeitig wurde entweder das Kurzzeitgedächtnis beansprucht oder eine nicht-gedächtnisbezogene Kontrollaufgabe durchgeführt. Die Ergebnisse zeigen eindeutig, dass RS signifikant abgeschwächt wird, sobald im Kurzzeitgedächtnis konkurrierende visuelle Informationen vorhanden sind. Daher kann davon ausgegangen werden, dass das Kurzzeitgedächtnis ebenfalls für Vorhersageprozesse relevant ist. In Studie 3b wurden zusätzlich Aufmerksamkeitsaspekte mithilfe einer Entdeckungsaufgabe (invertierte Gesichter erkennen) untersucht. Im Gegensatz zu den Ergebnissen von Studie 3a taucht RS auch während der Beanspruchung des Kurzzeitgedächtnisses wieder auf, sobald die Aufmerksamkeit auf die Gesichter gerichtet wird. Aufmerksamkeit kann somit der Abschwächung von RS, welche durch Beanspruchung des Kurzzeitgedächtnisses zustande kommt, entgegenwirken. Insgesamt spielen sowohl Aufmerksamkeit als auch das Kurzzeitgedächtnis eine wesentliche Rolle in Vorhersageprozessen, wie durch den RS-Effekt bestätigt werden konnte.

Zusammengefasst lassen sich aus allen Ergebnissen nachstehende Schlussfolgerungen ableiten: 1) Vorhersageprozesse, die durch den P(rep)-Effekt gemessen werden, sind von umfangreichen vorangegangene Erfahrungen mit den betreffenden Stimuli abhängig, welche 2) allerdings auch durch kurzfristige Lernerfahrungen beeinflusst werden können. 3) Weiterhin wird der Zusammenhang zwischen Vorhersageprozessen und früheren Erfahrungen durch Aufmerksamkeit und das Kurzzeitgedächtnis moduliert.

1 Introduction

The working mechanisms of the human brain are amongst the most magical, complex, and intriguing, but incompletely explored topics in Science. Vision is an important way for human beings to explore the world since vision often dominates over the other sensory modalities (Stokes and Biggs, 2004). The human brain rapidly and effortlessly processes a large variety of visual information in daily life that is important for survival. An increasingly popular hypothesis is that the brain does not merely receive inputs passively, but actively tries to model how future events might look like. Rao and Ballard (1999) first described the theory of predictive coding (PC) in the visual cortex, which contributed to a whole new understanding of brain activity in general. Their model states that the brain constantly attempts to predict the sensory inputs based on previous experience, and to minimize the prediction error (the mismatch between predictions and sensory inputs) (Kok and de Lange, 2015). According to this theory, prior experiences about the sensory inputs in our brain are necessary for optimal cognitive processes. However, it remains unclear how prior experiences modulates the top-down predictive processes in the ventral visual areas. To address this issue, the present thesis concentrates on the modulation of prior experiences on prediction-related effects associated with visual stimuli (i.e., faces, words, and objects).

Therefore, the PC theories are presented in detail in chapter 1.1, including theoretical backgrounds and basic concepts (chapter 1.1.1). The chapter also integrates PC theories in the more general framework of cognitive neuroscience (chapter 1.1.2).

Chapter 1.2 will provide both a general review about the well-known neuronal phenomena (Repetition suppression, repetition enhancement, expectation suppression and repetition probability effects in chapter 1.2.1) and a description of related models (bottom-up models in chapter 1.2.2 and top-down prediction model in chapter 1.2.3).

In chapter 1.3, some important influencing factors on prediction-related neural responses

in perceptual studies are introduced. These include, for example, prior experiences with the stimuli (chapter 1.3.1), stimulus familiarity and expectation (chapter 1.3.2), attention (chapter 1.3.3) and short-term memory (chapter 1.3.4).

Finally, chapter 1.4, closing the introduction section, states the objectives and questions of the present dissertation.

1.1 Predictive coding theories

Predictive coding (PC) is the central theory of the current thesis. This chapter will introduce theoretical backgrounds and basic concepts of predictive coding and will describe some prediction-related studies in cognitive neuroscience. Please note that predictive coding is also referred to as predictive processing and the two expressions will be used interchangeably in this thesis (Clark, 2015).

1.1.1 The predictive coding framework

The concept of predictive processing has a long tradition in Psychology and it is often associated with higher cognitive functions such as problem solving, decision making or language processing (e.g., Bendixen et al., 2012; Feuerriegel et al., 2021a; Huettig, 2015). Yet this notion is widespread in research on perception as a seemingly lower cognitive function as well (Helmholtz, 1867; Rock, 1983). Over a century ago, Helmholtz described perception as a knowledge-driven inference process based on probabilities (Helmholtz, 1867). In other words, perception is determined not only by sensory inputs, but also by our previous experiences with the world.

Compared to the traditional view of sensory processing, the PC theory describes the brain as an inference machine, which is trying to estimate the probability of future events based on priors, rather than passively processing the events themselves. Similarly, the idea that prior information is necessarily taken into account when computing the probability of an event, is well-known in Bayesian statistics. Bayes' theorem, named after Thomas Bayes, defines how to calculate the conditional probability of an event (Bayes and Price, 1763).

Bayes' theorem is mathematically expressed as follows:

Equation 1

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

In Equation 1, the $P(A|B)$ indicates the conditional probability of event A occurring based on the given event B. Therefore, $P(A|B)$ is also called the posterior probability of A given B, meaning the probability of A after considering the prior probability of B. Mathematically, $P(A|B)$ is estimated by the prior probability $P(A)$ and the standardized likelihood $\frac{P(B|A)}{P(B)}$.

Bayesian inference refers to a process of probability computing, which estimates the posterior probability of a latent event based on given data. Bayes' rule is considered to be the fundamental principle of predictive coding. Under this framework, perception could be explained as a combination of sensory inputs and prior experiences, which occurs based on the above-mentioned formula.

Another important principle of perceptual prediction is optimization, which is related to the explanation of the free energy principle in neuroscience (Friston, 2009, 2010; Friston and Kiebel, 2009). The free energy principle claims that any adaptive change in the brain will minimize free energy (Friston, 2009). In the case of perception, this means that minimizing the difference between sensory inputs and prior knowledge is the best way to efficiently process sensory inputs.

Predictive coding integrates the principles of Bayesian inference and that of free energy to describe how the human brain processes information. Although several algorithms implementing the principles of predictive coding have been proposed (for a review, see Spratling, 2017), I will mainly focus on Rao and Ballard's algorithm (Rao and Ballard, 1999) and the free energy principle (Friston, 2009, 2010; Friston and Kiebel, 2009) in the following section.

1.1.2 Predictive processing in cognitive neuroscience

This subchapter will describe the predictive processing in cognitive neuroscience. The related evidence will mainly be summarized from the perspectives of Rao and Ballard's model (chapter 1.1.2.1) and considering empirical evidence obtained with several methods (e.g., single-cell recordings, EEG, MEG, and fMRI; chapter 1.1.2.2).

1.1.2.1 The predictive coding model of Rao and Ballard

In cognitive neuroscience, the traditional view of perceptual processing considers the brain as a passive, stimulus-driven machine that does not actively create meaning by itself, but simply reacts to sensory inputs and unidirectionally transfers information from the lower-level cortex to the higher-level cortex. Accordingly, the brain is considered to be structured in hierarchically organized neural architectures, which processes information in a serial, 'bottom-up' fashion (Biederman, 1987; Hubel and Wiesel, 1965; Marr, 1982; Thorpe et al., 1996). However, an increasing number of studies indicate that the brain should be considered as a more active, adaptive, and bidirectional system (Churchland et al., 1994; for a review see Engel et al., 2001).

Rao and Ballard, whose work was deemed to become a cornerstone in this field, were the first describing how predictive coding occurs at the neural level. In their seminal paper (Rao and Ballard, 1999), they interpreted the underlying mechanism of an extra-classical receptive-field effect as an instance of predictive coding. They designed a bidirectional and hierarchical network performing natural scene processing, which implemented computational algorithms that would allow it to generate an internal model of its stimulus inputs. Importantly, compared to the previous Bayesian coding hypothesis (Knill and Pouget, 2004), this network displayed a hierarchical structure, which is an important feature of predictive coding.

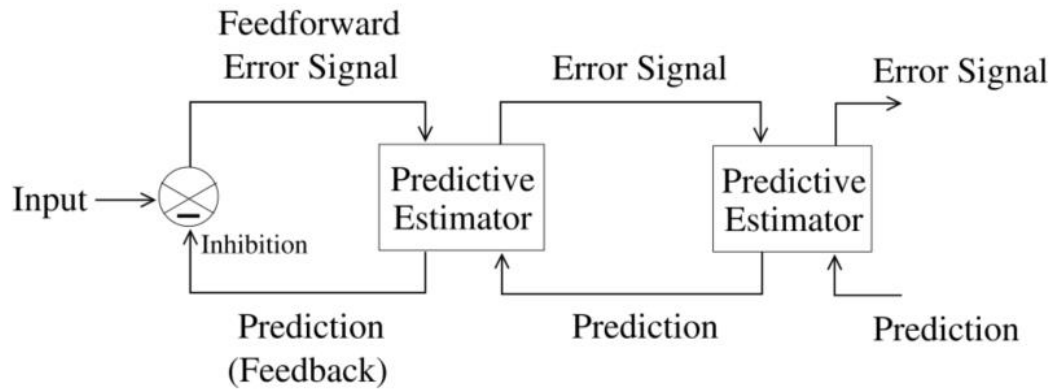


Figure 1. Schematic illustration of the hierarchical network for predictive coding with feedback and feedforward information flow. At each hierarchical level, feedback pathways carry predictions of neural activity at the lower level, whereas feedforward pathways carry residual errors between the predictions and actual neural activity. These errors are used by the predictive estimator (PE) at each level to correct its current estimate of the input signal and generate the next prediction. From Rao and Ballard (1999) with permission.

According to their predictive coding model (Figure 1 on page 5), perception is described as an iterative matching procedure of top-down predictions and bottom-up signal errors in a hierarchical system. This model is used to constantly predict sensory inputs and is continuously updated to match the inputs and their respective predictions. The sensory input itself represents the entry-level of the prediction system, then feedforward connections carry information about the actual input to the upper level in this hierarchical system. These bottom-up inputs are compared with top-down predictions (estimated in the predictive estimator). The prediction error resulting from this comparison signal is generated and used to compute updated predictions. Afterwards, the feedback connections transmit the new predictions from the higher- to the lower-order visual cortical areas, whereas the new prediction error is carried to the next upper level via feedforward connections. In other words, predictions and prediction errors are created in higher cortical areas when bottom-up inputs meet prior knowledge, then prediction and prediction error bidirectionally modulate perceptual processing. Optimally,

Introduction

the sensory signal will be fully ‘explained away’, when it is exactly matched with the top-down prediction (Clark, 2013; Friston, 2005).

According to this view, each level in the cortical hierarchy represents both predictions from the upper level and prediction error from the lower level (except the entry-level, which is the visual input itself). It is assumed that there are two categories of sub-neuronal populations: representation units and error units (Figure 2 on page 7). Representation units reflect the expected activities based on a hypothesis about forthcoming inputs and send the results to error units in the same level and the next lower level, while error units estimate the mismatch between the prediction and sensory inputs (prediction errors). In the meantime, the results are transmitted to representation units in the same level and the next upper level in this hierarchical prediction network. Several empirical studies support the separation of prediction and prediction error (e.g., Keller et al., 2018; Wacongne et al., 2012; for meta-analysis see Ficco et al., 2021; Siman-Tov et al., 2019). Linking to Helmholtz’s view, predictions would be mainly generated based on prior knowledge, while the size of prediction errors could depend on the precision of the prediction and the sensory inputs (e.g., Feldman and Friston, 2010).

At a more fine-grained level than the mere distinction between representation units and error units, it is thought that different neuronal populations perform different tasks in these loops (Bastos et al., 2012). For instance, deep pyramidal cells (representation units) send the predictions out and superficial pyramidal cells (error units) signal the prediction error (Feldman and Friston, 2010; Grother and Kovács, 2016). Similar feedforward and feedback loops continue throughout the sensory cortices to high-level areas, such as the FFA (Egner et al., 2010) or even the frontal cortex (Summerfield et al., 2006). The ultimate goal of these loops is to decrease the prediction error, in line with the optimization principle of free energy in Karl Friston’s unified brain theory (Friston, 2009, 2010).

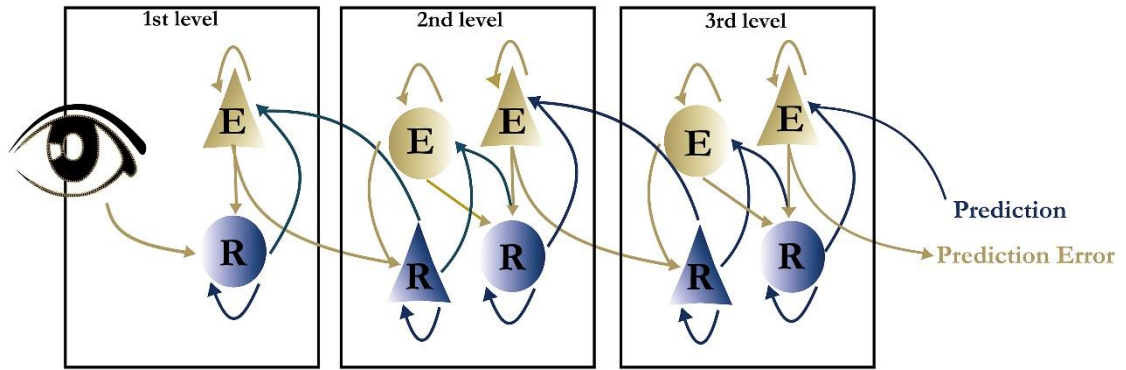


Figure 2. Illustration of the predictive coding model. The visual cortex is considered as a hierarchical structure (here consisting of three levels) containing feedback and feedforward loops. To optimize perception, deep pyramidal cells (representation units (R); blue circles and triangles) send out predictions about forthcoming perception, and superficial pyramidal cells (error units (E); gold circles and triangles) return the prediction error, the mismatch between the received predictions and the sensory input. Adapted from Grotheer and Kovács (2016), with permission.

What is the most important concept when we describe predictive coding under the free-energy principle (for a review see Friston, 2010)? It is optimization. The free-energy principle claims that “any self-organizing system that is at equilibrium with its environment must minimize its free energy” (Friston, 2010, p. 127). In other words, minimizing the prediction errors under the predictive framework can maximize the efficiency of cognitive processes. Friston and Kiebel (2009) demonstrated that predictive coding is in line with the free energy principle from a modelling computation perspective.

In summary, the view of a more active, adaptive, and bidirectional brain is substantiated in Rao and Ballard’s PC model (Rao and Ballard, 1999), which suggests that the information transmission in the sensory cortex adhere to a hierarchical and bidirectional network and follow the free energy principle (Friston, 2009, 2010; Friston et al., 2006; Friston and Kiebel, 2009).

1.1.2.2 Empirical evidence for the predictive coding in cognitive neuroscience

Perception has been modeled in a hierarchical and bidirectional network in Rao and Ballard's PC model (Rao and Ballard, 1999). It is the result of matching the bottom-up sensory inputs to prior expectation and minimizing their mismatch. This subchapter will describe the empirical evidence that supports this PC model using single-cell recordings and functional magnetic resonance imaging (fMRI) in monkeys and in human fMRI, electroencephalographic (EEG), and magnetoencephalographic (MEG) studies (for a review see Egner and Summerfield, 2013).

Early non-human primate studies revealed that the visual cortex is indeed hierarchically organized and highly interconnected, which is required for predictive coding (Felleman et al., 1991; Zeki and Shipp, 1988). Further, functional neuroimaging studies (e.g., Alink et al., 2010; den Ouden et al., 2010; Egner et al., 2010) and single-cell recordings in monkeys (Meyer and Olson, 2011) found stronger neuronal responses in the occipital cortex induced by the surprising stimuli relative to expected stimuli. For instance, Meyer and Olson (2011) trained macaque monkeys to learn six pairs of images, in which the second image can be predicted based on the first image in each pair. The neuronal activity in the inferotemporal (IT) cortex was reduced when the second image of the paired stimuli was predicted, as compared to when it was unpredicted. This modulation of top-down prior expectation in the perceptual processing supports the idea that predictive processes exist in non-human primates.

Furthermore, empirical evidence has also been reported in human studies in, which the probability of stimuli is manipulated. For instance, Wacongne and colleagues (2011) adopted a hierarchical auditory novelty paradigm and recorded human EEG and MEG responses to the fifth auditory stimulus in expected (xxxxx), mismatching (xxxxY) or omitted (xxxx) stimulus sequences. The probability of these three sequences was different in three types of blocks (expected (xxxxx), mismatching (xxxxY), and omitted (xxxx) blocks). Importantly, omitting the fifth tone should reflect the brain's hierarchical predictions. The rationale here is that, when a deviant tone is expected, its omission

represents a violation of two predictions: a local prediction of a tone plus a hierarchically higher expectation of its deviancy. The results showed the strongest mismatching response in the 'xxxxY' block, which was larger than that in the 'xxxxx' block. In addition, an omission of the fifth tone induced a greater prediction error than when a standard tone is expected. This provided evidence for a predictive and hierarchical system underlying the brain's response to auditory stimuli. Further fMRI studies support this idea. For instance, Egner et al. (2010) used three colored frames to cue participants to predict upcoming stimulus category (faces or houses). The maximal difference of brain activities for faces and houses was observed in the low expectation condition, and the response characteristics of the fusiform face area (FFA) could be explained by predictive coding, but not by feature detection models. Additional evidence supporting predictive coding in the visual cortex comes from a study by Kok and colleagues (2016). By using high-field (7T) fMRI, they found distinct laminar activation patterns for top-down (feedback) and bottom-up (feedforward) signals in the primary visual cortex (V1). In their study, participants were presented with two kinds of stimuli: (i) Kanizsa illusion figures, in which three out of four Pac-Man inducers were aligned, such that an illusory triangle could be perceived. (ii) No illusory figures, in which the same Pac-Man inducers were presented misaligned, thus not inducing any perceptual illusion. Kanizsa illusion figures allow the examination of neural responses to a shape that is not present but is induced by the surrounding shapes (i.e., the illusory triangle). The results showed that, whereas bottom-up stimulation activated all cortical layers, illusory figures led to stronger activity in deep layers of V1. This suggests that prediction units signal the presence of illusory figures, and these predictions originate from higher-level areas and terminate in deep layers of V1, where predictions are transferred via feedback connections. Instead, attenuated activity was found in middle and superficial layers, as compared to the deep layers, which could reflect the absence of bottom-up inputs (for a review about similar mechanisms in the auditory cortex, see Heilbron and Chait, 2018). Thereby different lines of empirical evidence offer converging support for the theory of PC, which suggests that the brain is constantly trying to predict the inputs it receives, and each region in the cortical sensory hierarchy represents both these predictions and the

mismatch between predictions and inputs (prediction error) (Kok and de Lange, 2015).

So far, the theoretical background, basic concepts, underlying mechanisms of the predictive coding framework and some related studies in cognitive neuroscience have been described. The following chapters will describe some specific phenomena related to predictive coding that have been extensively investigated in the past.

1.2 Fundamental mechanisms of neuronal phenomena in visual processing and related models

Perception is the interface between us and the external world (Efron, 1969). It is based on all kinds of sensory inputs and modulated by prior experiences. Vision is one of the primary ways in which we perceive our world for humans at least. Many scientific investigations of cognitive processes in the brain exploit known effects in the visual modality. For instance, scientists were able to investigate the mechanisms of attention, perception, and even consciousness based on change blindness phenomena (for a review see Simons and Rensink, 2005). This chapter will introduce some well-known neuronal phenomena that are potentially explained by the PC theory (Repetition suppression, repetition enhancement, expectation suppression, and repetition probability effects in chapter 1.2.1) and their underlying mechanism in some related models (bottom-up models in chapter 1.2.2 and top-down prediction model in chapter 1.2.3).

1.2.1 The neuronal phenomena of repetition and expectation

Most of our daily lives seem full of repetitive processes, and in contrast to a novel and changeable living environment, human seems to feel more comfortable in a stable and repetitive environment. In addition, repetition is a common way to learn new things. In this subchapter, I will describe four repetition/expectation-related neuronal phenomena (Repetition suppression, repetition enhancement, expectation suppression, and repetition probability effects) in neuroscience.

Introduction

The repetition suppression (RS) effect refers to the reduction of neural activities for repeated presentations of a given stimulus (Desimone, 1996). In other words, repeated stimuli induce smaller neural responses than novel stimuli. Please note that RS has also been called stimulus-specific adaptation (Sobotka and Ringo, 1994), or simply as adaptation (Grill-Spector and Malach, 2001). The RS effect has been widely observed in studies of monkey single-cell recordings (e.g., Baylis and Rolls, 1987; Gross et al., 1972), human EEG (e.g., Henson et al., 2004; Schendan and Kutas, 2003; Schweinberger et al., 1995), MEG (e.g., Deffke et al., 2007; Schweinberger et al., 2007; Fries et al., 2012), and neuroimaging (e.g., Larsson et al., 2016; Vuilleumier et al., 2002; for a review see Weigelt et al., 2008). Experimental paradigms inducing this effect have been extensively used to characterize neuronal populations involved in visual processing (Andrews and Ewbank, 2004; Grill-Spector and Malach, 2001; Schweinberger and Neumann, 2016). Especially, the fMRI-adaptation paradigm is regarded as a powerful tool to investigate the functional properties of human cortical neurons in a non-invasive way (Grill-Spector and Malach, 2001). In fact, in several studies, RS effects have been found in response to visual stimulus categories, such as faces (e.g., Bentin et al., 1996; Eimer et al., 2010; Henson, 2016; Kovács et al., 2006; Nemrodov and Itier, 2011; Schweinberger et al., 2002), words (e.g., Cao et al., 2015; Li et al., 2019; Mercure et al., 2011; Strother et al., 2016), scenes (e.g., Epstein et al., 2008) and objects (e.g., Buckner et al., 1998; Chouinard et al., 2008; Kaliukhovich and Vogels, 2011).

However, the repeated presentations of stimuli do not always produce RS. In some cases, repetition of stimuli results in an effect in the opposite direction, termed repetition enhancement (RE), i.e. stronger neuronal responses for repeated stimuli relative to unrepeated stimuli (for a review see Segaert et al., 2013). Interestingly, most previous studies showed the neural RE effect with a behavioral priming effect (e.g., Grill-Spector et al., 2000; Henson et al., 2002a, but see Kouider et al., 2010). The exact mechanisms of the opposite direction of repetition effect still remain unclear, but few factors such as stimulus recognition, learning, attention, expectation, explicit memory and the number of repetitions affect the direction of the repetition effect (Müller et al., 2013; for a review see

Segaert et al., 2013). In addition, several studies have found RE for novel and RS for familiar stimuli (e.g., Gagnepain et al., 2008; Henson et al., 2000). And the direction of repetition effects may change over the course of an experiment reflecting that a neural representation of a stimulus is learned (for a review see Nordt et al., 2016). For example, an EEG study by Grossmann et al. (2009) presented 6-month-olds infants with repetitive visual category (birds or fish), the results showed a RE effect for repeated stimuli in the familiarization phase, but a RS effect for repeated stimuli in the later test phase. Thus, the RE effect could appear in the initial object encoding, instead, the RS effect reflected later recognition of the familiar stimuli (Nordt et al., 2016). In other words, one hypothesis is that the direction of these two repetition effects could change with the accumulation of learning experiences, which could be related to the PC model as well (for a review see Segaert et al., 2013).

The expectation suppression (ES) effect was coined in reference to the term RS, and it describes the suppressed response to a stimulus that is more likely to occur than another stimulus (for a review see Feuerriegel et al., 2021b). In other words, ES refers to the reduction of neural activity for an expected stimulus compared to an unexpected stimulus (Todorovic and de Lange, 2012), or to a neutral or surprising one. As with RS, ES has been widely reported in monkey single-cell recordings (e.g., Ramachandran et al., 2016) and human EEG, MEG, and neuroimaging studies (e.g., Egner et al., 2010; Kok et al., 2012; Summerfield et al., 2011; Todorovic et al., 2011). Previous studies investigated ES using a variety of paradigms (e.g., Cue-based design, Statistical learning design, Block-based probability design, and Oddball design; Figure 3 on page 14). For instance, the paradigm implemented by Egner et al. (2010) is an example of the cue-based design (Figure 3A). In this study, participants were presented with three colored cues, which were associated with either face or house images with different probabilities (high, medium, and low). The results showed that the lowest BOLD response was observed when faces are more expected, which is consistent with ES in the FFA. As an example of a statistical learning design (Figure 3B), Manahova and colleagues (2018) presented participants with two kinds of sequences: (i) expected sequences, with which participants were familiarized

during a training session, and (ii) unexpected/random sequences, in which the order of images is irregular and unfamiliar. Also in this case, the lowest amplitudes were observed for expected, as compared to unexpected sequences. The oddball design (Figure 3C) is mostly used in studies using high temporal resolution techniques, such as EEG. In these studies, it is usually observed that deviant stimuli elicit more negative waveforms, as compared to stimuli presented with regularity. The difference in evoked-related potentials (ERP) waveforms between standard and deviant stimuli is known as mismatch negativity (MMN; Näätänen, 1992), which emerges due to a difference in expectations for standard and deviant stimuli (Kremlacek et al., 2016; but see Feuerriegel et al., 2021c for a surprise enhancement explanation). Lastly, in block-based probability designs (Figure 3D), stimulus repetition and alternation are manipulated in different block contexts (more details are provided in the paragraphs below). It is worth pointing out that the reason for describing RS and ES separately in this subsection is due to some previous studies finding a disassociation relationship between them (e.g., Grotheer and Kovács, 2015; Todorovic and de Lange, 2012). However, both the potential relationship and the underlying mechanism between ES and RS still remain unclear.

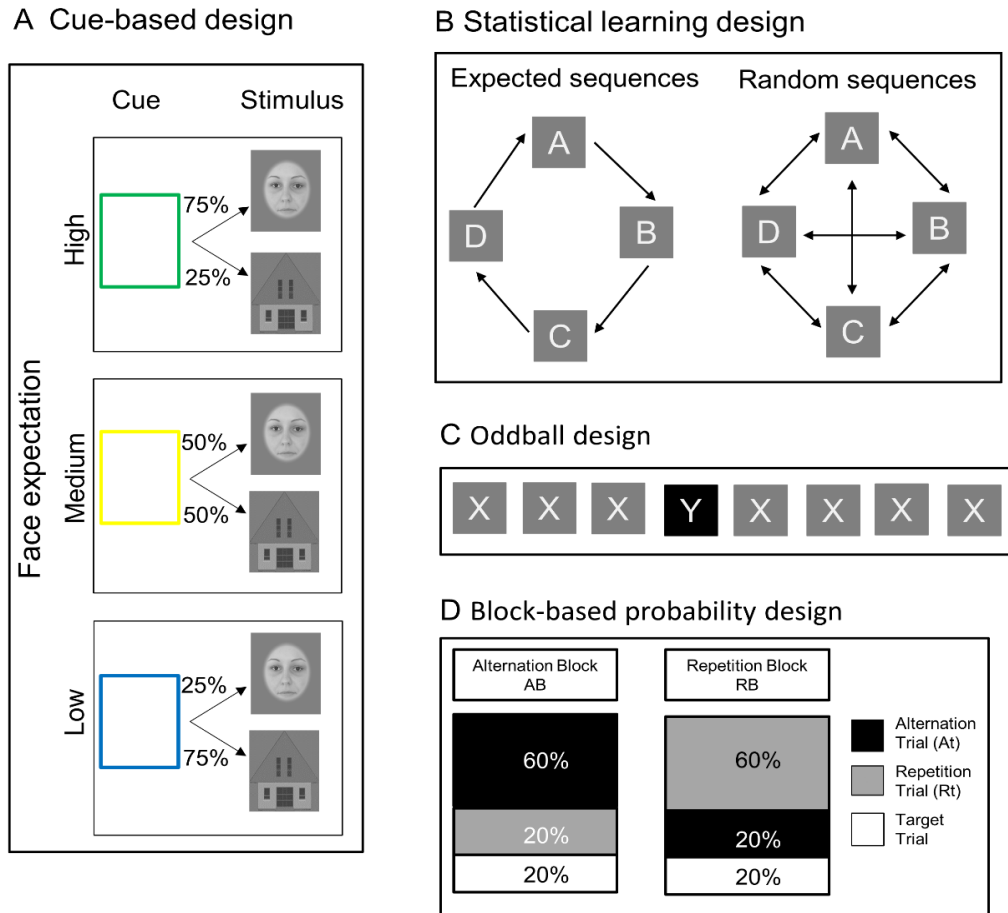


Figure 3. Examples for experimental designs investigating expectation suppression. A) Cue-based design (e.g., Egner et al., 2010). Participants were presented with three colored cues, which were associated with either face or house images with different probabilities (high (Green), medium (Yellow), and low (Blue)). B) Statistical learning design (e.g., Manahova et al., 2018). Participants are presented with expected sequences, whereby each image predicts the image that will appear next (familiarized during a training session), and random/unexpected sequences, whereby each image can follow each other image with equal probability. C) Oddball design (e.g., Kaliukhovich and Vogels., 2014). The sequence of stimuli is composed of standards (X) and deviants (Y). The deviant could randomly appear in each position in different sequences. D) Block-based probability design (e.g., Summerfield et al., 2008). Stimulus repetition and alternation are manipulated in the different blocks by probability (alternation and repetition blocks). Target trials are included to ensure that participants attend to stimuli, but are not analyzed in the end.

Introduction

The repetition probability ($P(\text{rep})$) effect actually belongs to the ES effect, but it is associated with RS as well, which reflects the modulation of repetition probability on RS. The $P(\text{rep})$ effect was first reported in an influential study (Summerfield et al., 2008). Summerfield and colleagues (2008) adopted a block-based probability design to investigate ES effect for faces. As the block-based probability design of that study is central to the present dissertation, I will describe it in detail. Summerfield et al. presented participants with pairs of face images in which the first stimulus could be either identical to (repetition trials) or different from the second one (alternation trials). The size of the second image in each pair of faces could be reduced by either 15% (non-target trial) or 60% (target trial). Participants were asked to pay attention to the size change of the paired face images and press the button when they detected the target trial. Importantly, the arrangement of repetition and alternation trials is not random in each block; rather they are grouped into blocks with high (repetition block, 60% repetition trials, 20% alternation trials, and 20% target trials) or low (alternation block, 20% repetition trials, 60% alternation trials, and 20% target trials) repetition probability (Figure 3D on Page 14). The main result of this study was that the RS effect in BOLD responses was modulated by repetition probability, i.e., a stronger RS was observed for faces in repetition blocks than in alternation blocks. This modulation effect of repetition probability on RS was called $P(\text{rep})$ effect in later studies (e.g., Grotheer and Kovács, 2014). This finding can be explained by the fact that repetition trials are more expected in repetition blocks than in alternation blocks. It represents a key piece of evidence for explaining the mechanism of RS within the framework of predictive coding. Later, this $P(\text{rep})$ effect has been corroborated in several studies using faces (e.g., Kovács et al., 2012; Larsson and Smith, 2012; Olkkonen et al., 2017).

Interestingly, the $P(\text{rep})$ effect for non-face stimuli tends to be less consistent. For example, Mayrhauser et al. (2014) reported the $P(\text{rep})$ effect for line-drawn objects. However, the effect was not replicated with every-day objects, neither with human neuroimaging nor with non-human primate single-cell recordings (Kaliukhovich and Vogel, 2011; Kovács et al., 2013). Later, Grotheer and Kovács (2014) using the same

paradigm observed the P(rep) effect for familiar Roman letters, but not for unfamiliar false fonts. They suggested that an important factor in determining the emergence of the P(rep) effect is the prior experiences which participants had with the stimuli. This possibility is related to the main research question of my work and will be expanded upon further in chapter 1.3.1. There are several reasons for describing the P(rep) effect separately from the ES effect. Firstly, this P(rep) effect is the focus of Study I and II and is considered as the indicator of predictive processing in the present thesis. Secondly and more importantly, the present thesis will give an explanation for the inconsistent results of the P(rep) effect for non-face stimuli from the prior experiences perspective.

1.2.2 Bottom-up models of repetition suppression

Since Gross and colleagues reported the RS phenomenon from single-unit recordings in the monkey inferotemporal (IT) cortex (Gross et al., 1967), the neural mechanism underlying RS is still a huge matter of debate. In this subchapter, four models relying on a primary role of bottom-up processing for explaining RS will be introduced (*Fatigue model*, *Sharpening model*, *Facilitation model*, *Synchronization model*) (Figure 4 on page 19).

The simplest mechanism underlying RS could be the *Fatigue model* (Figure 4A). According to this model, the reduced response to the repeated stimulus is due to the fatigue of neuronal populations (Miller and Desimone, 1994). In other words, a proportionally equivalent reduction of neural activity magnitude occurs in all neurons with repeated presentations of the same stimulus (for a review see Grill-Spector et al., 2006). Accordingly, this reduction in neuronal activity could result from firing rate adaptation, in which the reduction in a neuron's firing rate is proportional to its initial response (Li et al., 1993; Avidan et al., 2002; for a review see Grill-Spector et al., 2006). Within the model of neural fatigue, firing-rate dependent response fatigue could stem from mechanistically different sources: First, synaptic depression, a short-term synaptic plasticity phenomenon that is mainly presynaptic (for a review see Zucker and Regehr, 2002). Second, inherited adaptation effects, meaning that the RS in high-level visual regions could be inherited from adaptation in earlier visual regions (for a review see Vogels,

2016). However, the fatigue model fails to account for certain effects at behavioral level, for instance the improved behavioral performance which is seen in repetition priming (for a review see Grill-Spector et al., 2006).

The second model proposed to explain the RS effect is the *Sharpening model* (Desimone, 1996). According to this model, stimulus repetition results in reduced neurons numbers to respond to the same stimulus (Figure 4B). Thus, the “sharpening” of responses occurs by involving fewer neurons neural populations, which gradually respond more efficiently, by saving energy and increasing the speed of transmission. Therefore, the main difference between the Sharpening and Fatigue model is the different tuning of neurons. In the Sharpening model, most of the neurons that are optimally tuned to the repeated stimulus would show little even no response reductions, while in the Fatigue model a significant response reduction is encountered (for a review see Grill-Spector et al., 2006). This model has been used to explain priming (e.g., Wiggs and Martin, 1998; Zago et al., 2005). However, only limited evidence was found in other monkey single-cell recordings (e.g., De Baene and Vogel, 2010) and human neuroimaging (e.g., Alink et al., 2018) studies.

The third model of the RS effect is the *Facilitation model*, which also better accounts for the improvements reported in behavioral experiments. This model also hypothesizes higher efficiency processing for repeated stimuli (Figure 4C), but which is instead achieved via shorter latencies or shorter durations of neural firing (Henson and Rugg, 2003). That leads to time advanced neural activity showing a more rapid overall time course (for a review see Grill-Spector et al., 2006; Gotts et al., 2012). Empirical evidence supports the Facilitation model. Henson et al. (2002b) used a method for detecting differences in the latency of BOLD response and applied this method to an example dataset: initial versus repeated presentations of faces in a fame-judgment task. They found not only a reduced BOLD response to repeated stimuli but also a shorter latency to repeated stimuli. Another recent study supports this model using the precise spatial localization and high temporal resolution of electrocorticography (ECoG) to investigate RS in the human ventral temporal cortex (e.g., Rangarajan et al., 2020, but see Rodriguez et al., 2014). They presented participants with repeated and non-repeated images of faces

Introduction

with long-lagged intervals and many intervening stimuli between repeated stimuli. The results revealed the RS of the high-frequency broadband signal is associated with earlier peak responses. However, evidence in support of this model is not always consistent, especially when collected from non-human primates (e.g., De Baene and Vogels, 2010; Kar and Krekelberg, 2016). These inconsistent results seem to indicate that the facilitation model is not the basic source of RS, although it might contribute under certain specific circumstances (Gotts et al., 2012; Henson, 2012).

Another model of RS is the *Synchronization model*, which claims that the higher efficiency in neuronal processing is achieved mainly through enhanced neural synchronization (Gotts et al., 2012; Figure 4D). All these models need to account for the counterintuitive observation that a reduced neuronal response is associated with the improvement of behavioral performance. The synchronization model gives a possible explanation: inherent mechanisms which integrate presynaptic input in order to determine the postsynaptic action potential frequency (Gotts, 2003). Hence, there are relatively few studies that could strongly support it based on empirical evidence (for a review see Gotts et al., 2012).

In summary, the above-mentioned models have mainly explained the RS effect through the perspective of bottom-up processing, since they mainly focus on the RS effect from the local level, or concentrate on the activities of the neuronal or BOLD response of a few ROIs. Although it is necessary to explore brain functions at the local/micro level, the role of top-down modulation of prior knowledge in the brain cannot be neglected since our brain is considered as a more active, adaptive, and bidirectional system to process the sensory inputs. In other words, except for investigating and explaining the brain responses passively activated by the sensory inputs, researchers need to focus on how our brain dynamically processes the sensory inputs based on the prior expectations stored in the brain as well. Thus, more refined models, such as considering both bottom-up inputs and top-down expectations, should be structured to explain the RS effect (Bastos et al., 2012; Whitmire and Stanley, 2016). The following section will explain the underlying mechanism of RS from a predictive coding view.

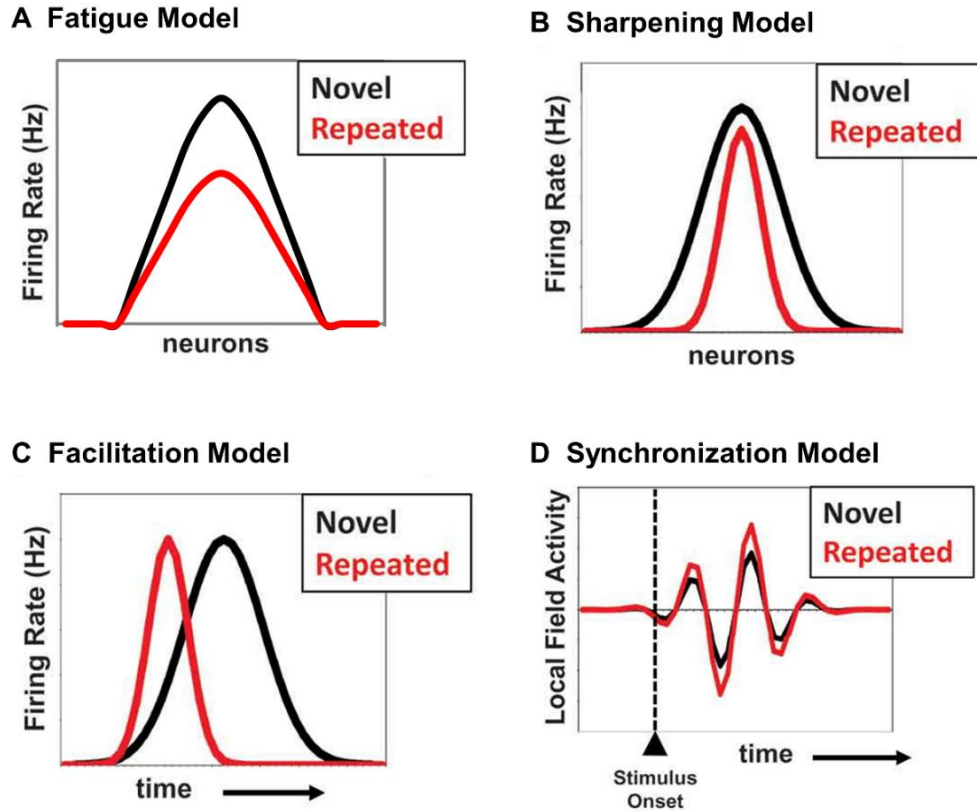


Figure 4. Schematic illustration of four different neuronal models for repetition suppression. Hypothetical neuronal responses to novel and repeated presentation of a stimulus are depicted in black and red curves, respectively. Firing rates could reduce due to fatigue of neurons (A, Fatigue Model), due to fewer neurons responding (B, Sharpening Model), due to speeding up of neuronal responses (C, Facilitation Model), or due to the increased synchronization of neuronal responses (D, Synchronization Model). Adapted from Gotts et al. (2012), with permission.

1.2.3 Repetition suppression as explained by the predictive coding model

The above-mentioned models for underlying mechanisms behind RS to repeated stimuli cannot explain all observed RS effects under specific conditions. Importantly, the underlying mechanism of RS based on these four models can also be explained within the predictive coding framework (Auksztulewicz and Friston, 2016; Ewbank and Henson, 2012; Grother and Kovács, 2016). For example, in terms of predictive coding, synchronization between neuronal populations could represent the computation of

predictions in representation units. With stimulus repetition, the computation of predictions (in each level) become more precise, leading to sharper and more facilitated neuronal responses (Henson, 2012; Amado, 2018). Thus, the relationship between predictive coding and four bottom-up models could be that the predictive coding framework is a theory about what the brain is doing, and those models (synchronization, fatigue, facilitation and sharpening) are specific proposals about how the brain does it (Friston, 2012).

The existence of the P(rep) effect (described in chapter 1.2.1) also supports the idea that RS can be explained by predictive coding. Summerfield and colleagues (2008) modulated the probability of different trials (repetition and alternation trials) in different context blocks (repetition and alternation blocks) and observed a stronger RS effect in repetition blocks. Therefore, the RS effect could be the product of perceptual inference and top-down modulations by expectation (i.e., higher expectation of encountering repetition trials in repetition blocks leads to a stronger RS than in the alternation blocks). In a following EEG study using the same paradigm, the same P(rep) effect was associated with a relatively late ERP component, around 300 ms after stimulus onset, in central (C3, C4, CP3, CP4, CPz, and Cz) electrodes (Summerfield et al., 2011). This effect was replicated and extended to other stimulus categories, such as Roman letters, words, objects (e.g., Grotheer and Kovács, 2014; Li and Kovács, 2022; Mayrhauser et al., 2014; but failed to be found in Kaliukhovich and Vogels (2011) and Kovács et al. (2013)).

Based on the available experimental evidence, Grotheer and Kovács (2016) proposed a two-step model of response suppression (Figure 5 on Page 21). On one hand, this model proposed how to explain RS and ES under the predictive coding framework. It suggests that RS is the expression of ‘low-level’ prediction error (ϵ) calculations compared to ES, which ES likely represents “higher-level” ϵ computations with expectation signals from frontal regions (Summerfield et al., 2006).

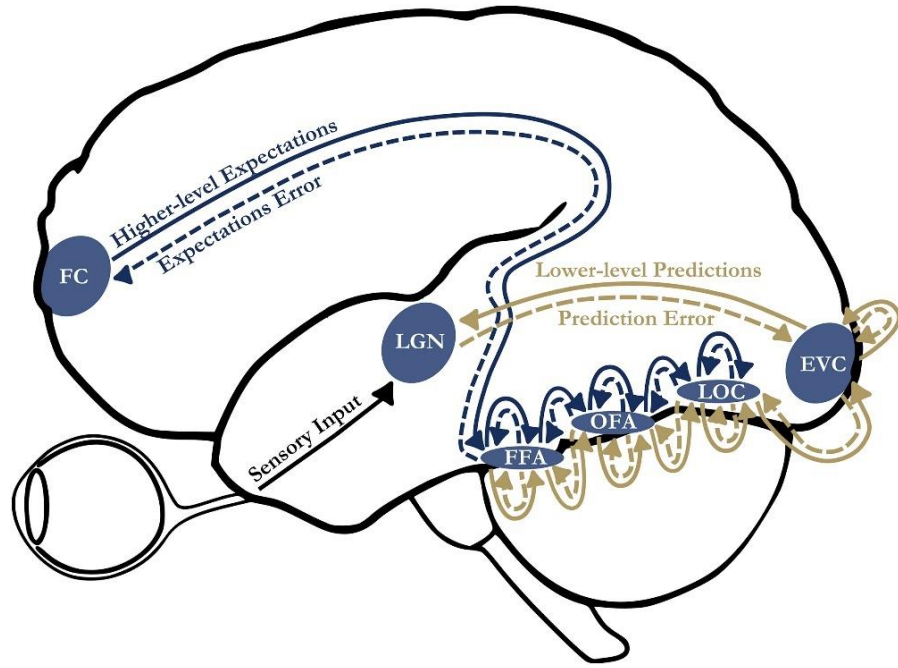


Figure 5. Schematic illustration of the two-stage model of response suppression in the ventral visual stream. Higher-level expectations (depicted in blue) originate in frontal regions and are transmitted to the FFA, the OFA and the LOC. Lower-level predictions (depicted in gold) originate in the FFA and are transmitted, via the OFA, the LOC and the EVC, to the LGN. Visual areas compute and return the prediction/expectation error. A reduction in prediction/expectation error magnitude expresses as RS/ES, respectively. Please note that the anatomical location of neuronal regions is estimated roughly. FC: frontal cortex, FFA: fusiform face area, OFA: occipital face area, LOC: lateral occipital complex, EVC: early visual cortex, LGN: lateral geniculate nucleus. Adapted from Grotheer and Kovács (2016), with permission.

On the other hand, it posits the independence of RS and ES, which has been found in previous studies (e.g., Grotheer and Kovács, 2015; Todorovic and de Lange, 2012). For instance, RS appears after stimulus onset, typically starting earlier than ES in EEG/MEG studies (e.g., Schweinberger et al., 2004; Summerfield et al., 2011, Todorovic and de Lange, 2012). In addition, Aukstulewicz and Friston (2016) explained that RS can be understood in terms of ‘explaining away’ sensory prediction error (Gotts et al., 2012) from the predictive coding perspective (Figure 6 on page 22). According to the authors, PC supports optimal categorization and learning of sensory inputs. With stimulus repetition, its prediction error is gradually optimized. PC also can be mapped onto

canonical cortical microcircuits whose superficial pyramidal cells encode prediction errors about potential causes, transmit error estimation to hierarchically higher areas via ascending connections (Bastos et al., 2012). Similarly, deep pyramidal cells send predictions via descending connections to lower levels.

Thus, both of the above-mentioned modeling and empirical evidence suggest that the RS effect could be explained by predictive coding, which more concentrates on the modulation of top-down expectations.

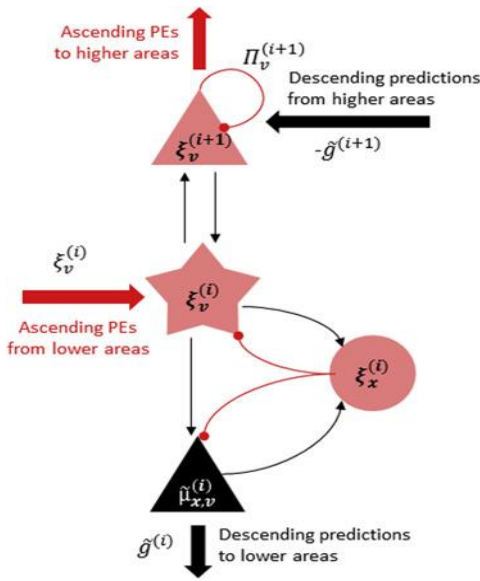


Figure 6. Mapping the predictive coding scheme onto a canonical microcircuit. Prediction errors about hidden causes from hierarchically lower areas are received by spiny stellate cells in the granular layer. The spiny stellate cells also receive inputs from inhibitory interneurons, encoding the prediction errors about hidden states (i.e., describing the dynamics at a given hierarchical level). These prediction errors are reconciled with descending predictions from hierarchically higher

areas received by the superficial pyramidal cells, which reciprocate the ensuing prediction errors. At the same time, predictions are reconciled in the deep pyramidal layers and relayed to hierarchically lower areas. Adapted from Aukstulewicz and Friston (2016), with permission.

1.3 Critical moderators of the prediction-related neural responses

PC has become an influential and popular theory in the domain of perception over the last two decades, mostly through the extensive empirical support it received. However, the factors modulating the behavioral and neural effects that are explained by the PC theory remain largely unexplored. This chapter will describe several factors which are important for predictive coding. For instance, prior experiences with the stimuli (Subchapter 1.3.1),

stimulus familiarity and expectation (Subchapter 1.3.2), attention (Subchapter 1.3.3), and short-term memory (Subchapter 1.3.4).

1.3.1 Prior experiences with the stimuli

‘Experience’ is an abstract and complicated concept and has been discussed in various disciplines, such as Epistemology, Phenomenology, Philosophy, and Psychology, to name a few. Examining this concept under all these perspectives is beyond the scope of this dissertation. In the context of the empirical studies, I conducted during my doctoral studies, the notion of experience is limited to perceptual and practical experience with the stimuli, as commonly described in Psychology (Silins, 2015). As described in Chapter 1.1, the prior experiences of perceived objects are not only central in Helmholtz’s perception but also in the predictive coding framework (Helmholtz, 1867, Kok and de Lange, 2015; Rao and Ballard, 1999). Generally, several kinds of experience are described in Psychology. For example, visual experience is observed through repetitive visual exposure. After a short-term training, participants have short-term experiences with trained objects. In contrast, if the training sessions last for several months or even years, participants would obtain long-term/extensive experience with the trained objects.

In addition to gaining experience through repeated visual exposure, perceptual learning (PL), is a popular and active practical approach, as it is considered to improve the perceptual experience for trained objects (For a review see Fahle, Poggio, T., and Poggio, T. A., 2002). With intensive training, participants can dramatically improve their performances in various perceptual tasks, ranging from visual features discrimination, to shape identification and object processing (e.g., Fiorentini and Berardi, 1980; Goldstone, 1998; Op de Beeck and Baker, 2010). The behavioral improvements obtained through training tasks are associated with changes in brain activity in the related cortical area (Grill-Spector et al., 2000; Op de Beeck et al., 2006; for a review see Bi and Fang, 2013). For instance, previous fMRI studies have found that PL can lead to a reduced BOLD signal response for trained stimuli (Henson et al., 2000; Reber et al., 1998; Sigman et al., 2005; but see Grill-Spector et al., 2000). Importantly, compared to untrained (novel) stimuli, the

Introduction

reduction of BOLD signal for trained stimuli can be regarded as a RS effect (Grill-Spector et al., 1999; Jiang et al., 2007). In other words, increased perceptual experience through PL or intensive training could lead to the appearance of the RS effect.

Additionally, perceptual expertise is another way of acquiring perceptual experience. For instance, Gauthier et al. (1999) asked participants to complete a series of training tasks over several days for a novel stimulus category, created ad-hoc for studies on object expertise, named “Greeble”. Alongside the improved performance after the training, they found increased brain activation for Greebles in face-sensitive areas, and no significant difference between faces and Greebles. This result suggests that these face-sensitive areas are not specifically responding to faces but also to well-trained non-face objects (Gauthier et al., 2000). This would support the domain-general hypothesis of the sensitive responses for faces in the FFA (e.g., Gauthier et al., 2003; Busey and Vanderkolk, 2005). Importantly, it has been suggested that the tight connection between predictions and perceptual expertise via analogies and associations (for a review see Cheung and Bar, 2012). In other words, predictions might depend on extensive experience (e.g., perceptual expertise) with perceived objects. To test this hypothesis, Grotheer and Kovács (2014) adopted a block-based probability design similar to that of Summerfield et al. 2008 (Figure 3D on page 14) to investigate the P(rep) effect for familiar Roman letters and unfamiliar false fonts. Their findings showed that the significant P(rep) effect is only observed for Roman letters, which suggests that perceptual expertise with stimuli is an important factor in determining the emergence of the P(rep) effect. Hence, in a study described in this thesis the same paradigm was used for other stimuli for which strong expertise is already developed, namely, words (details see Study 1 in the present thesis). Together, these results support the idea that predictive processes, measured by P(rep) modulations of RS, require prior experiences with the stimuli.

1.3.2 Stimulus familiarity and expectation

Prior experiences can modify how a visual stimulus is processed in our brain (Friston, 2005). Different from the previous subchapter, which emphasized the influence of prior

Introduction

experiences with the stimulus category for prediction related neural effects, this subchapter focuses more on the effect of stimulus familiarity and expectation on prediction related effects during the experiment. Prior experience is a generalized concept, containing many aspects of the experience for perceived/experimental objects, e.g., the extensive, long-term experience for stimulus categories, including rich semantic information or just visual experience of repeated exposure. Whereas stimulus familiarity and expectation are two sub-forms of prior experience, mainly refer to the short-term/temporal experience with experimental stimuli. In other words, stimulus familiarity and expectation belong to two specific prior experiences: stimulus familiarity indicates whether a stimulus has been seen before, and stimulus expectation refers to whether a stimulus could be expected to occur (Manahova et al., 2018). In general, familiarity is the impression that something has been seen in the past (i.e. "it looks familiar"), and there are different degrees of familiarity (Kovács, 2020). Prior experience is just a general, factual term to say that I was exposed to something. Familiarity is the subjective feeling that I was exposed to that thing. Expectations could be more related to experimental designs or tasks. For instance, in an old-new images of objects judgment task, participants could have prior experiences both for old and new objects, but they have learned the old images of objects only, therefore they are familiar with these specifically. Further, if participants were instructed to detect inverted stimuli in the test phase as well, the inverted images could be expected events regardless of they have been learned or not.

Many studies suggest that familiarity with a stimulus effectively modulates neural processing in various ways (e.g., Freedman et al., 2006; Huang et al., 2018; Manahova et al., 2020). For example, viewing an image of an object repeatedly leads to decreased spiking activity in the intertemporal (IT) cortex in monkeys (Miller et al., 1991) and reduced BOLD responses in the lateral occipital cortex (LOC) for familiar stimuli compared to unfamiliar stimuli (for a review see Grill-Spector et al., 2006). Similarly, the ES effect (as described previously) which suggests suppressed responses for expected in comparison to unexpected stimuli, has been reported in monkey single-cell recordings

(e.g., Meyer and Olson, 2011; Ramachandran et al., 2016) and human EEG, MEG, and neuroimaging studies (e.g., Egner et al., 2010; Kok et al., 2012; Summerfield et al., 2011; Todorovic et al., 2011). A recent study suggests that stimulus familiarity and expectation jointly modulate object processing in the LOC (Manahova et al., 2018). Specifically, Manahova et al. separately manipulated stimulus familiarity (familiar vs. novel) and expectation (expected vs. unexpected) and measured the neural activity using MEG while participants viewed object images. The results replicated both familiarity and expectation effects, and importantly, suggest that the familiarity effect (starting at 200 ms after stimulus onset) appears earlier than the expectation effect (starting at 500 ms after stimulus onset). In addition, they found that familiarity leads to a higher dynamic range (i.e., peak-to-trough difference) of sensory response, but expectation does not. Thus, these results suggest that stimulus familiarity and expectation could jointly modulate perception but in distinct time courses. One potential explanation is that familiarity mainly results in local changes in the LOC, which occur within a relatively short time (Vogels, 2016; Kaliukhovich and Vogels, 2011). Instead, expectation suppression occurs when the LOC receives the prediction error from higher-order areas, therefore requiring a longer time (Friston, 2005).

1.3.3 Attention

We are continuously exposed to an extensive amount of sensory information; therefore, attention plays a crucial role in selectively processing a relevant fraction of it. Similarly, several hypotheses about the latent causes of this sensory input are continuously computed by the brain and only the most appropriate should be selected for top-down transmission from the abundant corpus of prior knowledge about a situation. This is achieved by weighting which predictions are the most relevant at any time point, and attention is thought to be the potential mechanism allowing this. Thus, it is important to investigate how attention modulates predictive processing.

It has been well-established that attention increases the efficiency of information processing. Especially, attending to a stimulus, which is associated with increased

neuronal population selectivity (e.g., Maunsell, 2015; Murray and Wojciulik, 2004; Serences and Kastner, 2014; Treue, 2003). Summerfield and Egner (2009) have distinguished expectation and attention in visual cognition and discussed how their influences on visual perception overlap, differ and interact. Several empirical studies have attempted to manipulate either attention or expectation or to separate expectation from attention. The results of these studies suggest that attention is a prerequisite for perceptual expectations (e.g., Alink and Blank, 2021; Larsson and Smith, 2012; Richter and de Lange, 2019; but see Summerfield and Egner, 2009). For instance, Richter and de Lange (2019) manipulated attentional demands through two different tasks (attending either objects or distractors), while expectations were generated during a separate statistical learning session. Their results showed that the expectation suppression (ES) effect only appears when participants attend to objects and this effect is significantly attenuated when attention is drawn away from objects by distractors. Larsson and Smith (2012) found a similar pattern for the effect of attention on ES when using face stimuli, but interestingly their results also showed that the repetition suppression (RS) effect is attention-independent. However, it is still under debate how attention modulates expectation. In the classic predictive-coding based explanation, the significant ES effect in the attended condition is due to the reason that attention facilitates prediction error coding (Feldman and Friston, 2010; Friston, 2009), meaning that attention can enlarge the prediction error effect for unexpected relative to expected conditions. However, this explanation was challenged by attention theories (for a review see Alink and Blank, 2021), which propose that the ES effect in the attended condition is due to an enhanced stimulus saliency of unexpected stimuli. Therefore, the critical difference between these two explanations is that the former suggests that attention causally facilitates prediction-error coding, while the latter hints that attention effect of stimulus saliency leads to enhanced responses for unexpected stimuli.

The biggest challenge in investigating how attention modulates ES is the difficulty of orthogonalizing the attention and expectation effects. As described above, experimental manipulations of attention through different tasks are only possible for endogenous

attention (Larsson and Smith, 2012; Richter and de Lange, 2019), while stimulus saliency normally belongs to exogenous attention. Therefore, future studies can further distinguish between endogenous and exogenous attention. For instance, expectation can be manipulated by introducing different stimulus probabilities, endogenous attention can be manipulated by varying task-relevance, and exogenous attention can be manipulated using different stimulus saliency levels. Of course, besides the suggested manipulations, future experiments can combine multiple techniques. For example, recording the eye movements with eye-tracking, or separating attention and expectation using high temporal resolution methods (e.g., EEG, MEG). Since previous studies have found the exogenous attention effect precedes those of endogenous attention (Hickey et al., 2010), testing these interactions with the proposed techniques can be a promising research avenue (for a review see Summerfield and Egner, 2009).

1.3.4 Short-term memory

The German psychologist Hermann Ebbinghaus is usually considered the first scientist to scientifically investigate memory. Memory plays an extremely important role in our daily lives and is related to most cognitive processes. For instance, when you are doing a mathematical calculation, you need to extract some calculation rules you have learned before from your memory; You need to remember where you parked your car (spatial memory); You need to remember how to drive (procedural memory). According to an influential model of memory, short-term memory (STM) can hold a limited amount of information in a very accessible state temporarily (Cowan, 2008), differently from long-term memory and working memory (Atkinson and Shiffrin, 1968). Instead, long-term memory refers to a vast and long-term store of knowledge and a record of prior events (Atkinson and Shiffrin, 1971). Working memory refers to memory as it is used to plan and carry out behavior (Miller et al., 1960; Ruchkin et al., 2003).

According to the predictive coding framework, perception is the product of the brain's attempts to constantly match sensory inputs to prior knowledge (Kok and de Lange, 2015), involving the bidirectional streams of prediction error, transmitted from low level

to higher levels, and predictions, transmitted in the opposite direction (Clark, 2013; Friston, 2015). However, important questions about how the brain efficiently accomplishes these processes remain open. Especially, is predictive processing related to memory, and if so, how? Conventionally, we might assume the prediction-related priors are stored in our long-term memory and that when we have to perform relative cognitive processes, such prior expectation is retrieved from a long-term storage. However, given the high efficiency with which perception is accomplished, this hypothesis seems rather unlikely. Alternatively, Trapp and colleagues (2021) proposed that priors are retrieved from long-term memory and temporarily stored in short-term memory (STM). In this way, the matching process happens between sensory inputs and short-term memory. In most cognitive models, STM is regarded as a part of long-term memory that is engaged in the current task (e.g., Anderson et al., 2004; Cowan, 1998; Rochkin et al., 2003). Therefore, STM could be an optimal candidate to temporarily store priors that are then compared to sensory inputs, as it could hold relevant information in a highly accessible state for related cognitive processes (Cowan, 2008). Several studies provide evidence that STM is required for perceptual prediction processes. For instance, Travis and colleagues (2013) adopted a contextual cueing paradigm, in which participants searched a target in stimulus arrays, with some of these arrays being repeated several times during the experiment (participants are not explicitly informed), while others are new arrays. The results showed that participants tended to find the target in the repeated arrays faster than in the novel others. However, when participants had to maintain additional information in STM during the same search task, the performance facilitation performance was attenuated. This suggests that STM is required for storing repeated information and predicting the target. Furthermore, St John-Saaltink and colleagues (2015) demonstrated that the expectation suppression for predictable items is attenuated under STM load. They presented an auditory cue (a tone) prior to predicting the orientation of grating stimuli. The reduced BOLD response in V1 was found for predictable compared to non-predictable gratings, but this effect was attenuated when participants had to maintain additional, orientation-irrelevant information in their STM. Finally, Cashdollar et al. (2017) found that STM capacity is correlated with the neural correlates of prediction in

a task that required participants to attend to a series of images. Together, the presented empirical evidence suggests that STM plays an important role in predictive processing. Study III in the present thesis will further investigate the modulation of short-term memory on prediction-related effects in a high-level visual region.

In this subchapter, I described some critical moderators of predictive processing, such as prior experiences, stimulus familiarity and expectation, attention, and short-term memory. These factors are related to the topic of the current thesis under the framework of predictive coding. The author believes that integrating the effect of these factors into the theory of PC, the mechanisms underlying human perception would be better described.

1.4 Objectives and questions of the present thesis

The studies in the present thesis mainly focus on whether and how prior experiences modulate neural correlates of predictive processing in the ventral visual areas. Prediction-related neural effects are measured with fMRI and are elicited by presenting participants with different visual stimuli (faces, words, and objects). The definition of “prior experiences” in my works specifically refers to both the long-term experience of language learning and to the short-term experience achieved by perceptual learning. Two well-established phenomena are measured as the index of predictive processing (see chapter 1.2.1): repetition suppression (RS) (e.g., Grill-Spector et al., 2006; Grotheer and Kovács, 2016), and repetition probability (P(rep)) effects (e.g., Grotheer and Kovács, 2014; Summerfield et al., 2008). Four representative areas in the ventral visual stream are examined in detail (fusiform face area (FFA; Kanwisher et al, 1997a); occipital face area (OFA; Gauthier et al., 2000); visual word form area (VWFA; Cohen et al., 2000); and lateral occipital complex (LOC; Malach et al., 1995)), also see chapter 2.2). In addition, other important potential modulators on PC, such as attention and short-term memory, were investigated in this thesis. In general, this work attempts to clarify the influence of prior experiences on predictive processing of multiple stimulus categories and within stimulus-selective areas.

The first empirical contribution (Study I, chapter 3.1) investigates repetition probability effects ($P(\text{rep})$) for words. Previous studies have observed this effect for various sensory stimulus categories, such as faces, objects and Roman letters (e.g., Grotheer and Kovács, 2014; Mayrhauser et al., 2014; Summerfield et al., 2008). Especially, Grotheer and Kovács (2014) adopted the paradigm of Summerfield et al., (2008), using two non-face stimulus categories with distinct levels of prior experiences: familiar Roman letters and novel false fonts. The results showed the $P(\text{rep})$ effect for familiar Roman letters, but not the novel false fonts, which suggests that this $P(\text{rep})$ effect depends on the prior experiences of the participants with the stimuli. This study was the first demonstration that prior experience is important for the $P(\text{rep})$ effect, which is related to PC. Here, in our Study I, we tested the extent to which prior experiences affect the $P(\text{rep})$ effect and if it could be extended to more complex lexical stimuli (i.e. Chinese characters and German words) as well. Observing that $P(\text{rep})$ effects can occur for complex lexical stimuli would demonstrate that prior experience is a universal prerequisite for predicting multiple stimulus categories.

Study II (chapter 3.2) focuses on the modulation of short-term experience on the $P(\text{rep})$ effect for non-face objects. As mentioned above, the $P(\text{rep})$ effect was replicated for faces, but inconsistent findings were observed for non-face objects. For instance, Mayrhauser et al. (2014) reported the $P(\text{rep})$ effect for line-drawn objects, but such effects were not replicated for everyday objects or chairs in human neuroimaging studies (Kovács et al., 2013). Also, 3D artificial faces, fractal patterns and everyday objects in macaque single-cell recordings studies could not replicate the $P(\text{rep})$ effect (Kaliukhovich and Vogels, 2011; Vinken et al., 2018). This inconsistency could be due to differences in experimental species, and tasks, but also in prior experiences of the experimental stimuli in question. Importantly, the effect of prior experiences on the $P(\text{rep})$ modulation was investigated by Grotheer & Kovács (2014). The results showed that the $P(\text{rep})$ effect was only observed for familiar letters, but not for novel false fonts. This hints that observing the $P(\text{rep})$ effect depends on extensive experience with stimuli. However, it remains unknown whether the short-term experience of stimuli could modulate the $P(\text{rep})$ effect

Introduction

as well. Study II addresses this question by performing fMRI measurements before, and after ten-day perceptual learning (PL) sessions for non-face objects (cars).

Whereas the first two studies mainly focus on the modulation of prior experiences on the P(rep) effect, the third study (Study III, chapter 3.3) investigates how prior experiences modulates bottom-up sensory input in the brain. Two potential cognitive processes (short-term memory (STM) and attention) could contribute to the modulation. The influence of these factors on PC have been described in chapters 1.3.3 and 1.3.4. To address this question, Study 3a includes a classic stimulus repetition paradigm to measure RS, together with either concurrent STM load or a control condition. Instead, in Study 3b we then further manipulated the attention by introducing an active task. In combination with previous studies, the results from Study III may help clarify and contextualize how prior experiences modulates on the top-down predictive processing.

In sum, the central questions of the present dissertation are the following.

1. Does the repetition probability effect for words depend on long-term experience with the stimuli?
2. Does short-term experience affect the repetition probability effect for non-face objects?
3. How and to which extent do top-down predictions induced by prior experience modulate the repetition suppression effect?

2 General methods

The three studies presented in the present thesis are all based on the same technique (fMRI) with the stimulus repetition paradigm (Summerfield et al., 2008) used to investigate the RS and P(rep) effects of stimulus categories (faces, words, and objects) in the ventral visual stream. I will give a brief description of the fMRI technique in this section (Chapter 2.1). In addition, I will introduce some of the Regions-of-interest (ROIs) in the ventral visual system, on which the present studies are focused (Chapter 2.2).

2.1 Functional magnetic resonance imaging

Functional magnetic resonance imaging (fMRI) is a popular technique that has widely been used to investigate brain activity. Due to its non-invasiveness and its high spatial resolution, fMRI is considered to be an ideal technique to investigate brain function. The main principle of this technique is the coupled relationship between neuronal activation and cerebral blood flow (Buchbinder, 2016; Huettel et al., 2004). One important index in fMRI is the blood oxygen level-dependent (BOLD) signal. BOLD fMRI detects local increases in relative blood oxygenation that are assumed to be a direct consequence of neurotransmitter action, and thus reflect local neuronal activities (Matthews and Jezzard, 2004). These activity patterns can be related to cognitive functions by testing which regions showed more or less activity during different task conditions. In other words, when a brain area is engaged in a relative cognitive process, the exchange between oxygenated and deoxygenated blood would increase. This increase can be detected with fMRI and is then interpreted as related to the cognitive process that the task at hand required of the participant.

In this thesis, image data were preprocessed using SPM12 (Wellcome Department of Imaging Neuroscience, London, UK), which is a MATLAB-based software (The MathWorks). More details of this software and descriptions for the parameters in each preprocessing step are described in their manual (Ashburner et al., n.d.). All experiments

General methods

were performed with a 3-Tesla MR scanner (Siemens MAGNETOM Prisma, Erlangen, Germany), with 20-channel (Study I) and 64-channel (Studies II and III) head coils. High-resolution T1-weighted images (192 slices; TR = 2300 ms; TE = 3.03 ms; flip angle = 90 deg; 1 mm isotropic voxel size) were acquired to obtain 3D structural scans. The parameters of T2-weighted images were different in the three studies based on some specific requirements (details see the imaging parameters in each study).

In the following paragraph, each preprocessing step will be briefly described.

1) DICOM conversion. The scanner produces data in Digital Imaging and Communications in Medicine (DICOM) format. The DICOM format is not directly compatible with SPM, therefore conversion to the 3D image format (.nii) is needed. Additionally, the converted image data can be visualized as a 3D volume as well, rather than slice form.

2) Slice timing. The 3D volume is not acquired at the same time point but is composed of a sequence of 2D slices in a certain time span (repetition time, TR). Thus, differences in image acquisition time between slices should be corrected based on the number of slices, TR, slice order, and reference slice. After this step, these delays in different slices during acquisition can be corrected, and all slices of one volume can be interpolated to the same time point for which a reference slice is selected.

3) Spatial realignment. After the temporal correction of the data, the following steps are concerned with spatial corrections. Participants will inevitably move their heads during data acquisition, within and between runs. Although the amplitude of these movements can be very small, it will nevertheless affect the precision of spatial positioning of related brain areas. Thus, during spatial realignment, the software calculates six parameters reflecting the head motion relative to the reference volume (i.e., the first image). These six parameters are saved and used as additional regressors when the general linear model is specified. In other words, the images can be resliced based on these parameters.

4) Spatial co-registration. As the anatomical and functional images are different, especially in terms of spatial resolution and liability to noise in the form of distortions,

co-registration is necessary to match activity differences to anatomical locations. In this step, the preprocessed anatomical image is aligned to the mean functional volume to compare the difference of brain activities within a group of participants.

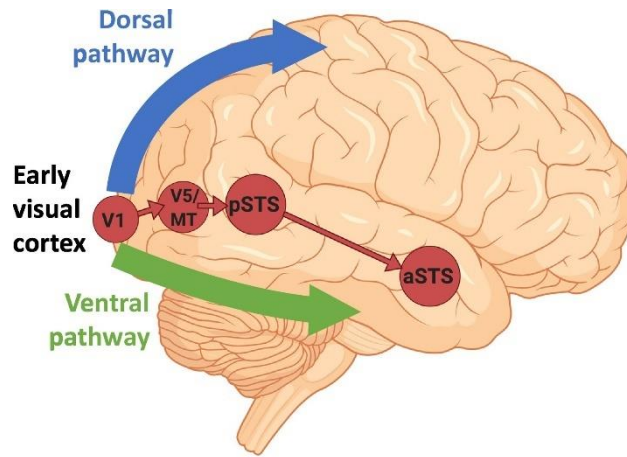
5) Spatial normalization. To ensure that different brains are comparable, different participants' brains are warped to a standard spatial system. This process happens in two steps. First, a deformation is estimated by deforming a template to match an individual scan; second, the spatially normalized images are then actually written, using the previously estimated deformation (Ashburner et al., n.d.). In our studies, we first normalized the individual anatomical images to the standard Montreal Neurological Institute (MNI) template, then we normalized all functional scans of the same participant. After the normalization, all participants' results can be compared in the standard space (MNI template). In our studies, all images were normalized to the MNI template ICBM152, which is the result of averaging 152 normal MRI scans (Mazziotta et al., 1995).

6) Smoothing. The final step of preprocessing is spatial smoothing to suppress noise and effect due to residual differences in functional and gyral anatomy during inter-subject averaging. This step will result in a loss of spatial resolution, but statistical power will instead be increased.

After the six preprocessing steps, the image data need to be mapped to experimental conditions. This is achieved by computing a general linear model (GLM), in which a function reflecting the expected hemodynamic response (canonical hemodynamic response function) is fit to the time-series data. The BOLD signal estimation based on this model, which also includes regressors representing different task conditions, can then be used for further analyses (e.g. ROI analysis or whole-brain analysis). Canonical hemodynamic response functions were extracted using MarsBaR 0.44 toolbox for SPM (Brett et al., 2002).

2.2 Regions-of-interest in the ventral visual system

Vision is one of the most important sensory channels enabling human to perceive the dynamic world. Visual information is captured by the eyes and received in the retina. Furthermore, it is transmitted by the optical nerves and processed through the lateral geniculate nucleus (LGN) before it enters to the primary visual cortex (e.g., V1, V2, etc.). Afterwards, the visual information is transferred to higher-level visual cortices (Tovée et al., 1996). Almost 40 years ago, Ungerleider and Mishkin (1982) proposed the two-visual pathway model of the primate visual cortex, assuming the presence of two visual streams (the ventral and the dorsal visual streams, Figure 7 on page 37). The ventral visual stream, also known as the ‘what’ pathway, is located in the occipital and temporal cortices and mainly computes the identity of all kinds of visual objects (e.g., characters, faces, scenes). Instead, the dorsal visual stream is known as the ‘where’ pathway, and it projects along the dorsal brain surface. It mainly processes the information about the location of visual objects in the field of view, and the actions related to these objects (e.g., reaching, throwing, grasping) (Milner and Goodale, 1995; Pitcher and Ungerleider, 2021). Recently, Pitcher and Ungerleider (2021) revised the two-visual pathway model and presented evidence for the existence of a third visual pathway on the lateral brain surface in both human and non-human primates (Pitcher and Ungerleider, 2021). The third pathway mainly processes dynamic information of visual objects, and it is responsive to social perception (e.g., Allison et al., 2000; Hein and Knight, 2008; Pitcher et al., 2014; Sliwinska et al., 2020).



Trends in Cognitive Sciences

Figure 7. Schematic illustration of the three-visual pathway model. The corticocortical connections of the third pathway (in red) are independent of the ventral pathway (in green) and the dorsal pathway (in blue). a/pSTS: anterior/posterior superior temporal sulcus. V1: primary visual cortex. V5: motion-selective area. MT: middle temporal visual area. Adapted from Pitcher and Ungerleider (2021), with permission.

The present chapter will mainly introduce four sensitive areas in the ventral visual stream, namely the fusiform face area (FFA), the occipital face area (OFA), the visual word form area (VWFA), the lateral occipital cortex (LOC). All these areas are visualized in Figure 8 on page 38. There are several reasons why the focus is put on these areas. First, they are thought to play an important role in visual object processing; Second, they are especially responsive to types of stimuli used in our studies; Last but not least, previous studies have reported the RS effect I discussed in relation to predictive coding (PC) in these areas.

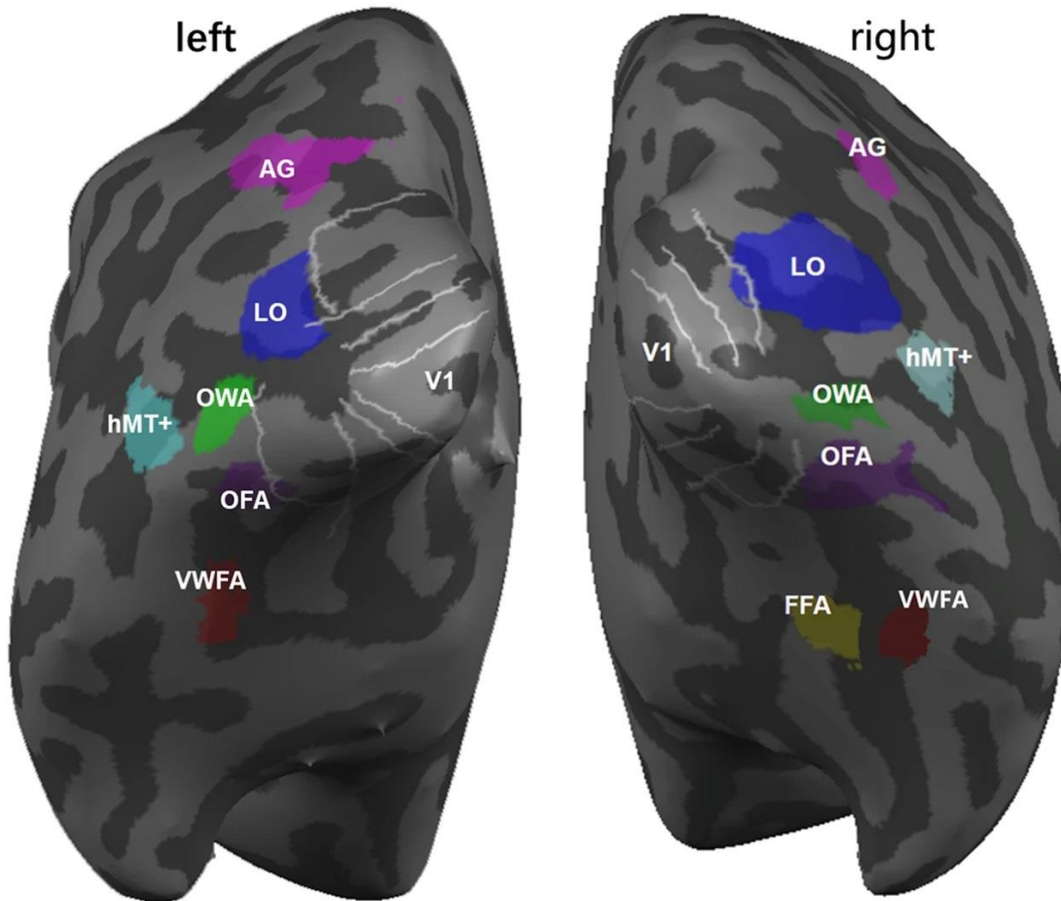


Figure 8. Relative spatial position of ROIs in a typical participant's inflated brain. From the perspective of the occipital lobe, the white lines indicate the boundaries of the retinotopic areas. LO: lateral occipital complex (named as LOC in the present thesis). AG: angular gyrus. hMT+: human motion-selective complex. OWA: occipital word sensitive area. OFA: occipital face area. VWFA: visual word form area. FFA: fusiform face area. V1: primary visual cortex. Left and right indicate left and right hemisphere. Adapted from Zhang et al. (2018), with permission.

Faces are commonly experienced, but simultaneously special stimuli in our daily lives. Many previous studies have reported serial-specific behavioral markers of perceptual expertise for faces, including the inversion effect (e.g., Yin, 1969), the composite effect (e.g., Young et al., 1987), and the left-side bias effect (e.g., Gilbert and Bakan, 1973). At the neural level, different kinds of information derived from faces are thought to be processed in a face perception network, which has been divided into 'core' and 'extended' systems (Haxby et al., 2000; Schweinberger and Neumann, 2016).

The fusiform face area (FFA) as one core region in the 'core' face system is located in the

fusiform gyrus and is extremely sensitive to faces (Figure 8 on Page 38; Kanwisher et al, 1997a; Zhang et al., 2018; for a meta-analysis see Müller et al., 2018; for a review see Kanwisher and Yovel, 2006). In other words, the FFA always shows a stronger activation for faces than for other objects. Normally, it is defined with a functional localizer contrast of faces vs objects (Grill-Spector et al., 2004), or other contrasts, such as faces vs objects and Fourier randomized images of faces (Grotheer and Kovács, 2015). With the development of neuroimaging techniques, researchers implemented some well-known criteria (precise anatomical location) to parcellate early visual areas (Felleman et al., 1991). Within the FFA, Weiner and Grill-Spector (2012) used the same criteria to define two subsections: a middle and a posterior FFA (mFFA and pFFA). The mFFA overlaps with the middle fusiform sulcus and is adjacent to the anterior and middle part of a limb-selective region on the occipitotemporal sulcus. Conversely, pFFA overlaps the posterior part of the fusiform gyrus and is located posteriorly to the above-mentioned limb-selective occipitotemporal region (Weiner and Grill-Spector, 2012). Schwarz et al. (2019) found that the best contrast to identify the pFFA is contrasting face vs landscapes. In contrast, the mFFA is activated more strongly for faces than object, houses, or landscapes. Recently, a MEG study found that the pFFA is activated earlier than mFFA, which suggests different functional roles of these two regions in face processing (Fan et al., 2020). However, further studies are needed to investigate whether and how these two regions are engaged in different aspects of face processing.

Besides mere face detection, the FFA is known to respond stronger to upright than inverted faces (Yovel and Kanwisher, 2005), which is closely associated with the behavioral face-inversion effect. These results relate to configural/holistic face processing (e.g., Maurer et al., 2007; Rossion, 2008; Rossion and Gauthier, 2002). Previous studies further reveal that the FFA plays an important role regarding face identification (for reviews see, e.g., Duchaine and Yovel, 2015; Gobbini and Haxby, 2007; Haxby et al., 2000). For example, a higher neuronal response was induced to the correctly identified faces than unfamiliar faces (Grill-Spector et al., 2004). Some neurological disorders can lead to reduced performance in face identity recognition, which is associated with impairments

of the anatomical and functional characteristics of the FFA. For example, prosopagnostic individuals, who cannot identify faces, often display lesions in this area (Barton et al., 2002; Furl et al., 2011). Other lesions in the fusiform gyrus can generate poor abilities in tasks involving face individuation (Wada and Yamamoto, 2001) and detection (Xu et al., 2014).

However, another line of evidence suggests that the FFA is not specifically responsive to faces, but in general to stimulus categories for which people develop a high expertise (Gauthier et al., 2000). This was observed by Gauthier et al. (1999), who asked participants to finish a series of training tasks where they acquired high familiarity with a novel stimulus category, the “Greeble”. The idea of the Greeble is that of an artificially generated category of visual objects by slightly varying the configuration of features. They have two ‘genders’ and five ‘families’. Each Greeble has its own ‘name’ like faces. These training tasks included gender inspection, gender categorization, individual inspection, naming, and verification. The training tasks aimed to train participants to identify Greebles at the individual level (identification), as fast and accurately as they categorized these stimuli at the more general ‘family’ level (categorization). After training, Gauthier et al. found increased brain activation for Greebles in the FFA, and no difference between faces and Greebles. This suggests that the FFA does not only respond to faces, but to any category we have expertise with. The hypothesis is that the FFA tends to respond to faces merely because expertise is an important factor that leads to specialization for faces in the face area. The same results were also found for other groups of objects for which expert professionals had extensive visual expertise, such as cars, birds, fingerprints (e.g., Gauthier et al., 2000). However, all findings combined leave the debate on the functional specialization of the FFA still open, while the role of this region in face perception is undisputed (e.g., Bukach et al., 2006; Kanwisher and Yovel, 2006; McKone et al., 2007; Xu, 2005).

The occipital face area (OFA; Gauthier et al., 2000) is also a core region of the face perception network and is located on the lateral surface of the occipital lobe and in the

vicinity of the inferior occipital gyrus (Figure 8 on page 38). It is more commonly defined with a functional localizer contrast of faces vs other non-face categories such as objects, or face vs objects and scrambled images (Kovács et al., 2008; Large et al., 2008; Yovel and Kanwisher, 2005). Although the OFA has not been as extensively studied as the FFA, it has been shown to be functionally distinct from other face-sensitive areas. In the face perception network, the OFA is thought to be important during the first stage in face processing: it is involved in face detection, it receives facial and face feature information from the lateral occipital cortex (Pitcher et al., 2007), and it transmits related information to other face-sensitive areas (Calder and Young, 2005; Pitcher et al., 2011). In other words, the OFA is more sensitive to physical/structural components of face perception than the FFA (Rotshtein et al., 2005). However, it is still under debate whether/how this area could play a role in the processing of higher-order facial information. For instance, Rossion and Colleagues (2003) reported about a prosopagnostic patient with lesions in the right inferior occipital gyrus, which suggests the OFA is crucial to face identification. Other studies using transcranial magnetic stimulation (TMS) over the OFA, reported a disruption in the discrimination of facial expressions (Pitcher et al., 2008). Recently, Ambrus and Colleagues (2017a) adopted state-dependent TMS to stimulate over the rOFA during a face priming paradigm. The results suggest that the OFA is causally involved in the formation of identity-specific face representations (Ambrus et al., 2017b). More recently, the OFA was also related to process face-related semantic information in a TMS study (Eick et al., 2020). To summarize, the OFA is not limited to the processing of low-level physical features and instead has a crucial role in the encoding of other face-related information.

The visual word form area (VWFA) is located in the left occipitotemporal sulcus and is thought to be involved in the visual and prelexical processing of written words (Figure 8 on page 38; e.g., Cohen et al., 2000; Dehaene and Cohen, 2011). It is commonly defined with a functional localizer contrast of words vs other non-words categories such as strings, line-drawings, fixation (e.g., Baker et al., 2007; Cohen et al., 2002; Glezer et al., 2009). Generally, the response of VWFA reflects strictly visual and prelexical stages of

word recognition (Dehaene et al., 2002), independently of the case and font of the stimulus words (Dehaene et al., 2001; Zhou et al., 2019). Literate humans are considered experts in the recognition of words of their native language. The hemispheric lateralization of this area is consistent with language lateralization (Cai et al., 2008, 2010). For instance, previous studies found that the word-sensitive N170, which is assumed to originate in the VWFA, was already left-lateralized in children after about one year of formal school education (e.g., Cao et al., 2011; Maurer et al., 2006). Accordingly, the processing of written words of the native language shows similarities to that of other stimulus categories in which humans are experts in, for example, familiar human faces (Davies-Thompson et al., 2016; for a review, see Moret-Tatay et al., 2020; Young and Burton, 2017). Anatomically, the VWFA is in the close neighborhood of the left FFA (Matsuo et al., 2015). In fact, it has been proposed that the VWFA emerges as a result of “cultural recycling” due to the potential re-use of neural structures previously related to face and object processing (Dehaene and Cohen, 2007, 2011). It has also been proposed that many structural, connecting and functional properties of the VWFA are inherited from evolutionarily older brain circuits, such as the face-processing network (Hannagan et al., 2021).

The lateral occipital complex (LO; named as LOC in the present thesis) is one key region of the object processing network in the ventral visual stream (Figure 8 on page 38; for a review see Grill-Spector et al., 2001). Malach et al. (1995) was the first to show a stronger activation of the LOC to images of everyday objects compared to texture stimuli. Since then, most studies consistently found that the object representations in LOC are independent of low-level visual properties and physical features (e.g., Silvano et al., 2010; Vuilleumier et al., 2002). LOC is also known to be selective to all visual objects (for a review see Grill-Spector, 2003). Additionally, the LOC is sensitive to the attributes of an object, such as shape (Kanwisher et al., 1997b), and position (Cichy et al., 2011). However, it is currently under debate how the LOC contributes to higher-level object perception (Chouinard et al., 2017; Roth and Zohary, 2015; Wang et al., 2016). Specifically, inconsistent results were reported regarding object individuation (Bona et al., 2018; Eger

et al., 2004, 2008; Vuilleumier et al., 2002) and categorization (Cichy et al., 2011; Grill-Spector et al., 1999; Sayres and Grill-Spector, 2008). For example, an fMRI study by Vuilleumier et al. (2002) reported repetition suppression in the LOC when objects with different shapes were presented repeatedly, but not when objects with the same name were shown. This result suggests that shape information is processed in the LOC while the semantic information about categorization is not. In contrast, by using multivariate pattern analysis (MVPA), Eger et al. (2008) found that the fMRI activation pattern of the LOC could be discriminated at a category-dependent level, but this region can also distinguish between different exemplars of a category of artificial objects. Recently, transcranial magnetic stimulation (TMS) of the right LOC (rLOC) led to reduced individuation of different exemplars of the same object category (Bona et al., 2018). Taken together, these findings suggest a dominant role of LOC in object identity encoding.

As described above, these four areas are located in the ventral visual system and play an important role in visual object processing. Some areas show category-selectivity, such as the FFA (especially sensitive to faces) and the VWFA (especially sensitive to words). Importantly, previous studies have already shown significant RS effect in the FFA (Gauthier et al., 2000; for a review see Henson, 2016), OFA (e.g., Fox et al., 2009; Kovács et al., 2008), VWFA (e.g., Barton et al., 2010; Glezer et al., 2015), and LOC (e.g., Grill-Spector and Malach, 2001). Even the P(rep) effect was reported in the FFA, OFA and LOC (e.g., Grotheer et al., 2014; Mayrhauser et al., 2014; Summerfield et al., 2008). Thus, these four areas are the best candidates to investigate how prior experiences modulate the top-down prediction processing in the ventral visual system.

3 Empirical contributions

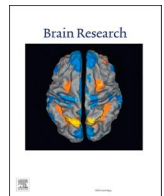
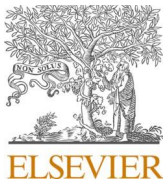
3.1 study I: Repetition probability effects for Chinese characters and German words in the visual word form area

This section corresponds to the manuscript that has been published in *Brain Research*:

Li, C., & Kovács, G. (2022). Repetition Probability Effects for Chinese Characters and German Words in the Visual Word Form Area. *Brain Research*, 147812. <https://doi.org/10.1016/j.brainres.2022.147812>

Main research question:

Does the repetition probability effect for words depend on long-term experience with the stimuli?



Research report

Repetition probability effects for Chinese characters and German words in the visual word form area

Chenglin Li, Gyula Kovács^{*}

Department of Biological Psychology and Cognitive Neurosciences, Institute of Psychology, Friedrich-Schiller-Universität Jena, Jena, Germany

ARTICLE INFO

Keywords:

Repetition suppression

VWFA

Visual word recognition

Predictive coding

ABSTRACT

The magnitude of repetition suppression (RS), measured by fMRI, is modulated by the probability of repetitions (P(rep)) for various sensory stimulus categories. It has been suggested that for visually presented simple letters this P(rep) effect depends on the prior practices of the participants with the stimuli. Here we tested further if previous experiences affect the neural mechanisms of RS, leading to the modulatory effects of stimulus P(rep), for more complex lexical stimuli as well. We measured the BOLD signal in the Visual Word Form Area (VWFA) of native Chinese and German participants and estimated the P(rep) effects for Chinese characters and German words. The results showed a significant P(rep) effect for stimuli of the mother tongue in both participant groups. Interestingly, Chinese participants, learning German as a second language, also showed a significant P(rep) modulation of RS for German words while the German participants who had no prior experiences with the Chinese characters showed no such effects. Our findings suggest that P(rep) effects on RS are manifest for visual word processing as well, but only for words of a language with which participants are highly familiar. These results support further the idea that predictive processes, estimated by P(rep) modulations of RS, require prior experiences.

1. Introduction

Over the last two decades, the theory of predictive coding became a popular framework to understand how the brain encodes sensory information (Friston, 2005; Rao and Ballard, 1999). To achieve effective and precise encoding, the predictive brain constantly matches the incoming sensory information to prior expectations, thereby minimizing their mismatch (the predictive error) (Kok and de Lange, 2015).

Recently, predictive error has been associated with the well-known reduction of the neural responses for repeated presentations of a given stimulus (Desimone, 1996). Repetition suppression (RS) is a widely studied neural phenomenon when the repetition of a stimulus leads to the attenuation of the neuronal responses compared to its first presentation (Henson and Rugg, 2003). Despite the permanent interest in the topic the precise neural mechanisms of the effect are still unclear (for a review see Grill-Spector et al., 2006). In recent years, RS has been related to predictive coding as repeating a stimulus may induce its enhanced expectation and the reduction of prediction error, expressed in RS (for a review, see Grotheer and Kovács, 2016). Summerfield et al

(2008) provided the first empirical evidence of the relationship between RS and predictive coding by modulating the statistical probability of stimulus repetitions [P(rep)] and using functional magnetic resonance imaging (fMRI). They compared the blood oxygenation level-dependent (BOLD) responses to face stimulus repetitions and alternations in two different contexts. A stronger RS effect was obtained in blocks, where the probability of repetition trials was higher when compared to blocks with less frequent repetitions. Later, the modulation of RS by P(rep) was confirmed in several subsequent studies using faces (Kovács et al., 2012; Larsson and Smith 2012). Interestingly, the P(rep) modulation of RS for non-face stimuli led to less consistent results. While some studies were able to find evidence of P(rep) modulations of RS for every-day objects (Mayrhauser et al., 2014), others failed to find it (Kovács et al., 2013). One possible explanation for these opposite results is that humans have different prior experiences with various stimulus categories, and this modulates the neural mechanisms, thereby the P(rep) effects of RS.

Indeed, it has been suggested that there is a tight connection between predictions and stimulus expertise, as gaining prior experiences leads to stronger associations among perceptual features or objects (for a review

^{*} Correspondence to: Gyula Kovács, Department of Biological Psychology and Cognitive Neurosciences, Institute of Psychology, Friedrich-Schiller-Universität Jena, Jena, Germany.

E-mail address: gyula.kovacs@uni-jena.de (G. Kovács).

<https://doi.org/10.1016/j.brainres.2022.147812>

Received 19 November 2021; Received in revised form 17 January 2022; Accepted 28 January 2022

Available online 1 February 2022

0006-8993/© 2022 Published by Elsevier B.V.

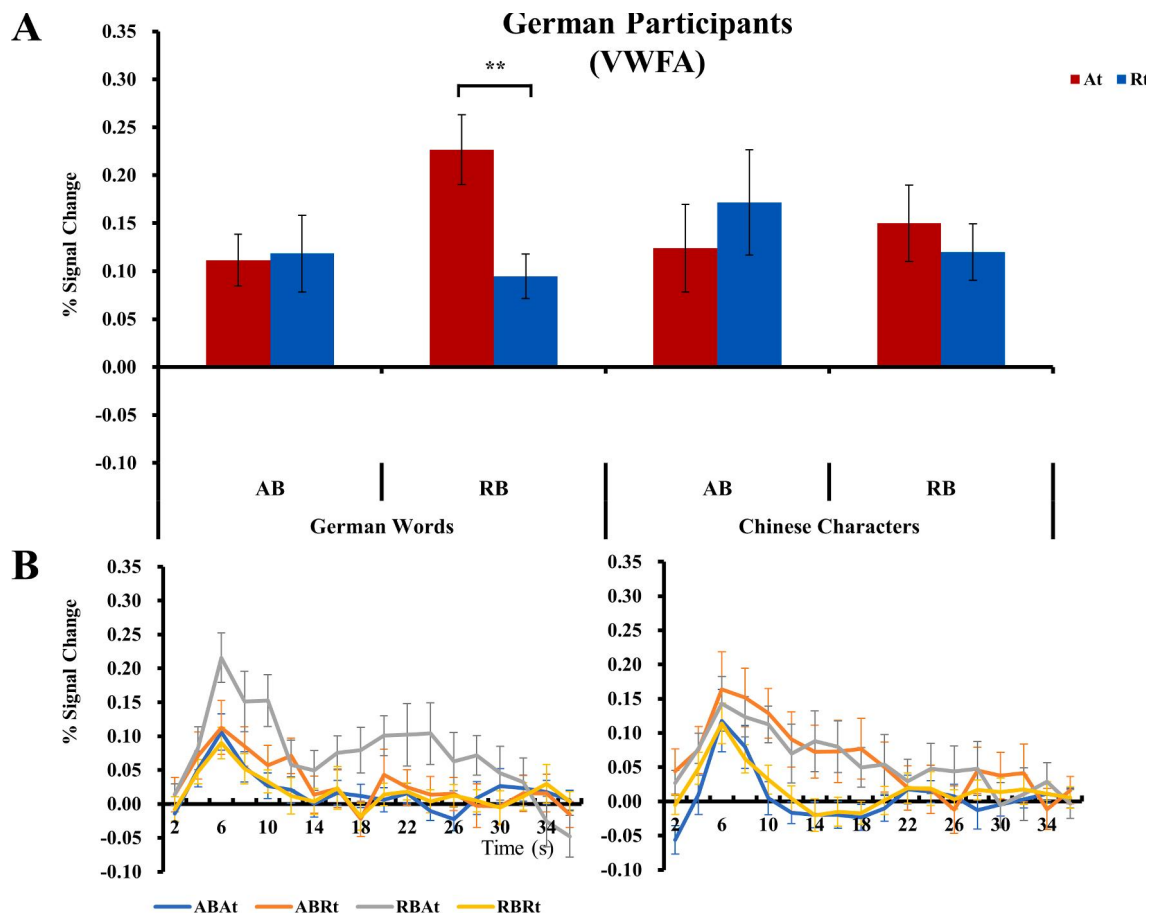


Fig. 1. The BOLD results of the VWFA for German participants. **A**, Average peak activation profiles of the VWFA for At and Rt trials for the German Words and for the Chinese Characters separately. **B**, The time course of fMRI activity in the VWFA for German Words and for Chinese Characters. The hemodynamic response functions were derived from an FIR model with 2 s time bins. Error bars represent standard errors of means. ** $p < 0.01$ (Fisher's *post hoc* comparisons).

see Cheung and Bar, 2012). To explicitly test this hypothesis, Grotheer and Kovács (2014) adopted the paradigm of Summerfield et al., (2008), using two non-face stimulus categories with different levels of prior expertise: individual Roman letters and novel pseudo-letters. The results revealed P(rep) modulations on RS in the Letter Form Area (LFA; Thesen et al., 2012) for the well-known Roman letters, but not for the unfamiliar pseudo-letters, suggesting that prior experiences might indeed alter the underlying neural mechanisms of RS. However, it remains unclear whether this P(rep) effect (a) is observed for other stimuli, such as words as well; (b) can also be observed in other cortical areas than the LFA and (c) if learning a new stimulus category changes the P(rep) effects after a given training period or if life-long experiences are required, such as the one we all have with conspecific faces or literate persons have with characters of the familiar alphabet.

To address these questions, we performed an fMRI experiment, estimating the P(rep) modulation of RS for Chinese characters and German words in two participant populations: literate German participants having no experience with the Chinese language and native Chinese participants learning German as a second language for at least two years. We focused on the visual word form area (VWFA), located in the left occipitotemporal sulcus as previous studies suggested that this area reflects the visual and pre-lexical processes of word recognition (Cohen et al., 2000; Dehaene and Cohen, 2011). Our findings suggest that the P(rep) modulation of RS is manifest for visual word processing as well, but only if the language is known to the participants. These results support further the theory that the neural mechanisms of RS are modulated by prior experiences.

2. Results

2.1. Behavioral performance

German participants detected the target stimuli on average with 93.1 and 94.1% (\pm SE: 2.0 and 1.0 %) accuracy (average reaction time: 1143 and 1153 ms (\pm SE: 24 and 25 ms)) during the German word and Chinese character runs, respectively. The performance was not significantly different for the different blocks for either language. Chinese participants showed a similar pattern (average performance: 90.4 and 93.3% (\pm SE: 3.3 and 1.7 %); average reaction time: 1117 and 1153 ms (\pm SE: 44 and 28 ms) during the German word and Chinese character runs, respectively) without any significant effect of blocks. In addition, no participant was aware of the distinct manipulations of P(rep) across the different blocks, based on their self-reported results after the fMRI measurement.

2.2. Neuroimaging results: German participants

2.2.1. VWFA

A three-way repeated measures ANOVA revealed no significantly different activations for the German words and Chinese characters in the VWFA of the German participants (main effect of stimulus language: $F(1, 19) = 0.02$, $p = 0.891$, $\eta_p^2 = 0.001$).

We found a significant main effect of trial type for the words of the mother tongue of the German participants (2-way ANOVA with block and trial as factors; main effect of trials: $F(1, 19) = 9.94$, $p = 0.005$, $\eta_p^2 = 0.34$), with stronger BOLD signals in At (average BOLD signal \pm SE:

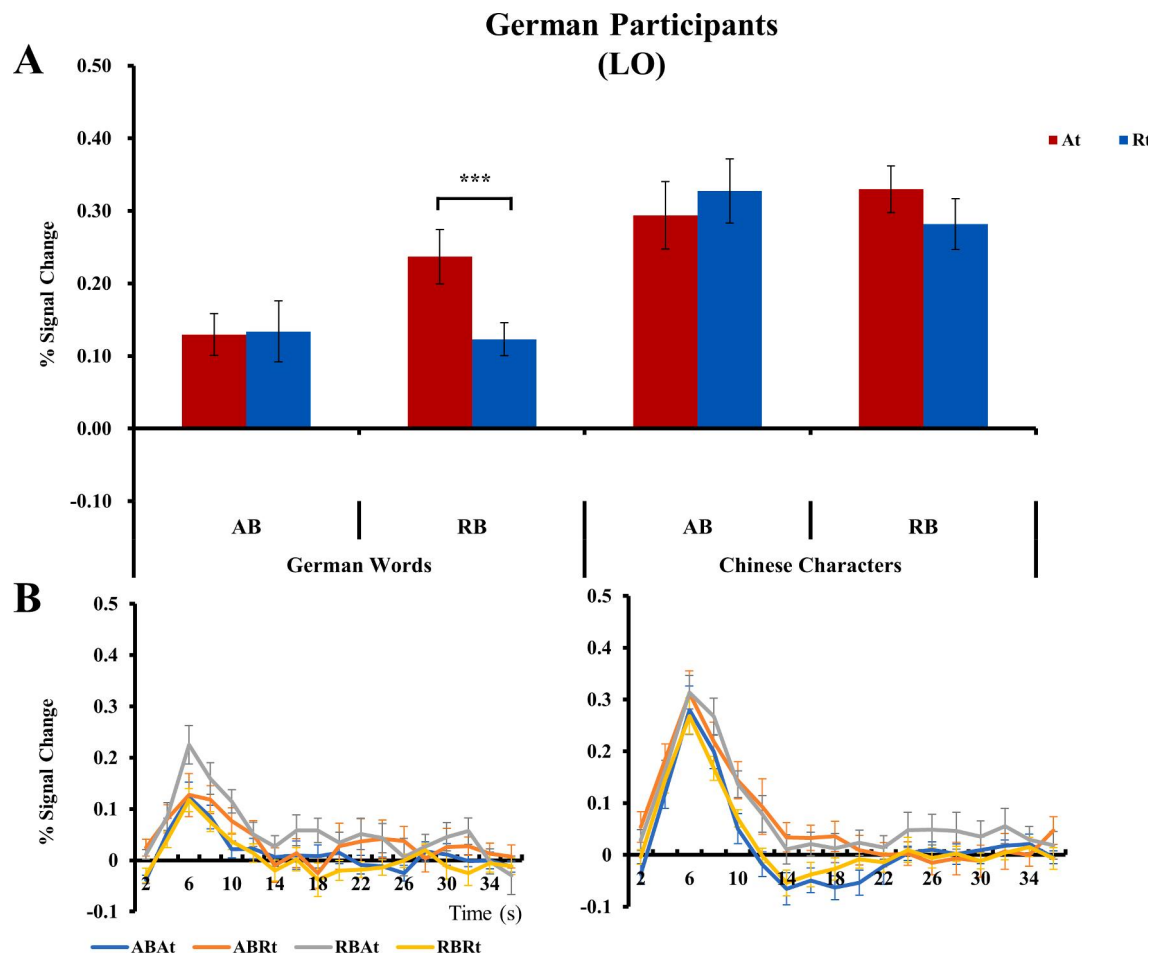


Fig. 2. The BOLD results of the LO for German participants. **A**, Average peak activation profiles of the LO for At and Rt trials for the German Words and for the Chinese Characters separately. **B**, The time course of fMRI activity in the LO for German Words and for Chinese Characters. The hemodynamic response functions were derived from an FIR model with 2 s time bins. Error bars represent standard errors of means. *** $p < 0.001$ (Fisher's *post hoc* comparisons).

0.169 ± 0.021) as compared to Rt trials (0.107 ± 0.028) (Fig. 1). More importantly, the interaction of block and trial type was also significant for German words ($F(1, 19) = 5.62, p = 0.028, \eta_p^2 = 0.23$): post-hoc tests revealed a significant RS effect in the repetition (Fisher *post hoc* test: $p = 0.005$), but not in the alternation blocks ($p = 0.869$). This suggests that P (rep) modulates the magnitude of RS strongly for words of the mother tongue in the VWFA. This finding extends the previous findings of Grotheer & Kovács, (2014) who found similar strong interactions for Roman letters in the LFA of German participants.

Interestingly, the same analysis showed no significant effects for Chinese characters (all p values ≥ 0.22), suggesting that neither repetition suppression, nor its modulation by P(rep) is present in the VWFA for the written symbols of an unknown language.

2.2.2. LO

The BOLD signal in the LO was significantly stronger for Chinese characters, as compared to German words in the German participants (main effect of stimulus language: $F(1, 19) = 43.45, p < 0.001, \eta_p^2 = 0.70$). This is in line with those previous findings which found stronger activations in the LO for Chinese characters, as compared to French words in French participants (Szwed et al., 2014).

Area LO did not show either a significant main effect of hemisphere ($F(1, 19) = 0.36, p = 0.554, \eta_p^2 = 0.02$), or any interaction of hemisphere with stimulus, trial or block (all p values ≥ 0.166). For the German words the main effect of trial showed a non-significant tendency ($F(1, 19) = 2.71, p = 0.116, \eta_p^2 = 0.13$) and there was a significant main effect of block ($F(1, 19) = 8.42, p = 0.009, \eta_p^2 = 0.31$) (Fig. 2). Similar to

the VWFA, the block \times trial interaction was also strongly significant in the LO ($F(1, 19) = 10.39, p = 0.004, \eta_p^2 = 0.35$) and post-hoc tests revealed a significant RS effect in the repetition (Fisher *post hoc* test: $p = 0.0001$), but not in the alternation blocks ($p = 0.819$). This suggests somewhat noisier, but similar processing of words of the mother tongue in the LO as in the VWFA.

For the Chinese characters neither the main effects of trial or block, nor their interaction approached significance (all p values ≥ 0.21). This suggests that the LO, just like the VWFA, processes words of the mother tongue and an unknown language differentially.

2.3. Neuroimaging results: Chinese participants

2.3.1. VWFA

The three-way repeated measure ANOVA revealed a significant main effect of stimulus language ($F(1, 19) = 10.63, p = 0.004, \eta_p^2 = 0.36$), with stronger BOLD signal for the Chinese characters (0.316 ± 0.046) as compared to the German words (0.191 ± 0.030) (Fig. 3) in the Chinese participants.

Unlike for the German participants, in the Chinese we found significant RS effects in the VWFA both for German words (main effect of trial type: $F(1, 19) = 18.51, p < 0.001, \eta_p^2 = 0.49$) and for Chinese characters (main effect of trial type: $F(1, 19) = 22.90, p < 0.001, \eta_p^2 = 0.55$). More importantly, P(rep) modulations on RS were found for the characters of the mother tongue (block \times trial for Chinese characters: $F(1, 19) = 4.69, p = 0.043, \eta_p^2 = 0.20$) as well as for the German words (block \times trial: $F(1, 19) = 6.72, p = 0.018, \eta_p^2 = 0.26$). Post-hoc tests revealed that strong

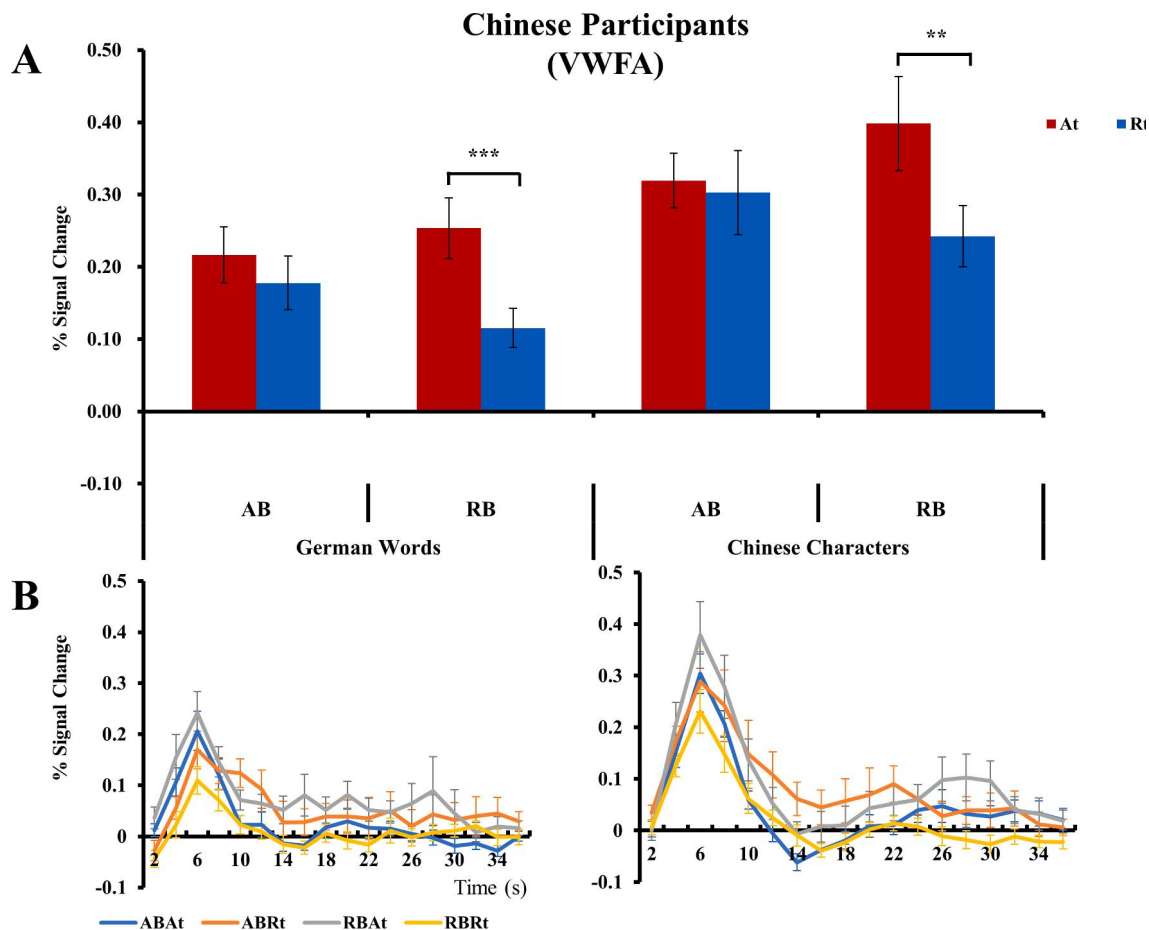


Fig. 3. The BOLD results of the VWFA for Chinese participants. **A**, Average peak activation profiles of the VWFA for At and Rt trials in the German Word and Chinese Character runs, respectively. **B**, The time course of fMRI activity in the VWFA for German Word and Chinese Character runs, respectively. The HRF curves were derived from an FIR model with 2 s time bins. Error bars represent standard errors of means. ** $p < 0.01$, *** $p < 0.001$ (Fisher's *post hoc* comparisons).

RS effects were observed in repetition blocks for both language stimuli (Fisher *post hoc* test: for German words: $p = 0.000064$, for Chinese characters: $p = 0.0028$), but not in the alternation block (Fisher *post hoc* test: for German words: $p = 0.168$, for Chinese characters: $p = 0.719$).

These results extend further the previous results with familiar Roman letters (Grotheer and Kovács, 2014) and words (present study) in the sense that the characters/words of the mother tongue elicit both a significant RS in the relevant processing areas as well as its modulation by P (rep), independently of language and character-set. Importantly, the modulatory effects extend towards a second language as well, acquired by the participants only recently.

2.3.2. LO

The activity of LO was significantly stronger for Chinese characters, as compared to German words in the Chinese participants as well (main effect of stimulus language: $F(1, 19) = 55.52$, $p < 0.001$, $\eta_p^2 = 0.75$). Area LO showed no main effect of hemisphere ($F(1, 19) = 0.95$, $p = 0.341$, $\eta_p^2 = 0.05$), or any interaction of hemisphere with stimulus, trial or block (all p values larger than 0.080) in the Chinese participants either. For the Chinese characters LO, just like the VWFA, showed a significant RS (main effect of trial: $F(1, 19) = 5.66$, $p = 0.028$, $\eta_p^2 = 0.23$) (Fig. 4). Also mirroring the findings of the VWFA, the interaction of RS with P(rep) was also significant for the characters of the mother tongue of the Chinese participants (block \times trial: $F(1, 19) = 5.04$, $p = 0.037$, $\eta_p^2 = 0.21$) and post-hoc tests revealed a significant RS effect in the repetition (Fisher *post hoc* test: $p = 0.0068$), but not in the alternation blocks ($p = 0.888$). For the German words LO showed a similar pattern. We observed a strong RS (main effect of trial: $F(1, 19) = 8.89$, $p = 0.008$,

$\eta_p^2 = 0.32$) as well as an interaction that just exceeded the 0.05 p value threshold (block \times trial: $F(1, 19) = 4.14$, $p = 0.056$, $\eta_p^2 = 0.18$) and post-hoc tests revealed a significant RS effect in the repetition (Fisher *post hoc* test: $p = 0.017$), but not in the alternation blocks ($p = 0.782$). Overall, the results obtained for the LO of Chinese participants are in line with those of the German participants in the sense that the LO and the VWFA seem to reflect similar P(rep) modulation of RS.

2.4. Whole-brain analyses

To test further whether the effects of stimulus, repetition, or repetition probability can also be observed outside the predefined ROIs of our study we performed a group-level whole-brain analysis, separately for the German and Chinese participants. We applied a rigorous threshold of $p < 0.05$ (FWE corrected) with a minimum cluster size of >50 voxels (Grotheer and Kovács, 2014). These analyses revealed a significant main effect of stimuli for both the German and Chinese participants: the activity of the middle occipital gyrus (for average coordinates see Supplementary Material, Table 2) was stronger for Chinese characters compared to German words. This is consistent with those previous studies which showed strong differences between the reading networks for Chinese and Latin alphabets (Szwed et al., 2014; Tan et al., 2005; Wu et al., 2012). No other statistically significant clusters were found when testing for the main effects of blocks, trials or their interaction. To make sure that no region is overlooked by the commonly applied, rigorous FWE corrected threshold we also used a more liberal ($p < 0.001_{\text{uncorrected}}$) threshold to reanalyze our data. In this case the At $>$ Rt contrast revealed one small cluster over the middle temporal gyrus (MNI

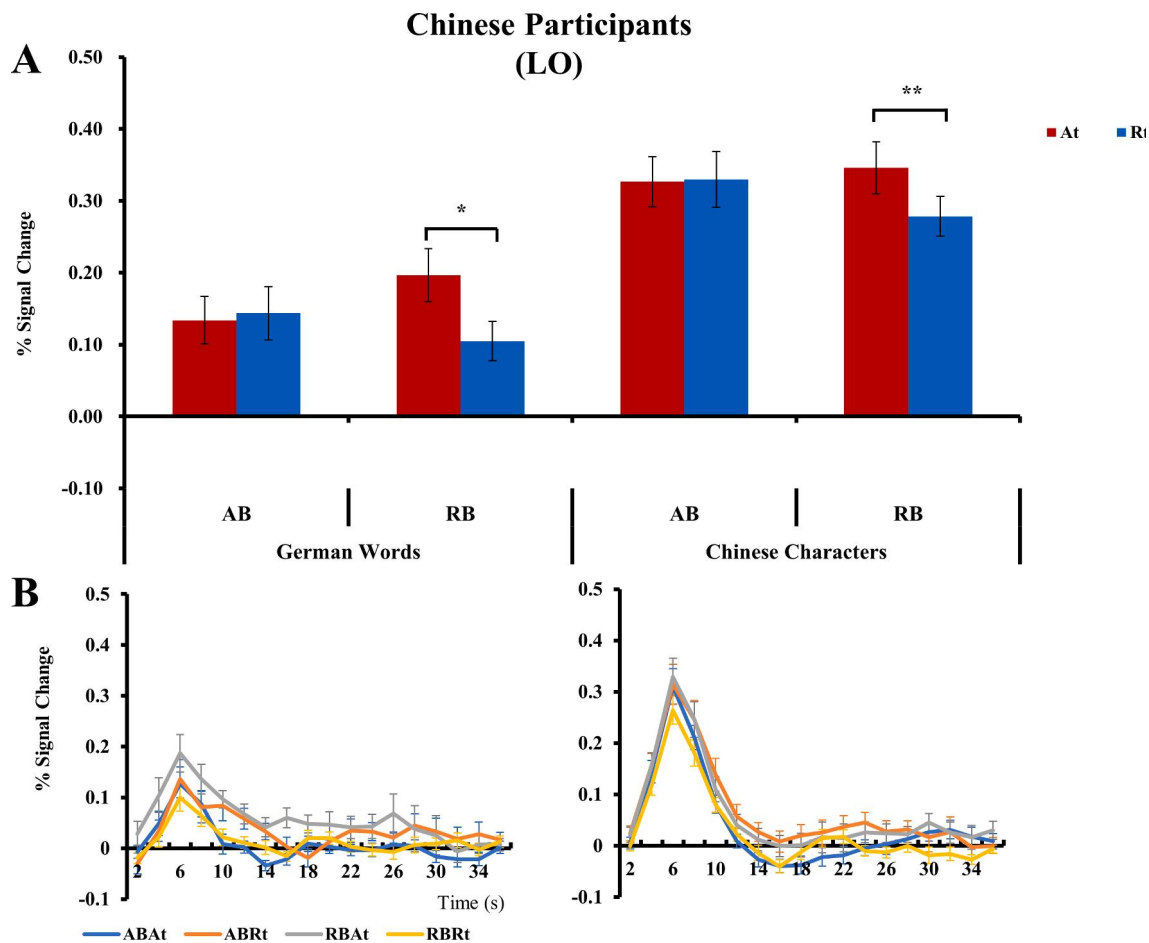


Fig. 4. The BOLD results of the LO for Chinese participants. **A**, Average peak activation profiles of the LO for At and Rt trials for the German Words and for the Chinese Characters separately. **B**, The time course of fMRI activity in the LO for German Words and for Chinese Characters. The hemodynamic response functions were derived from an FIR model with 2 s time bins. Error bars represent standard errors of means. * $p < 0.05$, ** $p < 0.01$ (Fisher's *post hoc* comparisons).

coordinates: $-58, -48, 8$; cluster size: 10 voxels) in Chinese participants, suggesting repetition related response reduction. All statistically significant clusters of stimulus effect are listed in [Supplementary Table 2](#).

3. Discussion

The main findings of the current study are the followings. (1) The VWFA and the LO show RS and its modulation by P(rep) for visually presented words/characters of the mother tongue but not an unknown language, in two different language populations. This result extends those previous findings which showed similar P(rep) effects within the fusiform face area and the letter form area. (2) The significant P(rep) effect for German words in Chinese participants and its lack for Chinese characters in Germans supports the idea that acquiring experience with a given stimulus category may be an important factor in determining P(rep) effects on RS. Further, it also suggests that learning a second language leads to changes of the neural mechanisms underlying RS in the VWFA.

To the best of our knowledge, the present study is the first to provide direct empirical evidence for the modulation of RS to written words by top-down probabilistic mechanisms. This result adds to the stimulus categories (Faces, [Summerfield et al., 2008](#); Objects: [Mayrhauser et al., 2014](#); Familiar Roman letters: [Grotheer and Kovács, 2014](#)) for which the significant P(rep) modulation of RS suggests the involvement of predictive coding mechanisms at least in human neuroimaging studies. Note however, that for every-day objects neither some neuroimaging

nor non-human primate single-cell recordings found such effects ([Kaliukhovich and Vogels, 2011](#); [Kovács et al., 2013](#)). Previously [Grotheer and Kovács \(2014\)](#) observed P(rep) effect in the LFA and LO for familiar Roman letters, but not for novel pseudo-letters. They suggested that an important factor in determining the emergence of P(rep) effect is the prior experience which the observer had with the stimuli. Indeed, the stimulus categories for which P(rep) modulation of RS was successfully shown are those where participants have typically extensive prior experience with (faces, words, and Roman letters). This conclusion is supported by the fact that prior experiences generally play a key role in forming predictions ([Kok and de Lange, 2015](#)).

The current study mainly focused on the VWFA, located in the left occipitotemporal sulcus and involved in the visual and prelexical processing of words ([Cohen et al., 2000](#); [Dehaene and Cohen, 2011](#)). Previous studies have already shown significant RS in the VWFA both by neuroimaging (e.g., [Barton et al., 2010](#); [Glezer et al., 2015](#)) and electrophysiological methods (e.g., [Cao et al., 2015](#); [Li et al., 2019](#); [Mercure et al., 2011](#)). Literate humans are considered as experts in the recognition of words of their mother tongue. Accordingly, the processing of written words of the mother tongue shows similarities to that of another stimulus categories of which humans are experts, for example familiar human faces ([Davies-Thompson et al., 2016](#); for a review, see [Moret-Tatay et al., 2020](#); [Young and Burton, 2017](#)). Therefore, it is not surprising that the VWFA is located over the occipito-temporal cortex ([Cohen and Dehaene, 2004](#)) in the close neighborhood of the fusiform face area (FFA; [Kanwisher et al., 1997](#)). In fact, it has been proposed that the VWFA emerges as the result of “cultural recycling” due to the

potential re-use of neural structures previously related to face and object processing (Dehaene and Cohen, 2007). As a result, it has also been proposed that many structural, connectivity and functional properties of the VWFA are inherited from evolutionarily older brain circuits, such as the face-processing network (Hannagan et al., 2021). Thus, it is not surprising that similar P(rep) modulation of RS has been found within the VWFA for written words and in the FFA for faces (Summerfield et al., 2008).

Previous studies, using the identical paradigm as the present work have found P(rep) modulations on RS for familiar Roman letters, but not for novel pseudo-letters (Grotheer and Kovács, 2014). Therefore, it has been suggested that an extensive prior experience is necessary to the P(rep) effect to emerge. Here we compared the P(rep) effects on RS for the written words of two very different languages as mother tongue. Thus, our results not only extended the P(rep) effects to visual word processing, but also show the same P(rep) effects in two separate participant groups and for two very different lexical stimuli: words composed of letters of the German alphabet and Chinese characters, which are logograms, representing words or morphemes. Our results are also in line with previous studies which found similar RS effects both for the written words of alphabetic and logographic languages (Cao et al., 2014; Maurer et al., 2005). Importantly, the current results also suggest that the P(rep) modulation of RS is a universal effect, which develops across languages and cultures and is independent of the alphabetical or logographic nature of the written characters.

Another result of the current study is that the Chinese participants showed a significant P(rep) effect on RS for German words, which they study as a second language for several years. Explicit learning is an obvious way to obtain experience and knowledge with something. For instance, perceptual learning can improve various perceptual abilities, ranging from discriminating visual features (for example contrast, orientation or shape) (Schoups et al., 1995; Yu et al., 2004; for a review, see Bi and Fang, 2013), to object recognition capacities (Op de Beeck and Baker, 2010). Previous studies indicated that the human visual system is generally more sensitive to learned compared to novel objects (Grill-Spector et al., 2000; Sigman et al., 2005; Op de Beeck et al., 2006). For instance, Song et al. (2010) trained two groups of participants, either in a visual association or a shape discrimination task. Their results showed a stronger learning effect in the VWFA as compared to the LO when trained in the association learning tasks, but a greater learning effect in the LO when trained in the shape discrimination task. This suggests that the VWFA is sensitive to top-down learning contexts during the encoding of objects. Our results suggest that the top-down effects, elicited by repetition probability modulations of stimulus blocks, are formed only after long-term accumulation of experiences with the stimuli.

One limitation of the current study is its unbalanced nature: we tested native German speaker, who were entirely naïve regarding Chinese characters as well as native Chinese speakers who were also exposed to German words in the last years. Ideally, in a fully balanced design, we should have tested Chinese participants who were never exposed to German words and Latin characters as well as Chinese-speaking German participants. On the one hand, as the study was conducted in Germany it was impossible to find Chinese participants who are entirely naïve regarding German words. In fact, in the light of the current educational system in China, where children are almost always exposed to western languages (and thereby to Latin characters) during their education we feel that the testing of such a population is not feasible in the future either. On the other hand, due to the scarcity of Chinese-speaking German participants in our experimental participant population (we only could record 5 native German participants, who learned Chinese for at least 2 years until the end of the current experiment) their data was also not analyzed here. However, two arguments suggest that the current data suggests the role of experience in modulating the neural mechanisms of RS. First, we found P(rep) modulations on RS for two fundamentally different languages, one with alphabetic

and another with logographic orthographies, suggesting its independence of the writing system. Second, the modulation effect transferred across writing systems, as the findings for German words in Chinese participants show. Nonetheless, for a full and language-independent support of the conclusion that P(rep) modulations of RS require experiences with the words of a particular language one should replicate the current experiment with a fully balanced design ideally in two different native populations with and without experiences with the other language. Alternatively, a longitudinal training study could also be performed in the future to investigate the modulation of the P(rep) effect during the learning of a foreign language.

In addition to the VWFA, the LO (Malach et al., 1995) is commonly considered as a key area involved in object processing, but is also responsive to faces (Lerner et al., 2001). The present study shows that the LO responds to German words and Chinese Characters as well. Previous studies found a significant modulation of P(rep) on RS in the area for various stimulus categories (faces: Kovács et al., 2012; Larsson and Smith, 2012; Familiar Roman letters: Grotheer and Kovács, 2014), but not for novel pseudo-letters. Consistent with these findings, the present study shows that the characters/words of the mother tongue induce a significant P(rep) modulation on RS in the LO, and these modulatory effects extend towards a second language acquired by the participants relatively recently. This indicates that P(rep) effects on RS in LO are not specific to one stimulus category, and that they depend on the prior experiences with the stimuli, such as the letters and words of the familiar alphabet and language for literate persons. Interestingly, some previous studies using the hemifield change paradigm found that the VWFA and LO reflect different sensitivity for repetition suppression in word recognition (Strother et al., 2016, 2017; Zhou et al., 2019). For instance, Zhou et al. (2019) found that the VWFA show both font-sensitive and font-invariant effects, while the LO reflects greater font sensitivity. As the modulatory pattern of prior experiences in the LO was similar to that in the VWFA, our current results suggest that the LO and the VWFA reflect similar neural processing stages.

We argue that that written word processing is an optimal cognitive process to test the validity of the idea of the predictive brain from an experimental perspective. Generally, word recognition includes visual, auditory as well as semantic processes. Previous studies suggested that predictive mechanisms are important in determining the RS, observed in the LFA for simple graphemes or letters (Grotheer and Kovács, 2014), in the superior temporal gyrus for word stress processing (Honbolygó et al., 2020) and in the superior temporal gyrus for lexical representations (Wang et al., 2021) as well as for semantic processes (Qin et al., 2021). These results combined with the current study, in which the involvement of predictive processes is suggested for visual word recognition in the VWFA, suggest that top-down probabilistic prediction effects are reflected across several stages of language processing. Future studies should address this issue in a more systematic manner by using for example artificial language learning.

In conclusion, the present study shows that top-down effects, expressed in probabilistic manipulations of RS can be observed in the VWFA/LO and this effect is only manifest for words of a language with which participants have extensive prior experiences. These findings support the idea that predictive processes, measured by P(rep) modulation, require extensive prior experiences.

4. Methods

4.1. Participants.

Twenty healthy German participants with normal or corrected-to-normal vision took part in the experiment (15 females; mean age = 22.4 years; SD = 3.6 years; four left-handed). These participants were native German speakers and had no experience with Chinese language.

Twenty-two healthy Chinese participants were also recruited who were living in Germany. Two of these were excluded from the analysis

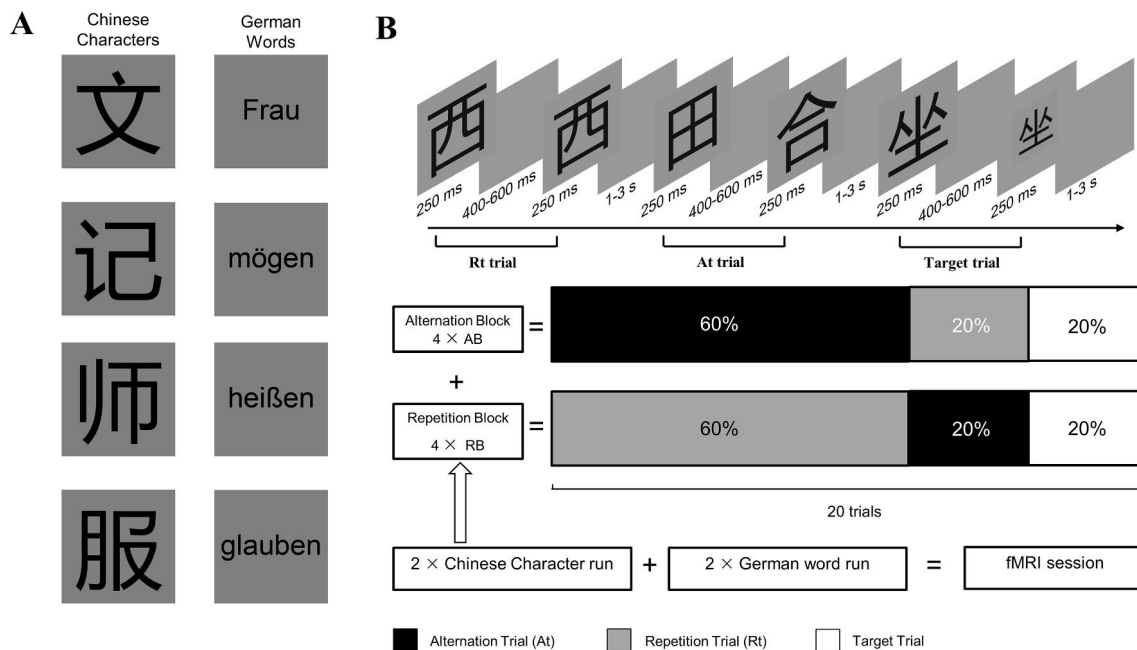


Fig. 5. A, Examples of the words stimuli (Left column, Chinese Characters. Right column, German Words) used in the experiment. B, Schematic illustration of the experimental design and task procedures. Rt-repeated trials, At-alternation trials, AB: alternation block, RB: repetition block.

due to their poor German knowledge as they recognized only 20% and 30% of the test words. The remaining 20 participants (12 females; mean age = 26.2 years; SD = 2.7 years; one left-handed) had normal or corrected-to-normal vision. On average, they have studied German for more than 4 years ($M \pm SD = 4.8 \pm 2.1$ years).

The sample size was determined, based on previous studies testing the same P(rep) effects for other stimulus categories and finding significant effects for faces (Grotheer and Kovács, 2014; Summerfield et al., 2008). Participants received partial course credits, monetary compensation, or their own 3D-printed brain-model as a compensation. All participants were informed about the experimental procedures and completed a written consent prior to the experiment, and they had no history of any neurological disorders. The research protocol was approved by the ethics committee of the Friedrich-Schiller-Universität Jena and conducted in accordance with the guidelines of the Declaration of Helsinki.

4.2. Stimuli and task

The experimental design of the present study was identical to the previous studies of our laboratory (Grotheer and Kovács, 2014), with the exception that 240 high-frequency German words and 240 high-frequency Chinese Characters were used as stimuli. The German words were selected from an online German linguistic database (dlexDB (available at www.dlexdb.de); Heister et al., 2011), ranging in length from 4 to 7 letters and were presented with a height of 1.2° visual angle from a viewing distance of 100 cm. The Chinese characters were selected from the modern Chinese frequency dictionary (Wang et al., 1986) with the number of strokes varying between 4 and 8 and subtending angles of $4.8^\circ \times 4.8^\circ$ from a viewing distance of 100 cm. The German words and Chinese characters were presented in standard typefaces (Sans Serif and Simhei fonts, respectively; for examples see Fig. 5A). No significant difference in word frequency between German words ($M = 0.59/1000$, range: 0.14/1000 – 4.93/1000) and Chinese characters ($M = 0.61/1000$, range: 0.17/1000 – 1.19/1000; $t(478) = 0.408$, $p = 0.683$) was present. All stimuli were presented centrally on a uniform gray background, back-projected via an LCD video projector onto a translucent circular screen. Psychtoolbox (Version 3.0.15) was used for stimulus presentation and behavioral response collection, controlled by Matlab R2013a

(The MathWorks).

The experimental procedures were identical to Grotheer and Kovács (2014), except for stimuli. Participants were asked to complete four experimental runs, two with German words and two with Chinese character stimuli, in a counterbalanced order. In each trial (Fig. 5B), a stimulus-pair was presented (exposition time: 250 ms each), separated by an inter-stimulus-interval (ISI) of 400–600 ms (randomized across trials), and randomly followed by an Inter-trial-interval (ITI) of 1,000 ms or 3,000 ms. The first stimulus (S1) was either identical to (Repetition trials, Rt) or different from the second one (S2; Alternation trials, At). In addition, two different types of blocks were created, each repeated four times within a single run. In the Repetition Blocks (RB), 75% of the trials were Rt while 25% were At. In contrast, the Alternation Blocks (AB) were composed of 75% At and 25% Rt (Summerfield et al., 2008). Therefore, each run included 160 trials of the four different conditions (ABAt, ABRT, RBAAt, and RBRT) in a randomized order. To avoid local feature adaptations, the size of either S1 or S2 (chosen randomly) was reduced by 18% during each trial. The design and parameters of the German word and Chinese character runs were identical, except for the stimuli. Participant's task was to maintain central fixation and signal the occurrence of stimuli which were 60% smaller than the rest. Such target trials composed 20% of all trials (equally for Rt and At trials) and were modelled explicitly during the analysis but were not analyzed further.

4.3. Imaging parameters and data analysis

Imaging data were collected on a 3-Tesla MRI scanner (Siemens MAGNETOM Prisma, Erlangen, Germany), with a 20-channel head coil. T2*-weighted images were collected with the following parameters: FOV = $204 \times 204 \text{ mm}^2$; 34 slices; TR = 2000 ms; TE = 30 ms; flip angle = 90° ; voxel size: $3 \times 3 \times 3 \text{ mm}^3$. High-resolution T1-weighted images (192 slices; TR = 2300 ms; TE = 3.03 ms; flip angle = 90° ; voxel size: $1 \times 1 \times 1 \text{ mm}^3$) were acquired to obtain 3D structural scans with the same 20-channel head coil. Data were pre-processed using SPM12 (Wellcome Department of Imaging Neuroscience, London, UK). Briefly, the functional images were slice-timed, realigned, co-registered to structural scans. The functional images were normalized to the MNI-152 space, resampled to $2 \times 2 \times 2 \text{ mm}$ resolution, and spatially smoothed using an 8-mm Gaussian kernel.

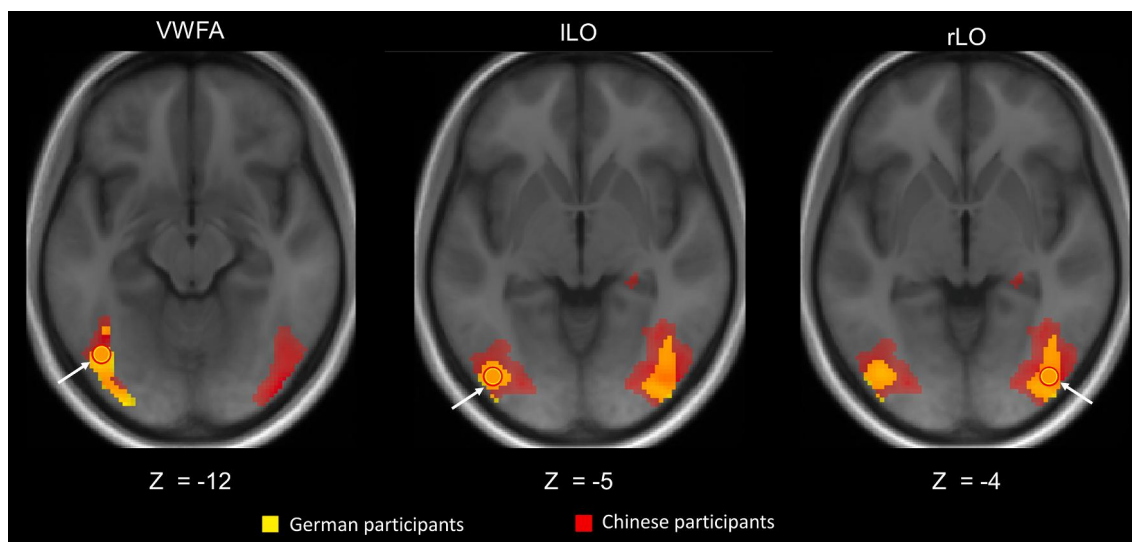


Fig. 6. Group-level ROIs in German and Chinese participants. Yellow and red areas indicate the results in German and Chinese participants, respectively. The circles mark the average coordinates for both groups.

Regions of interests (ROIs) were defined by additional functional localizer runs. Chinese characters, German words, line drawings of objects, and Fourier-randomized noise patterns were presented (230 ms exposure time; 20 ms inter-stimulus-interval) in blocks of 10 s, interrupted by breaks of 10 s and repeated four times. Each block included 40 images, which the size of 480x 480 pixels with a grey background. Participants were simply asked to pay attention to the stimuli and maintain fixation.

As previous studies, applying similar paradigms found that for faces the FFA while for Roman letters the LFA showed the largest effects (Grotheer and Kovács, 2014; Kovács et al., 2012, 2013; Summerfield et al., 2008), here we concentrated on the VWFA and the lateral occipital complex (LO; Malach et al., 1995). For the German participants the VWFA was localized individually by contrasting German words with noise images and its location was established as the local maximum from the t-maps with a threshold of $p_{FWE} < 0.05$ ($n = 11$) or $p < 0.001$ uncorrected ($n = 9$) on the single-subject level. The average MNI coordinates (\pm SE) for the VWFA were -44.8 (1.3), -65.0 (2.6), -11.0 (1.1). The individual coordinates are listed in the [Supplementary Table 1](#). In addition to the VWFA, we also estimated the BOLD signal in the LO as previous studies showed strong and significant $p(\text{rep})$ effects in this area for Roman letters (Grotheer and Kovács, 2014). The LO [average MNI coordinates (\pm SE): 42.1 (1.3), -78.0 (1.5), -3 (1.4) for the right, and -41.7 (1.2), -78.9 (1.4), -5.3 (1.6) for the left hemisphere] was localized individually by contrasting line drawing of objects with noise images and the location was established as the local maximum from the t-maps with a threshold of $p_{FWE} < 0.05$ on the single-subject level.

For the Chinese participants the VWFA was localized by contrasting Chinese characters with noise images individually and the location was established as the local maximum from the t-maps with a threshold of $p_{FWE} < 0.05$ ($n = 13$) or $p < 0.001$ uncorrected ($n = 7$) on the single-subject level. The mean MNI coordinates (\pm SE) for the VWFA were -43.6 (0.9), -65.7 (2.6), -13.8 (1.0). The coordinates of the VWFA did not differ significantly across participant groups (X coordinate: $t(38) = 0.448$, $p = 0.656$; Y coordinate: $t(38) = 0.205$, $p = 0.838$; Z coordinate: $t(38) = 2.015$, $p = 0.051$). The LO [average MNI coordinates (\pm SE): 41.8 (1.2), -79.0 (1.3), -4.3 (1.6) for right, and -42.3 (1.1), -78.7 (1.0), -4.7 (1.4) for left hemisphere] was localized individually by contrasting line drawing of objects with noise images and the location was established as the local maximum from the t-maps with a threshold of $p_{FWE} < 0.05$ on the single-subject level. The coordinates of the LO was also similar

across participant groups (all p values ≥ 0.539). Canonical hemodynamic response functions were extracted using MarsBaR 0.44 toolbox for SPM 12 (Brett et al., 2002). The group-level ROIs in German and Chinese participants were visualized using xjView toolbox (<http://www.alivelearn.net/xjview>) and the average locations of VWFA and LO are shown in Fig. 6, separately for German and Chinese participants. The analysis of these coordinates (between-subject t-tests, separately for x, y and z coordinates; $p > 0.05$ for every comparison) revealed that despite the fact that we used different contrasts to determine the location of these areas the VWFA and LO of both groups are overlapping in the two participant groups.

To directly compare with previous studies that used the same paradigm to investigate the repetition probability effects for objects and letters (Grotheer and Kovács, 2014; Kovács et al., 2013), the peak BOLD values were extracted from the event-related runs for the non-target trials in the VWFA and were analyzed by repeated-measures ANOVAs with stimulus language (German words vs. Chinese characters), block type (AB vs. RB) and trial type (At vs. Rt) as within-subject factors, separately for the German and Chinese participants. Additionally, since the general shape and visual complexity of the German words and Chinese characters are different (Tan et al., 2001), we performed another 2×2 repeated-measures ANOVAs with block type (AB vs. RB) and trial type (At vs. Rt), separately for German words and Chinese characters in the German and Chinese participants. For the LO, we conducted repeated-measures ANOVA with hemisphere (left vs. right), stimulus language (Chinese Characters vs. German words), block (AB vs. RB) and trial type (At vs. Rt) as within-subject factors, separately for the German and Chinese participants. All multiple comparisons of post-hoc tests were corrected by the Fisher's method.

CRedit authorship contribution statement

Chenglin Li: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. **Gyula Kovács:** Conceptualization, Supervision, Writing – review & editing, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by a Deutsche Forschungsgemeinschaft Grant [KO3918/5-1]. Chenglin Li was supported by the China Scholarship Council (CSC) scholarship (201808330399) during the study. The authors thank Sophie-Marie Rostalski for her comments on the paper; Marie Wächter and Wenbo Wang for their help in the stimuli collection and participant recruitment.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.brainres.2022.147812>.

References

- Barton, J.J., Fox, C.J., Sekunova, A., Iaria, G., 2010. Encoding in the visual word form area: An fMRI adaptation study of words versus handwriting. *J. Cognit. Neurosci.* 22 (8), 1649–1661. <https://doi.org/10.1162/jocn.2009.21286>.
- Bi, T., Fang, F., 2013. Neural plasticity in high-level visual cortex underlying object perceptual learning. *Frontiers in Biology* 8 (4), 434–443. <https://doi.org/10.1007/s11515-013-1262-2>.
- Brett, M., Johnsrude, I.S., Owen, A.M., 2002. The problem of functional localization in the human brain. *Nat. Rev. Neurosci.* 3 (3), 243–249. <https://doi.org/10.1038/nrn756>.
- Cao, X.H., Jiang, B., Li, C., He, Z.Q., 2014. Rapid adaptation effect of N170 for printed words. *Percept. Mot. Skills* 119 (1), 191–202. <https://doi.org/10.2466/24.22.PMS.119c15z6>.
- Cao, X., Ma, X., Qi, C., 2015. N170 adaptation effect for repeated faces and words. *Neuroscience* 294, 21–28. <https://doi.org/10.1016/j.neuroscience.2015.03.009>.
- Cheung, O.S., Bar, M., 2012. Visual prediction and perceptual expertise. *Int. J. Psychophysiol.* 83 (2), 156–163. <https://doi.org/10.1016/j.ijpsycho.2011.11.002>.
- Cohen, L., Dehaene, S., 2004. Specialization within the ventral stream: the case for the visual word form area. *Neuroimage* 22 (1), 466–476. <https://doi.org/10.1016/j.neuroimage.2003.12.049>.
- Cohen, L., Dehaene, S., Naccache, L., Lehéry, S., Dehaene-Lambertz, G., Hénaff, M.A., Michel, F., 2000. The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* 123 (2), 291–307. <https://doi.org/10.1093/brain/123.2.291>.
- Davies-Thompson, J., Johnston, S., Tashakkor, Y., Pancaroglu, R., Barton, J.J., 2016. The relationship between visual word and face processing lateralization in the fusiform gyrus: A cross-sectional study. *Brain Res.* 1644, 88–97. <https://doi.org/10.1016/j.brainres.2016.05.009>.
- Dehaene, S., Cohen, L., 2007. Cultural recycling of cortical maps. *Neuron* 56 (2), 384–398. <https://doi.org/10.1016/j.neuron.2007.10.004>.
- Dehaene, S., Cohen, L., 2011. The unique role of the visual word form area in reading. *Trends in Cognitive Sciences* 15 (6), 254–262. <https://doi.org/10.1016/j.tics.2011.04.003>.
- Desimone, R., 1996. Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci.* 93 (24), 13494–13499. <https://doi.org/10.1073/pnas.93.24.13494>.
- Friston, K., 2005. A theory of cortical responses. *Philos. Trans. R. Soc. B: Biological Sci.* 360 (1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>.
- Glezer, L.S., Kim, J., Rule, J., Jiang, X., Riesenhuber, M., 2015. Adding words to the brain's visual dictionary: novel word learning selectively sharpens orthographic representations in the VWFA. *J. Neurosci.* 35 (12), 4965–4972. <https://doi.org/10.1523/JNEUROSCI.4031-14.2015>.
- Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cognitive Sci.* 10 (1), 14–23. <https://doi.org/10.1016/j.tics.2005.11.006>.
- Grill-Spector, K., Kushnir, T., Hendler, T., Malach, R., 2000. The dynamics of object-selective activation correlate with recognition performance in humans. *Nat. Neurosci.* 3 (8), 837–843. <https://doi.org/10.1038/77754>.
- Grotheer, M., Kovács, G., 2014. Repetition probability effects depend on prior experiences. *J. Neurosci.* 34 (19), 6640–6646. <https://doi.org/10.1523/JNEUROSCI.5326-13.2014>.
- Grotheer, M., Kovács, G., 2016. Can predictive coding explain repetition suppression? *Cortex* 80, 113–124. <https://doi.org/10.1016/j.cortex.2015.11.027>.
- Hannagan, T., Agrawal, A., Cohen, L., & Dehaene, S. (2021). Simulating the emergence of the Visual Word Form Area: Recycling a convolutional neural network for reading. *bioRxiv*. <https://doi.org/10.1101/2021.02.15.431235>.
- Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., Kliegl, R., 2011. dlexDB—eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau* 62 (1), 10–20. <https://doi.org/10.1026/0033-3042/a000029>.
- Henson, R.N.A., Rugg, M.D., 2003. Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia* 41 (3), 263–270. [https://doi.org/10.1016/S0028-3932\(02\)00159-8](https://doi.org/10.1016/S0028-3932(02)00159-8).
- Honbolygó, F., Kóbor, A., Hermann, P., Kettinger, A.O., Vidnyánszky, Z., Kovács, G., Csépe, V., 2020. Expectations about word stress modulate neural activity in speech-sensitive cortical areas. *Neuropsychologia* 143, 107467. <https://doi.org/10.1016/j.neuropsychologia.2020.107467>.
- Kaliukhovich, D.A., Vogels, R., 2011. Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. *Cereb. Cortex* 21 (7), 1547–1558. <https://doi.org/10.1093/cercor/bhq207>.
- Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17 (11), 4302–4311. <https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997>.
- Kok, P., de Lange, F.P., 2015. Predictive coding in sensory cortex. In: Forstmann, B.U., Wagenmakers, E.-J. (Eds.), *An Introduction to Model-Based Cognitive Neuroscience*. Springer New York, New York, NY, pp. 221–244. https://doi.org/10.1007/978-1-4939-2236-9_11.
- Kovács, G., Iffland, L., Vidnyánszky, Z., Greenlee, M.W., 2012. Stimulus repetition probability effects on repetition suppression are position invariant for faces. *Neuroimage* 60 (4), 2128–2135. <https://doi.org/10.1016/j.neuroimage.2012.02.038>.
- Kovács, G., Kaiser, D., Kaliukhovich, D.A., Vidnyánszky, Z., Vogels, R., 2013. Repetition probability does not affect fMRI repetition suppression for objects. *J. Neurosci.* 33 (23), 9805–9812. <https://doi.org/10.1523/JNEUROSCI.3423-12.2013>.
- Larsson, J., Smith, A.T., 2012. fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cereb. Cortex* 22 (3), 567–576. <https://doi.org/10.1093/cercor/bhr119>.
- Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., Malach, R., 2001. A hierarchical axis of object processing stages in the human visual cortex. *Cereb. Cortex* 11 (4), 287–297. <https://doi.org/10.1093/cercor/11.4.287>.
- Li, C., Ma, X., Zhu, C., Cao, X., 2019. The recovery speed of category sensitive N170 responses to faces and Chinese characters. *Brain Res.* 1723, 146384. <https://doi.org/10.1016/j.brainres.2019.146384>.
- Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P. J., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc. Natl. Acad. Sci.* 92 (18), 8135–8139. <https://doi.org/10.1073/pnas.92.18.8135>.
- Maurer, U., Brem, S., Bucher, K., Brandeis, D., 2005. Emerging neurophysiological specialization for letter strings. *J. Cognit. Neurosci.* 17 (10), 1532–1552. <https://doi.org/10.1162/089892905774597218>.
- Mayrhauser, L., Bergmann, J., Crone, J., Kronbichler, M., 2014. Neural repetition suppression: evidence for perceptual expectation in object-selective regions. *Front. Hum. Neurosci.* 8, 225. <https://doi.org/10.3389/fnhum.2014.00225>.
- Mercure, E., Cohen Kadosh, K., Johnson, M., 2011. The N170 shows differential repetition effects for faces, objects, and orthographic stimuli. *Front. Hum. Neurosci.* 5, 6. <https://doi.org/10.3389/fnhum.2011.00006>.
- Moret-Tatay, C., Baixela Fortea, I., Grau Sevilla, M.D., 2020. Challenges and insights for the visual system: Are face and word recognition two sides of the same coin? *J. Neurolinguistics* 56, 100941. <https://doi.org/10.1016/j.jneuroling.2020.100941>.
- Op de Beeck, H.P., Baker, C.I., 2010. The neural basis of visual object learning. *Trends Cognitive Sci.* 14 (1), 22–30. <https://doi.org/10.1016/j.tics.2009.11.002>.
- Op de Beeck, H.P., Baker, C.I., DiCarlo, J.J., Kanwisher, N.G., 2006. Discrimination training alters object representations in human extrastriate cortex. *J. Neurosci.* 26 (50), 13025–13036. <https://doi.org/10.1523/JNEUROSCI.2481-06.2006>.
- Qin, L., Lyu, B., Shu, S.-u., Yin, Y., Wang, X., Ge, J., Siok, W.-T., Gao, J.-H., 2021. A heteromodal word-meaning binding site in the visual word form area under top-down frontoparietal control. *J. Neurosci.* 41 (17), 3854–3869. <https://doi.org/10.1523/JNEUROSCI.2771-20.2021>.
- Rao, R.P., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2 (1), 79–87. <https://doi.org/10.1038/4580>.
- Schoups, A.A., Vogels, R., Orban, G.A., 1995. Human perceptual learning in identifying the oblique orientation: retinotopy, orientation specificity and monocularly. *J. Physiology* 483 (3), 797–810. <https://doi.org/10.1113/jphysiol.1995.sp020623>.
- Sigman, M., Pan, H., Yang, Y., Stern, E., Silbersweig, D., Gilbert, C.D., 2005. Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron* 46 (5), 823–835. <https://doi.org/10.1016/j.neuron.2005.05.014>.
- Song, Y., Hu, S., Li, X., Li, W., Liu, J., 2010. The role of top-down task context in learning to perceive objects. *J. Neurosci.* 30 (29), 9869–9876. <https://doi.org/10.1523/JNEUROSCI.0140-10.2010>.
- Strother, L., Coros, A.M., Vilis, T., 2016. Visual cortical representation of whole words and hemifield-split word parts. *J. Cognit. Neurosci.* 28 (2), 252–260. https://doi.org/10.1162/jocn_a.00900.
- Strother, L., Zhou, Z., Coros, A.K., Vilis, T., 2017. An fMRI study of visual hemifield integration and cerebral lateralization. *Neuropsychologia* 100, 35–43. <https://doi.org/10.1016/j.neuropsychologia.2017.04.003>.
- Summerfield, C., Trittschuh, E.H., Monti, J.M., Mesulam, M.-M., Egner, T., 2008. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* 11 (9), 1004–1006. <https://doi.org/10.1038/nn.2163>.
- Szwed, M., Qiao, E., Jobert, A., Dehaene, S., Cohen, L., 2014. Effects of literacy in early visual and occipitotemporal areas of Chinese and French readers. *J. Cognit. Neurosci.* 26 (3), 459–475. https://doi.org/10.1162/jocn_a.00499.
- Tan, L.H., Laird, A.R., Li, K., Fox, P.T., 2005. Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: A meta-analysis. *Hum. Brain Mapp.* 25 (1), 83–91. <https://doi.org/10.1002/hbm.20134>.
- Tan, L.H., Liu, H.L., Perfetti, C.A., Spinks, J.A., Fox, P.T., Gao, J.H., 2001. The neural system underlying Chinese logograph reading. *Neuroimage* 13 (5), 836–846. <https://doi.org/10.1006/nimg.2001.0749>.
- Thesen, T., McDonald, C.R., Carlson, C., Doyle, W., Cash, S., Sherfey, J., Felsovalyi, O., Girard, H., Barr, W., Devinsky, O., Kuzniecky, R., Halgren, E., 2012. Sequential then

- interactive processing of letters and words in the left fusiform gyrus. *Nat. Commun.* 3 (1) <https://doi.org/10.1038/ncomms2220>.
- Wang, H., Chang, R.B., Li, Y.S., 1986. *Modern Chinese frequency dictionary*. Beijing Language Institute Press, Beijing. In Chinese.
- Wang, Y.C., Sohoglu, E., Gilbert, R.A., Henson, R.N., Davis, M.H., 2021. Predictive Neural Computations Support Spoken Word Recognition: Evidence from MEG and Competitor Priming. *J. Neurosci.* 41 (32), 6919–6932. <https://doi.org/10.1523/JNEUROSCI.1685-20.2021>.
- Wu, C.Y., Ho, M.H.R., Chen, S.H.A., 2012. A meta-analysis of fMRI studies on Chinese orthographic, phonological, and semantic processing. *Neuroimage* 63 (1), 381–391. <https://doi.org/10.1016/j.neuroimage.2012.06.047>.
- Young, A.W., Burton, A.M., 2017. Recognizing faces. *Current Directions in Psychological Science* 26 (3), 212–217. <https://doi.org/10.1177/0963721416688114>.
- Yu, C., Klein, S. A., & Levi, D. M. (2004). Perceptual learning in contrast discrimination and the (minimal) role of context. *Journal of Vision*, 4(3), 4-4. <https://doi.org/10.1167/4.3.4>.
- Zhou, Z., Vilis, T., Strother, L., 2019. Functionally separable font-invariant and font-sensitive neural populations in occipitotemporal cortex. *J. Cognit. Neurosci.* 31 (7), 1018–1029. https://doi.org/10.1162/jocn_a_01408.

Supplementary Table 1. Individual-participant VWFA and LO coordinates in German and Chinese participants

Participants	Number	VWFA	Left LO	Right LO
German	S01	-42, -64, -10	-44, -76, -10	42, -62, -16
	S02	-40, -58, -4	-38, -78, -4	40, -70, -10
	S03	-48, -60, -18	-40, -82, 4	46, -80, 2
	S04	-38, -48, -14	-44, -76, -2	42, -78, -2
	S05	-46, -52, -12	-36, -86, 10	38, -72, -8
	S06	-38, -68, -2	-50, -78, -4	44, -84, 4
	S07	-56, -66, -10	-34, -84, -18	38, -84, -2
	S08	-54, -46, -10	-42, -70, -8	54, -72, 0
	S09	-44, -86, -6	-46, -84, -6	40, -82, -2
	S10	-42, -70, -10	-48, -78, -2	38, -88, -4
	S11	-44, -76, -14	-54, -78, 4	52, -76, 2
	S12	-56, -66, -10	-36, -86, -14	36, -70, -12
	S13	-42, -76, -10	-38, -66, -12	46, -78, 4
	S14	-42, -58, -20	-44, -90, -2	40, -86, -2
	S15	-44, -68, -12	-36, -66, -10	40, -70, -8
	S16	-38, -82, -4	-40, -80, -6	38, -78, -4
	S17	-42, -48, -18	-36, -82, -14	30, -86, -2
	S18	-50, -76, -10	-42, -78, -8	44, -80, -12
	S19	-44, -76, -14	-38, -84, -6	48, -82, 8
	S20	-40, -56, -12	-48, -76, 2	46, -82, 4
Chinese	S01	-42, -54, -18	-40, -82, -12	42, -76, -12
	S02	-44, -50, -20	-40, -80, -2	42, -78, -6
	S03	-42, -52, -12	-44, -76, 2	52, -82, -2
	S04	-46, -56, -22	-52, -72, 0	44, -74, -14
	S05	-46, -68, -12	-40, -84, 0	44, -84, -2
	S06	-42, -64, -14	-30, -88, 4	46, -76, -16
	S07	-42, -74, -16	-42, -84, -12	42, -82, -6
	S08	-46, -68, -8	-48, -78, -12	44, -78, -4
	S09	-38, -83, -13	-42, -80, -10	40, -84, 8
	S10	-46, -58, -10	-40, -80, -2	42, -78, -6
	S11	-42, -78, -6	-44, -76, -2	34, -76, 6
	S12	-48, -78, -10	-50, -76, -10	42, -66, -2
	S13	-46, -54, -14	-40, -74, 6	44, -76, 6
	S14	-38, -78, -16	-42, -72, -4	40, -74, -4
	S15	-46, -76, -12	-38, -74, -4	40, -86, -4
	S16	-36, -82, -16	-36, -80, -16	30, -92, 0
	S17	-54, -72, -8	-44, -80, -4	48, -76, -10
	S18	-40, -50, -18	-40, -80, -6	42, -82, 2
	S19	-44, -54, -18	-48, -76, 2	32, -76, -16
	S20	-48, -66, -14	-46, -82, -12	46, -84, -4

Supplementary Table 2. Summary of significant activations identified from the group level whole-brain analysis

Contrast	Region	Hemisphere	German Participants		Chinese Participants		Threshold
			MNI	Cluster size	MNI	Cluster size	
			Coordinates	(Voxels)	Coordinates	(Voxels)	
Chinese Characters > German words	Middle occipital gyrus	R	28, -88, 12	2519	26, -90, 8	1152	P < 0.05, FWE
	Middle occipital gyrus	L	-16, -98, 6	1533	-18, -94, 6	555	P < 0.05, FWE
	Inferior occipital gyrus	L	-	-	-44, -78, -8	273	P < 0.05, FWE
At > Rt	Middle temporal l gyrus	L	-	-	-58, -48, 8	10	P < 0.001 uncorrected

3.2 study II: Short-term training attenuates the repetition probability effect for non-face objects

The version of the manuscript has been submitted to *Neuropsychologia*:

Li, C., & Kovács, G. Short-term experience reduces the repetition probability effect for non-face objects. (submitted)

Main research question:

Does short-term experience affect the repetition probability effect for
non-face objects?

Short-term training attenuates the repetition probability effect for non-face objects

Chenglin Li, Gyula Kovács

Department of Biological Psychology and Cognitive Neuroscience,
Friedrich-Schiller-Universität Jena, Germany

Corresponding author: Gyula Kovács

Email: gyula.kovacs@uni-jena.de

Abstract

The magnitude of repetition suppression (RS), measured by fMRI, is modulated by the probability of stimulus repetitions, which is called repetition probability [P(rep)] effect. The P(rep) effect refers a stronger RS effect in blocks, where the probability of repetition trials was higher when compared to blocks with less frequent repetition. Previous studies have suggested that this P(rep) effect depends on the extensive long-term experience of participants with the stimuli. Conversely, the modulation of short-term learning experience on the P(rep) effect remains largely unknown. To address this issue, we used fMRI and measured the P(rep) effect for unfamiliar faces and non-face objects before and after a 10-day perceptual learning (PL) session and the PL sessions were performed only for non-face objects (cars). The results showed a significant P(rep) effect for faces within the Fusiform Face Area (FFA) and for cars within the Lateral Occipital Complex (LO) in the pre-training fMRI measurement session. Following a 10-day PL session, participants exhibited huge behavioral improvements for the trained stimuli (cars). Surprisingly, the P(rep) effect was absent both for faces and cars in the post-training fMRI measurement session. These results support the idea that the predictive processes, measured by P(rep) modulation of RS, can be modulated by the short-term perceptual learning experience.

Key words: Expectation suppression; Repetition suppression; Perceptual learning; Face perception; Object perception

1. Introduction

How the brain efficiently encodes sensory information is always a hot topic in the field of cognitive psychology. Recently, the theory of predictive coding posits that the brain constantly attempts to match incoming sensory inputs to top-down predictions. Each region in the cortical sensory hierarchy represents both these predictions and the mismatch between predictions and inputs (prediction error) (de Lange et al., 2018; Friston, 2005; Kok and de Lange, 2015; Rao & Ballard, 1999).

This theory is one of the explanatory accounts of a well-established phenomenon, which is the repetition suppression (RS) effect (Desimone, 1996). RS refers to the reduction of neural responses for repeated stimuli compared to their first presentation (Henson and Rugg, 2003). This reduction is seen as an instance of predictive processing because stimulus repetition may enhance expectation or reduce prediction error (for a review see Grotheer and Kovács, 2016). Summerfield and colleagues (2008) first developed a paradigm to investigate the relationship between RS and predictive coding by manipulating the probability of stimulus repetition and alternation, and measured RS using functional magnetic resonance imaging (fMRI). They presented participants with pairs of face images in which the first stimulus could be either identical to (repetition trials) or different from the second one (alternation trials) in two different contexts, which the probability of repetition and alternation trials was differently distributed in two kinds of blocks. The results showed a stronger RS effect in blocks, where the probability of repetition trials was higher when compared to blocks with less frequent repetition. This finding can be explained by the fact that repetition trials are more expected in repetition contexts than in alternation contexts. Later, this modulation of repetition probability $[P(\text{rep})]$ on RS has been replicated in further studies using faces (Kovács et al., 2012, 2013; Larsson and Smith, 2012).

1 However, this P(rep) modulation on RS for non-face stimuli led to less
2 consistent results. For instance, in human neuroimaging studies, it was
3 obtained with Roman letters, line-drawing objects and words (Grotheer and
4 Kovács, 2014; Li and Kovács, 2022; Mayrhauser et al., 2014), but not with
5 every-day objects, chairs, novel false fonts (Grotheer and Kovács, 2014;
6 Kovács et al., 2013). In macaque single-cell recordings studies, this effect was
7 not replicated both with fractal patterns, everyday objects (Kaliukhovich &
8 Vogels, 2011) and with 3D artificial face images (Vinken et al., 2018). One
9 possible explanation for these inconsistent results is that participants have
10 different prior experiences with these different stimulus categories. Previous
11 studies have suggested that prior experience with stimuli is a prerequisite for
12 observing the P(rep) effect on RS (Grotheer and Kovács, 2014; Li and Kovács,
13 2022). In particular, the P(rep) effect could be robustly replicated for faces,
14 words, and Roman letters, with which participants had extensive experiences
15 with the stimuli, but not for other stimulus categories, with which participants
16 had no experience (i.e., novel false fonts, Chinese characters for German
17 participants, every-day objects for monkey), or insufficient experience (i.e.,
18 chairs). Therefore, it is necessary to test this hypothesis using a more familiar
19 general object category for human participants.

20 The above-mentioned studies suggest that the P(rep) effect depends on
21 the extensive long-term experience of participants with the stimuli (Grotheer
22 and Kovács, 2014; Li and Kovács, 2022). However, it remains largely unknown
23 the extent to which the P(rep) effect could be modulated by the short-term
24 learning experience with the stimuli. As far as we know, with intensive training,
25 perceptual abilities of participants such as discriminating visual features
26 (contrast, orientation, shape and so on) can be dramatically improved and
27 perceptual training allows to systematically provide participants with
28 experience about a particular stimulus category (for review see Bi and Fang,
29 2013). This learning process is referred to as perceptual learning (PL), which is

an approach to gain experience with certain stimuli (Fiorentini and Berardi, 1980; Goldstone, 1998; Op de Beeck and Baker, 2010). The behavioral improvements on training tasks have been shown to be closely related to the changes of brain activity (Grill-Spector et al., 2000; Op de Beeck et al., 2006; for a review see Bi and Fang, 2013). For instance, previous fMRI studies have found that PL can lead to decreased BOLD signal responses for trained stimuli (e.g., Henson et al., 2000; Reber et al., 1998; Sigman et al., 2005; but see Grill-Spector et al., 2000). Importantly, the reduction of BOLD signal for trained stimuli, as compared to untrained (novel) stimuli, can be regarded as a RS effect (Grill-Spector et al., 1999; Jiang et al., 2007). In other words, increased experience through PL training could lead to the appearance of the RS effect. Thus, the P(rep) modulation on RS could be modulated by the short-term experience through PL as well.

To address this question, we used fMRI and measured the P(rep) effect for two familiar stimulus categories (faces and cars) both in the pre- and post-training fMRI measurement sessions. The training session consisted of ten-day PL sessions, which participants gained perceptual experience for non-face objects (cars). The fMRI session adopted the same fMRI paradigm as previous studies (Summerfield et al., 2008). Briefly, we observed P(rep) effects for faces in the Fusiform Face Area (FFA; Kanwisher et al., 1997) and for cars in the Lateral Occipital Cortex (LO; Malach et al., 1995) in the pre-training session. Interestingly, the results in the post-training session showed that P(rep) effects were absent both for the trained (cars) and untrained stimuli (faces). We discuss these findings in the framework of the predictive coding theory.

2. Methods

2.1 Participants

A total of 19 healthy participants took part in this study in exchange for their own 3D-printed brain-model, partial course credits, or monetary

compensations. Participants were informed about the experimental procedures and gave written informed consent before joining the study. Three participants were excluded from the final analysis due to the interruption of training sessions caused by the COVID-19 pandemic. The remaining 16 participants (9 females; mean age, 23.2 years; SD, 3.4 years; one left-handed) had normal or corrected-to-normal vision. The experiment was approved by the ethics committee of the Friedrich-Schiller-Universität Jena and conducted in accordance with the guidelines of the Declaration of Helsinki.

2.2 Stimuli

Three-hundred and ninety-six grey-scale, digital images of unfamiliar Caucasian faces (198 female faces) were selected from a pool, used in previous studies of our laboratory (Kovács et al., 2012, 2013; Rostalski et al., 2019). All images were cropped into a circular grey-scale shape to remove external facial features (e.g., hair, ears, and jawline) and displayed a neutral expression. Images were resized to 480 × 480 pixels with a grey background.

One thousand one hundred and forty colored images of cars (Chrysler, Ford, Peugeot and Volkswagen, 285 different exemplar images for each brand) were selected from the public domain of the worldwide web. The background of these images was removed by the remove.bg software (remove.bg for Windows 1.4.2). Brand logos and names were removed from the front faces of the cars and their locations were filled out by using the content-aware fill function of GIMP 2.8.6. All images were resized to 480 × 480 pixels.

2.3 Experiment design and procedure

The experimental design is presented in Figure 1A. To investigate the effect of short-term perceptual training on RS and P(rep), perceptual learning (PL) sessions were performed for the car stimuli between a pre- and post-training fMRI recording session. Our training task included 10 sessions (2 weeks, 5 days per week with 40 minutes per day). During the training sessions

1 participants sat in a comfortable chair in a dimly lit room, in front of a 27-inch
2 LED display (2560 x 1440 pixels resolution; 60 Hz refresh rate, viewing
3 distance ~60 cm). Stimulus presentation and behavioral response collection
4 were performed by Psychtoolbox (Version 3.0.15) in Matlab R2017a (The
5 MathWorks).

6 Every training day, 120 car images (30 images per brand) were presented.
7 Half of these images (60; 15 images per brand) were always previously
8 unseen exemplars. Participants were asked to complete three runs of a
9 four-way category discrimination task, which they had to classify a car by
10 brand and press the corresponding buttons. Each run (240 trials) was
11 composed of 6 blocks which were separated by a 5-s-long break. Each trial
12 started with a black fixation cross (exposition time ranging from 500 ms to 700
13 ms), followed by the car images for 500 ms (Figure 1C). Next, a black fixation
14 point (maximum 2,000 ms) was presented until the participant issued a
15 response. The participant's task was to indicate the car brand as quickly and
16 accurately as possible. After pressing the corresponding buttons, participants
17 would get visual feedback ('Correct' or 'Incorrect') on the screen for 100 ms.
18 The mean performance of each car brand was presented at the end of each
19 run. To familiarize the participants with the training task, the logo of the car
20 images was not removed from the images during the very first run of the first
21 day, and the results of this run were excluded in the final data analysis.

22 The experimental paradigm during neuroimaging data acquisition (Figure
23 1B) was identical to what has previously been used to investigate P(rep)
24 modulations (Kovács et al., 2012, 2013; Larsson and Smith, 2012;
25 Summerfield et al., 2008). Stimuli were presented centrally on a gray
26 background, back-projected via an LCD video projector onto a translucent
27 circular screen. Psychtoolbox (Version 3.0.15) was used for stimulus
28 presentation and behavioral response collection, controlled by Matlab R2013a
29 (The MathWorks).

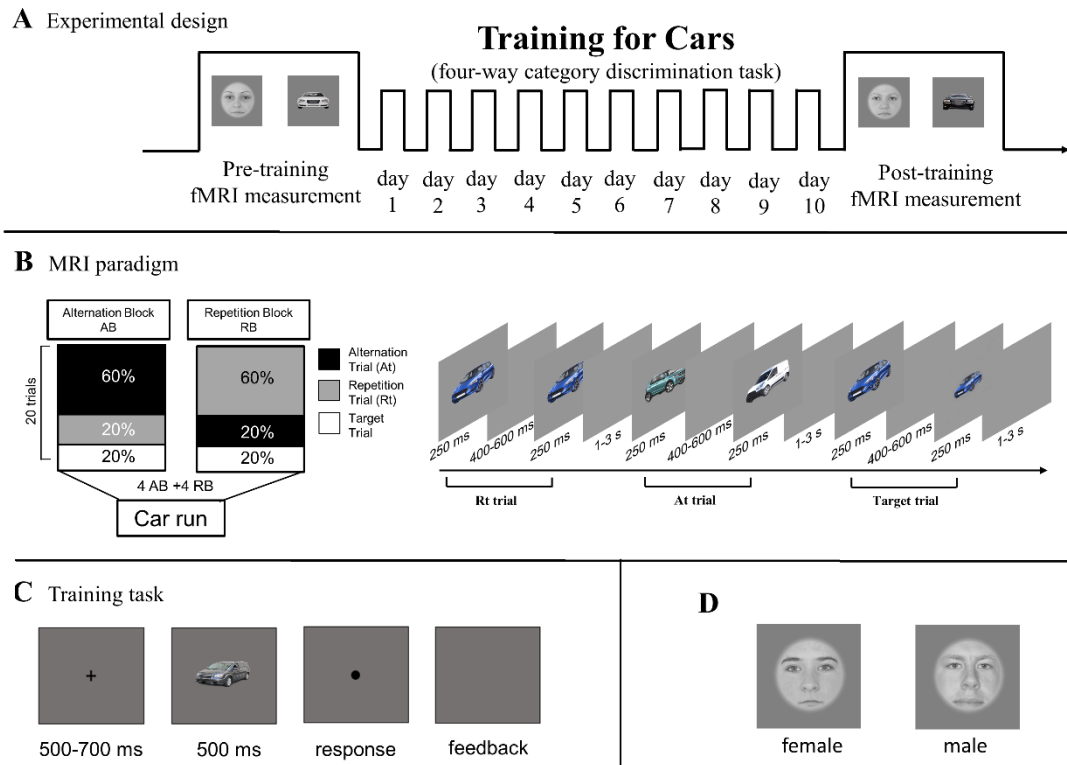


Figure 1. Schematic illustration of the experimental design and task procedures. **A**, A ten-day training session with a four-way car-brand discrimination task separated the pre- and post-training fMRI measurements. **B**, Each neuroimaging run was composed of four alternation (ABs) and four repetition blocks (RBs), which were presented randomly (taken car run as an example). A repetition (Rt) trial, an alternation (At) trial and target trial (detection of smaller stimuli) are illustrated. **C**, The trial structure during the training period. **D**, the exemplar images of face stimuli.

During the fMRI session, the participants completed four experimental runs, two with face and two with car stimuli, in a counterbalanced order. In each trial (Figure 1B, taken car run as an example), a stimulus-pair was presented (duration: 250 ms each), separated by an inter-stimulus-interval (ISI) of 400-600 ms (randomized across trials), and randomly followed by an Inter-trial-interval (ITI) of 1,000 ms or 3,000 ms. The first stimulus (S1) was either identical to (Repetition trials, Rt) or different from the second stimulus (S2; Alternation trials, At). In addition, two different types of blocks were given to participants, each repeated four times within a single run (160 trials). In the Repetition Blocks (RB), 75% of the trials were Repetition trials while 25% were

1 Alternation trials. In contrast, the Alternation Blocks (AB) were composed of 75%
2 Alternation trials and 25% Repetition trials (Summerfield et al, 2008).
3 Therefore, each run included 160 trials of the four different conditions (ABAt,
4 ABRT, RBAAt, and RBRT) in randomized order. To avoid local feature
5 adaptations, the size of either S1 or S2 (chosen randomly) was reduced by 18%
6 during each trial. To maintain participants' attention, the participants had to
7 press a button in target trials (20% per run), when the size difference between
8 S1 and S2 was 60%. The fMRI procedures were identical in the pre- and
9 post-training sessions, but had different stimuli. The design and parameters of
10 the pre- and post-training face sessions were the same as the car runs, with
11 the replacement of car stimuli by the face stimuli (Figure 1D).

12 **2.4 Imaging parameters and data analysis**

13 Scanning was performed by a 3-Tesla MRI scanner (Siemens
14 MAGNETOM Prisma, Erlangen, Germany). During the functional runs,
15 multi-band EPI sequences (MB acceleration factors = 4) were used with a
16 64-channel head coil. T2-weighted images were collected with the following
17 parameters: FOV = $224 \times 224 \text{ mm}^2$; 100 slices; TR = 2000 ms; TE = 33.6 ms;
18 flip angle = 90 deg; 1.4 mm isotropic voxel size. High-resolution T1-weighted
19 images (192 slices; TR = 2300 ms; TE = 3.03 ms; flip angle = 90 deg; 1 mm
20 isotropic voxel size) were acquired to obtain 3D structural scans with the same
21 64-channel head coil. Data were preprocessed using SPM12 (Wellcome
22 Department of Imaging Neuroscience, London, UK). Briefly, the functional
23 images were slice-timed, realigned, and co-registered to structural scans. The
24 functional images were normalized to the MNI-152 space, resampled to $1 \times 1 \times$
25 1 mm resolution and spatially smoothed using a 3-mm Gaussian kernel.

26 Regions of interests (ROIs) were defined by an additional functional
27 localizer run (Rostalski et al, 2020). Faces, cars, objects, and
28 Fourier-randomized noise patterns were presented (230 ms exposure time; 20
29 ms inter-stimulus-interval) in blocks of 10 s, interrupted by breaks of 10 s and

repeated four times. Each block included 40 images, which the size of 600x 600 pixels with a grey background. To ensure the successful definition of all ROIs, the localizer run was performed both in pre- and post- fMRI sessions for every participant, but the area locations were similar across sessions.

As previous studies, applying probabilistic repetition (P(rep)) paradigms found the largest P(rep) effect for faces in the FFA and for objects in the LO (Grotheer et al., 2014; Kovács et al., 2012, 2013; Mayrhauser et al., 2014), we concentrated on these two areas here. The FFA was localized by contrasting faces with object and noise images individually and the location was established as the local maximum from the t-maps with a threshold of $p_{FWE} < 0.05$ on the single-subject level. The average MNI coordinates (\pm SE) for the FFA were as follows: i) pre-training: left hemisphere, -40.8 (1.39), -56.2 (2.10), -18.4 (0.86), right hemisphere, 42.3 (1.27), -58.8 (2.36), -19.1 (0.82); ii) post-training: left hemisphere, -40.3 (1.37), -57.2 (2.39), -17.9 (0.93), right hemisphere, 41.3 (1.29), -59.4 (2.15), -18.4 (0.85). The location of LO was determined individually by contrasting object with noise blocks. The average MNI coordinates (\pm SE) for the LO were as follows: i) pre-training: left hemisphere, -45.1 (0.70), -79.9 (1.22), -4 (1.02), right hemisphere, 45.3 (1.11), -79.4 (1.38), -5.4 (1.12); ii) post-training: left hemisphere, -44.8 (0.96), -78.9 (1.26), -5.4 (1.55), right hemisphere, 43.9 (0.90), -81.1 (1.33), -5.8 (0.94). Canonical hemodynamic response functions were extracted using MarsBaR 0.44 toolbox for SPM 12 (Brett et al., 2002).

The peak BOLD values were extracted from the event-related runs and analyzed by repeated-measures ANOVAs with session (pre- vs. post-training), block type (AB vs. RB) and trial type (At vs. Rt) as within-subject factors. Additionally, we analyzed the mean performance for the detection of target trials using the same 2 x 2 x 2 repeated-measures ANOVA. To examine the effect of learning during training, a one-way repeated-measures ANOVA was performed for accuracies and reaction times with session number as a factor.

1 All multiple comparisons of *post-hoc* tests were corrected with the Fisher's
2 method.

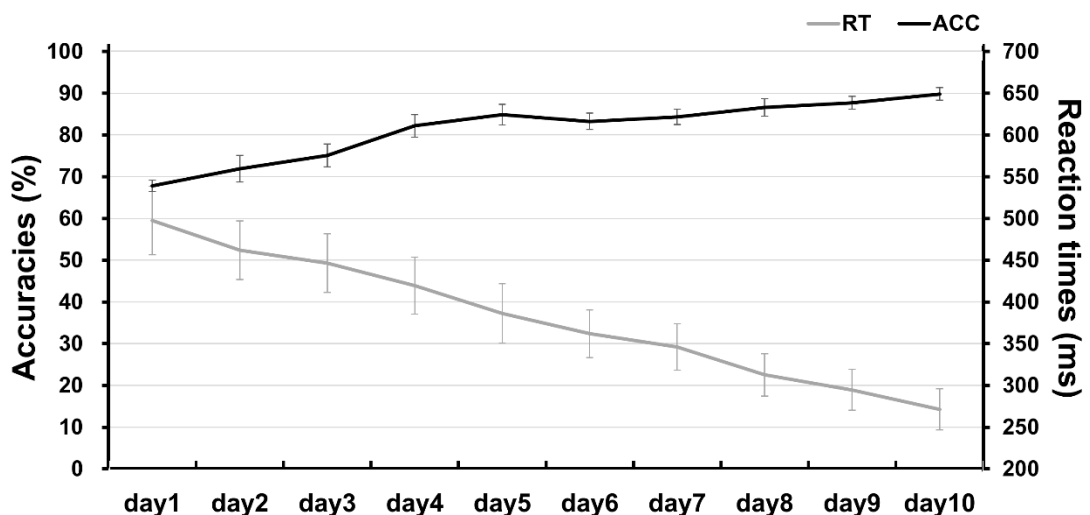
3 3. Results

4 3.1 Behavioral results

5 3.1.1 Training session

6 The training results are presented in Figure 2. The Greenhouse-Geisser
7 corrected results are reported because the Mauchly's test of sphericity
8 revealed unequal variances of differences both in accuracy ($\chi^2(44) = 142.545$,
9 $p < .001$) and reaction time ($\chi^2(44) = 119.919$, $p < .001$). A significant main
10 effect of training session was found both for accuracy ($F(2.02, 30.30) = 48.684$,
11 $p < 0.001$, $\eta_p^2 = 0.764$) and reaction time ($F(3.08, 46.17) = 33.571$, $p < 0.001$,
12 $\eta_p^2 = 0.691$). A *post-hoc* test revealed that there were significant improvements
13 both in accuracy (from day 2, $p = 0.006$) and reaction time (from day 3, $p =$
14 0.007).

15 Additionally, one-sample *t* tests were performed to compare the accuracy
16 of each day against chance level (0.25) (every day: $p < 0.001$). These results
17 suggest that participants got more accurate and faster in the brand
18 discrimination task over time.



19

20 Figure 2. The accuracies (black line) and reaction times (grey line) of the four-way
21 category discrimination task during the ten-day training session. Error bars represent

standard errors of means.

3.1.2 fMRI session

During the face runs participants detected the target stimuli on average with 85.35 and 86.62% (\pm SE: 4.53 and 4.54%) accuracy in the pre- and post-training sessions, respectively. The mean performance was 84.67 and 84.08% (\pm SE: 4.27 and 4.24%) in the pre- and post-training sessions in car runs, respectively. The four-way repeated measure ANOVA on behavioral data showed neither any significant main effect, nor any interaction effect (all $p \geq 0.103$). In addition, an informal survey after the post-training fMRI measurement revealed that none of the participants was aware of the manipulation of P(rep) across the different blocks.

3.2 Neuroimaging results

3.2.1 FFA responses to faces

We found a significant main effect of trial type ($F(1, 15) = 25.090$, $p < 0.001$, $\eta_p^2 = 0.626$), with lower BOLD signals in Rt trials (0.617 ± 0.068) than in Alt trials (0.805 ± 0.076) in the right FFA (Figure 3A and B). The three-way interaction between session, block type and trial type was significant ($F(1, 15) = 5.789$, $p = 0.029$, $\eta_p^2 = 0.278$). A post-hoc test revealed that the two-way interaction between block type and trial type was marginally significant ($F(1, 15) = 4.304$, $p = 0.055$, $\eta_p^2 = 0.223$) in the pre-training fMRI measurement session, where a stronger RS effect in the RB block (Fisher LSD *post hoc* test: $p = 0.00003$) than in the AB block was shown (Fisher LSD *post hoc* test: $p = 0.0298$). In the post-training fMRI session only a significant RS effect was observed in the AB block (Fisher LSD *post hoc* test: $p = 0.0046$) whereas, this interaction between block type and trial type was no longer significant in the post-training fMRI measurement session ($F(1, 15) = 0.527$, $p = 0.479$, $\eta_p^2 = 0.034$). No other main or interaction effects were significant. Altogether, these results suggested that P(rep) affects the magnitude of RS for faces in the right FFA only in the pre-training fMRI measurement session, which has been

similarly reported in previous studies (Kovács et al., 2012; Lasson & Smith 2012; Summerfield et al., 2008).

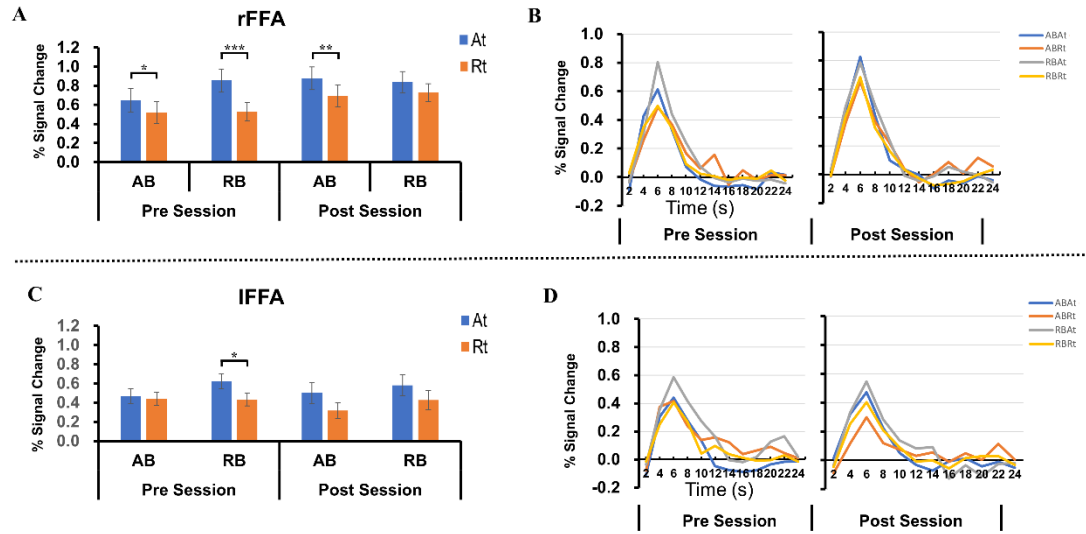


Figure 3. Average peak activation profiles (panels A and C) of the right and left FFA for At and Rt trials, and the time course (panels B and D) of fMRI activity in the right and left FFA for face runs. The HRF curves were derived from an FIR model with 2 s time bins. Error bars represent standard errors of means. $*p < .05$, $**p < .01$, $***p < .001$ (Fisher's *post hoc* comparisons).

We found a significant main effect of trial type ($F(1, 15) = 11.080$, $p = 0.005$, $\eta_p^2 = 0.425$), with lower BOLD signals in Rt trials (0.404 ± 0.047) compared to those in At trials (0.543 ± 0.062) in the left FFA (Figure 3C and D). No other main or interaction effects were significant.

3.2.2 LO responses to cars

We found a significant main effect of trial type ($F(1, 15) = 26.267$, $p < 0.001$, $\eta_p^2 = 0.637$), with lower BOLD signals in the Rt trials (0.724 ± 0.099) than in the At trials (0.892 ± 0.117) in the right LO (Figure 4A and B). Importantly, there was a significant main effect of session ($F(1, 15) = 5.171$, $p = 0.038$, $\eta_p^2 = 0.256$), with the reduction of BOLD signals in the post-training fMRI measurement session (0.647 ± 0.088) compared to those in the pre-training fMRI measurement session (0.969 ± 0.158). A three-way interaction between session, block type and trial type was significant ($F(1, 15) = 7.737$, $p = 0.014$, $\eta_p^2 = 0.340$). Post-hoc tests revealed that the two-way

interaction between block type and trial type was significant ($F(1, 15) = 13.788$, $p = 0.002$, $\eta_p^2 = 0.479$) in the pre-training fMRI measurement session, and that the RS effect was stronger in the RB block (Fisher LSD *post hoc* test: $p = 0.000004$) than that of in the AB block (Fisher LSD *post hoc* test: $p = 0.031$); but this interaction could not be found in the post-training fMRI measurement session ($F(1, 15) = 0.130$, $p = 0.723$, $\eta_p^2 = 0.009$), where only a significant RS effect was observed in the AB block (Fisher LSD *post hoc* test: $p = 0.0156$). No other main or interaction effects were significant. Altogether, these results suggest that P(rep) affects the magnitude of RS for cars in the right LO only in the pre-training fMRI measurement session, which was similar to the results of the face runs in the right FFA. However, the absence of the P(rep) effect for cars in the post-training fMRI measurement session was mainly caused by the reduction of the BOLD signals from short-term perceptual learning.

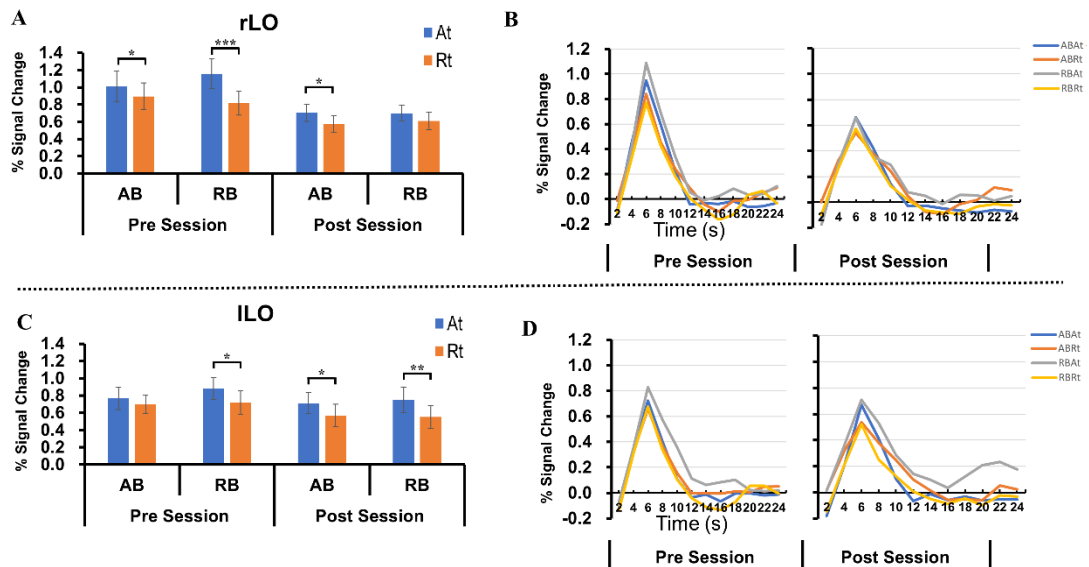


Figure 4. Average peak activation profiles (A, C) of the right and left LO for At and Rt trials, and the time course (B, D) of fMRI activity in the right and left LO for car runs. The HRF curves were derived from an FIR model with 2 s time bins. Error bars represent standard errors of means. $*p < .05$, $**p < .01$, $***p < .001$ (Fisher's *post hoc* comparisons).

There was a significant main effect of trial type ($F(1, 15) = 14.558$, $p = 0.002$, $\eta_p^2 = 0.493$) in the left LO (Figure 4C and D), with lower BOLD signals in

Rt trials (0.636 ± 0.102) than in At trials (0.778 ± 0.115). No other main or interaction effects were significant.

4. Discussion

We aimed to investigate the influence of short-term experience on neural effects reflecting expectation, as measured by P(rep) on the RS effect in category-responsive regions. The main results are summarized as follows: First, significant modulations of P(rep) on RS were found for faces in the right FFA and for cars in the right LO before the training sessions. Second, participants showed huge behavioral improvements during the training task over time. Third, the P(rep) effects were attenuated in opposite directions for faces and cars after the training sessions.

The modulation of P(rep) on RS has been found for various stimulus categories, such as faces, words, and Roman letters (e.g., Grotheer and Kovács, 2014; Larsson and Smith, 2012; Li and Kovács, 2022; Summerfield et al., 2008). Consistent with the results of previous studies, we replicated this effect for faces. However, there has been little agreement on P(rep) effects for objects using an identical paradigm with line-drawing of objects as stimuli (Mayrhauser et al., 2014), but not with chairs or everyday objects in human neuroimaging study (Kovács et al., 2013) and for fractal patterns and everyday objects in macaque single-cell recording experiments (Kaliukhovich and Vogels, 2011). These inconsistent results in object processing could be due to lots of factors related to between-study heterogeneity, such as different experimental species, stimulus categories, experimental tasks (a passive task was performed in macaques, unlike the active tasks in human neuroimaging studies) and so on. However, the results of Kovács et al. (2013) have excluded the factors of species and tasks as they adopt the identical paradigm and task with Summerfield et al. (2008) in human participants. Then, another possible factor causing this inconsistency could be the difference in the experience of stimulus categories: participants had different levels of prior experience with

1 experimental stimulus categories. Previous studies have demonstrated that
2 the P(rep) effect is experience-dependent (Grotheer and Kovács, 2014; Li and
3 Kovács, 2022). Thus, it is possible that the significant P(rep) effect for objects
4 in general was not present because participants had no experience with the
5 related stimulus categories (i.e., novel false fonts, Chinese characters for
6 German participants, every-day objects for monkey), or because existing
7 experience is not enough (i.e., chairs).

8 In the present study, we presented participants with another familiar
9 stimulus category (cars) to test this hypothesis (note that, more than other
10 objects, cars are a category with which some participants show extensive
11 expertise (Gauthier et al., 2000; McGigin et al., 2016)). Although we did not
12 test the participants' experience with cars before the experiment (i.e.,
13 Vanderbilt Expertise Test; McGugin et al., 2012), the accuracy levels were
14 significantly higher than the chance level in the category discrimination task on
15 day 1, which indicates that participants had extensive experience with cars.
16 This finding supports further the idea that the observation of the P(rep) effect
17 depends on extensive prior experience with this stimulus category.

18 Another potential reason why we observed significant P(rep) effects is
19 suggested by the results of Mayrhauser et al. (2014). They observed a
20 significant P(rep) effect for line-drawing objects, which require the extraction of
21 semantic information as they are abstract and stylized stimuli. Therefore, one
22 hypothesis is that the occurrence of the P(rep) effect requires lots of semantic
23 information from the higher-order cortex as well. Indeed, as compared to other
24 objects, car exemplars can individually be identified and abundant semantic
25 information are related to an individual identity information, like other expertise
26 stimuli (e.g., faces and words). Each car has its own color, shape, model,
27 brand, and other specific information, which helps us to identify it and
28 discriminate it from similar ones. More importantly, a car can be systematically
29 classified based on this information, especially by brand and model. The

1 hypothesis that retrieval of semantic information favors the emergence of
2 P(rep) effects should be tested in future studies.

3 While cumulating evidence suggests the essential role of PL for
4 perceptual processing (Brants et al., 2016; for a review see Bi and Fang, 2013),
5 its effect in P(rep) modulations on RS remains sparsely investigated. In the
6 present study, we investigated the extent to which the P (rep) effect is
7 modulated by the short-term experience from PL. Consistent with previous
8 studies on various perceptual tasks, such as orientation-, contrast-, motion
9 direction-, texture-discriminations, and object or face recognition (Furmanski
10 and Engel, 2000; Schoups et al., 1995; Yu et al., 2004; for reviews see Bi and
11 Fang, 2013), the present results confirm strong training effects of the car-brand
12 discrimination task, suggesting that PL is an effective method to obtain
13 perceptual experience, ranging from low-level feature detection to high-level
14 object recognition.

15 Last and importantly, the P(rep) effects were attenuated in opposite
16 directions for faces and cars after the training sessions. The reduction of
17 BOLD signal responses has also been observed in previous studies on other
18 object training studies, in which the BOLD signal responses for trained stimuli
19 were decreased (Henson et al., 2000; Reber et al., 1998; Sigman et al., 2005;
20 but see Grill-Spector et al., 2000). However, in these studies the attenuation
21 caused by the perceptual trainings was specific to the trained stimuli (Baeck
22 and Op de Beeck, 2010), but did not affect the untrained stimuli, which were
23 the same in both pre- and post-training sessions. Therefore, the absence of
24 P(rep) effect for trained stimuli (cars) may be due to the overall reduction of
25 neural response for the trained stimulus category. In other words, this overall
26 reduction of the BOLD signal responses for trained stimuli results in a P(rep)
27 effect that could not be detected in the post-training session, because the
28 P(rep) effect reflects the small difference between repetition and alternation
29 trials in different blocks.

1 Additionally, this rather intriguing result could be explained by the different
2 contributions of surprise and fulfilled expectations on expectation suppression
3 (ES) (Amado et al., 2016; Egner et al., 2010; Grotheer and Kovács, 2015; for a
4 review see Feuerriegel et al., 2021). Amado et al. (2016) modified the
5 paradigm of Summerfield et al (2008) to present pairs of faces (adult female,
6 adult male, or baby) that were repeated or alternated. Orthogonally to this, the
7 P(rep) was manipulated by the gender of the first face within each pair, thus,
8 the face repetition or alternation trials could either be expected (75%
9 probability), neutral (50% probability), or surprising (25% probability). The
10 BOLD signals results in FFA showed a larger response for surprising
11 compared to neutral face pairs in the alternation trials, suggesting the critical
12 role of surprise enhancement in P(rep) effects.

13 In the current study, we hypothesize that our results can be explained as
14 an instance of expectation suppression for trained stimuli (cars) and surprise
15 enhancement for untrained stimuli (faces). In the pre-training sessions, the
16 face and car runs are equally predicted for the participants as there was no
17 cue that induced more expectations for either face or car runs. However, the
18 expectations for the car runs would increase after the training sessions as the
19 training task was only related to cars. In other words, participants would more
20 expect the task in the post-training session to be more related to trained stimuli
21 (cars), while the face runs are surprising events in the post-training session.
22 Compared with the results of pre-training sessions, the overall BOLD signals of
23 the car runs were significantly decreased (expectation suppression) in the
24 post-training sessions, however, the overall BOLD signals of the face runs
25 were increased (surprise enhancement) in three conditions. As discussed
26 above, the P(rep) effect reflects the small difference between repetition and
27 alternation trials in different blocks, but these overall reductions or
28 enhancements of BOLD signals might result in the fact that the P(rep) effect
29 could not be detected. The attenuation of the P(rep) effects for face and

1 non-face stimuli was caused by this change of BOLD signals in the opposite
2 direction. This therefore suggests that the effect of the short-term training on
3 the P(rep) effects for cross-category stimuli can be explained by the predictive
4 coding theory.

5 This attenuation of P(rep) effect hints towards the fact that the influence of
6 prior experience with stimuli for the P(rep) effect might not be linear. Previous
7 studies showed that the P(rep) effect was only observed for stimuli with
8 extensive experience (faces, words, letters), but not for unfamiliar stimulus
9 categories (i.e., novel false fonts). We assume that the accumulation of
10 experience would prompt the emergence of the P(pre) effect that would not
11 change over time. However, based on the present findings, we speculate that
12 the P(rep) effect might have temporarily changed after a short-term training,
13 even for faces. In other words, it is likely that the P(rep) effect for faces could
14 have temporarily disappeared in our experiments, because faces might have
15 become unexpected stimuli for participants after the training session. Although
16 our findings could be explained under the framework of predictive coding, the
17 hypothesis that expectation effects could temporarily vary, remains only
18 speculative at present, and it should be systematically tested in future studies.

19 In conclusion, the current study reveals for the first time that the P(rep)
20 effect for non-face objects was abolished after a short-term training session.
21 This suggests that perceptual experiences from a short-term training modulate
22 the P(rep) effect on RS.

23 **Acknowledgements**

24 This work was supported by a Deutsche Forschungsgemeinschaft Grant
25 [KO3918/5-1]. Chenglin Li was supported by the China Scholarship Council
26 (CSC) scholarship (201808330399) during the study. We thank Marie Wächter
27 and Klemens Buttgereit for their help in the stimuli collection and thank
28 Sophie-Marie Rostalski for her help in the guideline of data analysis. The
29 authors thank Linda Ficco for her comments on this manuscript.

Reference

- Amado, C., Hermann, P., Kovács, P., Grotheer, M., Vidnyánszky, Z., & Kovács, G. (2016). The contribution of surprise to the prediction based modulation of fMRI responses. *Neuropsychologia*, 84, 105-112.
- Baeck, A., & Op de Beeck, H. P. (2010). Transfer of object learning across distinct visual learning paradigms. *Journal of vision*, 10(2), 17-17.
- Bi, T., & Fang, F. (2013). Neural plasticity in high-level visual cortex underlying object perceptual learning. *Frontiers in Biology*, 8(4), 434-443.
- Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature reviews neuroscience*, 3(3), 243-249.
- De Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception?. *Trends in cognitive sciences*, 22(9), 764-779.
- Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *Journal of Neuroscience*, 30(49), 16601-16608.
- Feuerriegel, D., Vogels, R., & Kovács, G. (2021). Evaluating the evidence for expectation suppression in the visual system. *Neuroscience and Biobehavioral Reviews*, 126, 368-381.
- Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature*, 287(5777), 43-44.
- Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, 360(1456), 815-836.
- Furmanski, C. S., & Engel, S. A. (2000). Perceptual learning in object recognition: Object specificity and size invariance. *Vision research*, 40(5), 473-484.
- Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature neuroscience*, 3(2), 191-197.

1 Goldstone, R. L. (1998). Perceptual learning. *Annual review of*
2 *psychology*, 49(1), 585-612.

3 Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R.
4 (1999). Differential processing of objects under various viewing
5 conditions in the human lateral occipital complex. *Neuron*, 24(1),
6 187-203.

7 Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics
8 of object-selective activation correlate with recognition performance in
9 humans. *Nature neuroscience*, 3(8), 837-843.

10 Grotheer, M., & Kovács, G. (2014). Repetition probability effects depend on
11 prior experiences. *Journal of Neuroscience*, 34(19), 6640-6646.

12 Henson, R. N. A., & Rugg, M. D. (2003). Neural response suppression,
13 haemodynamic repetition effects, and behavioural priming.
14 *Neuropsychologia*, 41(3), 263-270.

15 Henson, R., Shallice, T., & Dolan, R. (2000). Neuroimaging evidence for
16 dissociable forms of repetition priming. *Science*, 287(5456), 1269-1272.

17 Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., VanMeter, J., & Riesenhuber, M.
18 (2007). Categorization training results in shape-and category-selective
19 human neural plasticity. *Neuron*, 53(6), 891-903.

20 Kaliukhovich, D. A., & Vogels, R. (2011). Stimulus repetition probability does
21 not affect repetition suppression in macaque inferior temporal
22 cortex. *Cerebral Cortex*, 21(7), 1547-1558.

23 Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a
24 module in human extrastriate cortex specialized for face
25 perception. *Journal of neuroscience*, 17(11), 4302-4311.

26 Kok, P., & de Lange, F. P. (2015). Predictive coding in sensory cortex. In *An*
27 *introduction to model-based cognitive neuroscience* (pp. 221-244).
28 Springer, New York, NY.

29 Kovács, G., Iffland, L., Vidnyánszky, Z., & Greenlee, M. W. (2012). Stimulus

1 repetition probability effects on repetition suppression are position
2 invariant for faces. *Neuroimage*, 60(4), 2128-2135.

3 Kovács, G., Kaiser, D., Kaliukhovich, D. A., Vidnyánszky, Z., & Vogels, R.
4 (2013). Repetition probability does not affect fMRI repetition
5 suppression for objects. *Journal of Neuroscience*, 33(23), 9805-9812.

6 Larsson, J., & Smith, A. T. (2012). fMRI repetition suppression: neuronal
7 adaptation or stimulus expectation?. *Cerebral Cortex*, 22(3), 567-576.

8 Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W.
9 A., ... & Tootell, R. B. (1995). Object-related activity revealed by
10 functional magnetic resonance imaging in human occipital
11 cortex. *Proceedings of the National Academy of Sciences*, 92(18),
12 8135-8139.

13 Mayrhauser, L., Bergmann, J., Crone, J., & Kronbichler, M. (2014). Neural
14 repetition suppression: evidence for perceptual expectation in
15 object-selective regions. *Frontiers in Human Neuroscience*, 8, 225.

16 McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Gauthier, I.
17 (2012). The Vanderbilt Expertise Test reveals domain-general and
18 domain-specific sex effects in object recognition. *Vision research*, 69,
19 10-22.

20 Op de Beeck, H. P., & Baker, C. I. (2010). The neural basis of visual object
21 learning. *Trends in cognitive sciences*, 14(1), 22-30.

22 Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006).
23 Discrimination training alters object representations in human
24 extrastriate cortex. *Journal of Neuroscience*, 26(50), 13025-13036.

25 Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a
26 functional interpretation of some extra-classical receptive-field
27 effects. *Nature neuroscience*, 2(1), 79-87.

28 Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998). Cortical areas supporting
29 category learning identified using functional MRI. *Proceedings of the*

1 *National Academy of Sciences*, 95(2), 747-750.

2 Rostalski, S. M., Amado, C., & Kovács, G. (2019). Repetition suppression for
3 noisy and intact faces in the occipito-temporal cortex. *Frontiers in*
4 *psychology*, 10, 1348.

5 Rostalski, S. M., Amado, C., Kovács, G., & Feuerriegel, D. (2020). Measures
6 of repetition suppression in the Fusiform Face Area are inflated by
7 co-occurring effects of statistically learned visual
8 associations. *Cortex*, 131, 123-136.

9 Schoups, A. A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning
10 in identifying the oblique orientation: retinotopy, orientation specificity
11 and monocularly. *The Journal of physiology*, 483(3), 797-810.

12 Sigman, M., Pan, H., Yang, Y., Stern, E., Silbersweig, D., & Gilbert, C. D.
13 (2005). Top-down reorganization of activity in the visual pathway after
14 learning a shape identification task. *Neuron*, 46(5), 823-835.

15 Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T.
16 (2008). Neural repetition suppression reflects fulfilled perceptual
17 expectations. *Nature neuroscience*, 11(9), 1004.

18 Vinken, K., Op de Beeck, H. P., & Vogels, R. (2018). Face repetition probability
19 does not affect repetition suppression in macaque inferotemporal
20 cortex. *Journal of Neuroscience*, 38(34), 7492-7504.

21 Yu, C., Klein, S. A., & Levi, D. M. (2004). Perceptual learning in contrast
22 discrimination and the (minimal) role of context. *Journal of Vision*, 4(3),
23 4-4.

3.3 study III: 3a: Visual short-term memory load modulates repetition related fMRI signal adaptation

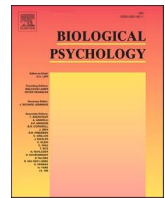
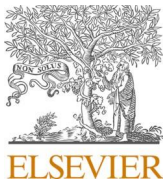
3b: Active attention counteracts the effects of increased load on visual short-term memory on repetition suppression

This section is composed of 3a and 3b. Section 3a corresponds to the manuscript that has been published in *Biological Psychology*. The section 3b is an unpublished experiment:

Li, C., Kovács, G., & Trapp, S. (2021). Visual short-term memory load modulates repetition related fMRI signal adaptation. *Biological Psychology*, 166, 108199. <https://doi.org/10.1016/j.biopsycho.2021.108199>

Main research question:

How and to which extent do top-down predictions induced by prior experience modulate the repetition suppression effect?



Visual short-term memory load modulates repetition related fMRI signal adaptation

Chenglin Li^a, Gyula Kovács^{a,1}, Sabrina Trapp^{b,c,1,*}

^a Department of Biological Psychology and Cognitive Neurosciences, Institute of Psychology, University of Jena, Jena, Germany

^b Department of Sport Science, University of Bielefeld, Bielefeld, Germany

^c Department of Psychology, University of Leipzig, Leipzig, Germany

ARTICLE INFO

Keywords:

Repetition suppression
Face perception
Fusiform face area
Predictive coding
Visual short-term memory

ABSTRACT

While several computational models have suggested how predictive coding could be implemented on an algorithmic level, reference to cognitive processes remains rather sparse. A crucial process might be elevating relevant prior information from long-term memory to render it highly accessible for subsequent comparison with sensory input. In many models, visual short-term memory (VSTM) is considered as information from long-term memory in a state of elevated activity. We measured the BOLD signal in face-specific cortical areas using repetition suppression (RS) paradigm. RS has been associated with predictive processing in previous studies. We show that RS within the fusiform face area is significantly attenuated when VSTM is loaded with other, non-facial visual information. Although an unequivocal inference is not possible, the data indicate a role of VSTM for predictive processes as indexed by expectation-related RS.

1. Introduction

Over the last two decades, predictive coding has become an influential model of how the brain processes sensory information (De Lange, Heilbron, & Kok, 2018; Friston, 2005; Rao & Ballard, 1999). The core assumption here is that incoming sensory information is compared to predictions hereof, and only the deviation or the *prediction error* is fed forward in the sensory hierarchy (Friston, 2010; Rao & Ballard, 1999). With *prediction*, we here refer to a process which operates according to conditional probabilities (i.e., what is likely to occur), and thereby differs from attention, a process which operates according to what is relevant (Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008). According to the predictive coding model, the brain attempts to match incoming, bottom-up sensory inputs with top-down predictions, and each region in the cortical sensory hierarchy represents both these predictions and the mismatch (prediction error) (Kok & de Lange, 2015). Although these ideas have been implemented computationally and pitted against behavioral or neural data in the past (e.g., Egner, Monti, & Summerfield, 2010; Jiang, Schmajuk, & Egner, 2012), explanations of how this is implemented from a cognitive point of view, e.g., by referring

to representation, memory, and retrieval processes, remain sparse.

Predictions about future events are possible because the environment consists of temporal and spatial regularities. However, not all regularities the brain has learnt are relevant for a specific perceptual task at hand. For instance, learnt regularities of an auditory sequence are only relevant when this sequence has to be predicted. What are the selection mechanisms that tag the relevant prior information from long-term memory as relevant for current perceptual inference? Several models assume that short-term memory is an activated part of long-term memory that is currently relevant to the individual or the task at hand (Anderson et al., 2004; Cowan, 1995; Oberauer, 2002; Ruchkin, Grafman, Cameron, & Berndt, 2003). Thus, it is possible that it is exactly this activated part of long-term memory that is required for perceptual inference – maintaining the predictive information in a highly accessible state to enable the comparison with incoming sensory input. Indeed, previous studies provide evidence that visual short-term memory (VSTM)² is required for paradigms that are assumed to tap into predictive processes. In the *contextual cueing paradigm*, participants have to find a target in layouts while some of the layouts are presented several times during the experiment, and thus allow participants to find the

* Correspondence to: Faculty of Psychology and Sport Science, University of Bielefeld, Bielefeld, Germany.

E-mail address: sabrina.trapp@uni-bielefeld.de (S. Trapp).

¹ These authors contributed equally.

² We use the term VSTM because our experimental task was related to storage of information only, as opposed to storage and manipulation, which is commonly associated with working memory. However, both concepts are often used synonymously.

position of these targets faster (Chun & Jiang, 1998). Several studies have reported an important role of STM for this paradigm. Specifically, when participants have to hold information in STM while performing this contextual cueing paradigm at the same time, the facilitation observed on repeated layouts is impeded, indicating that STM is required for predicting an object's location (Manginelli, Langer, Klose, & Pollmann, 2013; Travis, Mattingley, & Dux, 2013). The results of the contextual cueing paradigms are focused on spatial, rather than object or identity-based predictions. But it has also been demonstrated that VSTM capacity is correlated with neural correlates of object predictions during a probabilistic sequential task (Cashdollar, Ruhnu, Weisz, & Hasson, 2017). These findings indicate that VSTM may also be required for predicting the *identity* of upcoming sensory events. However, in this study, participants could only predict the occurrence of a given object category, but never its specific identity. Furthermore, the regularities of the stimulus sequences were not task-relevant, and thus processing of these regularities could not be determined. Finally, the results were only correlational, providing no causal evidence for the role of VSTM for identity predictions. St John-Saaltink, Utzerath, Kok, Lau, and de Lange (2015) demonstrated that the predictive processing of identities was attenuated under VSTM load. The authors presented auditory information (a tone) prior to the appearance of grating stimuli, and the tone deterministically predicted the orientation of the grating. The reduced BOLD response in V1 for predicted gratings was attenuated when participants had to keep additional, orientation-irrelevant information in their VSTM. However, this task required efforts in combining visual and auditory information. Executive or control processes are linked to working memory as well (Baddeley, 2003). Therefore, it cannot be excluded that the effect is more generic. Furthermore, the cue allowed participants to explicitly and consciously predict the occurrence of the upcoming stimulus, making the comparison with previous studies (which are focused on predictions based on implicit statistical regularities) difficult.

There are several ways by which the brain might make predictions about the environment: Using explicit cues, capitalizing upon implicit temporal or spatial regularities, (like in the contextual cueing paradigm), or by exploiting higher-level information, such as stimulus-context or semantic relationships (Bar, 2007; Turk-Browne, Scholl, Johnson, & Chun, 2010; Kimura, Kondo, Ohira, & Schröger, 2012; Kok, Failing, & de Lange, 2014; Bendixen, Schwartze, & Kotz, 2015; Jessen and Kotz, 2015; Trapp, Lepsien, Kotz, & Bar, 2016). In principle, a stimulus that is repeated over time allows for a prediction (of the repeated stimulus) as well. What happens on a neural level when stimuli are repeated? A very robust finding is that the BOLD response is reduced in several cortical areas, a phenomenon referred to as *repetition suppression* (RS; for a review see Grill-Spector, Henson, & Martin, 2006). While originally this effect was explained by neural fatigue, adaptation (for a discussion see Larsson & Smith, 2012) or sharpening of representations (Freedman, Riesenhuber, Poggio, & Miller, 2006), the data have been re-interpreted as reflecting a reduction of the magnitude of the predictive error (for a review, Grotheer & Kovács, 2016). In a seminal study, Summerfield et al. (2008) found that RS is enhanced when repetitions become more probable for faces; suggesting a link between RS and predictive processes. Since then, several modulating factors have been identified, and various other studies support the idea that RS is linked to predictive processing (Auksztulewicz & Friston, 2016; Kovács & Vogels, 2014), at least for face stimuli (Grotheer & Kovács, 2014). The present study aims to examine whether VSTM is required for visual predictions as elicited by the RS paradigms. If VSTM is required for predictive processes, the RS effect should be significantly attenuated when participants VSTM is filled.

2. Material and methods

2.1. Participants

Twenty healthy participants with normal or corrected-to-normal vision participated in the experiment (15 females; mean age, 23.4 years; SD, 2.8 years; one left-handed). Written informed consent was obtained from all participants and all participants received their own printed 3D brain model, course credit or monetary compensation for their participation. The experiment was approved by the ethics committee of the Friedrich-Schiller-Universität Jena and conducted in accordance with the guidelines of the Declaration of Helsinki.

2.2. Stimuli

One hundred and eight gray-scale images of unfamiliar Caucasian faces (54 female faces) were selected from a set of faces used in previous studies by one of the authors (Kovács, Iffland, Vidnyánszky, & Greenlee, 2012; Kovács, Kaiser, Kaliukhovich, Vidnyánszky, & Vogels, 2013). All images displayed a neutral expression and were cropped into a circular frame to remove the external facial features. All images were resized to 350×370 pixel. Twenty Chinese characters (Simhei font) were used for the VSTM task, with a size of 130×130 pixel. All of these characters were unfamiliar to the participants.

2.3. Procedure

We used a simple stimulus repetition paradigm to measure RS, together with either concurrent VSTM load or a control condition (color discrimination task). The experimental procedures are illustrated in Fig. 1. Psychtoolbox (Version 3.0.15) was used for stimulus presentation and behavioral response collection, controlled by Matlab R2013a (The MathWorks, Natick, MA, USA).

Participants completed four experimental runs, each run included 72 trials, and those included four different trial types (VSTM load and alternating face image-pairs: VSTM_Alt, VSTM load and repeated face image-pairs: VSTM_Rep, No VSTM load and alternating face image-pairs: Control_Alt, and No VSTM load and repeated face image-pairs: Control_Rep). In each run, those four kinds of trials (18 trials per trial type) were randomly presented. In each single trial, a test display composed of a fixation cross and six Chinese characters (VSTM condition) or six black 'X' characters (control condition; size: $1.4^\circ \times 1.4^\circ$) were presented around the center of the screen (fixation-cross – stimulus distance: 4.5°), simultaneously for 2000 ms and followed by an inter-stimulus interval, jittered among 1000, 2000 or 3000 ms. Next, a face stimulus-pair was presented (exposition time: 250 ms each), separated by an inter-stimulus-interval (ISI) of 300–500 ms (randomized across trials). The face pairs could either present the same face (repeated-Rep) or depict two different same-sex persons (alternated-Alt). The stimulus pairs were followed by a blank period and a choice screen (500–700 ms). The probe screen contained either a Chinese character (VSTM) or a red/blue X (control) presented for 500 ms. The trials were separated by an Inter-trial-interval (ITI) of 2000 or 3000 ms. In the VSTM condition, participants had to memorize the Chinese characters of the first, test-display screen as accurately as possible, and were required to indicate at the end of the trial, by pressing a button, whether the character of the probe-screen was present in the initial screen or not (present/absent decision). In the control condition, participants were merely asked to discriminate the color (Red/ Blue) of the probe ('X') by pressing a button. Each experimental run lasted about 10 min.

2.4. Imaging parameters and data analysis

Neuroimaging was performed with a 3-Tesla MR scanner (Siemens MAGNETOM Prisma, Erlangen, Germany). During the functional runs, to obtain a temporal resolution of 1 s, a 64-channel head coil was used

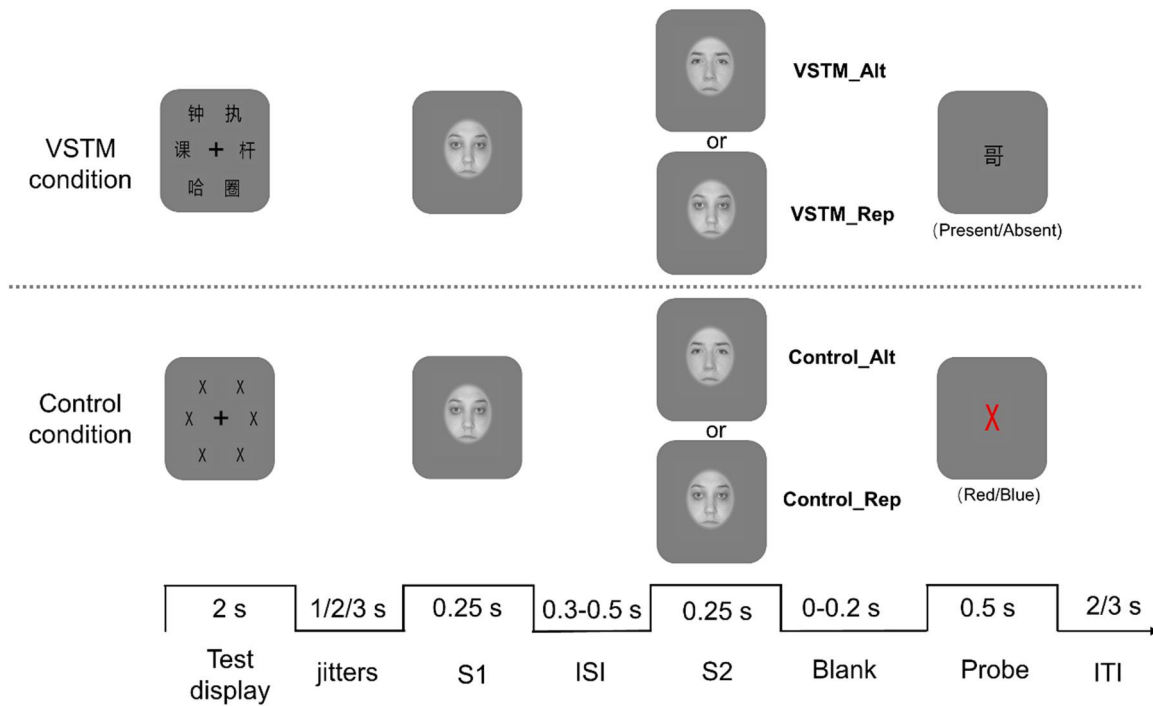


Fig. 1. Schematic illustration of the trial structure. The test display could either be a fixation cross and six Chinese characters that participants had to remember (VSTM condition) or six black Xs that participants were just asked to watch (control condition). Independently of this manipulation, the stimulus pairs presented hereafter could either depict the same face (repeated-Rep) or two different, same-sex persons (alternated-Alt). The probe at the end of each trial contained either a Chinese character (and participants had to make a present/absent decision) or a red/blue X (and participants simply had to make a color discrimination task).

with multi-band EPI sequences (MB acceleration factors = 8). T2 * weighted images were collected with the following parameters: FOV = 200 × 200 mm²; 64 slices; TR = 1000 ms; TE = 45.8 ms; flip angle = 90 deg; 2 mm isotropic voxel size. High-resolution T1 * weighted images (192 slices; TR = 2300 ms; TE = 3.03 ms; flip angle = 90 deg; 1 mm isotropic voxel size) were acquired to obtain a 3D structural scan with the same 64-channel head coil. Data were pre-processed using SPM12 (Wellcome Department of Imaging Neuroscience, London, UK). Briefly, the functional images were slice-timed, realigned, co-registered to structural scan, afterward the functional images were normalized to the MNI-152 space and smoothed using a 6-mm Gaussian kernel.

Regions of interests (ROIs) were defined by separate functional localizer runs (Rostalski, Amado, Kovács, & Feuerriegel, 2020). Faces, objects, and Fourier-randomized noise patterns were presented (4 Hz; 230 ms exposition time; 20 ms inter-stimulus-interval) in blocks of 10 s, interrupted by breaks of 10 s and repeated four times. A passive view task was performed in the localizer run. The rFFA was localized by contrasting face blocks with object and noise blocks individually and the location was established as the local maximum from the t-maps with a threshold of $p < 0.05$ family-wise error corrected ($n = 18$) or $p < 0.0001$ uncorrected ($n = 2$) on the single-subject level. The mean MNI coordinates (\pm SD) for the rFFA were 39.3 (3.85), -51.2 (5.25), -16 (4.15), average cluster size (\pm SE) for rFFA was 413 (203) voxels. As previous studies, applying probabilistic repetition paradigms with faces, found that the right FFA (rFFA) shows the largest modulation of RS by predictions (Grotheer, Hermann, Vidnyánszky, & Kovács, 2014; Kovács et al., 2012, 2013) here we concentrated on the signal modulation within the rFFA. The results of other face and object selective ROIs are presented in the [supplementary material](#). Canonical hemodynamic response functions were extracted using MarsBaR 0.44 toolbox for SPM (Brett, Johnsrude, & Owen, 2002). The peak BOLD values were extracted and analyzed by repeated measurement ANOVAs with the factors VSTM load (VSTM vs. Control) and trial type (alternated vs. repeated) as within-subject factors.

The behavioral results of performance and reaction time were analyzed similarly by two-way repeated-measures ANOVAs with VSTM load (VSTM vs. Control) and trial type (alternated vs. repeated) as within-subject factors. All *post-hoc* measurements were corrected by the Bonferroni method (Field, 2009).

To further prevent accidentally rejecting the null hypothesis, Bayesian repeated measures ANOVAs were also performed. Here, the inclusion Bayes Factors (BF) are reported, which provide the quantified evidence in the data for including a predictor in a model (van den Bergh et al., 2020). We compared only “matched” models (JASP Team, 2020). This model comparison approach can provide a straightforward interpretation of the main effects and interactions in a multi-factor design. To show the results in a concise format, inclusion BFs are presented alongside the frequentist ANOVA results (Li et al., 2021).

3. Results

3.1. Behavioral results

The behavioral results are presented in Fig. 2. A significant main effect of VSTM load was found both for accuracy ($F(1, 19) = 607.971$, $p < 0.001$, $\eta_p^2 = 0.970$, $BF_{10} > 1000$) and reaction time data ($F(1, 19) = 79.481$, $p < 0.001$, $\eta_p^2 = 0.807$, $BF_{10} > 1000$). The results showed that the accuracy was higher ($97.2 \pm 0.6\%$; average and standard error) and the reaction time was shorter (613 ± 28 ms) in the control as compared to the VSTM condition ($59.9 \pm 1.6\%$; reaction time = 913 ± 43 ms). There was no significant main effect of trial type on accuracy ($F(1, 19) = 0.002$, $p = 0.968$, $\eta_p^2 < 0.001$, $BF_{10} = 0.224$) or reaction time data ($F(1, 19) = 0.001$, $p = 0.971$, $\eta_p^2 < 0.001$, $BF_{10} = 0.235$).

3.2. Repetition suppression modulation in the rFFA

There was no significant main effect of VSTM load ($F(1, 19) = 0.251$, $p = 0.622$, $\eta_p^2 = 0.013$, $BF_{10} = 0.258$) and trial type ($F(1, 19) = 0.694$,

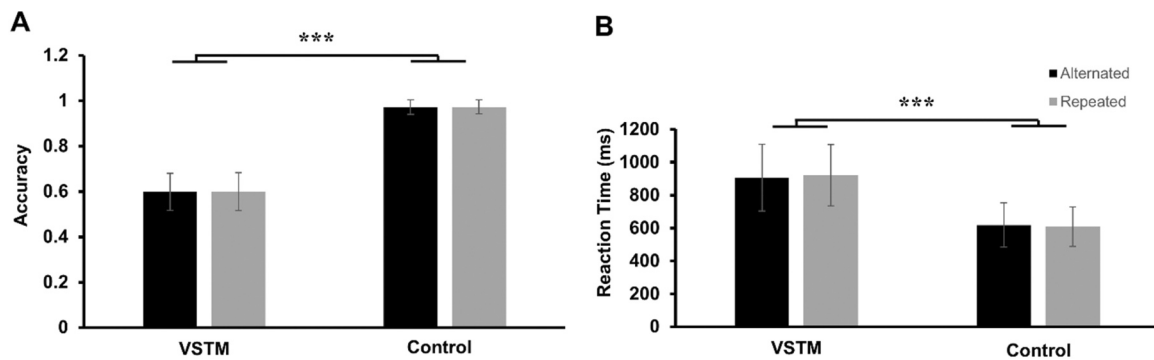


Fig. 2. Accuracy (A) and reaction time (B) for the VSTM and control conditions, separately for alternated (black) and repeated (grey) trials. Error bars represent standard errors of means. *** $p < 0.001$.

$p = 0.415$, $\eta_p^2 = 0.035$, $BF_{10} = 0.301$). We found a significant interaction between VSTM load and trial type for the peak BOLD signal ($F(1, 19) = 4.534$, $p = 0.047$, $\eta_p^2 = 0.193$, $BF_{10} = 1.379$). Importantly, this RS effect was significantly attenuated ($t(19) = 0.793$, $p = 0.437$, *Cohen's* $d = 0.13$, $BF_{10} = 0.142$) in the VSTM condition. A *post-hoc* test revealed that there was a significant RS effect in the control condition ($t(19) = 2.298$, $p = 0.033$, *Cohen's* $d = 0.34$, $BF_{10} = 3.768$), i.e., repeated trials led to lower responses as compared to alternating ones, confirming previous results (e.g., Grotheer et al., 2014; Grotheer & Kovács, 2015; Kovács et al., 2012, 2013; Rostalski et al., 2020). Fig. 3.

4. Discussion

While accumulating evidence suggests the necessity of VSTM for spatial predictions (Manginelli et al., 2013; Travis et al., 2013), its role for object or identity predictions remains sparsely addressed. In the current study, we tested if VSTM is required for visual predictions as assessed by an RS paradigm. Consistent with previous studies (e.g., Kovács et al., 2012, 2013; Rostalski et al., 2020), we found strong and robust RS effects in the rFFA without additional VSTM load. Importantly, RS was significantly attenuated when participants' VSTM was engaged. Thus, our results suggest that VSTM plays an important role in the ability to access and/or use stored identify-specific stimulus

regularities for predictions.

Importantly, however, it has to be mentioned that other factors can account for the reduction of RS in the VSTM condition relative to the control condition.

First, a limitation of cognitive resources under VSTM load could be a potential, more generic explanation of the reduced RS under VSTM load. In other words, when VSTM was engaged, there remained no additional generic capacity for processing the face stimuli. Several neuroimaging studies showed that high perceptual load modulates the neural activity of task-irrelevant distractors (e.g., Rees, Russell, Frith, & Driver, 1999; Yi, Woodman, Widders, Marois, & Chun, 2004; for a review, Lavie, 2005). For instance, Yi et al. (2004) presented faces in the center of the screen with a background scene image. Participants were instructed to ignore the background scene and perform a one-back task on the central faces. The results showed that the observed RS effect for the scene images within the parahippocampal place area (PPA) is abolished in this high perceptual load condition. Similarly, it is possible that RS in the rFFA under the VSTM condition was reduced for the same reason in the current study. However, as opposed to the study by Yi et al. (2004), the faces in this study were presented at the center, and participants were not asked to ignore them. Future research should directly test this hypothesis by matching the overall generic load in the two conditions.

Second, there is great consensus that attention is a necessary prerequisite for short-term memory activity. Specifically, it has been suggested that attention serves as a *rehearsal mechanism* in visuo-spatial working memory (Awh & Jonides, 2001). In other words, attention contributes to the maintenance of internal information over a brief period of time (e.g., Awh, Jonides, & Reuter-Lorenz, 1998; Awh, Vogel, & Oh, 2006; Jha, 2002; Trapp & Lepsien, 2012). It has even been proposed that VSTM and attention should no longer be considered as separate systems or concepts as they rely, to a large extent, on the same limited resources (Kiyonaga & Egner, 2013). Importantly, attention is also known to modulate the magnitude of expectation-related RS (Kok, Jehee, & De Lange, 2012; Larsson & Smith, 2012; Todorovic, van Ede, Maris, & de Lange, 2011). When participant's attention was diverted away from the RS stimuli, a large, expectation-related component of the response reduction was eliminated (Larsson & Smith, 2012). Thus, it is possible that under VSTM load, the allocation of attention was partly withdrawn from the face stimuli during the RS part of the task, reducing the overall expectation-related RS effect. In principle, this could be experimentally addressed by having an attention task that has no memory or maintenance component as a control condition. For instance, during presentation of RS face stimuli, participants have to track a ball in the periphery of the screen or make a forced choice decision on some other visual stimuli. However, this experimental approach has drawbacks – if the interaction (i.e., the RS attenuation under VSTM) diminishes in this design, it could mean that (i) the RS attenuation under VSTM is just due to attentional withdrawal, and not linked to VSTM maintenance activity, or (ii) the RS attenuation in the VSTM load

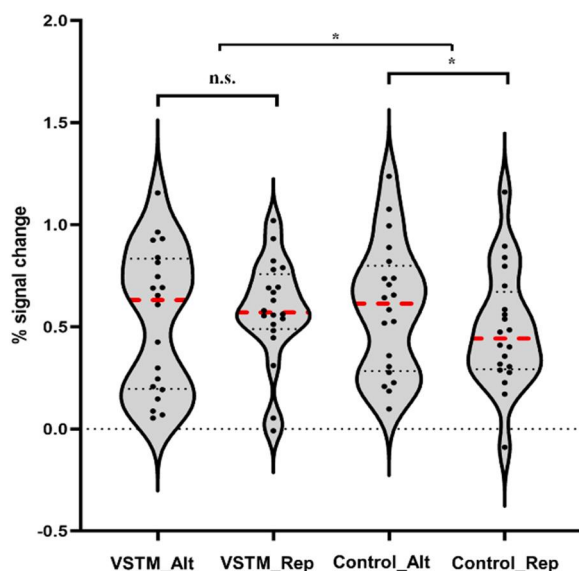


Fig. 3. The BOLD signal of the right FFA for the VSTM and control conditions, separately for alternating and repeated trials. Dots represent individual data points; red dotted lines show the group median. Shaded areas depict the distribution of data for each condition. * $p < 0.05$, n.s. indicates no significant.

condition is due to maintenance being consumed by holding content in VSTM, and the RS attenuation in the attention control condition is due to attentional withdrawal (which had been reported before). Previous studies showed that tasks with different short-term memory loads also involve different amounts of eye movements during rehearsal (Droll & Hayhoe, 2007), and suggest that eye gazes tracks attention shifts in VSTM (Van Ede, Chekroud, & Nobre, 2019). Accordingly, eye gaze differences as a variant of overt attention could be another alternative explanation for the present data. This could in principle be addressed by a combined fMRI and eye-tracking experiment. Note, however, that the inherent difficulty remains present here as well, as covert attention can be allocated differently, even when the overt attention component is kept equal.

Our study rests on the assumption that by repeating a stimulus, deviations of predictions are being reduced. There are a lot of studies which provide evidence for this conjecture (Aukstulewicz & Friston, 2016; Grotheer & Kovács, 2014; Kovács & Vogels, 2014; Summerfield et al., 2008). Importantly, there are studies which suggest that repetition and expectation effects are subserved by dissociable neural processes (e. g., Todorovic & de Lange, 2012). Based on these findings, it has been discussed that the underlying mechanisms are different, in that expectation can be considered as sharpening neural populations (Press & Yon, 2019) whereas repetition acts as a form of local scaling (Alink, Abdulrahman, & Henson, 2018). With the current design, it is not possible to unequivocally decide which mechanism was affected by loaded VSTM. However, previous evidence suggests that RS is independent of attention, whereas the expectation-suppression is not (Larsson & Smith, 2012), thus it is very likely that the latter component was affected by our manipulation.

While the current data do not allow unequivocal inference on the exact causes or mechanisms of the attenuation of RS by loaded VSTM, they still demonstrate that VSTM seems to be central for full manifestation of RS. To the best of our knowledge, this has not been shown before. Importantly, the data also highlight the importance of shedding light on cognitive processes that are relevant for predictive processes. Short-term memory is an important cognitive faculty and had recently been associated with predictive processing on conceptual grounds (Trapp, Parr, Friston, & Schroeger, in press). The discussion of its core components (encoding, maintenance, retrieval) for predictive process has been rather neglected so far.

While past research on predictive processing has largely focused on general effects, there is now raising interest in interindividual differences (e.g., Siegelman, Bogaerts, Christiansen, & Frost, 2017). Therefore, our data are important, as there are inter-individual differences in VSTM capacity limitation (Cowan, 2001), that may interact with RS as well. Furthermore, depressed individuals show peculiarities in short-term memory processing, just as other clinically relevant disorders (Joormann & Gotlib, 2008). It would be worthily to assess whether and to what extend RS is attenuated in these disorders, whether the attenuation can be explained by a disturbed short-term memory, and if the RS attenuation under VSTM load can support early detection of depression or other clinical disorders.

Acknowledgments

This work was supported by a Deutsche Forschungsgemeinschaft Grant given to G.K. (KO3918/5-1) and S.T. (TR 1543/2-1). Chenglin Li was supported by the China Scholarship Council (CSC) scholarship (201808330399) during the study. We thank Marie Wächter for helping with participants' recruitment and Sophie-Marie Rostalski for guidance in data analysis.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.biopsycho.2021.108199](https://doi.org/10.1016/j.biopsycho.2021.108199).

References

- Alink, A., Abdulrahman, H., & Henson, R. N. (2018). Forward models demonstrate that repetition suppression is best modelled by local neural scaling. *Nature Communications*, 9(1), 1–10. <https://doi.org/10.1038/s41467-018-05957-0>
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4), 1036–1060. <https://doi.org/10.1037/0033-295X.111.4.1036>
- Aukstulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex*, 80, 125–140. <https://doi.org/10.1016/j.cortex.2015.11.024>
- Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, 5(3), 119–126. [https://doi.org/10.1016/S1364-6613\(00\)01593-X](https://doi.org/10.1016/S1364-6613(00)01593-X)
- Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology-Human Perception and Performance*, 24(3), 780–790. <https://doi.org/10.1037/0096-1523.24.3.780>
- Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience*, 139(1), 201–208. <https://doi.org/10.1016/j.neuroscience.2005.08.023>
- Baddeley, A. D. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829–839. <https://doi.org/10.1038/nrn1201>
- Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11(7), 280–289. <https://doi.org/10.1016/j.tics.2007.05.005>
- Bendixen, A., Schwartz, M., & Kotz, S. A. (2015). Temporal dynamics of contingency extraction from tonal and verbal auditory sequences. *Brain and Language*, 148, 64–73. <https://doi.org/10.1016/j.bandl.2014.11.009>
- van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E., Derks, K., & Wagenmakers, E. (2020). A tutorial on conducting and interpreting a Bayesian ANOVA in JASP. *L'Année Psychologique*, 1(1), 73–96. <https://doi.org/10.3917/anpsy1.201.0073>
- Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature reviews Neuroscience*, 3(3), 243–249. <https://doi.org/10.1038/nrn756>
- Cashdollar, N., Ruhnau, P., Weisz, N., & Hasson, U. (2017). The role of working memory in the probabilistic inference of future sensory events. *Cerebral Cortex*, 27, 2955–2969. <https://doi.org/10.1093/cercor/bhw138>
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36(1), 28–71. <https://doi.org/10.1006/cogp.1998.0681>
- Cowan, N. (1995). *Attention and memory: an integrated framework*. New York: Oxford University Press.
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral Brain Sciences*, 24, 87–114.
- Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology: Human Perception and Performance*, 33(6), 1352–1365. <https://doi.org/10.1037/0096-1523.33.6.1352>
- Van Ede, F., Chekroud, S. R., & Nobre, A. C. (2019). Human gaze tracks attentional focusing in memorized visual space. *Nature Human Behaviour*, 3(5), 462–470. <https://doi.org/10.1038/s41562-019-0549-y>
- Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *Journal of Neuroscience*, 30, 16601–16608. <https://doi.org/10.1523/JNEUROSCI.2770-10.2010>
- Field, A. P. (2009). *Discovering statistics using SPSS: (and sex, drugs and rock 'n' roll)* (3rd ed.). Los Angeles: SAGE Publications.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cerebral Cortex*, 16(11), 1631–1644. <https://doi.org/10.1093/cercor/bhj100>
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B Biological Sciences*, 360(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11, 127–138. <https://doi.org/10.1038/nrn2787>
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, 10(1), 14–23. <https://doi.org/10.1016/j.tics.2005.11.006>
- Grotheer, M., Hermann, P., Vidnyánszky, Z., & Kovács, G. (2014). Repetition probability effects for inverted faces. *NeuroImage*, 102, 416–423. <https://doi.org/10.1016/j.neuroimage.2014.08.006>
- Grotheer, M., & Kovács, G. (2014). Repetition probability effects depend on prior experiences. *Journal of Neuroscience*, 34(19), 6640–6646. <https://doi.org/10.1523/JNEUROSCI.5326-13.2014>
- Grotheer, M., & Kovács, G. (2015). The relationship between stimulus repetitions and fulfilled expectations. *Neuropsychologia*, 67, 175–182. <https://doi.org/10.1016/j.neuropsychologia.2014.12.017>
- Grotheer, M., & Kovács, G. (2016). Can predictive coding explain repetition suppression? *Cortex*, 80, 113–124. <https://doi.org/10.1016/j.cortex.2015.11.027>
- JASP Team (2020). *JASP (Version 0.14.1)* [Computer software].
- Jessen, S., & Kotz, S. A. (2015). Affect differentially modulates brain activation in uni- and multisensory body-voice perception. *Neuropsychologia*, 66, 134–143. <https://doi.org/10.1016/j.neuropsychologia.2014.10.038>
- Jha, A. P. (2002). Tracking the time-course of attentional involvement in spatial working memory: an event-related potential investigation. *Cognitive Brain Research*, 15(1), 61–69. [https://doi.org/10.1016/S0926-6410\(02\)00216-1](https://doi.org/10.1016/S0926-6410(02)00216-1)

- Jiang, J., Schmajuk, N., & Egner, T. (2012). Explaining neural signals in human visual cortex with an associative learning model. *Behavioral Neuroscience*, 126(4), 575–581. <https://doi.org/10.1037/a0029029>
- Joormann, J., & Gotlib, I. H. (2008). Updating the contents of working memory in depression: interference from irrelevant negative material. *Journal of Abnormal Psychology*, 117(1), 182–192. <https://doi.org/10.1037/0021-843X.117.1.182>
- Kimura, M., Kondo, H., Ohira, H., & Schröger, E. (2012). Unintentional temporal context-based prediction of emotional faces: an electrophysiological study. *Cerebral Cortex*, 22, 1774–1785. <https://doi.org/10.1093/cercor/bhr244>
- Kiyonaga, A., & Egner, T. (2013). Working memory as internal attention: toward an integrative account of internal and external selection processes. *Psychonomic Bulletin and Review*, 20, 228–242. <https://doi.org/10.3758/s13423-012-0359-y>
- Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience*, 26(7), 1546–1554. https://doi.org/10.1162/jocn_a_00562
- Kok, P., Jehee, J. F., & De Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, 75(2), 265–270. https://doi.org/10.1162/jocn_a_00562
- Kok, P., & de Lange, F. P. (2015). *Predictive coding in sensory cortex. In An introduction to model-based cognitive neuroscience* (pp. 221–244). New York, NY.: Springer.
- Kovács, G., Iffland, L., Vidnyánszky, Z., & Greenlee, M. W. (2012). Stimulus repetition probability effects on repetition suppression are position invariant for faces. *Neuroimage*, 60(4), 2128–2135. <https://doi.org/10.1016/j.neuroimage.2012.02.038>
- Kovács, G., Kaiser, D., Kaliukhovich, D. A., Vidnyánszky, Z., & Vogels, R. (2013). Repetition probability does not affect fMRI repetition suppression for objects. *Journal of Neuroscience*, 33(23), 9805–9812. <https://doi.org/10.1523/JNEUROSCI.3423-12.2013>
- Kovács, G., & Vogels, R. (2014). When does repetition suppression depend on repetition probability? *Frontiers in Human Neuroscience*, 8, 685. <https://doi.org/10.3389/fnhum.2014.00685>
- De Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, 22(9), 764–779. <https://doi.org/10.1016/j.tics.2018.06.002>
- Larsson, J., & Smith, A. T. (2012). fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cerebral Cortex*, 22(3), 567–576. <https://doi.org/10.1093/cercor/bhr119>
- Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, 9(2), 75–82. <https://doi.org/10.1016/j.tics.2004.12.004>
- Li, C., Wang, Z., Bao, H., Wang, J., Chen, S., & Cao, X. (2021). The left-side bias is not unique to own-race face processing. *Attention, Perception, & Psychophysics*, 1–9. <https://doi.org/10.3758/s13414-021-02264-7>
- Manginelli, A. A., Langer, N., Klose, D., & Pollmann, S. (2013). Contextual cuing under working memory load: selective interference of visuospatial load with expression of learning. *Attention, Perception, and Psychophysics*, 75, 1103–1117. <https://doi.org/10.3758/s13414-013-0466-5>
- Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology-Learning Memory and Cognition*, 28(3), 411–421. <https://doi.org/10.1037/0278-7393.28.3.411>
- Press, C., & Yon, D. (2019). Perceptual prediction: rapidly making sense of a noisy world. *Current Biology*, 29(15), R751–R753. <https://doi.org/10.1016/j.cub.2019.06.054>
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87. <https://doi.org/10.1038/4580>
- Rees, G., Russell, C., Frith, C. D., & Driver, J. (1999). Inattention blindness versus inattentional amnesia for fixated but ignored words. *Science*, 286(5449), 2504–2507. <https://doi.org/10.1126/science.286.5449.2504>
- Rostalski, S. M., Amado, C., Kovács, G., & Feuerriegel, D. (2020). Measures of repetition suppression in the Fusiform Face Area are inflated by co-occurring effects of statistically learned visual associations. *Cortex*, 131, 123–136. <https://doi.org/10.1016/j.cortex.2020.07.010>
- Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: a state of activated long-term memory. *Behavioral and Brain Sciences*, 26(6), 709–728. <https://doi.org/10.1017/s0140525x03000165>
- Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*, 372(1711), Article 20160059.
- St John-Saaltink, E., Utzerath, C., Kok, P., Lau, H. C., & de Lange, F. P. (2015). Expectation suppression in early visual Cortex depends on task set. *PLoS One*, 10(6), Article e0131172. <https://doi.org/10.1371/journal.pone.0131172>
- Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, 11, 1004–1006. <https://doi.org/10.1038/nn.2163>
- Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *Journal of Neuroscience*, 31(25), 9118–9123. <https://doi.org/10.1523/JNEUROSCI.1425-11.2011>
- Todorovic, A., & de Lange, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *The Journal of Neuroscience*, 32, 13389–13395. <https://doi.org/10.1523/JNEUROSCI.2227-12.2012>
- Trapp, S., & Lepsien, J. (2012). Attentional orienting to mnemonic representations: reduction of load-sensitive maintenance-related activity in the intraparietal sulcus. *Neuropsychologia*, 50(12), 2805–2811. <https://doi.org/10.1016/j.neuropsychologia.2012.08.003>
- Trapp, S., Lepsien, J., Kotz, S., & Bar, M. (2016). Prior probability modulates baseline activity in category-specific areas. *Cognitive, Affective, & Behavioral Neuroscience*, 16, 135–144. <https://doi.org/10.3758/s13415-015-0373-4>
- Trapp, S., Parr, T., Friston, K., Schroeger, E. (in press). *Current Directions in Psychological Sciences The Predictive Brain Must Have A Limitation In Short-Term Memory Capacity*.
- Travis, S. L., Mattingley, J. B., & Dux, P.-E. (2013). On the role of working memory in spatial contextual cueing. *Journal of Experimental Psychology Learning, Memory, and Cognition*, 39, 208–219. <https://doi.org/10.1037/a0028644>
- Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *The Journal of Neuroscience*, 30, 11177–11187. <https://doi.org/10.1523/JNEUROSCI.0858-10.2010>
- Yi, D. J., Woodman, G. F., Widders, D., Marois, R., & Chun, M. M. (2004). Neural fate of ignored stimuli: dissociable effects of perceptual and working memory load. *Nature Neuroscience*, 7(9), 992–996. <https://doi.org/10.1038/nn1294>

Supplementary

Results

A two-by-two repeated measures ANOVA with factors VSTM load (VSTM, Control) and trial type (alternated vs. repeated) was performed on peak BOLD signals in the IFFA ($n = 18$), rOFA ($n = 19$), IOFA ($n = 17$), rLO ($n = 19$), and ILO ($n = 19$). The mean MNI coordinates (\pm SD) for those ROIs were IFFA: -41.0 (3.83), -55.8 (10.76), -15.2 (4.71), rOFA: 39.3 (5.17), -65.9 (8.58), -13.2 (5.05), IOFA: -39.2 (5.00), -69.4 (9.87), -14.2 (4.74), rLO: 44.0 (4.32), -77.2 (6.64), -4.0 (5.37), and ILO: -42.7 (4.53), -78.8 (4.96), -2.1 (5.68), respectively. Bayesian repeated measures ANOVAs were also performed for all ROIs.

In the IFFA, there was a significant interaction between the factors VSTM load and trial type ($F(1, 17) = 8.754$, $p = 0.009$, $\eta_p^2 = 0.340$, $BF_{10} = 2.474$). A *post-hoc* tests revealed that there was a significant RS effect in the control task ($t(17) = 3.373$, $p = 0.004$, *Cohen's d* = 0.48), but the RS effect was absent in the VSTM task ($t(17) = 0.793$, $p = 0.439$, *Cohen's d* = 0.18).

In the rOFA, there was a significant interaction between the factors VSTM load and trial type ($F(1, 18) = 6.084$, $p = 0.024$, $\eta_p^2 = 0.253$, $BF_{10} = 0.954$). A *post-hoc* tests revealed that there was a significant RS effect in the control task ($t(18) = 3.139$, $p = 0.006$, *Cohen's d* = 0.43), but the RS effect was absent in the VSTM tasks ($t(18) = 0.686$, $p = 0.502$, *Cohen's d* = 0.10).

There was no significant main effect or interaction (all $F_s \leq 1.535$, all $BF_{10} \leq 0.580$) in the IOFA.

In the rLO, there was a significant main effect of VSTM load ($F(1, 18) = 5.491, p = 0.031, \eta_p^2 = 0.234, BF_{10} = 35.168$). the BOLD signal of VSTM task ($M = 0.404, SD = 0.292$) was stronger than control task ($M = 0.165, SD = 0.410$).

In the lLO, there was a significant main effect of VSTM load ($F(1, 18) = 5.881, p = 0.026, \eta_p^2 = 0.246, BF_{10} = 71.875$). The BOLD signal of VSTM task ($M = 0.360, SD = 0.344$) was stronger than control task ($M = 0.131, SD = 0.349$).

3b: Active attention counteracts the effects of increased load on visual short-term memory on repetition suppression

Chenglin Li¹, Gyula Kovács^{1*}Sabrina Trapp^{2,3*}

¹ Department of Psychology, University of Jena, Jena, Germany

² Department of Sport Science, University of Bielefeld, Bielefeld, Germany

³ Department of Psychology, University of Leipzig, Leipzig, Germany

*These authors contributed equally

Abstract

The results of study 3a showed that repetition suppression (RS) within the fusiform face area is significantly attenuated when visual short-term memory (VSTM) is loaded with additional visual information. It suggests the important role of VSTM for predictive processes as indexed by expectation-related RS. However, it remains unclear whether this attenuated effect is related to attention. In this control experiment, we explicitly manipulated attention through an inverted-face detection task. In contrast with the results in study 3a, when participants attend to faces, the RS effect in the fusiform face area appears again despite the increased load on the VSTM. This suggests that active attention can effectively counteract the reduction of RS effect due to increased visual load.

Interim discussion

In Study 3a, we aimed to test whether visual short-term memory (VSTM) would be required for implicit predictions, as assessed by a typical repetition suppression (RS) paradigm. The data in Study 3a showed a significant interaction between the magnitude of RS and VSTM load, in the sense that the response suppression within the right FFA was significantly attenuated when participants had to maintain additional, non-facial information in VSTM (Li et al., 2021). However, in that experiment, participants had no specific face-related task during the presentation of the repeated and alternating face images in either condition. Thus, as attention could be directed away from the face stimuli in the VSTM trials, it is possible that the observed modulatory effect of short-term memory load on RS effects is due to differentially allocated attentional resources. Indeed, several studies suggested a close link between attention and VSTM (for a review see Awh & Jonides, 2001; Awh et al., 2006). Importantly, Larsson and Smith (2012) showed that diverting attention away from the stimuli eliminates a large, expectation-related component of the response reductions in several cortical areas. This supports the idea that the absence of RS effects in our previous study could simply be due to participants not paying sufficient attention to faces during the high load condition. To test this potential explanation, we performed a second experiment where we explicitly controlled for attention across the previous VSTM and Control conditions equally, by instructing participants to attend to the faces and to signal the occurrence of inverted faces.

Methods

Participants.

Twenty-three healthy participants took part in this experiment. All of them were informed about the experimental procedures and gave written informed consent beforehand. Importantly, none of them participated in Study 3a. They received their own printed 3D brain model, course credit or monetary compensation for their participation. Two participants were excluded from the

analysis due to excessive head movements during image acquisition. The remaining 21 participants (16 females; mean age, 23.8 years; SD, 2.9 years; two left-handed) had normal or corrected-to-normal vision. The same ethical guidelines were followed as in Study 3a.

Stimuli. The stimuli were identical to those of Study 3a.

Procedure and data analysis.

The setup, imaging parameters and the experimental procedure of nontarget trials were identical to those of Study 3a. To control participants' attention during the presentation of face pairs, we embedded 8 target trials (12 %) into each run (equally in the VSTM and Control conditions). For these target trials, the first or the second face of the pair was inverted, and participants were required to report which of the two faces was inverted by pressing one of two buttons directly after the presentation of the second face. In nontarget trials, the presentations and the tasks were identical to those of Study 3a. There were different buttons allocated to the attentional and to the memory and color discrimination tasks, two buttons for the attentional task, other two for the memory and color discrimination tasks.

We applied the same methods to analyze the behavioral results as in Study 3a. The target trials were excluded from any further analysis of the BOLD signal (Summerfield et al, 2008).

The peak BOLD signal of the right FFA (rFFA) was analyzed by a repeated measurements ANOVA with the factors of VSTM load (VSTM vs. Control) and trial type (alternated vs. repeated) for the nontarget trials, as in Study 3a.

The rFFA was localized identically as in Study 3a (one participant was excluded from the final analysis because this area could not be identified reliably). The mean MNI coordinates (\pm SD) for the rFFA of this experiment were 42.2 (3.93), -54.8 (8.12), -16.5 (4.18).

Results

Behavioral results.

The behavioral results are presented in Figure 1, separately for the

non-target (A, B) and target trials (C, D). In nontarget trials, a significant main effect of VSTM load was found for performance ($F(1, 19) = 676.991, p < 0.001, \eta_p^2 = 0.973$) and reaction times ($F(1, 19) = 86.441, p < 0.001, \eta_p^2 = 0.820$). On average, performance in the Control tasks was better ($96.0 \pm 1.4\%$) and reaction times were shorter ($M = 823 \pm 44$ ms) than in the VSTM tasks (performance: $M = 57.8 \pm 1.6\%$; reaction time: $M = 1137 \pm 64$ ms).

However, there was no significant main effect or interaction in the target trials (all $ps > 0.136$). This suggests that, participants were allocating attention to the face stimuli equally in both conditions, when instructed to detect inverted faces.

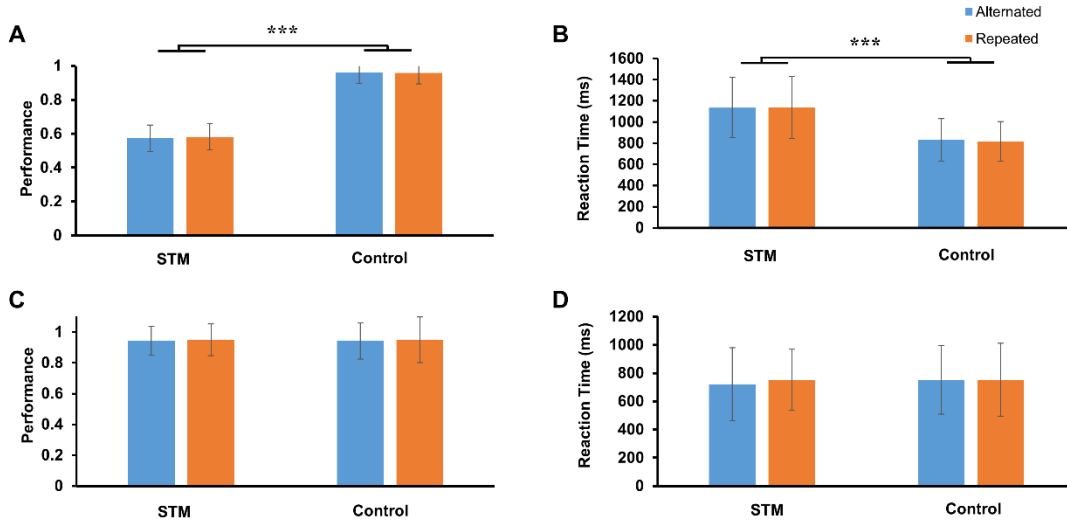


Figure 1. Performance (A, C) and reaction times (B, D) of Experiment 3b for different visual short-term memory conditions (VSTM and Control) separately for the non-target trials (A, B) and for the attentional target trials (C, D). Error bars represent standard errors of means. Blue and orange bars represent alternated and repeated trials, respectively. *** $p < .001$.

Repetition suppression modulation in the rFFA.

The results of the BOLD measurements of the rFFA are presented in Figure 2. We found a significant RS effect (main effect of trial-type: $F(1, 19) = 7.992, p = 0.011, \eta_p^2 = 0.296$) in the sense that the BOLD response is stronger in the

alternation (0.454 ± 0.065) than in the repetition trials (0.392 ± 0.071). Conversely, both the main effect of VSTM-load ($F(1, 19) = 0.091$, $p = 0.766$, $\eta_p^2 = 0.005$) and the VSTM-load x trial-type interaction ($F(1, 19) = 0.952$, $p = 0.342$, $\eta_p^2 = 0.048$) were not significant. This suggests that VSTM load has no effect on RS when participants' attention is allocated to the stimuli.

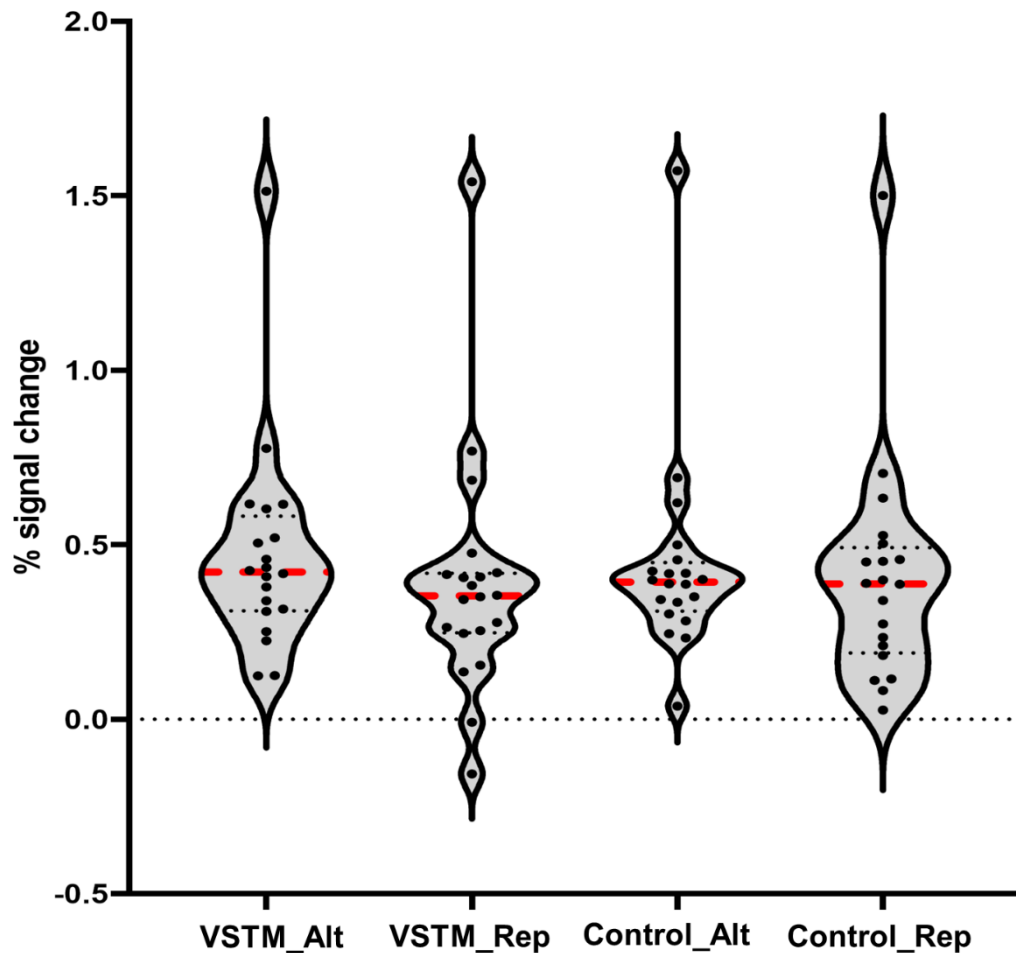


Figure 2. The BOLD signal of the right FFA for the VSTM and Control conditions, separately for alternating and repeated trials when attention is controlled. Dots represent individual data points; red dotted lines show the group median. Shaded areas depict the distributions of data for each condition.

Discussion

This control experiment was designed to shed more light on the mechanisms

of VSTM modulation of RS that we demonstrated in *Study 3a*. Specifically, we aimed at clarifying whether the attenuated RS is rather due to lower attentional resources to faces or to the heavily loaded VSTM mechanisms. The results of this control experiment demonstrated that the attenuated RS effect under VSTM load vanishes if participants had to attend to the face stimuli actively, indicating that the RS modulation, we found in *Study 3a* is likely due to the differential allocation of attentional resources between the two conditions.

Previous studies showed that there is, indeed, a close link between attention and VSTM (for a review see Awh and Jonides, 2001; Awh et al., 2006). For instance, it has been suggested that attention plays an important role in the maintenance of internal information over a brief period of time (e.g., Awh et al., 1998; Trapp and Lepsien, 2012). It has even been proposed that attention and VSTM should no longer be considered as separate systems, or they rely on the same limited resources (Kiyonaga and Egner, 2013). Crucially, Larsson and Smith (2012) showed that attention is also important for observing expectation-related RS. When participants' attention was diverted away from the stimuli, a large, expectation-related component of the response reduction disappeared. Thus, it is possible that the RS effect was no longer present because participants were not paying attention to faces when the load on VSTM was high.

However, to further investigate the influence of attention and VSTM on expectation-related RS effect separately, a completely orthogonal design between these two factors should be tested. For instance, an additional control manipulation, participants are asked to divert their attention from face stimuli in both VSTM and Control conditions, should be conducted. If the RS effect diminishes in both conditions, it could mean that the attenuated RS in study 3a is completely due to attention. Conversely, if the interaction effect (i.e., as Study 3a, the RS attenuation only under VSTM load condition) is replicated, it could mean that it is due to the limitation of short-term memory capacity.

Additionally, besides the suggested manipulations, future experiments can

combine multiple techniques. For example, recording the eye movements with eye-tracking, or separating attention and expectation using high temporal resolution methods (e.g., EEG, MEG). Since previous studies have found the exogenous attention effect precedes those of endogenous attention (Hickey et al., 2010), testing these interactions with these techniques can be a promising research avenue.

Acknowledgments

This work was supported by a Deutsche Forschungsgemeinschaft Grant given to G.K. [KO3918/5-1] and S.T. (TR 1543/2-1). Chenglin Li was supported by the China Scholarship Council (CSC) scholarship (201808330399) during the study. We thank Marie Wächter for helping with participants' recruitment, and Sophie-Marie Rostalski for guidance in data analysis, and Linda Ficco for her comments.

References

- Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, 5(3), 119-126. [https://doi.org/10.1016/S1364-6613\(00\)01593-X](https://doi.org/10.1016/S1364-6613(00)01593-X)
- Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology-Human Perception and Performance*, 24(3), 780-790. <https://doi.org/10.1037/0096-1523.24.3.780>
- Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience*, 139(1), 201-208. <https://doi.org/10.1016/j.neuroscience.2005.08.023>
- Hickey, C., Van Zoest, W., & Theeuwes, J. (2010). The time course of exogenous and endogenous control of covert attention. *Experimental brain research*, 201(4), 789-796. <https://doi.org/10.1007/s00221-009-2094-9>

Empirical contributions

- Kiyonaga, A. & Egner, T. (2013). Working memory as internal attention: toward an integrative account of internal and external selection processes. *Psychonomic Bulletin and Review*, 20:228-242.
<https://doi.org/10.3758/s13423-012-0359-y>
- Larsson, J., & Smith, A. T. (2012). fMRI repetition suppression: neuronal adaptation or stimulus expectation?. *Cerebral Cortex*, 22(3), 567-576.
<https://doi.org/10.1093/cercor/bhr119>
- Li, C., Kovács, G., & Trapp, S. (2021). Visual short-term memory load modulates repetition related fMRI signal adaptation. *Biological Psychology*, 166, 108199. <https://doi.org/10.1016/j.biopsycho.2021.108199>
- Summerfield C., Trittschuh E. H., Monti J. M., Mesulam M. M., Egner T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*. 11 1004–1006. <https://doi.org/10.1038/nn.2163>
- Trapp, S., & Lepsien, J. (2012). Attentional orienting to mnemonic representations: Reduction of load-sensitive maintenance-related activity in the intraparietal sulcus. *Neuropsychologia*, 50(12), 2805-2811.
<https://doi.org/10.1016/j.neuropsychologia.2012.08.003>

4 Discussion

4.1 Summary of the current studies

The aim of the current dissertation is to elucidate whether and how prior experiences modulates neural correlates of predictive processing in the ventral visual areas in humans. The definition of “prior experiences” in my work specifically refers to both the long-term experience of language learning and to the short-term experience achieved by perceptual learning. Three studies were designed to address these issues. Two well-established neural effects (P(rep) and RS effects) are elicited by presenting participants with different visual stimuli (faces, words, and objects) and measured as an index of predictive processing with fMRI. The first two studies investigated the P(rep) effect to elucidate the importance of prior experiences on predictive processing, while the third one attempted to explain how this modulation could be conducted in the brain. As all studies have been discussed in the related manuscripts respectively, this subchapter will focus on the overarching results, their relations to each other and the conclusions we can draw from it.

The first study (Study I, chapter 3.1) aimed to investigate the effect of prior experiences on predictive processing by testing the repetition probability (P(rep)) effects for words. Grotheer and Kovács (2014) found this effect for familiar Roman letters in the Letter Form Area (LFA; Thesen et al., 2012), but not for novel false fonts. This suggests that this P(rep) effect depends on prior experiences of the participants with the stimuli. In study I, we further tested whether this conclusion can be extended to more complex lexical words. We adapted a block-based probability design (see figure 3D; Summerfield et al., 2008) to measure the BOLD signal in the VWFA of native Chinese and German participants and to estimate the P(rep) effects for Chinese characters and German words. The results showed a significant P(rep) effect for stimuli of the mother tongue in both Chinese and German participants, which suggests that the P(rep) effect can be extended to words in the VWFA. Interestingly, Chinese participants, learning German as a second language, also show a

significant P(rep) effect for German words while German participants who had no prior experiences with Chinese characters showed no such effect. This further suggests that the P(rep) effect only emerges for words of a language with which participants are highly familiar. Thus, this study demonstrates the idea that predictive processes require extensive prior experiences with perceived stimuli.

In contrast to extensive, semantic prior experiences, Study II (chapter 3.2) aimed to investigate the modulation of short-term experience on predictive processing by testing the P(rep) effect for non-face objects. As mentioned above, the P(rep) effect was robustly replicated for faces, words, and Roman letters, which participants had extensive experiences with (e.g., Summerfield et al., 2008; Grotheer and Kovács, 2014; Li and Kovács, 2022), but not for novel false fonts and unfamiliar words (Grotheer and Kovács, 2014; Li and Kovács, 2022). However, inconsistent findings were observed for objects. For instance, Mayrhauser et al. (2014) reported the P(rep) effect for line-drawn objects, but such effects were not replicated for everyday objects or chairs in human neuroimaging studies (Kovács et al., 2013) and for 3D artificial faces, fractal patterns and everyday objects in macaque single-cell recording studies (Kaliukhovich and Vogels, 2011; Vinken et al., 2018). To explain this inconsistency from the point of view of prior experiences, we measured the P(pre) effects for cars (a familiar category of everyday objects) pre and post a ten-day perceptual learning (PL) session. The results showed that the P(rep) effect is significantly attenuated after short-term PL, which suggests that short-term experience could temporarily modulate the P(rep) effect as well.

The results of the first two studies suggest that the predictive processes, measured via the P(rep) effect, are modulated by both long-term and short-term prior experiences. However, it remains largely unclear how exactly prior experiences modulate predictive processes. Especially, although feedback loops are thought to play a role at the level of single neuronal connections, it is still unclear which cognitive and psychological processes modulate the effect of prior experiences in predictive processing. To address this question, the third study (Study III, chapter 3.3) attempted to investigate the degree

to which two potential cognitive factors (short-term memory and attention) could be the modulators of this process. Study III used a classic stimulus repetition paradigm to measure RS, together with either concurrent VSTM load or a control condition. Attention was manipulated by an active detection task in Study 3b. The results of Study 3a showed that RS was significantly attenuated when VSTM is loaded with other non-face visual information. Study 3b further showed that this attenuated RS effect could be counteracted by active attention allocation. Therefore, these results suggest that the modulation of prior experiences on predictive processes is dependent on short-term memory and attention.

In summary, the importance of prior experiences on predictive processing was demonstrated in Studies I and II: Long-term experience with the stimuli is the prerequisite of the appearance of the predictive process while short-term visual experience could temporarily modulate predictive processes as well. Study III further revealed that this modulation is related to two important cognitive processes: short-term memory and attention. These results will be discussed below in the relationship between RS, ES and experience (chapter 4.2), a potential model about the relationship between experience and P(rep) effect (chapter 4.3), and a framework of a preliminary, cognitive predictive coding (chapter 4.4).

4.2 The role of experience in repetition and expectation suppression

According to the theory of PC, the process of matching between top-down predictions and bottom-up sensory inputs is crucial (Friston, 2005; Kok and de Lange, 2015). Prior experiences about sensory inputs in our brain is necessary to form predictions. These prior experiences could include both implicit statistical regularities learnt during an experiment and the cumulative, long-term experience with the stimulus category itself. The present thesis focused more on the latter, and investigated the influence of prior experiences on predictive processes, estimated by P(rep) modulations of RS. Based on the results of previous studies and the current thesis, we propose the different influences of

Discussion

experience in RS and ES: RS is experience-independent, but ES is experience-dependent.

Conceptually, RS indicates the difference of the neural response between repeated and alternated or even the first presentation of a given stimulus (Desimone, 1996). Instead, ES identifies the difference between expected and unexpected or neutral stimuli (for a review see Feuerriegel et al., 2021b). There are various aspects reasons for supporting the differences between RS and ES. For instance, non-primates single-cell recording studies failed to find prediction-related effects for unfamiliar objects (Kaliuhovich and Vogels, 2011; Vinken et al., 2018), but RS was still found in these studies. In human studies, RS appears in an earlier time window, typically starting between 100 – 250 ms after stimulus onset in the visual modality (Henson et al., 2004; Schendan and Kutas, 2003; Schweinberger et al., 2004; Schweinberger and Neumann, 2016) and between 40 - 60 ms in the auditory modality (Todorovic and de Lange, 2012), whereas ES appears later (300ms after stimulus onset and between 100-200 ms in the visual and auditory modalities, respectively)(Summerfield et al., 2011; Todorovic and de Lange, 2012). Grotheer and Kovács (2015) confirmed the separation between RS and ES using fMRI, and speculated that the slight temporal delay of ES as compared to RS might reflect that ES require top-down expectations from the frontal cortex (Brodmann Area 7). Therefore, Grotheer and Kovács (2016) proposed a two-stage model of response suppression, which suggests that RS is the expression of ‘low-level’ prediction error (ϵ) calculations compared to ES, which ES likely represents “higher-level” ϵ computations with expectation signals from frontal regions (Summerfield et al., 2006). In light of this separation, I will now examine the effects of prior experiences on RS and ES separately.

When considering the influence of experience, the RS effect could appear regardless of experience with stimuli. For instance, a stronger RS effect in non-primate single-cell recordings studies when monkeys had no extensive experience with stimuli (e.g., Kaliukhovich and Vogels, 2011; Vinken et al., 2018), and in human neuroimaging studies, e.g., when participants view novel false fonts (Grotheer and Kovács, 2014). However, the results of ES and P(rep) (a kind of ES effect as described in chapter 1.2.1) effects for

various stimulus categories suggest that ES is experience-dependent. For instance, Meyer and Olson (2011) trained macaque monkeys to learn six pairs of images, in which the second image can be predicted based on the first image in each pair. The ES effect for these trained stimuli was obtained in the IT cortex, where neuronal activity was reduced when the trailing image was predicted, as compared to when it was unpredicted. Grotheer and Kovács (2014) observed the P(rep) effect for familiar Roman letters, but not for novel false fonts. The first study in the present thesis adopted the identical paradigm to compare this effect for two kinds of words (Chinese characters and German words) in two groups of participants (Chinese and German), the results showed that P(rep) effects were observed only for words of a language with which participants had extensive experiences. These results further suggest that prior experiences with stimuli is important for observing a prediction-related effect – the P(rep) effect. It is worth pointing out that this conclusion - ES is experience-dependent - is only demonstrated in the monkey training study (Meyer and Olson, 2011) by using block-based probability designs (Summerfield et al., 2008) in human studies (Grotheer and Kovács, 2014; Li and Kovács, 2022). For other designs, I would hypothesize that prior experience is also important for the occurrence of ES. In other designs, the content of this prior experience can also not be just the occurrence of the stimulus category itself, but rather other expectation-related experiences. For instance, in statistical learning designs, prior experiences might refer to the experience of sequence, because the expected sequence was actually learned during the training session (Manahova et al., 2018); In cue-based designs, prior experiences could be the relationship between cues and stimuli. The predictions of related cues for subsequent stimuli were actively learned during the stimulus presentation (Egner et al., 2010), while the ES was dependent on these expectations. In addition, Vogels (2016) reviewed the results of response suppression in macaque IT and suggested that the RS effect could be the expression of synaptic depression or inherited adaptation effects, which implies that RS could be just a local change, independent of the participation of top-down experience. Overall, these experiments addressed the question of whether prior experiences modulate predictive processes. Taking their results together, it can be

concluded that prior experience is necessary for predictive processes, but not for the RS effect to occur.

4.3 The relationship between experience and repetition probability effect

The last subchapter discussed the different influences of experience on RS and ES, and suggested that ES is experience-dependent, but RS is experience-independent. This subchapter especially focuses on the relationship between experience and the repetition probability $P(\text{rep})$ effect. Several reasons prompted the discussion of their relationship in this section: First, the $P(\text{rep})$ effect is the main indicator of predictive processes in the present thesis; second, this effect was widely investigated for various stimulus categories in humans and non-primates. Third, inconsistent results regarding this effect could be explained from the perspective of prior experiences. Thus, this subchapter will review the studies of the $P(\text{rep})$ effect and attempt to explain the inconsistent results from the point of view of experience. Afterwards, I will attempt to fill in the gaps, connect the evidence at hand, and propose a preliminary model about the relationship between experience and the $P(\text{rep})$ effect.

The modulation of $P(\text{rep})$ on RS effects has been described in the introduction part (chapter 1.2.1). This effect is considered as a robust and classical indicator to explain the RS effect by prediction coding. The corresponding paradigm is called block-based probability design (see figure 3D on page 14), which manipulates the probability of repetition and alternation trials in different blocks. A stronger RS effect observed in repetition blocks, the probability of repetition trials is higher when compared to alternation blocks, indicates the modulations of top-down expectation on RS, which is known as the $P(\text{rep})$ effect. This effect was firstly reported in a human fMRI study using human faces as stimuli. It was replicated for face stimuli in independent laboratories (e.g., Kovács et al., 2012; Larsson and Smith, 2012). However, it failed to appear in non-primates single-cell recordings studies, neither using fractal patterns and natural

objects (including human and animal faces, human and animal bodies, animals, indoor and outdoor environments, natural landscapes, and manmade objects) as stimuli (Kaliuhovich and Vogels, 2011), nor using 3D artificial faces generated by FaceGEN Modeler as stimuli (Vinken et al., 2018). To exclude the difference of species and tasks, Kovács and colleagues (2013) found no P(rep) effect for everyday objects and chairs in the LO in human fMRI studies as well. This suggested that P(rep) effects might change across the stimulus categories and this variance could be due to differences in selectivity for faces and objects in face and object regions respectively (Kovács and Vogels, 2014). Subsequently, Grotheer and Kovács (2014) used the Roman letters and novel false fonts as stimuli to investigate the P(rep) effect, their results excluded this difference between faces and objects areas. Because they found the P(rep) effect for Roman letters in the LFA and LO, but not for novel false fonts. their results suggest that the P(rep) effect depend on prior experiences with stimuli. This conclusion was confirmed in the first study of the present thesis. We only observed the P(rep) effect in the VWFA and LO for words of a language with which participants had extensive experience. For other objects, Study II in the present thesis found the P(rep) effect in the right LO for highly familiar objects (cars) as well, which could explain the fact that no P(rep) effect was found in Kovács et al. (2013) study. A hypothesis could be that participants had various degrees of prior experiences with everyday objects. Alternatively, the prior experiences they had with stimuli (chairs) was not enough to induce the P(rep) effect. it is worth pointing out that there is a specific case, where Mayrhauser et al. (2014) found the P(rep) effect for line-drawing of objects in the left LO. On the one hand, we may need to receive more relevant experience from some higher-level regions when we recognize line drawing of objects compared to when we recognize objects based on the real images. Receiving more experience for line drawing of objects from higher-level regions still can be explained under the framework of the influence of prior experiences on predictive processes; on the other hand, based on their results that the P(rep) effect was only observed in the left LO, one possibility was that participants were verbalizing the abstract line drawing of objects from some language areas in the left hemisphere when they view the images (Kovács and Vogels, 2014). If so,

Discussion

language information/knowledge could be considered as prior experiences as well. Taken together, all findings of P(pre) effects could be explained from the perspective of prior experiences with stimulus categories. Thus, I attempt to propose a preliminary model of the relationship between experience and the P(rep) effect in the following.

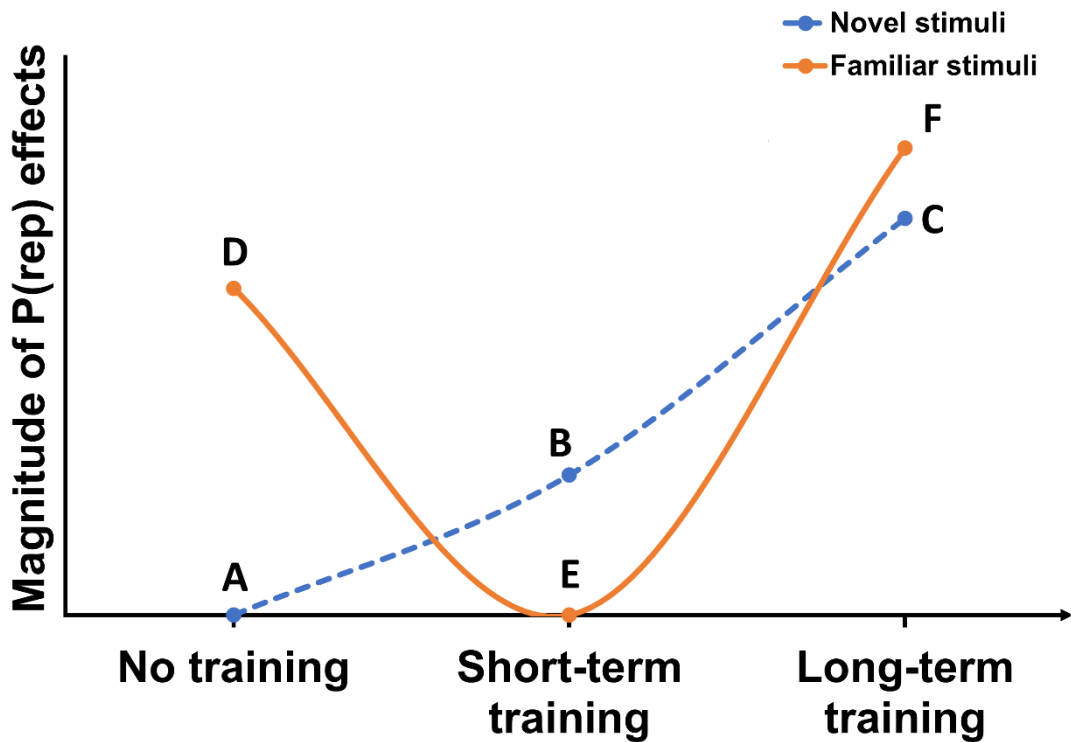


Figure 9. Schematic illustration of the preliminary model of the relationship between experience and P(rep) effect. The horizontal axis shows three kinds of experiences based on the length of training. The vertical axis indicates the magnitude of P (rep) effects. The blue dotted line indicates the magnitude of P(rep) effect for novel stimuli with different levels of experience. The orange line indicates the magnitude of P(rep) effect for familiar stimuli with different levels of experience. The letters of A - F indicate different conditions.

To integrate the above-mentioned results of the current thesis in this model (Figure 9 on page 110) and relate them to the second study, I separate the perceived stimuli into two categories based on participants' experience (Novel and Familiar stimuli). At the same time, I simply divide the prior experiences with stimuli into three levels based on how

long the perceptual training on these stimuli has been (No training, short-term training, long-term training/learning). Although not all points have been demonstrated by experiments, I would like to list the related evidence or potential proposals for each point. First, for novel stimuli, there seems to be no P(rep) effect when we have not had got any training (point A), this is supported by the results of novel false fonts in Grotheer and Kovács's (2014) study and the results of German participants for Chinese characters in Study I (Li and Kovács, 2022). Second, when participants have extensive experience with a novel stimulus category (point C), the P(rep) effect would be present, which has been supported by the results of Roman letters (Grotheer and Kovács, 2014), words of the mother tongue, and words of the second language (Study I; Li and Kovács, 2022). However, the magnitude of P(rep) effects when participants just undergo a short-term training for a novel stimulus category remains unclear (point B). In this model, I assume that it will be between A and C. This hypothesis should be tested in future experiments.

Furthermore, for familiar stimuli, with which participants had extensive experience, a strong P(rep) effect would be obtained (point D). This has been robustly demonstrated for faces, Roman letters, words, and cars (e.g., Grotheer and Kovács, 2014; Li and Kovács, 2022; Summerfield et al., 2008; study II in the present thesis). This P(rep) effect temporarily disappears when participants perform short-term training for this familiar stimulus category (point E). The data of Study II in the present thesis could support this hypothesis. However, it has not been explored whether the P(rep) effect could be further enhanced for familiar stimuli. In this model, I assume it could be enhanced after long-term training (point F) which should be tested further. It is worth pointing out that this enhancement is not unlimited. For example, the P(rep) effect showed no difference between Chinese characters and German words in Chinese participants (Study I), which means that the P(rep) effect would not increase indefinitely when prior experiences with the stimuli accumulate to a certain extent.

This preliminary model could summarize the outcome of all studies investigated the P(rep) effect and relate the magnitude of the P(rep) effect with the amount and kind of

Discussion

prior. However, two key questions should be investigated to reveal the underlying mechanism between prior experiences and the P(rep) effect in future studies.

First, why does the P(rep) effect for familiar objects disappear after a short-term training (Point E related to the results of Study II)? Since previous studies suggested the P(rep) effect depends on prior experiences with the stimuli (Grotheer and Kovács, 2014; Li and Kovács, 2022), we might assume that, with the accumulation of experience with experimental stimuli, a stronger P(rep) effect would appear, and remain stable after its emergence. However, the results of Study II showed that the P(rep) effect for familiar objects could be temporarily abolished after a short-term perceptual training, even for untrained stimuli (faces). Two potential reasons could explain this attenuation of P(rep) effects for familiar stimuli. i) for the trained stimuli, it may be due to the general training effect, which refers to the reduction of BOLD signal responses for the trained compared to novel/untrained stimuli demonstrated in perceptual training studies (e.g., Baek and Op de Beeck, 2010; Henson et al., 2000; Reber et al., 1998; Sigman et al., 2005). Thus, the absence of P(rep) effect for trained familiar stimuli (cars) in Study II could be due to the overall reduction of neural response for the trained stimulus category. In other words, this overall reduction of the BOLD signal responses for trained stimuli results in the difference between repetition and alternation trials in different blocks being too small to be detected. ii) for the untrained stimuli, it could be due to the top-down expectations changed after the training session. Previous studies adopting cue-based designs showed that the expectation effect for familiar stimuli (faces and houses) could be changed under different conditions with different expectation probabilities (Egner et al., 2010). This result indirectly supports the absence of the P(rep) effect for untrained familiar stimuli. In Study II, in the pre-training sessions, the face and car runs are equally predictable for participants as there was no cue that can induce more expectations to either face or car runs. However, the expectations of the car runs would increase after the training sessions as the training task was only related to cars. Thus, according to the PC theory, the overall BOLD signals of the car runs would be decreased (expectation suppression) in the post-training sessions, as compared with the results of pre-training sessions. However,

the overall BOLD signals of the face runs would be increased (surprise enhancement). This hypothesis was confirmed by the results in Study II. Please note the increased BOLD signals of the face runs in the post-training sessions was only a trend, not statistically significant. Thus, future studies should attempt to replicate this result and investigate the effects of expectations that are mainly the product of task design.

Second, the definition and classification of experience are rough. They should be refined in future studies. For instance, important questions are how much experience is enough to result in the emergence of the P(rep) effect, and which kinds of experience are more important for the P(rep) effect. For more details, please see the first point of subchapter 4.6.

Overall, this model could describe the relationship between experience and the P(rep) effect, but it still needs to be proved and improved. I will discuss it further in chapter 4.6.

4.4 The modulation of top-down prediction for repetition suppression

The previous sections have discussed the crucial role of prior experiences in predictive coding. However, it remains largely unknown how this modulation works in the brain. Study III attempted to answer this question based on the mechanism of short-term memory (STM) and attention. According to the results of study III, we tried to integrate the memory and attention processes into the framework of predictive coding (PC) (Figure 10 on page 114). We proposed that extensive prior experiences with upcoming events is stored in long-term memory. This information in long-term memory can be temporarily extracted to short-term memory, to then form prediction in a highly accessible state. This hypothesis is consistent with the idea that STM is regarded as a part of long-term memory that is engaged by the current task (e.g., Anderson et al., 2004; Cowan, 1998; Rochkin et al., 2003). In addition, Study 3a in the present thesis and several previous studies provided the empirical evidence that STM is required for perceptual prediction processes (e.g., Cashdollar et al., 2017; St John-Saaltink et al., 2015; Travis et al., 2013; details see chapter 1.3.4 on page 28). Under this assumption, on the one hand, it can explain why the

RS effect was significantly attenuated under VSTM load in Study 3a: the capacity of STM was loaded with additional information, which partially impaired the estimation of predictions. On the other hand, it can explain why extensive prior experiences is important for prediction: prior experiences in the long-term memory can provide sufficient information that is then temporarily store in STM to calculate predictions for the current predictive task (Trapp et al., 2021).

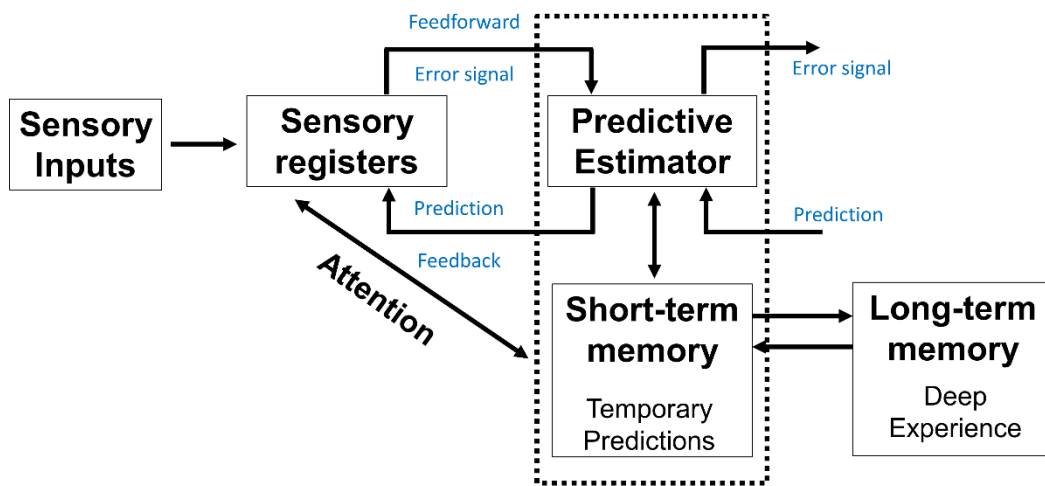


Figure 10. Schematic illustration of a hierarchical network for predictive coding with memory and attention. The hierarchical network of predictive coding is identical to Rao and Ballard (1999). Deep experience is stored in long-term memory, and can be temporarily extracted to STM, to then form predictions in a highly accessible state. Attention as a modulator facilitates prediction error coding in the predictive estimator or affects a stimulus saliency during bottom-up processing.

The results of Study 3b showed that, RS was restored in the VSTM condition when participants actively attended to stimuli. This result dovetails with a wealth of research findings suggesting the crucial role of attention in prediction. It has been well established that attention increases the efficiency of information processing. Especially, attending to a stimulus is associated with increased neuronal population selectivity (e.g., Maunsell, 2015; Murray and Wojciulik, 2004; Serences and Kastner, 2014; Treue, 2003). Importantly, several studies have demonstrated the crucial role of attention in predictive processing

(e.g., Alink and Blank, 2021; Larsson and Smith, 2012; Richter and de Lange, 2019; more details have been described in chapter 1.3.3 on page 26). Three potential explanations about the modulation of attention in predictive processing were proposed: First, attention facilitates prediction error coding (Feldman and Friston, 2010; Friston, 2009), meaning that attention can enlarge the prediction error effect for unexpected relative to expected conditions. Second, the expectation effect in the attended condition is due to an enhanced stimulus saliency of unexpected stimuli (for a review see Alink and Blank, 2021). Third, prediction might provide an anchor for attentional processing, meaning that the subsequent attention orientation could depend on the process of computing predictions (Hsu et al., 2014 cited in Ficco et al., 2021). Therefore, in our refined framework of PC, we assumed the main role of attention is that of a modulator, which means it both facilitates prediction error coding in the predictive estimator (top-down) or it produces an increased stimulus saliency effect (bottom-up). This can explain why the RS effect appeared in the VSTM condition when participants attended to the face stimuli in Study 3b. In other words, when STM is loaded with other information, attention still can be a potential driver for the expectation effects. After all, our brain can flexibly perform complex tasks through the cooperation of multiple networks. It is worth pointing out that there is a great consensus that attention is a necessary prerequisite for STM activity. Especially, it has been suggested that attention serves as a rehearsal mechanism in visuo-spatial working memory (Awh and Jonides, 2001). In other words, attention contributes to the maintenance of information over a short time (Awh et al., 1998; Awh et al., 2006). Some studies have even proposed that visual STM and attention should no longer be considered as separate systems (e.g., Kiyonaga and Egner, 2013). Therefore, the close link between attention and STM could be an important reason to integrate attention and STM into the PC theory. Of course, future studies should test the contribution of attention and STM on PC respectively.

In summary, based on the findings of Study III and related studies, we proposed that STM and attention may be two important modulators of the effect of prior experiences on predictive processes. However, they play different roles: STM is mainly responsible for

forming predictions based on prior experiences; attention mainly facilitates prediction error coding or reflects the stimulus saliency effect.

4.5 Limitations of the present thesis

As the limitations of each study have been pointed out in the respective manuscripts, this subchapter will focus on the overarching limitations in the present thesis and how to avoid them in future studies.

First, a general limitation is the presence of bias in the sampling process. Well-known concerns in experimental research are that of achieving sufficient power, and that to render a proper randomized selection process. If the sampling process is biased in some way, the conclusions will inevitably suffer from it. In all studies of the present thesis, we randomly sampled from undergraduate and graduate students, thus our results would mainly represent the performance of undergraduate and graduate students. If one plans to extend the present conclusions to other populations, it will be necessary to diversify the sampling process, and make it more inclusive. Another issue is that of sample size. The best approach to decide the needed sample size reliably in a power analysis is running a pilot experiment. Another option is to calculate it based on the effect size of previous studies, such as using G* Power (Faul et al., 2007). However, due to limited resources, this estimation is not always used for fMRI experiments. Especially, due to the COVID-19 pandemic situation, large sample sizes are more difficult to achieve. In our studies, we decided to take the sample size of published studies, that used the same paradigm and investigated the same phenomena, as a reference. Based on the effect size and statistical power of the results, our conclusions were built on the higher effect size. In addition, in Study 3a, we performed a Bayesian ANOVA, besides a traditional frequentist ANOVA. The results of the Bayes factor also support our conclusion. Of course, in future studies, we should recruit more participants to ensure the results are more representative.

Second, my work also presents some methodological limitations. In all studies, we

Discussion

measured the BOLD signal with fMRI because such technique has obvious advantages, especially, high spatial resolution and non-invasiveness. However, one disadvantage is the lower temporal resolution when compared to other techniques, such as EEG and MEG. Although the fMRI can be used to distinguish RS and ES (Grotheer and Kovács, 2015) as well, it would improve these conclusions by using a high temporal resolution technique (Todorovic and de Lange, 2012). This would shed light on the temporal characteristics of RS and ES. With the technological development of the fMRI technique, its temporal resolution can be improved, meaning the timing parameters (i.e., TR) of fMRI can be reduced. For example, we use a shorter TR in Study 3a (TR = 1000 ms). In some studies, it can even be shortened to approximately 500 ms with multi-slice acquisition at the same timepoint (Xu et al., 2013). In addition, our studies mainly focus on the difference in brain activities in different conditions.

Another potential limitation is that in our studies, we did not correlate the behavioral results with BOLD signal responses. For instance, the task in Studies I and II just ensured participants attended to stimuli, and the brain responses of these responded trials were not analyzed further. From the perspective of Psychological Science, it would be necessary to connect brain activity with participants' behavior. Future studies may employ tasks that explicitly require participants to anticipate future stimuli. Finally, in regards to the data analysis, we only conducted ROI analysis for all studies, and the whole-brain analysis for Study I. Other methods should be considered as well, such as, multivariate pattern analysis (Haxby et al., 2001). Therefore, in the future studies, we should not only focus on the univariate changes of BOLD signals during the modulation of prior experiences on predictive processes, but also pay attention to its temporal dynamics, activity patterns, and attempt to connect these brain activities to behavioral results by using multiple analysis methods.

Third, another limitation could be the limited range of paradigms we used. There are various designs to demonstrate that predictive processes exist in our brain (for a review see Feuerriegel et al., 2021b). However, for better comparison to previous studies, our

studies in the present thesis mainly adopted block-based probability designs. Although the results of these studies can address our research questions, which is whether prior experiences are important for the predictive process, we still have to further test this conclusion in other paradigms. In this way, it could be better understood how prior experiences modulate predictive processes in different paradigms.

4.6 Open questions and future directions

The aim of the present thesis is to investigate whether and how prior experiences modulate predictive processing in the ventral visual areas. Although the conclusion supports the idea that prior experience is a prerequisite for predictive processing, and this process could be achieved through the modulation of short-term memory and attention, I will propose some ideas for follow-up studies based on the findings of the present thesis.

First, the minimum amount of experience necessary to observe these effects should be quantified. According to the introduction in chapter 1.3.1, we know that ‘experience’ is a complicated concept. The main reason is that it is simply not easy to quantify experience, and there is no standard quantifiable measurement of experience. In previous and present studies only a simple and rough classification of experience is made, such as no experience, short-term experience, and long-term experience. In our preliminary model in chapter 4.3, we only can classify experience into three levels. If we will be able to quantify experience in a more fine-grained manner, this model will be refined correspondingly. Another important issue related to experience is that the *quality* of experience matters too. For instance, participants being exposed to several examples of cars might experience the stimulus ‘car’ differently in the end, depending on task instructions, e.g. identification of car brand, categorization by color, inferences about the car owner, typicality ratings, etc. Accordingly, expectation effects might vary depending on qualitative aspects. Therefore, future studies can comprehensively consider various qualitative factors affecting experience acquisition, especially task instructions and

subjective reports of participants.

Second, a promising avenue for future research is offered by longitudinal training. Training is a dependable method to obtain experience. According to the conclusion of the present thesis, prior experiences with the stimuli are necessary for predictive processing. Especially, Study I provided empirical evidence to support that the words could be a reliable exemplar in the training experiments. For instance, one could manipulate how a new language is learnt. Xue et al. (2006) designed a logographic artificial language for Korean characters to conduct visual form, phonological, and semantic training in Chinese participants, who have no experience with Korean characters. The results showed a U-shaped change of the brain responses in the VWFA. The modulation of prior experiences on predictive processes will be further clarified if future studies will conduct a long-term longitudinal training using words as stimuli. This would be especially important to investigate how different experiences modulate top-down predictive processes.

Last but not least, future studies could investigate the presence of a dynamic prediction network. The brain is a dynamic, hierarchical predictive machine (for a review see Badcock et al., 2019). When our eyes catch an object, its processing will rapidly involve many brain areas. However, this thesis and previous studies mainly focused on a few areas and ignored the connection between each area. Future studies should integrate various high-resolution technologies to record brain activity and behavioral responses during task performance. In particular, when using faces as stimuli, researchers should not only focus on face-sensitive areas, but also consider the role of other networks such as attentional network and memory networks.

Discussion

5. Conclusion

The current dissertation is aimed to elucidate whether and how prior experiences modulates neural correlates of predictive processing in the ventral visual areas in humans. We specifically focus on prediction-related neural effects (P(rep) and RS effects) for different visual stimulus categories (faces, words, and objects) measured with fMRI. Here we mainly investigated the influence of two kinds of experiences (long-term experience from language learning and short-term experience from perceptual learning) on the P(rep) effect (Studies I and II). Furthermore, we investigated the crucial role of STM and attention on prediction-related effects to explain how prior experiences modulates neural effects of predictive processing (Study III).

The main findings of the three studies presented in this dissertation are the followings. First, the P(rep) effect is only manifest for words of a language with which participants had extensive prior experiences (Study I). Second, the P(rep) effect for familiar objects can be temporarily abolished by short-term perceptual learning (Study II). Finally, the attenuation/emergence of RS could be modulated by STM and attention (Study III). Summarizing these results leads to the following conclusion: the predictive processes, as measured by the P(rep) effect, require extensive prior experience with stimulus categories, but these can also be modulated by the short-term learning experience. More importantly, STM and attention are considered are two important modulators of prior experiences on predictive processes.

References

- Alink, A., & Blank, H. (2021). Can expectation suppression be explained by reduced attention to predictable stimuli?. *NeuroImage*, 231, 117824. <https://doi.org/10.1016/j.neuroimage.2021.117824>
- Alink, A., Abdulrahman, H., & Henson, R. N. (2018). Forward models demonstrate that repetition suppression is best modelled by local neural scaling. *Nature communications*, 9(1), 1-10. <https://doi.org/10.1038/s41467-018-05957-0>
- Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience*, 30(8), 2960-2966. <https://doi.org/10.1523/JNEUROSCI.3730-10.2010>
- Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in cognitive sciences*, 4(7), 267-278. [https://doi.org/10.1016/s1364-6613\(00\)01501-1](https://doi.org/10.1016/s1364-6613(00)01501-1)
- Amado, C. P. (2017). *Prediction related phenomena of visual perception* (Doctoral dissertation, Friedrich-Schiller-Universität Jena).
- Ambrus, G. G., Dotzer, M., Schweinberger, S. R., & Kovács, G. (2017a). The occipital face area is causally involved in the formation of identity-specific face representations. *Brain Structure and Function*, 222(9), 4271-4282. <https://doi.org/10.1007/s00429-017-1467-2>
- Ambrus, G. G., Windel, F., Burton, A. M., & Kovács, G. (2017b). Causal evidence of the involvement of the right occipital face area in face-identity acquisition. *NeuroImage*, 148, 212-218. <https://doi.org/10.1016/j.neuroimage.2017.01.043>
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological review*, 111(4), 1036. <https://doi.org/10.1037/0033-295x.111.4.1036>
- Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and

References

- changeable aspects of faces in the human temporal lobe. *Neuroimage*, 23(3), 905-913. <https://doi.org/10.1016/j.neuroimage.2004.07.060>
- Ashburner, J., Chen, C., Moran, R., Henson, R., Phillips, C., n.d. SPM12 Manual The FIL Methods Group (and honorary members) 1–508.
- Atkinson, R.C. and Shiffrin, R.M. (1968) Human memory: a proposed system and its control processes. In: Spence K.W. and Spence J.T. (Eds.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 2. Academic Press, New York, pp. 89–195. [https://doi.org/10.1016/s0079-7421\(08\)60422-3](https://doi.org/10.1016/s0079-7421(08)60422-3)
- Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex*, 80, 125-140. <https://doi.org/10.1016/j.cortex.2015.11.024>
- Avidan, G., Hasson, U., Hendler, T., Zohary, E., & Malach, R. (2002). Analysis of the neuronal selectivity underlying low fMRI signals. *Current Biology*, 12(12), 964-972. [https://doi.org/10.1016/s0960-9822\(02\)00872-2](https://doi.org/10.1016/s0960-9822(02)00872-2)
- Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in cognitive sciences*, 5(3), 119-126. [https://doi.org/10.1016/S1364-6613\(00\)01593-X](https://doi.org/10.1016/S1364-6613(00)01593-X)
- Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 780. <https://doi.org/10.1037/0096-1523.24.3.780>
- Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience*, 139(1), 201-208. <https://doi.org/10.1016/j.neuroscience.2005.08.023>
- Baeck, A., & de Beeck, H. P. O. (2010). Transfer of object learning across distinct visual learning paradigms. *Journal of vision*, 10(2), 17-17. <https://doi.org/10.1167/10.2.17>
- Baker, C. I., Liu, J., Wald, L. L., Kwong, K. K., Benner, T., & Kanwisher, N. (2007). Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proceedings of the National Academy of Sciences*, 104(21), 9087-9092.

References

<https://doi.org/10.1073/pnas.0703300104>

- Barton, J. J., Press, D. Z., Keenan, J. P., & O'Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology*, 58(1), 71-78. <https://doi.org/10.1212/wnl.58.1.71>
- Barton, J. J., Fox, C. J., Sekunova, A., & Iaria, G. (2010). Encoding in the visual word form area: An fMRI adaptation study of words versus handwriting. *Journal of Cognitive Neuroscience*, 22(8), 1649-1661. <https://doi.org/10.1162/jocn.2009.21286>
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, 76(4), 695-711. <https://doi.org/10.1016/j.neuron.2012.10.038>
- Baylis, G. C., & Rolls, E. T. (1987). Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. *Experimental brain research*, 65(3), 614-622. <https://doi.org/10.1007/bf00235984>
- Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: a review. *International Journal of Psychophysiology*, 83(2), 120-131. <https://doi.org/10.1016/j.ijpsycho.2011.08.003>
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience*, 8(6), 551-565. <https://doi.org/10.1162/jocn.1996.8.6.551>
- Bi, T., & Fang, F. (2013). Neural plasticity in high-level visual cortex underlying object perceptual learning. *Frontiers in Biology*, 8(4), 434-443. <https://doi.org/10.1007/s11515-013-1262-2>
- Bona, S., Silvanto, J., & Cattaneo, Z. (2018). TMS over right OFA affects individuation of faces but not of exemplars of objects. *Neuropsychologia*, 117, 364-370. <https://doi.org/10.1016/j.neuropsychologia.2018.06.024>
- Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature reviews neuroscience*, 3(3), 243-249. <https://doi.org/10.1038/nrn756>

References

- Buckner, R. L., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D., ... & Dale, A. M. (1998). Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. *Neuron*, 20(2), 285-296. [https://doi.org/10.1016/s0896-6273\(00\)80456-0](https://doi.org/10.1016/s0896-6273(00)80456-0)
- Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: the power of an expertise framework. *Trends in cognitive sciences*, 10(4), 159-166. <https://doi.org/10.1016/j.tics.2006.02.004>
- Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision research*, 45(4), 431-448. <https://doi.org/10.1016/j.visres.2004.08.021>
- Cai, Q., Lavidor, M., Brysbaert, M., Paulignan, Y., & Nazir, T. A. (2008). Cerebral lateralization of frontal lobe language processes and lateralization of the posterior visual word processing system. *Journal of Cognitive Neuroscience*, 20(4), 672-681. <https://doi.org/10.1162/jocn.2008.20043>
- Cai, Q., Paulignan, Y., Brysbaert, M., Ibarrola, D., & Nazir, T. A. (2010). The left ventral occipito-temporal response to words depends on language lateralization but not on visual familiarity. *Cerebral Cortex*, 20(5), 1153-1163. <https://doi.org/10.1093/cercor/bhp175>
- Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6(8), 641-651. <https://doi.org/10.1038/nrn1724>
- Cao, X., Li, S., Zhao, J., & Weng, X. (2011). Left-lateralized early neurophysiological response for Chinese characters in young primary school children. *Neuroscience letters*, 492(3), 165-169. <https://doi.org/10.1016/j.neulet.2011.02.002>
- Cao, X., Ma, X., & Qi, C. (2015). N170 adaptation effect for repeated faces and words. *Neuroscience*, 294, 21-28. <https://doi.org/10.1016/j.neuroscience.2015.03.009>
- Cashdollar, N., Ruhnau, P., Weisz, N., & Hasson, U. (2017). The role of working memory in the probabilistic inference of future sensory events. *Cerebral Cortex*, 27(5),

References

- 2955-2969. <https://doi.org/10.1093/cercor/bhw138>
- Cheung, O. S., & Bar, M. (2012). Visual prediction and perceptual expertise. *International Journal of Psychophysiology*, 83(2), 156-163. <https://doi.org/10.1016/j.ijpsycho.2011.11.002>
- Chouinard, P. A., Meena, D. K., Whitwell, R. L., Hilchey, M. D., & Goodale, M. A. (2017). A TMS investigation on the role of lateral occipital complex and caudal intraparietal sulcus in the perception of object form and orientation. *Journal of cognitive neuroscience*, 29(5), 881-895. https://doi.org/10.1162/jocn_a_01094
- Chouinard, P. A., Morrissey, B. F., Köhler, S., & Goodale, M. A. (2008). Repetition suppression in occipital-temporal visual areas is modulated by physical rather than semantic features of objects. *Neuroimage*, 41(1), 130-144. <https://doi.org/10.1016/j.neuroimage.2008.02.011>
- Cichy, R. M., Chen, Y., & Haynes, J. D. (2011). Encoding the identity and location of objects in human LOC. *Neuroimage*, 54(3), 2297-2307. <https://doi.org/10.1016/j.neuroimage.2010.09.044>
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181-204. <https://doi.org/10.1017/S0140525X12000477>
- Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.
- Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123(2), 291-307. <https://doi.org/10.1093/brain/123.2.291>
- Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, 125(5), 1054-1069. <https://doi.org/10.1093/brain/awf094>
- Cowan, N. (1998). *Attention and memory: An integrated framework*. Oxford University Press.

References

- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory?. *Progress in brain research*, 169, 323-338. [https://doi.org/10.1016/s0079-6123\(07\)00020-9](https://doi.org/10.1016/s0079-6123(07)00020-9)
- Davies-Thompson, J., Johnston, S., Tashakkor, Y., Pancaroglu, R., & Barton, J. J. (2016). The relationship between visual word and face processing lateralization in the fusiform gyri: A cross-sectional study. *Brain research*, 1644, 88-97. <https://doi.org/10.1016/j.brainres.2016.05.009>
- De Baene, W., & Vogels, R. (2010). Effects of adaptation on the stimulus selectivity of macaque inferior temporal spiking activity and local field potentials. *Cerebral cortex*, 20(9), 2145-2165. <https://doi.org/10.1093/cercor/bhp277>
- Deffke, I., Sander, T., Heidenreich, J., Sommer, W., Curio, G., Trahms, L., & Lueschow, A. (2007). MEG/EEG sources of the 170-ms response to faces are co-localized in the fusiform gyrus. *Neuroimage*, 35(4), 1495-1501. <https://doi.org/10.1016/j.neuroimage.2007.01.034>
- Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, 56(2), 384-398. <https://doi.org/10.1016/j.neuron.2007.10.004>
- Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in cognitive sciences*, 15(6), 254-262. <https://doi.org/10.1016/j.tics.2011.04.003>
- Dehaene, S., Le Clec'H, G., Poline, J. B., Le Bihan, D., & Cohen, L. (2002). The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport*, 13(3), 321-325. <https://doi.org/10.1097/00001756-200203040-00015>
- Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature neuroscience*, 4(7), 752-758. <https://doi.org/10.1038/89551>
- den Ouden, H. E., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *Journal of Neuroscience*, 30(9), 3210-3219. <https://doi.org/10.1523/JNEUROSCI.4458-09.2010>

References

- Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences*, 93(24), 13494-13499. <https://doi.org/10.1073/pnas.93.24.13494>
- Duchaine, B., & Yovel, G. (2015). A revised neural framework for face processing. *Annual review of vision science*, 1, 393-416. <https://doi.org/10.1146/annurev-vision-082114-035518>
- Efron, R. (1969). What is perception?. In *Proceedings of the Boston Colloquium for the Philosophy of Science 1966/1968* (pp. 137-173). Springer, Dordrecht. https://doi.org/10.1007/978-94-010-3378-7_4
- Eger, E., Ashburner, J., Haynes, J. D., Dolan, R. J., & Rees, G. (2008). fMRI activity patterns in human LOC carry information about object exemplars within category. *Journal of cognitive neuroscience*, 20(2), 356-370. <https://doi.org/10.1162/jocn.2008.20019>
- Eger, E., Henson, R. N. A., Driver, J., & Dolan, R. J. (2004). BOLD repetition decreases in object-responsive ventral visual areas depend on spatial attention. *Journal of neurophysiology*, 92(2), 1241-1247. <https://doi.org/10.1152/jn.00206.2004>
- Egner, T., & Summerfield, C. (2013). Grounding predictive coding models in empirical neuroscience research. *Behavioral and Brain Sciences*, 36(3), 210-211. <https://doi.org/10.1017/s0140525x1200218x>
- Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *Journal of Neuroscience*, 30(49), 16601-16608. <https://doi.org/10.1523/JNEUROSCI.2770-10.2010>
- Eick, C. M., Kovács, G., Rostalski, S. M., Röhrig, L., & Ambrus, G. G. (2020). The occipital face area is causally involved in identity-related visual-semantic associations. *Brain Structure and Function*, 225(5), 1483-1493. <https://doi.org/10.1007/s00429-020-02068-9>
- Eimer, M., Kiss, M., & Nicholas, S. (2010). Response profile of the face-sensitive N170 component: a rapid adaptation study. *Cerebral cortex*, 20(10), 2442-2452. <https://doi.org/10.1093/cercor/bhp312>

References

- Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, 2(10), 704-716. <https://doi.org/10.1038/35094565>
- Epstein, R. A., Parker, W. E., & Feiler, A. M. (2008). Two kinds of fMRI repetition suppression? Evidence for dissociable neural mechanisms. *Journal of Neurophysiology*, 99(6), 2877-2886. <https://doi.org/10.1152/jn.90376.2008>
- Fahle, M., Poggio, T., & Poggio, T. A. (Eds.). (2002). *Perceptual learning*. MIT Press.
- Fan, X., Wang, F., Shao, H., Zhang, P., & He, S. (2020). The bottom-up and top-down processing of faces in the human occipitotemporal cortex. *Elife*, 9, e48764. <https://doi.org/10.7554/eLife.48764>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175-191. <https://doi.org/10.3758/bf03193146>
- Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in human neuroscience*, 4, 215. <https://doi.org/10.3389/fnhum.2010.00215>
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 1(1), 1-47. <https://doi.org/10.1093/cercor/1.1.1-a>
- Feuerriegel, D., Blom, T., & Hogendoorn, H. (2021a). Predictive activation of sensory representations as a source of evidence in perceptual decision-making. *Cortex*, 136, 140-146. <https://doi.org/10.1016/j.cortex.2020.12.008>
- Feuerriegel, D., Vogels, R., & Kovács, G. (2021b). Evaluating the evidence for expectation suppression in the visual system. *Neuroscience & Biobehavioral Reviews*, 126, 368-381. <https://doi.org/10.1016/j.neubiorev.2021.04.002>
- Feuerriegel, D., Yook, J., Quek, G. L., Hogendoorn, H., & Bode, S. (2021c). Visual mismatch responses index surprise signalling but not expectation suppression. *Cortex*, 134, 16-29. <https://doi.org/10.1016/j.cortex.2020.10.006>
- Ficco, L., Mancuso, L., Manuello, J., Teneggi, A., Liloia, D., Duca, S., ... & Cauda, F. (2021).

References

- Disentangling predictive processing in the brain: a meta-analytic study in favour of a predictive network. *Scientific Reports*, 11(1), 1-14. <https://doi.org/10.1038/s41598-021-95603-5>
- Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature*, 287(5777), 43-44. <https://doi.org/10.1038/287043a0>
- Fox, C. J., Moon, S. Y., Iaria, G., & Barton, J. J. (2009). The correlates of subjective perception of identity and expression in the face network: an fMRI adaptation study. *Neuroimage*, 44(2), 569-580. <https://doi.org/10.1016/j.neuroimage.2008.09.011>
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cerebral Cortex*, 16(11), 1631-1644. <https://doi.org/10.1093/cercor/bhj100>
- Friese, U., Supp, G. G., Hipp, J. F., Engel, A. K., & Gruber, T. (2012). Oscillatory MEG gamma band activity dissociates perceptual and conceptual aspects of visual object processing: A combined repetition/conceptual priming study. *Neuroimage*, 59(1), 861-871. <https://doi.org/10.1016/j.neuroimage.2011.07.073>
- Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, 360(1456), 815-836. <https://doi.org/10.1098/rstb.2005.1622>
- Friston, K. (2009). The free-energy principle: a rough guide to the brain?. *Trends in cognitive sciences*, 13(7), 293-301. <https://doi.org/10.1016/j.tics.2009.04.005>
- Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, 11(2), 127-138. <https://doi.org/10.1038/nrn2787>
- Friston, K. (2012). Predictive coding, precision and synchrony. *Cognitive neuroscience*, 3(3-4), 238-239. <https://doi.org/10.1080/17588928.2012.691277>
- Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical transactions of the Royal Society B: Biological sciences*, 364(1521), 1211-1221. <https://doi.org/10.1098/rstb.2008.0300>
- Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of*

References

- physiology-Paris*, 100(1-3), 70-87.
<https://doi.org/10.1016/j.jphysparis.2006.10.001>
- Furl, N., Garrido, L., Dolan, R. J., Driver, J., & Duchaine, B. (2011). Fusiform gyrus face selectivity relates to individual differences in facial recognition ability. *Journal of cognitive neuroscience*, 23(7), 1723-1740.
<https://doi.org/10.1162/jocn.2010.21545>
- Gagnepain, P., Chételat, G., Landeau, B., Dayan, J., Eustache, F., & Lebreton, K. (2008). Spoken word memory traces within the human auditory cortex revealed by repetition priming and functional magnetic resonance imaging. *Journal of Neuroscience*, 28(20), 5281-5289.
<https://doi.org/10.1523/JNEUROSCI.0565-08.2008>
- Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature neuroscience*, 6(4), 428-432.
<https://doi.org/10.1038/nn1029>
- Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3(2), 191-197. <https://doi.org/10.1038/72140>
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature neuroscience*, 2(6), 568-573. <https://doi.org/10.1038/9224>
- Gilbert, C., & Bakan, P. (1973). Visual asymmetry in perception of faces. *Neuropsychologia*, 11(3), 355-362.
[https://doi.org/10.1016/0028-3932\(73\)90049-3](https://doi.org/10.1016/0028-3932(73)90049-3)
- Glezer, L. S., Jiang, X., & Riesenhuber, M. (2009). Evidence for highly selective neuronal tuning to whole words in the "visual word form area". *Neuron*, 62(2), 199-204.
<https://doi.org/10.1016/j.neuron.2009.03.017>
- Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, 45(1), 32-41.
<https://doi.org/10.1016/j.neuropsychologia.2006.04.015>

References

- Goldstone, R. L. (1998). Perceptual learning. *Annual review of psychology*, 49(1), 585-612.
<https://doi.org/10.1146/annurev.psych.49.1.585>
- Gotts, S. J. (2003). Mechanisms underlying enhanced processing efficiency in neural systems. *Unpublished Ph. D. thesis, Carnegie Mellon University, Pittsburgh, PA.*
- Gotts, S. J., Chow, C. C., & Martin, A. (2012). Repetition priming and repetition suppression: A case for enhanced efficiency through neural synchronization. *Cognitive Neuroscience*, 3(3-4), 227-237. <https://doi.org/10.1080/17588928.2012.670617>
- Grill-Spector, K. (2003). The neural basis of object perception. *Current opinion in neurobiology*, 13(2), 159-166. [https://doi.org/10.1016/s0959-4388\(03\)00060-6](https://doi.org/10.1016/s0959-4388(03)00060-6)
- Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta psychologica*, 107(1-3), 293-321.
[https://doi.org/10.1016/s0001-6918\(01\)00019-1](https://doi.org/10.1016/s0001-6918(01)00019-1)
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in cognitive sciences*, 10(1), 14-23.
<https://doi.org/10.1016/j.tics.2005.11.006>
- Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, 7(5), 555-562. <https://doi.org/10.1038/nn1224>
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision research*, 41(10-11), 1409-1422.
[https://doi.org/10.1016/s0042-6989\(01\)00073-6](https://doi.org/10.1016/s0042-6989(01)00073-6)
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, 24(1), 187-203.
[https://doi.org/10.1016/s0896-6273\(00\)80832-6](https://doi.org/10.1016/s0896-6273(00)80832-6)
- Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature neuroscience*, 3(8), 837-843. <https://doi.org/10.1038/77754>
- Gross, C. G., Rocha-Miranda, C. D., & Bender, D. B. (1972). Visual properties of neurons in

References

- inferotemporal cortex of the macaque. *Journal of neurophysiology*, 35(1), 96-111.
<https://doi.org/10.1152/jn.1972.35.1.96>
- Gross, C. G., Schiller, P. H., Wells, C., & Gerstein, G. L. (1967). Single-unit activity in temporal association cortex of the monkey. *Journal of Neurophysiology*, 30(4), 833-843. <https://doi.org/10.1152/jn.1967.30.4.833>
- Grossmann, T., Gliga, T., Johnson, M. H., & Mareschal, D. (2009). The neural basis of perceptual category learning in human infants. *Journal of Cognitive Neuroscience*, 21(12), 2276-2286. <https://doi.org/10.1162/jocn.2009.21188>
- Grotheer, M., & Kovács, G. (2014). Repetition probability effects depend on prior experiences. *Journal of Neuroscience*, 34(19), 6640-6646.
<https://doi.org/10.1523/JNEUROSCI.5326-13.2014>
- Grotheer, M., & Kovács, G. (2015). The relationship between stimulus repetitions and fulfilled expectations. *Neuropsychologia*, 67, 175-182.
<https://doi.org/10.1016/j.neuropsychologia.2014.12.017>
- Grotheer, M., & Kovács, G. (2016). Can predictive coding explain repetition suppression?. *Cortex*, 80, 113-124. <https://doi.org/10.1016/j.cortex.2015.11.027>
- Grotheer, M., Hermann, P., Vidnyánszky, Z., & Kovács, G. (2014). Repetition probability effects for inverted faces. *NeuroImage*, 102, 416-423.
<https://doi.org/10.1016/j.neuroimage.2014.08.006>
- Hannagan, T., Agrawal, A., Cohen, L., & Dehaene, S. (2021). Simulating the emergence of the Visual Word Form Area: Recycling a convolutional neural network for reading. *bioRxiv*. <https://doi.org/10.1101/2021.02.15.431235>
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539), 2425-2430.
<https://doi.org/10.1126/science.1063736>
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223-233.
[https://doi.org/10.1016/s1364-6613\(00\)01482-0](https://doi.org/10.1016/s1364-6613(00)01482-0)

References

- Heilbron, M., & Chait, M. (2018). Great expectations: is there evidence for predictive coding in auditory cortex?. *Neuroscience*, 389, 54-73.
<https://doi.org/10.1016/j.neuroscience.2017.07.061>
- Hein, G., & Knight, R. T. (2008). Superior temporal sulcus—it's my area: or is it?. *Journal of cognitive neuroscience*, 20(12), 2125-2136.
<https://doi.org/10.1162/jocn.2008.20148>
- Helmholtz, H.v., 1867. Handbuch der physiologischen Optik, Voss, Leipzig. 874 pp
- Henson, R. N. (2012). Repetition accelerates neural dynamics: In defense of facilitation models. *Cognitive Neuroscience*, 3(3-4), 240-241.
<https://doi.org/10.1080/17588928.2012.689962>
- Henson, R. N. (2016). Repetition suppression to faces in the fusiform face area: A personal and dynamic journey. *Cortex*, 80, 174-184.
<https://doi.org/10.1016/j.cortex.2015.09.012>
- Henson, R. N. A., & Rugg, M. D. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia*, 41(3), 263-270.
[https://doi.org/10.1016/s0028-3932\(02\)00159-8](https://doi.org/10.1016/s0028-3932(02)00159-8)
- Henson, R. N. A., Shallice, T., Gorno-Tempini, M. L., & Dolan, R. J. (2002a). Face repetition effects in implicit and explicit memory tests as measured by fMRI. *Cerebral Cortex*, 12(2), 178-186. <https://doi.org/10.1093/cercor/12.2.178>
- Henson, R. N., Price, C. J., Rugg, M. D., Turner, R., & Friston, K. J. (2002b). Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. *Neuroimage*, 15(1), 83-97. <https://doi.org/10.1006/nimg.2001.0940>
- Henson, R. N., Rylands, A., Ross, E., Vuilleumier, P., & Rugg, M. D. (2004). The effect of repetition lag on electrophysiological and haemodynamic correlates of visual object priming. *Neuroimage*, 21(4), 1674-1689.
<https://doi.org/10.1016/j.neuroimage.2003.12.020>
- Henson, R., Shallice, T., & Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. *Science*, 287(5456), 1269-1272.

References

- <https://doi.org/10.1126/science.287.5456.1269>
- Hickey, C., Van Zoest, W., & Theeuwes, J. (2010). The time course of exogenous and endogenous control of covert attention. *Experimental brain research*, 201(4), 789-796. <https://doi.org/10.1007/s00221-009-2094-9>
- Huang, G., Ramachandran, S., Lee, T. S., & Olson, C. R. (2018). Neural correlate of visual familiarity in macaque area V2. *Journal of Neuroscience*, 38(42), 8967-8975. <https://doi.org/10.1523/JNEUROSCI.0664-18.2018>
- Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of neurophysiology*, 28(2), 229-289. <https://doi.org/10.1152/jn.1965.28.2.229>
- Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging* (Vol. 1). Sunderland, MA: Sinauer Associates.
- Huettig, F. (2015). Four central questions about prediction in language processing. *Brain research*, 1626, 118-135. <https://doi.org/10.1016/j.brainres.2015.02.014>
- Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., VanMeter, J., & Riesenhuber, M. (2007). Categorization training results in shape-and category-selective human neural plasticity. *Neuron*, 53(6), 891-903. <https://doi.org/10.1016/j.neuron.2007.02.015>
- Kaliukhovich, D. A., & Vogels, R. (2011). Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. *Cerebral cortex*, 21(7), 1547-1558. <https://doi.org/10.1093/cercor/bhq207>
- Kaliukhovich, D. A., & Vogels, R. (2014). Neurons in macaque inferior temporal cortex show no surprise response to deviants in visual oddball sequences. *Journal of Neuroscience*, 34(38), 12801-12815. <https://doi.org/10.1523/JNEUROSCI.2154-14.2014>
- Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1476), 2109-2128. <https://doi.org/10.1098/rstb.2006.1934>
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997a). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of*

References

- neuroscience*, 17(11), 4302-4311.
<https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997>
- Kanwisher, N., Woods, R. P., Iacoboni, M., & Mazziotta, J. C. (1997b). A locus in human extrastriate cortex for visual shape analysis. *Journal of Cognitive Neuroscience*, 9(1), 133-142. <https://doi.org/10.1162/jocn.1997.9.1.133>
- Kar, K., & Krekelberg, B. (2016). Testing the assumptions underlying fMRI adaptation using intracortical recordings in area MT. *Cortex*, 80, 21-34. <https://doi.org/10.1016/j.cortex.2015.12.011>
- Keller, G. B., & Morsic-Flogel, T. D. (2018). Predictive processing: a canonical cortical computation. *Neuron*, 100(2), 424-435. <https://doi.org/10.1016/j.neuron.2018.10.003>
- Kiyonaga, A., & Egner, T. (2013). Working memory as internal attention: Toward an integrative account of internal and external selection processes. *Psychonomic bulletin & review*, 20(2), 228-242. <https://doi.org/10.3758/s13423-012-0359-y>
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12), 712-719. <https://doi.org/10.1016/j.tins.2004.10.007>
- Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Current Biology*, 26(3), 371-376. <https://doi.org/10.1016/j.cub.2015.12.038>
- Kok, P., Jehee, J. F., & De Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, 75(2), 265-270. <https://doi.org/10.1016/j.neuron.2012.04.034>
- Kok, P., & Lange, F. P. D. (2015). Predictive coding in sensory cortex. In *An introduction to model-based cognitive neuroscience* (pp. 221-244). Springer, New York, NY. https://doi.org/10.1007/978-1-4939-2236-9_11
- Kouider, S., de Gardelle, V., Dehaene, S., Dupoux, E., & Pallier, C. (2010). Cerebral bases of subliminal speech priming. *Neuroimage*, 49(1), 922-929.

References

- <https://doi.org/10.1016/j.neuroimage.2009.08.043>
- Kovács, G. (2020). Getting to know someone: Familiarity, person recognition, and identification in the human brain. *Journal of cognitive neuroscience*, 32(12), 2205-2225. https://doi.org/10.1162/jocn_a_01627
- Kovács, G., & Vogels, R. (2014). When does repetition suppression depend on repetition probability?. *Frontiers in Human Neuroscience*, 8, 685. <https://doi.org/10.3389/fnhum.2014.00685>
- Kovács, G., Cziraki, C., Vidnyánszky, Z., Schweinberger, S. R., & Greenlee, M. W. (2008). Position-specific and position-invariant face aftereffects reflect the adaptation of different cortical areas. *Neuroimage*, 43(1), 156-164. <https://doi.org/10.1016/j.neuroimage.2008.06.042>
- Kovács, G., Iffland, L., Vidnyánszky, Z., & Greenlee, M. W. (2012). Stimulus repetition probability effects on repetition suppression are position invariant for faces. *Neuroimage*, 60(4), 2128-2135. <https://doi.org/10.1016/j.neuroimage.2012.02.038>
- Kovács, G., Kaiser, D., Kaliukhovich, D. A., Vidnyánszky, Z., & Vogels, R. (2013). Repetition probability does not affect fMRI repetition suppression for objects. *Journal of Neuroscience*, 33(23), 9805-9812. <https://doi.org/10.1523/JNEUROSCI.3423-12.2013>
- Kovács, G., Zimmer, M., Bankó, É., Harza, I., Antal, A., & Vidnyánszky, Z. (2006). Electrophysiological correlates of visual adaptation to faces and body parts in humans. *Cerebral Cortex*, 16(5), 742-753. <https://doi.org/10.1093/cercor/bhj020>
- Kremláček, J., Kreegipuu, K., Tales, A., Astikainen, P., Poldver, N., Näätänen, R., & Stefanics, G. (2016). Visual mismatch negativity (vMMN): A review and meta-analysis of studies in psychiatric and neurological disorders. *Cortex*, 80, 76-112. <https://doi.org/10.1016/j.cortex.2016.03.017>
- Large, M. E., Cavina-Pratesi, C., Vilis, T., & Culham, J. C. (2008). The neural correlates of change detection in the face perception network. *Neuropsychologia*, 46(8), 2169-2176. <https://doi.org/10.1016/j.neuropsychologia.2008.02.027>

References

- Larsson, J., & Smith, A. T. (2012). fMRI repetition suppression: neuronal adaptation or stimulus expectation?. *Cerebral Cortex*, 22(3), 567-576.
<https://doi.org/10.1093/cercor/bhr119>
- Larsson, J., Solomon, S. G., & Kohn, A. (2016). fMRI adaptation revisited. *Cortex*, 80, 154-160. <https://doi.org/10.1016/j.cortex.2015.10.026>
- Li, C., Ma, X., Zhu, C., & Cao, X. (2019). The recovery speed of category sensitive N170 responses to faces and Chinese characters. *Brain research*, 1723, 146384.
<https://doi.org/10.1016/j.brainres.2019.146384>
- Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of neurophysiology*, 69(6), 1918-1929.
<https://doi.org/10.1152/jn.1993.69.6.1918>
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., ... & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, 92(18), 8135-8139. <https://doi.org/10.1073/pnas.92.18.8135>
- Manahova, M. E., Mostert, P., Kok, P., Schoffelen, J. M., & De Lange, F. P. (2018). Stimulus familiarity and expectation jointly modulate neural activity in the visual ventral stream. *Journal of Cognitive Neuroscience*, 30(9), 1366-1377.
https://doi.org/10.1162/jocn_a_01281
- Manahova, M. E., Spaak, E., & de Lange, F. P. (2020). Familiarity increases processing speed in the visual system. *Journal of cognitive neuroscience*, 32(4), 722-733.
https://doi.org/10.1162/jocn_a_01507
- Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. Freeman, San Francisco.
[https://doi.org/10.1016/0022-2496\(83\)90030-5](https://doi.org/10.1016/0022-2496(83)90030-5)
- Matsuo, T., Kawasaki, K., Kawai, K., Majima, K., Masuda, H., Murakami, H., ... & Hasegawa, I. (2015). Alternating zones selective to faces and written words in the human ventral occipitotemporal cortex. *Cerebral Cortex*, 25(5), 1265-1277.
<https://doi.org/10.1093/cercor/bht319>

References

- Matthews, P. M., & Jezzard, P. (2004). Functional magnetic resonance imaging. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(1), 6-12.
<https://jnnp.bmj.com/content/75/1/6>
- Maunsell, J. H. (2015). Neuronal mechanisms of visual attention. *Annual review of vision science*, 1, 373-391. <https://doi.org/10.1146/annurev-vision-082114-035431>
- Maurer, D., O'craven, K. M., Le Grand, R., Mondloch, C. J., Springer, M. V., Lewis, T. L., & Grady, C. L. (2007). Neural correlates of processing facial identity based on features versus their spacing. *Neuropsychologia*, 45(7), 1438-1451.
<https://doi.org/10.1016/j.neuropsychologia.2006.11.016>
- Maurer, U., Brem, S., Kranz, F., Bucher, K., Benz, R., Halder, P., ... & Brandeis, D. (2006). Coarse neural tuning for print peaks when children learn to read. *Neuroimage*, 33(2), 749-758.
<https://doi.org/10.1016/j.neuroimage.2006.06.025>
- Mayrhauser, L., Bergmann, J., Crone, J., & Kronbichler, M. (2014). Neural repetition suppression: evidence for perceptual expectation in object-selective regions. *Frontiers in Human Neuroscience*, 8, 225.
<https://doi.org/10.3389/fnhum.2014.00225>
- Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., & Lancaster, J. (1995). A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage*, 2(2), 89-101. <https://doi.org/10.1006/nimg.1995.1012>
- McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces?. *Trends in cognitive sciences*, 11(1), 8-15.
<https://doi.org/10.1016/j.tics.2006.11.002>
- Mercure, E., Cohen Kadosh, K., & Johnson, M. (2011). The N170 shows differential repetition effects for faces, objects, and orthographic stimuli. *Frontiers in human neuroscience*, 5, 6. <https://doi.org/10.3389/fnhum.2011.00006>
- Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences*, 108(48), 19401-19406. <https://doi.org/10.1073/pnas.1112895108>

References

- Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science*, 263(5146), 520-522. <http://www.jstor.org/stable/2883033>
- Miller, E. K., Li, L., & Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science*, 254(5036), 1377-1379. <https://doi.org/10.1126/science.1962197>
- Miller, G.A., Galanter, E. and Pribram, K.H. (1960) Plans and the structure of behavior. Holt, Rinehart and Winston, Inc., New York. <https://www.jstor.org/stable/667610>
- Milner, D., & Goodale, M. (2006). *The visual brain in action* (Vol. 27). OUP Oxford.
- Moret-Tatay, C., Fortea, I. B., & Sevilla, M. D. G. (2020). Challenges and insights for the visual system: Are face and word recognition two sides of the same coin?. *Journal of Neurolinguistics*, 56, 100941. <https://doi.org/10.1016/j.jneuroling.2020.100941>
- Müller, N. G., Strumpf, H., Scholz, M., Baier, B., & Melloni, L. (2013). Repetition suppression versus enhancement—it's quantity that matters. *Cerebral cortex*, 23(2), 315-322. <https://doi.org/10.1093/cercor/bhs009>
- Müller, V. I., Höhner, Y., & Eickhoff, S. B. (2018). Influence of task instructions and stimuli on the neural network of face processing: An ALE meta-analysis. *Cortex*, 103, 240-255. <https://doi.org/10.1016/j.cortex.2018.03.011>
- Murray, S. O., & Wojciulik, E. (2004). Attention increases neural selectivity in the human lateral occipital complex. *Nature neuroscience*, 7(1), 70-74. <https://doi.org/10.1038/nn1161>
- Näätänen, R. (1992). *Attention and Brain Function* (1st ed.). Routledge. <https://doi.org/10.4324/9780429487354>
- Nemrodov, D., & Itier, R. J. (2011). The role of eyes in early face processing: a rapid adaptation study of the inversion effect. *British journal of psychology*, 102(4), 783-798. <https://doi.org/10.1111/j.2044-8295.2011.02033.x>
- Nordt, M., Hoehl, S., & Weigelt, S. (2016). The use of repetition suppression paradigms in developmental cognitive neuroscience. *Cortex*, 80, 61-75. <https://doi.org/10.1016/j.cortex.2016.04.002>

References

- Olkkonen, M., Aguirre, G. K., & Epstein, R. A. (2017). Expectation modulates repetition priming under high stimulus variability. *Journal of vision*, 17(6), 1-16. <https://doi.org/10.1167/17.6.10>
- Op de Beeck, H. P., & Baker, C. I. (2010). The neural basis of visual object learning. *Trends in cognitive sciences*, 14(1), 22-30. <https://doi.org/10.1016/j.tics.2009.11.002>
- Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *Journal of Neuroscience*, 26(50), 13025-13036. <https://doi.org/10.1523/jneurosci.2481-06.2006>
- Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends in Cognitive Sciences*, 25(2), 100-110. <https://doi.org/10.1016/j.tics.2020.11.006>
- Pitcher, D., Duchaine, B., & Walsh, V. (2014). Combined TMS and fMRI reveal dissociable cortical pathways for dynamic and static face perception. *Current Biology*, 24(17), 2066-2070. <https://doi.org/10.1016/j.cub.2014.07.060>
- Pitcher, D., Garrido, L., Walsh, V., & Duchaine, B. C. (2008). Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions. *Journal of Neuroscience*, 28(36), 8929-8933. <https://doi.org/10.1523/jneurosci.1450-08.2008>
- Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental brain research*, 209(4), 481-493. <https://doi.org/10.1007/s00221-011-2579-1>
- Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology*, 17(18), 1568-1573. <https://doi.org/10.1016/j.cub.2007.07.063>
- Ramachandran, S., Meyer, T., & Olson, C. R. (2016). Prediction suppression in monkey inferotemporal cortex depends on the conditional probability between images. *Journal of neurophysiology*, 115(1), 355-362. <https://doi.org/10.1152/jn.00091.2015>

References

- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1), 79-87. <https://doi.org/10.1038/4580>
- Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Sciences*, 95(2), 747-750. <https://doi.org/10.1073/pnas.95.2.747>
- Richter, D., & de Lange, F. P. (2019). Statistical learning attenuates visual activity only for attended stimuli. *Elife*, 8, e47869. <https://doi.org/10.7554/eLife.47869>
- Rock, I., 1983. *The Logic of Perception*. MIT Press, Cambridge MA. 341 pp
- Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta psychologica*, 128(2), 274-289. <https://doi.org/10.1016/j.actpsy.2008.02.003>
- Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces?. *Behavioral and cognitive neuroscience reviews*, 1(1), 63-75. <https://doi.org/10.1177/1534582302001001004>
- Rossion, B., Caldara, R., Seghier, M., Schuller, A. M., Lazeyras, F., & Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain*, 126(11), 2381-2395. <https://doi.org/10.1093/brain/awg241>
- Roth, Z. N., & Zohary, E. (2015). Position and identity information available in fMRI patterns of activity in human visual cortex. *Journal of Neuroscience*, 35(33), 11559-11571. <https://doi.org/10.1523/jneurosci.0752-15.2015>
- Rotshtein, P., Henson, R. N., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature neuroscience*, 8(1), 107-113. <https://doi.org/10.1038/nn1370>
- Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain sciences*, 26(6), 709-728. <https://doi.org/10.1017/S0140525X03000165>
- Sayres, R., & Grill-Spector, K. (2008). Relating retinotopic and object-selective responses

References

- in human lateral occipital cortex. *Journal of neurophysiology*, 100(1), 249-267.
<https://doi.org/10.1152/jn.01383.2007>
- Schendan, H. E., & Kutas, M. (2003). Time course of processes and representations supporting visual object identification and memory. *Journal of Cognitive Neuroscience*, 15(1), 111-135. <https://doi.org/10.1162/089892903321107864>
- Schwarz, L., Kreifelts, B., Wildgruber, D., Erb, M., Scheffler, K., & Ethofer, T. (2019). Properties of face localizer activations and their application in functional magnetic resonance imaging (fMRI) fingerprinting. *Plos one*, 14(4), e0214997. <https://doi.org/10.1371/journal.pone.0214997>
- Schweinberger, S. R., & Neumann, M. F. (2016). Repetition effects in human ERPs to faces. *Cortex*, 80, 141-153. <https://doi.org/10.1016/j.cortex.2015.11.001>
- Schweinberger, S. R., Huddy, V., & Burton, A. M. (2004). N250r: a face-selective brain response to stimulus repetitions. *NeuroReport*, 15(9), 1501-1505. <https://doi.org/10.1097/01.wnr.0000131675.00319.42>
- Schweinberger, S. R., Kaufmann, J. M., Moratti, S., Keil, A., & Burton, A. M. (2007). Brain responses to repetitions of human and animal faces, inverted faces, and objects—an MEG study. *Brain research*, 1184, 226-233. <https://doi.org/10.1016/j.brainres.2007.09.079>
- Schweinberger, S. R., Pfütze, E. M., & Sommer, W. (1995). Repetition priming and associative priming of face recognition: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(3), 722-736. <https://doi.org/10.1037/0278-7393.21.3.722>
- Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., & Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Cognitive Brain Research*, 14(3), 398-409. [https://doi.org/10.1016/s0926-6410\(02\)00142-8](https://doi.org/10.1016/s0926-6410(02)00142-8)
- Segaert, K., Weber, K., de Lange, F. P., Petersson, K. M., & Hagoort, P. (2013). The suppression of repetition enhancement: a review of fMRI studies. *Neuropsychologia*, 51(1), 59-66.

References

- <https://doi.org/10.1016/j.neuropsychologia.2012.11.006>
- Serences, J. T., & Kastner, S. (2014). A multi-level account of selective attention. In A. C. Nobre & S. Kastner (Eds.), *The Oxford handbook of attention* (pp. 76–104). Oxford University Press.
- Sigman, M., Pan, H., Yang, Y., Stern, E., Silbersweig, D., & Gilbert, C. D. (2005). Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron*, 46(5), 823-835. <https://doi.org/10.1016/j.neuron.2005.05.014>
- Silins, N. (2015). Perceptual experience and perceptual justification.
- Silvanto, J., Schwarzkopf, D. S., Gilaie-Dotan, S., & Rees, G. (2010). Differing causal roles for lateral occipital cortex and occipital face area in invariant shape recognition. *European Journal of Neuroscience*, 32(1), 165-171. <https://doi.org/10.1111/j.1460-9568.2010.07278.x>
- Siman-Tov, T., Granot, R. Y., Shany, O., Singer, N., Hendler, T., & Gordon, C. R. (2019). Is there a prediction network? Meta-analytic evidence for a cortical-subcortical network likely subserving prediction. *Neuroscience & Biobehavioral Reviews*, 105, 262-275. <https://doi.org/10.1016/j.neubiorev.2019.08.012>
- Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in cognitive sciences*, 9(1), 16-20. <https://doi.org/10.1016/j.tics.2004.11.006>
- Sliwinska, M. W., Bearpark, C., Corkhill, J., McPhillips, A., & Pitcher, D. (2020). Dissociable pathways for moving and static face perception begin in early visual cortex: Evidence from an acquired prosopagnosic. *Cortex*, 130, 327-339. <https://doi.org/10.1016/j.cortex.2020.03.033>
- Sobotka, S., & Ringo, J. L. (1994). Stimulus specific adaptation in excited but not in inhibited cells in inferotemporal cortex of macaque. *Brain research*, 646(1), 95-99. [https://doi.org/10.1016/0006-8993\(94\)90061-2](https://doi.org/10.1016/0006-8993(94)90061-2)
- Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and cognition*, 112, 92-97. <https://doi.org/10.1016/j.bandc.2015.11.003>
- St. John-Saaltink, E., Utzerath, C., Kok, P., Lau, H. C., & De Lange, F. P. (2015). Expectation

References

- suppression in early visual cortex depends on task set. *PLoS One*, 10(6), e0131172.
<https://doi.org/10.1371/journal.pone.0131172>
- Stokes, D., & Biggs, S. (2014). The dominance of the visual. *Perception and its Modalities*, 350-378.
- Strother, L., Coros, A. M., & Vilis, T. (2016). Visual cortical representation of whole words and hemifield-split word parts. *Journal of cognitive neuroscience*, 28(2), 252-260.
https://doi.org/10.1162/jocn_a_00900
- Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in cognitive sciences*, 13(9), 403-409.
<https://doi.org/10.1016/j.tics.2009.06.003>
- Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science*, 314(5803), 1311-1314. <https://doi.org/10.1126/science.1132028>
- Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature neuroscience*, 11(9), 1004-1006. <https://doi.org/10.1038/nn.2163>
- Summerfield, C., Wyart, V., Mareike Johnen, V., & De Gardelle, V. (2011). Human scalp electroencephalography reveals that repetition suppression varies with expectation. *Frontiers in Human Neuroscience*, 5, 67.
<https://doi.org/10.3389/fnhum.2011.00067>
- Todorovic, A., & de Lange, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *Journal of Neuroscience*, 32(39), 13389-13395.
<https://doi.org/10.1523/JNEUROSCI.2227-12.2012>
- Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *Journal of Neuroscience*, 31(25), 9118-9123.
<https://doi.org/10.1523/JNEUROSCI.1425-11.2011>
- Tovée, M. J. (1996). An introduction to the visual system. Cambridge University Press.

References

- Trapp, S., Parr, T., Friston, K., & Schröger, E. (2021). The predictive brain must have a limitation in short-term memory capacity. *Current Directions in Psychological Science*, 30(5), 384-390. <https://doi.org/10.1177/09637214211029977>
- Travis, S. L., Mattingley, J. B., & Dux, P. E. (2013). On the role of working memory in spatial contextual cueing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(1), 208. <https://doi.org/10.1037/a0028644>
- Treue, S. (2003). Visual attention: the where, what, how and why of saliency. *Current opinion in neurobiology*, 13(4), 428-432. [https://doi.org/10.1016/s0959-4388\(03\)00105-3](https://doi.org/10.1016/s0959-4388(03)00105-3)
- Ungerleider, L. G. (1982). Two cortical visual systems. *Analysis of visual behavior*, 549-586. MIT Press.
- Vinken, K., Op de Beeck, H. P., & Vogels, R. (2018). Face repetition probability does not affect repetition suppression in macaque inferotemporal cortex. *Journal of Neuroscience*, 38(34), 7492-7504. <https://doi.org/10.1523/jneurosci.0462-18.2018>
- Vogels, R. (2016). Sources of adaptation of inferior temporal cortical responses. *Cortex*, 80, 185-195. <https://doi.org/10.1016/j.cortex.2015.08.024>
- Vuilleumier, P., Henson, R. N., Driver, J., & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature neuroscience*, 5(5), 491-499. <https://doi.org/10.1038/nn839>
- Wacongne, C., Changeux, J. P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience*, 32(11), 3665-3678. <https://doi.org/10.1523/JNEUROSCI.5003-11.2012>
- Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences*, 108(51), 20754-20759. <https://doi.org/10.1073/pnas.1117807108>
- Wada, Y., & Yamamoto, T. (2001). Selective impairment of facial recognition due to a haematoma restricted to the right fusiform and lateral occipital region. *Journal of*

References

- Neurology, Neurosurgery & Psychiatry*, 71(2), 254-257.
<https://doi.org/10.1136/jnnp.71.2.254>
- Wang, B., Guo, J., Yan, T., Ohno, S., Kanazawa, S., Huang, Q., & Wu, J. (2016). Neural responses to central and peripheral objects in the lateral occipital cortex. *Frontiers in human neuroscience*, 10, 54. <https://doi.org/10.3389/fnhum.2016.00054>
- Weigelt, S., Muckli, L., & Kohler, A. (2008). Functional magnetic resonance adaptation in visual neuroscience. *Reviews in the Neurosciences*, 19(4-5), 363-380.
<https://doi.org/10.1515/revneuro.2008.19.4-5.363>
- Weiner, K. S., & Grill-Spector, K. (2012). The improbable simplicity of the fusiform face area. *Trends in Cognitive Sciences*, 16(5), 251-254.
<https://doi.org/10.1016/j.tics.2012.03.003>
- Whitmire, C. J., & Stanley, G. B. (2016). Rapid sensory adaptation redux: a circuit perspective. *Neuron*, 92(2), 298-315.
<https://doi.org/10.1016/j.neuron.2016.09.046>
- Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current opinion in neurobiology*, 8(2), 227-233.
[https://doi.org/10.1016/s0959-4388\(98\)80144-x](https://doi.org/10.1016/s0959-4388(98)80144-x)
- Xu, J., Moeller, S., Auerbach, E. J., Strupp, J., Smith, S. M., Feinberg, D. A., ... & Uğurbil, K. (2013). Evaluation of slice accelerations using multiband echo planar imaging at 3 T. *Neuroimage*, 83, 991-1001. <https://doi.org/10.1016/j.neuroimage.2013.07.055>
- Xu, X., Biederman, I., & Shah, M. P. (2014). A neurocomputational account of the face configural effect. *Journal of Vision*, 14(8), 1-9. <https://doi.org/10.1167/14.8.9>
- Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex*, 15(8), 1234-1242. <https://doi.org/10.1093/cercor/bhi006>
- Xue, G., Chen, C., Jin, Z., & Dong, Q. (2006). Language experience shapes fusiform activation when processing a logographic artificial language: an fMRI training study. *Neuroimage*, 31(3), 1315-1326.
<https://doi.org/10.1016/j.neuroimage.2005.11.055>
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of experimental psychology*, 81(1),

References

141. <https://doi.org/10.1037/h0027474>
- Young, A. W., & Burton, A. M. (2017). Recognizing faces. *Current Directions in Psychological Science*, 26(3), 212-217. <https://doi.org/10.1177/0963721416688114>
- Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational Information in Face Perception. *Perception*, 16(6), 747-759. <https://doi.org/10.1068/p160747>
- Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current biology*, 15(24), 2256-2262. <https://doi.org/10.1016/j.cub.2005.10.072>
- Zago, L., Fenske, M. J., Aminoff, E., & Bar, M. (2005). The rise and fall of priming: how visual exposure shapes cortical representations of objects. *Cerebral Cortex*, 15(11), 1655-1665. <https://doi.org/10.1093/cercor/bhi060>
- Zeki, S., & Shipp, S. (1988). The functional logic of cortical connections. *Nature*, 335(6188), 311-317. <https://doi.org/10.1038/335311a0>
- Zhang, B., He, S., & Weng, X. (2018). Localization and functional characterization of an occipital visual word form sensitive area. *Scientific reports*, 8(1), 1-11. <https://doi.org/10.1038/s41598-018-25029-z>
- Zhou, Z., Vilis, T., & Strother, L. (2019). Functionally separable font-invariant and font-sensitive neural populations in occipitotemporal cortex. *Journal of Cognitive Neuroscience*, 31(7), 1018-1029. https://doi.org/10.1162/jocn_a_01408
- Zucker, R. S., & Regehr, W. G. (2002). Short-term synaptic plasticity. *Annual review of physiology*, 64(1), 355-405. <https://doi.org/10.1146/annurev.physiol.64.092501.114547>

List of figures

NO.	SECTION	SHORT DESCRIPTION	PAGE
1	Introduction	Figure 1. Schematic illustration of hierarchical network for predictive coding with feedback and feedforward information flow.	5
2	Introduction	Figure 2. Illustration of the predictive coding model.	7
3	Introduction	Figure 3. Examples for experimental designs investigating expectation suppression.	14
4	Introduction	Figure 4. Schematic illustration of four different neuronal models for repetition suppression.	19
5	Introduction	Figure 5. Schematic illustration of the two-stage model of response suppression in the ventral visual stream.	21
6	Introduction	Figure 6. Mapping the predictive coding scheme onto a canonical microcircuit.	22
7	General methods	Figure 7. Schematic illustration of the three-visual pathway model.	37
8	General methods	Figure 8. Relative spatial position of ROIs in a typical participant's inflated brain.	38
9	Study I	Fig. 1. The BOLD results of the VWFA for German participants.	50
10	Study I	Fig. 2. The BOLD results of the LO for German participants.	51
11	Study I	Fig. 3. The BOLD results of the VWFA for Chinese participants.	52
12	Study I	Fig. 4. The BOLD results of the LO for Chinese participants.	53
13	Study I	Fig. 5. Examples of the word stimuli and Schematic illustration of the experimental design and task procedures.	55
14	Study I	Fig. 6. Group-level ROIs in German and Chinese participants.	56
15	Study II	Figure 1. Schematic illustration of the experimental design and task procedures.	69
16	Study II	Figure 2. The accuracies and reaction times of the four-way category discrimination task during the ten-day training session.	72
17	Study II	Figure 3. Average peak activation profiles and HRF curves for faces.	74
18	Study II	Figure 4. Average peak activation profiles and HRF curves for cars	77
19	Study III-3a	Fig. 1. Schematic illustration of the trial structure.	89

List of figures

NO.	SECTION	SHORT DESCRIPTION	PAGE
20	Study III-3a	Fig. 2. Accuracy (A) and reaction time (B) for the VSTM and control conditions, separately for alternated (black) and repeated (grey) trials.	90
21	Study III-3a	Fig. 3. The BOLD signal of the right FFA for the VSTM and control conditions, separately for alternating and repeated trials.	90
22	Study III-3b	Figure 1. Performance (A, C) and reaction times (B, D) of Experiment 3b for different visual short-term memory conditions (VSTM and Control) separately for the non-target trials (A, B) and for the attentional target trials (C, D).	98
23	Study III-3b	Figure 2. The BOLD signal of the right FFA for the VSTM and Control conditions, separately for alternating and repeated trials when attention is controlled.	99
24	Discussion	Figure 9. Schematic illustration of the preliminary model of the relationship between experience and P(rep) effect.	110
25	Discussion	Figure 10. Schematic illustration of a hierarchical network for predictive coding with memory and attention.	114

List of tables

List of tables

NO.	SECTION	SHORT DESCRIPTION	PAGE
1	Study I	Supplementary Table 1. Individual-participant VWFA and LO coordinates in German and Chinese participants	59
2	Study I	Supplementary Table 2. Summary of significant activations identified from the group level whole-brain analysis	60

List of Abbreviations

A

AB alternation block
AG..... angular gyrus
At alternation trial
ANOVA analysis of variance
aSTS anterior superior temporal sulcus area

B

BF..... Bayes factors
BOLD..... blood oxygen level dependent

C

Control_Alt alternation trials in the visual short-term memory condition
Control_Rep repetition trials in the visual short-term memory condition

D

DCM dynamic causal modeling
DICOM digital imaging and communications in medicine format

E

E error units
ECoG electrocorticography
EEG electroencephalography
ERP event-related potential
ES expectation suppression
EVC early visual cortex

F

FC frontal cortex
FFA..... fusiform face area
fMRI..... functional magnetic resonance imaging
FWE..... family-wise error
FWHM full-width at half-maximum

G

GLM general linear model

H

HRF hemodynamic response function
hMT+..... human motion-selective complex

I

ISI inter-stimulus interval
ITI inter-trial interval

L

LGN..... lateral geniculate nucleus
LFA letter form area
lFFA left fusiform face area

List of Abbreviations

<i>LO</i>	lateral occipital cortex
<i>lLO</i>	left lateral occipital cortex
<i>LOC</i>	lateral occipital complex
<i>lOFA</i>	left occipital face area
M	
<i>M</i>	mean
<i>MEG</i>	magnetoencephalography
<i>mFFA</i>	middle fusiform face area
<i>MMN</i>	mismatch negativity
<i>MNI</i>	Montreal Neurological Institute
<i>MP-RAGE</i>	magnetization prepared rapid gradient echo
<i>MRI</i>	magnetic resonance imaging
<i>MT</i>	middle temporal visual area
<i>MVPA</i>	multivariate pattern analysis
O	
<i>OFA</i>	occipital face area
<i>OWA</i>	occipital word sensitive area
P	
<i>PC</i>	predictive coding
<i>pFFA</i>	posterior fusiform face area
<i>PL</i>	predictive learning
<i>P(rep)</i>	repetition probability
<i>pSTS</i>	posterior superior temporal sulcus area
R	
<i>R</i>	representation units
<i>RB</i>	repetition block
<i>RE</i>	repetition enhancement
<i>rFFA</i>	right fusiform face area
<i>rLO</i>	right lateral occipital cortex
<i>rOFA</i>	right occipital face area
<i>ROI</i>	region of interest
<i>RS</i>	repetition suppression
<i>RT</i>	reaction time
<i>Rt</i>	repetition trial
S	
<i>S1</i>	first stimulus
<i>S2</i>	second stimulus
<i>SD</i>	standard deviation
<i>SE</i>	standard error
<i>STM</i>	short-term memory
<i>STS</i>	superior temporal sulcus
T	
<i>TE</i>	time of echo
<i>TMS</i>	transcranial magnetic stimulation

List of Abbreviations

TRrepetition time

V

V1 primary visual cortex

V2 secondary visual cortex

V5 motion-selective area

VSTM visual short-term memory

VSTM_Alt alternation trials in the visual short-term memory condition

VSTM_Rep repetition trials in the visual short-term memory condition

Author contributions

Funding: All experiments in this Dissertation were supported by a Deutsche Forschungsgemeinschaft (DFG) Grant (grant number KO 3918/5-1). Chenglin Li was supported by China Scholarship Council (CSC) scholarship (201808330399) during the study. Study III was supported by a DFG Grant (grant number TR 1543/2-1) as well.

Study I. Repetition probability effects for Chinese characters and German words in the visual word form area

The idea of the first study was developed by Prof. Dr. Gyula Kovács and me. I programmed the software. Dr. Sophie-Marie Rostalski gave me the help of initial setting of fMRI sequence. I conducted the experiment and collected the data presented in this study. Marie Wächter and Wenbo Wang gave me the help with stimuli collecting and participants' recruitment. I analyzed the data under supervision of Prof. Dr. Gyula Kovács. The manuscript was written by me and revised by Prof. Dr. Gyula Kovács. I submitted the manuscript and made all the revisions with Prof. Dr. Gyula Kovács.

Study II. Short-term training attenuates the repetition probability effect for non-face objects

The idea of the second study was developed by Prof. Dr. Gyula Kovács and me. With the help of Marie Wächter and Lukas Korn, I collected and processed stimulus material. I wrote the experimental program and collected both the behavioral data in training sessions and fMRI data under initial guidance from Dr. Sophie-Marie Rostalski. I performed all analysis. I wrote the manuscript under the guidance of Prof. Dr. Gyula Kovács. I submitted the manuscript. Linda Ficco gave some useful comments.

Study III. 3a: Visual short-term memory load modulates repetition related fMRI signal adaptation

3b: Active attention counteracts the effects of increased load on visual short-term

Author contributions

memory on repetition suppression

The third study is a cooperation between Dr. Sabrina Trapp and Prof. Dr. Gyula Kovács. The idea was developed by them. I wrote the program and implemented it with the fMRI sequence under supervision of Dr. Sophie-Marie Rostalski. I collected the data with the help of Marie Wächter for participant's recruitment. The results presented in this article were analyzed by me. Dr. Sabrina Trapp, Prof. Dr. Gyula Kovács and me wrote and revised the manuscript of Study 3a. Dr. Sabrina Trapp submitted it to journal. I wrote the first draft of the manuscript of Study 3b. Prof. Dr. Gyula Kovács help me with all revisions. Hannah Klink and Linda Ficco gave some useful comments.



Chenglin Li

Acknowledgements

Today, 1261 days after I arrived in Germany. I am full with the complex mood when I am writing this section, including feeling grateful, happy, excited, etc. More importantly, there are many unforgettable episodes that come to my mind about I performed this work. I have to write down a few words to thank many people who pushed me to this point, although words might not be able to express all of my appreciation.

First and foremost, I would like to thank my super nice supervisor Prof. Dr. Gyula Kovács for giving me an opportunity to start my PhD projects in his research team and lovely department. I still remembered the situation that I got positive feedback when I sent my first email to you for searching for PhD position. Then, you became my supervisor. I have learned a lot from you. I thank you for all your guidance, advice, ideas, research insights, and patience – especially with my English. You encouraged me not only in this thesis but also in other projects, especially, you give me feedback very faster. I think it makes me more confident and braver to talk research questions with others. In addition, thank you for your efforts in connecting some cooperation for me. I believe all of these ‘PRIOR EXPERIENCES’ will affect my ‘EXPECTED WORKS’ as well. Furthermore, a special thanks to Prof. Dr. Stefan Schweinberger for investing his time to evaluate my work, I think it is my honor that you are willing to be my examiner.

I would like to thank other Colleagues in BPCN group, Charlotta Eick, Sophie Rostalski, Hannah Klink, Géza Ambrus, Kathrin Wiese. I really miss the happy time of we shared a room before Covid-19. Especially, Thank Sophie to guide me on how to set MRI sequences, how to analyze data in SPM, etc. Your help is crucial for my work. Thank Charlotta to help me check the grammar in her busy schedule and encourages me. Thank you, Hannah, help me to translate the summary to German and check the English grammar. Thank you, Kathrin, help me with lots of office works, such as ordering my PC. Géza, I still learn a lot from you, such as your passion for Python. All of my colleagues, even if we will work in different places, just hope we still can keep post.

Acknowledgements

I would also like to thank other temporary members or HiWis in our group, Alexia Dalski, Johannes Lehnert, Klemens Buttgerit, Lukas Korn, Marie Wächter, Matthäus Rudolph, Rico Stecher and Wenbo Wang. I enjoyed working with you. Especially, Marie and Wenbo give me lots of help for recruiting and arranging participants. At the same time, I want to thank all participants in this work, you give an important contribution to science. A special thanks go to Ines, who stayed with me for each participant in the scanner center.

I want to thank others who do not belong to our department as well, Kristina Ogonyants, Bettina Kamchen, and Kathrin Rauscher. You give me lots of help even it is not a part of this work. Especially, Kristina helped me solve lots of problems when I arrived in Jena, such as search a flat. I want to thank other PhD students of our brilliant PhD-Club, who share our projects and experience in daily struggles with Science. I really enjoy talking with you. A special thanks to Linda Ficco for her help me with checking the grammar and giving me lots of helpful feedback, Italian-Chinese friendship forever.

Last but not least, I am very grateful for my family and friends. You always support me in different ways. My dear grandmother, I am so sad that I cannot say goodbye to you, and you cannot see my thesis, just hope you are doing well in paradise I will miss you forever. My dear parents, I will be back in China and stay with you soon. My parents-in-law, thank you give me the best gift in the world. My brother and sister, thank you for always concerning with my daily life in Germany. A very very very special thank you goes to my great wife; you always stand my back and support me without any conditions, love you forever. I would also like to thank you, my daughter, you give me lots of happiness and motivated me further. My friends, please wait for me. I will visit you when I am back.

A short sentence to myself: this dissertation is just a start, please try your best to do anything in the future. Keep Work Moving Forward!

Chenglin Li

Jena/2022.3.17

Statutory declaration

I, Chenglin Li, hereby confirm that I am familiar with the current Course of Examination for doctoral candidates of the Faculty of Social and Behavioral Sciences.

I declare that the dissertation 'Prior experience modulates top-down predictive processing in the ventral visual areas' is a result of my personal work and that no other than the indicated sources have been used for its completion. Furthermore, I assure that all quotations and statements that have been inferred literally or in a general manner from published or unpublished writings are marked as such. In the selection and evaluation of the material as well as the production of the manuscripts, I was supported by the co-authors (free of charge), I have stated in the section "Contributions of authors". This work has not been used, neither completely nor in parts, for academic grading or to pass any previous examination or has been handed in as a dissertation in another university. I am aware of the current doctoral regulations of the Faculty of Social and Behavioral Sciences. The assistance of a commercial doctoral agency has not been utilized and no paid services related to the content of the submitted thesis have been used.

Jena, 15. 03.2022



Chenglin Li

Statutory declaration

Curriculum vitae

The CV is not included in this electronic version for protecting personal data.