Neural correlates of the use of prior knowledge in predictive coding

Vom Fachbereich Biologie der Technischen Universität Darmstadt
zur
Erlangung des akademischen Grades
eines Doctor rerum naturalium
genehmigte Dissertation von
M.Sc. Interdisciplinary Neuroscience
Alla Brodski-Guerniero
aus Kiew (Ukraine):

1. Referent: Prof. Dr. Ralf Galuske, Technische Universität Darmstadt, Darmstadt
2. Referent: PD Dr. Matthias Munk, Technische Universität Darmstadt, Darmstadt
3. Referent: Prof. Dr. Michael Wibral, Goethe-Universität, Frankfurt am Main

Tag der Einreichung: 15.11.2017
Tag der mündlichen Prüfung: 26.01.2018

Darmstadt 2018
D 17

„Das Gedächtnis ist unser wichtigstes Sinnesorgan.“ – Gerhard Roth

Моей бабушке (Für meine Oma)

Acknowledgements

The research presented in this thesis was carried out at the MEG Unit in the Brain Imaging Center of the Goethe University in Frankfurt.

I would like to thank Prof. Dr. Michael Wibral for the supervision of my thesis and the opportunity to be part of his group and to conduct my research at the MEG Unit.

I would also like to thank Prof. Dr. Ralf Galuske for being my supervisor at the technical university of Darmstadt and PD Dr. Matthias Munk for agreeing to be co-supervisor of this thesis.

Thanks to my colleagues at the MEG Unit: Michael Schaum, Patricia Wollstadt, Cerisa Stawowsky, Ivo Peres, Georg-Friedrich Paasch, Nina Merkel, Anya Dietrich, Yuranny Cabral-Calderin and Edoardo Pinzuti. You created a really nice atmosphere in the lab, in which everybody was helping out the others. I am happy to work with you!

I would like to thank the Ernst Ludwig Ehrlich Studienwerk (ELES) for financial funding and ideational support during the research for this thesis.

Thank you, to PD Bernhard Gaese for helpful advices whenever it became necessary.

I would also like to thank all co-authors of the publications in this thesis and also all the participants of the MEG studies.

Specially, I would like to thank my husband Salvatore and also my friends and family. Your support makes everything a lot easier!

Abstract

Every day, we use our sensory organs to perceive the environment around us. However, our perception not only depends on sensory information, but also on information already present in our brains, i.e. prior knowledge acquired by previous experience. The idea that prior knowledge is required for efficient perception goes back to Hermann von Helmholtz (1867). He raised the hypothesis that perception is a knowledge-driven inference process, in which prior knowledge allows to infer the (uncertain) causes of our sensory inputs. According to the currently very prominent “predictive coding theory” (e. g. Rao and Ballard, 1999; Friston, 2005, 2010; Hawkins and Blakeslee, 2005; Clark, 2012; Hohwy, 2013) this inference process is realized in our brains by using prior knowledge to build internal predictions for incoming information.

Despite the increasing popularity of predictive coding theory in the last decade (see Clark, 2012 and comments to his article), previous research in the field has left out several important aspects: 1. The neural correlates of the use of prior knowledge are still widely unexplored; 2. Neurophysiological evidence for the neural implementation of predictive coding is limited and 3. Assumption-free approaches to study predictive coding mechanism are missing.

In the present work, I try to fill these gaps using three studies with magnetoencephalographic (MEG) recordings in human participants:

Study 1 (n = 48) investigates how prior knowledge from life-long experience influences perception. The results demonstrate that prediction errors induced by the violation of predictions based on life-long experience with faces are reflected in increased high-frequency gamma band activity (> 68 Hz).

For studies 2 and 3, neurophysiological analysis is combined with information-theoretic analysis methods. These allow investigating the neural correlates of predictive coding with only few prior assumptions. In particular, the information-theoretic measure active information storage (AIS; Lizier et al., 2012; Wibral et al., 2014) can quantify how much information is maintained in neural activity (predictable information). I use AIS in order to study the neural correlates of activated prior knowledge in study 2 and 3.

Study 2 (n = 52) assesses how prior knowledge is pre-activated in task relevant states to become usable for predictions. I find that pre-activation of prior knowledge for predictions about faces increases alpha and beta band related predictable information as measured by AIS in content specific brain areas.

Study 3 (n patients = 19; n controls = 19) explores whether predictive coding related mechanism are impaired in autism spectrum disorder (ASD). The results show that alpha and beta band related predictable information is reduced in the brain of ASD patients, in particular in the posterior part of the default mode network. These findings indicate reduced use or precision of prior knowledge in ASD.

In summary, the results presented in the present work illustrate the neural correlates of the use of prior knowledge in the predictive coding framework. They provide neurophysiological evidence for the link of prediction errors and fast neural activity (study 1, gamma band) as well as predictions and slower neural activity (study 2 and 3, alpha and beta band). These findings are in line with a theoretical proposal for the neural implementation of predictive coding theory (Bastos et al., 2012). Further, by application of AIS analysis (study 2 and 3) the present work introduces the largely assumption-free usage of information-theoretic measures to study the neural correlates of predictive coding in the human brain. In future, analysis of predictable information as measured by AIS may be applied to a broad variety of experiments studying predictive coding and also for research on neuropsychiatric disorders as has been demonstrated for ASD.

Table of contents

Acknowledgements V

Abstract VII

Table of contents IX

List of figures XIII

List of publications with contributions XVII

1. General introduction 19
   - 1.1. Prior knowledge has a key role in perception 19
   - 1.2. Perception as knowledge-driven inference 20
   - 1.3. History of the term “predictive coding” 21
   - 1.4. Basic principles of predictive coding theory 21
   - 1.5. Predictive coding models for autism spectrum disorder 23
   - 1.6. Neural implementation of predictive coding 24
   - 1.7. Shortcomings of previous research in the predictive coding field and contribution of present research 25
     - 1.7.1. Neural correlates of the use of prior knowledge in predictive coding are still widely unexplored 26
     - 1.7.2. Neurophysiological evidence for the neural implementation of predictive coding is limited 27
     - 1.7.3. Assumption-free approaches to study predictive coding algorithms are missing 28

2. The faces of predictive coding 31
   - 2.1. Abstract 31
   - 2.2. Introduction 32
   - 2.3. Methods 34
     - 2.3.1. Experimental strategy 34
     - 2.3.2. Subjects 35
     - 2.3.3. Stimuli 35
     - 2.3.4. Stimulus presentation 36
     - 2.3.5. Task and Instructions 36
     - 2.3.6. Data acquisition and exclusion criteria 37
     - 2.3.7. Statistical analysis of behavioral data 38
     - 2.3.8. MEG data analysis 39
   - 2.4. Results 46
     - 2.4.1. Behavioral analysis 46
     - 2.4.2. Neural responses 48
   - 2.5. Discussion 56

2.5.1. Violations decrease accuracy and increase reaction times 56
2.5.2. Cortical source power changes in high-frequency GBA reflect PEs 57
2.5.3. GBA additionally reflects attentional effects 59
2.5.4. Increased GBA for violations is associated with slower processing. 59
2.5.5. Conclusion 60
3. Information-theoretic evidence for predictive coding in the face-processing system 61
3.1. Abstract... 61
3.2. Significance statement.. 62
3.3. Introduction 62
3.4. Methods 64
3.4.1. Basic concept and testable hypotheses 64
3.4.2. Subjects 66
3.4.3. Stimuli and stimulus presentation.. 67
3.4.4. Task and Instructions.. 68
3.4.5. Data acquisition.. 69
3.4.6. Statistical analysis of behavioral data.. 70
3.4.7. MEG-data preprocessing 70
3.4.8. Sensor level spectral analysis 71
3.4.9. Source grid creation 71
3.4.10. Source time course reconstruction.. 72
3.4.11. Definition of active information storage.. 72
3.4.12. Analysis of predictable information using active information storage 73
3.4.13. AIS Statistics 74
3.4.14. Correlation analysis of spectral properties and AIS 75
3.4.15. Correlation analysis of reaction times and AIS 75
3.4.16. Decoding analysis 76
3.4.17. Definition of transfer entropy (and Granger analysis).................... 77
3.4.18. Analysis of information transfer using transfer entropy and Granger causality analysis 78
3.5. Results 79
3.5.1. Behavioral results.. 79
3.5.2. Definition of frequency bands.. 80
3.5.3. Analysis of predictable information.. 81
3.5.4. Correlation of single trial power and single trial predictable information 82
3.5.5. Decoding prediction content from single trial AIS values................. 84
3.5.6. Analysis of information transfer 84

3.5.7. Correlation of predictable information and reaction times 85
3.6. Discussion 87
3.6.1. Activated prior knowledge for faces shows as predictable information in content specific areas 87
3.6.2. Maintenance of activated prior knowledge about faces is reflected by increased alpha/beta power 89
3.6.3. Face predictions are transferred in a top-down manner 89
3.6.4. Pre-activation of prior knowledge about faces facilitates performance 91
4. Predictable information is reduced in ASD – a predictive coding study 93
4.1. Abstract 94
4.2. Introduction 94
4.3. Methods 96
4.3.1. Participants 96
4.3.2. Assessment instruments across groups 97
4.3.3. Autism-specific assessment instruments 97
4.3.4. Data acquisition 98
4.3.5. MEG-data analysis 99
4.4. Results 104
4.4.1. Group characteristics 104
4.4.2. Analysis of average predictable information 105
4.4.3. Whole brain analysis of predictable information 106
4.4.4. Complexity analysis 108
4.4.5. Correlation of predictable information and power in individual epochs 109
4.4.6. Correlation of predictable information and ADI-R scores 110
4.5. Discussion 111
4.5.1. Average predictable information is reduced in ASD 112
4.5.2. Predictable information at PCC, SMG and Prec is reduced in ASD 112
4.5.3. Complexity is not reduced in PCC, SMG and Prec 113
4.5.4. Reduced predictable information in the light of predictive coding accounts of ASD 114
4.5.5. Predictable information in ASD is associated with symptom severity in the domain of restricted, repetitive, and stereotyped behaviors and interests (ADI-R rit) 116
4.5.6. Conclusion 117
4.5.7. Acknowledgements 117
5. General discussion 119
5.1. Neural correlates of the use of prior knowledge in predictive coding 119

    5.2. Neurophysiological evidence for the neural implementation of predictive coding 121
    5.3. An assumption-free approach to study predictive coding algorithms 124
    5.4. Limitations of the present work and outlook for future predictive coding research 127
    5.5. Conclusion 129
6. Bibliography........ 131
7. Zusammenfassung........ 145
8. Ehrenwörtliche Erklärung 149
9. Curriculum Vitae........ 151

List of figures

Figure 1.1. What is this? (Solution in the text); Roth (1997: 261- 263). 19
Figure 1.2. Schematic illustration of the model for neural implementation of predictive coding proposed in Bastos et al. (2012)... 25
Figure 2.1. Graphical depiction of stimulus timing and the five stimulus categories. 37
Figure 2.2. Behavioral analysis of hit rates and reaction times of correct responses. 48
Figure 2.3. Sensor level frequency analysis. 50
Figure 2.4. Statistical analysis on beamformer estimated MEG source power in the high gamma frequency range (68-144 Hz). 53
Figure 2.5. Post-hoc analysis on beamformer estimated MEG source power in the high gamma frequency range (68-144 Hz). 54
Figure 2.6. Correlation analysis for high-frequency gamma power and reaction times. 55
Figure 3.1. Central idea of the study and experimental design......................... 66
Figure 3.2. Behavioral results. 80
Figure 3.3. Sensor-level frequency analysis - defining frequency bands.......... 81
Figure 3.4. Statistical analysis of predictable information (measured by AIS) at the MEG source level. 82
Figure 3.5. Correlation of predictable information contrast maps and source power contrast maps. 84
Figure 3.6. Analysis of information transfer in the prestimulus interval 85
Figure 3.7. Correlation analysis for predictable information and reaction times. 86
Figure 4.1. Comparison of average AIS between groups. 106
Figure 4.2. Statistical comparison of AIS between groups at the MEG source level. 108
Figure 4.3. Post-hoc complexity analysis.. 109
Figure 4.4. Linear regression analysis of AIS and ADI-R scores. 111
Figure 5.1. Modified version of Bastos' proposal (2012) for a neural implementation of predictive coding (schematic). 125

List of tables

Table 2.1. Behavioral analysis: ANOVA and post-hoc test results 48
Table 2.2. Effects on beamformer reconstructed source power in the high gamma frequency range (68-144 Hz); corresponding time, anatomical regions and MNI coordinates of peak voxels 52
Table 3.1. Correlation of single trial power and single trial predictable information (measured by AIS) in the face prediction network 83
Table 4.1. Summary of group characteristics.. 104
Table 4.2. Correlation of epoch-by-epoch AIS values and power in PCC, SMG and Prec..110
Table 4.3. Peak voxels for the significant cluster in the regression analysis of AIS and ADI-R rit (see Figure 4.4.)..110

List of publications with contributions

1. *Brodski, A., Paasch, G. F., Helbling, S., & Wibral, M. (2015). The faces of predictive coding. Journal of Neuroscience, 35:8997-9006.*

Contributions: *AB* – study design, recruitment of participants, data acquisition, design of analysis pipeline, data analysis and writing of the manuscript; *GFP* – study design, recruitment of participants, data acquisition, design of analysis pipeline, data analysis and writing of the manuscript; *SH* – provision of novel statistical tools and description of these tools in the manuscript; *MW* – study design, design of analysis pipeline and manuscript revision.

2. *Brodski-Guerniero, A., Paasch, G. F., Wollstadt, P., Özdemir, I., Lizier, J. T., & Wibral, M. (2017). Information-theoretic evidence for predictive coding in the face-processing system. Journal of Neuroscience, 37:8273-8283.*

Contributions: *ABG* – study design, recruitment of participants, data acquisition, design of analysis pipeline, data analysis and writing of the manuscript, *GFP* – data acquisition; *PW* – provision of novel analytic tools and manuscript revision; *IÖ* – provision of novel analytic tools; *JTL* – provision of novel analytic tools and manuscript revision; *MW* – study design, design of analysis pipeline and manuscript revision.

3. *Brodski-Guerniero, A., Naumer, M., Moliadze, V., Chan, J., Althen, H., Ferreira-Santos F., Schlitt, S., Kitzerow, J., Schütz, M., Langer, A., Kaiser, J., Freitag C.M., Wibral, M. (2017). Predictable information is reduced in autism spectrum disorder: A predictive coding study. Submitted to Human Brain Mapping.*

Contributions: *ABG* - design of analysis pipeline, data analysis and writing of the manuscript; *MJN* - study design and manuscript revision; *VM* - study design and data acquisition; *JC* – study design and data acquisition; *HA* - phenotypic analysis and manuscript revision; *FFS* – discussion of analysis pipeline and manuscript revision; *SS* – diagnostic assessment; *JK* – diagnostic assessment; *MS* - recruitment of participants and data acquisition; *AL* – recruitment of participants and data acquisition; *JK* – study design and manuscript revision; *CMF* - study design and manuscript revision; *MW* – design of analysis pipeline and manuscript revision.

All published material has been reproduced with permission from the publisher.

1. General introduction

1.1. Prior knowledge has a key role in perception

These days it is a well-established notion that our perception is not exclusively based on environmental stimuli, but substantially depends on previous experience and contextual factors. A compelling example for this notion is given by the German neuroscientist Gerhard Roth (Figure 1.1.):

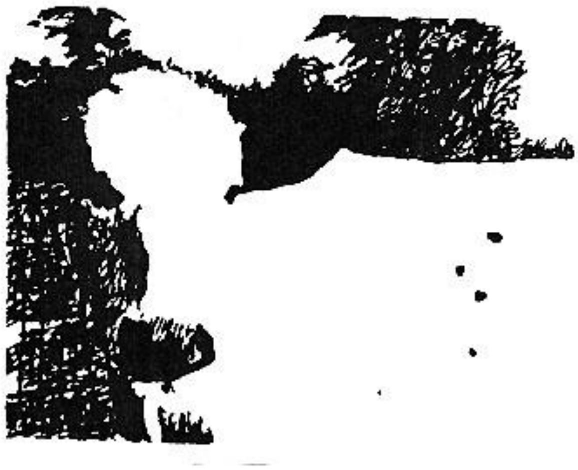

Figure 1.1. What is this? (Solution in the text); Roth (1997: 261- 263).

Most people have difficulties in recognizing what is shown in the picture when seeing it for the first time. After 10 or 20 minutes of exposure some can finally identify the animal portrayed in the picture, while others still remain unable to recognize it. However, when several hints are provided (e.g. think of “milk” and “Milka chocolate”) and after pointing on details of the outline, almost everyone is able to recognize the identity of the animal after a few minutes only. This can be explained by the fact that these hints activate pieces of knowledge in (visual) memory, which match to the information in the picture and thereby speed up the process of considering possible interpretations of the visual input (see e.g. Fenske et al., 2006). After repeated exposure to the picture, the recognition process becomes almost automatic and after an extensive exposure period, it becomes almost impossible to see anything else than a cow in the picture. Importantly, Gerhard Roth's example shows that a sensory input not resulting in any meaningful perception in the beginning, can transform into a stable and meaningful percept through experience. This parallels the well-established finding that previous experience with a stimulus facilitates perception – which is for instance demonstrated by impaired detection performance when objects are presented in unusual contexts (Biederman et al., 1973; see Bar, 2004 for a review).

1.2. *Perception as knowledge-driven inference*

The idea that prior knowledge from previous experience is required for efficient perception is not new; it can be traced back to Hermann von Helmholtz in the 1860s. Von Helmholtz proposed the idea of perception as unconscious knowledge-driven inference, in which the brain uses prior knowledge from memory to infer the (high-level) causes of its sensory inputs (Von Helmholtz, 1867). Helmholtz's idea is based on the general problem that the brain does not have direct access to the external causes resulting in its (sensory) inputs. Moreover, there is no one-to-one mapping from external causes to sensory inputs. This is exemplified by the fact that the retina has only limited possibilities of representation for an unlimited amount of causes. Consequently, the same representation on the retina can be caused by a lot of different causes. Additionally, the same cause can lead to different representations on the retina depending on the context (see e.g. Kersten et al., 2004, Figure 1 for an illustration of this problem).

It is assumed that the brain deals with this uncertainty by incorporating prior knowledge from previous experience in a (approximately) Bayesian fashion into the perceptual process (Kersten et al., 2004; Knill and Pouget, 2004). This allows the brain to combine the prior probability of a potential cause with a likelihood-term (probability of the observed sensory information given the cause) to build its posterior (probability of the cause given the observed sensory information) in order to determine the most probable cause for its sensory input. A promising candidate for realization of this knowledge driven inference in the brain is predictive processing (e. g. Mumford, 1992; Rao et al., 1999; Friston, 2005, 2010; Hawkins and Blakeslee, 2005; Clark, 2012; Hohwy, 2013), which became a highly prominent motif in neuroscience research during the last decade. Different variants of predictive processing in the brain have been suggested (see Spratling, 2017). The most popular variant is the (hierarchical) predictive coding theory (Rao and Ballard, 1999; Friston, 2005, 2010).

1.3. History of the term “predictive coding”

The term “predictive coding” was originally introduced in the context of television and radio transmission (Harrison, 1952; Oliver, 1952). In this context, it referred to an efficient strategy to transmit information over a channel with limited capacity. An ideal transmission strategy should get rid of signal redundancy and should also include a signal coding for which only a small dynamic range is required. To illustrate these requirements, television can be considered as an example. In television broadcasting, subsequent images are mostly very similar. So, instead of sending nearly the same image twice or more often it is more efficient to signal only the difference between the last and the actual image. Signaling only the difference and not the same and thus predictable part of the image constitutes a simple predictive code.

Such a predictive code was also proposed to be a basic principle in the brain: According to predictive coding theory the information passed feed-forward in the cortical hierarchy is limited to the difference between the predicted and the actual incoming information – the so called “prediction error” (Clark, 2012). The basic principles of predictive coding theory are described below.

1.4. Basic principles of predictive coding theory

Predictive coding theory proposes that the brain builds a generative model of the world based on the statistical regularities in the environment. This generative model maps from external causes to sensory consequences and can be inverted to predict the incoming information. The prediction with the best fitting to the incoming information is identified by an iterative process of prediction error minimization between hierarchical layers. Mathematically this can be formulated as finding the parameters of the generative model which minimize the sum of the squared error between the actual incoming information and the prediction. The prediction is represented in “prediction units” and is transmitted via feedback connections (top-down)[1] to areas lower in the cortical hierarchy, where predicted and actual information are compared. A potential mismatch between predicted and actual information is represented in “error units” and is transmitted as prediction error via cortical feed-forward connections (bottom-up)[2] to brain areas higher in the cortical hierarchy, where it may induce a modification or update of the original prediction. This procedure can be repeated in several loops until the prediction error is minimized and the most likely causes for the incoming information have been inferred (Rao and Ballard, 1999; Friston, 2005, 2010).

This principle has been recently expanded by the concept of precision-weighting (e.g. Friston, 2009; Feldman and Friston, 2010). According to this concept a stronger or weaker weighting of the sensory input or prediction error compared to the prediction depends on their corresponding “precision”. To give an example, a sound in a noisy environment would usually be not considered as reliable and would therefore not be associated with a strong precision, while the same sound in a silent environment would be potentially associated with a higher precision. This regulation of precision is also supposed to constitute a mechanism for the role of attention in the predictive coding framework (Friston, 2009; Feldman and Friston, 2010), allowing to boost the precision when attention is directed to a stimulus.

[1,2] Strictly speaking, the terms feedback and feed-forward refer to the anatomical connections, while the terms top-down and bottom-up refer to psychological concepts. However, for practical purposes these terms will be used interchangeably in this thesis.

Imbalances in the precision-weighting system may result in an excessive reliance on either top-down predictions or bottom-up incoming information and have been associated with neuropsychiatric disorders – in particular with autism spectrum disorder (ASD; Friston et al., 2013; Lawson et al., 2014).

1.5. Predictive coding models for autism spectrum disorder

A relation of impairments in patients with ASD and aberrant predictive coding mechanisms was first proposed by Pellicano and Burr (2012a). Specifically, Pellicano and Burr hypothesized that the prior knowledge (or in short, the “prior”) used for predictions is less precise in ASD patients, resulting in a reduced influence of prior knowledge on perception. According to this view, perception in ASD patients compared to healthy humans more heavily relies on the sensory information. This view is in line with hypotheses of a general reduction of top-down control in ASD (Frith, 2003).

Pellicano and Burr’s proposal started an intensive discussion in which most researchers acknowledged the general idea of impaired predictive coding mechanisms in ASD (Brock, 2012; Friston et al., 2013; Teufel et al., 2013; van Boxtel and Lu, 2013; Lawson et al., 2014; Van de Cruys et al., 2014). However, some of the follow-up articles questioned Pellicano and Burr’s theory of attenuated influence of priors in ASD (Brock, 2012; Teufel et al., 2013; Van de Cruys et al., 2014). Their authors suggested that rather the influence of bottom-up information (sensory input, prediction error) could be enhanced in ASD patients. This enhancement of bottom-up influences in ASD patients might be either caused by reduced sensory noise (Brock, 2012; however see Pellicano and Burr, 2012b) or by the unduly high precision of prediction errors (Van de Cruys et al., 2014). In turn, other accounts proposed that the use of top-down or bottom-up information is not unilaterally altered, but rather that the flexible precision-weighting of top-down and bottom-up information is abnormal in ASD (Friston et al., 2013, Lawson et al., 2014). According to these proposals, the excessive influence of prediction errors in ASD results from the failure to attenuate the sensory gain of prediction error units by top-down gain control (Lawson et al., 2014).

1.6. Neural implementation of predictive coding

Differentiating between competing models of predictive coding in ASD with the help of neurophysiological recordings requires knowledge of the neural implementation of predictive coding mechanisms in the human brain. A proper predictive coding model at the implementational level (see Marr, 1982), ought to combine the physiology and anatomy of the human cortex with the “ingredients” of predictive coding theory: In the predictive coding framework, prediction errors are propagated via feed-forward connections which are known to originate in superficial cortical layers. Consequently, “error units” signaling the prediction errors should be preferentially located at superficial cortical layers. On the other hand, predictions are propagated via feedback connections which are known to originate in deep cortical layers. Hence, “prediction units” signaling the predictions should be preferentially located at deep cortical layers. In addition to this spatial segregation, a spectral segregation of predictions and prediction errors is suggested. This is based on the spectral predominance of gamma frequencies in the superficial cortical levels and alpha and beta frequencies in the deep cortical layers (Roopun et al., 2006, 2008; Buffalo et al., 2011; Xing et al., 2012). Additionally, a spectral segregation is supported by physiological findings in monkeys (Bastos et al., 2015) and humans (Michalareas et al., 2016) linking feed-forward and feedback connections to information transfer in the gamma and alpha/beta frequency band, respectively.

In line with a spatial and spectral segregation of predictions and prediction errors, Bastos and colleagues recently proposed a theoretical model for the neural implementation for predictive coding (2012; Figure 1.2.).

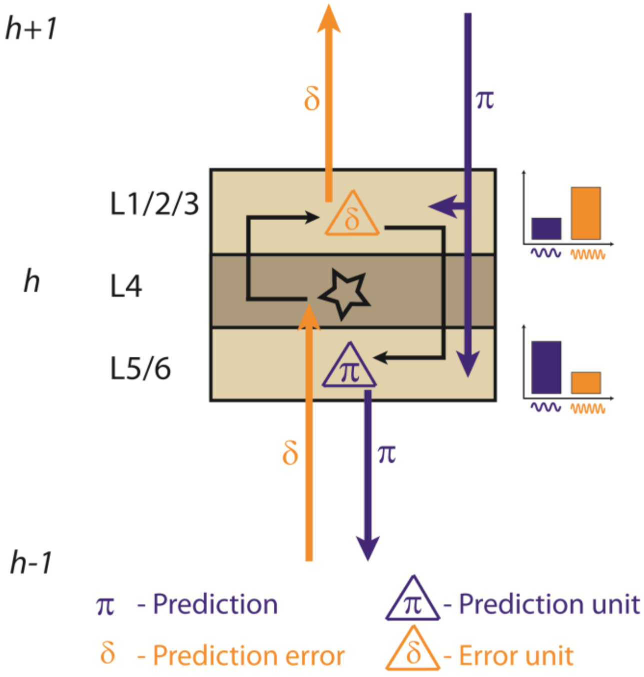

Figure 1.2. Schematic illustration of the model for neural implementation of predictive coding proposed in Bastos et al. (2012). Error units in the superficial cortical layers, in which high frequencies dominate, send prediction error signals in gamma frequencies to areas higher up in the cortical hierarchy (h-1 to h and h to h+1). Prediction units in the deep cortical layers, in which the low frequencies dominate, send prediction signals in beta frequencies to areas lower in the cortical hierarchy (h+1 to h and h to h-1). Orange indicates high frequencies, blue indicates low frequencies.

In Bastos' model, prediction errors express in fast neural activity (> 30 Hz, gamma frequency band) at superficial cortical layers, while predictions express at deep cortical layers in lower frequencies, presumably in the beta frequency band (~ 12 - 30 Hz).

1.7. Shortcomings of previous research in the predictive coding field and contribution of present research

Although predictive coding theory became a highly popular research topic in neuroscience within the last ten years (see e.g. Clark, 2012 and comments to his article), previous research in the field has not addressed several important aspects, yet:

1. The neural correlates of the use of prior knowledge in predictive coding are still widely unexplored.

2. Neurophysiological evidence for the neural implementation of predictive coding is limited.

3. Assumption-free approaches to study predictive coding algorithms are missing, i.e. many experimental tests of predictive coding theory rely on *ad hoc* beliefs about what the brain should actually predict in a given situation.

In the remainder of the general introduction I am going to discuss these shortcomings and describe how I addressed them in the present work.

1.7.1. The neural correlates of the use of prior knowledge in predictive coding are still widely unexplored

The use of prior knowledge is inevitably an essential part of predictive coding theory, as it facilitates the fundamental differentiation between predicted and unpredicted incoming information. It is also well known that the predictive coding principle can account for several behavioral and cognitive phenomena, in which prior knowledge plays a role – like priming, mismatch negativity, repetition suppression and binocular rivalry (Friston, 2005; Hohwy et al., 2008). Nevertheless, the neural correlates of the use of prior knowledge in predictive coding remain fairly unexplored to date. In the present work, I tried to fill this gap by conducting three studies to investigate the neural correlates of the use of prior knowledge in the predictive coding framework:

1. In study 1 (“The faces of predictive coding” published in *The Journal of Neuroscience*, 2015, chapter 2) I investigated how prior knowledge from life-long experience influences our perception. Hereby, I focused on the neural correlates of prediction errors – induced by a mismatch of sensory input and predictions based on prior knowledge about faces from life-long experience.

2. In study 2 (“Information-theoretic evidence for predictive coding in the face-processing system” published in *The Journal of Neuroscience*, 2017, chapter 3) I investigated how prior knowledge is activated in task relevant states to

become relevant for predictions. Here, I focused on the pre-activation of relevant prior knowledge for face predictions.

These two studies used Mooney stimuli (Mooney, 1957), which are (degraded) black and white versions of faces (and houses in study 2 only). Mooney stimuli are well suited to study the use of prior knowledge in the predictive coding context, as prior knowledge from memory is inevitably required for their successful recognition (Kemelmacher-Shlizerman et al., 2008)

3. Last, in study 3 ("Active information storage is reduced in autism spectrum disorder – a predictive coding study" submitted to *Human Brain Mapping*, chapter 4) I investigated whether the use of prior knowledge is reduced in ASD. Thereby, I tested the hypothesis that predictive coding related mechanisms are disturbed in ASD patients.

All three studies used magnetoencephalography (MEG) recordings. The very good spatial resolution of MEG allowed whole-brain reconstruction of time courses in source space; the excellent temporal resolution further allowed studying the neural correlates of the use of prior knowledge in neural source activity proper.

1.7.2. Neurophysiological evidence for the neural implementation of predictive coding is limited

Although Bastos' theory (2012) is to date the most relevant proposal for the neuronal implementation of predictive coding in the human cortex, neurophysiological evidence for Bastos' suggestion of a separate fast frequency channel for prediction errors and a slower frequency channel for predictions remains rare until now. Fortunately, the use of MEG recordings also enables performing a spectral analysis for the neural correlates of prediction errors and predictions, respectively – and thereby testing Bastos' hypothesis. Thus, in all of the three studies in the present work I also investigated whether the spectral profile of prediction errors or predictions (prior knowledge) was in line with

Bastos’ hypothesis of an association of prediction errors with high and predictions with lower frequencies.

1.7.3. Assumption-free approaches to study predictive coding algorithms are missing

When interested in the use of predictions, studying the neural correlates of predictive coding usually requires making assumptions about the particular brain areas being involved, and about what these should predict in a given situation. Partly, this information may be acquired from the literature in the field. For instance, van Pelt and colleagues (2016) studied the prediction of causal events based on a network of brain areas known to be involved in causal inference. However, defining brain areas based on other studies might be often misleading, as even small changes in the experimental design or setting can lead to an involvement of different brain areas. To overcome this problem, I used information-theoretic analysis methods for study 2 and 3, which allow investigating the neural correlates of predictive coding in terms of fundamental components of information dynamics, i.e. *information storage* and *information transfer*. This approach facilitates to describe the neural correlates of predictive coding with only few prior assumptions (Wibral et al., 2015) and allows to find the brain areas involved in representing the prior knowledge for predictions without a-priori defining these brain areas. In particular, the concept of *information storage* was of relevance for studying the activation of prior knowledge (chapter 3 and 4), while the concept of *information transfer* was of relevance for studying the propagation of predictions (chapter 3). The definition and application of these two information-theoretic concepts is outlined below:

1. *Information Storage*: We can differentiate between passive and active storage in the brain (Zipser et al., 1993). While passive storage refers to information stored in physiological parameters like synaptic weights, active storage refers to information maintained in neural activity. In the predictive coding framework, knowledge previously stored passively needs to become

activated and to be maintained in neural activity in order to be transferred to other brain areas and to predict the incoming information. This active type of storage can be measured with the information-theoretic measure active information storage (AIS). AIS measures the mutual information $I(X_{t-};X_t)$ between the past $X_{t-} = \{X_{t-1}, X_{t-2}, ...\}$, and present state X_t of a (neural) signal (see methods part in chapter 3 for details). AIS can quantify how much information for a given time step of a neural signal has been stored in its past state or, in other words, how much information is maintained in neural activity (predictable information). Thus, analysis of predictable information as measured by AIS was applied to quantify the amount of prior knowledge activated in neural activity (chapter 3 and 4, healthy controls and ASD patients).

2. *Information transfer*: In subsequent processing steps, activated prior knowledge may serve predictions which are transferred to other brain areas. To study this information transfer I used the information-theoretic measure transfer entropy (TE, Schreiber, 2000; Vicente et al., 2011; Wibral et al., 2011). TE measures the conditional mutual information $I(X_t;Y_{t-u}|X_{t-})$ between the future X_t of a (neural) target signal X and source signal Y conditional on the past of the target signal X_{t-}, where u is the physical delay from source to target (see methods part in chapter 3 for details). TE quantifies how much information is present in the target, which is already known from the source but new to the target. In other words, TE quantifies how much information has been transferred from source to target. This allowed us to study how predictions based on prior knowledge were transferred between brain areas (chapter 3).

The results of all studies are summarized and discussed in the general discussion (chapter 5).

2. The faces of predictive coding

Authors

Alla Brodski[a,1], Georg-Friedrich Paasch[a,1], Saskia Helbling[b], Michael Wibral[a]

Affiliations

[a] MEG Unit, Brain Imaging Center, J.W. Goethe University, 60528,

Frankfurt a.M., Germany.

[b] Institute of Medical Psychology, J.W. Goethe University, 60528,

Frankfurt a.M., Germany.

[1] These authors contributed equally to this work

Acknowledgements

We thank William Phillips for comments on an earlier version of this manuscript. This work was supported by grants from the LOEWE Neuronale Koordination Forschungsschwerpunkt Frankfurt (NeFF).

2.1. Abstract

Recent neurophysiological accounts of predictive coding hypothesized that a mismatch of prediction and sensory evidence – a prediction error (PE) – should be signaled by increased gamma band activity (GBA) in the cortical area where prediction and evidence are compared. This hypothesis contrasts with alternative accounts where violated predictions should lead to reduced neural responses.

We tested these hypotheses by violating predictions about face orientation and illumination direction in a Mooney face detection task, while recording magnetoencephalographic responses in a large sample of 48 human subjects. The investigated predictions – acquired via life-long experience – are known to be processed at different time-points and brain regions during face recognition.

Behavioral responses confirmed the induction of PEs by our task. Beamformer source analysis revealed an early PE signal for unexpected orientation in visual brain areas followed by a PE signal for unexpected illumination in areas involved in 3-D shape from shading and spatial working memory. Both PE signals were reflected by increases in high-frequency (68-140 Hz) GBA. In high-frequency GBA we observed also a late interaction effect in visual brain areas, probably corresponding to a high-level PE signal. In addition, increased high-frequency GBA for expected illumination was observed in brain areas involved in attention to internal representations.

Our results strongly support the hypothesis that increased GBA signals PEs. Additionally, GBA may represent attentional effects.

2.2. Introduction

The view of the brain as a “predictive machine” has gained considerable popularity in the last decade (e.g. Hawkins and Blankeslee, 2005; Clark, 2012; Hohwy, 2013). This notion implies that the brain relies on statistical regularities in the environment to construct internal predictions of its sensory inputs to facilitate perception. In many cases these statistical regularities are extracted from life-long experience and form priors residing in implicit long-term memory. Yet, the mechanisms underlying the integration of experience-based information and sensory evidence during the perceptual process are still a matter of debate (e.g. Mumford, 1992; Rao and Ballard, 1999; Kersten et al., 2004; Friston, 2005, 2010; Grossberg, 2007, 2012; Spratling, 2008; Kay and Phillips, 2011). Opposing theories propose either signal suppression (Grossberg, 2007, 2012;

Carpenter and Grossberg, 2010) or signal enhancement (Mumford, 1992; Rao and Ballard, 1999; Friston, 2005) in case of a mismatch of sensory evidence and information learned from previous experience.

According to predictive coding theory (Rao and Ballard, 1999) in particular, a mismatch between predictions based on priors from our experience and incoming information should result in a prediction error (PE), reflected by increased neural activity.

Anatomically, PEs are supposed to be propagated by feed-forward connections (Rao and Ballard, 1999), originating in superficial cortical layers (e.g. Barone et al., 2000). As gamma band activity (GBA) is prominent in the superficial layers of the cortical microcircuit (Buffalo et al., 2011; also see Wang, 2010 for a review), it has been suggested that the bottom-up propagation of PE signals is reflected in GBA (Arnal and Giraud, 2012; Bastos et al., 2012).

To test the hypothesis that PEs are reflected by increased neural activity – versus alternative accounts that favor suppression of activity in case of violated predictions, we used MEG because of its high temporal and spatial resolution. This enabled the investigation of timing, anatomical location and magnitude of PE signals at distinct hierarchical levels. Moreover, direct access to electrophysiological activity by MEG allowed us to specifically test whether GBA is the carrier of PE signals.

First evidence for PE signaling in GBA has been provided in recent MEG studies (Arnal et al., 2011; Todorovic et al., 2011; Bauer et al., 2014). The present study is however to our knowledge the first one to test this hypothesis for priors from life-long experience while providing the spatial resolution to investigate PEs at different hierarchical levels and a high statistical power due to the large sample size of 48 subjects (see Button et al., 2013 for a review on the problems caused by small sample sizes in neuroscience).

We induced PEs by a mismatch between the sensory input in a Mooney face (Mooney and Ferguson, 1951) detection task and predictions based on priors from lifelong visual experience. The investigated priors “upright face orientation” and “illumination from the top” are supposed to be processed in different brain areas as well as at different time-points during the face recognition process (Cavanagh, 1991), which we expect to be reflected by time-shifted PEs.

2.3. Methods

2.3.1. Experimental strategy

To investigate the neural correlates of PE signals we collected MEG responses while subjects performed a Mooney face detection task (Mooney and Ferguson, 1951). Mooney stimuli can not be recognized without relying on predictions based on priors from our life-long experience (Moore and Cavanagh, 1998; Kemelmacher-Shlizerman et al., 2008). Here, we focussed on two important priors for Mooney faces: First, faces normally appear in upright orientation (‘orientation prior’; Yin, 1969; Valentine, 1988). Second, a scene is normally illuminated by a single light source from the top (‘illumination prior’; Brewster, 1847; Sun and Perona, 1998; Adams, 2007; Gerardin et al., 2010).

To induce PEs, the presented stimuli were made incompatible with the orientation prior, the illumination prior or both priors. To this end, we presented upright (UP) or inverted (IN) Mooney faces illuminated from the top (TP) or from the bottom (BT) which resulted in a 2x2 full factorial design (factors orientation and illumination) with four Mooney face conditions: UPTP, UPBT, INTP and INBT. To counter a potential response bias, additional sham stimuli with matched image statistics were presented that did not contain a face.

In order to formulate hypotheses about the expected timing of neural PE responses we draw on a behaviourally well validated process model for Mooney face recognition by Cavanagh (1991). Cavanaghs model suggests that the stimulus orientation should be processed before the illumination direction is evaluated. Hence, we assume that the PE response for the violation of the orientation prior should precede the PE response for the violation of the illumination prior.

2.3.2. Subjects

59 subjects participated in the MEG experiment. Subjects had normal or corrected-to-normal visual acuity and were right handed according to the Edinburgh Handedness Inventory scale (Oldfield, 1971). Each subject gave written informed consent before the beginning of the experiment. Subjects were paid 10€ per hour. 11 subjects had to be excluded from further analysis; 1 subject was not able to tolerate the structural MRI scan; 5 subjects were excluded due to excessive movement or due to an insufficient amount of remaining trials after artefact rejection. 5 more subjects were excluded based on their behavioral performance (see exclusion criteria below). 48 participants (average age: 25.04 years, 22 males) remained and were considered for behavioral and neurophysiological analysis. The large sample size of 48 subjects was chosen to reduce the risk of false positives, as suggested by Button and colleagues (2013).

The local ethics committee (Johann Wolfgang Goethe University, Frankfurt, Germany) approved of the experimental procedure.

2.3.3. Stimuli

Two-tone images, known as Mooney face stimuli (Mooney, 1957), were created by transforming all shades of gray in photographs of upright (UP) faces into either black or white. To investigate the violation of the orientation prior, Mooney face orientation was inverted (IN). To investigate the violation of the illumination prior, the illumination source was set to light from the bottom (BT), while light from the top (TP) corresponded to the expected illumination direction.

There was no significant difference in average local luminance between any of the four Mooney face conditions ($p > 0.55$)

In addition, scrambled ‘No-Face’ stimuli (SCR) were created from each of the

Mooney face conditions by displacing white or black patches within the given background. Thereby all low-level information was maintained but the facial configuration disappeared. The scrambled stimuli served as sham stimuli to avoid a response bias towards detecting faces. Examples of the stimuli can be seen in Figure 2.1.

All stimuli were resized to a resolution of 510 x 650 pixels. All stimulus manipulations were performed with the program GIMP (GNU Image Manipulation Program, 2.4, free software foundation, Inc., Boston, Massachusetts, USA).

2.3.4. Stimulus presentation

A projector with a refresh rate of 60 Hz was used to display the stimuli at the center of a translucent screen (background set to gray, 145 cd/m^2). Stimulus presentation was controlled using the Presentation software package (Version 9.90, Neurobehavioral Systems).

Stimuli were presented in a pseudo-randomized order for a short time window of 0.2 seconds with a vertical visual angle of 20.8 and a horizontal visual angle of 16.2 degrees (white stimulus parts, 1140 cd/m^2; black stimulus parts, 30 cd/m^2). To avoid effects of fatigue, the overall experiment was divided into six blocks (134 stimuli per block) and subjects were allowed to take short breaks between blocks. In each block, 20 Mooney face stimuli of each face condition were presented together with No-Face stimuli in a 3:2 (exact ratio 2.96:2) ratio to counteract response bias; resulting in 80 Mooney face stimuli and 54 Scramble stimuli. The inter-trial-interval between stimulus presentations was randomly jittered from 3.5 to 4.5 seconds.

2.3.5. Task and Instructions

Subjects performed a face detection task on two-tone images and responded by pressing one of two buttons. The button assignment for a “Face” or “No-Face” response was counterbalanced across subjects (n=24 right index finger for

'Face' response). Subjects were instructed to respond only once and as precisely and quickly as possible. The subjects were informed about the ratio (3:2) of "Faces" to "No-Faces" in the presentation. Between stimulus presentations subjects were instructed to fixate a white cross on the center of the gray screen. Further, they were instructed to maintain fixation during the whole block. In addition, subjects were asked to suppress eye blinks during stimulus presentation and to avoid any movement during the acquisition session. Before data acquisition, subjects performed a test block of two minutes with stimuli not used during the actual task.

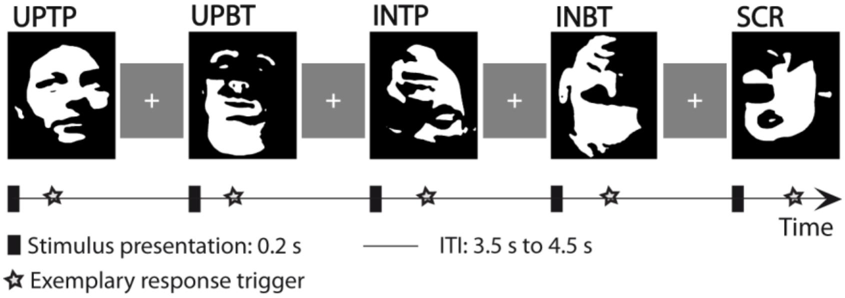

Figure 2.1. Graphical depiction of stimulus timing and the five stimulus categories. UPTP: Upright faces with illumination from the top; UPBT: Upright faces with illumination from the bottom; INTP: Inverted faces with illumination from the top; INBT: Inverted faces with illumination from the bottom; SCR: Scrambled Mooney stimuli, not representing a face; ITI = Intertrial interval

2.3.6. Data acquisition and exclusion criteria

MEG data acquisition was performed in line with recently published guidelines for MEG recordings (Gross et al., 2012). MEG signals were recorded using a whole-head system (Omega 2005; VSM MedTech Ltd.) with 275 channels. The signals were recorded continuously at a sampling rate of 1200 Hz in a synthetic third-order gradiometer configuration and were filtered online with fourth-order Butterworth filters with 300 Hz low pass and 0.1 Hz high pass.

Before and after each block the subject's head position relative to the gradiometer array was determined using three localization coils, one at the nasion and the other two located 1 cm anterior to the tragus of each ear on the nasion-tragus plane. Blocks with a head movement exceeding 5 mm were discarded from further MEG data analysis.

For artefact detection the horizontal and vertical electrooculogram (EOG) was recorded via four electrodes; two were placed distal to the outer canthi of the left and right eye (horizontal eye movements) and the other two were placed above and below the right eye (vertical eye movements and blinks). The impedance of each electrode was measured with an electrode impedance meter (Astro-Med, Inc Grass Instrument Division, W.Warwick RI USA) and was kept below 15 kΩ.

Structural magnetic resonance (MR) images were obtained with a 3T Siemens Allegra or Trio scanner (Siemens Medical Solutions, Erlangen, Germany) using a standard T1 sequence (3-D magnetization -prepared -rapid-acquisition gradient echo sequence, 176 slices, 1 x 1 x 1 mm voxel size). For the structural scans vitamin E pills were placed at the former positions of the MEG localization coils for co-registration of MEG data and magnetic resonance images. Behavioral responses were recorded using a fiberoptic response pad (Photon Control Inc. LUMItouch™ Response System) in combination with the Presentation software (Version 9.90, Neurobehavioral Systems). Participants were excluded from further analysis if a response bias was detected (5 of 59 subjects). For response bias detection we calculated the normalized c criterion ($c_{(n)}$, Green and Swets, 1966) from the performance of each participant. A mean response bias deviating more than two standard deviations from zero was chosen as the rejection criterion.

2.3.7. Statistical analysis of behavioral data

Responses were classified as correct or incorrect based on the subject's first answer. For the hit rate analysis, the accuracy for each condition was calculated. For the reaction time analysis only correct responses were considered.

Hit rates (HRs) and reaction times (RTs) were subjected to separate 2x2 repeated-measurements permutation ANOVAs (Anderson and Ter Braak, 2003; Suckling and Bullmore, 2004). To test whether the standard F-statistics obtained for the main effects and the interaction were likely to have occurred by chance, the condition labels of the original data were permuted across conditions. The F-value of the original data was then tested against an empirical distribution of F-values constructed from 5000 data sets with such randomly permuted condition labels. Each main effect and the interaction were tested separately. F-values larger than the 95th percentile of the distribution of F-values obtained for the permuted data sets were considered to be significant at an alpha level of 0.05. For the main effects, condition labels were permuted between the two levels of the tested factor within each subject, but permutations were restricted to occur within the level of the other factor, e.g. for the orientation effect labels for UPTP and INTP were considered to be exchangeable, but labels of UPTP and INBT were not exchangeable. By keeping the labels of the other factor fixed, we aimed to avoid any confounds due to the variability introduced by the factor not currently of interest. For calculation of the interaction effect, condition labels were permuted across levels of both factors within subjects. In contrast to standard F-tests, non-parametric permutation tests avoid the assumption of normality and are therefore recommended when testing non-Gaussian data as they are frequently encountered in behavioral measurements.

For post-hoc testing, a Wilcoxon signed rank test was performed for each simple effect and a sequential Bonferroni Holm correction (Holm, 1979) was applied to account for multiple comparisons (uncorrected alpha level = 0.05).

2.3.8. MEG-data analysis

Preprocessing

Data analysis was performed with Matlab (RRID:nlx_153890; MATLAB 2008b, MathWorks, Inc.) and the open source Matlab toolbox Fieldtrip (RRID:nlx_143928; Oostenveld et al., 2011; Version 2012 01-05).

Trials were defined from 0.55 s before to 0.55 s after stimulus onset. The time-point of the stimulus onset was adjusted to take the projector delay into account.

Trials containing sensor jump-, eye movement-, or muscle-artefacts were rejected using automatic FieldTrip artefact rejection routines. In addition, EOG channels were checked manually for horizontal and vertical eye movements.

Only trials with correct behavioral responses were taken into account for MEG data analysis.

To avoid potential effects of button-press related motor activity, we analysed only data up to 0.350 s after stimulus onset.

Spectral analysis at the sensor level

A multi-taper approach (Percival and Walden, 1993) based on Slepian sequences (dpss; Slepian, 1978) was used for time-frequency transformation. The transformation was applied in an interval from 2 to 150 Hz in 2 Hz steps and in a time window of 0.400 s – 0.050 s before (baseline) and 0 – 0.350 s after stimulus onset (task).

For each frequency, we considered an adaptive sliding time-window with a width of 7 divided by the frequency in Hz and an adaptive frequency smoothing, with a factor of 0.2 times the frequency, resulting in 2 tapers for each frequency. Time frequency representations (TFR) for the combined face conditions (UPTP, UPBT, INTP and INBT) were averaged over time to obtain an average frequency representation for the task and baseline period, respectively. To identify frequency bands for subsequent beamformer analysis, we compared the spectral power in the task interval for all subjects and the combined face conditions with the baseline spectral power using a dependent-sample permutation t-test and a cluster-based correction method (Maris and Oostenveld, 2007) to account for multiple comparisons across frequency and sensors. Clusters were defined as (spatially and spectrally) adjacent samples whose t-values exceeded a critical threshold corresponding to an uncorrected alpha level of 0.05. Cluster sizes were defined by taking the sum of t-values of a given cluster. During the randomization procedure labels of task and baseline data were randomly reassigned within each subject. Cluster sizes observed for the original data set were then tested against the distribution of cluster sizes obtained from 1000 permuted data sets. Cluster values larger than the 95th percentile of the distribution of cluster sizes obtained for the permuted data sets were considered to be significant. We found a significant positive and a significant negative cluster (Fig. 2.3.). To delineate frequency bands for these clusters, we identified the points of maximum curvature in the spectrum by visual inspection. Based on the points of maximum curvature (excluding the maximum turning points for positive values and minimum turning points for negative values), we determined four non-overlapping frequency intervals for subsequent beamformer source analysis: 1. 14-28 Hz (beta); 2. 28-56 Hz (low gamma); 3. 56-68 Hz (mid gamma); 4. 68-144 Hz (high gamma).

Note that current recommendations for best practice favour source level statistics over statistics at the sensor level (Gross et al., 2012), we therefore only performed the minimally necessary statistics for a choice of frequency bands at the sensor level, while all other (orthogonal) statistical tests were performed at beamformer source level.

Source grid creation

To create individual source grids we transformed the anatomical MR images to a standard T1 template from the SPM8 toolbox (http://www.fil.ion.ucl.ac.uk/spm) in MNI space (Collins et al., 1994) obtaining an individual transformation matrix for each subject. We then warped a regular 3-D dipole grid based on the standard T1 template (spacing 10mm) with the inverse of the transformation matrix, to obtain an individual dipole grid for each subject in subject space. This way, each specific grid point was located at the same brain area for each subject, which allowed us to perform source analysis with individual head models as well as multi-subject statistics for all grid locations. Lead fields at those grid locations were computed for the individual subjects with a realistic single shell forward model (Nolte, 2003).

Beamformer source power analysis

Beamformer source analysis was performed using the DICS (dynamic imaging of coherent sources) algorithm; a frequency domain beamformer (Gross et al., 2001) implemented in the FieldTrip toolbox. While the DICS algorithm was designed to compute source coherence estimates, we used real valued filter coefficients only and thus restricted our analysis to the local source power (see also Grützner et al., 2010). The real part of the filters reflects the propagation of the magnetic fields from sources to sensors, as this process is supposed to happen instantaneously (e.g. Nunez and Srinivasan, 2006). Beamformer analysis uses an adaptive spatial filter to estimate the power at every specific location of the brain. The spatial filter is constructed from the individual lead fields and the cross spectral density matrix for each subject. Cross spectral density matrices were computed for the task period of 0 to 0.350 s after stimulus onset and the baseline period of 0.400 to 0.050 s before stimulus onset in four bands based on the statistical analysis of spectral power at the sensor level (spectral smoothing indicated in brackets): 21 Hz (± 7Hz), 42 Hz (± 14 Hz), 62 Hz (± 6 Hz), 106 Hz (± 38 Hz). Cross-spectral density matrix calculation was performed using the FieldTrip toolbox with the multi-taper-method (Percival and Walden, 1993) using 3, 4, 9 or 26 Slepian tapers (Slepian, 1978), depending on the required spectral smoothing. We used a regularization of 5% (Brookes et al., 2008).

Beamformer filters were computed as “common filters” based on the activation and baseline data across all conditions. Using common filters for activation and baseline and all conditions allows for subsequent testing for differences between conditions; using common filters ensures that differences in source activity do not reflect differences between filters.

Spatial filtering of the sensor data for source statistics was then performed by projecting single trials through the common filter for each condition, task and baseline separately.

Source Statistics

We used an equal amount of trials for the beamformer analysis for each subject in all conditions, to make sure that statistical differences were not caused by a different numbers of trials. When the trial number differed across conditions for a subject, the minimal amount of trials across conditions was selected randomly from the available trials in each condition.

Statistical testing was performed in two steps: At the first level, we computed a within-subject t-test on the single trial data to obtain a test statistic for task vs. baseline source activity for each condition (dual state beamformer, Huang et al., 2004). At the second level, the resulting t-values for each grid point and condition across all subjects were subjected to a 2x2 repeated-measurements permutation ANOVA with factors stimulus orientation and illumination direction. Hereby, we aimed to identify the consistent effects of condition-dependent source-power changes across subjects. To account for multiple comparisons across voxels, a cluster-based correction method (Maris and Oostenveld, 2007) was used. Clusters were defined to be adjacent voxels whose F-values exceeded a critical threshold corresponding to an uncorrected alpha level of 0.05. Cluster sizes were defined the same way as for the sensor level statistics and were then tested against the distribution of cluster sizes obtained from 5000 permuted data sets. Permutation strategies for main effects and the interaction were identical to the ones applied to the behavioral data. Cluster values larger than the 95th percentile of the distribution of cluster sizes obtained for the permuted data sets were considered to be significant. For illustration of the effects in bar charts, the t-values of the significant voxels in each cluster were averaged for each condition and over all subjects.

Both, the statistical procedure for the cluster-based analysis as well as the beamformer analysis parameters chosen for source power reconstruction were very similar to the approach applied by Gruetzner and colleagues (2010). Gruetzner and colleagues were able to show a close correspondence of the beamformer source locations recovered from MEG data and the locations revealed by fMRI in a Mooney faces task, supporting the validity of the method.

Post-hoc source analysis

To characterise the effects in more detail by examining the frequency and time ranges at which the conditions underlying the significant effects differed, a post-hoc analysis was performed. For this purpose, the source time courses of all significant voxels obtained by the permutation ANOVA were extracted. To that end, raw data were filtered in a broad frequency range (8 Hz high pass, 150 Hz low pass). Then, we calculated a time-domain beamformer filter (LCMV, linear constrained minimum variance, Van Veen et al., 1997) based on task and baseline intervals of all conditions (“common filters”, Nieuwenhuis et al., 2008). For each source location three orthogonal filters were computed (x, y, z direction). To obtain the source time courses, the broadly filtered raw data was projected through the LCMV filter. Subsequently, the 3-D direction carrying the largest variance, indicating the dominant dipole orientation, was identified using a singular value decomposition.

For each source time course a time-frequency transformation was applied with the same parameters as for the sensor level analysis but only in the relevant frequency range (high gamma frequency range). Source time-frequency spectral power was transformed to relative change values by subtracting the average baseline power at each frequency and by subsequently dividing by it.

To determine the time and frequency ranges of the differential activations underlying the main or interaction effects, time-frequency transformations were averaged across voxels within each significant cluster of the permutation ANOVA and subjected to a post-hoc dependent samples permutation t-test. When investigating the main effects, we additionally averaged over the two levels of the other (i.e. currently not tested) factor. For example, for the main effect of orientation, we calculated the mean of inverted stimuli (INTP and INBT) and the mean of upright stimuli (UPTP and UPBT) across all voxels and contrasted the resulting TFR with the permutation t-test. Condition labels were randomly reassigned within each subject between the two levels of the tested factor during the randomization procedure. For the main effects of illumination the mean of stimuli illuminated from the bottom (UPBT and INBT) and the mean of stimuli illuminated from the top (UPTP and INTP) were contrasted. For the interaction effect we first calculated the orientation difference for stimuli illuminated from the bottom (UPBT-INBT) and from the top (UPTP-INTP) and contrasted the resulting difference TFR using the permutation t-test. To account for multiple comparisons across frequency and time bins a cluster-based correction method (Maris and Oostenveld, 2007) was used. For one of the effects the post-hoc test did not reach significance with the cluster-based correction method and only uncorrected t-values are reported (Figure 2.5.C).

To obtain the time-points of the strongest differential activation for each effect, the difference in the averaged TFR between the two levels of the tested factor (e.g. upright and inverted stimuli for the orientation effect) was further averaged over the relevant frequency range and plotted over time. Only the peaks in the significant time ranges identified by the post-hoc tests are reported. For one of the effects, for which the post-hoc test did not reach significance with the cluster-based correction method, both main peaks are reported.

Correlation of high-frequency gamma band activity (GBA) with reaction times

Pearson’s correlations were calculated in order to assess the relationship between per-subject mean reaction times and baseline corrected high-frequency GBA averaged over the significant cluster obtained for each effect.

Before correlation, RT and GBA for each subject were averaged over upright (UPTP, UPBT) and inverted (INTP, INBT) conditions for the orientation effect, conditions illuminated from the top (UPTP, INTP) and from the bottom (UPBT, INBT) for the illumination effect and congruent (UPTP, INBT) and incongruent (UPBT, INTP) conditions for the interaction effect.

To focus on the effects of potential PEs, we subtracted each subjects' mean of GBA at the significant source locations across the four face conditions as well the subjects' mean RT across the four face conditions from the individual GBA and RT values, respectively. This subtraction corrects for individual differences in GBA (see Hoogenboom et al., 2006) as well as in behavioural speed between subjects (see Kanai and Rees, 2011, e.g. related to variations in the myelination of motor fibers.

2.4. Results

2.4.1. Behavioral analysis

To assess the behavioral effects of the violation of the orientation and illumination prior, we analysed the hit rates (HR) and the reaction times (RT) of correct responses by means of a permutation ANOVA (see Methods). Post-hoc Wilcoxon Signed Rank tests were used to investigate the simple effects underlying the interactions for HR and RT (Figure 2.2.). Statistical results are summarized in Table 2.1.

Hit rates

Subjects made fewest mistakes in detecting faces when both priors were met (avg. $HR_{(UPTP)}$ = 94.38%) and made most mistakes when both priors were violated (avg. $HR_{(INBT)}$ = 68.84%), suggesting the induction of PEs by our task design.

The permutation ANOVA revealed a main effect of orientation (p = 0.0002) and illumination (p = 0.0002), as well as an interaction between the two factors (p = 0.0002). Higher HR were found for the upright (UP) than for the inverted (IN) Mooney faces. Also, higher HR were found for the Mooney faces illuminated from the top (TP) than for the Mooney faces illuminated from the bottom (BT).

Post-hoc tests revealed that violating the orientation prior led to a decrease in HR for faces illuminated from the top ($p = 6.6 \times 10^{-9}$; avg. $HR_{(UPTP)}$ – avg. $HR_{(INTP)}$ = 13.3%,) as well as for faces illuminated from the bottom ($p = 1.63 \times 10^{-9}$; avg. $HR_{(UPBT)}$ – avg. $HR_{(INBT)}$ = 24.1%). HR also decreased, when the illumination prior was violated for upright ($p = 0.046$; avg. $HR_{(UPBT)}$ – avg. $HR_{(UPTP)}$ = 1.8%) and inverted Mooney faces ($p = 2.14 \times 10^{-8}$; avg. $HR_{(INBT)}$ – avg. $HR_{(INTP)}$ = 12.2%).

Reaction times

Subjects responded fastest when both priors were met (avg. $RT_{(UPTP)}$ = 0.614 s) and responded slowest when both were violated (avg. $RT_{(INBT)}$ = 0.723 s), which is also in line with the induction of PEs by our task design.

We found main effects of orientation and illumination for the reaction times ($p = 0.0002$), as well as an interaction between the two factors ($p = 0.0002$). Shorter RT were found for the upright (UP) than for the inverted (IN) Mooney faces. Also, RT were shorter for the Mooney faces illuminated from the top (TP) than for the Mooney faces illuminated from the bottom (BT).

Violating the orientation prior led to increases in RT for faces illuminated from the top ($p = 1.63 \times 10^{-9}$; avg. $RT_{(INTP)}$ – avg. $RT_{(UPTP)}$ = 0.0710), and for faces illuminated from the bottom ($p = 1.11 \times 10^{-8}$; avg. $RT_{(INBT)}$ – avg. $RT_{(UPBT)}$ = 0.0899 s) as revealed by the post-hoc Wilcoxon Signed rank tests.

Further, an increase in RT was detected when the illumination prior was violated for the upright Mooney faces ($p = 0.0035$; avg. $RT_{(UPBT)}$ – avg. $RT_{(UPTP)}$ = 0.0190 s). The violation of the illumination prior had an even more severe effect on the detection of Mooney faces in inverted orientation ($p = 2.22 \times 10^{-7}$; avg. $RT_{(INBT)}$ – avg. $RT_{(INTP)}$ = 0.0379 s).
The orientation effect on RT as well as HR was stronger than the illumination effect (RT: $p = 3.23 \times 10^{-8}$; HR: $p = 1.63 \times 10^{-9}$).

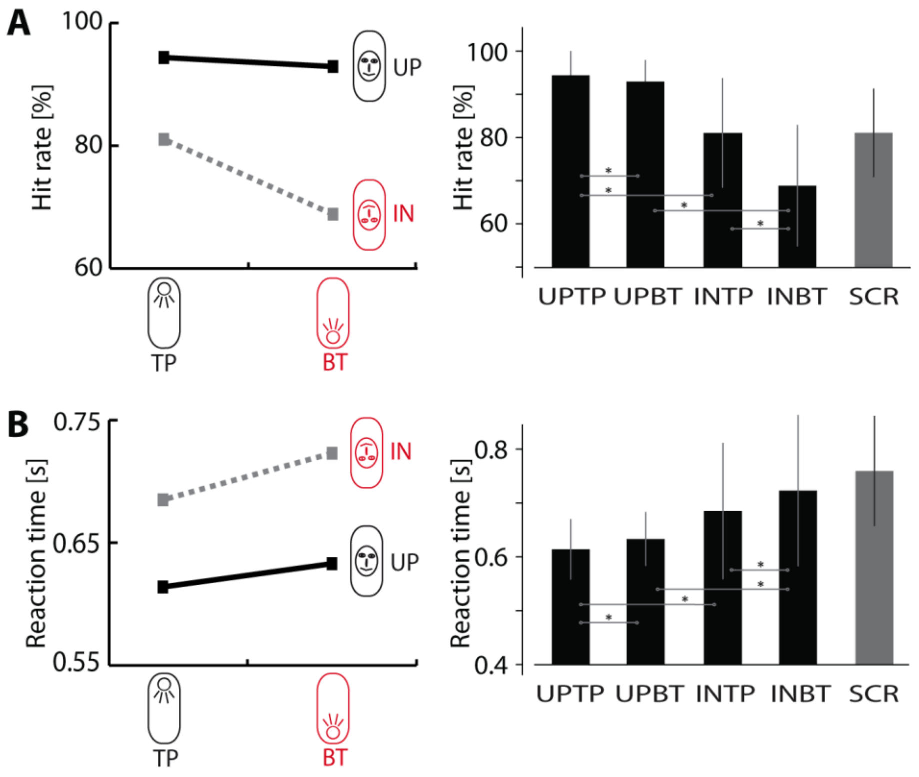

Figure 2.2. Behavioral analysis of hit rates and reaction times of correct responses. Interaction plots (left) and bar plots (right) for hit rates and reaction times (n = 48) ***A)*** Hit rates decreased, when the orientation prior and/or the illumination prior were violated*.* ***B)*** Reaction times increased when the orientation prior and/or the illumination prior were violated. Error bars indicate one standard deviation of the mean. Asterisks indicate significant results of Post-hoc Wilcoxon signed rank tests, Bonferroni-Holm corrected for multiple comparisons.

2.4.2. Neural responses

We performed a time-resolved beamformer source analysis of MEG activity to assess the PE responses in source space that corresponded to the violations of illumination and orientation priors. To this end we first identified the relevant frequency bands for beamformer analysis by statistically comparing the sensor activity in the task interval (0-350 ms) for all face conditions and correct trials with the baseline activity.

Table 2.1. Behavioral analysis: ANOVA and post-hoc test results

	Hit rate	*Reaction times*
Mean		
UPTP	94.38 %	614 ms
UPBT	92.58 %	633 ms
INTP	81.08 %	685 ms
INBT	68.47 %	723 ms
Permutation ANOVA (p-values)		
Main effect Orientation	$2x10^{-4*}$	$2x10^{-4*}$
Main effect Illumination	$2x10^{-4}$*	$2x10^{-4*}$
Interaction effect	$2x10^{-4*}$	$2x10^{-4*}$
Post-hoc Wilcoxon signed rank test (p-values)		
UPTP vs. UPBT	0.0416*	0.0035*
UPTP vs. INTP	$6.60x10^{-9*}$	$1.63x10^{-9*}$
UPBT vs. INBT	$1.63x10^{-9*}$	$1.11x10^{-8*}$
INTP vs. INBT	$2.14x10^{-8*}$	$2.22x10^{-7*}$
* = significant, corrected for multiple comparisons		

This analysis revealed a cluster with task-related increases in activity over occipital, parietal and temporal sensors and a cluster with task-related decreases over frontal, parietal and temporal sensors (Figure 2.3.). The spectral profile of the two clusters was used to determine four non-overlapping frequency intervals for beamformer source analysis: 1. 14-28 Hz (beta); 2. 28-56 Hz (low gamma); 3. 56-68 Hz (mid gamma); 4. 68-144 Hz (high gamma).

Note that all later statistical comparisons were carried out in source space as this was strongly recommended in the recently published guidelines for MEG analyses (Gross et al., 2012). Moreover, we note that all subsequent statistical comparisons were orthogonal to the one used for identifying the frequency bands of interests, i.e. there is no double-dipping (Kriegeskorte et al., 2009).

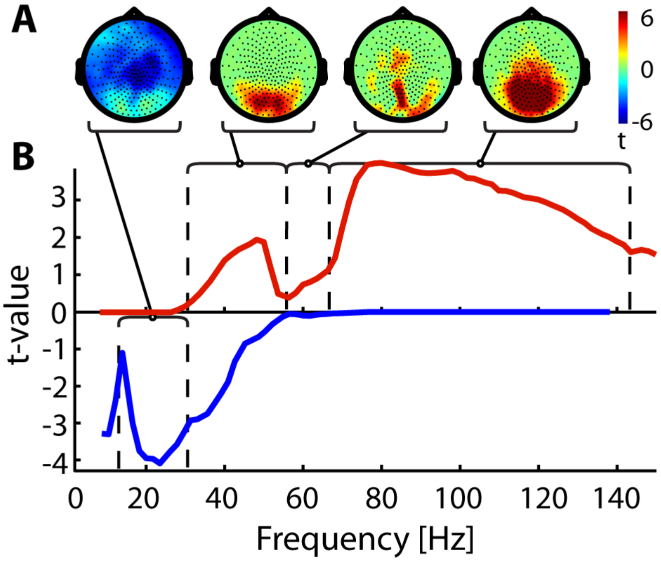

Figure 2.3. Sensor level frequency analysis. Significant clusters at the sensor level identified by frequency analysis for the four Mooney face conditions (task vs. baseline interval, t-values masked by 0.05, cluster correction, n = 48). ***A)*** Topographic plots of the activity for each identified frequency range. Note the two spatial clusters with task-related decreases (blue colors) and increases (red colors). ***B)*** Power spectra for the two clusters, with task related increases in power (red) and task-related decreases (blue). Black dashed lines frame the frequency ranges of interest for subsequent beamformer source power analysis.

High gamma frequency range (68-144 Hz):

Orientation effect

In the high gamma frequency range, we observed a main effect of orientation (cluster-based permutation ANOVA, p = 0.0154) at the occipital pole (V2), right superior occipital gyrus, left middle occipital gyrus as well as left fusiform gyrus (Figure 2.4.A left column; see Table 2.2. for MNI coordinates of peak voxels). At these areas, power in comparison to baseline was higher for inverted Mooney faces than for upright Mooney faces (Fig. 2.4.A right column). Post-hoc analysis revealed two significant clusters, the first one peaking at 80 ms and the second one at 270 ms after stimulus onset (Fig. 2.5.A bottom row). The orientation effect involved the high gamma frequency range from 76 - 120 Hz (Fig. 2.5.A middle row).

Illumination effect

We found a main effect of illumination (cluster-based permutation ANOVA, p = 0.012) in a cluster located in right superior frontal gyrus (SFG) / superior frontal sulcus (SFS), medial frontal cortex (MFC) and anterior cingulate gyrus (ACG) (Fig. 2.4.B left column; see Table 2.2 for MNI coordinates of peak voxels). At these locations power in comparison to baseline was higher for Mooney faces with illumination from the bottom than for Mooney faces with illumination from the top (Fig. 2.4.B right column). Post-hoc analysis revealed a significant frequency range from 78 - 112 Hz and a peak time at around 120 ms after stimulus onset (Fig. 2.5.B middle and bottom row).

A second cluster for the main effect of illumination (cluster-based permutation ANOVA, p = 0.011) had a maximum located at right supramarginal gyrus (SMG) in the inferior parietal lobule, but extended also to the inferior temporal gyrus (Fig. 2.4.C left column; see Table 2.2. for MNI coordinates of peak voxels). At these locations, power in comparison to baseline was higher for Mooney faces with illumination from the top than for Mooney faces with illumination from the bottom (Fig. 2.4.C right column). This difference peaked at around 135 ms and 310 ms after stimulus onset and was most pronounced between 75 - 144 Hz (Fig. 2.5.C middle and bottom row).

Table 2.2. Effects on beamformer reconstructed source power in the high gamma frequency range (68-144 Hz); corresponding time, anatomical regions and MNI coordinates of peak voxels

Effect	*Time*	*Anatomic region* *(L = left; R = right)*	*MNI coordinates (x,y,z)*
1. Orientation effect	80 ms / 270 ms	L Occipital pole (V2)	-10, -70, 0
		R Superior occipital gyrus	30, -80, 30
		L Middle occipital gyrus	-40, -80, 20
		L Fusiform gyrus	-40, -70, -10
2. Illumination effect I	120 ms	Superior frontal gyrus/ Superior frontal sulcus	-30, 30, 50
			10, 0, 70
		Medial frontal cortex	0, 30, 50
			-10, 50, 20
		Anterior cingulate gyrus	-10, 30, 30
			10, 10 , 30
3. Illumination effect II	135 ms / 310 ms	R Supramarginal gyrus	40, -40, 20
		R Inferior temporal gyrus	60, -50, -10
4. Interaction effect	210 ms	L Superior parietal lobe/ Precuneus	-10, -60, 70
		R V2	10, -70, 20
		R Inferior occipital gyrus	10, -90, -20
		R Lingual gyrus	10, -40, -10
		R Cerebellum	20, -80, -50

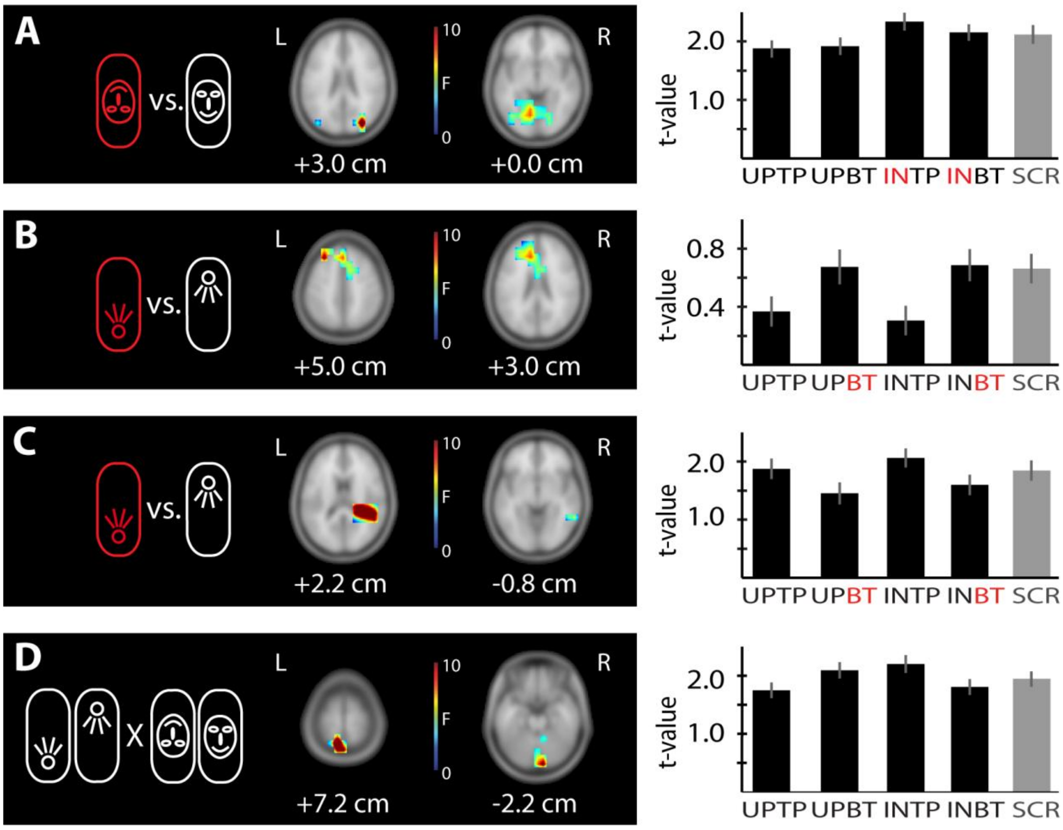

Figure 2.4. Statistical analysis on beamformer estimated MEG source power in the high gamma frequency range (68-144 Hz). ***A)*** Main effect of orientation; ***B)*** Main effect of illumination *I*; ***C)*** Main effect of illumination *II*; ***D)*** Interaction effect of orientation and illumination. ***Left****:* Results of the 2-factorial permutation ANOVA on beamformer estimated source power (permutation F-values masked by $p < 0.05$, cluster correction, $n = 48$; z-value below each brain slice). Two representative slices are shown. For MNI coordinates of the peak voxels see Table 2.2. Contrasts are indicated by icons. L = left; R = right. ***Right****:* Mean t-values (task vs. baseline contrast) for the significant cluster shown on the left, in the four face conditions and the scrambled condition. Error bars indicate one standard error of the mean.

Interaction effect

In the high gamma frequency range also an interaction effect of the factors illumination and orientation (cluster-based permutation ANOVA, $p = 0.002$) was observed. The cluster was located at left superior parietal lobe (SPL) / precuneus, occipital pole (V2), right inferior occipital gyrus, right lingual gyrus and the right cerebellum (Fig. 2.4.D left column; see Table 2.2 for MNI coordinates of peak voxels). Here, source power in comparison to baseline was higher for the UPBT and INTP condition than for the INBT and UPTP condition (Fig. 2.4.D right column). The interaction effect involved a significant frequency interval from about 68 to 96 Hz and had a peak at 210 ms after stimulus onset (Fig. 2.5.D middle and bottom row).

Beta (14-28 Hz), low (28-56 Hz) and mid gamma (56-68 Hz) frequency range:

No significant main or interaction effects were found in the beta, mid and low gamma frequency range.

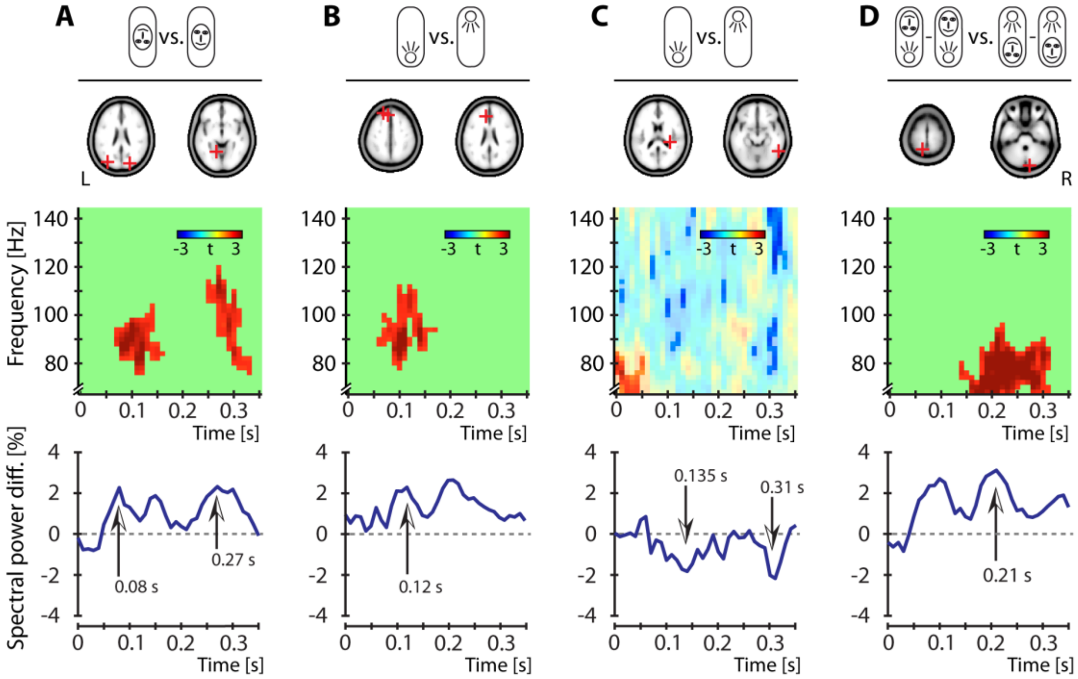

Figure 2.5. Post-hoc analysis on beamformer estimated MEG source power in the high gamma frequency range (68-144 Hz). Post-hoc time frequency analysis for significant voxels of the 2-factorial permutation ANOVA shown in Figure 2.4. ***A)*** Main effect of orientation; ***B)*** Main effect of illumination *I*; ***C)*** Main effect of illumination *II*; ***D)*** Interaction effect of orientation and illumination. ***Top*****:** Contrasts and peak source locations of significant cluster. ***Middle:*** Time-frequency representation of post-hoc permutation t-test. **A**, **B** and **D**: Cluster correction, t-values masked by p < 0.05 **C**: Uncorrected, t-values corresponding to p < 0.05 are highlighted**.** ***Bottom****:* Mean relative high-frequency gamma-band power difference over time. The arrows highlight the peaks within the significant time periods. Please note that effect size might be exaggerated as only significant voxels were selected. L = left; R = right.

Correlation of high-frequency gamma band responses and RT

Correlation of high-frequency gamma band responses and RT revealed a significant positive correlation at the source locations of the orientation effect (r = 0.37, p = 0.00019, Figure 2.6.A) and the first illumination effect (r = 0.43, p = 8 x 10^{-6}, Figure 2.6.B). A significant negative correlation was found at the source locations of the interaction effect (r = -0.32, p = 0.0011, Figure 2.6.D) and a tendency towards a negative correlation was found at the locations of the second illumination effect (r = -0.17, p = 0.09, Figure 2.6.C).

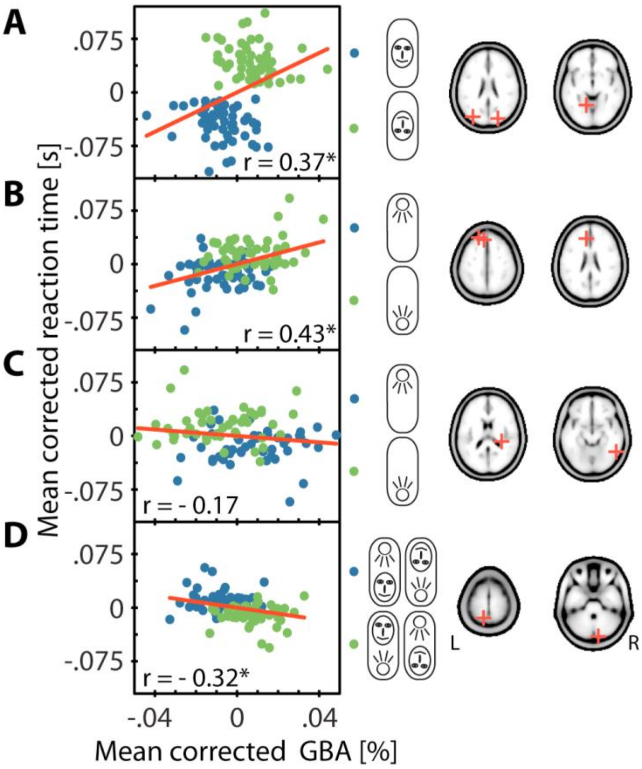

Figure 2.6. Correlation analysis for high-frequency gamma power and reaction times. Scatter plots displaying the correlation of per-subject mean values (see methods for details) of high-frequency GBA with reaction times at the source locations of the ***A)*** Main effect of orientation, ***B)*** Main effect of illumination *I*, ***C)*** Main effect of illumination *II* and ***D)*** Interaction effect of orientation and illumination. Peak source locations are indicated on the right. Each subjects' mean of GBA across conditions at the indicated source locations and the mean reaction time across conditions was subtracted before correlation to focus on the effects of potential PEs (n=48). Asterisks indicate significant correlation. Linear regression lines are shown in red for each effect. L = left; R = right.

2.5. Discussion

We tested whether prediction errors (PEs) are reflected by increased neural activity vs. the alternative of reduced neural activity for violated predictions. Using MEG with its direct access to electrophysiological activity allowed testing specifically whether PEs are signaled in gamma-band activity (GBA). PEs were induced by the violation of two priors based on lifelong visual experience – upright face orientation and illumination from the top. Deviations from these priors were embedded in a Mooney face detection task (Mooney, 1957).

Behavioral findings confirmed the successful induction of PEs by our task. In addition, neuronal activity at task-specific brain locations was increased when priors were violated, in line with the concept of PEs in predictive coding theory (Rao and Ballard, 1999). Importantly, this increase in neuronal activity was indeed observed in GBA (> 68 Hz), the frequency range thought to be associated with the bottom-up propagation of PEs (Arnal and Giraud, 2012; Bastos et al., 2012). These findings strongly support the notion that increased (high-frequency) GBA reflects PEs. No PE signals were found in any of the lower frequency bands, suggesting that PEs are mainly represented in high-frequency GBA.

However, for the violation of the illumination prior we additionally found *decreased* GBA in posterior parietal brain areas, which may represent decreased attention to internal mnemonic representations (Wagner et al., 2005). Hence, we suggest that the high-frequency GBA not only signals PEs, but also attentional effects – in line with previous results (e.g. Fries et al., 2001).

2.5.1. Violations decrease accuracy and increase reaction times

Behavioral responses were slower and more inaccurate when priors were violated. This is in line with other behavioral phenomena accounted for by predictive coding such as priming and global precedence (Friston, 2005) and validates that our task design successfully induced PEs.

Notably, the violation of the orientation prior had a higher impact on hit rates and reaction times than the violation of the illumination prior. This difference may be explained as follows. While a robust inversion effect is found in face perception (Yin, 1969 for photographic faces; Rodriguez et al., 1999 for Mooney faces), the illumination prior varies substantially between individuals (Adams, 2007) and can be altered with experience (Adams et al., 2004). Thus, the stronger behavioural effect of the violation of the orientation prior is in line with a precision-weighting of PEs (Friston and Kiebel, 2009; Adams et al., 2013) based on the higher precision of the orientation prior than the illumination prior.

2.5.2. Cortical source power changes in high-frequency GBA reflect PEs

For the violation of the orientation prior we expected that the neural correlate of a PE should arise before any illumination effect, and that it would be signaled by GBA increases. Indeed, at 80 ms after stimulus onset – before any effect of illumination – we observed the first of two significant clusters of increased high-frequency GBA for the violation of the orientation prior in early visual areas. These areas have been linked to contour integration (e.g. Kourtzi et al., 2003). The contour-integration role of these areas combined with the early latency of the orientation effect supports its interpretation as reflecting PEs arising for unexpected face orientations. This is because contour processing areas are suitable candidate locations for an orientation PE as an early (2-D) contour match to internal templates was suggested as the first stage of Mooney face recognition (Cavanagh, 1991). Since the stimulus contour pattern of the inverted faces does not match the expected template contour pattern of upright faces, a specific PE in contour processing brain areas is supposed to arise for inverted stimuli at this early processing stage.

An orientation-related PE could also arise in areas tuned to specific, illumination-invariant, coarse-grained luminance contrasts in faces, because these seem to play a role in face processing (Ohayon et al., 2012). This specific tuning was reported in the macaque middle face patch (MFP), making its homologue, the fusiform face area a candidate for orientation PEs. However, MFP cells seem to be preferentially active for contrasts matching environmental priors, additionally requiring embedding of the contrast in a face-like pattern. This latter condition is not well met in Mooney stimuli, potentially reducing any effects of changes in luminance contrasts with orientation in our study.

For the violation of the illumination prior we expected that the correlate of a PE should arise after the first orientation-related effect. Again, we expected this PE to be signaled by increased GBA. We observed increased high-frequency GBA for violation of the illumination prior at 120 ms after stimulus onset, and thus 40 ms after the first orientation effect. This effect was located in MFC, SFS and ACG. Both timing and location of this effect support its interpretation as an illumination-related PE. This is because the illumination direction strongly influences the shading pattern of an image and shading cues are the only cues available in Mooney faces to reconstruct the 3-D shape (Kemelmacher-Shlizerman et al., 2008). PEs are therefore likely to arise in areas involved in the processing 3-D shape from shading cues, such as the MFC (Taira et al., 2001). Additionally, SFS may be used to keep shading cues in working memory (Courtney et al., 1998) and ACG may support error detection (Botvinick et al., 2004). Thus, we interpret this illumination effect as a PE signal for the unexpected illumination.

We also observed an interaction effect with increased GBA for the UPBT and INTP conditions at precuneus, V2 and lingual gyrus, which all three are involved in (global) shape processing (Fink et al., 1997; Hegdé and Van Essen, 2000; Tanskanen et al., 2008). This interaction effect occurred at 210 ms after stimulus onset. At this late time-point, the process model of Cavanagh (1991) suggests that the shape of the sensory input is supposed to be evaluated based on the interaction of light and 3-D structure. The combination of these two properties of a scene can also be predicted based on prior experience. We expect upright face orientation to be combined with illumination from the top and – as it is probably more common to see photographs of inverted faces than actual inverted faces – we expect inverted face orientation to be combined with illumination from the bottom. This expected combination of orientation and illumination is violated in the INTP and UPBT conditions. Therefore, we interpret this late interaction effect at precuneus, V2 and lingual gyrus as a PE at a higher conceptual level.

2.5.3. GBA additionally reflects attentional effects

We observed a second illumination effect peaking at 135 ms and 310 ms after stimulus onset. For this illumination effect, we found a decrease of GBA for violation of the illumination prior mainly in SMG. Activity in this area may reflect deployment of attention to internal mnemonic representations – as stated in the attention to memory hypothesis (AtoM, Wagner et al., 2005). Accordingly, the SMG usually shows decreased BOLD fMRI activity for less familiar information (Wagner et al., 2005; Ciaramelli et al., 2008) – potentially corresponding to unusual illumination conditions here. To link these fMRI findings to our MEG results, we draw on the well established positive correlation of the BOLD-fMRI signal with GBA in MEG (Brookes et al., 2005). Taking this correlation into account, the observed decreased GBA for the stimuli with the less familiar illumination direction in the SMG may be an AtoM effect rather than a PE.

Thus, our results suggest that high-frequency GBA does not exclusively signal PEs, but also reflects attention. This attentional interpretation could be reconciled with an interpretation as a PE by the recent proposal that attention itself is implemented via gain modulation of PE units (Feldman and Friston, 2010). As our study was not designed to test this specific hypothesis, the interplay of attention, PEs and GBA remains to be investigated.

2.5.4. Increased GBA for violations is associated with slower processing

High-frequency GBA at the locations of the orientation effect and the first illumination effect showed a positive correlation with RT. This relationship is compatible with longer RT reflecting the PE for violation of the orientation and illumination prior.

In contrast, the negative relationship of GBA and RT at the locations of the second illumination effect suggests that here increased GBA rather speeds up processing. This is in line with our interpretation that this effect does not represent a PE and also with a general negative correlation of GBA and RT from previous reports (e.g. Hoogenboom et al., 2010).

The above interpretation of the interaction effect as a high-level PE, however, is questioned by the negative correlation of GBA with RT at these locations. Nevertheless, it’s possible that the consistency violation inducing the interaction is not performance relevant.

2.5.5. Conclusion

Our results strongly support the notion that PEs are signaled by increased high-frequency GBA (> 68 Hz) for violation of priors from life-long experience.

3. Information-theoretic evidence for predictive coding in the face-processing system

Authors

Alla Brodski-Guerniero[1], Georg-Friedrich Paasch[1], Patricia Wollstadt[1], Ipek Özdemir[1], Joseph T.Lizier[2], Michael Wibral[1]

Affiliations

[1]MEG Unit, Brain Imaging Center, J.W. Goethe University, Frankfurt a.M., Germany

[2]Complex Systems Research Group and Centre for Complex Systems, Faculty of Engineering & IT, The University of Sydney, NSW 2006, Australia

Acknowledgements

ABG received support by Ernst Ludwig Ehrlich Studienwerk (BMBF scholarship for graduate students). GFP received support by Villigst Studienwerk (BMBF scholarship for graduate students). The authors thank Saskia Helbling for fruitful discussions and for making the permutation ANOVA code available. The authors acknowledge support by SP3 of the Human Brain Project (EU Grant 604102).

3.1. Abstract

Predictive coding suggests that the brain infers the causes of its sensations by combining sensory evidence with internal predictions based on available prior knowledge. However, the neurophysiological correlates of (pre-)activated prior knowledge serving these predictions are still unknown. Based on the idea that such pre-activated prior knowledge must be maintained until needed we measured the amount of maintained information in neural signals via the active information storage (AIS) measure. AIS was calculated on whole-brain beamformer-reconstructed source time-courses from magnetoencephalography (MEG) recordings of 52 human subjects during the baseline of a Mooney

face/house detection task. Pre-activation of prior knowledge for faces showed as alpha- and beta-band related AIS increases in content specific areas; these AIS increases were behaviourally relevant in brain area FFA. Further, AIS allowed decoding of the cued category on a trial-by-trial basis. Our results support accounts showing that activated prior knowledge and the corresponding predictions are signalled in low-frequency activity (< 30 Hz).

3.2. Significance statement

Our perception is not only determined by the information our eyes/retina and other sensory organs receive from the outside world, but strongly depends also on information already present in our brains like prior knowledge about specific situations or objects. A currently popular theory in neuroscience, predictive coding theory, suggests that this prior knowledge is used by the brain to form internal predictions about upcoming sensory information. However, neurophysiological evidence for this hypothesis is rare – mostly because this kind of evidence requires making strong a-priori assumptions about the specific predictions the brain makes and the brain areas involved. Using a novel, assumption-free approach we find that face-related prior knowledge and the derived predictions are represented in low-frequency brain activity.

3.3. Introduction

In the last decade, predictive coding theory has become a dominant paradigm to organize behavioral and neurophysiological findings into a coherent theory of brain function (George and Hawkins, 2009; Friston, 2010; Huang and Rao, 2011; Clark, 2012; Hohwy, 2013). Predictive coding theory proposes that the brain constantly makes inferences about the state of the outside world. This is supposed to be accomplished by building hierarchical internal predictions based on prior knowledge which are compared to incoming information in order to continuously adapt these internal models (Mumford, 1992; Rao et al., 1999; Friston, 2005, 2010)

The postulated use of predictions for inference requires several preparatory steps: First, task relevant prior knowledge passively stored in synaptic weights needs to be transferred into activated prior knowledge, i.e. information stored in neural activity (see Zipser et al., 1993 for a distinction of active/passive storage). Subsequently, (pre-)activated prior knowledge needs to be maintained until needed and transferred as a prediction in top-down direction to a lower cortical area, where it will be matched with incoming information (e.g. Mumford, 1992; Friston, 2005, 2010).

With respect to the neural correlates of activated prior knowledge and predictions we know that the prediction of specific features or object categories increases fMRI BOLD activity in the brain region at which the feature or category is usually processed (Puri et al., 2009; Esterman and Yantis, 2009; Kok et al., 2014). However, little is known about how the maintenance of pre-activated prior knowledge and the corresponding transfer of predictions are actually implemented in neural activity proper.

As a first step towards resolving this issue a microcircuit theory of predictive coding has been put forward, suggesting internal predictions to be processed in deep cortical layers and to manifest and to be transferred in low-frequency neural activity (< 30 Hz) along descending fiber systems (Bastos et al., 2012).

This theory is in line with the findings of a spectral predominance of low-frequency neural activity in deep cortical layers (Buffalo et al., 2011) and the physiological findings linking feedback connections to alpha/beta frequency channels in monkeys (Fries et al., 2015) and humans (Michalareas et al., 2016). Recently, this microcircuit theory of predictive coding gained experimental support by neurophysiological studies showing the predictability of events to be associated with neural power in alpha (Bauer et al., 2014; Sedley et al., 2016) or beta frequencies (Pelt et al., 2016).

However, representation and signalling of pre-activated prior knowledge serving predictions has been difficult to investigate with classical analysis methods. One reason is that classical analysis methods require a-priori assumptions about which predictions specific brain areas are going to make – assumptions which might be very challenging to make beyond early sensory cortices and for complex experimental designs (Wibral et al., 2014, section 4.4, p. 9). Moreover, classical analysis methods do not allow quantifying the *amount* of pre-activated prior knowledge for predictions, as for instance diminished neural activity measured by fMRI, MEG/EEG may still come with less or more information being maintained in these signals. To overcome these problems we studied the maintenance and signalling of pre-activated prior knowledge for predictions using the information-theoretic measures of active information storage (AIS, see Methods in Lizier et al., 2012; also see Gómez et al., 2014 for an application to MEG), and transfer entropy (TE, Schreiber, 2000; Vicente et al., 2011). AIS measures the amount of information in the future of a process predicted by its past (predictable information) while TE measures the amount of directed information transfer between two processes (see Methods for details).

Using these information-theoretic measures we investigated the pre-activation of prior knowledge for face predictions in neural source activity reconstructed from MEG recordings of 52 human subjects. In order to induce the pre-activation of face-related prior knowledge, subjects were instructed to detect Faces in two-tone stimuli (Mooney and Ferguson, 1951; Cavanagh, 1991).

3.4. Methods

3.4.1. Basic concept and testable hypotheses

To study the neural correlates of pre-activated prior knowledge for face predictions we used the information-theoretic measures active information storage (AIS) and transfer entropy (TE) – measuring predictable information (see Methods in Lizier et al., 2012) and information transfer (Schreiber, 2000; Vicente et al., 2011), respectively.

The use of AIS and TE in our study is based on the following rationale: Since the brain will usually not know exactly when a prediction will be needed, it will maintain activated prior knowledge related to the content of the prediction over time. If there is a reliable neural code that maps between content and activity, maintained activated prior knowledge must be represented as maintained information content in neural signals, measurable by AIS (Figure 3.1.A).

Importantly, we do not suggest that predictable information in neural signals as measured by AIS measures the predictability of external events. Rather, we suggest that AIS can be used as a measure to detect increased predictable information in specific brain areas. This predictable information is bound to rise when prior knowledge is pre-activated based on perceptual demands and thereby becomes available for predictions.

Further, predictions based on prior knowledge are supposed to be transferred to hierarchically lower brain areas, where they can be matched with incoming information. This information transfer thus must be measurable via TE.

From this basic concept we derived five testable hypotheses about AIS and TE in the predictive coding framework:

1. When activated prior knowledge is maintained, predictable information as measured by AIS is supposed to be high in brain areas specific to the content of the predictions.
2. If the microcircuit theory of predictive coding is correct, maintenance of pre-activated prior knowledge should be reflected in alpha/beta frequencies, i.e., predictable information and alpha/beta power should correlate.
3. If maintenance of relevant prior knowledge is reflected by predictable information on a trial-by-trial basis, the content of predictions should be also decodable from AIS information on a trial-by-trial basis.
4. Information transfer related to predictions (i.e. signalling of pre-activated prior knowledge measured by TE) should occur in a top-down direction from brain areas showing increased predictable information, and should be reflected in alpha/beta band Granger causality.
5. As predictions based on pre-activated prior knowledge are known to facilitate performance, predictable information is supposed to correlate with behavioural parameters, if it reflects the relevant pre-activated prior knowledge.

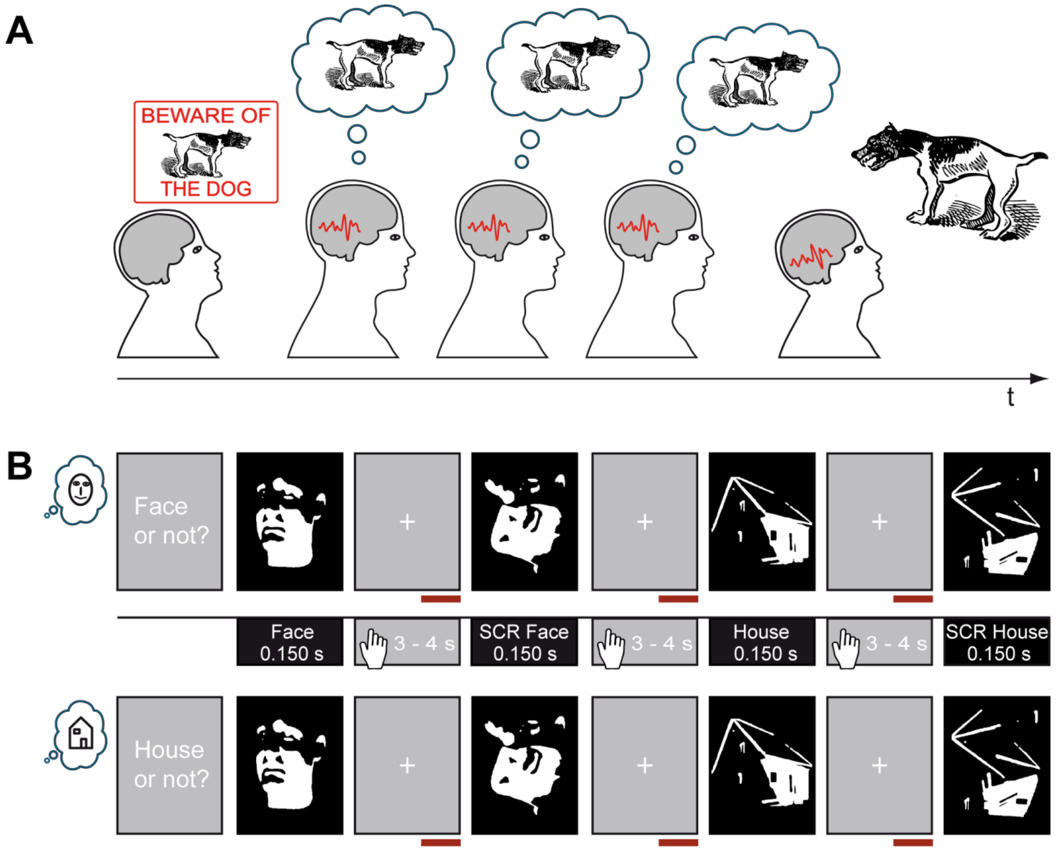

Figure 3.1. Central idea of the study and experimental design. ***A)*** Typically, pre-activated prior knowledge related to the content of a prediction has to be maintained as the brain will not know exactly when it will be needed. If there is a reliable neural code that maps between content and activity, maintained activated prior knowledge should lead to brain signals that are themselves predictable over time (here the brain signals are depicted as identical, although the relation between past and future will almost certainly be much more complicated). ***B)*** Exemplary stimulus presentation in Face blocks (top) and in House blocks (bottom). Face and House icons on the left indicate Face and House blocks, respectively. Middle**:** Depiction of stimulus categories and timing. The beginning of the response time window is indicated by the hand icon. Red horizontal bars mark the analysis interval. Figure elements obtained from OpenCliparts Library (http://www.openclipart.org) and modified. SCR – scrambled Mooney stimuli, not representing a face or house.

3.4.2. Subjects

57 subjects participated in the MEG experiment. 5 of these subjects had to be excluded due to excessive movements, technical problems, or unavailability of anatomical scans. 52 subjects remained for the analysis (average age: 24.8 years, SD 2.8, 23 males). Each subject gave written informed consent before the beginning of the experiment and was paid 10€ per hour for participation. The local ethics committee (Johann Wolfgang Goethe University clinics, Frankfurt, Germany) approved of the experimental procedure. All subjects had

normal or corrected-to-normal visual acuity and were right handed according to the Edinburgh Handedness Inventory scale (Oldfield, 1971). The large sample size subjects was chosen to reduce the risk of false positives, as suggested by (Button et al., 2013).

3.4.3. Stimuli and stimulus presentation

Photographs of faces and houses were transformed into two-tone (black and white) images known as Mooney stimuli (Mooney and Ferguson, 1951). Mooney stimuli were used based on the rationale that recognition of two-tone stimuli cannot be accomplished without relying on prior knowledge from previous experience, as is evident for example from the late onset of two-tone image recognition capabilities during development (> 4 years of age, Mooney, 1957) and from theoretical considerations (Kemelmacher-Shlizerman et al., 2008).

In order to increase task difficulty, in addition to Mooney faces and houses also scrambled stimuli (SCR) were created from each of the resulting Mooney faces and Mooney houses by displacing the white or black patches within the given background. Thereby all low-level information was maintained but the configuration of the face or house was destroyed. Examples of the stimuli can be seen in Figure 3.1.B.

All stimuli were resized to a resolution of 591x754 pixels. Stimulus manipulations were performed with the program GIMP (GNU Image Manipulation Program, 2.4, free software foundation, Inc., Boston, Massachusetts, USA).

A projector with a refresh rate of 60 Hz (resolution 1024x768 pixels) was used to display the stimuli at the center of a translucent screen (background set to gray, 145 cd/m^2). Stimulus presentation during the experiment was controlled using the Presentation software package (Version 9.90, Neurobehavioral Systems).

The experiment consisted of eight blocks of seven minutes. In each block 120 stimuli were presented (30 Mooney faces, 30 Mooney houses, 30 SCR faces, 30 SCR houses) in a randomized order. Stimuli were presented for 150 ms with a vertical visual angle of 24.1 and a horizontal visual angle of 18.8 degrees. The inter-trial-interval between stimulus presentations was randomly jittered from 3 to 4 seconds (in steps of 100 ms).

3.4.4. Task and Instructions

Subjects performed a detection task for faces or houses (Figure 3.1.B). Each of the eight experimental blocks started with the presentation of a written instruction; four of the experimental blocks started with the instruction “Face or not?” while for the other four experimental blocks started with the instruction “House or not?”. The former are referred to as “Face blocks” and the latter as “House blocks”. Face and House blocks were presented in alternating order. The same blocks of stimuli were presented as Face blocks for half of the subjects, while for the other half of the subjects these experimental blocks appeared as House blocks and vice versa. This way, the initial block was alternated between subjects (i.e. half of the subjects started with Face blocks and the other half with House blocks). Importantly, as the blocks contained the same face, house, SCR face and SCR house stimuli the only difference between face and house blocks was in the subjects’ instruction.

To avoid accidental serial effects, the order of blocks was reversed for half of the subjects. Subjects responded by pressing one of two buttons directly after stimulus presentation. The button assignment for a ‘Face’ or ‘No-Face’ response in Face blocks and 'House' or 'No-House' in House blocks was counterbalanced across subjects (n=26 right index finger for ‘Face’ response).

Between stimulus presentations, subjects were instructed to fixate a white cross on the center of the gray screen. Further, they were instructed to maintain fixation during the whole block and to avoid any movement during the acquisition session. Before data acquisition, subjects performed Face and House test blocks of two minutes with stimuli not used during the actual task. During the test blocks subjects received feedback on whether their response was correct or not. No feedback was provided during the actual task.

3.4.5. Data acquisition

MEG data acquisition was performed in line with recently published guidelines for MEG recordings (Gross et al., 2012). MEG signals were recorded using a whole-head system (Omega 2005; VSM MedTech Ltd.) with 275 channels. The signals were recorded continuously at a sampling rate of 1200 Hz in a synthetic third-order gradiometer configuration and were filtered online with fourth-order Butterworth filters with 300 Hz low pass and 0.1 Hz high pass.

Subjects' head position relative to the gradiometer array was recorded continuously using three localization coils, one at the nasion and the other two located 1 cm anterior to the left and right tragus on the nasion-tragus plane for 43 of the subjects and at the left and right ear canal for 9 of the subjects.

For artefact detection the horizontal and vertical electrooculogram (EOG) was recorded via four electrodes; two were placed distal to the outer canthi of the left and right eye (horizontal eye movements) and the other two were placed above and below the right eye (vertical eye movements and blinks). In addition, an electrocardiogram (ECG) was recorded with two electrodes placed at the left and right collar bones of the subject. The impedance of each electrode was kept below 15 kΩ.

Structural magnetic resonance (MR) images were obtained with either a 3T Siemens Allegra or a Trio scanner (Siemens Medical Solutions, Erlangen, Germany) using a standard T1 sequence (3-D magnetization-prepared-rapid-acquisition gradient echo sequence, 176 slices, 1 x 1 x 1 mm voxel size). For the structural scans vitamin E pills were placed at the former positions of the MEG localization coils for co-registration of MEG data and magnetic resonance images.

Behavioral responses were recorded using a fiberoptic response pad (Photon Control Inc. Lumitouch Control ™ Response System) in combination with the Presentation software (Version 9.90, Neurobehavioral Systems).

3.4.6. Statistical analysis of behavioral data

Responses were classified as correct or incorrect based on the subject’s first answer. For hit rate analysis the accuracy for each condition was calculated. For reaction time analysis only correct responses were considered.

Post-hoc Wilcoxon signed rank tests were performed on hit rates as well as reaction times. To account for multiple testing, Bonferroni correction was applied (uncorrected alpha = 0.05).

3.4.7. MEG-data preprocessing

MEG Data analysis was performed with Matlab (RRID:nlx_153890; Matlab 2012b, The Mathworks, Inc.) using the open source Matlab toolbox Fieldtrip (RRID:nlx_143928; Oostenveld et al., 2011; Version 2013 11-11) and custom Matlab scripts.

Only trials with correct behavioral responses were taken into account for MEG data analysis. The focus of data analysis was on the prestimulus intervals from 1 s to 0.050 s before stimulus onset. Trials containing sensor jump-, or muscle-artefacts were rejected using automatic FieldTrip artefact rejection routines. Line noise was removed using a discrete Fourier transform filter at 50,100 and 150 Hz. In addition, independent component analysis (ICA; Makeig et al., 1996) was performed using the extended infomax (runica) algorithm implemented in fieldtrip/EEGLAB. ICA components strongly correlated with EOG and ECG channels were removed from the data. Finally, data was visually inspected for residual artefacts.

In order to minimize movement related errors, the mean head position over all experimental blocks was determined for each subject. Only trials in which the head position did not deviate more than 5 mm from the mean head position were considered for further analysis.

As artefact rejection and trial rejection based on the head position may result in different trial numbers for Face and House blocks, after trial rejection the minimum amount of trials across Face and House blocks was selected randomly from the available trials in each block (stratification).

3.4.8. Sensor level spectral analysis

Spectral analysis at the sensor level was performed in order to determine the subdivision of the power spectrum in frequency bands (see Brodski et al., 2015 for a similar approach). As we aimed to identify frequency bands based on stimulus related increases or decreases, respectively, new data segments were cut from -0.35 to -0.05s before stimulus onset for the time interval of “baseline” and from 0.05 s to 0.35 after stimulus onset for the interval of “task”. Before spectral transformation a single Hanning taper was applied to the data. The spectral transformation was calculated in an interval from 4 to 150 Hz using a fast Fourier approach. Average spectra of task and baseline periods were contrasted over subjects using a dependent-sample permutation t-metric with a cluster based correction method (Maris and Oostenveld, 2007) to account for multiple comparisons. Adjacent samples whose *t*-values exceeded a threshold corresponding to an uncorrected α-level of 0.05 were defined as clusters. The resulting cluster sizes were then tested against the distribution of cluster sizes obtained from 1000 permuted datasets (i.e. labels “task” and “baseline” were randomly reassigned within each of the subjects). Cluster sizes larger than the 95th percentile of the cluster sizes in the permuted datasets were defined as significant.

3.4.9. Source grid creation

In order to create individual source grids we transformed the anatomical MR images to a standard T1 MNI template from the SPM8 toolbox (http://www.fil.ion.ucl.ac.uk/spm) – obtaining an individual transformation matrix for each subject. We then warped a regular 3-D dipole grid based on the standard T1 template (spacing 15 mm resulting in 478 grid locations) with the inverse of each subjects’ transformation matrix, to obtain an individual dipole grid for each subject in subject space. This way, each specific grid point was located at the same brain area for each subject, which allowed us to perform source analysis with individual head models as well as multi-subject statistics for all grid locations. Lead-fields at those grid locations were computed for the individual subjects with a realistic single shell forward model (Nolte, 2003) taking into account the effects of the ICA component removal in pre-processing.

3.4.10. Source time course reconstruction

To enable a whole brain analysis of active information storage (AIS), we reconstructed the source time courses for all 478 source grid locations.

For source time course reconstruction we calculated a time-domain beamformer filter (linear constrained minimum variance, LCMV; Van Veen et al., 1997) based on broadband filtered data (8 Hz high pass, 150 Hz low pass) from the prestimulus interval (-1 s to -0.050 s) of Face blocks as well as House blocks (use of common filters – see Gross et al., 2012, page 357).

For each source location three orthogonal filters were computed (x, y, z direction). To obtain the source time courses, the broadly filtered raw data was projected through the LCMV filters resulting in three time courses per location. On these source time courses we performed a singular value decomposition to obtain the time course in direction of the dominant dipole orientation. The source time course in direction of the dominant dipole orientation was used for calculation of active information storage (AIS).

3.4.11. Definition of active information storage

We assume that the reconstructed source time courses for each brain location can be treated as realizations $\{x_1, \ldots, x_t, \ldots, x_N\}$ of a random process $X = \{X_1, \ldots, X_t, \ldots, X_N\}$, which consists of a collection of random variables, X_t, ordered by some integer t. AIS then describes how much of the information in the next time step t of the process is predictable from its immediate past state (Lizier et al., 2012). This is defined as the mutual information

$$A_x = \lim_{k \to \infty} I\left(X_{t-1}^{k}; X_t\right) = \lim_{k \to \infty} \sum_{x_t, x_{t-1}^{k}} p\left(x_t, x_{t-1}^{k}\right) \log \frac{p\left(x_{t-1}^{k}, x_t\right)}{p\left(x_{t-1}^{k}\right) p(x_t)} \qquad (1)$$

where I is the mutual information and $p(.)$ are the variables' probability density functions. Variable X_{t-1}^{k} describes the past *state* of X as a collection of past random variables $X_{t-1}^{k} = \left\{X_{t-1}, \ldots, X_{t-1-(k*\tau)}\right\}$, where k is the embedding dimension, i.e., the number of time steps used in the collection, and τ the embedding delay between these time steps. For practical purposes, k has to be set to a finite value k_{max}, such that the history before time point $t - k_{\mathrm{max}} * \tau$ does (statistically) not further improve the prediction of X_t from its past (Lizier et al., 2012).

Predictable information as measured by AIS indicates that a signal is both rich in information and predictable at the same time. Note that neither a constant signal (predictable but low information content) nor a memory-less stochastic process (high information content but unpredictable) will exhibit high AIS values. In other words, a neural process with high AIS must visit many different possible states (rich dynamics); yet visit these states in a predictable manner with minimal branching of its trajectory (this is the meaning of the log ratio of equation (1)). As such, AIS is a general measure of information that is maintained in a process, and could here reflect any form of memory based on neural activity. AIS is linked specifically to activated prior knowledge in our study via the experimental manipulation that alternately activates face- or house-specific prior knowledge, and by investigating the difference in AIS between the two conditions.

3.4.12. Analysis of predictable information using active information storage

The history dimension (k_{max}; range 3 to 6) and optimal embedding delay parameter (tau; range 0.2 to 0.5 in units of the autocorrelation decay time) was determined for each source location separately using Ragwitz' criterion (Ragwitz and Kantz, 2002), as implemented in the TRENTOOL toolbox (Lindner et al., 2011). To avoid a bias in estimated values based on different history dimensions, we chose the maximal history dimension across Face and House blocks for each source location (median k_{max} over source locations and subjects = 4).

The actual spacing between the time-points in the history was the median across trials of the output of Ragwitz' criterion for the embedding delay tau (Lindner et al., 2011).

Based on the assumption of stationarity in the prestimulus interval, AIS was computed on the embedded data across all available time points and trials. This was done separately for each source location and condition in every subject.

Computation of AIS was performed using the Java Information Dynamics Toolkit (Lizier, 2014). A minimum of 68400 samples entered the AIS analysis for each subject, block type and source location (minimum of 57 trials, approx. 1 sec time interval, sampling rate 1200 Hz). AIS was estimated with 4 nearest neighbours in the joint embedding space using the Kraskov-Stoegbauer-Grassberger (KSG) estimator (Kraskov et al., 2004; algorithm 1), as implemented in the open source Java Information Dynamics Toolkit (JIDT; Lizier, 2014).

Computation of AIS was performed at the Center for Scientific Computing (CSC) Frankfurt, using the high-performance computing Cluster FUCHS (https://csc.uni-frankfurt.de/index.php?id=4), which enabled the computationally demanding calculation of AIS for the whole brain across all subjects as well as Face and House blocks (478 x 52 x 2 = 49712 computations of AIS).

3.4.13. AIS Statistics

In order to determine the source locations in which AIS values were increased when subjects held face information in memory, a within-subject permutation t-metric was computed. Here, AIS values for each source location across all subjects were contrasted for Face blocks and House blocks. The permutation test was chosen as the distribution of AIS values is unknown and not assumed to be Gaussian. To account for multiple comparisons across the 478 source locations, a cluster-based correction method (Maris and Oostenveld, 2007) was used. Clusters were defined as adjacent voxels whose t-values exceeded a critical threshold corresponding to an uncorrected alpha level of 0.01. In the randomization procedure labels of Face block and House block data were randomly reassigned within each subject. Cluster sizes were tested against the distribution of cluster sizes obtained from 5000 permuted data sets. Cluster values larger than the 95th percentile of the distribution of cluster sizes obtained for the permuted data sets were considered to be significant.

3.4.14. Correlation analysis of spectral properties and AIS

We investigated the relationship of spectral power in the prestimulus interval and AIS values on the single trial level. Before calculation of single trial spectral power, a single Hanning taper was applied to each prestimulus epoch. Then, single trial spectra were computed with the fast Fourier approach, were averaged over all epochs and subdivided in the predefined frequency bands for each subject. Next, Spearman's rho was computed for correlation of the median single trial spectral power in the predefined frequency bands with the single trial AIS values in order to obtain individual correlation values. Median correlation values over both block types were computed for each subject. In order to test the significance of the correlation analysis, for each subject the epochs were randomly permuted 5000 times and correlation was re-calculated also for the permuted data sets. For each subject an original correlation value larger (or smaller) than 99.99997% (threshold Bonferroni adjusted for the 52*5*6 multiple comparisons) of the correlation values obtained for the permuted data sets was considered to be significant. At the second level we used a binomial test to assess whether the number of subjects showing significant correlations (for one source and frequency range) could be explained by chance. Median correlation values over subjects and their significance based on the binomial test are reported.

We also calculated a correlation of two t-value maps: (1) the mean AIS contrast and (2) a mean power contrast. For both t-value maps the dependent samples t-metric Face blocks vs. House blocks was computed over all 52 subjects and all 478 source locations inside the brain. For the power t-value map, source power in the alpha (8-14 Hz) and beta (14-32 Hz) frequency band was reconstructed with the DICS (dynamic imaging of coherent sources, Gross et al., 2001) algorithm as implemented in the FieldTrip toolbox using real valued filter coefficients only (see also Grützner et al., 2010).

3.4.15. Correlation analysis of reaction times and AIS

Last, we assessed the relationship of AIS values and reaction times for each subject. To this end, before the correlation analysis, for each subject mean reaction times and mean AIS values in the brain areas of interest for Face and House blocks were subtracted from each other. This allowed accounting for differential behavioral speed between subjects. The correlation of the difference in AIS values and the difference in reaction times was calculated via Spearman skipped correlations using the Robust Correlation Toolbox (Pernet et al., 2013). Calculation of skipped correlations includes identifying and removing bivariate outliers (Rousseeuw, 1984; Rousseeuw and Driessen, 1999; Verboven and Hubert, 2005). This can provide a more robust measure, which has been recommended for brain-behaviour correlation analyses (Rousselet and Pernet, 2012). The uncorrected alpha level was set to 0.05. For each correlation bootstrap confidence intervals (CIs) were computed based on 1000 resamples. In order to account for multiple comparisons across brain areas, bootstrap CIs were adjusted using Bonferroni correction. If the adjusted CI did not encompass 0, the correlation was considered as significant.

3.4.16. Decoding analysis

To investigate whether prediction content (i.e. face or house block) can be decoded from individual trial AIS values, we applied a multivariate analysis using support vector machines (SVMs) with the libsvm toolbox (Chang and Lin, 2011; available at http://www.csie.ntu.edu.tw/~cjlin/libsvm). For each subject the linear SVM classifier was trained using 70% randomly chosen trials as training data. However, the training data contained always the same amount of trials for face and house blocks, respectively. Parameters for the SVMs were optimized in a three-fold cross-validation procedure for the training data only. Subsequently, the classifier was tested using the data from the remaining 30% of the trials with the best parameters obtained from the training procedure, thereby ensuring strict separation of training and testing data (Nowotny, 2014).

This procedure was repeated 10 times. We report the median accuracy value for each subject. In order to test the significance of the median accuracy value, for each subject the labels of face blocks and house blocks were randomly permuted 500 times for each of the 10 training and testing sets and the median over the 10 accuracy values was calculated also for the permuted data sets. A median accuracy value larger than the 99.999% (threshold Bonferroni adjusted for the 52 multiple comparisons) of the median accuracy values obtained for the permuted data sets was considered to be significant, corresponding to an uncorrected alpha level of 0.05.

3.4.17. Definition of transfer entropy (and Granger analysis)

Transfer entropy (TE, Schreiber, 2000) was applied to investigate the information transfer between the brain areas identified with AIS analysis. For links with significant information transfer, we post-hoc studied the spectral fingerprints of these links using spectral Granger analysis (Granger, 1969).

Both, TE and Granger analysis are implementations of Wiener's principle (Wiener, 1956) which in short can be rephrased as follows: If the prediction of the future of one time series X, can be improved in comparison to predicting it from the past of X alone by adding information from the past of another time series Y, then information is transferred from Y to X.

TE is an information-theoretic, model-free implementation of Wiener's principle and can be used, in contrast to Granger analysis, in order to study linear as well as non-linear interactions (e.g. Chang and Lin, 2011) and was previously applied to broadband MEG source data (Wibral et al., 2011). TE is defined as a conditional mutual information

$$TE_{Y \to X} = \lim_{j,k \to \infty} I(\, X_t; Y_{t-u}^{j} | X_{t-1}^{k})$$

$$= \lim_{j,k \to \infty} \sum\nolimits_{x_t, x_{t-1,}^{k} y_{t-u}^{j}} p(x_t, x_{t-1,}^{k} y_{t-u}^{j}) \ \log \frac{(p(x_t|, x_{t-1,}^{k} y_{t-u}^{j})}{p(x_t|, x_{t-1,}^{k})} \qquad (2)$$

where X_t describes the future of the target time series X, X_{t-1}^{k}describes the past state of X, and Y_{t-u}^{j} describes the past state of the source time series Y . As for the calculation of AIS, past states are defined as collections of past random variables with number of time steps j and k and a delay τ. The parameter u accounts for a physical delay between processes Y and X (Wibral et al., 2013) and can be optimized by finding the maximum TE over a range of assumed values for u.

3.4.18. Analysis of information transfer using transfer entropy and Granger causality analysis

We performed TE analysis with the open-source Matlab toolbox TRENTOOL (Lindner et al., 2011), which implements the KSG-estimator (Kraskov et al., 2004; Frenzel and Pompe, 2007; Gómez-Herrero et al., 2015) for TE estimation. We used ensemble estimation (Wollstadt et al., 2014; Gómez-Herrero et al., 2015), which estimates TE from data pooled over trials to obtain more data and hence more robust TE-estimates. Additionally, we used Faes' correction method to account for volume conduction (Faes et al., 2013).

In the TE analysis the same time intervals (prestimulus) and embedding parameters as for AIS analysis were used. TE values for Face blocks and House blocks were contrasted using a dependent-sample permutation t-metric for statistical analysis across subjects. In the statistical analysis, Bonferroni correction was used to account for multiple comparisons across links (uncorrected alpha level 0.05). As for AIS, the history dimension for the past states was set to finite values; we here set $j_{\max} = k_{\max}$ and used the values obtained during AIS estimation for the target time series of each signal combination.

For the significant TE links post-hoc nonparametric bivariate Granger causality analysis in the frequency domain (Dhamala et al., 2008) was computed. Using the nonparametric variant of Granger causality analysis avoids choosing an autoregressive model order, which may easily introduce a bias. In the nonparametric approach Granger causality is computed from a factorization of the spectral density matrix, which is based on the direct Fourier transform of the time series data (Dhamala et al., 2008). The Wilson algorithm was used for factorization (Wilson, 1972). A spectral resolution of 2 Hz and a spectral smoothing of 5 Hz were used for spectral transformation using the multitaper approach (Percival and Walden, 1993) (9 Slepian tapers). We were interested in the differences of Granger spectral fingerprints of Face and House blocks, however we also wanted to make sure that the Granger values for these differences significantly differed from noise. For that reason we created two additional “random” conditions by permuting the trials for the Face block and the House block condition for each source separately. Two types of statistical comparisons were performed for the frequency range between 8 and 150 Hz and each of the significant TE links: 1. Granger values in Face blocks were contrasted with Granger values in House blocks using a dependent-samples permutation t-metric 2. Granger values in Face blocks / House blocks were contrasted with the random Face block condition / random House block condition using another dependent-samples permutation t-metric. For the first test a cluster-correction was used to account for multiple comparisons across frequency (Maris and Oostenveld, 2007). Adjacent samples which uncorrected p-values were below 0.01 were considered as clusters. 5000 permutations were performed and the alpha value was set to 0.05. Frequency intervals in the Face block vs. House block comparison were only considered as significant if all included frequencies also reached significance in the comparison with the random conditions using a Bonferroni correction. Last, Bonferroni correction was also applied to account for multiple comparisons across links.

3.1. Results

3.5.1. Behavioral results

We found no differences between Face blocks and House blocks for hit rates (avg. hitrate Face blocks 93.9%; avg. hitrate House blocks 94.6%; Wilcoxon Signed rank test $p = 0.57$) and reaction times of correct responses (avg. mean reaction times Face blocks 0.545 s, avg. reaction times House blocks 0.546 s; Wilcoxon Signed rank test $p = 0.85$). For both block types subjects showed decreased hit rates and increased reactions times for the instructed intact stimulus (i.e. face in Face blocks and house in House blocks) compared to the non-instructed intact stimulus (house in Face blocks and face in House blocks), as the instructed intact stimuli had to be distinguished from a similar distractor (SCR stimuli; Figure 3.2.). Also, slower reaction times were found for the instructed intact stimulus vs. the non-instructed scrambled stimulus for both block types. Moreover, for both block types subjects showed lower hit rates for houses than SCR houses (see Figure 3.2. for results of behaviour analysis).

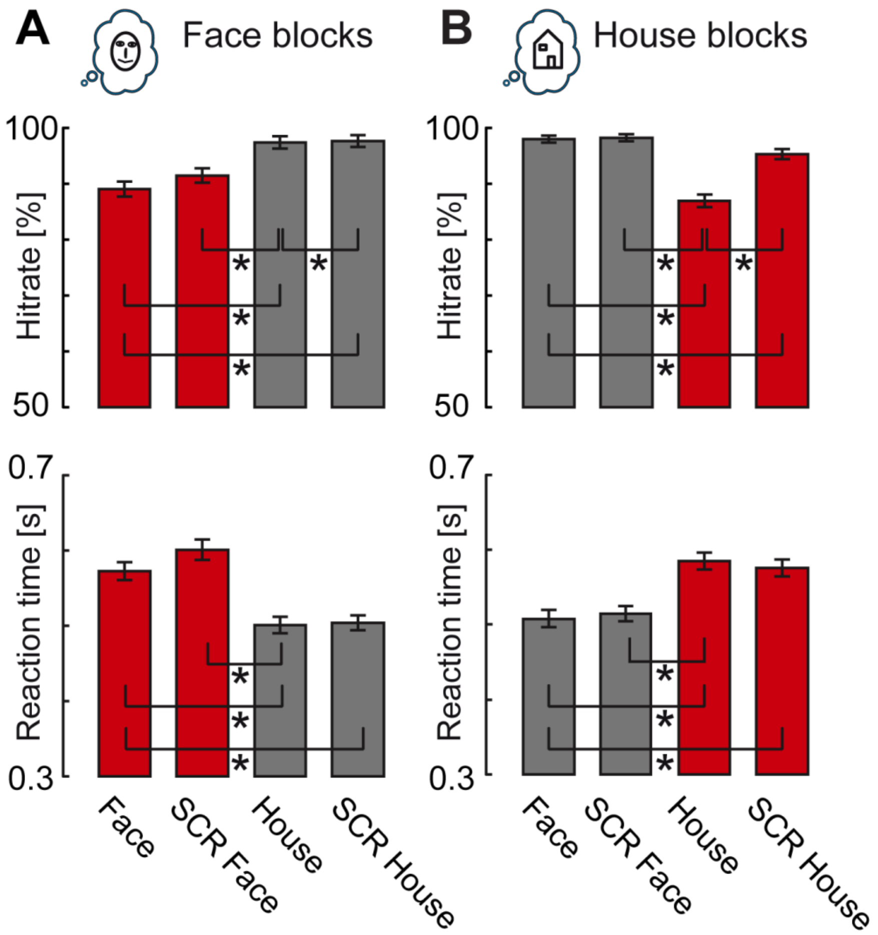

Figure 3.2. Behavioral results. Depiction of hit rates and reaction times of correct responses for ***(A)*** Face blocks and ***(B)*** House blocks. Equivalent conditions in different block types are marked in red and grey, respectively. Asterisks indicate significant differences based on Wilcoxon signed-rank tests within block type (n = 52; Bonferroni corrected for multiple comparisons). Error bars indicate standard error. SCR – scrambled Mooney stimuli.

3.5.2. Definition of frequency bands

Following the same approach as Brodski and colleagues (2015), we defined frequency bands for subsequent neural analysis based on the significant clusters of a task vs. baseline contrast at the MEG sensor level. This analysis was based on the spectra of all conditions for both block types and revealed one positive cluster with task-related increases in activity and one negative cluster with task related decreases in activity (Figure 3.3.). Based on the spectral profile of the two significant clusters, the following six frequency bands were defined for further analysis: (1) 8–14 Hz (alpha); (2) 14–32 Hz (beta); (3) 32–50 Hz (low gamma), (4) 50–60 Hz (mid gamma), (5) 60–100 Hz (high gamma) and (6) 100-150 Hz (very high gamma).

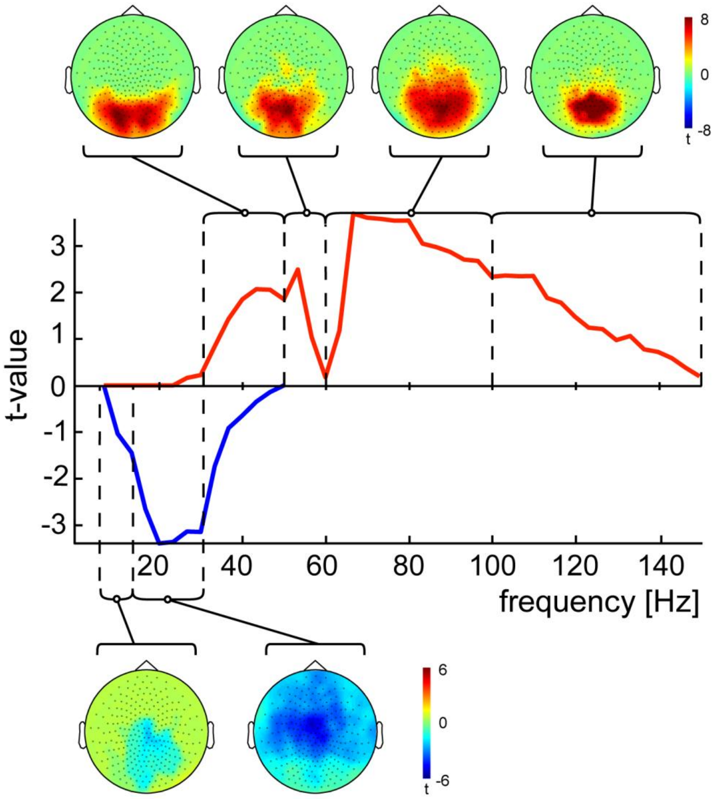

Figure 3.3. Sensor-level frequency analysis – defining frequency bands. Middle: Power spectra for all of the significant clusters (one positive and one negative cluster) at the sensor level (permutation t-metric, contrast [0.05s 0.35s] vs. [-0.35s -0.05s] around stimulus onset, t values masked by $p < 0.05$, cluster correction, $n = 52$). Frequency analysis at the sensor level was calculated using both blocks types jointly. Task-related increases in power are shown in red (positive cluster) and task-related decreases in blue (negative cluster). Black dashed lines frame the identified frequency ranges. Top and bottom: Topographical plots of the task-related increases or decreases for each defined frequency range.

3.5.3. Analysis of predictable information

Statistical comparisons of AIS values between Face blocks and House blocks in the prestimulus interval revealed increased AIS values for Face blocks in clusters in fusiform face area (FFA), anterior inferior temporal cortex (aIT), occipital face area (OFA), posterior parietal cortex (PPC) and primary visual cortex (V1) (Figure 3.4.). We referred to these five brain areas as “face prediction network” and subjected it to further analyses. In contrast to this finding of a face prediction network, we did not find brain areas showing significantly higher AIS values in House blocks compared to Face blocks. This is similar to highly cited previous studies that failed to find prediction effects for houses in the brain in contrast to faces (e.g. Summerfield et al., 2006a, 2006b; Trapp et al., 2015).

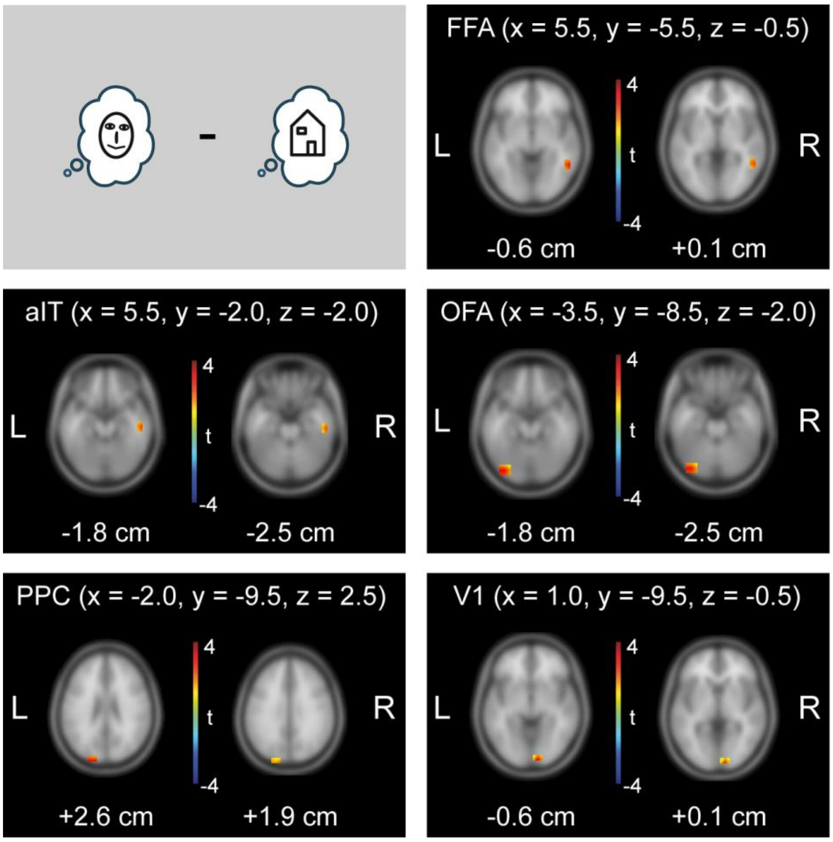
-
FFA (x = 5.5, y = -5.5, z = -0.5)
L
R
4
t
-4
-0.6 cm
+0.1 cm
aIT (x = 5.5, y = -2.0, z = -2.0)
L
R
4
t
-4
-1.8 cm
-2.5 cm
OFA (x = -3.5, y = -8.5, z = -2.0)
L
R
4
t
-4
-1.8 cm
-2.5 cm
PPC (x = -2.0, y = -9.5, z = 2.5)
L
R
4
t
-4
+2.6 cm
+1.9 cm
V1 (x = 1.0, y = -9.5, z = -0.5)
L
R
4
t
-4
-0.6 cm
+0.1 cm

Figure 3.4. Statistical analysis of predictable information (measured by AIS) at the MEG source level. Results of whole-brain dependent samples permutation t-metric contrasting Face blocks and House blocks (n=52, t-values masked by $p < 0.05$, cluster correction). Peak voxel coordinates in MNI space are shown at the top for each brain location; z-values are displayed below each brain slice. OFA = occipital face area; FFA = fusiform face area; aIT= anterior inferior temporal cortex; PPC = posterior parietal cortex; V1 = primary visual cortex.

3.5.4. Correlation of single trial power and single trial predictable information

In order to investigate the neurophysiological correlates of activated prior knowledge identified via AIS analysis, a correlation analysis of single trial power in distinct frequency bands with single trial AIS was conducted. Correlation analysis revealed significant positive correlations in the alpha and beta frequency bands (Table 3.1.). This means that alpha and beta band activity is the most likely carrier of activated prior knowledge. Additionally, for two of the brain areas we also found a weak negative correlation of single trial very high gamma power and AIS. However, the tiny effect size of the very high gamma correlation questions the relevance of this effect. We will therefore only discuss the findings in the alpha and beta band.

Table 3.1. Correlation of single trial power and single trial predictable information (measured by AIS) in the face prediction network

	FFA	*aIT*	*V1*	*OFA*	*PPC*
Alpha (8-14 hz)	rho = 0.46*	rho = 0.46*	rho = 0.49*	rho = 0.47*	rho = 0.47*
Beta (14-32 hz)	rho = 0.33*	rho = 0.34*	rho = 0.31*	rho = 0.33*	rho = 0.3*
Low gamma (32-50 hz)	rho = 0.07	rho = 0.07	rho = 0.08	rho = 0.07	rho = 0.09
Mid gamma (50-60 hz)	rho = 0.03	rho = 0.01	rho = 0.02	rho = 0.02	rho = 0.04
High gamma (60-100 hz)	rho = -0.007	rho = -0.02	rho = 0.01	rho = 0.003	rho = 0.05
Very high gamma (100-150 Hz)	rho = -0.13	rho =-0.16*	rho= -0.12	rho = -0.13*	rho = -0.11

** = significant, based on binomial test*

While we found a significant correlation of single trial power and predictable information in the alpha and beta band, the contrast map based on mean beamformer reconstructed source power over all source grid points for Face and House blocks (t-values obtained from dependent sample t-metric over all 52 subjects) did not correlate with the mean AIS contrast map for both, alpha and beta power (alpha rho = 0.043, p = 0.33; beta rho = 0.05, p = 0.21; Figure 3.5.). This suggests that AIS analysis provides additional information not directly provided by a spectral analysis. In other words, while AIS seems to be carried by alpha/beta-band activity, not all alpha/beta-band activity contributes to AIS.

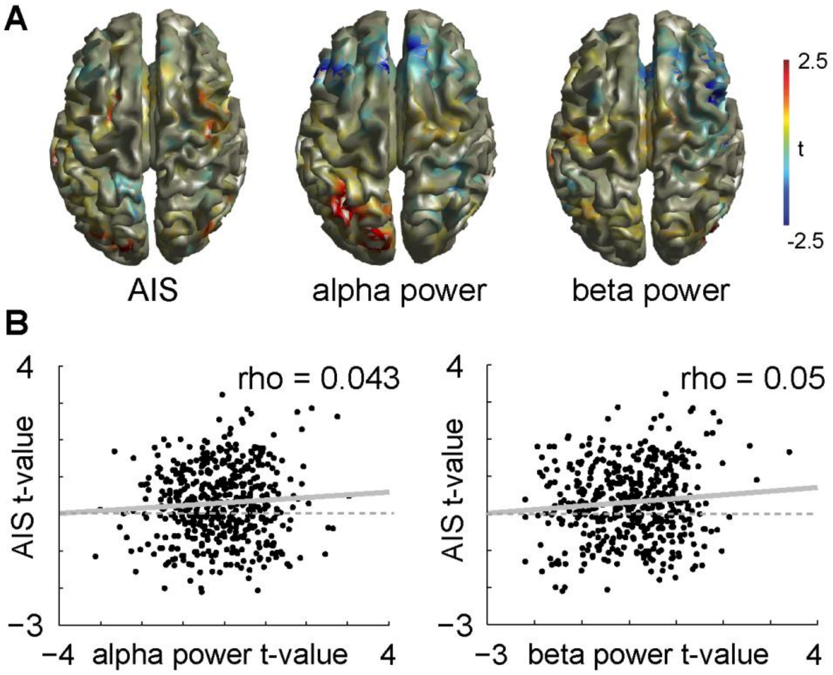

Figure 3.5. Correlation of predictable information contrast maps and source power contrast maps*. A)* Illustration of the t-value maps of the dependent samples t-metric for the Face block vs. House block contrast (n = 52, no correction) on the cortical surface. ***B)*** Scatter plots of the relationship of the alpha/beta contrast and the AIS contrast. Each dot represents a source location within the brain. Spearman correlation values are displayed at the top right corner of each plot (n = 478). Linear regression lines are included in gray (solid).

3.5.5. Decoding prediction content from single trial AIS values

To study whether face or house predictions can be decoded from AIS values of the face prediction network on a trial-by-trial basis, support vector machines were used (Chang and Lin, 2011). Cross-validated decoding performance reached a maximum of 65.2% (mean performance 53.5%, SD 3.9% over subjects). When Bonferroni correcting for the high number of subjects tested (n = 52), for 22 of the 52 subjects performance was still significantly better than for permuted datasets ($p < 0.05/52$). Note, that this fraction is much higher than would have been expected by chance ($p = 1.1 \times 10^{-52}$, binomial test).

3.5.6. Analysis of information transfer

To understand how activated prior knowledge is communicated within the cortical hierarchy, we assessed the information transfer within the face prediction network in the prestimulus interval by estimating transfer entropy (TE, Schreiber, 2000) on source time courses for Face blocks and House blocks, respectively. Statistical analysis revealed significantly increased information transfer for Face blocks from aIT to FFA (p = 0.0001, Bonferroni correction) and from PPC to FFA (p = 0.0014, Bonferroni correction). For House blocks information transfer was increased in comparison to Face blocks from brain area V1 to PPC (p = 0.0014, Bonferroni correction) (Figure 3.6.).

Post-hoc frequency-resolved granger causality analysis did not reveal any significant effects.

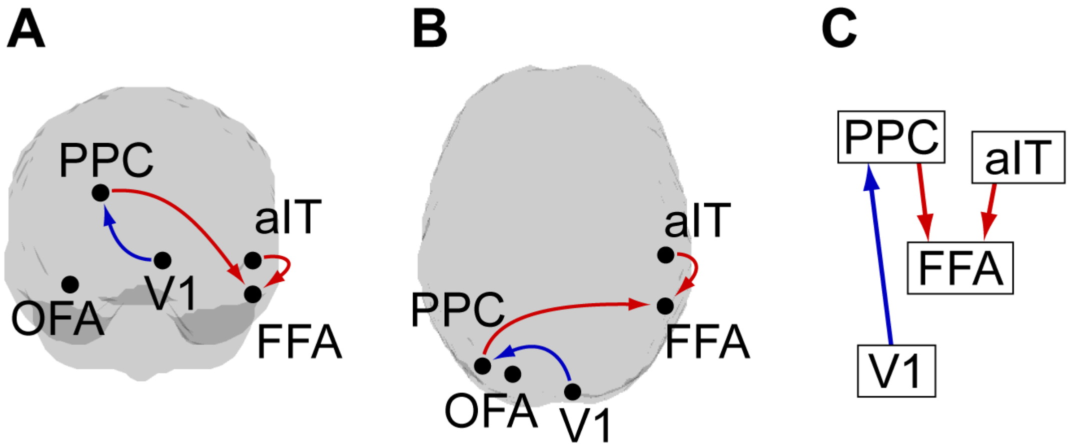

Figure 3.6. Analysis of information transfer in the prestimulus interval. Results of dependent sample permutation t-tests on transfer entropy (TE) values (Face blocks vs. House blocks, n = 52, p < 0.05, Bonferroni corrected). Red arrows indicate increased information transfer for Face blocks; blue arrows indicate increased information transfer for House blocks. Illustration of the resulting network in ***A)*** a view from the back of the brain, ***B)*** view from the top of the brain, ***C)*** depiction of the network hierarchy (based on the hierarchy in Zhen et al., 2013; Michalareas et al., 2016).

3.5.7. Correlation of predictable information and reaction times

In order to study the association of predictable information and behaviour, we correlated the per subject difference of AIS values between Face blocks and House blocks with the per subject difference in reaction times. This analysis was performed for the three brain areas between which we found increased information transfer during Face blocks (FFA, aIT and PPC). For these brain areas we tested the hypothesis that predictable information for face blocks was associated with performance, i.e. reaction times during Face blocks. Negative correlation values were found for all of the three brain areas, however only brain area FFA reached significance when correcting for multiple comparisons (Figure 3.7.): FFA robust Spearman's rho -0.41, robust confidence interval (CI) after correcting for multiple comparisons [-0.68 -0.066]; aIT robust Spearmans rho = -0.12, CI [-0.4554 0.245]; PPC robust Spearman's rho -0.21 CI [-0.5480 0.1178].

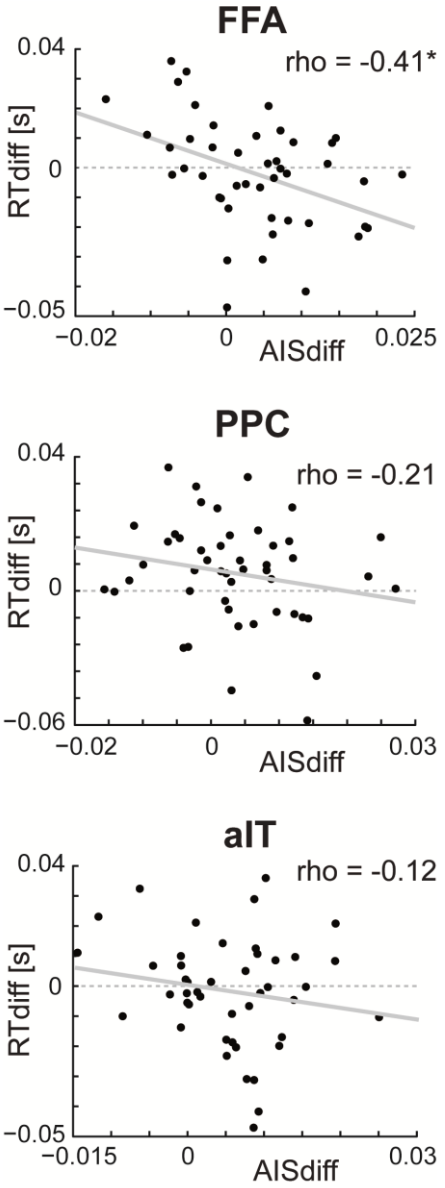

Figure 3.7. Correlation analysis for predictable information and reaction times. Scatter plots displaying the (skipped) correlation of per subject AIS difference values (Face blocks – House blocks) with per subject reaction time difference values (Face blocks – House blocks). Robust Spearman correlation values are displayed at the top right corner of each plot. Asterisks indicate significant correlation, using Bonferroni correction of bootstrap confidence intervals. Linear regression lines are included in gray (solid).

3.6. Discussion

We tested the hypothesis that the neural correlates of prior knowledge activated for use as an internal prediction must show as predictable information in the neural signals carrying that activated prior knowledge. This hypothesis is based on the rationale that the content of activated prior knowledge must be maintained until the knowledge or the prediction derived from it is used. The fact that activated prior knowledge has a specific content then mandates that increases in predictable information should be found in brain areas specific to processing the respective content. This is indeed what we found when investigating the activation of prior knowledge about faces during face detection blocks. In these blocks, predictable information was selectively enhanced in a network of well-known face processing areas. At these areas prediction content was decodable from the predictable information on a trial-by-trial basis and increased predictable information was related to improved task performance in brain area FFA. Given this established link between the activation of prior knowledge and predictable information we then tested current neurophysiological accounts of predictive coding suggesting that activated prior knowledge should be represented in deep cortical layers and at alpha or beta-band frequencies and should be communicated as a prediction along descending fiber pathways (Bastos et al., 2012). Indeed, predictable information within the network of brain areas related to activated prior knowledge of faces was associated with alpha and beta-band frequencies and information transfer within this network was increased in top-down direction– in accordance with the theory.

We will next discuss our findings with respect to their implications for current theories of predictive coding.

3.6.1. Activated prior knowledge for faces shows as predictable information in content specific areas

We found increased predictable information as reflected by increased AIS values in Face blocks in the prestimulus interval in FFA, OFA, aIT, PPC and V1.

Out of these five brain areas FFA, OFA and aIT are well known to play a major role in face processing (Kanwisher et al., 1997; Kriegeskorte et al., 2007; Tsao et al., 2008; Pitcher et al., 2011).

It might seem surprising that predictable information for Face blocks was not increased within superior temporal gyrus (STS), a brain area which has been recently identified as a key region for the prediction of face identities in a face identity recognition task (Apps and Tsakiris, 2013). This finding may be explained by the specific role of STS in face processing – mainly processing facial *identities* and emotional expressions (Winston et al., 2004; Fox et al., 2009). In contrast, the STS may play a lesser role in the pure face detection task of our design where neither identities nor emotional expressions were of relevance.

In addition to increased predictable information in well-known face processing areas we also found increased predictable information in Face blocks in PPC. We consider the increase in predictable information in PPC also as content-specific, because regions in PPC have been recently linked to high-level visual processing of objects like faces (Pashkam and Xu, 2014) and activation of PPC has been repeatedly observed during the recognition of Mooney faces by us and others (Dolan et al., 1997; Grützner et al., 2010; Brodski et al., 2015).

In sum, our finding of increased predictable information for Face blocks in FFA, OFA, aIT and PPC confirms our hypothesis that activation of face prior knowledge elevates predictable information in content specific areas. Additionally, our results suggest that predictable information in content-specific areas is associated with the corresponding prediction on a trial-by-trial basis – by decoding the anticipated category (Face or House block) from trial-by-trial AIS values at the face prediction areas.

However, while we found increased predictable information in content specific areas for Face blocks, we did not find brain areas showing increased predictable information for House blocks. Similarly, Summerfield and colleagues (2006b) observed in a face/house discrimination task increased activation in FFA, when a house was misperceived as a face – but failed to see increased activation in parahippocampal place area (PPA), a scene/house responsive region, when a face was misperceived as a house. The authors suggest that this might be related to the fact that PPA is less subject to top-down information than FFA – as faces have much more regularities potentially utilizable for top-down mechanisms than the natural scenes that PPA usually responds to. Additionally, because of their strong social relevance (e.g. Farah et al., 1995) faces capture attention disproportionally (e.g. Vuilleumier and Schwartz, 2001). Thus, also face predictions/templates may be prioritized in comparison to other templates e.g. for houses (Esterman and Yantis, 2009; Puri et al., 2009; Van Belle et al., 2010).

3.6.2. Maintenance of activated prior knowledge about faces is reflected by increased alpha/beta power

We found a positive single-trial correlation of AIS with alpha/beta power for all face prediction areas. This finding supports the assumption that the maintenance of activated prior knowledge as indexed by AIS is related to alpha and beta frequencies. Congruently with our findings, Mayer and colleagues (2015) recently showed that activation of prior knowledge about previously seen letters is associated with increased power in alpha frequencies in the prestimulus interval. Also, Sedley and colleagues (2016) observed that the update of predictions, which also requires access to maintained activated knowledge, is associated with increased power in beta frequencies.

Extending these previous findings, we are the first to report that single-trial low frequency activity strongly correlates with the momentary amount of activated prior knowledge in content specific brain areas. Specifically, our results demonstrate that the current amount of activated prior knowledge usable as predictions for face detection is associated with neural activity in the alpha and beta frequency range, supporting the hypothesis of a popular microcircuit theory of predictive coding (Bastos et al., 2012).

3.6.3. Face predictions are transferred in a top-down manner

In Face blocks we observed increased information transfer to FFA from aIT as well as from PPC, both areas located higher in the processing hierarchy than

FFA (e.g. Zhen et al., 2013; Michalareas et al., 2016). Thus, FFA seems to have the role of a convergence center to which information from higher cortical areas is transferred in order to prepare for rapid face detection.

Closely related to our findings Esterman and Yantis (2009) observed that anticipation effects for faces in FFA (and houses in PPA) were associated with increased activity in a posterior IPS region (part of the PPC) extending to the occipital junction. However, to our knowledge our study is the first to report face-related anticipatory top-down information transfer from PPC and aIT to FFA.

Top-down information transfer in face processing regions in a preparatory interval before face detection is in general supportive of the predictive coding account (Mumford, 1992; Rao et al., 1999; Friston, 2005, 2010), that suggests a top-down propagation of predictions. This top-down information transfer of predictions is probably associated with a low-frequency channel (Bastos et al., 2012) – in contrast to the bottom-up propagation of prediction errors, which has been linked to a high-frequency channel (Bastos et al., 2012; Brodski et al., 2015). The spectral dissociation between the transfer of predictions and of prediction errors frequencies is in line with physiological findings in monkeys and humans (Bastos et al., 2015; Michalareas et al., 2016) and received recent support from a MEG study investigating the (spectrally resolved) information transfer during the prediction of causal events (Pelt et al., 2016). Our spectrally resolved granger causality analysis did not contradict this view, yet results failed to reach statistical significance.

In addition to the two top-down links showing increased information transfer for Face blocks, we observed a bottom-up link from V1 to PPC with increased information transfer for House blocks. As we did not find a prediction network for houses and our analysis was thus only performed in the brain areas of the face prediction network, one can only speculate on the function of this bottom-up information transfer. It is possible that it indicates that house detection was rather performed in a bottom-up manner for instance by first identifying low level features that distinguish houses from their scrambled counterparts.

3.6.4. Pre-activation of prior knowledge about faces facilitates performance

Across subjects we found elevated predictable information in FFA in Face blocks in contrast to House blocks to be associated with shorter reaction times for Face blocks compared to House blocks. This suggests that especially pre-activation of prior knowledge about faces in FFA facilitates processing and speeds up face detection, as also suggested by FFA effects in previous fMRI studies (Esterman and Yantis, 2009; Puri et al., 2009). Our study is however the first to demonstrate that the size of the facilitatory effect on perceptual performance depends on the quantity of activated prior knowledge for faces in FFA, measurable as the difference in AIS between face and house block for each subject. Differential size of the faciliatory effect between subjects and the associated differences in the quantity of activated prior knowledge in FFA may be related to the differential ability in maintaining an object specific representation (see Ranganath et al., 2004).

4. Predictable information is reduced in ASD – a predictive coding study

Short title: Altered predictive coding mechanisms in ASD

Authors: Alla Brodski-Guerniero (1), Marcus J. Naumer (2), Vera Moliadze (3, 4), Jason Chan (2, 3, 5), Heike Althen (3), Fernando Ferreira-Santos (6), Sabine Schlitt (3), Janina Kitzerow (3), Magdalena Schütz (2, 3), Anne Langer (2, 3), Jochen Kaiser (2), Christine M. Freitag (3), Michael Wibral (1)

(1) MEG Unit, Brain Imaging Center, Goethe University, Frankfurt am Main, Germany
(2) Institute of Medical Psychology, Faculty of Medicine, Goethe University, Frankfurt am Main, Germany
(3) Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Autism Research and Intervention Center of Excellence, University Hospital Frankfurt, Goethe University, Frankfurt am Main, Germany
(4) Department of Medical Psychology and Medical Sociology, Schleswig-Holstein University Hospital (UKSH), Christian-Albrechts-University, Kiel, Germany
(5) School of Applied Psychology, University College Cork, Cork, Ireland
(6) Laboratory of Neuropsychophysiology, Faculty of Psychology and Education Sciences, University of Porto, Porto, Portugal

Key words*: Predictive coding theory, ASD, information theory, active information storage, prior knowledge, magnetoencephalography, MEG, default mode network*

4.1. Abstract

The neurophysiological underpinnings of the non-social symptoms of Autism Spectrum Disorder (ASD) which include sensory and perceptual atypicalities remain inconclusive. Well-known accounts of less dominant top-down influences and more dominant bottom-up processes compete to explain these characteristics. These accounts have been recently embedded in the popular frame work of predictive coding theory. In order to differentiate between competing accounts, we studied altered information dynamics in ASD by quantifying predictable information in neural signals. Predictable information in neural signals measures the amount of stored information that is used for the next time step of a neural process. Thus, predictable information limits the (prior) information which might be available for other brain areas, e.g. to build predictions for upcoming sensory information. We studied predictable information in neural signals based on resting state magnetoencephalography (MEG) recordings of 19 ASD patients and 19 neurotypical controls aged between 14 and 27 years. Using whole-brain beamformer source analysis we found reduced predictable information in ASD patients across the whole brain, but in particular in posterior regions of the default mode network. In these regions, predictable information was positively associated with source power in the alpha and beta frequency range. Predictable information in precuneus and cerebellum was negatively associated with non-social symptom severity, indicating a clinical relevance of the analysis of predictable information for research in ASD. Our findings are in line with the assumption that use or precision of prior knowledge is reduced in ASD patients.

4.2. Introduction

Autism Spectrum Disorder (ASD) is a developmental disorder with an estimated prevalence of about one in 68 children (Christensen, 2016). The disorder is characterized by deficits in social communication together with restricted, repetitive and stereotyped patterns of behaviors and interests as well as hypo- or hyperreactivity to sensory input (American Psychiatric Association, 2013).

Despite the first descriptions of the disorder by Kanner (1943) and Asperger (1944) dating back more than 70 years, the neurophysiological mechanisms underlying the symptoms of ASD have remained largely unknown. Historically, there have been attempts to elicit specific core underlying cognitive mechanisms, such as the “Theory of Mind” hypothesis (Baron-Cohen et al., 1985) assumed to underlie impaired social-cognitive function or executive function impairments (Russell, 1997) as well as “weak central coherence” (Happé and Frith, 2006) assumed to underlie stereotyped and repetitive behavior as well as sensory aspects. Examples of sensory and perceptual atypicalities in ASD are decreased susceptibility to visual illusions (Happé, 1996) as well as the superior performance in perceptual tasks requiring a focus on local features compared to global features (e.g. Shah and Frith, 1983; Plaisted et al., 1998; Joseph et al., 2009). Recent accounts of ASD further confirm these perceptual characteristics as a key element towards a comprehensive theory of ASD and propose to elucidate the non-social symptoms of the disorder within the framework of predictive coding theory (Pellicano and Burr, 2012a; Lawson et al., 2014). Predictive coding theory (Rao et al., 1999; Friston, 2005, 2010; Clark, 2012) suggests that perception is a process of hierarchical probabilistic inference, in which the brain uses prior knowledge from life-long experience for building internal predictions. These predictions are combined with incoming sensory information in order to infer the state of the outside world. A mismatch between top-down propagated predictions and sensory evidence results in a bottom-up propagated prediction error. Influence on perception of the prediction error depends on so-called precision-weighting (Friston, 2009; Friston and Kiebel, 2009) – i.e. the weight that is given to the prediction error compared to the prediction / prior knowledge. Predictive coding accounts of perception in ASD can be formalized as changes in information processing in terms of a reduced influence of prior knowledge (Pellicano and Burr, 2012a), a relative imbalance of prior knowledge and prediction error (Friston et al., 2013; Lawson et al., 2014) or a mere overweighing of prediction error/sensory input (Brock, 2012; Van de Cruys et al., 2014). In order to differentiate between these accounts, altered information dynamics in ASD may be assessed via the three fundamental component operations of information processing, i.e. *information storage*, *transfer* and *modification* (Langton, 1990; Lizier et al., 2012; Gómez et al., 2014; Wibral et al., 2015). In particular, quantifying *information storage* in neural signals may be a useful tool for testing the hypothesis of reduced use of prior knowledge in ASD (Gómez et al., 2014) as the use of prior knowledge for predictions requires that (passively) stored information is re-expressed in neural activity (active storage – see Zipser et al., (1993) for a distinction between passive and active storage). Information storage in neural processes is mirrored by the fact that information from the past of a neural process predicts a certain fraction of information in the future of this process (e.g. Gómez et al., 2014; Wibral et al., 2014). This predictable information provides the upper bound of the information potentially becoming useful as predictions for the brain.

In the present study we compared predictable information as measured by the information-theoretic measure active information storage (Lizier et al., 2012) for young patients diagnosed with ASD and neurotypical controls based on neural signals reconstructed from resting state magnetoencephalography (MEG) recordings.

We hypothesized that predictable information would be reduced in patients with ASD and that reduced predictable information would further be associated with severity in one or more of the symptom domains in ASD.

4.3. Methods

4.3.1. Participants

Nineteen male patients diagnosed with autism spectrum disorder (ASD) according to ICD-10 (World Health Organization, 1992), i.e. autism (F84.0), Asperger Syndrome (F84.5) or atypical autism (F84.1), and nineteen male, neurotypical controls (NTC) aged 14 – 27 years participated in the present study. Exclusion criteria for both groups were an IQ below 70, history of or current diagnosis of schizophrenia or bipolar disorder, current depressive episode, severe anxiety disorder, tic disorder, illegal drug use, and a chronic medical or neurological condition. All participants showed normal or corrected to normal vision. Neurotypical individuals had to score below the clinical cut-off of all first order scales of the Youth Self Report (YSR; (Achenbach and Edelbrock, 1991; Deutsche Child Behaviour Checklist, 1998a) or Young Adult Self Report (YASR 18-30; Achenbach, 1990; Deutsche Child Behaviour Checklist, 1998b). The ethics committee of the Medical Faculty of the University of Frankfurt approved the experimental study. Participants and/or their parents gave written informed consent before the experiment and received monetary compensation.

ASD patients were recruited through the Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Frankfurt, Goethe-University and via ASD related websites. NTC were recruited from local schools and by notices on the university campus.

4.3.2. Assessment instruments across groups

In- and exclusion criteria were assessed using checklists as well as a semi-standardised medical history interview. IQ was measured by the Culture Fair Intelligence Test (CFT 20-R; Weiß, 2006). The German version of the Youth Self Report (YSR) and the German version of the Young Adult Self Report (YASR 18-30) were implemented to describe severity of current psychopathology in both groups.

The socio-economic status (SES) of the respective family was computed based on the mean occupational status of both parents. The occupational status ranged from 1 to 5 (1 = unskilled worker; 5 = highly skilled, leading position). Handedness was assessed according to the Edinburgh Handedness Inventory scale (Oldfield, 1971), in which positive values indicate right-handedness and negative values indicate left-handedness.

4.3.3. Autism-specific assessment instruments

Patients were diagnosed according to ICD-10 criteria (World Health Organization, 1992), employing a semi-structured clinical interview, the German version of the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000; see Rühl et al., 2004 for the German version), and the Autism Diagnostic Interview–Revised (ADI-R; Rutter et al., 2003; see Bölte et al., 2006 for the

German version) administered by experienced clinicians (psychiatrists, clinical psychologists). The ADOS is a direct observation measure, assessing social communication and stereotyped, restricted behaviour within a social interaction situation. The ADI-R is a standardized interview for caregivers of autistic individuals and encompasses the three domains of “social interaction”, “communication”, and “restrictive, repetitive and stereotyped behaviours and interests”. The ADI-R could not be obtained in five of the ASD patients.

4.3.4. Data acquisition

For each participant, magnetoencephalography (MEG) resting state recordings were obtained for five minutes each with eyes open (fixating) and eyes closed, respectively. Only analysis of the data obtained from the resting state recordings with eyes closed will be reported in the present study.

The acquisition of the MEG data was performed in line with the guidelines for “good practice” of MEG recordings (Gross et al., 2012). A whole-head system (Omega 2005; VSM MedTech, Port Coquitlam, BC, Canada) with 275 axial gradiometers was used to record the MEG signals. Signals were recorded continuously at a sampling rate of 1200 Hz in a synthetic third-order gradiometer configuration and filtered online with fourth-order Butterworth filters with a 300 Hz low pass and a 0.1 Hz high pass (Data Acquisition Software Version 5.4.0, VSM MedTech, BC, Canada). During the complete recording participants’ head position relative to the gradiometer array was localized via three head localization coils that were placed on the nasion and 1 cm anterior of the tragus of each ear. In order to detect artifacts, the horizontal and vertical electrooculogram (EOG) and the electrocardiogram (ECG) were recorded via six electrodes. These were placed distal to the outer canthi of both eyes to record horizontal eye movements, above and below the right eye to record blinks and vertical eye movements and below both collarbones to record the ECG. The impedance of each electrode was kept below 15 kΩ, as measured with an electrode impedance meter (Astro-Med Electrode Impedance Meter, Model F-EZM5, Grass Technologies, Natus Neurology Inc., Warwick RI, USA).

Structural MR images were obtained with a 3 T Siemens Allegra or Trio scanner

(Siemens Medical Solutions) using a standard T1 sequence (3D MPRAGE sequence, 176 slices, 1 × 1 × 1 mm voxel size). Before acquisition of the structural images, vitamin E pills were placed at the former positions of the MEG head localization coils to enable co-registration of MEG data and structural MR images.

4.3.5. MEG data analysis

Preprocessing

MEG data analysis was performed with MATLAB (MATLAB 2012; The MathWorks) and the open source MATLAB toolbox FieldTrip (Oostenveld et al., 2011; Version 2013 11-11). During preprocessing, the continuous recordings of five minutes were split into data epochs of 1 s each. Line noise was removed using a discrete Fourier transform filter at 50,100 and 150 Hz. Further, FieldTrip artifact-rejection routines were used to automatically reject epochs containing muscle or sensor jump artifacts. For further cleaning of the data, independent component analysis (ICA; Makeig et al., 1996) was performed using the extended infomax (runica) algorithm implemented in fieldtrip/EEGLAB. ICs displaying a strong correlation with EOG and ECG channels were removed from the data. Additionally, data was visually inspected for residual artefacts.

In order to minimize movement related inaccuracies, the mean head position in the resting state datasets was calculated for each participant and only epochs in which the head position did not deviate more than 5 mm from the mean head position were considered for analysis.

Source grid creation

To perform MEG source analysis with individual head models, individual source grids were created by transforming the structural MR image for each participant to a T1 MNI template (http://www.fil.ion.ucl.ac.uk/spm). This way an individual transformation matrix was obtained for each participant. Next, the inverse of each participants’ transformation matrix was warped with a regular dipole grid (based on T1 template, spacing 15mm, resulting in 478 grid locations inside the brain), thereby obtaining a dipole grid for each participant in participant space. Using this approach every brain area was located at the same grid point for all participants allowing calculation of multi-participant statistics. A realistic single shell forward model (Nolte, 2003) was used to compute the lead-fields for each grid location.

Source time course reconstruction

In order to enable a whole brain analysis of active information storage (AIS), we reconstructed the source time courses for all 478 source grid locations inside the brain. Whole-brain source time course reconstruction was performed using a time-domain beamformer filter (linear constrained minimum variance; LCMV, Van Veen et al., 1997) applied on MEG sensor data filtered broadly with 8 Hz high pass and 150 Hz low pass. For each of the 478 source grid location three orthogonal filters in x, y, and z direction were computed and the sensor data were projected through the LCMV filters. From the resulting three time courses per location via singular value decomposition the time course in direction of the dominant dipole orientation was obtained and used for calculation of active information storage (AIS).

Analysis of active information storage

AIS describes how much of the information in the next time step of a process is predictable from its immediate past state (Lizier et al., 2012). High AIS values indicate that a signal is both rich in information and predictable at the same time. Detailed definition of AIS is given in Lizier et al., (2012; methods part) and Wibral et al., (2014; see also Gómez et al., 2014 and Brodski-Guerniero et al., 2017 for applications on MEG data).

In order to determine the history dimension and optimal embedding delay parameter for AIS computation, the Ragwitz’ criterion (Ragwitz and Kantz, 2002) as implemented in the TRENTOOL toolbox (Lindner et al., 2011) was used for each participant and each of the source locations separately. As differences in history dimension may induce a bias on the estimated values, we chose the history dimension of 6 over all participants and source locations for computation of AIS. This means that 6 samples were chosen, spaced at an embedding delay interval that was individually determined per subject based on that subject's signal autocorrelation decay time and the optimization via the Ragwitz criterion. AIS was computed with 4 nearest neighbours (recommended by Kraskov et al., 2004) in the joint embedding space using the Kraskov-Stoegbauer-Grassberger estimator (Kraskov et al., 2004; algorithm 1), as implemented in the open source Java Information Dynamics Toolkit (Lizier, 2014). Also, as a different number of data points may induce a bias in the estimation, AIS was computed on embedded data only up to the minimal number of data points over participants (number of data points entering the analysis: 149987; number of epochs: 127, 1 s length, sampling rate 1200 Hz). AIS values were averaged across time points for each source location and participant before statistical analysis.

Statistical analysis on AIS

For investigation of the mean difference in AIS between ASD patients and NTC for each participant, the AIS values were further averaged over all 478 source locations and a Wilcoxon rank sum test was performed to test for AIS differences between groups. In order to examine a potential correlation of AIS and age, Pearson’s r and Spearman’s rho were calculated for each group separately.

For finding the specific source locations at which AIS values differed between groups, an independent samples permutation t-test was performed across all potential source locations over the whole brain. To account for multiple comparisons across the 478 source locations, a cluster-based correction method (Maris and Oostenveld, 2007) was used. Clusters were defined as adjacent grid points whose t-values exceeded a critical threshold corresponding to an uncorrected alpha level of 0.05. For these clusters we defined cluster values as the sum of t-values in a particular cluster. Cluster values were tested against the distribution of cluster values obtained from 5000 permuted data sets. Significance was assessed based on an alpha value of 0.05. For the significant clusters, the brain areas showing (local) peaks in the t-map were reported. For these locations, we also calculated Spearman's and Pearson's correlations of AIS with age.

In order to assess the relationship of AIS values and ADI-R ratings of symptom severity (Bölte et al., 2006) linear regression analysis with AIS as response variable and the ADI-R algorithm scores as the predictor variable was calculated for all 478 brain locations. To account for multiple comparisons across brain locations, a cluster-based correction method was performed on the t-values of the beta coefficients (t = beta/standard error). Clusters significance was assessed in the same way as for the independent sample t-statistic described above (5000 permutations, alpha 0.05). Linear regression analysis was performed separately for each of the three ADI-R domains “communication” (ADI-R com), “social interactions” (ADI-R soc) and “restrictive, repetitive and stereotyped behaviours and interests” (ADI-R rit). Please note that ADI-R scores were available for 14 of the 19 ASD patients only, and thus this part of the analysis is based on a smaller sample size than the rest.

Correlation analysis of AIS and beamformer reconstructed source power

We further investigated the relationship of AIS and spectral power in individual epochs using Spearman's correlations. The frequency bands for correlation analysis were determined based on the averaged and normalized spectrum (by multiplication with frequency to account for the 1/f shape of the non-normalized spectrum) of the peak sources showing significant differences in AIS between groups. The spectrum was calculated from the reconstructed source time courses using a multitaper approach (Percival and Walden, 1993) with 2 Slepian tapers (Slepian, 1978) for a frequency interval from 8 to 150 Hz in 2 Hz steps. The averaged spectrum over sources, participants and epochs revealed three frequency bands: 8-14 Hz (alpha); 14-36 Hz (beta) and 36-150 Hz (gamma). Epoch-by-epoch power in these frequency bands was used for calculation of the Spearman's correlation with epoch-by-epoch AIS. For each participant, Spearman's rho was computed for correlation of the median epoch-by-epoch power in the predefined frequency bands with the epoch-by-epoch AIS. In order to test the significance of the correlation, for each participant the epochs were randomly permuted 5000 times and correlation was re-calculated for the permuted data sets. For each participant an original correlation value larger (or smaller) than 99.9998% (threshold Bonferroni adjusted for the 38*3*3 multiple comparisons) of the correlation values obtained for the permuted data sets was considered to be significant. At the second level, a binomial test was used to assess whether the number of participants showing significant correlations could be explained by chance. The median correlation values over participants and their significance based on the binomial test are reported.

Complexity analysis

As the average information contained in a signal (i.e. entropy; Shannon, 2001) defines the upper bound of AIS (Lizier et al., 2012; Wollstadt et al., 2016), we also quantified the differential entropy (e.g. Cover and Thomas, 2012) at the brain locations showing significantly decreased AIS for patients. Thereby we aimed to investigate whether decreased AIS in ASD was also associated with decreased (differential) entropy values. Differential entropy was calculated from the continuous signals using the Kozachenko-Leonenko estimator (Kozachenko and Leonenko, 1987) with 4 nearest neighbours as implemented in the Java Information Dynamic Toolkit (Lizier, 2014). To avoid a bias for the estimation of entropy based on a differential amount of epochs, the minimal number of 127 epochs was also used for entropy estimation of all participants. Wilcoxon rank sum test was used to access the differences in differential entropy between the ASD patients and the NTC group.

Statistical analysis using Bayesian Statistics

As non-significant effects were found for the difference in entropies between groups as well as for the correlation of AIS values and age, we additionally report Bayes factors (BF; Jeffreys, 1998) to clarify these findings. BFs allow direct quantification of the weight of evidence in favor of the null or the alternative hypothesis (Dienes, 2014) – a measure that cannot be obtained just by failing to reject the null-hypothesis in a frequentist approach. BFs were computed with the BayesFactor package (Morey et al., 2015 p.2) in R (R Core

Team, 2016). Default (medium width) prior settings for linear regression were used (Jeffrey-Zellner-Siow mixture of g-priors, Rouder and Morey, 2012, see also Liang et al., 2008, section 3), not favoring the null or alternative hypothesis in advance. For equal priors for the null and the alternative hypothesis a BF of for instance 3 indicates that the posterior odds are 3:1 in favor of the alternative hypothesis, i.e. that the alternative hypothesis is three times more probable than the null hypothesis given the data and the prior probabilities of both hypotheses. Three types of BF comparisons were performed in the present study:

1. For average AIS (over the whole brain) as a response variable we compared BFs for a linear regression with age as a predictor variable, with a linear regression with group (i.e. ASD or NTC) as a predictor variable.
2. For AIS in difference areas (mean over the brain areas showing a significant difference in AIS between the NTC and the ASD group) as a response variable we compared BFs for a linear regression with age and group as predictor variables with a linear regression including only group as a predictor variable. Note that here the factor group was included in the “null model” as the brain areas were pre-selected based on a group comparison in the independent samples t-test.
3. For differential entropy as the response variable we compared BFs for a linear regression with group as predictor variable with a linear regression “null model” including the intercept only.

4.4. Results

4.4.1. Group characteristics

Nineteen ASD patients and nineteen NTC participated in the experiment. The patients were diagnosed with either high functioning autism (n = 12), Asperger (n = 6) or atypical autism (n = 1). Eight of the participants in the ASD group received medication (2x Risperidon, 3x psychostimulant, 2x SSRI, 1x Risperidon, psychostimulant and SSRI). Main characteristics for both groups are summarized in Table 4.1.

NTC and ASD were well matched with regard to IQ (p = 0.861), handedness (p = 0.388), and socio-economic status (SES) (p = 0.097). The ASD group was younger than the NTC group (p = 0.041), so age was controlled for during analysis. As expected from the in- and exclusion criteria, several first and second order scales of the Y(A)SR (t-values) differed between groups.

Table 4.1. Summary of group characteristics

	ASD (mean ± SD)	*NTC (mean ± SD)*	*Statistics (Wilcoxon rank sum test)*
IQ	109.4 (±16.4)	109.6 (±18.6)	p = 0.861
Age	18.7 years (±3.4)	21.6 years (±3.8)	p = 0.041
Handedness	69.5 (±47.5)	65.5 (±51.6)	p = 0.388
SES	3.3 (±1.03)	3.9 (±0.79)	p = 0.097
Y(A)SR SR	61.6 (±11.9)	51.9 (±3.3)	p = 0.006
Y(A)SR KB	57.3 (±10.3)	52.6 (±4.3)	p = 0.169
Y(A)SR ADP	57.4 (±9.5)	51.5 (±2.5)	p = 0.036
Y(A)SR SP	59.7 (±13.6)	52.8 (±6.2)	p = 0.06
Y(A)SR SZ	60.3 (±10.9)	50.4 (±1.8)	p = 0.0005
Y(A)SR AP	59.6 (±10.7)	52.3 (±3.4)	p = 0.013
Y(A)SR AV	53.6 (±5.8)	51.7 (±3.5)	p = 0.153
Y(A)SR DV	53.3 (±4.7)	52.4 (±3.9)	p = 0.577
Y(A)SR INT	57.1 (±13.1)	47.2 (±6.5)	p = 0.019
Y(A)SR EXT	49.8 (±7.7)	45.7 (±8.5)	p = 0.156

SES = socio-economic status; SR = withdrawn; KB = somatic complaints; ADP = anxious/depressed; SP = social problems; SZ = thought problems; AP = attention problems; DV = delinquent behavior; AV = aggressive behavior; INT = internalizing; EXT = externalizing

4.4.2. Analysis of average predictable information

Comparison of average predictable information as measured by active information storage (AIS) between the ASD and NTC group revealed a significantly reduced mean AIS for the ASD group (Wilcoxon rank sum test, p = 0.031; median ± SD: ASD 2.02 ± 0.08; NTC 2.08 ± 0.06; Figure 4.1.)

In order to exclude that the group difference in average predictable information was related to age differences between participants, a correlation analysis of AIS and age was performed for each group. No significant correlation with age was found for any of the groups (ASD Spearman's correlation rho = 0.36, p = 0.135; Pearson correlation r = 0.19, p = 0.428; n = 19; NTC Spearman's correlation rho = 0.06, p = 0.815; Pearson correlation r = 0.15, p = 0.530, n =

19). To further clarify these non-significant results, we also calculated the ratio of Bayes Factors for a linear regression with average AIS as a response variable and group (i.e. ASD or NTC) as a predictor variable and a linear regression with the same response variable and age as a predictor variable. The resulting Bayes Factor ratio of 3.15 indicated that the observed average AIS values were more than three times more likely to occur when group was considered the predictor than when age was considered the predictor. In other words, the calculated Bayes Factor ratio indicated that group was a better predictor on average AIS than age. Thus, it is unlikely that the observed group differences on average AIS were based on age differences only.

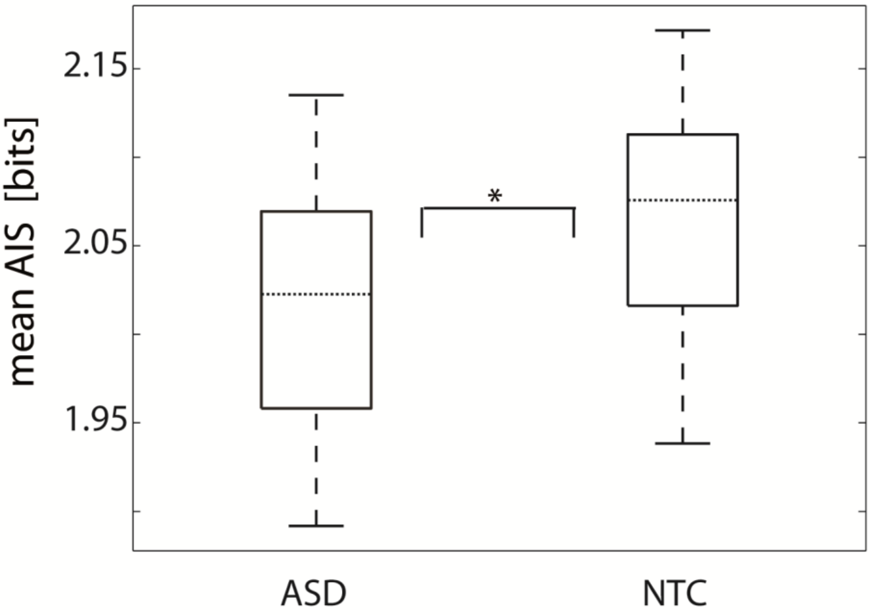

Figure 4.1. Comparison of average AIS between groups. Each boxplot shows the distribution of averaged AIS values for all participants within the ASD or NTC group, respectively (n ASD = 19, n NTC = 19). These values have been obtained by averaging AIS for all 478 sources inside the brain for each subject. Horizontal dotted lines mark the median of each group. The asterisk indicates a significant difference in average AIS between groups (Wilcoxon rank sum test p = 0.03).

4.4.3. Whole brain analysis of predictable information

Spatially resolved comparison of AIS between the ASD and NTC group at the MEG source level revealed a significant group difference in posterior cingulate cortex (PCC), supramarginal gyrus (SMG) and precuneus (Prec) (Figure 4.2.A).

At these areas AIS was significantly reduced for ASD compared to NTC (Figure 4.2.B).

In order to exclude age-related effects, again correlations with age as well as Bayes Factor comparisons were performed. No significant correlation was found for AIS (averaged over the three sources) and age for any of the groups (ASD Spearman's correlation rho = -0.04, p = 0.873; Pearson correlation r = 0.312, p = 0.19; n = 19; NTC Spearman correlation rho = 0.16, p = 0.499; Pearson correlation r = 0.02, p = 0.927, n = 19). Further, Bayes Factors were computed based on a linear regression with AIS as response variable and group as a predictor variable, and a linear regression with the same response variable and group as well as age as predictors. Please note that here the factor group was included in both models, as the brain areas had been pre-selected based on the group comparison. The resulting Bayes Factor ratio of 3.25 indicated that the observed AIS values were more than three times more likely to occur with group as a predictor than with group as well as age as predictors. This means that a model with group as the only predictor predicted the AIS values better than when age was included as an additional predictor. Thus, similar to the global AIS effect, group differences in AIS at the identified specific sources were not likely to be driven by differential age.

4.4.4. Complexity analysis

In order to study whether lower AIS in ASD might be associated with decreased signal complexity, we also computed a measure of entropy in PCC, SMG and Prec. None of these areas showed significant differences in entropy between groups (Wilcoxon rank sum test, PCC p = 0.599, SMG p = 0.350, Prec p = 0.884, Figure 4.3.). Additionally, a Bayes factor of 0.34 for a linear regression with entropy (averaged over the three brain areas) as response variable and group as predictor variable indicated that the present entropy values were 2.94 (1/0.34) times more likely to be observed in case of the validity of the null hypothesis, i.e. when group is not a proper predictor for entropy. This suggests that there was no difference in entropy between groups and thus differences in AIS for these source locations were not likely to be based on differences in signal complexity in these areas.

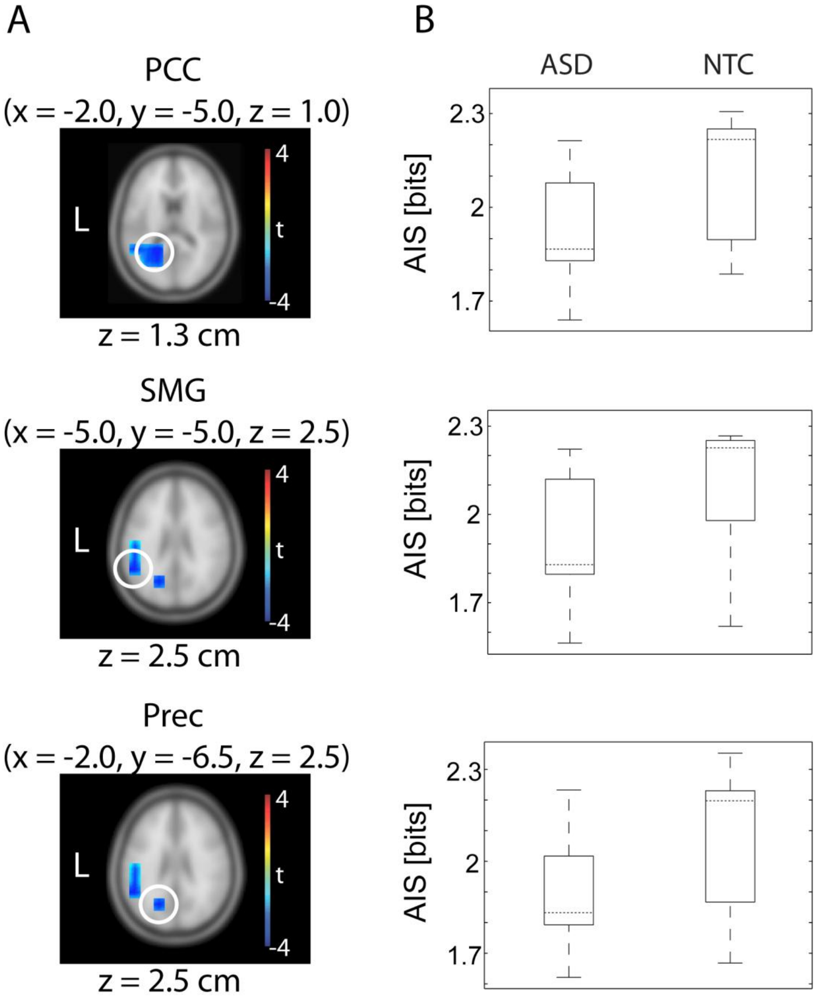

Figure 4.2. Statistical comparison of AIS between groups at the MEG source level. ***Left:*** Results of whole-brain independent samples permutation t-metric contrasting the ASD and NTC group (n ASD = 19, n NTC = 19, t-values masked by $p < 0.05$, cluster correction). Peak brain locations are highlighted with white circles. For each brain location MNI coordinates are shown at the top. An exemplary brain slice is shown for each brain location; z-values are displayed below each brain slice. ***Right:*** Illustration of the distribution of AIS values for each brain location and for the ASD and NTC group, respectively. Dotted horizontal lines mark the median of each group. PCC = posterior cingulate cortex; SMG = supramarginal gyrus; Prec = precuneus.

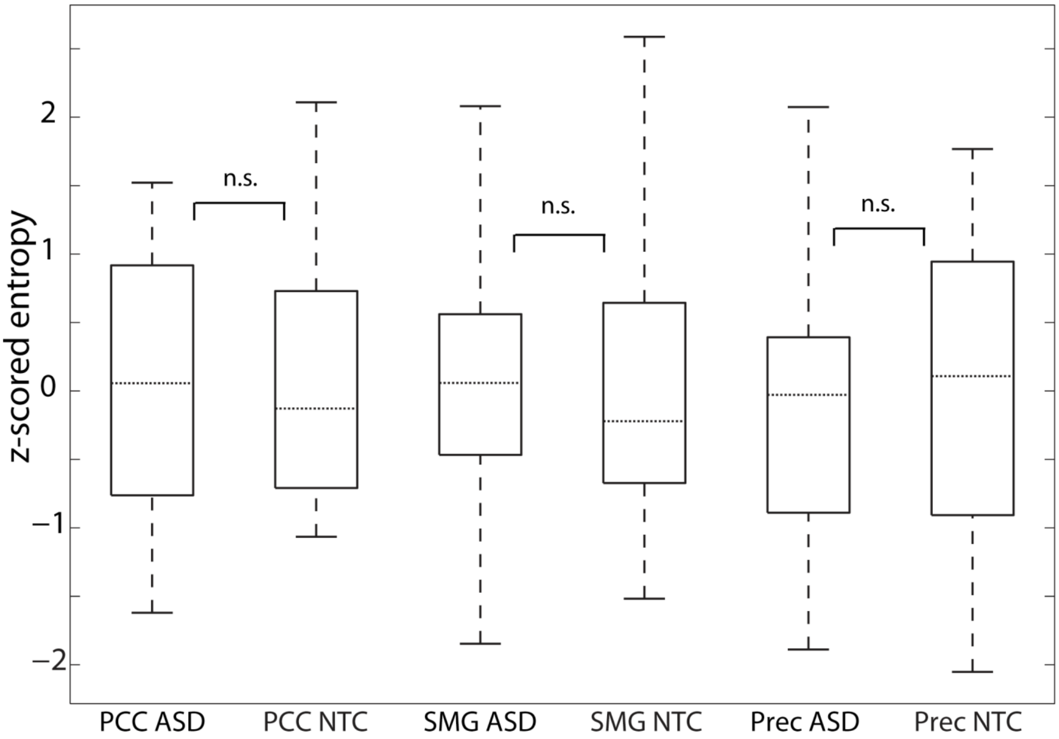

Figure 4.3. Post-hoc complexity analysis. Boxplots illustrate the distribution of entropy values across participants. Entropies are displayed as z-scores across all estimates for both groups. Dotted horizontal lines mark the median of each group. n.s. = not significant based on Wilcoxon rank sum test. PCC = posterior cingulate cortex; SMG = supramarginal gyrus; Prec = Precuneus.

4.4.5. Correlation of predictable information and power in individual epochs

In order to study the relation of AIS and spectral power over individual epochs in PCC, SMG and Prec, we correlated single epoch AIS values with the power in different frequency bands during the same epochs (Table 4.2.). Correlation analysis revealed a strong positive correlation in the alpha band and a moderate positive correlation in the beta band.

Table 4.2. Correlation of epoch-by-epoch AIS values and power in PCC, SMG and Prec

	PCC	*SMG*	*Prec*
8-14 Hz (alpha)	rho = 0.69*	rho = 0.71*	rho = 0.74*
14-36 Hz (beta)	rho = 0.35*	rho = 0.35*	rho = 0.40*
36-150 Hz (gamma)	rho = -0.13	rho = -0.13	rho = -0.12

* = significant correlation, based on binomial test

4.4.6. Correlation of predictable information and ADI-R scores

In order to further assess whether predictable information relates to autistic traits, we performed a whole brain linear regression analysis of AIS and the three ADI-R algorithm scores for the ASD group (n = 14 as ADI-R scores were not available for all of the patients). Regression analysis was performed separately for the three domains: communication (ADI-R com), social interactions (ADI-R soc) and restrictive, repetitive and stereotyped behaviours and interests (ADI-R rit). There was a significant cluster for the regression of AIS and ADI-R rit (Figure 4.4.). This cluster encompassed four peak areas, including sources in the cerebellum and the precuneus. The source in precuneus was located slightly more anterior compared to the precuneus area found in the group comparison of AIS (see MNI coordinates in Table 4.3.). All of the brain areas in the significant cluster showed negative t-values, indicating a negative relationship of AIS and ADI-R rit. In other words, lower AIS in these areas was associated with higher ADI-R rit scores, i.e. a higher degree of impairment in this domain. No significant clusters were found for the regression of AIS and ADI-R com or ADI-R soc.

Table 4.3. Peak voxels for the significant cluster in the regression analysis of AIS and ADI-R rit (see Figure 4.4.)

No	*MNI Coordinates*	*Label*
1	x = -5, y = -65, z = -5	Left Lingual gyrus / Cerebellum (culmen)
2	x = -20, y = -50, z = 25	Left Precuneus
3	x = -20, y = -50, z = -35	Left Cerebellum
4	x = -5, y = -65, z = -50	Left Cerebellum (Lobule VIII)

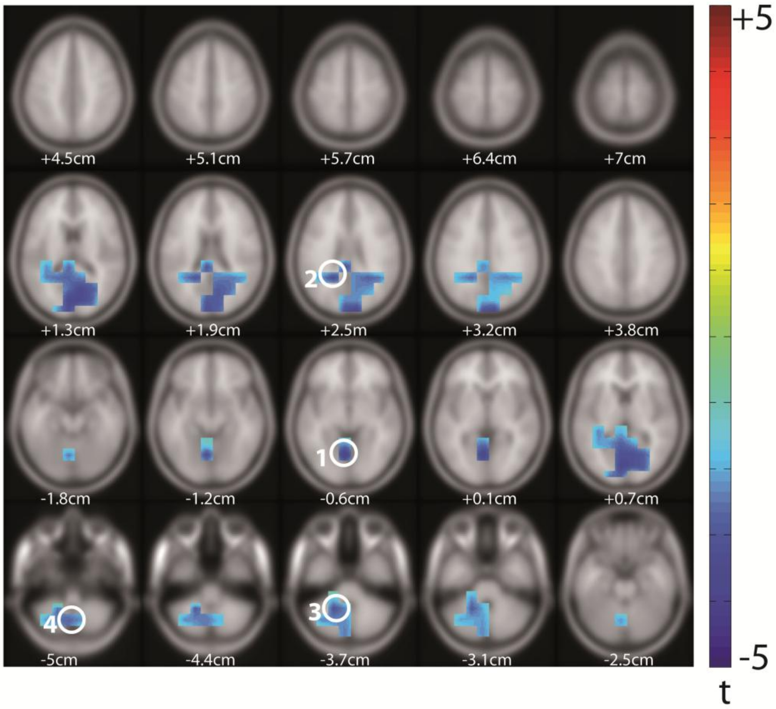

Figure 4.4. Linear regression analysis of AIS and ADI-R scores. Results of linear (permutation) regression with ADI-R rit scores as predictor variable and AIS as response variable (ASD only, n = 14, t-values masked by $p < 0.05$, cluster correction). Z-values are shown below each brain slice. Peak voxels are highlighted with white circles and numbers; corresponding MNI coordinates and labels are given in Table 4.3.

4.5. Discussion

In line with the theory of impaired predictive coding mechanisms in autism spectrum disorder (ASD) we tested the hypothesis that predictable information in neural signals is reduced in ASD patients. In line with this hypothesis we found average predictable information to be reduced in individuals with ASD compared to neurotypical controls (NTC) during resting state magnetoencephaography (MEG) recordings. In addition to the reduction of average predictable information in ASD, we found specific reductions of predictable information in posterior cingulate cortex (PCC), supramarginal gyrus (SMG) and precuneus (Prec). Here, predictable information was strongly associated with alpha band activity (8-14 Hz) and moderately with beta band activity (14-36 Hz). Importantly, none of the group differences in predictable information could be accounted for by differences in age between the participants or by differences in signal complexity between groups. In addition, ASD patients showed a negative relationship between predictable information and symptom severity in the domain of restricted, repetitive, and stereotyped behaviors and interests (ADI-R rit) in several brain areas including the cerebellum, suggesting a potential clinical relevance of predictable information in ASD research.

In the following we will relate our results to previous findings and predictive coding accounts of perception in ASD.

4.5.1. Average predictable information is reduced in ASD

Our finding of reduced average predictable information in ASD is in line with a recent report by Gómez et al. (2014). They studied predictable information using the AIS measure (Lizier et al., 2012) in ASD patients and NTC in the pre-stimulus interval of a face detection task. In all but one of the twelve studied brain regions they found at least a tendency towards a reduction of AIS in ASD. The present study confirms this general finding of reduced predictable information in ASD – while overcoming several shortcomings of the predecessor study: First, the present larger sample of 38 (19 patients, 19 controls) improved statistical power in contrast to the small sample size of 22 (10 patients, 12 controls) in the previous study (Gómez et al., 2014). Second, in contrast to the region-of-interest approach by Gómez and colleagues, applying a whole-brain approach allowed to study the (average) differences in AIS between groups based on a large amount of sources (~500 inside the brain) covering all potential brain areas. Thus, the present study demonstrated that *overall* AIS is reduced in patients with ASD.

4.5.2. Predictable information at PCC, SMG and Prec is reduced in ASD

Whole-brain analysis of AIS additionally enabled to determine the brain areas at which predictable information was particularly reduced for ASD patients. This was the case for PCC, SMG and Prec. These three brain areas belong to the default mode network (DMN; Mason et al., 2007; Buckner et al., 2008; Raichle, 2015), which is known to be engaged during passive (internally focused) tasks or epochs - corresponding to the resting state design of the current study.

Atypicalities in the DMN for ASD patients have been reported as reduced activation of the DMN nodes (Kennedy et al., 2006) or as altered, mainly diminished connectivity between the nodes (e.g. Cherkassky et al., 2006; Weng et al., 2010; Washington et al., 2014). However, hyperconnectivity of posterior nodes has also been reported, namely to medial and anterior temporal lobe regions (Monk et al., 2009; Lynch et al., 2013). Further, ASD atypicalties in the DMN may also be related to anatomical differences such as a relative increase of gray matter volume in several brain areas including the PCC in ASD patients (Waiter et al., 2004). Interestingly, internal thoughts which mainly involve activity in the DMN (e.g. Buckner et al., 2008) are also reported to differ considerably between NTC and ASD patients (Hurlburt et al., 1994). Extending previous findings of ASD-related atypicalities in the DMN, our results show that for ASD patients also the amount of predictable information is particularly reduced in posterior nodes of the DMN during resting state periods. It should be noted that reductions of predictable information in ASD may appear in brain regions other than the DMN, when the task requires specific predictions (see Gómez et al., 2014).

4.5.3. Complexity is not reduced in PCC, SMG and Prec

Noteworthy, decreased predictable information within the DMN as assessed by AIS was not associated with decreased signal complexity in these areas. This is of importance as the AIS measure quantifies both the complexity and predictability of neural processes. High AIS values are observed for predictable signals, rich cortical dynamics or a combination of both. In particular, complexity, which can be measured as signal entropy (i.e. average information content), defines the upper limit for AIS. Thus, in principle, reduced AIS in ASD patients could also have resulted from a reduced signal complexity. Indeed, reduced EEG signal complexity has been observed during a visual matching task for ASD patients (Catarino et al., 2011) as well as during resting state recordings for children with a high risk of developing ASD (Bosl et al., 2011). However, these previous findings were not replicated in our study using differential entropy as a measure of complexity. Our Bayesian analysis favored the hypothesis that there is no difference in entropy and thus complexity between ASD patients and NTC for the three DMN areas. Note that at the descriptive level even the sign of the marginal differences in entropy between groups was not the same over brain areas. These findings suggest that neural signals in the DMN for patients and controls were equally rich in cortical dynamics; however, they were structured in a less predictable manner for the ASD group. In the following, we relate our findings to predictive coding accounts of ASD.

4.5.4. Reduced predictable information in the light of predictive coding accounts of ASD

Recently it has been argued that the DMN may play a key role in predictive coding by acting as a top level of the predictive hierarchy, being responsible for initiating predictions that cascade down to categorize sensory input and drive motor activity (Barrett and Satpute, 2013; Barrett, 2017). This high impact in predictive processing is in line with the key functions associated with the DMN like episodic memory retrieval (posterior parietal and PCC regions), future planning, self-referential thoughts (dorsomedial pre-frontal cortex; PFC), and integrating sensory and interoceptive signals (ventromedial PFC) (e.g. Buckner et al., 2008; Whitfield-Gabrieli and Ford, 2012; Raichle, 2015). In particular, the posterior regions of the DMN like the Prec and PCC have been linked to memory-related processing, i.e. retrieving information from memory and anticipating the future (Wagner et al., 2005; Bar, 2007; Buckner et al., 2008). Bar (2007) linked the DMN even more directly to the continuous generation of (memory-based) predictions in the brain. Extending this line of thought, Fiser et al. (2010) have suggested that spontaneous brain activity (as for instance within the DMN during resting state recordings) reflects the historically informed prior beliefs about the world (see also Sadaghiani et al., 2010). Based on these suggestions our findings of decreased predictable information in ASD within the DMN seamlessly fit into the account of reduced use of prior knowledge or “weaker” prior beliefs in ASD (Pellicano and Burr, 2012a). As the neural signals in the DMN for ASD patients were structured in a less predictable manner during rest, we might speculate that this reflects the impairment of ASD patients to represent the high-level regularities of the environment in their spontaneous activity.

The fact that this effect was found at posterior nodes of the DMN could indicate that there is a deficit in retrieving information from memory (Raichle, 2015) in order to generate appropriate predictions. One hypothetical mechanism for this deficit could be related to the previously reported functional hyperconnectivity between the PCC and medial and anterior temporal lobe regions (Lynch et al., 2013): if too much or non-specific information is retrieved more or less at random during rest, it may impair the ability to generate stable predictions that would show as AIS.

Aside from this speculation, our data strongly suggest that the brain areas receiving the information from the DMN will need to deal with information that is less predictable. This may result in difficulties to learn (or change) new predictive models, leading to even more unreliable representations of the world. Reduced reliability or precision of prior knowledge for forming of top-down propagated predictions may further result in an imbalance of bottom-up and top-down influences in ASD (Friston et al., 2013; Lawson et al., 2014). A relative increase of the influence of bottom-up propagated prediction error may lead to the feeling of being overwhelmed by sensory information (Grandin, 1992), as weak or imprecise predictions will be less efficient in explaining away sensory inputs, thus leaving more feed-forward sensory information to be processed by relatively limited central resources; the relative imbalance also allows to explain the apparent paradox that ASD patients often perceive self-produced sounds not as equally unpleasant as external, unexpected sounds (Kanner, 1943).

While our results are fully compatible with predictive coding accounts of perception in ASD, which highlight the reduced use or (relative) precision of priors (Pellicano and Burr, 2012a; Friston et al., 2013; Lawson et al., 2014), they are not easily explained by accounts of merely increased bottom-up precision in ASD (Brock, 2012; Van de Cruys et al., 2014). This is because increased bottom-up precision should not be associated with decreased predictable information.

Furthermore, our finding of the strong association of predictable information in resting state recordings and neural activity in low frequencies adds support to the hypothesis that low frequencies are the carrier of top-down propagated information within the predictive coding framework (Bastos et al., 2012). Importantly, despite its correlation with low frequencies, predictable information also yields additional insights not immediately accessible by traditional spectral analysis (Gómez et al., 2014; Wollstadt et al., 2016; Brodski-Guerniero et al., 2017).

4.5.5. Predictable information in ASD is associated with symptom severity in the domain of restricted, repetitive, and stereotyped behaviors and interests (ADI-R rit)

Studying predictable information in ASD may not only help to distinguish between competing theoretical accounts but may also have a potential clinical relevance. This is supported by the finding of a significant negative relationship between predictable information and the symptom severity in the domain of restricted, repetitive, and stereotyped behaviors and interests (ADI-R rit). The ADI-R rit domain captures mainly stereotypical motor behaviors like hand flapping as well as the insistence on sameness and routines. Thus, in contrast to deficits in social interactions (ADI-R soc) and communication (ADI-R com), the ADI-R rit domain of ASD symptoms is presumably most closely related to perceptual atypicalities and also predictive coding mechanisms in ASD. In fact, the behavioral abnormalities captured in this domain might represent techniques to control the exaggerated prediction error resulting from the imbalance of top-down and bottom-up information flow (Friston et al., 2013; Lawson et al., 2014) and reduce the anxiety associated with the inability to predict upcoming events (Sinha, 2002).

The significant cluster for the regression of ADI-R rit scores and AIS included prominent peaks in the cerebellum – a brain area in which anatomical abnormalities (e.g. decreased number of Purkinje cells) have been observed most consistently in ASD (e.g. Brambilla et al., 2003; see also Fatemi et al., 2012 for a review). Even more closely related to our findings, a significant negative correlation between rates of repetitive behavior and area measures of cerebellar vermis lobules VI – VII has been previously reported (Pierce and Courchesne, 2001). Based on this finding we might speculate that anatomical abnormalities in the cerebellum would also show a correlation with the AIS measure. However, this remains to be tested in future investigations.

4.5.6. Conclusion

Resting state neural activity in patients with ASD shows less predictable patterns compared to controls. This is particularly the case for posterior regions of the default mode network. Further, in cerebellum and precuneus signal predictability is negatively associated with symptom severity in the domain of restricted and repetitive behaviors.

4.5.7. Acknowledgements

ABG received support by Ernst Ludwig Ehrlich Studienwerk (BMBF scholarship for graduate students). MW acknowledges support by SP3 of the Human Brain Project (EU Grant 604102).

5. General Discussion

In the following part, I am going to review the findings of the present work with respect to previous shortcomings in predictive coding research. As specified in the general introduction, these are the largely unexplored neural correlates of the use of prior knowledge in predictive coding (5.1.); the limited neurophysiological evidence for the neural implementation of predictive coding in the human brain (5.2.) and the lack of assumption-free approaches to study predictive coding mechanisms (5.3.). In the last paragraph of this chapter I will discuss potential limitations of the present work and give an outlook for future research (5.4.).

5.1. The neural correlates of the use of prior knowledge in predictive coding

In the present work, I introduced three studies exploring the neural correlates of the use of prior knowledge in the predictive coding framework. Each of the studies was designed to investigate key aspects of the way prior knowledge is represented, processed or applied to predict upcoming information in the human brain. All studies were based on magnetoencephalographic (MEG) recordings of human participants.

In study 1 (chapter 2) a Mooney (Mooney, 1957) face detection task was used to induce prediction errors by the mismatch between sensory input and predictions. To this end I violated two predictions based on prior knowledge about faces from life-long experience: 1. Upright face orientation and 2. Illumination from the top. Violation of the predicted face orientation and illumination direction resulted in decreased accuracy and increased reaction times for face detection, confirming a successful induction of prediction errors in the experimental design. At the MEG source level the mismatch of sensory input and predictions based on life-long experience resulted in an early prediction error for the unexpected orientation and later prediction error for the unexpected illumination direction. The prediction error for the unexpected orientation was observed at visual brain areas. The prediction error for the unexpected illumination direction was observed at brain areas involved in spatial working memory, reconstruction of 3-D shape from shading and error detection. Both prediction errors were reflected by increased high-frequency gamma band activity (> 68 Hz). Furthermore, for both prediction errors, high-frequency gamma band activity was positively correlated with participants' reaction times during face detection. This positive correlation suggests that the prediction errors for unexpected orientation and illumination direction reflected in high-frequency gamma band activity slowed down the processing. In addition to the two prediction errors, I observed increased high-frequency gamma band activity at mid-latency for the expected illumination direction at brain areas processing attention to internal representations. Last, a late interaction effect for violation of both expectations in high-frequency gamma band activity was located to visual brain areas, potentially representing a high-level prediction error.

In study 2 (chapter 3) a Mooney face/house detection task with a block design was used to induce the (pre-)activation of relevant prior knowledge for face predictions. The amount of activated prior knowledge for face predictions was quantified as predictable information at the MEG source level with the information-theoretic measure active information storage (AIS; Lizier et al., 2012). By application of AIS analysis to whole-brain source time courses in the pre-stimulus interval, I found that pre-activated prior knowledge for faces showed as increased predictable information in content-specific brain areas. These content-specific areas included the fusiform face area (FFA), occipital face area, posterior parietal cortex, and anterior inferior temporal lobe. In particular in brain area FFA the increase in predictable information was negatively correlated with reaction times, suggesting a behavioral relevance of the amount of pre-activated prior knowledge at the main face processing region. In addition, the increase in predictable information in all content-specific areas was positively associated with alpha and beta band activity (< 34 Hz) on a trial-by-trial basis. Further, the trial-by-trial AIS values in content-specific brain areas allowed differentiating trials with face predictions from trials with house predictions by a classifier approach. Last, application of the information-theoretic measure transfer entropy (TE; Schreiber, 2000; Vicente et al., 2011) to the source time courses revealed top-down information transfer from posterior parietal cortex and anterior inferior temporal lobe to FFA in the pre-stimulus interval. This information transfer suggests that face predictions based on pre-activated prior knowledge were transferred to FFA in a top-down manner in order to prepare for face detection.

In study 3 (chapter 4) a resting state design was used to compare the amount of activated prior knowledge in the brain of patients diagnosed with autism spectrum disorder (ASD) and healthy controls. As in study 2, the amount of activated prior knowledge was quantified as predictable information by the AIS method. Using whole-brain AIS analysis at the MEG source level, I found that average predictable information was reduced for ASD patients. Moreover, for the ASD patients predictable information was particularly reduced at the posterior nodes of the default mode network i.e. the posterior cingulate cortex, supramarginal gyrus and precuneus. Similar to the results of study 2, predictable information at these areas was positively associated with neural activity in the alpha and beta frequency bands (< 36 Hz) on a trial-by-trial basis. In addition, for the ASD patients the amount of reduction in predictable information at the precuneus and cerebellum was correlated with symptom severity in the domain of restricted and repetitive behaviors, indicating a potential role of predictable information as a biomarker in ASD research.

5.2. Neurophysiological evidence for the neural implementation of predictive coding

In the currently most influential proposal for a neural implementation of predictive coding in the human brain, Bastos and colleagues (2012) suggested that prediction errors should be linked to fast neural activity above 30 Hz, i.e. the gamma frequency band. I was able to provide neurophysiological evidence for this hypothesis in study 1. In this study, I found that prediction errors induced by the violation of expectations from life-long experience were indeed linked with fast neural activity. The findings indicate that specifically the high-frequency gamma band activity above 68 Hz is associated with prediction errors. The general association of prediction errors and gamma band activity is in line with a bottom-up propagation of prediction errors from error units in superficial cortical layers – when the predominance of gamma band activity in these superficial layers (Buffalo et al., 2011) is considered. However, as described in the previous section, in addition to the high-frequency gamma band activity reflecting prediction errors I found an effect in the exact same frequency band which most likely did not represent a prediction error. This high-frequency gamma power increase occurred for the expected and not unexpected illumination and was not associated with changes in reaction times. As I found sources for this effect in brain areas known to be involved in processing of attention to memory representations, I suggest that high-frequency gamma band activity can also carry (this type of) attentional effects. In the predictive coding context, attentional effects are often conceptualized as regulation of the precision-weighting of prediction errors vs. predictions (Feldmann and Friston, 2009, Friston, 2010). Recently, this precision-weighting mechanism has also been linked to oscillations in alpha frequencies (Sedley et al., 2016). Contrary to these findings, the present results show that attentional effects can also be associated with gamma band activity – a relationship which is well documented in earlier studies (Herrmann et al., 2004; Uhlhaas et al., 2011).

Importantly, gamma band activity is linked to feed-forward propagation in the cortex (Bastos et al., 2015; Michalareas et al., 2016). Thus, the attentional effect in the gamma band conflicts with conventional views of predictive coding in which feed-forward propagation is limited to prediction errors (Clark, 2012). However, the following alternative explanation can also be considered: Gamma band activity dominates in the superficial layers of the cortex but is not limited to them (see Bastos et al. 2012, Figure 1C; Xing et al., 2012, Figure 3). Hence, the gamma band activity which I observed for the attentional effect might also represent a feedback signal originating from deeper cortical layers. Future research combining MEG or electroencephalography (EEG) with high resolution functional magnetic resonance imaging (fMRI), in which the different cortical layers can be resolved (e.g. Goense et al., 2012) may further clarify this interpretation.

In addition to the association of high-frequency neural activity and prediction errors, Bastos and colleagues (2012) also proposed an association of lower frequency activity with predictions – presumably in the beta frequency range (~12 – 30 Hz in the literature). This is line with other reports, in which the beta frequency range has been in the focus for signaling of top-down predictions (Bressler and Richter, 2015) or the status quo in general (Engel and Fries, 2010).

In the present work, I found a strong association of beta (~14 – 34 Hz) as well as alpha band activity (8 – 14 Hz) with the amount of activated prior knowledge in study 2 and 3. In study 2, this activated prior knowledge was closely linked to face predictions. In line with a potential link of predictions and neural activity in the alpha frequency band, Mayer and colleagues (2015) recently demonstrated that pre-stimulus alpha band activity can implement the prediction content for different letters. Similarly, Bauer and colleagues (2015) showed that pre-stimulus alpha power is associated with the predictability of an upcoming target. These findings as well as the present results suggest that the power in the alpha frequency band is a likely candidate to implement predictions in the human brain – in addition to the power in the beta frequency band.

In sum, the present results support Bastos' theory of a representation of prediction errors in high and predictions in lower frequency activity – however with some modifications to the original proposal (Figure 5.1.).

In addition to the modifications based on the present results, Bastos' model may also be complemented by two types of feedback influences: The first type of influence which is missing in Bastos' model is a potentially crucial modulatory influence of the prediction units of a higher cortical area on the prediction units of a lower cortical area – the back-propagation activated calcium signaling (BAC firing; Larkum, 2013). The BAC firing mechanism is based on the following physiological background: Feedback connections from layer 5/6 (higher area) mainly target the apical tuft dendrites of layer 5 pyramidal neurons in layer 1 (lower area). Near the apical tuft, calcium dependent action potentials can be triggered. These can cause a sustained depolarization and thereby a high-frequency bursting of axonal action potentials. However, the threshold for a calcium dependent action potential is only reached when the synaptic input to the apical dendrites coincided with a back-propagating spike from the cell soma, i.e. when feedback input to the apical dendrites coincides with feed-forward input to the soma. Thus, strong bursting of layer 5 pyramidal cells can only occur if feedback and feed-forward input can be integrated.

The second feedback pathway which is missing in Bastos' proposal is the "extra-descending pathway" (Mumford, 1992). It includes superficial cells of a higher cortical areas projecting to deep cortical layers of the lower cortical area. This extra-descending pathway allows prediction errors of the higher cortical level to influence the predictions of the lower cortical area – maybe in order to enable the lower area to reinterpret the data in case of substantial error signals.

I added both missing feedback influences with dashed lines to the circuit in Figure 5.1.

5.3. An assumption-free approach to study predictive coding algorithms

In the general introduction, I suggested that quantification of predictable information with the information-theoretic measure AIS may provide a largely assumption-free approach to study the neural correlates of prior knowledge in predictive coding. This assumption was confirmed in study 2 in which I found increased predictable information in face processing areas when faces were predicted. Importantly, calculation of predictable information with AIS does not require a-priori assumptions about the brain areas involved in making specific predictions as it allows finding these brain areas exclusively by the properties of their neural signals. This information-theoretic approach can be applied to a variety of designs in predictive coding research and also any type of neurophysiological recordings, at the circuit, layer or even single cell level.

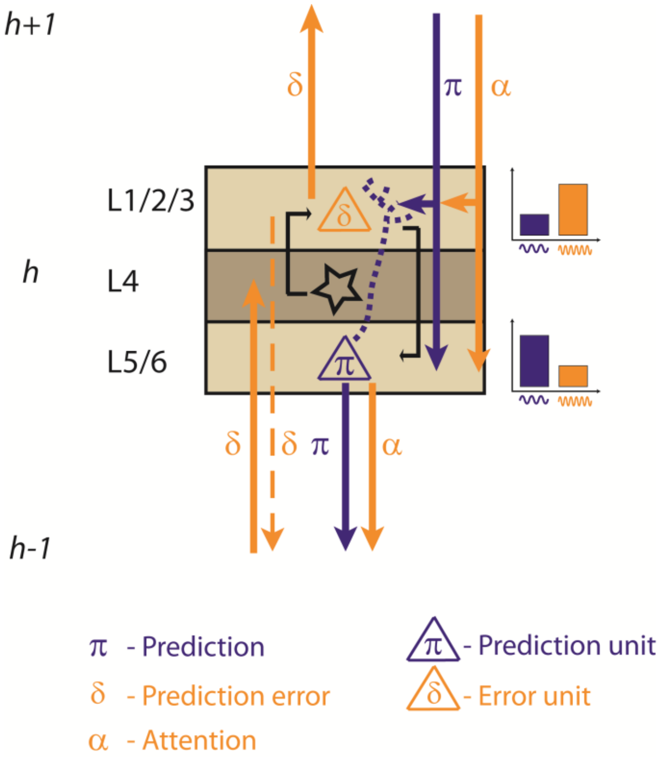

Figure 5.1. Modified version of Bastos' proposal (2012) for a neural implementation of predictive coding (schematic). Error units in the superficial cortical layers, in which high frequencies dominate, send prediction error signals in gamma frequencies to areas higher up in the cortical hierarchy (h-1 to h and h to h+1). Prediction units in the deep cortical layers, in which low frequencies dominate, send prediction signals in alpha and beta frequencies to areas lower in the cortical hierarchy (h+1 to h and h to h-1). Additionally, cells in the deep cortical layers send attentional signals in gamma frequencies to areas lower in the cortical hierarchy (h+1 to h and h to h-1). Orange indicates high frequencies, blue indicates low frequencies. The dashed orange line indicates the extra-descending pathway; the dashed blue line indicates the dendritic tree of a layer five neuron for illustration of the BAC firing pathway.

Moreover, this approach does not only allow to find the brain areas involved in predictive processing, but also to quantify the *amount* of activated information (prior knowledge) at the analyzed brain areas – a quantity which is not directly available from neural activity as recorded with fMRI, EEG or MEG. The quantitative nature of predictable information as measured by AIS allows straightforward correlation analysis with behavioral parameters or parameters obtained from standardized questionnaires as well as comparisons between groups of participants. Such a group comparison of predictable information has for instance been applied in study 3 to compare a group of ASD patients and a group of neurotypical controls. In this study, I found that predictable information was reduced for ASD patients, in particular in the posterior nodes of the default mode network. The information-theoretic calculation of predictable information by AIS allowed an interpretation of these results in terms of an impairment of predictive coding mechanisms in ASD disease: Independent of the nature of the underlying signal, high AIS values are observed when a signal is complex and has a predictable structure. As complexity was not differing between ASD patients and controls in the posterior nodes of the default mode network, reduced AIS at these areas suggested that the information stored in the neural signals at these brain areas was structured in a less predictable manner for ASD patients. This can be interpreted as the failure to represent environmental regularities in spontaneous brain activity, which has been suggested by Fiser and colleagues (2010). Further, the reduced amount of predictable information in the default mode network of the ASD brain should lead to the consequence that the brain areas receiving top-down predictions from the default mode network need to deal with less predictable, less reliable information. Hence, the present findings are in line with a reduced reliability, precision or use of prior knowledge in patients with ASD (Pellicano and Burr, 2012). A reduced amount of reliable information for predictions might also result in the failure to inhibit the gain of error units in superficial cortical layers and thus lead to an exaggerated precision-weighting of prediction errors compared to predictions (Lawson et al, 2014). Thus, the present findings are also compatible with an imbalance of precision-weighting in patients with ASD (Friston, 2013, Lawson et al., 2014). Importantly, the AIS analysis allowed to rule out accounts of merely increased bottom-up influences (Brock, 2012; Van de Cruys et al., 2014), as these are not supposed to be associated with reduced predictable information in the default mode network of ASD patients.

Within the same study, the information-theoretic approach allowed also to find a negative correlation of predictable information with symptom severity in the domain of repetitive and stereotyped behaviors and interests. Previous accounts of ASD reported correlations between rates of repetitive behavior and anatomical abnormalities in the cerebellum (Pierce and Courchesne, 2001). Hence, the correlation between repetitive behavior and predictable information in the cerebellum in study 3 suggests that a reduced amount of activated prior knowledge and anatomical abnormalities in the cerebellum might be linked.

Summing up, quantifying the information storage in neural signals with the AIS method has a broad variety of applications in neuroscience research and can in particular provide a largely assumption-free approach to study the mechanisms of predictive coding in the human brain.

5.4. Limitations of the present work and outlook for future predictive coding research

As any measurement technique, MEG is also subject to technical limitations. While it is sensitive to brain sources with a tangential orientation (more precisely the tangential components of a source), radial source components are barely represented (Hämäläinen et al., 1993). Therefore, the MEG technique mostly captures the brain sources at the walls of the cortical sulci. When brain areas of interest are located at the bottom of cortical sulci or at the top of the cortical gyri, they are not well representable with the MEG technique. This might be an alternative explanation why I was not able to find any house prediction areas in study 3. This problem might be resolved by a combination of EEG, MEG recordings and AIS analysis, which should also allow finding prediction areas at the top of the cortical gyri and at the bottom of cortical sulci.

The MEG technique is also limited in its spatial resolution. Although the spatial resolution is supposed to lie within the mm range (Hari et al., 1988), MEG is not sensitive enough to resolve individual cortical layers. In future, MEG (or EEG) recordings can be combined with high resolution fMRI recordings to clarify the contribution of different layers in predictive message passing.

In addition to the technical limitations, many open questions remain regarding the neural correlates of predictive coding in the human brain. For example, evidence for distinct neural subpopulations for the representation of errors and predictions is still missing (Clark, 2012). In order to classify cell populations as error or prediction units, AIS and TE values may be subjected to correlation analyses (see Wibral et al., 2015). For the error units, the sum of AIS in the incoming signals is supposed to be *negatively* correlated with outgoing information transfer, as this means that these cells have increased output for unpredictable information. For the prediction units, the sum of AIS in the incoming signals is supposed to be *positively* correlated with outgoing information transfer, which means that these cells have increased output when incoming information is more predictable (Wibral et al., 2015, Figure 4).

Another open question is how predictions and the underlying models are updated by evidence accumulation or by prediction errors, respectively. To study these different types of prediction updates, analysis of predictable information as measured by AIS can also be applied. It might be hypothesized that predictable information would gradually increase when predictions are updated by evidence accumulation – as more information can be used for the predictions. However, predictable information might decrease when the predictions are updated by prediction errors as previously stored information for the predictions should be discarded.

In summary, information-theoretic analysis methods may help to fill the gaps in predictive coding research and test the claims made in theoretical considerations (e.g. Friston, 2010, Bastos et al., 2012; Clark, 2012). Especially the claim of predictive coding as a universal and fundamental functional principle of the brain (e.g. Friston, 2010; Huang and Rao, 2011) requires also the accumulation of neural evidence in different modalities, different experimental designs and for different groups of participants.

To give an example, the face detection task in study 1 could be repeated with ASD patients and additional AIS analysis. This would allow the investigation of questions that can’t be answered in a resting state design, e.g. how predictions are combined with sensory evidence in ASD.

Application of the analysis of predictable information to other neuropsychiatric disorders than ASD may also become a potential focus of future research. Interestingly, current accounts suggest that schizophrenic symptoms like hallucinations might also be caused by an impaired precision-weighting of predictions and sensory evidence (Adams et al., 2013). Thus, the next step could be analyzing predictable information in patients diagnosed with schizophrenia to investigate this hypothesis.

5.5. Conclusion

In this work, I investigated the neural correlates of the use of prior knowledge in predictive coding with three MEG studies. The results of these studies provide neurophysiological evidence for the neural implementation of predictive coding theory as proposed by Bastos et al. (2012) – in particular for an alpha/beta frequency channel for predictions and a gamma frequency channel for prediction errors. Furthermore, by application of information-theoretic measures to quantify predictable information in neural signals of healthy participants as well as ASD patients, I introduced a largely assumption-free method to study the neural correlates of predictive coding in the healthy and diseased human brain.

6. Bibliography

Achenbach T (1990) Infant Assessment Unit-Young Adult Self Report. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach TM, Edelbrock CS (1991) Youth Self-report and Profile. University of Vermont, Department of psychiatry.

Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ (2013) The Computational Anatomy of Psychosis. Front Psychiatry.

Adams WJ (2007) A common light-prior for visual search, shape, and reflectance judgments. J Vis 7.

Adams WJ, Graf EW, Ernst MO (2004) Experience can change the “light-from-above” prior. Nat Neurosci 7:1057–1058.

Anderson M, Ter Braak C (2003) Permutation tests for multi-factorial analysis of variance. J Stat Comput Simul 73:85–113.

Apps MAJ, Tsakiris M (2013) Predictive codes of familiarity and context during the perceptual learning of facial identities. Nat Commun 4:2698.

Arnal LH, Giraud AL (2012) Cortical oscillations and sensory predictions. Trends Cogn Sci.

Arnal LH, Wyart V, Giraud AL (2011) Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. Nat Neurosci 14:797–801.

Asperger H (1944) Die „Autistischen Psychopathen” im Kindesalter. Eur Arch Psychiatry Clin Neurosci 117:76–136.

American Psychiatric Association (2013): Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub.

Arbeitsgruppe Deutsche Child Behavior Checklist (1998a): Fragebogen für Jugendliche; deutsche Bearbeitung der Youth Self-Report Form der Child Behavior Checklist (YSR). Einführung und Anleitung zur Handauswertung mit deutschen Normen, bearbeitet von M. Döpfner, J. Plück, S. Bölte, K. Lenz, P. Melchers & K. Heim (2. Aufl.). Köln: Arbeitsgruppe Kinder-, Jugend- und Familiendiagnostik (KJFD).

Arbeitsgruppe Deutsche Child Behavior Checklist. (1998b): Fragebogen für junge Erwachsene (YASR). Köln: Arbeitsgruppe Kinder-, Jugend- und Familiendiagnostik (KJFD)

Bar M (2004) Visual objects in context. Nat Rev Neurosci 5:617–629.

Bar M (2007) The proactive brain: using analogies and associations to generate predictions. Trends Cogn Sci 11:280–289

Baron-Cohen S, Leslie AM, Frith U (1985) Does the autistic child have a “theory of mind”? Cognition 21:37–46.

Barone P, Batardiere A, Knoblauch K, Kennedy H (2000) Laminar Distribution of Neurons in Extrastriate Areas Projecting to Visual Areas V1 and V4 Correlates with the Hierarchical Rank and Indicates the Operation of a Distance Rule. J Neurosci 20:3263–3281.

Barrett LF (2017) The theory of constructed emotion: an active inference account of interoception and categorization. Soc Cogn Affect Neurosci 12:1–23.

Barrett LF, Satpute AB (2013) Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. Curr Opin Neurobiol 23:361–372.

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding. Neuron 76:695–711.

Bastos AM, Vezoli J, Bosman CA, Schoffelen J-M, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H, Fries P (2015) Visual areas exert feedforward and feedback influences through distinct frequency channels. Neuron 85:390–401.

Bauer M, Stenner M-P, Friston KJ, Dolan RJ (2014) Attentional Modulation of Alpha/Beta and Gamma Oscillations Reflect Functionally Distinct Processes. J Neurosci 34:16117–16125.

Biederman I, Glass AL, Stacy EW (1973) Searching for objects in real-world scenes. J Exp Psychol 97:22.

Bölte S, Rühl D, Schmötzer G, Poustka F (2006) Diagnostisches interview für autismus-revidiert (ADI-R). Bern Huber.

Bosl W, Tierney A, Tager-Flusberg H, Nelson C (2011) EEG complexity as a biomarker for autism spectrum disorder risk. BMC Med 9:18.

Botvinick MM, Cohen JD, Carter CS (2004) Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn Sci 8:539–546.

Brambilla P, Hardan A, di Nemi SU, Perez J, Soares JC, Barale F (2003) Brain anatomy and development in autism: review of structural MRI studies. Brain Res Bull 61:557–569.

Bressler SL, Richter CG (2015) Interareal oscillatory synchronization in top-down neocortical processing. Curr Opin Neurobiol 31:62–66.

Brewster SD (1847) LXVII. On the conversion of relief by inverted vision. Lond Edinb Dublin Philos Mag J Sci 30:432–437.

Brock J (2012) Alternative Bayesian accounts of autistic perception: comment on Pellicano and Burr. Trends Cogn Sci 16:573–574.

Brodski A, Paasch G-F, Helbling S, Wibral M (2015) The Faces of Predictive Coding. J Neurosci 35:8997–9006.

Brodski-Guerniero A, Paasch G-F, Wollstadt P, Özdemir I, Lizier JT, Wibral M (2017) Information-theoretic evidence for predictive coding in the face-processing system. J Neurosci 37: 8273–8283.

Brookes MJ, Gibson AM, Hall SD, Furlong PL, Barnes GR, Hillebrand A, Singh KD, Holliday IE, Francis ST, Morris PG (2005) GLM-beamformer method demonstrates stationary field, alpha ERD and gamma ERS co-localisation with fMRI BOLD response in visual cortex. Neuroimage 26:302–308.

Brookes MJ, Vrba J, Robinson SE, Stevenson CM, Peters AM, Barnes GR, Hillebrand A, Morris PG (2008) Optimising experimental design for MEG beamformer imaging. NeuroImage 39:1788–1802.

Buckner RL, Andrews-Hanna JR, Schacter DL (2008) The Brain's Default Network. Ann N Y Acad Sci 1124:1–38.

Buffalo EA, Fries P, Landman R, Buschman TJ, Desimone R (2011) Laminar differences in gamma and alpha coherence in the ventral stream. Proc Natl Acad Sci 108:11262–11267.

Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR (2013) Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci.

Carpenter GA, Grossberg S (2010) Adaptive Resonance Theory. CASCNS Tech Rep Ser 0 .

Catarino A, Churches O, Baron-Cohen S, Andrade A, Ring H (2011) Atypical EEG complexity in autism spectrum conditions: a multiscale entropy analysis. Clin Neurophysiol 122:2375–2383.

Cavanagh P (1991) What's up in top-down processing. Represent Vis Trends Tacit Assumpt Vis Res:295–304.

Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol TIST 2:27.

Cherkassky VL, Kana RK, Keller TA, Just MA (2006) Functional connectivity in a baseline resting-state network in autism. Neuroreport 17:1687–1690.

Christensen DL (2016) Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. MMWR Surveill Summ 65

Ciaramelli E, Grady CL, Moscovitch M (2008) Top-down and bottom-up attention to memory: A hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval. Neuropsychologia 46:1828–1851.

Clark A (2012) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci.

Collins DL, Neelin P, Peters TM, Evans AC, others (1994) Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 18:192.

Courtney SM, Petit L, Maisog JM, Ungerleider LG, Haxby JV (1998) An Area Specialized for Spatial Working Memory in Human Frontal Cortex. Science 279:1347–1351.

Cover TM, Thomas JA (2012) Elements of information theory. John Wiley & Sons.

Dhamala M, Rangarajan G, Ding M (2008) Estimating Granger causality from Fourier and wavelet transforms of time series data. Phys Rev Lett 100:018701.

Dienes Z (2014) Using Bayes to get the most out of non-significant results. Front Psychol 5:781.

Dolan RJ, Fink GR, Rolls E, Booth M, Holmes A, Frackowiak RSJ, Friston KJ, others (1997) How the brain learns to see objects and faces in an impoverished context. Nature 389:596–598.

Engel AK, Fries P (2010) Beta-band oscillations—signalling the status quo? Curr Opin Neurobiol 20:156–165.

Esterman M, Yantis S (2009) Perceptual expectation evokes category-selective cortical activity. Cereb Cortex:bhp188.

Faes L, Nollo G, Porta A (2013) Compensated transfer entropy as a tool for reliably estimating information transfer in physiological time series. Entropy 15:198–219.

Farah MJ, Tanaka JW, Drain HM (1995) What causes the face inversion effect? J Exp Psychol Hum Percept Perform 21:628–634.

Fatemi SH, Aldinger KA, Ashwood P, Bauman ML, Blaha CD, Blatt GJ, Chauhan A, Chauhan V, Dager SR, Dickson PE, others (2012) Consensus paper: pathological role of the cerebellum in autism. The Cerebellum 11:777–807.

Feldman H, Friston KJ (2010) Attention, Uncertainty, and Free-Energy. Front Hum Neurosci 4

Fenske MJ, Aminoff E, Gronau N, Bar M (2006) Top-down facilitation of visual object recognition: object-based and context-based contributions. Progress in brain research 155:3–21.

Fink GR, Halligan PW, Marshall JC, Frith CD, Frackowiak RS, Dolan RJ (1997) Neural mechanisms involved in the processing of global and local aspects of hierarchically organized visual stimuli. Brain 120:1779–1791.

Fiser J, Berkes P, Orbán G, Lengyel M (2010) Statistically optimal perception and learning: from behavior to neural representations. Trends Cogn Sci 14:119–130.

Fox CJ, Moon SY, Iaria G, Barton JJS (2009) The correlates of subjective perception of identity and expression in the face network: An fMRI adaptation study. NeuroImage 44:569–580.

Frenzel S, Pompe B (2007) Partial mutual information for coupling analysis of multivariate time series. Phys Rev Lett 99:204101.

Fries P, Reynolds JH, Rorie AE, Desimone R (2001) Modulation of Oscillatory Neuronal Synchronization by Selective Visual Attention. Science 291:1560–1563.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc B Biol Sci 360:815–836.

Friston K (2009) The free-energy principle: a rough guide to the brain? Trends Cogn Sci 13:293–301.

Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11:127–138.

Friston K, Kiebel S (2009) Predictive coding under the free-energy principle. Philos Trans R Soc B Biol Sci 364:1211–1221.

Friston KJ, Lawson R, Frith CD (2013) On hyperpriors and hypopriors: comment on Pellicano and Burr. Trends Cogn Sci 17:10–1016.

Frith C (2003) What do imaging studies tell us about the neural basis of autism. Autism Neural Basis Treat Possibilities:149–176.

Frith U (1989) Autism: Explaining the enigma. Wiley Online Library.

Frith U, Happé F (1994) Autism: beyond “theory of mind.” Cognition 50:115–132.

George D, Hawkins J (2009) Towards a Mathematical Theory of Cortical Micro-circuits. PLoS Comput Biol 5:e1000532.

Gerardin P, Kourtzi Z, Mamassian P (2010) Prior knowledge of illumination for 3D perception in the human brain. Proc Natl Acad Sci 107:16309–16314.

Goense J, Merkle H, Logothetis NK (2012) High-resolution fMRI reveals laminar differences in neurovascular coupling between positive and negative BOLD responses. Neuron 76:629–639.

Gómez C, Lizier JT, Schaum M, Wollstadt P, Grützner C, Uhlhaas P, Freitag CM, Schlitt S, Bölte S, Hornero R, others (2014) Reduced predictable information in brain signals in autism spectrum disorder. Front Neuroinformatics 8.

Gómez-Herrero G, Wu W, Rutanen K, Soriano MC, Pipa G, Vicente R (2015) Assessing coupling dynamics from an ensemble of time series. Entropy 17:1958–1970.

Grandin T (1992) An inside view of autism. High Funct Individ Autism:105–126.

Granger CWJ (1969) Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica 37:424–438.

Green DM, Swets JA (1966) Signal detection theory and psychophysics. Wiley New York.

Gross J, Baillet S, Barnes GR, Henson RN, Hillebrand A, Jensen O, Jerbi K, Litvak V, Maess B, Oostenveld R (2012) Good-practice for conducting and reporting MEG research. NeuroImage

Gross J, Kujala J, Hamalainen M, Timmermann L, Schnitzler A, Salmelin R (2001) Dynamic imaging of coherent sources: Studying neural interactions in the human brain. Proc Natl Acad Sci U S A 98:694–699.

Grossberg S (2007) Towards a unified theory of neocortex: laminar cortical circuits for vision and cognition. Prog Brain Res 165:79–104.

Grossberg S (2012) ADAPTIVE RESONANCE THEORY How a brain learns to consciously attend, learn, and recognize a changing world. Neural Netw

Grützner C, Uhlhaas PJ, Genc E, Kohler A, Singer W, Wibral M (2010) Neuroelectromagnetic Correlates of Perceptual Closure Processes. J Neurosci 30:8342–8352.

Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV (1993) Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev Mod Phys 65:413–497.

Happé F, Briskman J, Frith U (2001) Exploring the cognitive phenotype of autism: weak “central coherence” in parents and siblings of children with autism: I. Experimental tests. J Child Psychol Psychiatry 42:299–307.

Happé F, Frith U (2006) The weak coherence account: detail-focused cognitive style in autism spectrum disorders. J Autism Dev Disord 36:5–25.

Happé FG (1996) Studying weak central coherence at low levels: children with autism do not succumb to visual illusions. A research note. J Child Psychol Psychiatry 37:873–877.

Hari R, Joutsiniemi S-L, Sarvas J (1988) Spatial resolution of neuromagnetic records: theoretical calculations in a spherical model. Electroencephalogr Clin Neurophysiol Potentials Sect 71:64–72.

Harrison CW (1952) Experiments with linear prediction in television. Bell Syst Tech J 31:764–783.

Hawkins J, Blakeslee S (2005) On intelligence. Holt Paperbacks.

Hegdé J, Van Essen DC (2000) Selectivity for complex shapes in primate visual area V2. J Neurosci Off J Soc Neurosci 20:RC61.

Herrmann CS, Munk MHJ, Engel AK (2004) Cognitive functions of gamma-band activity: memory match and utilization. Trends Cogn Sci 8:347–355.

Hohwy J (2013) The Predictive Mind. Oxford University Press.

Hohwy J, Roepstorff A, Friston K (2008) Predictive coding explains binocular rivalry: an epistemological review. Cognition 108:687–701.

Holm S (1979) A Simple Sequentially Rejective Multiple Test Procedure. Scand J Stat 6:65–70.

Hoogenboom N, Schoffelen J-M, Oostenveld R, Fries P (2010) Visually induced gamma-band activity predicts speed of change detection in humans. NeuroImage 51:1162–1167.

Hoogenboom N, Schoffelen J-M, Oostenveld R, Parkes LM, Fries P (2006) Localizing human visual gamma-band activity in frequency, time and space. Neuroimage 29:764–773.

Huang MX, Shih JJ, Lee RR, Harrington DL, Thoma RJ, Weisend MP, Hanlon F, Paulson KM, Li T, Martin K, others (2004) Commonalities and differences among vectorized beamformers in electromagnetic source imaging. Brain Topogr 16:139–158.

Huang Y, Rao RP (2011) Predictive coding. Wiley Interdiscip Rev Cogn Sci 2:580–593.

Hurlburt RT, Happe F, Frith U (1994) Sampling the form of inner experience in three adults with Asperger syndrome. Psychol Med 24:385–395.

Jeffreys H (1998) The theory of probability. OUP Oxford.

Joseph RM, Keehn B, Connolly C, Wolfe JM, Horowitz TS (2009) Why is visual search superior in autism spectrum disorder? Dev Sci 12:1083–1096.

Kanai R, Rees G (2011) The structural basis of inter-individual differences in human behaviour and cognition. Nat Rev Neurosci 12:231–242.

Kanner L (1943) Autistic disturbances of affective contact. Available at: http://neurodiversity.com/library_kanner_1943.pdf [Accessed March 2, 2017].

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311.

Kay JW, Phillips WA (2011) Coherent infomax as a computational goal for neural systems. Bull Math Biol 73:344–372.

Kemelmacher-Shlizerman I, Basri R, Nadler B (2008) 3D shape reconstruction of Mooney faces. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pp 1–8.

Kennedy DP, Redcay E, Courchesne E (2006) Failing to deactivate: resting functional abnormalities in autism. Proc Natl Acad Sci 103:8275–8280.

Kersten D, Mamassian P, Yuille A (2004) Object perception as Bayesian inference. Annu Rev Psychol 55:271–304.

Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. TRENDS Neurosci 27:712–719.

Kok P, Failing MF, de Lange FP (2014) Prior expectations evoke stimulus templates in the primary visual cortex. J Cogn Neurosci 26:1546–1554.

Kourtzi Z, Tolias AS, Altmann CF, Augath M, Logothetis NK (2003) Integration of local features into global shapes-monkey and human fMRI studies. Neuron 37:333–346.

Kozachenko LF, Leonenko NN (1987) Sample estimate of the entropy of a random vector. Probl Peredachi Informatsii 23:9–16.

Kraskov A, Stögbauer H, Grassberger P (2004) Estimating mutual information. Phys Rev E 69:066138.

Kriegeskorte N, Formisano E, Sorger B, Goebel R (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. Proc Natl Acad Sci 104:20600–20605.

Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 12:535–540.

Langton CG (1990) Computation at the edge of chaos: phase transitions and emergent computation. Phys Nonlinear Phenom 42:12–37.

Larkum M (2013) A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. Trends Neurosci 36:141–151.

Lawson RP, Rees G, Friston KJ (2014) An aberrant precision account of autism. Front Hum Neurosci 8:302.

Liang F, Paulo R, Molina G, Clyde MA, Berger JO (2008) Mixtures of g priors for Bayesian variable selection. J Am Stat Assoc 103:410–423.

Lindner M, Vicente R, Priesemann V, Wibral M (2011) TRENTOOL: A Matlab open source toolbox to analyse information flow in time series data with transfer entropy. BMC Neurosci 12:119.

Lizier JT (2014) JIDT: an information-theoretic toolkit for studying the dynamics of complex systems. Comput Intell 1:11.

Lizier JT, Prokopenko M, Zomaya AY (2012) Local measures of information storage in complex distributed computation. Inf Sci 208:39–54.

Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, Pickles A, Rutter M (2000) The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord 30:205–223.

Lynch CJ, Uddin LQ, Supekar K, Khouzam A, Phillips J, Menon V (2013) Default mode network in childhood autism: posteromedial cortex heterogeneity and relationship with social deficits. Biol Psychiatry 74:212–219.

Makeig S, Bell AJ, Jung T-P, Sejnowski TJ, others (1996) Independent component analysis of electroencephalographic data. Adv Neural Inf Process Syst:145–151.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190.

Marr D (1982) Vision San Francisco. W H Freeman Co.

Mayer A, Schwiedrzik CM, Wibral M, Singer W, Melloni L (2015) Expecting to See a Letter: Alpha Oscillations as Carriers of Top-Down Sensory Predictions. Cereb Cortex:bhv146.

Michalareas G, Vezoli J, van Pelt S, Schoffelen J-M, Kennedy H, Fries P (2016) Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. Neuron .

Monk CS, Peltier SJ, Wiggins JL, Weng S-J, Carrasco M, Risi S, Lord C (2009) Abnormalities of intrinsic functional connectivity in autism spectrum disorders. Neuroimage 47:764–772.

Mooney CM (1957) Age in the development of closure ability in children. Can J Psychol Can Psychol 11:219–226.

Mooney CM, Ferguson GA (1951) A new closure test. Can J Psychol Can Psychol 5:129–133.

Moore C, Cavanagh P (1998) Recovery of 3D volume from 2-tone images of novel objects. Cognition 67:45–71.

Morey RD, Rouder JN, Jamil T, Morey MRD (2015) Package 'BayesFactor' . Available at: ftp://alvarestech.com/pub/plan/R/web/packages/BayesFactor/BayesFactor.pdf

Mumford D (1992) On the computational architecture of the neocortex. Biol Cybern 66:241–251.

Nieuwenhuis ILC, Takashima A, Oostenveld R, Fernández G, Jensen O (2008) Visual areas become less engaged in associative recall following memory stabilization. NeuroImage 40:1319–1327.

Nolte G (2003) The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. Phys Med Biol 48:3637–3652.

Nowotny T (2014) Two challenges of correct validation in pattern recognition. Comput Intell 1:5.

Nunez PL, Srinivasan R (2006) A theoretical basis for standing and traveling brain waves measured with human EEG with implications for an integrated consciousness. Clin Neurophysiol Off J Int Fed Clin Neurophysiol 117:2424–2435.

Ohayon S, Freiwald WA, Tsao DY (2012) What makes a cell face selective? The importance of contrast. Neuron 74:567–581.

Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9:97–113.

Oliver BM (1952) Efficient coding. Bell Labs Tech J 31:724–750.

Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:1.

Pashkam MV, Xu Y (2014) Decoding visual object representation in human parietal cortex. J Vis 14:1307–1307.

Pellicano E, Burr D (2012a) When the world becomes “too real”: a Bayesian explanation of autistic perception. Trends Cogn Sci 16:504–510.

Pellicano E, Burr D (2012b) Response to Brock: noise and autism. Trends Cogn Sci 16:574–575.

Pelt S van, Heil L, Kwisthout J, Ondobaka S, Rooij I van, Bekkering H (2016) Beta- and gamma-band activity reflect predictive coding in the processing of causal events. Soc Cogn Affect Neurosci:nsw017.

Percival DB, Walden AT (1993) Spectral Analysis for Physical Applications. Cambridge University Press.

Pernet CR, Wilcox RR, Rousselet GA (2013) Robust correlation analyses: false positive and power validation using a new open source Matlab toolbox. Front Psychol 3:606.

Pierce K, Courchesne E (2001) Evidence for a cerebellar role in reduced exploration and stereotyped behavior in autism. Biol Psychiatry 49:655–664.

Pitcher D, Walsh V, Duchaine B (2011) The role of the occipital face area in the cortical face perception network. Exp Brain Res 209:481–493.

Plaisted K, O’Riordan M, Baron-Cohen S (1998) Enhanced Discrimination of Novel, Highly Similar Stimuli by Adults with Autism During a Perceptual Learning Task. J Child Psychol Psychiatry 39:765–775.

Puri AM, Wojciulik E, Ranganath C (2009) Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. Brain Res 1301:89–99.

Ragwitz M, Kantz H (2002) Markov models from data by simple nonlinear time series predictors in delay embedding spaces. Phys Rev E 65:056201.

Raichle ME (2015) The brain’s default mode network. Annu Rev Neurosci 38:433–447.

Ranganath C, Cohen MX, Dam C, D’Esposito M (2004) Inferior Temporal, Prefrontal, and Hippocampal Contributions to Visual Working Memory Maintenance and Associative Memory Retrieval. J Neurosci 24:3917–3925.

Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87.

Rodriguez E, George N, Lachaux JP, Martinerie J, Renault B, Varela FJ (1999) Perception’s shadow: long-distance synchronization of human brain activity. Nature 397:430–433.

Roopun AK, Kramer MA, Carracedo LM, Kaiser M, Davies CH, Traub RD, Kopell NJ, Whittington MA (2008) Period Concatenation Underlies Interactions between Gamma and Beta Rhythms in Neocortex. Front Cell Neurosci 2 .

Roopun AK, Middleton SJ, Cunningham MO, LeBeau FE, Bibbig A, Whittington MA, Traub RD (2006) A beta2-frequency (20–30 Hz) oscillation in nonsynaptic networks of somatosensory cortex. Proc Natl Acad Sci 103:15646–15650.

Roth G (1997) Das Gehirn und seine Wirklichkeit. Kognitive Neurobiologie und ihre philosophischen Konsequenzen. Suhrkamp Taschenbuch Wissenschaft.

Rouder JN, Morey RD (2012) Default Bayes factors for model selection in regression. Multivar Behav Res 47:877–903.

Rousseeuw PJ (1984) Least median of squares regression. J Am Stat Assoc 79:871–880.

Rousseeuw PJ, Driessen KV (1999) A fast algorithm for the minimum covariance determinant estimator. Technometrics 41:212–223.

Rousselet GA, Pernet CR (2012) Improving standards in brain-behavior correlation analyses. Front Hum Neurosci 6:119.

Rühl D, Bölte S, Feineis-Matthews S, Poustka F (2004) Diagnostische Beobachtungsskala für Autistische Störungen (ADOS). Bern Huber.

Russell JE (1997) Autism as an executive disorder. Oxford University Press.

Rutter M, Le Couteur A, Lord C (2003) Autism diagnostic interview-revised. Los Angel CA West Psychol Serv 29:30.

Sadaghiani S, Hesselmann G, Friston KJ, Kleinschmidt A (2010) The relation of ongoing brain activity, evoked neural responses, and cognition. Front Syst Neurosci 4:20.

Schreiber T (2000) Measuring information transfer. Phys Rev Lett 85:461.

Sedley W, Gander PE, Kumar S, Kovach CK, Oya H, Kawasaki H, Iii MAH, Griffiths TD (2016) Neural signatures of perceptual inference. eLife 5:e11476.

Shah A, Frith U (1983) An Islet of Ability in Autistic Children: A Research Note. J Child Psychol Psychiatry 24:613–620.

Shannon CE (2001) A mathematical theory of communication. ACM SIGMOBILE Mob Comput Commun Rev 5:3–55.

Sinha P (2002) Qualitative representations for recognition. In: Biologically motivated computer vision, pp 249–262. Springer.

Slepian D (1978) Prolate spheroidal wave functions, Fourier analysis and uncertainty. Bell Syst Tech J 57:1371–1429.

Spratling MW (2008) Reconciling Predictive Coding and Biased Competition Models of Cortical Function. Front Comput Neurosci 2 .

Spratling MW (2017) A review of predictive coding algorithms. Brain Cogn 112:92–97.

Suckling J, Bullmore E (2004) Permutation tests for factorially designed neuroimaging experiments. Hum Brain Mapp 22:193–205.

Summerfield C, Egner T, Greene M, Koechlin E, Mangels J, Hirsch J (2006a) Predictive Codes for Forthcoming Perception in the Frontal Cortex. Science 314:1311–1314.

Summerfield C, Egner T, Mangels J, Hirsch J (2006b) Mistaking a house for a face: neural correlates of misperception in healthy humans. Cereb Cortex N Y N 1991 16:500–508.

Sun J, Perona P (1998) Where is the sun? Nat Neurosci 1:183–184.

Taira M, Nose I, Inoue K, Tsutsui K (2001) Cortical areas related to attention to 3D surface structures based on shading: an fMRI study. Neuroimage 14:959–966.

R Core Team (2016) A language and environment for statistical computing. R Foundation for statistical computing, 2015; Vienna, Austria.

Teufel C, Subramaniam N, Fletcher PC (2013) The role of priors in Bayesian models of perception. Front Comput Neurosci 7.

Todorovic A, Ede F van, Maris E, Lange FP de (2011) Prior Expectation Mediates Neural Adaptation to Repeated Sounds in the Auditory Cortex: An MEG Study. J Neurosci 31:9118–9123.

Trapp S, Lepsien J, Kotz SA, Bar M (2015) Prior probability modulates anticipatory activity in category-specific areas. Cogn Affect Behav Neurosci:1–10.

Tsao DY, Moeller S, Freiwald WA (2008) Comparing face patch systems in macaques and humans. Proc Natl Acad Sci 105:19514–19519.

Uhlhaas PJ, Pipa G, Neuenschwander S, Wibral M, Singer W (2011) A new look at gamma? High-(> 60 Hz) γ-band activity in cortical networks: function, mechanisms and impairment. Prog Biophys Mol Biol 105:14–28.

Valentine T (1988) Upside-down faces: A review of the effect of inversion upon face recognition. Br J Psychol 79:471–491.

Van Belle G, De Graef P, Verfaillie K, Busigny T, Rossion B (2010) Whole not hole: Expert face recognition requires holistic perception. Neuropsychologia 48:2620–2629.

van Boxtel JJA, Lu H (2013) A predictive coding perspective on autism spectrum disorders. Front Psychol 4:19.

Van de Cruys S, Evers K, Van der Hallen R, Van Eylen L, Boets B, de-Wit L, Wagemans J (2014) Precise minds in uncertain worlds: predictive coding in autism. Psychol Rev 121:649.

Van Veen BD, Van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng 44:867–880.

Verboven S, Hubert M (2005) LIBRA: a MATLAB library for robust analysis. Chemom Intell Lab Syst 75:127–136.

Vicente R, Wibral M, Lindner M, Pipa G (2011) Transfer entropy—a model-free measure of effective connectivity for the neurosciences. J Comput Neurosci 30:45–67.

Von Helmholtz H (1867) Handbuch der physiologischen Optik. Voss.

Vuilleumier P, Schwartz S (2001) Emotional facial expressions capture attention. Neurology 56:153–158.

Wagner AD, Shannon BJ, Kahn I, Buckner RL (2005) Parietal lobe contributions to episodic memory retrieval. Trends Cogn Sci 9:445–453.

Waiter GD, Williams JHG, Murray AD, Gilchrist A, Perrett DI, Whiten A (2004) A voxel-based investigation of brain structure in male adolescents with autistic spectrum disorder. NeuroImage 22:619–625.

Wang X-J (2010) Neurophysiological and Computational Principles of Cortical Rhythms in Cognition. Physiol Rev 90:1195–1268.

Washington SD, Gordon EM, Brar J, Warburton S, Sawyer AT, Wolfe A, Mease-Ference ER, Girton L, Hailu A, Mbwana J, others (2014) Dysmaturation of the default mode network in autism. Hum Brain Mapp 35:1284–1296.

Weiß RH (2006) Grundintelligenztest skala 2—revision CFT 20-R [culture fair intelligence test scale 2—revision]. Hogrefe Gött.

Weng S-J, Wiggins JL, Peltier SJ, Carrasco M, Risi S, Lord C, Monk CS (2010) Alterations of resting state functional connectivity in the default network in adolescents with autism spectrum disorders. Brain Res 1313:202–214.

Whitfield-Gabrieli S, Ford JM (2012) Default mode network activity and connectivity in psychopathology. Annu Rev Clin Psychol 8:49–76.

Wibral M, Lizier JT, Priesemann V (2015) Bits from Brains for Biologically-Inspired Computing. Name Front Robot AI.

Wibral M, Lizier JT, Vögler S, Priesemann V, Galuske R (2014) Local active information storage as a tool to understand distributed neural information processing. Front Neuroinformatics.

Wibral M, Pampu N, Priesemann V, Siebenhühner F, Seiwert H, Lindner M, Lizier JT, Vicente R (2013) Measuring information-transfer delays. PloS One 8:e55809.

Wibral M, Rahm B, Rieder M, Lindner M, Vicente R, Kaiser J (2011) Transfer entropy in magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks. Prog Biophys Mol Biol 105:80–97.

Wiener N (1956) The theory of prediction. Mod Math Eng N Y McGraw-Hill:165–190.

Wilson GT (1972) The factorization of matricial spectral densities. SIAM J Appl Math 23:420–426.

Winston JS, Henson RNA, Fine-Goulden MR, Dolan RJ (2004) fMRI-Adaptation Reveals Dissociable Neural Representations of Identity and Expression in Face Perception. J Neurophysiol 92:1830–1839.

Wollstadt P, Martínez-Zarzuela M, Vicente R, Díaz-Pernas FJ, Wibral M (2014) Efficient transfer entropy analysis of non-stationary neural time series. PLoS One 9:e102833.

Wollstadt P, Sellers KK, Rudelt L, Priesemann V, Hutt A, Fröhlich F, Wibral M (2016) The relation of local entropy and information transfer suggests an origin of isoflurane anesthesia effects in local information processing. ArXiv Prepr ArXiv160808387

World Health Organization (1992) The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization.

Xing D, Yeh C-I, Burns S, Shapley RM (2012) Laminar analysis of visually evoked activity in the primary visual cortex. Proc Natl Acad Sci 109:13871–13876.

Yin RK (1969) Looking at upside-down faces. J Exp Psychol 81:141.

Zhen Z, Fang H, Liu J (2013) The hierarchical brain network for face recognition. PloS One 8:e59886.

Zipser D, Kehoe B, Littlewort G, Fuster J (1993) A spiking network model of short-term active memory. J Neurosci 13:3406–3420.

7. Zusammenfassung

Während der Wahrnehmung unserer Umgebung greifen wir ständig auf gespeichertes Wissen aus unserer bisherigen Erfahrung zurück. Die Idee, dass solch ein Vorwissen für unsere Wahrnehmung essentiell ist, geht zurück auf Hermann von Helmholtz (1867). Dieser stellte die Theorie auf, dass Wahrnehmung einen Vorgang „unbewusster Schlussfolgerung“ darstellt, in dem Vorwissen es dem Gehirn erleichtert, die Ursachen für seinen sensorischen Input zu erschließen. Dieser Schlussfolgerungsvorgang soll im Gehirn mit Hilfe des „Predictive Coding“ Prinzips umgesetzt werden (z.B. Mumford, 1992; Rao et al., 1999; Friston, 2005, 2010; Hawkins and Blakeslee, 2005; Clark, 2012; Hohwy, 2013). So besagt die zurzeit populärste Variante der Predictive Coding Theorie („Rao und Ballard Version“ in Spratling, 2017), dass das Gehirn basierend auf Vorwissen Vorhersagen (*predictions*) in höheren kortikalen Arealen generiert, die dann an hierarchisch tiefer gelegene kortikale Areale geleitet werden. In den tiefer gelegenen kortikalen Arealen wird die durch die Vorhersage erwartete neuronale Repräsentation mit der tatsächlichen Repräsentation verglichen. Diskrepanzen resultieren dabei in einem Vorhersagefehler (*prediction error*), der wiederum aufwärts an die höheren kortikalen Areale geleitet wird, in denen als Konsequenz die Vorhersage angepasst wird. Dabei können mehrere Schleifendurchläufe stattfinden, bis der Vorhersagefehler minimiert und somit die wahrscheinlichste Ursache für die eingehende Information bestimmt wurde.

Hinsichtlich der neuronalen Implementierung der Predictive Coding Theorie im Gehirn stellten Bastos und Kollegen vor wenigen Jahren die Hypothese auf, dass schnelle neuronale Aktivität im Gamma Frequenzbereich (> 30 Hz) die Weiterleitung eines Vorhersagefehlers an höhere kortikale Areale widerspiegelt, wohingegen langsamere neuronale Aktivität (< 30 Hz) mit der Weiterleitung von Vorhersagen an tiefere kortikale Areale in Zusammenhang stehen soll (Bastos et al., 2012).

Obwohl die Predictive Coding Theorie in den letzten Jahren deutlich an Popularität gewonnen hat (siehe z.B. Clark, 2012), und sogar die Hypothese aufkam, sie könne eine universelle Erklärung der Gehirnfunktion liefern (Friston, 2010; Huang and Rao, 2011), wirft die bisherige Forschung auf diesem Gebiet noch viele offene Fragen auf. Drei maßgebliche Lücken, die es in der Predictive Coding Forschung zu schließen gilt, sind die folgenden:

1. Die neuronalen Mechanismen, die dem Einfluss von Vorwissen zugrunde liegen, sind nach wie vor weitgehend unerforscht.
2. Bislang liegen nur wenige neurophysiologischen Evidenzen für die neuronale Implementierung des Predictive Coding Prinzips vor.
3. Um die dem Predictive Coding Prinzip unterliegenden Mechanismen zu untersuchen, fehlen nach wie vor Methoden, welche sich nicht auf Vorannahmen stützen, z.B. über die spezifischen beteiligten Areale.

Diese Lücken versuche ich in der vorliegenden Arbeit mit Hilfe von drei Studien zu schließen. In diesen drei Studien wird die neuronale Aktivität der Teilnehmer mit Magnetoenzephalographie (MEG) erfasst, was eine zeitaufgelöste Analyse der neuronalen Quellensignale ermöglicht.

In Studie 1 („The faces of predictive coding“, publiziert im *Journal of Neuroscience*, 2015; n = 48) untersuche ich, wie sich Vorwissen aus unserer lebenslangen Erfahrung auf unsere Wahrnehmung auswirkt. Dabei liegt der Fokus auf den neuronalen Korrelaten von Vorhersagefehlern. Diese Vorhersagefehler werden durch die Verletzung von Vorhersagen erzeugt, die auf unserer (visuellen) Erfahrung mit Gesichtern basieren: 1. Die aufrechte Orientierung der Gesichter; 2. Die Beleuchtung von oben. Die Analyse der Verhaltensdaten aus Studie 1 zeigt, dass die Verletzung dieser Vorhersagen die Wahrnehmung von Gesichtern sowohl verlangsamt als auch erschwert. Auf MEG Quellebene beobachte ich einen frühen Vorhersagefehler für die unerwartete Orientierung in visuellen Gehirnarealen, sowie einen späteren Vorhersagefehler für die unerwartete Beleuchtungsrichtung in Arealen, die mit räumlichem Arbeitsgedächtnis, Rekonstruktion von Form aus Schattenwurf (*shape-from-shading*) und Fehlererkennung in Zusammenhang stehen. In Einklang mit der Theorie von Bastos et al. (2012) spiegeln sich beide Vorhersagefehler in erhöhter hochfrequenter Gamma-Band Aktivität (> 64 Hz) wider.

Für die nachfolgenden Studien wird die „klassische“ neurophysiologische Analyse durch informationstheoretische Analysemethoden ergänzt. Diese ermöglichen es, die neuronalen Mechanismen des Predictive Coding Prinzips mit nur wenigen Vorannahmen zu untersuchen (siehe Wibral et al, 2015). Hierbei verwende ich primär das informationstheoretische Verfahren *Active Information Storage* (AIS, Lizier et al., 2012). AIS lässt sich mit aktivitätsgetragenem Informationsspeicher übersetzen und ermöglicht zu quantifizieren, in welchen Arealen Information für den nachfolgenden Verarbeitungsschritt eines Prozesses aufrechterhalten wird. Deshalb verwende ich AIS in Studie 2 und 3, um in allen Gehirnarealen die „Menge“ an Vorwissen zu quantifizieren, die in neuronaler Aktivität aufrechterhalten wird.

In Studie 2 („Information-theoretic evidence for predictive coding in the face-processing system“, publiziert in *Journal of Neuroscience*, 2017; n = 52) untersuche ich, wie Vorwissen über Gesichter in Abhängigkeit von Relevanz in unserem Gehirn aktiviert und für Vorhersagen genutzt wird. Die Anwendung des informationstheoretischen Verfahrens AIS auf MEG Quellenebene zeigt, dass aktiviertes Vorwissen über Gesichter als erhöhtes AIS in Arealen der Gesichtsverarbeitung sichtbar wird. Insbesondere im fusiformen Gesichtsareal (FFA) ist die AIS Erhöhung negativ mit den Reaktionszeiten der Teilnehmer korreliert, was auf eine Verhaltensrelevanz der Menge des aktivierten Vorwissens in FFA hindeutet. Zusätzlich ist in allen Gesichtsverarbeitungs-Arealen die Zunahme von AIS mit langsamer neuronaler Aktivität im Alpha und Beta Band assoziiert (ca. 8 bis 30 Hz) – übereinstimmend mit der Theorie von Bastos et al. (2012). Zuletzt demonstriert die Anwendung der informationstheoretischen Methode Transfer Entropy (TE, Schreiber, 2000; Vicente et al., 2011; Wibral et al., 2011), dass vor Stimulus Präsentation von dem hinteren Parietallappen (*posterior parietal cortex*) und von dem vorderen Bereich des unteren Schläfenlappens (*anterior inferior temporal lobe*) Informationen an FFA übertragen werden. Dieser Informationstransfer deutet daraufhin, dass aktiviertes Vorwissen über Gesichter zu FFA transportiert wird, um sich auf die Gesichtserkennung vorzubereiten.

In Studie 3 („Predictable information is reduced in autism spectrum disorder – a predictive coding study“, eingereicht bei *Human Brain Mapping*; n = 38) untersuche ich, ob Predictive Coding Mechanismen in Patienten mit Autismus-Spektrum-Störung verändert sind. Dazu vergleiche ich die Aktivierung von Vorwissen bei Patienten mit Autismus und gesunden Kontrollprobanden. Vergleichbar zu Studie 2, wird in Studie 3 das informationstheoretische Verfahren AIS verwendet, um auf MEG Quellenebene die Menge des aktivierten Vorwissens zu berechnen. Ich finde, dass AIS im Mittel bei Patienten mit Autismus reduziert ist. Insbesondere finde ich eine Reduktion von AIS bei Autismus Patienten im hinteren Bereich des Ruhezustandsnetzwerks (*default mode network*). Vergleichbar zu Studie 2, sind auch in Studie 3 die AIS Werte mit langsamer neuronaler Aktivität im Alpha und Beta Frequenzband korreliert. Zuletzt finde ich in der Patientengruppe, dass die Reduktion von AIS in Precuneus und Kleinhirn mit der Symptomschwere im Bereich restriktiver, repetitiver und stereotyper Verhaltensweisen korreliert. Die Resultate von Studie 3 legen Veränderungen der Predictive Coding Mechanismen bei Patienten mit Autismus nahe.

Zusammengefasst illustrieren die in dieser Arbeit präsentierten Ergebnisse die neuronalen Mechanismen, die dem Einfluss von Vorwissen zugrunde liegen. Sie liefern sowohl neurophysiologische Evidenzen für den Zusammenhang von hoch-frequenter neuronaler Aktivität und Vorhersagefehlern (Studie 1), als auch tiefer-frequenter neuronaler Aktivität und Vorhersagen/Vorwissen (Studie 2 und 3). Damit unterstützen die vorliegenden Studien die von Bastos und Kollegen vorgeschlagene neuronale Implementierung des Predictive Coding Prinzips im Gehirn (Bastos et al., 2012). Durch die Anwendung des AIS Verfahrens führt diese Doktorarbeit zudem eine neue informationstheoretische Methode ein, die in Zukunft erleichtern sollte, die Mechanismen der Predictive Coding Theorie nahezu ohne Vorannahmen zu untersuchen. So könnte in Zukunft das AIS Verfahren für ein breites Spektrum an mit Predictive Coding assoziierten Paradigmen sowie für die Untersuchung verschiedener neuropsychiatrischer Erkrankungen wie z.B. Schizophrenie angewendet werden.

8. Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.
Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

Darmstadt, den …………

…………………………...

9. Curriculum Vitae

Personal details:

Name:	Alla Brodski-Guerniero (née Brodski)
Date of Birth:	17.11.1987

Education:

Since 1.4.2013	**PhD Student (MEG Unit, Brain Imaging Center, Frankfurt University)** Frankfurt University / Technical University of Darmstadt (Supervisor Prof. M. Wibral / Prof. R. Galuske)
10.2010 – 10.2012	**Master of Science in Interdisciplinary Neuroscience with distinction**, Interdisciplinary Center for Neuroscience Frankfurt (ICNF), grade 1.0 Master's thesis at the MEG unit, Brain Imaging Center, Dept. of medical science, Frankfurt University
10.2007 – 07.2010	**Bachelor of Science in Biological Sciences** Frankfurt University, grade 1.1 Bachelor's thesis at the working group "Neurobiology and Biosensors", Dept. of biological sciences, Frankfurt University
09.1998 – 06.2007	**Abitur (equivalent to A level)**, 2007, grade 1.0; Ziehenschule – Gymnasium (academic high school), Frankfurt

Professional experience:

Since 1.03.2017	**Data scientist** at savedroid AG, Frankfurt
Since 1.11.2012	**Research assistant** at the MEG Unit, Brain Imaging Center, Dept. of medical science, Frankfurt University
10.2011 – 01.2012	**Internship in Cognitive Neuroscience laboratory,** Dept. of Psychology, Frankfurt University
Summer term 2011	**Student assistant** in the working group “Neurobiology and Biosensors“, Dept. of biological sciences, Frankfurt University

Publications:

-**A. Brodski,** G.-F. Paasch, S. Helbling, and M. Wibral, “The Faces of Predictive Coding”, *J. Neurosci.*, vol. 35, no. 24, pp. 8997–9006, Jun. 2015.

-V. Moliadze, E. Lyzhko, L. Böcher, **A. Brodski**, T. Gurashvili, C. M. Freitag, and M. Siniatchkin, “P46. Neuronal mechanisms of error monitoring in motivational context in healthy children and adolescents”, *Clinical Neurophysiology*, vol. 126, no. 8, p. e118, 2015.

-S. C. Reitz, S.-M. Hof, V. Fleischer, **A. Brodski,** A. Gröger, R.-M. Gracien, A. Droby, H. Steinmetz, U. Ziemann, F. Zipp, and others, “Multi-parametric quantitative MRI of normal appearing white matter in multiple sclerosis, and the effect of disease activity on T2”, *Brain imaging and behavior*, pp. 1–10, 2016.

-**A. Brodski-Guerniero,** G.-F. Paasch, P. Wollstadt, I. Özdemir, J. T. Lizier, und M. Wibral, “Information-theoretic evidence for predictive coding in the face-processing system“, *J Neurosci* .vol. 37, no. 34, pp. 8273-8283, Aug. 2017.

Scholarships and Awards:

1.4.2013 – 31.3.2016	PhD scholarship of the Ernst Ludwig Ehrlich foundation (ELES)
6.2014 and 6.2015	Best poster award at the Biennial Meeting of the Rhine-Main Neuroscience network (2014) and at the “Psychologie and Gehirn” meeting (2015)
10.2011 – 10.2012	scholarship of the Gerhard C. Starck foundation for highly talented students
1.1.2011	award for one of the best Bachelor’s degrees in Biological Sciences in 2010 at Frankfurt University

Conference Posters (selection):

1. **Alla Brodski**, Ipek Özdemir, Georg-Friedrich Paasch, Joseph Lizier, Michael Wibral (2015). *Information theory reveals neural correlates of predictions – a magnetoencephalography study.* 2015 Society for Neuroscience annual meeting, Chicago (USA)

2. Susanne Eisenhauer, **Alla Brodski**, Michael Wibral, Georg-Friedrich Paasch. *Internal models of face identities in predictive coding*. 2015 Psychologie und Gehirn Meeting, Frankfurt am Main (Germany)

3. **Alla Brodski**, Georg-Friedrich Paasch, Saskia Helbling, Michael Wibral (2014). *Neural correlates of prediction errors in visual perception*. 2014 Biennial Meeting of the Rhine-Main Neuroscience network, Oberwesel (Germany)