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

[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

Title:

Activating task relevant prior knowledge increases active information storage in content specific brain areas

Abstract (max 150 words)

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 for 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 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. Moreover, top-down transfer of predictions estimated by transfer entropy was associated with beta frequencies. Our results support accounts that activated prior knowledge and the corresponding predictions are signalled in low frequency activity (<30 Hz).

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).

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 at each level of the cortical hierarchy in order to continuously adapt and update 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 represented in neural activity in order to make this knowledge available to other parts of the brain (see Zipser et al., 1993 for a distinction of active and passive storage). Subsequently, (pre-)activated prior knowledge needs to be maintained until needed and to be constantly transferred as a prediction in top-down direction to a lower area of the cortical hierarchy, where it will be matched with the 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, only 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 recently a promising microcircuit theory of predictive coding has been put forward, suggesting internal predictions to be processed in deep cortical layers and to manifest in low frequency neural activity (<30 Hz) (Bastos et al., 2012). Accordingly this microcircuit theory also suggests that predictions are transferred via neural activity propagating along descending fiber systems in the same frequencies.

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 specific anatomic pathways to specific Granger causality signatures, in particular feedback connections to alpha/beta frequency channels in monkeys (Bastos et al., 2015) and humans (Michalareas et al., 2016)

Recently, the microcircuit theory of predictive coding gained experimental support by neurophysiological studies showing the predictability of events (i.e. precision of predictions) to be

66 associated with neural power in the alpha (Bauer et al., 2014; Sedley et al., 2016) or beta
67 frequency band (van Pelt et al., 2016)

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

82 The use of AIS and TE is based on the following rationale: Since the brain will usually not know
83 exactly when a prediction will be needed, it will maintain activated prior knowledge related to the
84 content of the prediction over time. If there is a reliable neural code that maps between content and
85 activity, maintained activated prior knowledge must be represented as maintained information
86 content in neural signals, measurable by AIS (Figure 1). Further, predictions based on prior
87 knowledge are supposed to be transferred to lower brain areas, where they can be matched with
88 incoming information. This information transfer is measurable with TE.
89 From this basic concept we can derive four testable hypotheses about AIS and TE in the predictive
90 coding framework:
91 1. When activated prior knowledge is maintained, predictable information as measured by AIS is
92 supposed to be high in brain areas specific to the content of the predictions.
93 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. 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.

4. 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.

These hypotheses were tested here on neural source activity reconstructed from magnetoencephalography (MEG) recordings of 52 human subjects during the performance of a two-tone (Mooney and Ferguson, 1951; Cavanagh, 1991) Face/House detection task (Figure 2). In this task subjects were instructed to detect faces during half of the experimental blocks (Face blocks) and houses during the other half (House blocks) in a stream of two-tone images containing faces, houses and scrambled versions thereof. These instructions were given to subjects at the beginning of each block of stimulus presentations to induce the pre-activation of face- or house-related prior knowledge, respectively. This task was designed 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).

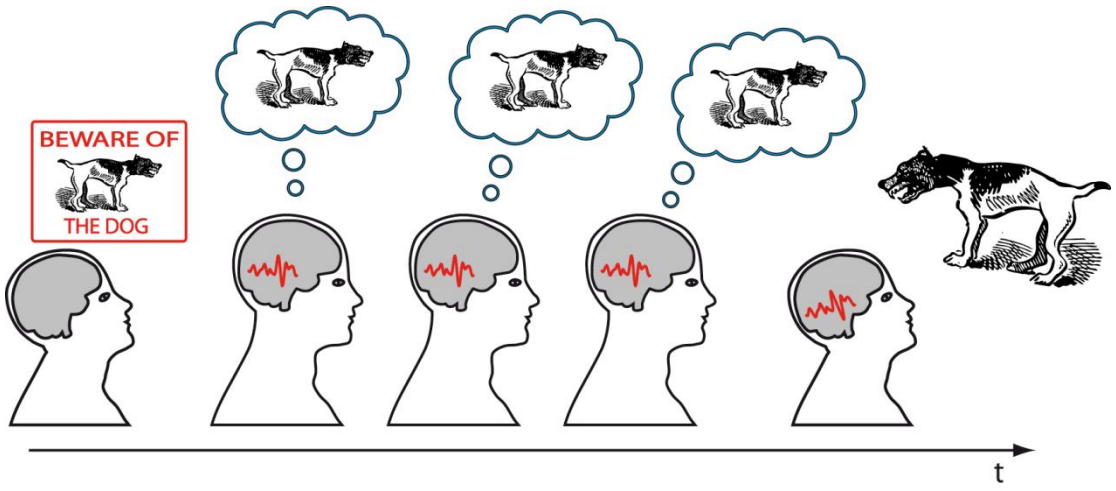


Figure 1. Central idea of the study. Typically, pre-activated prior knowledge related to the content of a prediction has to be maintained as the brain will not know exactly when it will be needed. If there is a reliable neural code that maps between content and activity, maintained activated prior knowledge 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). Figure elements obtained from OpenCliparts Library (http://www.opencliparts.org) and modified.

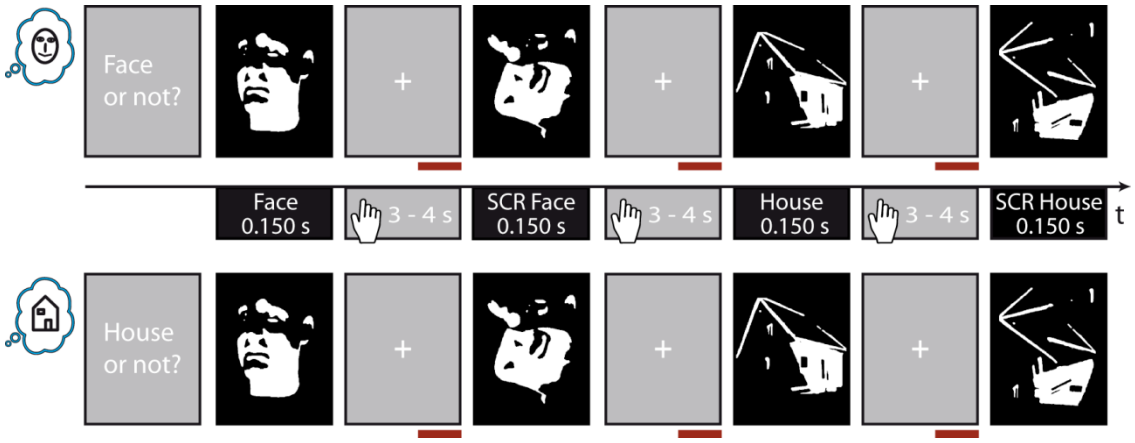


Figure 2. Experimental design. *Top* and *bottom*: Exemplary stimulus presentation in Face blocks *(top)* and in House blocks (*bottom*). Face and House icons on the left indicate Face and House blocks, respectively. *Middle***:** Depiction of stimulus categories and timing. The beginning of the response time window is indicated by the hand icon. Red horizontal bars mark the analysis interval. SCR - Scrambled Mooney stimuli, not representing a face or house.

Results

Behavioral results

We found no differences between Face blocks and House blocks for hitrates (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). Subjects showed equivalent behavioural patterns for both block types, for instance increased reaction times for the instructed stimulus conditions as these stimuli had to be distinguished from a similar distractor (SCR stimuli) (see Figure 3 for the analysis of behavioural differences between stimulus conditions within block types).

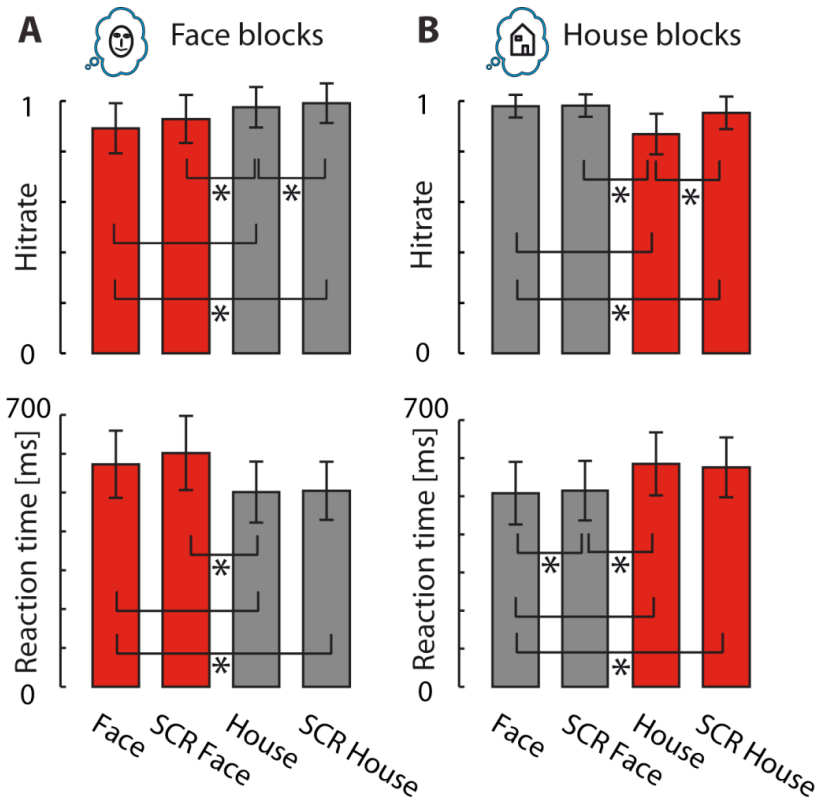


Figure 3. Behavioral results. Depiction of hitrates and reaction times of correct responses for (A) Face blocks and (B) House blocks. Equivalent conditions in different block types are marked in red and grey, respectively. Asterisks indicate significant differences based on Wilcoxon signed-rank tests within block type (n = 52; Bonferroni-Holms corrected for multiple comparisons). Error bars indicate standard deviation. SCR – scrambled Mooney stimuli.

Analysis of predictable information

To find the neural correlates of activated prior knowledge we focussed our analysis of the MEG recordings on the prestimulus interval up to 1 second before stimulus onset, as this time interval is not confounded by the responses to the visual stimuli. For this time range we used a LCMV beamformer to reconstruct the source time courses for the whole brain (1.5 cm grid spacing, resulting in 478 brain locations). Each source time course was then subjected to analysis of active information storage (AIS, Lizier et al., 2012), quantifying predictable information in the signals. Statistical comparisons of AIS values between Face blocks and House blocks 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 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.

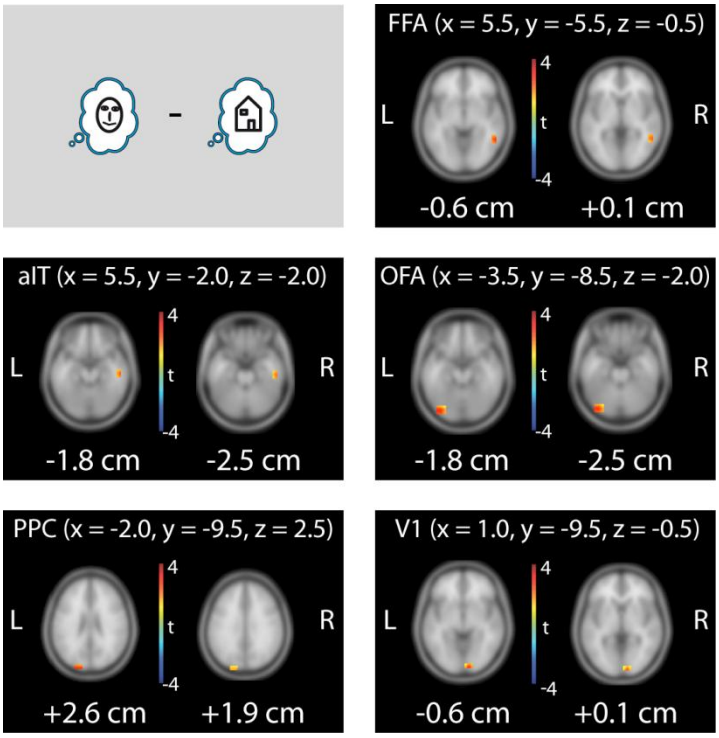


Figure 4. Statistical analysis of predictable information (measured by AIS) at the MEG source level. Results of whole-brain dependent samples permutation t-metric contrasting Face blocks and House blocks (n=52, t-values masked by $p<0.05$, cluster correction). Peak voxel coordinates in MNI space are 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

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. To this end, frequency bands were defined by a statistical comparison of MEG sensor-level activity in the task vs. baseline interval (jointly for Face and House Blocks, see Methods for details). Correlation analysis revealed a strong positive correlation in the alpha and beta frequency bands, only very small mostly positive correlations in the low and mid-gamma frequency bands and a small negative correlation for the high-gamma frequency band (Table 1). This means that alpha and beta band activity is the most likely carrier of activated prior knowledge.

While we found a significant correlation of single trial power and predictable information in the

alpha and beta band, the contrast map over all source grid points for Face and House blocks (t-values obtained from dependent sample t-metric over subjects) did not correlate for neither alpha nor beta power with the AIS map (alpha rho = 0.043, p = 0.33; beta rho = 0.05, p = 0.21). This suggests that AIS analysis provides additional information not directly provided by a spectral analysis. In sum, while AIS seems to be carried by alpha/beta-band activity, not all alpha/beta band activity contributes to AIS.

Table 1: Correlation of single trial power and single trial predictable information (measured by AIS) in the face prediction network

	OFA	**aIT**	**PPC**	**V1**	**FFA**
8-14 Hz (alpha)	***rho = 0.58*** *p < 0.001*	***rho = 0.54*** *p < 0.001*	***rho = 0.59*** *p < 0.001*	***rho = 0.61*** *p < 0.001*	**rho = 0.58** *p < 0.001*
14-32 Hz (beta)	***rho = 0.53*** *p < 0.001*	***rho = 0.52*** *p < 0.001*	***rho = 0.53*** *p <0.001*	***rho = 0.52*** *p < 0.001*	***rho = 0.54*** *p < 0.001*
32-56 Hz (low gamma)	rho = 0.13 *p < 0.001*	rho = 0.10 *p < 0.001*	rho = 0.17 *p < 0.001*	rho = 0.14 *p < 0.001*	rho = 0.13 *p < 0.001*
56-64 Hz (Mid-gamma)	rho = 0.06 *p < 0.001*	rho = -0.004 *p=0.22*	rho = 0.07 *p < 0.001*	rho = 0.07 *p < 0.001*	rho = 0.02 *p < 0.001*
64-150 Hz (High-gamma)	rho = -0.07 *p < 0.001*	rho = -0.14 *p <0.001*	rho = -0.04 *p < 0.001*	rho =-0.04 *p <0.001*	rho=-0.09 *p < 0 .001*

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, fdr correction) and from PPC to FFA (p = 0.0014, fdr correction). For House blocks information transfer was increased in comparison to Face blocks from brain area V1 to PPC (p=0.0014, fdr correction) (Figure 5).

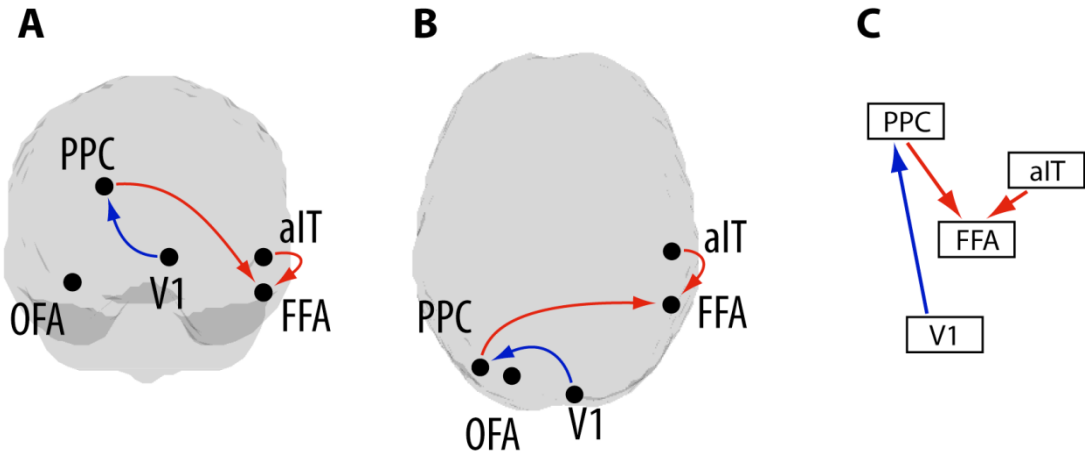


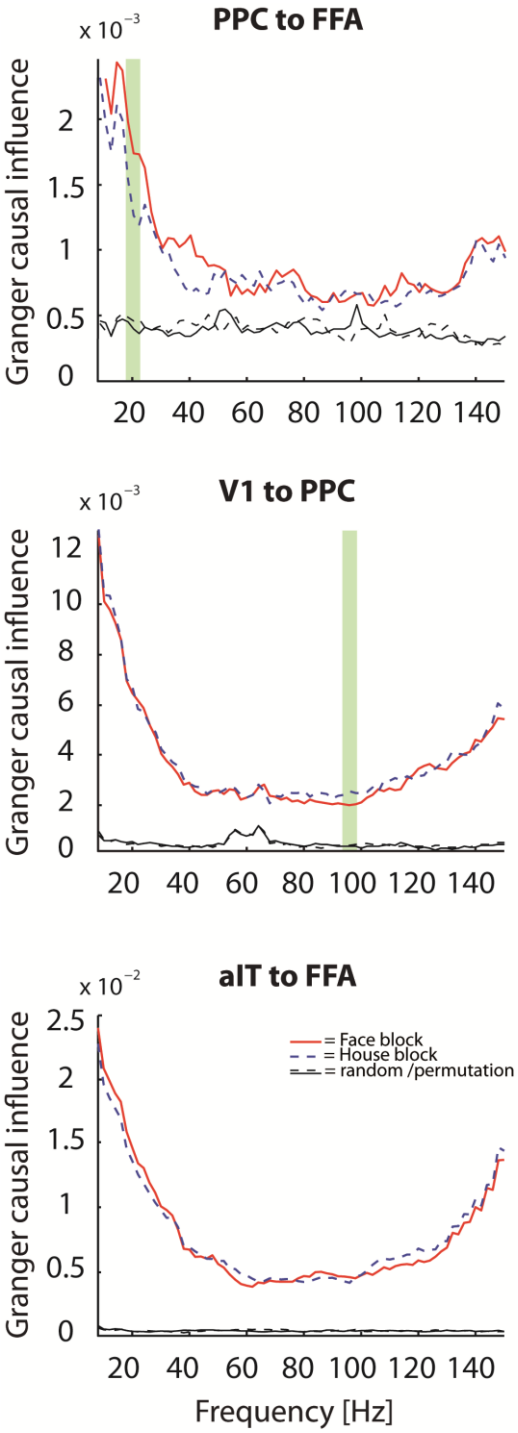
Figure 5. Analysis of information transfer in the prestimulus interval. Results of dependent sample permutation t-tests on transfer entropy (TE) values (Face blocks vs House blocks, n = 52, $p < 0.05$, fdr corrected). Red arrows indicate increased information transfer for Face blocks, blue arrows indicate 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).

Post-hoc frequency resolved Granger causality

In order to investigate whether information transfer differences in Face and House blocks were reflected in specific frequency bands, we post-hoc performed a non-parametric spectral Granger causality analysis on the three links identified with transfer entropy analysis. For the link from PPC to FFA we found stronger Granger causality for Face blocks than House blocks in a cluster between 18 and 22 Hz (Figure 6, p=0.045, cluster correction for frequencies, uncorrected for the number of links in this post hoc test). The link from V1 to PPC showed a stronger Granger causal influence for House blocks than Face blocks between 94 and 98 Hz (Figure 6, p=0.042, cluster correction for frequencies, uncorrected for the number of links in this post hoc test). Using cluster correction, the link from aIT to FFA did not show significant differences in Granger causal influence.

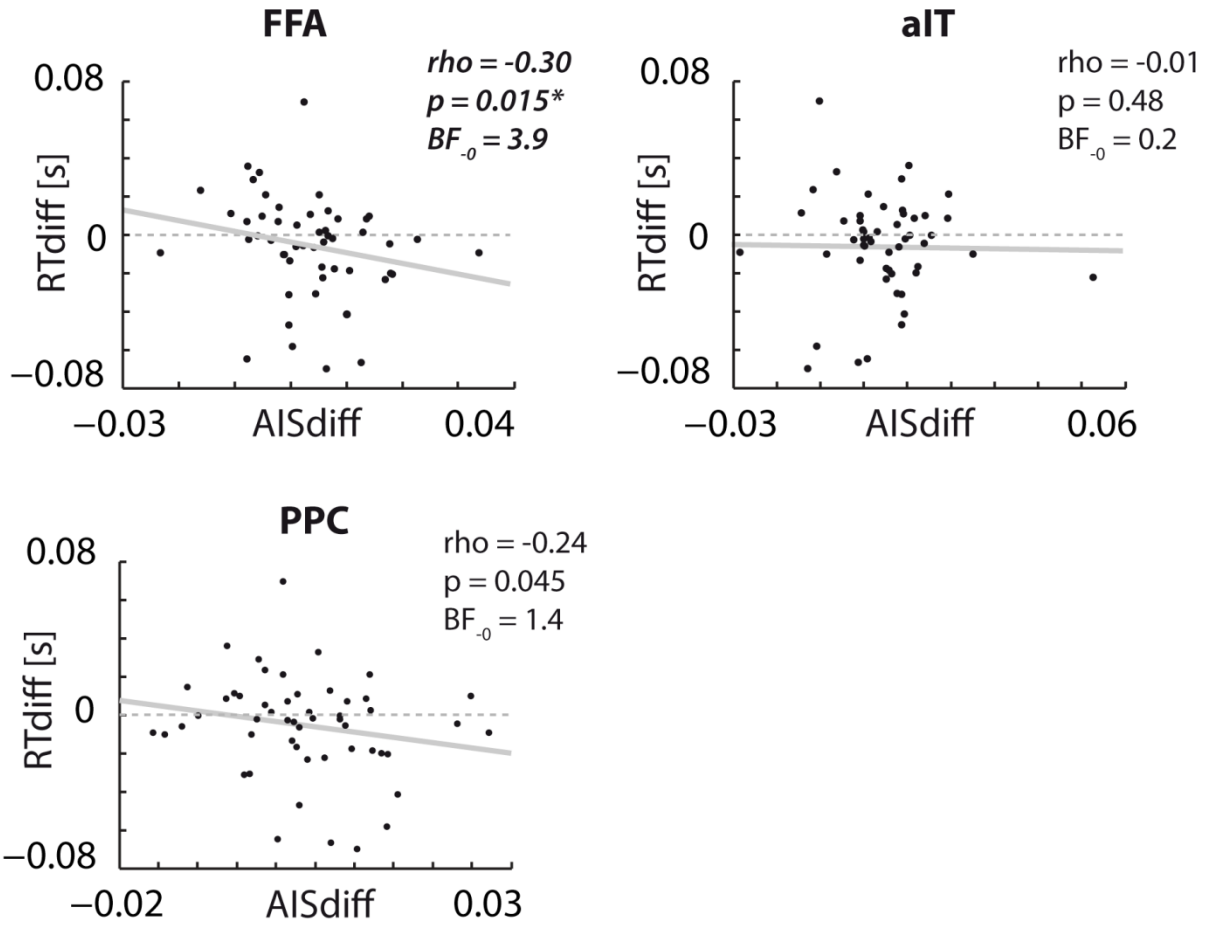
Correlation of predictable information and reaction times

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 higher predictable information for face blocks was associated with faster performance, i.e. decreased reaction times during Face blocks. Indeed, negative correlation values were found for all of the three brain areas (Figure 7). However, only the correlation for FFA reached significance when correcting for the number of comparisons (rho= 0.30, p = 0.015; 0.015 x 3 <0.05). In order to be able to interpret also the non-significant results we additionally report the Bayes factor (BF_{-0}) for the three correlations. We find a BF_{-0} of 3.89 for brain area FFA, a BF_{-0} of 0.213 for aIT and a BF_{-0} of 1.41 for PPC. As Bayes factors above three can be considered reliable evidence for the alternative hypothesis (here: negative correlation), while Bayes factors below 1/3 can be considered reliable evidence for the null hypothesis (here: no or positive correlation) (Jeffreys, 1998), our findings indicate that predictable information was behaviourally relevant in area FFA but not in area aIT. It remains inconclusive whether predictable information in brain area PPC was behaviourally relevant.



231

232 **Figure 6. Frequency resolved Granger causality – post-hoc test for TE analysis.** Granger
233 causality for Face blocks and House blocks as well as random/permutation conditions. Green
234 shaded regions indicate significant differences between Face and House blocks with cluster
235 correction (dependent samples permutation t-test, n = 52, $p<0.05$). Frequency ranges were only
236 considered, if granger values for both block types in these frequencies also significantly differed
237 from the random conditions (dependent samples permutation t-tests, n = 52, Bonferroni-Holms
238 correction).



239

240 **Figure 7. Correlation analysis for predictable information (measured by AIS) and reaction**
241 **times.** Scatter plots displaying the correlation of per subject AIS difference values (Face blocks –
242 House blocks) with per subject reaction time difference values (Face blocks – House blocks).
243 Spearman correlation values, p-values and Bayes factors (BF_{-0}) are indicated at the top right
244 corner of each plot (*n* = 52). Asterisks indicates significant correlation, using Bonferroni correction
245 for multiple comparisons (p x 3 <0.05). Linear regression lines are included in gray (solid).

246 ***Discussion***

247 Here we tested the hypothesis that the neural correlates of prior knowledge that has been
248 activated for use as an internal prediction must show as predictable information in the neural
249 signals carrying that activated prior knowledge. This hypothesis is based on the rationale that the
250 content of activated prior knowledge must be maintained until the knowledge or the prediction
251 derived from it is used. The fact that activated prior knowledge has a specific content then
252 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 in intervals preceding stimulus presentation in a face detection task. In this task predictable information was selectively enhanced in a network of brain areas known for their role in face processing, and 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 which suggest 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 also in alpha/beta-band frequencies (Bastos et al., 2012). Indeed, within the network of brain areas related to activated prior knowledge of faces, information transfer was increased in top-down direction and related to Granger-causality in the beta band – in accordance with the theory. In the following we will detail the exact interpretation of increases in predictable information and discuss our findings in relation to specific predictions of predictive coding theory in more detail.

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) in the Methods section). 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. This manipulation should increase AIS in content-specific brain areas, and this is indeed what we found. Hence, in the study of predictive coding theory, the observation of increased AIS must be seen as a necessary consequence of the maintenance of activated prior knowledge. Yet, the observation of increased AIS is not sufficient to indicate activated prior knowledge – any process requiring some form of information to be stored in neural activity (e.g. working memory) must lead to an observation of increased AIS when analysing the

relevant signal. Nevertheless, AIS is very useful tool to separate activated prior knowledge and predictions from other processes in predictive coding theory - especially from (unpredictable) error processing.

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

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 as well as 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). Hence, increased predictable information for Face blocks in these brain areas supports the hypothesis of prior knowledge being activated in content specific brain areas.

In addition to increased predictable information in the 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 posterior parietal cortex have been recently linked to high level visual processing of objects like faces (Pashkam and Xu, 2014) and an activation in the posterior parietal cortex has been repeatedly observed during the recognition of Mooney faces (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.

These findings are in agreement with suggestions that content specific brain areas are in general activated in anticipation of a specific stimulus (Stokes et al., 2009; Peelen and Kastner, 2011; Kok et al., 2014) and that the FFA in particular is activated in anticipation of faces (Esterman and Yantis, 2009; Puri et al., 2009; Egner et al., 2010). Most importantly however, our results add to

307 previous findings by demonstrating that anticipatory activity in face specific areas is indeed related
308 to pre-activation and extended maintenance of content specific prior knowledge, reflected by the
309 predictable character of the signal itself, as measurable with AIS.

310 Note that an increase in predictable information (AIS) is not necessarily predicted by accounting for
311 preparatory processing as attentional feature/object selection in the sense of a subthreshold gain
312 control mechanism (e.g. Corbetta et al., 1990). This is because gain control does not necessarily
313 imply that any additional preparatory activity evoked by increased gains is temporally self-
314 predicting. If our observation of increased predictable information is indeed the consequence of an
315 attentional mechanism, e.g. altered baseline firing rates in content specific neurons (Desimone
316 and Duncan, 1995, see also Hillyard et al., 1998), then our results would imply that this additional
317 activity must highly patterned and predictable in the temporal domain. In other words, in this latter
318 case the boundary between attention-related activity changes that increase information storage
319 and a pre-activation of specific content in the sense of predictive coding theories may become
320 blurred. Hence, a unified information theoretic description in terms of predictable information,
321 measured by AIS, as an algorithmic quantity in the sense of Marr (1982) may serve our
322 understanding better.

323 Our results also complement previous findings by demonstrating that prior knowledge about faces
324 is pre-activated not only in FFA but in an extended face processing network. Since it has been
325 suggested that e.g. FFA, OFA and aIT code for different aspects of face processing (e.g. Haxby et
326 al., 2000; Pitcher et al., 2011), our results might indicate that different kinds of prior knowledge
327 about faces, e.g. related to the general face configuration, but also to facial features or familiar
328 identities is pre-activated in face detection tasks.

329 However, while we found increased predictable information in content specific areas for Face
330 blocks, we did not find brain areas showing increased predictable information for House blocks.
331 This is similar to other highly cited studies that failed to find prediction effects for houses in the
332 brain in contrast to faces (e.g. Summerfield et al., 2006a, 2006b; Trapp et al., 2015). For instance,
333 in a face/house discrimination task Summerfield and colleagues (2006b) observed increased

activation in FFA, when a house was misperceived as a face. However, they 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. Alternatively, they suggest, subjects may have employed a face prediction strategy by matching all input to face templates and judging a "No-face" to be a house. However, this explanation seems unlikely to apply in our design as a separate No-face (SCR) condition existed. Hence, in contrast to the design of Summerfield and colleagues, a No-face did not necessarily have to be a house in our design. Nevertheless, because of their strong social relevance (e.g. Farah et al., 1995) faces disproportionally capture attention (e.g. Vuilleumier and Schwartz, 2001). Thus, also face predictions/templates may be prioritized in comparison to other templates e.g. for houses. In general, it is possible that perceptual templates are more in use for expert categories like faces, for which holistic processing can be exploited (e.g. Esterman and Yantis, 2009; Puri et al., 2009; Van Belle et al., 2010).

In contrast to the absence of brain areas showing increased predictable information for house blocks in our study, in some fMRI studies researchers were able to observe an anticipatory effect for houses in PPA (Esterman and Yantis, 2009; Puri et al., 2009), indicating that prediction effects for houses are in general possible. Thus, it is worth considering that the house-responsive region PPA may not be well detectable with MEG, as it is a deep-lying source of activity. In fact, with the exception of a recent MEG/fMRI study of Baldauf and Desimone (2014) studying selective attention to faces or houses, to our knowledge, no MEG study was able to find activation in PPA for houses. Yet, even for the study of Baldauf and Desimone (2014) the brain locations showing an increased MEG activation when houses were attended noticeably differed from the PPA location obtained by fMRI.

Last, we note that the effect of increased AIS for Face blocks in content specific areas but not for House blocks cannot be simply explained by differential bottom-up information as the same Mooney images were presented in Face and House blocks and only the prestimulus interval was

analyzed.

In addition to elevated AIS in content specific areas for face blocks, we also found increased AIS for Face blocks in V1. This finding was surprising, as we did not expect prediction effects for faces to show up as "early" in the processing hierarchy as primary visual areas. Nevertheless, preparation effects in V1 have been previously observed (e.g Smith and Muckli, 2010; Peelen and Kastner, 2011; Kok et al., 2014) and may indicate that subjects also prepared to detect faces by low level features in our task. However, anticipation of low level features is not expected to be engaged more strongly for face than house preparation. Consequently, one may speculate that the house prediction was not used as consistently as the face prediction in our task - in line with our inability to find brain areas associated with increased AIS values for House blocks.

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

We found a strong positive single-trial correlation of AIS with alpha/beta power for all face prediction areas. This finding supports the assumption that especially the maintenance of activated prior knowledge as indexed by AIS is related to alpha and beta frequencies.

Previously, both alpha as well as beta frequencies have been linked with maintaining items in working memory (Jensen et al., 2002; Busch and Herrmann, 2003; Buschman and Miller, 2007; Kaiser et al., 2007) but have also been associated with more general functions related to maintenance of information in the brain like mental imagery (alpha, Cooper et al., 2006), implementation of an attentional set (beta, Bressler and Richter, 2015) and maintenance of the current cognitive state (beta, Engel and Fries, 2010)

More specifically and 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 the alpha frequency range 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 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. Face predictions are transferred in a top-down manner and via beta frequencies

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.

Both brain areas, PPC as well as aIT have been associated with memory processing (Erickson and Desimone, 1999; Sakai and Miyashita, 1991; Wagner et al., 2005) and in particular with memory related to faces or face templates (Dolan et al., 1997; Barton, 2008) making them plausible candidates for top-down preparatory information transfer to FFA. 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 a preparatory interval 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. However, the concrete role of the top-down signals from PPC as well as aIT or the prediction “content” being transferred to FFA is unknown so far. One possibility would be that PPC, well known to be active for familiar stimuli (e.g. Wagner et al., 2005), may signal a set of faces with typical properties (i.e. feature combinations from the typical set). aIT, highly active for individual well known faces (e.g. Haxby et al., 2000), may signal the individually most probable face or faces to FFA in order to facilitate face detection. Thus, both serve predictions of high

statistical validity yet for different purposes (face detection vs. identification).

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.

Information transfer in top-down direction was associated with Granger causality in the beta frequency band (PPC to FFA), while information transfer in bottom-up direction was associated with Granger causality in the high gamma frequency band (V1 to PPC).

The association of top-down information transfer with beta frequencies and bottom-up information transfer with gamma frequencies is in line with recent physiological findings in monkeys and humans (Bastos et al., 2015; Michalareas et al., 2016) and has been linked with predictive coding in Bastos' microcircuit model (Bastos et al., 2012), resulting in the hypothesis of predictions being transferred top-down via low frequency channels and prediction errors bottom-up via high frequency channels. In accord with this hypothesis, our group has recently shown that prediction errors are communicated in the high frequency gamma band (Brodski et al., 2015). Our present finding of top-down information transfer in low beta frequencies during anticipation of faces adds support to the microcircuit model hypothesis of a low frequency channel for the top-down propagation of predictions (Bastos et al., 2012). In particular, low beta frequencies (<22 Hz) as observed in our experiment have been proposed as the main rhythm for inter-areal top-down processing, possibly conveying behavioural context like predictions or attentional set to lower-level sensory neurons (Bressler and Richter, 2015).

In line with our findings, the spectral dissociation between the transfer of predictions and of prediction errors recently received additional support from a MEG study applying Granger causality analysis for the investigation of information transfer during the prediction of causal events (van Pelt

et al., 2016). Van Pelt and colleagues recorded MEG activity while subjects watched bowling action animations with more or less predictable combinations of throwing direction and outcomes. Van Pelts' Granger causality findings suggest that top-down transfer of predictions from frontal to parietal to temporal regions is dominated by the beta-band while bottom up prediction error transfer from temporal to parietal to frontal regions is dominated by neural activity in the gamma frequency band.

It should be noted that van Pelt and colleagues defined their network of interest for Granger causality analysis based on the prior assumption of the involvement of these brain areas in causal inference, suggesting that predictions and prediction errors related to causal events should be transferred within that network. In contrast, a Granger causality analysis based on a network defined with AIS as performed in our study allows finding the brain areas involved in predictive processing without relying on prior assumptions about their function and thus the specific predictions transferred from specific brain areas. This makes AIS analysis applicable for a large variety of predictive coding designs, independent of the fact whether prediction areas can be defined in advance or not.

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. Inter-individual differences between subjects in the quantity of activated prior knowledge in FFA and thus also performance may be related to the differential ability in maintaining an object specific representation (see Ranganath et al., 2004).

Conclusion: Pre-activation of task relevant prior knowledge for predictions shows as alpha/beta related active information storage in content specific areas and is behaviourally relevant. Top-down prediction transfer from content specific areas is associated with neural activity in the low beta frequency band (18-22 Hz).

Methods

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 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 and colleagues (2013).

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).

In order to increase task difficulty additionally 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 2.

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 was used to display the stimuli at the center of a translucent screen (background set to gray, 145 cd/m²). 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).

Task and Instructions

Subjects performed a detection task for faces or houses (Figure 2). 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 House block). 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, order of blocks was reversed for half of the subjects.

Subject 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' block 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 a feedback whether their response was correct or not. No feedback was provided during the actual task.

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).

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 hitrates as well as reaction times. To account for multiple testing, sequential Bonferroni-Holm correction (Holm, 1979) was applied (uncorrected alpha = 0.05).

MEG-data preprocessing

MEG Data analysis was performed with the open source Matlab toolbox Fieldtrip (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).

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, before spectral analysis new data segments were cut from -0.55 to 0.55 s around stimulus onset. For spectral analysis we used a multitaper approach (Percival and Walden, 1993) based on Slepian sequences (Slepian, 1978). The spectral transformation was applied in an interval from 4 to 150 Hz in 2 Hz steps in time steps of 0.01 s and using two slepian tapers for each frequency. For each subject, time-frequency representations were averaged for Face blocks and House blocks as well as within the time interval of “baseline” (-0.35 s – 0.05 s) and “task” (0.05 s – 0.35 s), respectively. Average spectra of task and baseline period 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.

Following the same approach as Brodski and colleagues (2015) based on the significant clusters of the task vs. baseline statistics five frequency bands were defined for further analysis: (1) 8–14 Hz (alpha); (2) 14–32 Hz (beta); (3) 32–56 Hz (low gamma); (4) 56–64 Hz (mid gamma) and (5) 64–150 Hz (high gamma) (Figure 8).

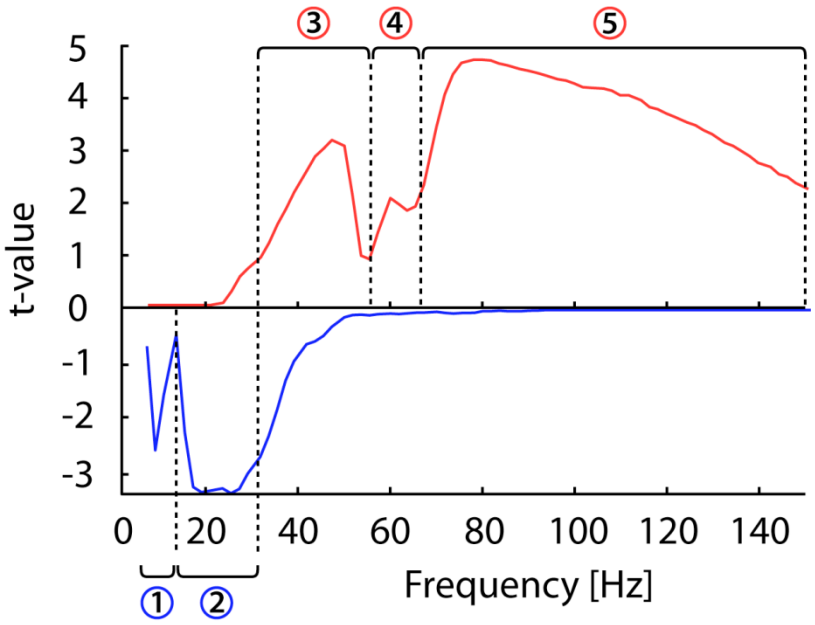


Figure 8. Sensor-level frequency analysis - defining frequency bands. Power spectra for all of the significant clusters (one positive and one negative cluster) at the sensor level (permutation t-metric, contrast [0.05s 0.35s] vs [-0.35s -0.05s] around stimulus onset, *t* values masked by $p < 0.05$, 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.

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 15mm 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.

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 (LCMV, linear constrained minimum variance; (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).

Definition of active information storage

We assume that the reconstructed source time courses for each brain location can be treated as realizations $\{x_1, \dots, x_t, \dots, x_N\}$ of a random process $X = \{X_1, \dots, X_t, \dots, X_N\}$, which consists of a collection of random variables, X_t, ordered by some integer t. AIS then describes how much of the information 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

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

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 $X_{t-1}^k = \left\{X_{t-1}, \dots, X_{t-1-(k*\tau)}\right\}$, 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 $t - k_{\max} * \tau$ does not further improve (statistically) the prediction of X_t from its past (Lizier et al., 2012).

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).

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.

Correlation analysis

We investigated the relationship of spectral power in the prestimulus interval and AIS values on the single trial level. Before calculation of the correlation coefficient, single trial spectral power in each of the predefined frequency bands and single trial AIS values were z-normalized for each subject. These values were appended for Face and House blocks, pooled over all subjects and Spearman's rho was calculated. Then, trials were shuffled 1000 times for spectral power and AIS values separately within each subject and correlation analysis was repeated for each randomization. Original correlation values larger than the 95^{th} percentile of the distribution of correlation values in the shuffled data were considered as significant. This statistical procedure conforms to a permutation test of the correlation where permutations are restricted within the levels of the factor subjects.

We also calculated the correlation of t-values on AIS (based on the dependent sample t-metric, contrast Face blocks vs. House blocks) for all grid points at the source level with the t-values obtained from the same contrast based on beamformer reconstructed source power in the alpha (8-14 Hz) and beta (14-32 Hz) frequency band. Source power was reconstructed with the DICS

693 (dynamic imaging of coherent sources, Gross et al., 2001) algorithm as implemented in the
694 FieldTrip toolbox using real values filter coefficients only (see also Grützner et al., 2010).
695 Last, we accessed the relationship of AIS values and reaction times for each subject. To this end
696 for each subject mean reaction times and mean AIS values in the brain areas of interest for Face
697 and House blocks were subtracted from each other. This allowed accounting for differential
698 behavioral speed between subjects. The correlation of the difference in AIS values and the
699 difference in reaction times was calculated using Spearman's rho and a one sided test for the
700 hypothesis that the correlation was negative, i.e. that higher AIS values in face blocks were
701 associated with faster performance. Significance was accessed using a permutation distribution:
702 AIS differences were shuffled between subjects 5000 times independent of differences in reaction
703 times and Spearman correlation was recomputed. The p-values were calculated as the proportion
704 of permutations in which a more extreme (more negative) correlation value than the original
705 correlation value was observed.
706 In order to facilitate interpretation of the results of the correlation of AIS and reaction times, we
707 further report Bayes factors (BF) for the correlations. For calculation of the BFs we used the
708 software JASP (JASP TEAM, 2016), applying a Bayesian correlation using Kendall's tau (van
709 Doorn et al., 2016) using JASP standard parameters for the distributions resulting in uniform prior
710 probabilities for H1 and H0. The Bayes factor represents the weight of evidence of one hypothesis
711 over another (Jeffreys, 1998). Bayes factors above 3 and Bayes factors below 1/3 indicate reliable
712 evidence for H1 or H0, respectively (Jeffreys, 1998). For equal priors of one-half for the null and
713 the alternative hypothesis a BF of 3 indicates that the posterior odds are 3:1 in favor of the
714 alternative hypothesis, i.e. that the alternative hypothesis is three times more probable than the null
715 hypothesis given the data and the prior probabilities of both hypotheses. We report the Bayes
716 factor (BF_{10} = BF_{-0}) for the hypothesis H1 (or H-) that AIS differences are negatively associated
717 with reaction time differences, compared to the null hypothesis H0 that AIS is uncorrelated or
718 positively correlated with reaction times.

719

720

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 Wieners 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. Vicente et al., 2011) and was previously applied to broadband MEG source data (Wibral et al., 2011). TE is defined as a conditional mutual information

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

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

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.

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, FDR 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 difference 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

776 included frequencies also reached significance in the comparison with the random conditions using
777 a Bonferroni-Holms correction to account for multiple comparisons.
778

References

Baldauf, D., Desimone, R., 2014. Neural Mechanisms of Object-Based Attention. Science 344, 424–427. doi:10.1126/science.1247003

Barton, J.J., 2008. Structure and function in acquired prosopagnosia: lessons from a series of 10 patients with brain damage. J. Neuropsychol. 2, 197–225.

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., Friston, K.J., 2012. Canonical microcircuits for predictive coding. Neuron 76, 695–711.

Bastos, A.M., Vezoli, J., Bosman, C.A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J.R., 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, K.J., Dolan, R.J., 2014. Attentional Modulation of Alpha/Beta and Gamma Oscillations Reflect Functionally Distinct Processes. J. Neurosci. 34, 16117–16125.

Bressler, S.L., Richter, C.G., 2015. Interareal oscillatory synchronization in top-down neocortical processing. Curr. Opin. Neurobiol. 31, 62–66.

Brodski, A., Paasch, G.-F., Helbling, S., Wibral, M., 2015. The Faces of Predictive Coding. J. Neurosci. 35, 8997–9006. doi:10.1523/JNEUROSCI.1529-14.2015

Buffalo, E.A., Fries, P., Landman, R., Buschman, T.J., Desimone, R., 2011. Laminar differences in gamma and alpha coherence in the ventral stream. Proc. Natl. Acad. Sci. 108, 11262–11267.

Busch, N.A., Herrmann, C.S., 2003. Object-load and feature-load modulate EEG in a short-term memory task. Neuroreport 14, 1721–1724.

Buschman, T.J., Miller, E.K., 2007. Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. Science 315, 1860–1862. doi:10.1126/science.1138071

Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci.

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

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

Cooper, N.R., Burgess, A.P., Croft, R.J., Gruzelier, J.H., 2006. Investigating evoked and induced electroencephalogram activity in task-related alpha power increases during an internally directed attention task. Neuroreport 17, 205–208.

Corbetta, M., Miezin, F.M., Dobmeyer, S., Shulman, G.L., Petersen, S.E., 1990. Attentional modulation of neural processing of shape, color, and velocity in humans. Science 248, 1556–1559. doi:10.1126/science.2360050

Desimone, R., Duncan, J., 1995. Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. 18, 193–222.

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

Dolan, R.J., Fink, G.R., Rolls, E., Booth, M., Holmes, A., Frackowiak, R.S.J., Friston, K.J., 1997. How the brain learns to see objects and faces in an impoverished context. Nature 389, 596–598.

Egner, T., Monti, J.M., Summerfield, C., 2010. Expectation and surprise determine neural population responses in the ventral visual stream. J. Neurosci. 30, 16601–16608.

Engel, A.K., Fries, P., 2010. Beta-band oscillations—signalling the status quo? Curr. Opin. Neurobiol. 20, 156–165.

Erickson, C.A., Desimone, R., 1999. Responses of macaque perirhinal neurons during and after visual stimulus association learning. J. Neurosci. 19, 10404–10416.

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, M.J., Tanaka, J.W., Drain, H.M., 1995. What causes the face inversion effect? J. Exp. Psychol. Hum. Percept. Perform. 21, 628–634. doi:10.1037/0096-1523.21.3.628

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

Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11, 127–138. doi:10.1038/nrn2787

Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. B Biol. Sci. 360, 815–836. doi:10.1098/rstb.2005.1622

George, D., Hawkins, J., 2009. Towards a Mathematical Theory of Cortical Micro-circuits. PLoS Comput Biol 5, e1000532. doi:10.1371/journal.pcbi.1000532

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

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

Granger, C.W.J., 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica 37, 424–438. doi:10.2307/1912791

Gross, J., Baillet, S., Barnes, G.R., Henson, R.N., 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. doi:10.1073/pnas.98.2.694

Grützner, C., Uhlhaas, P.J., Genc, E., Kohler, A., Singer, W., Wibral, M., 2010. Neuroelectromagnetic Correlates of Perceptual Closure Processes. J. Neurosci. 30, 8342–8352. doi:10.1523/JNEUROSCI.5434-09.2010

Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. Trends Cogn. Sci. 4, 223–233. doi:10.1016/S1364-6613(00)01482-0

Hillyard, S.A., Vogel, E.K., Luck, S.J., 1998. Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. Philos. Trans. R. Soc. Lond. B Biol. Sci. 353, 1257–1270.

Hohwy, J., 2013. The Predictive Mind. Oxford University Press.

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

Huang, Y., Rao, R.P., 2011. Predictive coding. Wiley Interdiscip. Rev. Cogn. Sci. 2, 580–593.

Jeffreys, H., 1998. The theory of probability. OUP Oxford.

Jensen, O., Gelfand, J., Kounios, J., Lisman, J.E., 2002. Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. Cereb. Cortex 12, 877–882.

Kaiser, J., Heidegger, T., Wibral, M., Altmann, C.F., Lutzenberger, W., 2007. Alpha synchronization during auditory spatial short-term memory. Neuroreport 18, 1129–1132.

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

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.

Kok, P., Failing, M.F., de Lange, F.P., 2014. Prior expectations evoke stimulus templates in the primary visual cortex. J. Cogn. Neurosci. 26, 1546–1554.

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.

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. doi:10.1186/1471-2202-12-119

Lizier, J.T., 2014. JIDT: an information-theoretic toolkit for studying the dynamics of complex systems. Comput. Intell. 1, 11. doi:10.3389/frobt.2014.00011

Lizier, J.T., Prokopenko, M., Zomaya, A.Y., 2012. Local measures of information storage in complex distributed computation. Inf. Sci. 208, 39–54. doi:10.1016/j.ins.2012.04.016

Makeig, S., Bell, A.J., Jung, T.-P., Sejnowski, T.J., 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. doi:10.1016/j.jneumeth.2007.03.024

Marr, D., 1982. Vision San Francisco. W H Freeman Co.

Mayer, A., Schwiedrzik, C.M., 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.

Mooney, C.M., 1957. Age in the development of closure ability in children. Can. J. Psychol. Can. Psychol. 11, 219–226. doi:10.1037/h0083717

Mooney, C.M., Ferguson, G.A., 1951. A new closure test. Can. J. Psychol. Can. Psychol. 5, 129–133. doi:10.1037/h0083540

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

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.

Oldfield, R.C., 1971. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9, 97–113. doi:10.1016/0028-3932(71)90067-4

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, M.V., Xu, Y., 2014. Decoding visual object representation in human parietal cortex. J. Vis. 14, 1307–1307.

Peelen, M.V., Kastner, S., 2011. A neural basis for real-world visual search in human occipitotemporal cortex. Proc. Natl. Acad. Sci. 108, 12125–12130.

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. doi:10.1093/scan/nsw017

Percival, D.B., Walden, A.T., 1993. Spectral Analysis for Physical Applications. Cambridge University Press.

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.

Puri, A.M., 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. doi:10.1103/PhysRevE.65.056201

Ranganath, C., Cohen, M.X., 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. doi:10.1523/JNEUROSCI.5053-03.2004

Rao, R.P.N., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. 2, 79–87.

Sakai, K., Miyashita, Y., 1991. Neural organization for the long-term memory of paired associates. Nature 354, 152–155. doi:10.1038/354152a0

Schreiber, T., 2000. Measuring information transfer. Phys. Rev. Lett. 85, 461.

Sedley, W., Gander, P.E., Kumar, S., Kovach, C.K., Oya, H., Kawasaki, H., Iii, M.A.H., Griffiths, T.D., 2016. Neural signatures of perceptual inference. eLife 5, e11476. doi:10.7554/eLife.11476

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

Smith, F.W., Muckli, L., 2010. Nonstimulated early visual areas carry information about surrounding context. Proc. Natl. Acad. Sci. 107, 20099–20103. doi:10.1073/pnas.1000233107

Stokes, M., Thompson, R., Cusack, R., Duncan, J., 2009. Top-down activation of shape-specific population codes in visual cortex during mental imagery. J. Neurosci. 29, 1565–1572.

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. doi:10.1126/science.1132028

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. doi:10.1093/cercor/bhi129

JASP Team, 2016. JASP (Version 0.8. 0.0). [Computer Software]

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

Tsao, D.Y., Moeller, S., Freiwald, W.A., 2008. Comparing face patch systems in macaques and humans. Proc. Natl. Acad. Sci. 105, 19514–19519.

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 Doorn, J., Ly, A., Marsman, M., Wagenmakers E-J., in press. Bayesian Inference for Kendall's Rank Correlation Coefficient. Am. Stat.

Van Veen, B.D., 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. doi:10.1109/10.623056

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

969 Vuilleumier, P., Schwartz, S., 2001. Emotional facial expressions capture attention. Neurology 56,
970 153–158.

971 Wagner, A.D., Shannon, B.J., Kahn, I., Buckner, R.L., 2005. Parietal lobe contributions to episodic
972 memory retrieval. Trends Cogn. Sci. 9, 445–453.

973 Wibral, M., Lizier, J.T., Vögler, S., Priesemann, V., Galuske, R., 2014. Local active information
974 storage as a tool to understand distributed neural information processing. Front.
975 Neuroinformatics 8. doi:10.3389/fninf.2014.00001

976 Wibral, M., Pampu, N., Priesemann, V., Siebenhühner, F., Seiwert, H., Lindner, M., Lizier, J.T.,
977 Vicente, R., 2013. Measuring information-transfer delays. PloS One 8, e55809.

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

981 Wiener, N., 1956. The theory of prediction. Mod. Math. Eng. N. Y. McGraw-Hill 165–190.

982 Wilson, G.T., 1972. The factorization of matricial spectral densities. SIAM J. Appl. Math. 23, 420–
983 426.

984 Wollstadt, P., Martínez-Zarzuela, M., Vicente, R., Díaz-Pernas, F.J., Wibral, M., 2014. Efficient
985 transfer entropy analysis of non-stationary neural time series. PLoS One 9, e102833.

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

988 Zipser, D., Kehoe, B., Littlewort, G., Fuster, J., 1993. A spiking network model of short-term active
989 memory. J. Neurosci. 13, 3406–3420.
990