

1 **Title**

2 Information theoretic evidence for predictive coding in the face processing system

3

4 **Abbreviated Title**

5 Neural representation of face predictions

6 **Authors**

7 Alla Brodski-Guerniero<sup>1</sup>, Georg-Friedrich Paasch<sup>1</sup>, Patricia Wollstadt<sup>1</sup>, Ipek Özdemir<sup>1</sup>, Joseph  
8 T.Lizier<sup>2</sup>, Michael Wibral<sup>1</sup>

9

10

11 **Affiliations**

12 <sup>1</sup>MEG Unit, Brain Imaging Center, J.W. Goethe University, Frankfurt a.M., Germany; <sup>2</sup>Complex  
13 Systems Research Group and Centre for Complex Systems, Faculty of Engineering & IT, The  
14 University of Sydney, NSW 2006, Australia

15

16 Corresponding author

17 Prof. Dr. Michael Wibral

18 MEG Unit, Brain Imaging Center, Goethe Universität

19 Heinrich-Hoffmann Str. 10, 60528 Frankfurt

20 Telefon: +49 (0)69 6301 83193

21 E-Mail: [wibral@bic.uni-frankfurt.de](mailto:wibral@bic.uni-frankfurt.de)

22

23 Number of pages: 36

24 Number of figures: 7

25 Number of tables: 1

26 Number of words: Abstract: 167; Introduction: 635; Discussion: 1487

27 The authors declare no competing financial interests.

28

29 **Acknowledgements:** ABG received support by Ernst Ludwig Ehrlich Studienwerk (BMBF  
30 scholarship for graduate students). GFP received support by Villigst Studienwerk (BMBF  
31 scholarship for graduate students).

32

## 33 **Abstract**

34 Predictive coding suggests that the brain infers the causes of its sensations by combining sensory  
35 evidence with internal predictions based on available prior knowledge. However, the  
36 neurophysiological correlates of (pre-)activated prior knowledge serving these predictions are still  
37 unknown. Based on the idea that such pre-activated prior knowledge must be maintained until  
38 needed we measured the amount of maintained information in neural signals via the active  
39 information storage (AIS) measure. AIS was calculated on whole-brain beamformer-reconstructed  
40 source time-courses from magnetoencephalography (MEG) recordings of 52 human subjects  
41 during the baseline of a Mooney face/house detection task. Pre-activation of prior knowledge for  
42 faces showed as alpha- and beta-band related AIS increases in content specific areas; these AIS  
43 increases were behaviourally relevant in brain area FFA. Further, AIS allowed decoding of the  
44 cued category on a trial-by-trial basis. Moreover, top-down transfer of predictions estimated by  
45 transfer entropy was associated with beta frequencies. Our results support accounts that activated  
46 prior knowledge and the corresponding predictions are signalled in low-frequency activity (<30 Hz).

## 47 **Significance statement**

48  
49 Our perception is not only determined by the information our eyes/retina and other sensory organs  
50 receive from the outside world, but strongly depends also on information already present in our  
51 brains like prior knowledge about specific situations or objects. A currently popular theory in  
52 neuroscience, predictive coding theory, suggests that this prior knowledge is used by the brain to  
53 form internal predictions about upcoming sensory information. However, neurophysiological  
54 evidence for this hypothesis is rare – mostly because this kind of evidence requires making strong  
55 a-priori assumptions about the specific predictions the brain makes and the brain areas involved.  
56 Using a novel, assumption-free approach we find that face-related prior knowledge and the derived  
57 predictions are represented and transferred in low-frequency brain activity.

## 58 59 **Introduction**

60 In the last decade, predictive coding theory has become a dominant paradigm to organize  
61 behavioral and neurophysiological findings into a coherent theory of brain function (George and  
62 Hawkins, 2009; Friston, 2010; Huang and Rao, 2011; Clark, 2012; Hohwy, 2013). Predictive

63 coding theory proposes that the brain constantly makes inferences about the state of the outside  
64 world. This is supposed to be accomplished by building hierarchical internal predictions based on  
65 prior knowledge which are compared to incoming information in order to continuously adapt these  
66 internal models (Mumford, 1992; Rao et al., 1999; Friston, 2005, 2010)

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

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

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

83 This theory is in line with the findings of a spectral predominance of low-frequency neural activity in  
84 deep cortical layers (Buffalo et al., 2011) and the physiological findings linking feedback  
85 connections to alpha/beta frequency channels in monkeys (Bastos et al., 2015) and humans  
86 (Michalareas et al., 2016).

87 Recently, this microcircuit theory of predictive coding gained experimental support by  
88 neurophysiological studies showing the predictability of events to be associated with neural power  
89 in alpha (Bauer et al., 2014; Sedley et al., 2016) or beta frequencies (Pelt et al., 2016).

90 However, representation and signalling of pre-activated prior knowledge serving predictions has  
91 been difficult to investigate with classical analysis methods. One reason is that classical analysis  
92 methods require a-priori assumptions about which predictions specific brain areas are going to  
93 make - assumptions which might be very challenging to make beyond early sensory cortices and  
94 for complex experimental designs (Wibral et al., 2014, section 4.4, p. 9). Moreover, classical  
95 analysis methods do not allow quantifying the *amount* of pre-activated prior knowledge for  
96 predictions, as for instance diminished neural activity measured by fMRI, MEG/EEG may still come  
97 with less or more information being maintained in these signals. To overcome these problems we  
98 studied the maintenance and signalling of pre-activated prior knowledge for predictions using the  
99 information-theoretic measures of active information storage (AIS, see Methods in Lizier et al.,  
100 2012; also see Gómez et al., 2014 for an application to MEG), and transfer entropy (TE,  
101 Schreiber, 2000; Vicente et al., 2011a). AIS measures the amount of information in the future of a  
102 process predicted by its past (predictable information) while TE measures the amount of directed  
103 information transfer between two processes (see Methods for details).

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

108

## 109 **Methods**

### 110 ***Basic concept and testable hypotheses***

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

115 The use of AIS and TE in our study is based on the following rationale: Since the brain will usually  
116 not know exactly when a prediction will be needed, it will maintain activated prior knowledge  
117 related to the content of the prediction over time. If there is a reliable neural code that maps

118 between content and activity, maintained activated prior knowledge must be represented as  
119 maintained information content in neural signals, measurable by AIS (Figure 1).

120 Importantly, we do not suggest that predictable information in neural signals as measured by AIS  
121 measures the predictability of external events. Rather, we suggest that AIS can be used as a  
122 measure to detect increased predictable information in specific brain areas. This predictable  
123 information is bound to rise (see Fig. 1) when prior knowledge is pre-activated based on perceptual  
124 demands and thereby becomes available for predictions.

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

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

130 1. When activated prior knowledge is maintained, predictable information as measured by AIS is  
131 supposed to be high in brain areas specific to the content of the predictions.

132 2. If the microcircuit theory of predictive coding is correct, maintenance of pre-activated prior  
133 knowledge should be reflected in alpha/beta frequencies, i.e., predictable information and  
134 alpha/beta power should correlate.

135 3. If maintenance of relevant prior knowledge is reflected by predictable information on a trial-by-  
136 trial basis, the content of predictions should be also decodable from AIS information on a trial-by-  
137 trial basis.

138 4. Information transfer related to predictions (i.e. signalling of pre-activated prior knowledge  
139 measured by TE) should occur in a top-down direction from brain areas showing increased  
140 predictable information, and should be reflected in alpha/beta band Granger causality.

141 5. As predictions based on pre-activated prior knowledge are known to facilitate performance,  
142 predictable information is supposed to correlate with behavioural parameters, if it reflects the  
143 relevant pre-activated prior knowledge.

#### 144 **Subjects**

145 57 subjects participated in the MEG experiment. 5 of these subjects had to be excluded due to

146 excessive movements, technical problems, or unavailability of anatomical scans. 52 subjects  
147 remained for the analysis (average age: 24.8 years, SD 2.8, 23 males). Each subject gave written  
148 informed consent before the beginning of the experiment and was paid 10€ per hour for  
149 participation. The local ethics committee (Johann Wolfgang Goethe University clinics, Frankfurt,  
150 Germany) approved of the experimental procedure. All subjects had normal or corrected-to-normal  
151 visual acuity and were right handed according to the Edinburgh Handedness Inventory scale  
152 (Oldfield, 1971). The large sample size subjects was chosen to reduce the risk of false positives,  
153 as suggested by (Button et al., 2013).

#### 154 ***Stimuli and stimulus presentation***

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

161 In order to increase task difficulty, in addition to Mooney faces and houses also scrambled stimuli  
162 (SCR) were created from each of the resulting Mooney faces and Mooney houses by displacing  
163 the white or black patches within the given background. Thereby all low-level information was  
164 maintained but the configuration of the face or house was destroyed. Examples of the stimuli can  
165 be seen in Figure 2.

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

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

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

### 178 ***Task and Instructions***

179 Subjects performed a detection task for faces or houses (Figure 2). Each of the eight experimental  
180 blocks started with the presentation of a written instruction; four of the experimental blocks started  
181 with the instruction “Face or not?” while for the other four experimental blocks started with the  
182 instruction “House or not?”. The former are referred to as “Face blocks” and the latter as “House  
183 blocks”. Face and House blocks were presented in alternating order. The same blocks of stimuli  
184 were presented as Face blocks for half of the subjects, while for the other half of the subjects these  
185 experimental blocks appeared as House blocks and vice versa. This way, the initial block was  
186 alternated between subjects (i.e. half of the subjects started with Face blocks and the other half  
187 with House blocks). Importantly, as the blocks contained the same face, house, SCR face and  
188 SCR house stimuli the only difference between face and house blocks was in the subjects’  
189 instruction.

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

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

200 **Data acquisition**

201 MEG data acquisition was performed in line with recently published guidelines for MEG recordings  
202 (Gross et al., 2012). MEG signals were recorded using a whole-head system (Omega 2005; VSM  
203 MedTech Ltd.) with 275 channels. The signals were recorded continuously at a sampling rate of  
204 1200 Hz in a synthetic third-order gradiometer configuration and were filtered online with fourth-  
205 order Butterworth filters with 300 Hz low pass and 0.1 Hz high pass.

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

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

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

221 Behavioral responses were recorded using a fiberoptic response pad (Photon Control Inc.  
222 Lumitouch Control <sup>TM</sup> Response System) in combination with the Presentation software (Version  
223 9.90, Neurobehavioral Systems).

224



## 225 ***Statistical analysis of behavioral data***

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

229 Post-hoc Wilcoxon signed rank tests were performed on hitrates as well as reaction times. To  
230 account for multiple testing, sequential Bonferroni-Holm correction (Holm, 1979) was applied  
231 (uncorrected alpha = 0.05).

## 232 ***MEG-data preprocessing***

233 MEG Data analysis was performed with Matlab (RRID:nlx\_153890; Matlab 2012b, The Mathworks,  
234 Inc.) using the open source Matlab toolbox Fieldtrip (RRID:nlx\_143928; Oostenveld et al., 2011);  
235 Version 2013 11-11) and custom Matlab scripts.

236 Only trials with correct behavioral responses were taken into account for MEG data analysis. The  
237 focus of data analysis was on the prestimulus intervals from 1 s to 0.050 s before stimulus onset.  
238 Trials containing sensor jump-, or muscle-artefacts were rejected using automatic FieldTrip artefact  
239 rejection routines. Line noise was removed using a discrete Fourier transform filter at 50,100 and  
240 150 Hz. In addition, independent component analysis (ICA; (Makeig et al., 1996) was performed  
241 using the extended infomax (runica) algorithm implemented in fieldtrip/EEGLAB. ICA components  
242 strongly correlated with EOG and ECG channels were removed from the data. Finally, data was  
243 visually inspected for residual artefacts.

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

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

## 250 **Sensor level spectral analysis**

251 Spectral analysis at the sensor level was performed in order to determine the subdivision of the  
252 power spectrum in frequency bands (see Brodski et al., 2015 for a similar approach). As we aimed  
253 to identify frequency bands based on stimulus related increases or decreases, respectively, before  
254 spectral analysis new data segments were cut from -0.55 to 0.55 s around stimulus onset. For  
255 spectral analysis we used a multitaper approach (Percival and Walden, 1993) based on Slepian  
256 sequences (Slepian, 1978). The spectral transformation was applied in an interval from 4 to 150 Hz  
257 in 2 Hz steps in time steps of 0.01 s and using two slepian tapers for each frequency. For each  
258 subject, time-frequency representations were averaged for Face blocks and House blocks as well  
259 as within the time interval of “baseline” (-0.35 s – 0.05 s) and “task” (0.05 s – 0.35 s), respectively.  
260 Average spectra of task and baseline period were contrasted over subjects using a dependent-  
261 sample permutation t-metric with a cluster based correction method (Maris and Oostenveld, 2007)  
262 to account for multiple comparisons. Adjacent samples whose *t*-values exceeded a threshold  
263 corresponding to an uncorrected  $\alpha$ -level of 0.05 were defined as clusters. The resulting cluster  
264 sizes were then tested against the distribution of cluster sizes obtained from 1000 permuted  
265 datasets (i.e. labels “task” and “baseline” were randomly reassigned within each of the subjects).  
266 Cluster sizes larger than the 95th percentile of the cluster sizes in the permuted datasets were  
267 defined as significant.

268 Following the same approach as (Brodski et al., 2015) based on the significant clusters of the task  
269 vs. baseline statistics five frequency bands were defined for further analysis: (1) 8–14 Hz (alpha);  
270 (2) 14–32 Hz (beta); (3) 32–56 Hz (low gamma); (4) 56–64 Hz (mid gamma) and (5) 64–150 Hz  
271 (high gamma) (Figure 3).

## 272 **Source grid creation**

273 In order to create individual source grids we transformed the anatomical MR images to a standard  
274 T1 MNI template from the SPM8 toolbox (<http://www.fil.ion.ucl.ac.uk/spm>) - obtaining an individual  
275 transformation matrix for each subject. We then warped a regular 3-D dipole grid based on the  
276 standard T1 template (spacing 15mm resulting in 478 grid locations) with the inverse of each

277 subjects' transformation matrix, to obtain an individual dipole grid for each subject in subject space.  
278 This way, each specific grid point was located at the same brain area for each subject, which  
279 allowed us to perform source analysis with individual head models as well as multi-subject  
280 statistics for all grid locations. Lead-fields at those grid locations were computed for the individual  
281 subjects with a realistic single shell forward model (Nolte, 2003) taking into account the effects of  
282 the ICA component removal in pre-processing.

### 283 **Source time course reconstruction**

284

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

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

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

### 297 **Definition of active information storage**

298 We assume that the reconstructed source time courses for each brain location can be treated as  
299 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  
300 collection of random variables,  $X_t$ , ordered by some integer  $t$ . AIS then describes how much of the  
301 information the next time step  $t$  of the process is predictable from its immediate past state  
302 (Lizier et al., 2012). This is defined as the mutual information

$$303 \quad A_x = \lim_{k \rightarrow \infty} I(X_{t-1}^k; X_t) = \lim_{k \rightarrow \infty} \sum_{x_t, x_{t-1}^k} p(x_t, x_{t-1}^k) \log \frac{p(x_{t-1}^k, x_t)}{p(x_{t-1}^k)p(x_t)} \quad (1)$$

304

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

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

### 321 ***Analysis of predictable information using active information storage***

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

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

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

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

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

### 343 ***AIS Statistics***

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

356

## 357 **Correlation analysis**

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

368 We also calculated the correlation of t-values computed from AIS (based on the dependent sample  
369 t-metric, contrast Face blocks vs. House blocks) for all grid points at the source level with the t-  
370 values obtained from the same contrast based on beamformer reconstructed source power in the  
371 alpha (8-14 Hz) and beta (14-32 Hz) frequency band. Source power was reconstructed with the  
372 DICS (dynamic imaging of coherent sources, Gross et al., 2001) algorithm as implemented in the  
373 FieldTrip toolbox using real values filter coefficients only - see also Grützner et al., 2010).

374 Last, we accessed the relationship of AIS values and reaction times for each subject. To this end  
375 before the correlation analysis for each subject mean reaction times and mean AIS values in the  
376 brain areas of interest for Face and House blocks were subtracted from each other. This allowed  
377 accounting for differential behavioral speed between subjects. The correlation of the difference in  
378 AIS values and the difference in reaction times was calculated via Spearman skipped correlations  
379 using the Robust correlation Toolbox (Pernet et al., 2013). Calculation of skipped correlations  
380 includes identifying and removing bivariate outliers (Rousseeuw, 1984; Rousseeuw and Driessen,  
381 1999; Verboven and Hubert, 2005). This can provide a more robust measure, which has been  
382 recommended for brain-behaviour correlation analyses (Rousselet and Pernet, 2012). The  
383 uncorrected alpha level was set to 0.05. For each correlation bootstrap confidence intervals (CIs)

384 were computed based on 1000 resamples. In order to account for multiple comparisons across  
385 brain areas, bootstrap CIs were adjusted using Bonferroni correction. If the adjusted CI did not  
386 encompassed 0, the correlation was considered as significant.

387

### 388 **Decoding analysis**

389

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

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

406

407

### 408 **Definition of transfer entropy (and Granger analysis)**

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

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

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

$$\begin{aligned} 421 \quad TE_{Y \rightarrow X} &= \lim_{j,k \rightarrow \infty} I(X_t; Y_{t-u}^j | X_{t-1}^k) \\ 422 \quad &= \lim_{j,k \rightarrow \infty} \sum_{x_t, x_{t-1}^k, y_{t-u}^j} p(x_t, x_{t-1}^k, y_{t-u}^j) \log \frac{p(x_t | x_{t-1}^k, y_{t-u}^j)}{p(x_t | x_{t-1}^k)} \end{aligned} \quad (2)$$

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

#### 428 **Analysis of information transfer using transfer entropy and Granger causality analysis**

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

435 In the TE analysis the same time intervals (prestimulus) and embedding parameters as for AIS  
436 analysis were used. TE values for Face blocks and House blocks were contrasted using a  
437 dependent-sample permutation t-metric for statistical analysis across subjects. In the statistical  
438 analysis, FDR correction was used to account for multiple comparisons across links (uncorrected  
439 alpha level 0.05). As for AIS, the history dimension for the past states was set to finite values; we



440 here set  $j_{\max} = k_{\max}$  and used the values obtained during AIS estimation for the target time series  
441 of each signal combination.

442 For the significant TE links post-hoc nonparametric bivariate Granger causality analysis in the  
443 frequency domain (Dhamala et al., 2008) was computed. Using the nonparametric variant of  
444 Granger causality analysis avoids choosing an autoregressive model order, which may easily  
445 introduce a bias. In the nonparametric approach Granger causality is computed from a  
446 factorization of the spectral density matrix, which is based on the direct Fourier transform of the  
447 time series data (Dhamala et al., 2008). The Wilson algorithm was used for factorization (Wilson,  
448 1972). A spectral resolution of 2 Hz and a spectral smoothing of 5 Hz were used for spectral  
449 transformation using the multitaper approach (Percival and Walden, 1993) (9 Slepian tapers). We  
450 were interested in the differences of Granger spectral fingerprints of Face and House blocks,  
451 however we also wanted to make sure that the Granger values for these difference significantly  
452 differed from noise. For that reason we created two additional “random” conditions by permuting  
453 the trials for the Face block and the House block condition for each source separately. Two types  
454 of statistical comparisons were performed for the frequency range between 8 and 150 Hz and each  
455 of the significant TE links: 1. Granger values in Face blocks were contrasted with Granger values  
456 in House blocks using a dependent-samples permutation t-metric 2. Granger values in Face  
457 blocks/House blocks were contrasted with the random Face block condition / random House block  
458 condition using another dependent-samples permutation t-metric. For the first test a cluster-  
459 correction was used to account for multiple comparisons across frequency (Maris and Oostenveld,  
460 2007). Adjacent samples which uncorrected p-values were below 0.01 were considered as  
461 clusters. 5000 permutations were performed and the alpha value was set to 0.05. Frequency  
462 intervals in the Face block vs. House block comparison were only considered as significant if all  
463 included frequencies also reached significance in the comparison with the random conditions using  
464 a Bonferroni-Holms correction to account for multiple comparisons.

## 465 **Results**

### 466 ***Behavioral results***

467 We found no differences between Face blocks and House blocks for hitrates (avg. hitrate Face  
468 blocks 93.9%; avg. hitrate House blocks 94.6%; Wilcoxon Signed rank test  $p=0.57$ ) and reaction  
469 times of correct responses (avg. mean reaction times Face blocks 0.545 s, avg. reaction times  
470 House blocks 0.546 s; Wilcoxon Signed rank test  $p=0.85$ ). Subjects showed equivalent behavioural  
471 patterns for both block types, for instance increased reaction times for the instructed stimulus  
472 conditions as these stimuli had to be distinguished from a similar distractor (SCR stimuli) (see  
473 Figure 4 for the analysis of behavioural differences between stimulus conditions within block  
474 types).

### 475 ***Analysis of predictable information***

476 Statistical comparisons of AIS values between Face blocks and House blocks in the prestimulus  
477 interval revealed increased AIS values for Face blocks in clusters in fusiform face area (FFA),  
478 anterior inferior temporal cortex (aIT), occipital face area (OFA), posterior parietal cortex (PPC)  
479 and primary visual cortex (V1) (Figure 5). We referred to these five brain areas as “face prediction  
480 network” and subjected it to further analyses. In contrast to this finding of a face prediction network,  
481 we did not find brain areas showing significantly higher AIS values in House blocks compared to  
482 Face blocks. This is similar to highly cited previous studies that failed to find prediction effects for  
483 houses in the brain in contrast to faces (e.g. Summerfield et al., 2006a, 2006b; Trapp et al., 2015).

484

## 485 ***Correlation of single trial power and single trial predictable information***

486 In order to investigate the neurophysiological correlates of activated prior knowledge identified via  
487 AIS analysis, a correlation analysis of single trial power in distinct frequency bands with single trial  
488 AIS was conducted. Correlation analysis revealed a strong positive correlation in the alpha and  
489 beta frequency bands, only very small mostly positive correlations in the low and mid-gamma  
490 frequency bands and a small negative correlation for the high-gamma frequency band (Table 1).  
491 Note that although the correlations in the higher frequency bands were partly significant, the effect  
492 size was much higher in the alpha and beta frequency band. This means that alpha and beta band  
493 activity is the most likely carrier of activated prior knowledge.

494 While we found a significant correlation of single trial power and predictable information in the  
495 alpha and beta band, the contrast map over all source grid points for Face and House blocks (t-  
496 values obtained from dependent sample t-metric over subjects) did not correlate with the AIS  
497 contrast map for both, alpha and beta power (alpha rho = 0.043, p = 0.33; beta rho = 0.05, p =  
498 0.21). This suggests that AIS analysis provides additional information not directly provided by a  
499 spectral analysis. In sum, while AIS seems to be carried by alpha/beta-band activity, not all  
500 alpha/beta band activity contributes to AIS.

## 501 ***Decoding prediction content from single trial AIS values***

502 To study whether face or house predictions can be decoded from AIS values of the face prediction  
503 network on a trial-by-trial basis, support vector machines were used (Chang and Lin, 2011). Cross-  
504 validated decoding performance reached a maximum of 65.2% (mean performance 53.5%, SD  
505 3.9% over subjects). When bonferroni correcting for the high number of subjects tested (n=52), for  
506 22 of the subjects performance was still significantly better than for permuted datasets (p <  
507 0.05/52). Note, that this is much more than would have been expected by chance (p =  $1.1 \times 10^{-52}$ ,  
508 binomial test).

## 509 ***Analysis of information transfer***

510 To understand how activated prior knowledge is communicated within the cortical hierarchy, we

511 assessed the information transfer within the face prediction network in the prestimulus interval by  
512 estimating transfer entropy (TE, Schreiber, 2000) on source time courses for Face blocks and  
513 House blocks, respectively. Statistical analysis revealed significantly increased information transfer  
514 for Face blocks from aIT to FFA ( $p=0.0001$ , fdr correction) and from PPC to FFA ( $p = 0.0014$ , fdr  
515 correction). For House blocks information transfer was increased in comparison to Face blocks  
516 from brain area V1 to PPC ( $p=0.0014$ , fdr correction) (Figure 6).

517

#### 518 ***Post-hoc frequency resolved Granger causality***

519 In order to investigate whether information transfer differences in Face and House blocks were  
520 reflected in specific frequency bands, we post-hoc performed a non-parametric spectral Granger  
521 causality analysis on the three links identified with transfer entropy analysis. For the link from PPC  
522 to FFA we found stronger Granger causality for Face blocks than House blocks in a cluster  
523 between 18 and 22 Hz (Figure 7,  $p=0.045$ , cluster correction for frequencies, uncorrected for the  
524 number of links in this post hoc test). The link from V1 to PPC showed a stronger Granger causal  
525 influence for House blocks than Face blocks between 94 and 98 Hz (Figure 7,  $p=0.042$ , cluster  
526 correction for frequencies, uncorrected for the number of links in this post hoc test). Using cluster  
527 correction, the link from aIT to FFA did not show significant differences in Granger causal  
528 influence.

#### 529 ***Correlation of predictable information and reaction times***

530 In order to study the association of predictable information and behaviour, we correlated the per  
531 subject difference of AIS values between Face blocks and House blocks with the per subject  
532 difference in reaction times. This analysis was performed for the three brain areas between which  
533 we found increased information transfer during Face blocks (FFA, aIT and PPC). For these brain  
534 areas we tested the hypothesis that predictable information for face blocks was associated with  
535 performance, i.e. reaction times during Face blocks. Negative correlation values were found for all  
536 of the three brain areas, however only brain area FFA reached significance when correcting for  
537 multiple comparisons: FFA robust Spearman's rho -0.41, robust confidence interval (CI) after  
538 correcting for multiple comparisons [-0.68 -0.066]; aIT robust Spearman's rho = -0.12, CI [-0.4554  
539 0.245]; PPC robust Spearman's rho -0.21 CI [-0.5480 0.1178].

540 **Discussion**

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

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

561

562 **1. Activated prior knowledge for faces shows as predictable information in content specific**  
563 **areas**

564 We found increased predictable information as reflected by increased AIS values in Face blocks in  
565 the prestimulus interval in FFA, OFA, aIT, PPC and V1. Out of these five brain areas FFA, OFA  
566 and aIT are well known to play a major role in face processing (Kanwisher et al., 1997;  
567 Kriegeskorte et al., 2007; Tsao et al., 2008; Pitcher et al., 2011).

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

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

580 However, while we found increased predictable information in content specific areas for Face  
581 blocks, we did not find brain areas showing increased predictable information for House blocks.  
582 Similarly, in a face/house discrimination task Summerfield and colleagues (2006b) observed  
583 increased activation in FFA, when a house was misperceived as a face. However, they failed to  
584 see increased activation in parahippocampal place area (PPA), a scene/house responsive region,  
585 when a face was misperceived as a house. The authors suggest that this might be related to the  
586 fact that PPA is less subject to top-down information than FFA – as faces have much more  
587 regularities potentially utilizable for top-down mechanisms than the natural scenes that PPA  
588 usually responds to. Additionally, because of their strong social relevance (e.g. Farah et al., 1995)

589 faces capture attention disproportionately (e.g. Vuilleumier and Schwartz, 2001). Thus, also face  
590 predictions/templates may be prioritized in comparison to other templates e.g. for houses  
591 (Esterman and Yantis, 2009; Puri et al., 2009; Van Belle et al., 2010).

## 592 **2. Maintenance of activated prior knowledge about faces is reflected by increased** 593 **alpha/beta power**

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

597 Congruently with our findings, Mayer and colleagues (2015) recently showed that activation of prior  
598 knowledge about previously seen letters is associated with increased power in alpha frequencies in  
599 the prestimulus interval. Also, Sedley and colleagues (2016) observed that the update of  
600 predictions, which also requires access to maintained activated knowledge, is associated with  
601 increased power in beta frequencies.

602 Extending these previous findings, our results demonstrate that the activated prior knowledge  
603 usable as predictions for face detection is associated with neural activity in the alpha and beta  
604 frequency range (Bastos et al., 2012).

## 605 **3. Face predictions are transferred in a top-down manner and via beta frequencies**

606 In Face blocks we observed increased information transfer to FFA from aIT as well as from PPC,  
607 both areas located higher in the processing hierarchy than FFA (e.g. Zhen et al., 2013;  
608 Michalareas et al., 2016). Thus, FFA seems to have the role of a convergence center to which  
609 information from higher cortical areas is transferred in order to prepare for rapid face detection.

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

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

621 Our findings further demonstrated that information transfer in top-down direction was associated  
622 with Granger causality in the beta frequency band (PPC to FFA), while information transfer in  
623 bottom-up direction was associated with Granger causality in the high gamma frequency band (V1  
624 to PPC).

625 The association of top-down information transfer with beta frequencies and bottom-up information  
626 transfer with gamma frequencies is in line with recent physiological findings in monkeys and  
627 humans (Bastos et al., 2015; Michalareas et al., 2016) and has been linked with predictive coding  
628 in Bastos' microcircuit model (Bastos et al., 2012), resulting in the hypothesis of predictions being  
629 transferred top-down via low-frequency channels and prediction errors bottom-up via high  
630 frequency channels. In accord with this hypothesis, our group has recently shown that prediction  
631 errors are communicated in the high frequency gamma band (Brodski et al., 2015). Our present  
632 finding of top-down information transfer in low beta frequencies during anticipation of faces adds  
633 support to the microcircuit model hypothesis of a low-frequency channel for the top-down  
634 propagation of predictions (Bastos et al., 2012).

635 In line with our findings, the spectral dissociation between the transfer of predictions and of  
636 prediction errors recently received additional support from a MEG study applying Granger causality  
637 analysis for the investigation of information transfer during the prediction of causal events (Pelt et  
638 al., 2016). It should be noted that van Pelt and colleagues defined their network of interest for  
639 Granger causality analysis based on the prior assumption of the involvement of these brain areas  
640 in causal inference. In contrast, in our study defining the network via condition-specific AIS



641 increases allowed finding the brain areas involved in predictive processing without relying on prior  
642 assumptions about their function.

#### 643 **4. Pre-activation of prior knowledge about faces facilitates performance**

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

655

## 656 **References**

- 657 Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for  
658 predictive coding. *Neuron* 76:695–711.
- 659 Bastos AM, Vezoli J, Bosman CA, Schoffelen J-M, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H,  
660 Fries P (2015) Visual areas exert feedforward and feedback influences through distinct frequency  
661 channels. *Neuron* 85:390–401.
- 662 Bauer M, Stenner M-P, Friston KJ, Dolan RJ (2014) Attentional Modulation of Alpha/Beta and Gamma  
663 Oscillations Reflect Functionally Distinct Processes. *J Neurosci* 34:16117–16125.
- 664 Brodski A, Paasch G-F, Helbling S, Wibral M (2015) The Faces of Predictive Coding. *J Neurosci* 35:8997–  
665 9006.
- 666 Buffalo EA, Fries P, Landman R, Buschman TJ, Desimone R (2011) Laminar differences in gamma and  
667 alpha coherence in the ventral stream. *Proc Natl Acad Sci* 108:11262–11267.
- 668 Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR (2013) Power failure: why  
669 small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* Available at:  
670 <http://www.nature.com/nrn/journal/vaop/ncurrent/full/nrn3475.html> [Accessed May 8, 2013].
- 671 Cavanagh P (1991) What's up in top-down processing. *Represent Vis Trends Tacit Assumpt Vis Res*:295–  
672 304.
- 673 Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol*  
674 *TIST* 2:27.
- 675 Clark A (2012) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav*  
676 *Brain Sci* Available at: <http://bi.snu.ac.kr/Courses/aplc12/3-2.pdf> [Accessed May 22, 2013].
- 677 Dhamala M, Rangarajan G, Ding M (2008) Estimating Granger causality from Fourier and wavelet transforms  
678 of time series data. *Phys Rev Lett* 100:018701.
- 679 Dolan RJ, Fink GR, Rolls E, Booth M, Holmes A, Frackowiak RSJ, Friston KJ (1997) How the brain learns to  
680 see objects and faces in an impoverished context. *Nature* 389:596–598.
- 681 Esterman M, Yantis S (2009) Perceptual expectation evokes category-selective cortical activity. *Cereb*  
682 *Cortex*:bhp188.
- 683 Faes L, Nollo G, Porta A (2013) Compensated transfer entropy as a tool for reliably estimating information  
684 transfer in physiological time series. *Entropy* 15:198–219.
- 685 Farah MJ, Tanaka JW, Drain HM (1995) What causes the face inversion effect? *J Exp Psychol Hum Percept*  
686 *Perform* 21:628–634.
- 687 Frenzel S, Pompe B (2007) Partial mutual information for coupling analysis of multivariate time series. *Phys*  
688 *Rev Lett* 99:204101.
- 689 Friston K (2005) A theory of cortical responses. *Philos Trans R Soc B Biol Sci* 360:815–836.
- 690 Friston K (2010) The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11:127–138.
- 691 George D, Hawkins J (2009) Towards a Mathematical Theory of Cortical Micro-circuits. *PLoS Comput Biol*  
692 5:e1000532.

- 693 Gómez C, Lizier JT, Schaum M, Wollstadt P, Grützner C, Uhlhaas P, Freitag CM, Schlitt S, Bölte S, Hornero  
694 R, Wibral M (2014) Reduced predictable information in brain signals in autism spectrum disorder.  
695 Front Neuroinformatics 8 Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3924322/>  
696 [Accessed March 4, 2015].
- 697 Gómez-Herrero G, Wu W, Rutanen K, Soriano MC, Pipa G, Vicente R (2015) Assessing coupling dynamics  
698 from an ensemble of time series. Entropy 17:1958–1970.
- 699 Granger CWJ (1969) Investigating Causal Relations by Econometric Models and Cross-spectral Methods.  
700 Econometrica 37:424–438.
- 701 Gross J, Baillet S, Barnes GR, Henson RN, Hillebrand A, Jensen O, Jerbi K, Litvak V, Maess B, Oostenveld  
702 R (2012) Good-practice for conducting and reporting MEG research. NeuroImage Available at:  
703 <http://www.sciencedirect.com/science/article/pii/S1053811912009895> [Accessed April 5, 2013].
- 704 Gross J, Kujala J, Hamalainen M, Timmermann L, Schnitzler A, Salmelin R (2001) Dynamic imaging of  
705 coherent sources: Studying neural interactions in the human brain. Proc Natl Acad Sci U S A  
706 98:694–699.
- 707 Grützner C, Uhlhaas PJ, Genc E, Kohler A, Singer W, Wibral M (2010) Neuroelectromagnetic Correlates of  
708 Perceptual Closure Processes. J Neurosci 30:8342–8352.
- 709 Hohwy J (2013) The Predictive Mind. Oxford University Press.
- 710 Holm S (1979) A Simple Sequentially Rejective Multiple Test Procedure. Scand J Stat 6:65–70.
- 711 Huang Y, Rao RP (2011) Predictive coding. Wiley Interdiscip Rev Cogn Sci 2:580–593.
- 712 Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex  
713 specialized for face perception. J Neurosci 17:4302–4311.
- 714 Kemelmacher-Shlizerman I, Basri R, Nadler B (2008) 3D shape reconstruction of Mooney faces. In:  
715 Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pp 1–8  
716 Available at: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4587769](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4587769) [Accessed August 24,  
717 2012].
- 718 Kok P, Failing MF, de Lange FP (2014) Prior expectations evoke stimulus templates in the primary visual  
719 cortex. J Cogn Neurosci 26:1546–1554.
- 720 Kraskov A, Stögbauer H, Grassberger P (2004) Estimating mutual information. Phys Rev E 69:066138.
- 721 Kriegeskorte N, Formisano E, Sorger B, Goebel R (2007) Individual faces elicit distinct response patterns in  
722 human anterior temporal cortex. Proc Natl Acad Sci 104:20600–20605.
- 723 Lindner M, Vicente R, Priesemann V, Wibral M (2011) TRENTOOL: A Matlab open source toolbox to analyse  
724 information flow in time series data with transfer entropy. BMC Neurosci 12:119.
- 725 Lizier JT (2014) JIDT: an information-theoretic toolkit for studying the dynamics of complex systems. Comput  
726 Intell 1:11.
- 727 Lizier JT, Prokopenko M, Zomaya AY (2012) Local measures of information storage in complex distributed  
728 computation. Inf Sci 208:39–54.
- 729 Makeig S, Bell AJ, Jung T-P, Sejnowski TJ (1996) Independent component analysis of  
730 electroencephalographic data. Adv Neural Inf Process Syst:145–151.
- 731 Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods  
732 164:177–190.

- 733 Mayer A, Schwiedrzik CM, Wibral M, Singer W, Melloni L (2015) Expecting to See a Letter: Alpha  
734 Oscillations as Carriers of Top-Down Sensory Predictions. *Cereb Cortex*:bhv146.
- 735 Michalareas G, Vezoli J, van Pelt S, Schoffelen J-M, Kennedy H, Fries P (2016) Alpha-Beta and Gamma  
736 Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas.  
737 *Neuron* Available at: <http://www.sciencedirect.com/science/article/pii/S0896627315011204>  
738 [Accessed February 26, 2016].
- 739 Mooney CM (1957) Age in the development of closure ability in children. *Can J Psychol Can Psychol*  
740 11:219–226.
- 741 Mooney CM, Ferguson GA (1951) A new closure test. *Can J Psychol Can Psychol* 5:129–133.
- 742 Mumford D (1992) On the computational architecture of the neocortex. *Biol Cybern* 66:241–251.
- 743 Nolte G (2003) The magnetic lead field theorem in the quasi-static approximation and its use for  
744 magnetoencephalography forward calculation in realistic volume conductors. *Phys Med Biol*  
745 48:3637–3652.
- 746 Nowotny T (2014) Two challenges of correct validation in pattern recognition. *Comput Intell* 1:5.
- 747 Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory.  
748 *Neuropsychologia* 9:97–113.
- 749 Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: open source software for advanced  
750 analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011:1.
- 751 Pashkam MV, Xu Y (2014) Decoding visual object representation in human parietal cortex. *J Vis* 14:1307–  
752 1307.
- 753 Pelt S van, Heil L, Kwisthout J, Ondobaka S, Rooij I van, Bekkering H (2016) Beta- and gamma-band activity  
754 reflect predictive coding in the processing of causal events. *Soc Cogn Affect Neurosci*:nsw017.
- 755 Percival DB, Walden AT (1993) *Spectral Analysis for Physical Applications*. Cambridge University Press.
- 756 Pernet CR, Wilcox RR, Rousselet GA (2013) Robust correlation analyses: false positive and power validation  
757 using a new open source Matlab toolbox. *Front Psychol* 3:606.
- 758 Pitcher D, Walsh V, Duchaine B (2011) The role of the occipital face area in the cortical face perception  
759 network. *Exp Brain Res* 209:481–493.
- 760 Puri AM, Wojciulik E, Ranganath C (2009) Category expectation modulates baseline and stimulus-evoked  
761 activity in human inferotemporal cortex. *Brain Res* 1301:89–99.
- 762 Ragwitz M, Kantz H (2002) Markov models from data by simple nonlinear time series predictors in delay  
763 embedding spaces. *Phys Rev E* 65:056201.
- 764 Ranganath C, Cohen MX, Dam C, D'Esposito M (2004) Inferior Temporal, Prefrontal, and Hippocampal  
765 Contributions to Visual Working Memory Maintenance and Associative Memory Retrieval. *J Neurosci*  
766 24:3917–3925.
- 767 Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-  
768 classical receptive-field effects. *Nat Neurosci* 2:79–87.
- 769 Rousseeuw PJ (1984) Least median of squares regression. *J Am Stat Assoc* 79:871–880.
- 770 Rousseeuw PJ, Driessen KV (1999) A fast algorithm for the minimum covariance determinant estimator.  
771 *Technometrics* 41:212–223.

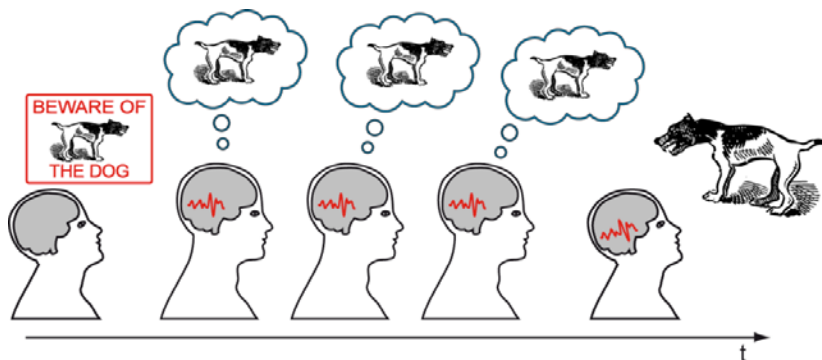
- 772 Rousselet GA, Pernet CR (2012) Improving standards in brain-behavior correlation analyses. *Front Hum*  
773 *Neurosci* 6:119.
- 774 Schreiber T (2000) Measuring information transfer. *Phys Rev Lett* 85:461.
- 775 Sedley W, Gander PE, Kumar S, Kovach CK, Oya H, Kawasaki H, Iii MAH, Griffiths TD (2016) Neural  
776 signatures of perceptual inference. *eLife* 5:e11476.
- 777 Slepian D (1978) Prolate spheroidal wave functions, Fourier analysis and uncertainty. *Bell Syst Tech J*  
778 57:1371–1429.
- 779 Summerfield C, Egner T, Greene M, Koechlin E, Mangels J, Hirsch J (2006a) Predictive Codes for  
780 Forthcoming Perception in the Frontal Cortex. *Science* 314:1311–1314.
- 781 Summerfield C, Egner T, Mangels J, Hirsch J (2006b) Mistaking a house for a face: neural correlates of  
782 misperception in healthy humans. *Cereb Cortex N Y N* 1991 16:500–508.
- 783 Trapp S, Lepsien J, Kotz SA, Bar M (2015) Prior probability modulates anticipatory activity in category-  
784 specific areas. *Cogn Affect Behav Neurosci*:1–10.
- 785 Tsao DY, Moeller S, Freiwald WA (2008) Comparing face patch systems in macaques and humans. *Proc*  
786 *Natl Acad Sci* 105:19514–19519.
- 787 Van Belle G, De Graef P, Verfaillie K, Busigny T, Rossion B (2010) Whole not hole: Expert face recognition  
788 requires holistic perception. *Neuropsychologia* 48:2620–2629.
- 789 Van Veen BD, Van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of brain electrical activity via  
790 linearly constrained minimum variance spatial filtering. *IEEE Trans Biomed Eng* 44:867–880.
- 791 Verboven S, Hubert M (2005) LIBRA: a MATLAB library for robust analysis. *Chemom Intell Lab Syst* 75:127–  
792 136.
- 793 Vicente R, Wibral M, Lindner M, Pipa G (2011) Transfer entropy—a model-free measure of effective  
794 connectivity for the neurosciences. *J Comput Neurosci* 30:45–67.
- 795 Vuilleumier P, Schwartz S (2001) Emotional facial expressions capture attention. *Neurology* 56:153–158.
- 796 Wibral M, Lizier JT, Vögler S, Priesemann V, Galuske R (2014) Local active information storage as a tool to  
797 understand distributed neural information processing. *Front Neuroinformatics* 8 Available at:  
798 <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3904075/> [Accessed March 19, 2015].
- 799 Wibral M, Pampu N, Priesemann V, Siebenhühner F, Seiwert H, Lindner M, Lizier JT, Vicente R (2013)  
800 Measuring information-transfer delays. *PLoS One* 8:e55809.
- 801 Wibral M, Rahm B, Rieder M, Lindner M, Vicente R, Kaiser J (2011) Transfer entropy in  
802 magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks.  
803 *Prog Biophys Mol Biol* 105:80–97.
- 804 Wiener N (1956) The theory of prediction. *Mod Math Eng N Y McGraw-Hill*:165–190.
- 805 Wilson GT (1972) The factorization of matricial spectral densities. *SIAM J Appl Math* 23:420–426.
- 806 Wollstadt P, Martínez-Zarzuela M, Vicente R, Díaz-Pernas FJ, Wibral M (2014) Efficient transfer entropy  
807 analysis of non-stationary neural time series. *PLoS One* 9:e102833.
- 808 Zhen Z, Fang H, Liu J (2013) The hierarchical brain network for face recognition. *PLoS One* 8:e59886.
- 809 Zipser D, Kehoe B, Littlewort G, Fuster J (1993) A spiking network model of short-term active memory. *J*  
810 *Neurosci* 13:3406–3420.

811 **Tables & Figures**

812 Table 1: Correlation of single trial power and single trial predictable information (measured by AIS)  
 813 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

814

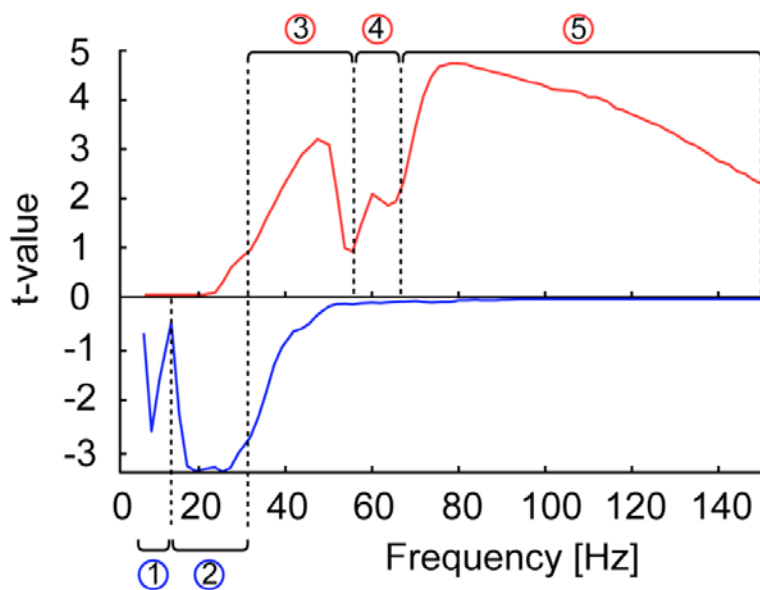


815

816 **Figure 1. Central idea of the study.** Typically, pre-activated prior knowledge related to the  
 817 content of a prediction has to be maintained as the brain will not know exactly when it will be  
 818 needed. If there is a reliable neural code that maps between content and activity, maintained  
 819 activated prior knowledge should lead to brain signals that are themselves predictable over time  
 820 (here the brain signals are depicted as identical, although the relation between past and future will  
 821 almost certainly be much more complicated). Figure elements obtained from OpenCliparts Library  
 822 (<http://www.opencliparts.org>) and modified.

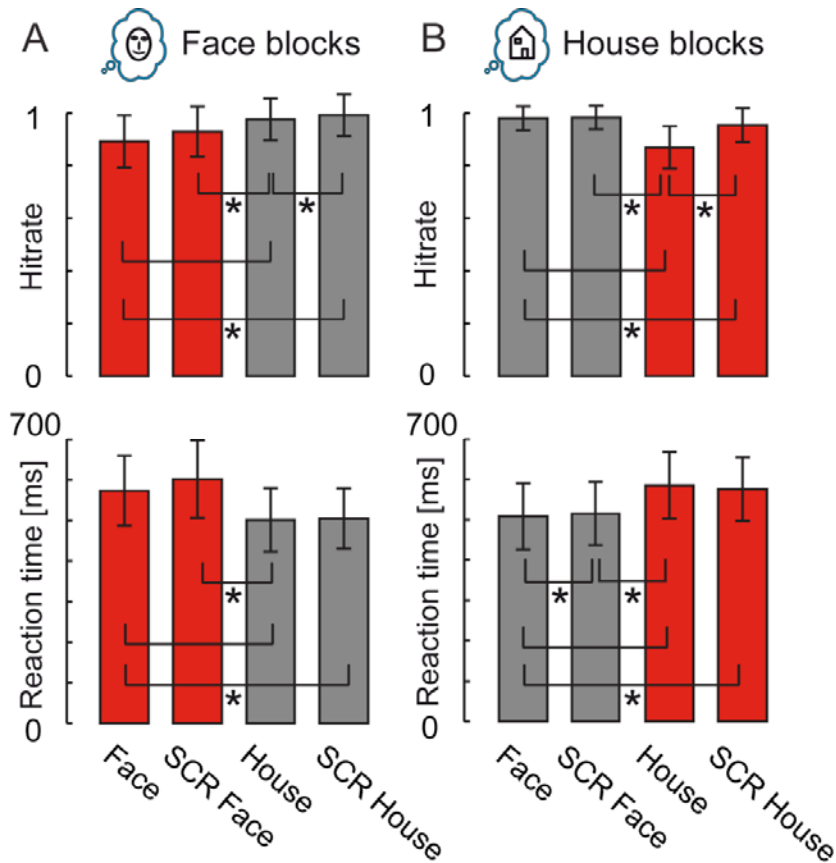


824 **Figure 2. Experimental design.** *Top and bottom:* Exemplary stimulus presentation in Face blocks  
825 (*top*) and in House blocks (*bottom*). Face and House icons on the left indicate Face and House  
826 blocks, respectively. *Middle:* Depiction of stimulus categories and timing. The beginning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



830 **Figure 3. Sensor-level frequency analysis - defining frequency bands.** Power spectra for all of  
831 the significant clusters (one positive and one negative cluster) at the sensor level (permutation t-  
832 metric, contrast [0.05s 0.35s] vs [-0.35s -0.05s] around stimulus onset,  $t$  values masked by  $p <$   
833 0.05, cluster correction,  $n = 52$ ). Frequency analysis at the sensor level was calculated using both  
834 blocks types jointly. Task-related increases in power are shown in red (positive cluster) and task-  
835 related decreases in blue (negative cluster). Black dashed lines frame the identified frequency

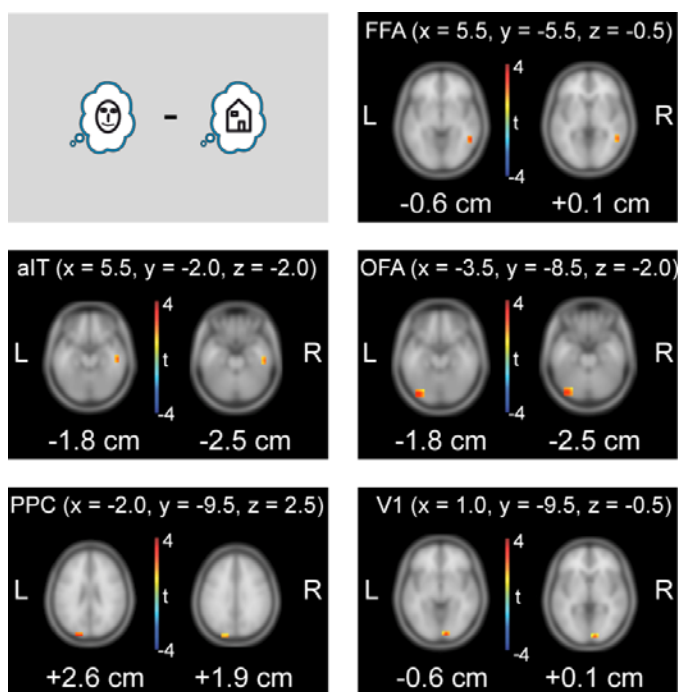
836 ranges.



837

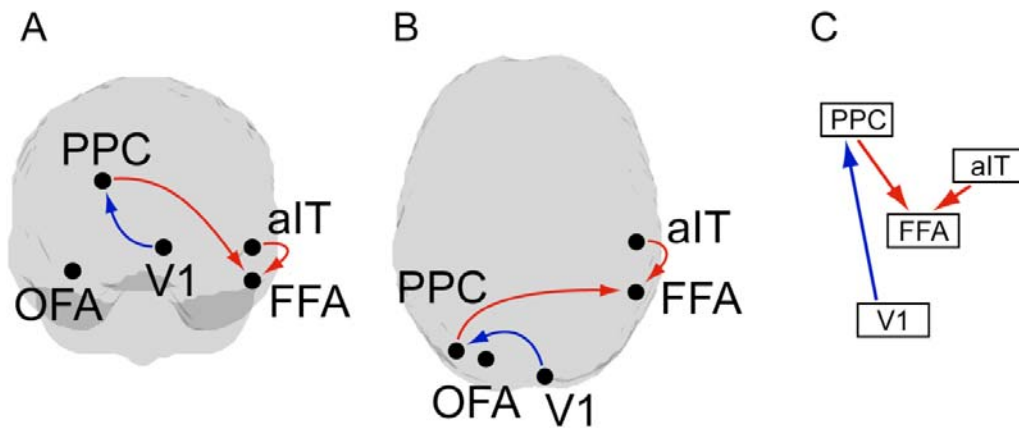
838 **Figure 4. Behavioral results.** Depiction of hitrates and reaction times of correct responses for (A)  
839 Face blocks and (B) House blocks. Equivalent conditions in different block types are marked in red  
840 and grey, respectively. Asterisks indicate significant differences based on Wilcoxon signed-rank  
841 tests within block type ( $n = 52$ ; Bonferroni-Holms corrected for multiple comparisons). Error bars  
842 indicate standard deviation. SCR – scrambled Mooney stimuli.





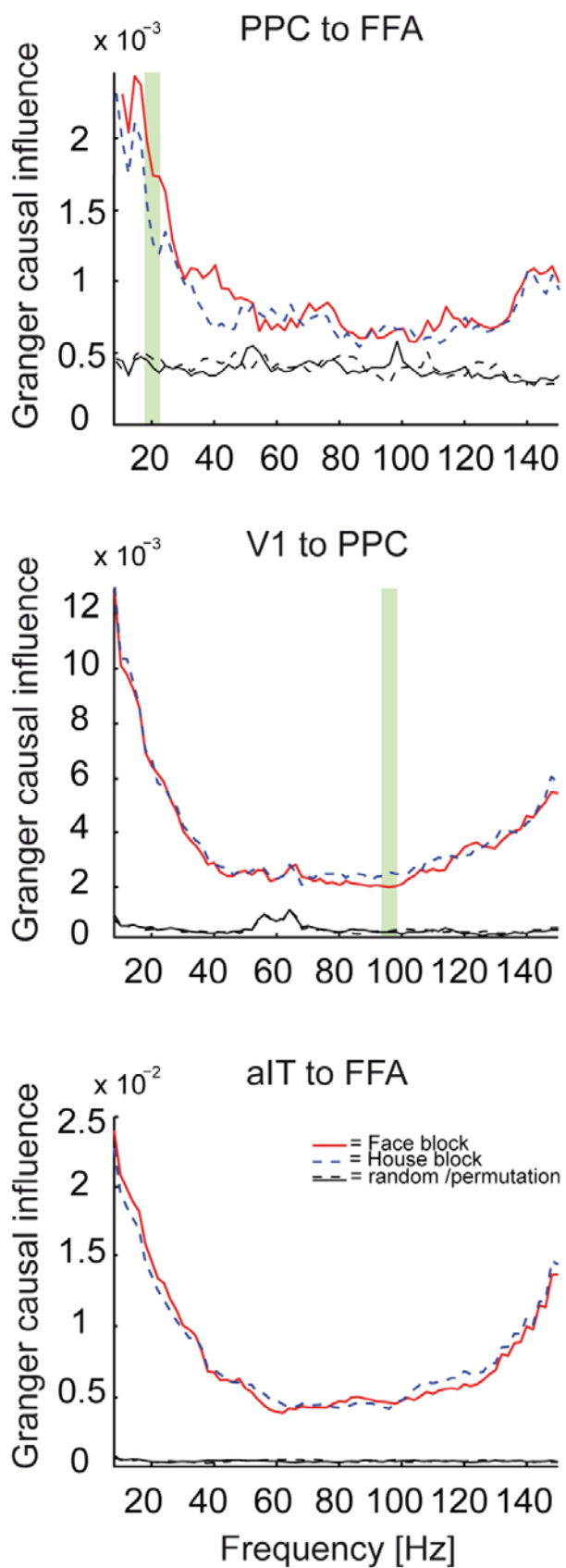
844 **Figure 5. Statistical analysis of predictable information (measured by AIS) at the MEG**  
845 **source level.** Results of whole-brain dependent samples permutation t-metric contrasting Face  
846 blocks and House blocks (n=52, t-values masked by  $p < 0.05$ , cluster correction). Peak voxel  
847 coordinates in MNI space are shown at the top for each brain location; z-values are displayed  
848 below each brain slice. OFA = occipital face area; FFA = fusiform face area; aIT= anterior inferior  
849 temporal cortex; PPC = posterior parietal cortex; V1 = primary visual cortex

850



851

852 **Figure 6. Analysis of information transfer in the prestimulus interval.** Results of dependent  
853 sample permutation t-tests on transfer entropy (TE) values (Face blocks vs House blocks,  $n = 52$ ,  
854  $p < 0.05$ , fdr corrected). Red arrows indicate increased information transfer for Face blocks, blue  
855 arrows indicate increased information transfer for House blocks. Illustration of the resulting network  
856 in A) a view from the back of the brain, B) view from the top of the brain, C) depiction of the  
857 network hierarchy (based on the hierarchy in (Zhen et al., 2013; Michalareas et al., 2016)).



858

859 **Figure 7. Frequency resolved Granger causality – post-hoc test for TE analysis.** Granger

860 causality for Face blocks and House blocks as well as random/permutation conditions. Green  
861 shaded regions indicate significant differences between Face and House blocks with cluster  
862 correction (dependent samples permutation t-test,  $n = 52$ ,  $p < 0.05$ ). Frequency ranges were only  
863 considered as significant, if granger values for both block types in these frequencies also  
864 significantly differed from the random conditions (dependent samples permutation t-tests,  $n = 52$ ,  
865 Bonferroni-Holms correction).

866