

1 Title: Prefrontal cortex activation reflects efficient exploitation of higher-order statistical structure

2 Abbreviated title: Exploitation of 1<sup>st</sup>- and 2<sup>nd</sup>-order information

3

4 Authors: Christiane Ahlheim<sup>1,2</sup>, Anne-Marike Schiffer<sup>3,4</sup>, Ricarda I. Schubotz<sup>1,2</sup>

5 <sup>1</sup>Institute of Psychology, Westfälische Wilhelms-Universität, 48149 Münster, Germany

6 <sup>2</sup>Max Planck Institute for Neurological Research, 50931 Cologne, Germany

7 <sup>3</sup>Department of Experimental Psychology, University of Oxford, OX1 3UD Oxford, United Kingdom

8 <sup>4</sup>Department of Life Sciences, Brunel University, Uxbridge, Middlesex UB8 3PH, UK

9

10 Corresponding author:

11 Christiane Ahlheim

12 Fliegerstr. 21,

13 48149 Münster

14 c.ahlheim@ucl.ac.uk

15

16

17

18 **This manuscript was accepted for publication at Journal of Cognitive**

19 **Neuroscience, June 2, 2016**

20 **<http://www.mitpressjournals.org/loi/jocn>**

21

22 Abstract

23 Since everyday actions are statistically structured, knowing which action a person has just completed allows  
24 predicting the most likely next action step. Taking even more than the preceding action into account improves  
25 this predictability, but also causes higher processing costs. Using fMRI, we investigated whether observers  
26 exploit 2<sup>nd</sup>-order statistical regularities preferentially if information on possible upcoming actions provided by  
27 1<sup>st</sup>-order regularities is insufficient. We hypothesized that anterior prefrontal cortex balances whether or not 2<sup>nd</sup>-  
28 order information should be exploited. Participants watched videos of actions that were structured by 1<sup>st</sup>- and  
29 2<sup>nd</sup>-order conditional probabilities. Information provided by the 1<sup>st</sup> and by the 2<sup>nd</sup> order was manipulated  
30 independently. BOLD activity in the action observation network was more attenuated the more information on  
31 upcoming actions was provided by 1<sup>st</sup>- order structure, reflecting expectation suppression for more predictable  
32 actions. Activation in posterior parietal sites decreased further with 2<sup>nd</sup>-order information, but increased in  
33 temporal areas. As expected, 2<sup>nd</sup>-order information was integrated more when less 1<sup>st</sup>-order information was  
34 provided, and this interaction was mediated by anterior prefrontal cortex (BA 10). Observers spontaneously  
35 used both the present and the preceding action to predict the upcoming action, and integration of the preceding  
36 action was enhanced when the present action was uninformative.

37

38 Keywords: action observation, anterior prefrontal cortex, BA 10, information theory, statistical learning

39

## 40 1. Introduction

41 Humans use knowledge about structural regularities to shape their expectations about upcoming events  
42 (Bubic, von Cramon, & Schubotz, 2010; Friston & Kiebel, 2009; Kok, Brouwer, van Gerven, & de Lange,  
43 2013; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008; Turk-Browne, Scholl, Johnson, & Chun,  
44 2010). A good example of this ability is action observation: actions provide a conditional structure of sequential  
45 action steps, so that knowing about a preceding action step improves predictability of the upcoming action  
46 (Zacks, Kurby, Eisenberg, & Haroutunian, 2011). Therefore, it appears that the more preceding action steps an  
47 observer takes into account, the more accurate the prediction will be. For instance, we do expect that a person  
48 will put a tea bag into a mug after switching on a kettle, but we do not if we observed that person putting a  
49 descaler into the kettle right before. Here, the 1<sup>st</sup>-order conditional probability of “putting a tea bag in a mug”  
50 after observing “switching on a kettle” is modulated by taking one additional previous action step into account,  
51 which constitutes a 2<sup>nd</sup>-order conditional probability. However, retrieving this 2<sup>nd</sup>-order information comes with  
52 processing costs, and may thus not always be worth the investment. This leads to the question: do observers  
53 always consider as many preceding action steps as possible to optimize their predictions, or do they only do so if  
54 their expectation is hardly informed by the directly preceding action? We know that humans do not take into  
55 account all available sources of information to make optimal decisions, but often jump to conclusions, taking  
56 heuristic shortcuts (Gigerenzer & Goldstein, 1996). A basic question in human cognition concerns this cost-  
57 benefit ratio: How much information processing is invested (as a cost) to optimize expectations and behavior (as  
58 a benefit)?

59 Behavioral and functional MRI (fMRI) findings strongly suggest predictive mechanisms are engaged  
60 during action observation. Humans are particularly fast and accurate at recognizing actions, even if visual  
61 information is sparse (Blake & Shiffrar, 2007) or parts of the action are occluded (Stadler, Schubotz, & von  
62 Cramon, 2011; Zacks et al., 2011). The so-called action observation network (AON), including premotor cortex,  
63 inferior parietal lobule, and posterior temporo-occipital regions (Caspers, Zilles, Laird, & Eickhoff, 2010) shows  
64 reduced activation for expected compared to unexpected actions (expectation suppression, see Summerfield &  
65 de Lange 2014; Summerfield et al. 2008). For instance, AON activation is attenuated by previous encounters of  
66 an action (Schiffer, Ahlheim, Ulrichs, & Schubotz, 2013), successful inference of action goals (Wurm, Hrkać,  
67 Morikawa, & Schubotz, 2014), or predictive regularities between action steps (Ahlheim, Stadler, & Schubotz,  
68 2014; Schubotz, Wurm, Wittmann, & von Cramon, 2014). This shows that the human brain exploits previous  
69 action steps to prepare for upcoming action steps. However, it is so far unknown how many previous action

70 steps are considered to improve predictability, and whether this occurs as a function of the uncertainty regarding  
71 the next action step.

72 In general, the predictability of an upcoming event depends on the degree of structure that underlies the  
73 event sequence, and knowledge of this structure allows for more accurate predictions. Using various paradigms  
74 and stimuli, it has been shown that humans spontaneously learn about 1<sup>st</sup>-order structures defined by conditional  
75 probabilities between successive items, which can be accessed directly through pairwise associations. Humans  
76 use knowledge of those probabilities to prepare for upcoming stimuli, both in abstract stimulus sequences as  
77 well as actions (Ahlheim et al., 2014; Baldwin, Andersson, Saffran, & Meyer, 2008; Fiser & Aslin, 2002;  
78 Swallow & Zacks, 2008; Turk-Browne, Scholl, Chun, & Johnson, 2009). However, most everyday actions are  
79 not guided by simple 1<sup>st</sup>-order conditional probabilities, but involve higher-order (e.g., 2<sup>nd</sup>-order structures).  
80 Contrary to 1<sup>st</sup>-order information, 2<sup>nd</sup>-order information cannot be assessed directly, but requires retrieving  
81 information about the event  $t-2$  from memory, and integrating it with the 1<sup>st</sup>-order information. This integration  
82 is necessary, as the event  $t-2$  alone does not constitute the 2<sup>nd</sup> order, but only in combination with the event  $t-1$ .  
83 While the beneficial effects of 1<sup>st</sup>-order regularities on neural processing and behavior are uncontroversial, it  
84 remains unclear whether and how 2<sup>nd</sup>-order regularities influence behavior and prediction of upcoming events,  
85 and how this depends on concurrently available 1<sup>st</sup>-order information. Findings are mixed, as some studies do  
86 not show an effect of higher-order structures (Gureckis & Love, 2010), while others show that learning of  
87 higher-order structures is slower (Remillard, 2008), or not different from 1<sup>st</sup>-order learning (Domenech &  
88 Dreher, 2010). Research in amnesic patients revealed a specific deficit in the learning of higher-order  
89 conditional structures, whereas learning of 1<sup>st</sup>-order associations remained intact (Curran, 1989). This suggests  
90 that the hippocampal formation, which is frequently damaged in amnesia, specifically contributes to learning of  
91 higher-order compared to lower-order structures, additionally to its critical role in episodic memory and  
92 associative knowledge (Fortin, Agster, & Eichenbaum, 2002; Kumaran & Maguire, 2009; Strange & Dolan,  
93 2001).

94 In order to account for the mixed findings on learning of higher-order structures, it has been suggested  
95 that humans are biased towards attending to lower-order structures, and only attend to higher-order structures if  
96 the information provided by the lower-order structure is insufficient to reliably predict the upcoming event  
97 (Gureckis & Love, 2010). It is so far unclear whether the same principle holds for action observation, and which  
98 neural structures could underpin this process of integration of predictive information. Recent findings indicate  
99 that the search and use of further information is orchestrated by the lateral BA 10 (Badre, Doll, Long, & Frank,  
100 2012; Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006). Badre et al. (2012) showed that activation in the BA



101 10 increases with relative uncertainty about a potential action outcome, but only in participants that showed a  
102 so-called explorative behavior, i.e. participants that were searching for additional information from unknown  
103 choices. This links the BA 10 to explorative choice. In a similar vein, Daw et al. (2006) showed that activation  
104 in the lateral BA 10 is higher for explorative, or information-gathering, choices. Exploration can be understood  
105 as search for information, and higher activation in the BA 10 is also frequently observed during episodic or  
106 source memory retrieval tasks (Ramnani & Owen, 2004), that is, when information needs to be gathered from  
107 memory. Furthermore, the BA 10 has been associated with the integration of different sources of information  
108 (Nee, Jahn, & Brown, 2013).

109 In the present fMRI study, we tested the hypothesis that observers' exploitation of 2<sup>nd</sup>-order statistical  
110 information in action sequences depends on how much information was already provided by the 1<sup>st</sup> order. We  
111 used fMRI to test whether information from an observed action's 2<sup>nd</sup>-order statistical structure is used the more  
112 the less informative the action's 1<sup>st</sup>-order statistical structure is and whether this cost-efficient integration of  
113 information would be signified by BA 10 activity.

114 We presented observers with videos of action sequences structured by 1<sup>st</sup>- and 2<sup>nd</sup>-order conditional  
115 probabilities. That is, the probability of a given action step  $t$  was to a quantifiable amount determined by the  
116 preceding action step  $t-1$  (1<sup>st</sup>-order statistical structure) and to another amount by the combination of the  
117 preceding ( $t-1$ ) and the last but one preceding action step  $t-2$  (2<sup>nd</sup>-order statistical structure). Importantly, the  
118 amount of information provided by 1<sup>st</sup>- and by 2<sup>nd</sup>-order structure was varied independently. This enabled us to  
119 estimate both effects independently and also their interaction. We modeled the BOLD effect at the beginning of  
120 action  $t$  as a function of the amount of information provided by the action  $t-1$  alone and by the combination of  
121 action  $t-1$  and  $t-2$ . We expected three effects:

122 1) First, we expected to replicate findings from our previous studies (Ahlheim et al., 2014; Wurm et al.,  
123 2014), showing that facilitating the prediction of the upcoming action step leads to attenuation of activity in the  
124 AON. The more informative action  $t-1$ , the better the prediction of the upcoming action  $t$ . Accordingly, we  
125 expected the BOLD response in the action observation network to decrease with the amount of information  
126 provided by action  $t-1$ .

127 2) At the same point in time, integrating information from action  $t-2$  with information from action  $t-1$   
128 can effectively modulate expectations based on the relation between the actions  $t-1$  and  $t$ , and thereby increase  
129 predictability of action  $t$ . Unlike 1<sup>st</sup>-order information, 2<sup>nd</sup>-order information cannot be accessed through direct  
130 associations between stimuli, but requires action  $t-2$  to be retrieved from working memory and integrated with  
131 action  $t-1$ . Moreover, previous encounters of a particular combination of preceding action steps need to be

132 retrieved from long-term memory in order to derive information on upcoming actions from the combination. We  
133 expected the retrieval and integration of 2<sup>nd</sup>-order information to be reflected in the hippocampal formation, due  
134 to its role in learning of higher-order sequences (Curran, 1989; Fortin et al., 2002; Kumaran & Maguire, 2009;  
135 Strange & Dolan, 2001). Activation of the hippocampus has furthermore been found to correlate positively with  
136 amount of information provided on an upcoming event (Harrison, Duggins, & Friston, 2006). We assumed that  
137 this effect generalizes to higher-order structures and hypothesized that activation in the hippocampal formation  
138 will correlate positively with the amount of information provided by the 2<sup>nd</sup> order. Furthermore, we expected  
139 use of 2<sup>nd</sup>-order information to draw on the AON. Here, we considered two potential scenarios. First, given that  
140 the exploitation of 2<sup>nd</sup>-order information improves predictability of the upcoming action, it can be expected to  
141 result in a further attenuation of the AON, paralleling the effect of 1<sup>st</sup>-order information, and pointing towards  
142 an interpretation of AON activity as reflecting a gain in predictability. Alternatively, activation in the AON  
143 could also be expected to increase with the amount of 2<sup>nd</sup>-order information. This is because the more  
144 information is provided by the 2<sup>nd</sup>-order structure, the more the predictions based on the 1<sup>st</sup>-order change and  
145 thus, integrating 2<sup>nd</sup>-order information is more demanding. This pattern would point towards sensitivity of the  
146 AON to the integration costs of 2<sup>nd</sup>-order information with the previously provided 1<sup>st</sup>-order information.

147 3) Lastly, we were particularly interested in the question as to how exploitation of 2<sup>nd</sup>-order  
148 information depends on the amount of information already provided by the 1<sup>st</sup>-order – that is, which brain areas  
149 show a stronger modulation by 2<sup>nd</sup>-order information when 1<sup>st</sup>-order is low compared to when it is high. We  
150 hypothesized that integration of 2<sup>nd</sup>-order information should be especially enhanced when action  $t-1$  alone was  
151 less informative about the upcoming action  $t$  and the need for further information is high. Thus, we expected a  
152 stronger modulation of the BOLD-signal by the 2<sup>nd</sup>-order information for trials with low compared to high 1<sup>st</sup>-  
153 order information. We expected Brodmann Area 10 at the frontal pole to show this interaction effect, as it has  
154 not only be reported to be activated by integration of information (Nee et al., 2013) but also to orchestrate  
155 uncertainty-driven search for information (Badre et al., 2012; Daw et al., 2006).

156

157

## 158 2. Methods

### 159 2.1 Participants

160 Twenty-two healthy, right-handed participants volunteered for the study and were paid 80 € for their  
161 participation. The local ethics committee of the University of Münster approved the experimental protocol and  
162 written informed consent was obtained from each participant. Three participants had to be excluded after  
163 completing the experiment, one because of poor performance in the control task (score below two SD from  
164 mean), and two because of self-reported inattentiveness and sleep during the fMRI session. All following  
165 analyses are based on the data of the remaining 19 participants (mean age  $25.35 \pm 2.13$  years, 14 females).

166

### 167 2.2 Stimuli and Task

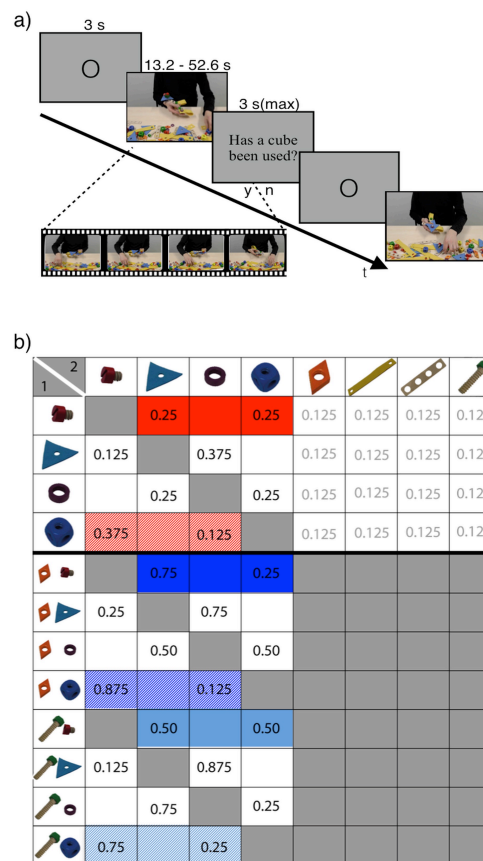
168 We employed a paradigm that required constant monitoring of sequences of action steps that were  
169 structured by 1<sup>st</sup>- and 2<sup>nd</sup>-order conditional probabilities. To construct sequential actions devoid of semantic  
170 expectations, we used eight objects from the constructional toy Baufix® and defined the grasping and  
171 manipulation of an object as one action step. Overall, we created a total of 140 action sequences, ranging from  
172 four to nine action steps. Base-rate probability of occurrence was nearly identical for all action steps, ranging  
173 from 12% to 14%. Therefore, predictions of upcoming action steps could not reliably be based on frequency.

174 To prevent participants from episodically remembering entire video clips as a basis for prediction we  
175 shot every sequence in seven versions, each with different starting scaffolds, which consisted of various  
176 different mounted objects (see Figure 1a for an illustration of the video clips).

177 Action videos were displayed on a grey background in the middle of a computer screen. A fixation  
178 circle with a duration of 3 s, or adjusted length after question trials, preceded all videos. Within the videos, onset  
179 asynchronies of the single action steps ranged from 1.28 s to 12.24 s (mean 4.39 s).

180 Approximately half of the video clips (64 of 140 during the training, 32 of 70 during the fMRI session)  
181 were followed by questions trials. Here, participants were required to answer questions concerning the previous  
182 video, e.g., “Has a long screw been used?”. Responses were given via computer mouse with the right button  
183 (i.e., middle finger of the right hand) corresponding to the answer “no” and the left button (i.e., right index  
184 finger) corresponding to “yes”. Half of the questions required a positive answer and all participants responded  
185 according to the same response contingencies. Questions were presented for 3 s or until the first response, and  
186 had to be answered within 3 s (see Figure 1a). The duration of the fixation circle following responses was  
187 adapted to compensate for different response times and could range from 2 to 5 s. Questions were followed by a  
188 feedback of 2 s indicating correct (“+”), incorrect (“-“), or delayed (“/”) responses.

Figure 1



a) Illustration of the trial course. A fixation circle preceded each video and 46% of the videos were followed by a two-alternative forced choice question. Feedback on correctness of responses was only given during the training sessions. b) Excerpt of the employed transition matrix. Rows 1-4 show 1<sup>st</sup>-order conditional probabilities between action steps, rows 5-12 show 2<sup>nd</sup>-order conditional probabilities. Objects in rows depict the preceding objects of the transition. Red marked are two examples for possible 1<sup>st</sup>-order transitions with high or low information. Transitions with high information provided by the 1<sup>st</sup>-order structure are marked with criss-cross lines (red for 1<sup>st</sup>-order conditional probabilities, light or dark blue for 2<sup>nd</sup>-order conditional probabilities). Light blue fields show exemplary transitions with low, dark blue fields with high modulatory influence of the 2<sup>nd</sup>-order structure.

### 189 2.2.1 Markov Matrix

190 The succession of action steps within the sequences followed pre-defined 1<sup>st</sup>- and 2<sup>nd</sup>-order conditional  
 191 probabilities (see Figure 1b for an excerpt of the transition matrix). First-order conditional probability refers to  
 192 the probability of each action step based on the immediately preceding action, ranging from 12.5% to 37.5%  
 193 (rows 1-4 in the transition matrix, Figure 1b). The larger the difference between probabilities of the possible  
 194 upcoming actions, the more information about the upcoming action was provided by the 1<sup>st</sup>-order structure. For  
 195 instance, the blue cube provided more 1<sup>st</sup>-order information than the short screw, as it allowed for a better  
 196 prediction of the upcoming action. Paralleling the 1<sup>st</sup>-order, the 2<sup>nd</sup>-order conditional probability refers to the  
 197 probability of each action step based on the combination of the two preceding actions, ranging from 12.5% to  
 198 87.5% (rows 5-12 in the transition matrix, Figure 1b). Here, the larger the difference between probabilities of  
 199 the possible upcoming actions, i.e., between all actions within one row of the matrix, the more information was  
 200 provided by the 2<sup>nd</sup>-order structure. For instance, if a screw nut preceded the short screw, it provided much  
 201 information on the upcoming action: the previously balanced probabilities on the 1<sup>st</sup>-order structure would  
 202 become biased, and putting the triangle would become the most likely action step. Contrary to that, a long screw  
 203 preceding the short screw provided little information, as the probability ratio between the next possible actions  
 204 stays the same. As can be seen from the matrix, the amount of information provided by the 2<sup>nd</sup>-order structure

205 varied independently of the information provided by the 1<sup>st</sup>-order structure. This feature of the statistical  
206 structure is important as it allowed us to test if the amount of information provided by the 1<sup>st</sup> order affects  
207 exploitation of the 2<sup>nd</sup> order as an additional source of information.

208

209

### 210 2.3 Experimental Procedure

211 Prior to the fMRI scan, each participant completed three 90-minute training sessions on three  
212 successive days to acquire implicit knowledge of the statistical structure. Since we wanted to test if human  
213 observers spontaneously attend to different levels of statistical structure, participants did not receive explicit  
214 learning instructions at any point either in training or during the fMRI session, and were not told that there was a  
215 certain systematic concerning the structure of the action sequences. Participants were familiarized with the eight  
216 different objects as well as with the type of question they would be asked before they started the training  
217 sessions.

218 The course of the fMRI session was identical to the training session, but no feedback was provided  
219 after question trials. To account for the limits in maximal duration of fMRI sessions, only 70 out of the 140  
220 action sequences were presented, resulting in approximately 45 minutes of fMRI scan. The selected 70  
221 sequences were a representative sample of the total set of sequences, while ensuring that rare action  
222 combinations (i.e. with low 1<sup>st</sup>- or 2<sup>nd</sup>-order conditional probabilities) occurred with sufficient frequency.

223 To test our prediction that participants would be capable of learning both 1<sup>st</sup>- and 2<sup>nd</sup>-order conditional  
224 probabilities, we implemented two post-scanner tests to assess participants' knowledge of the action syntax.

225 The first computer-based post-test was a serial reaction time task (SRTT, Nissen & Bullemer 1987)  
226 wherein pictures of the eight Baufix objects occurred at different locations on the screen. Unknown to the  
227 participants, the succession of the objects was defined by the same statistical structure as in the main  
228 experiment. Participants had to press a button, specifically assigned to each of the objects on an eight-button  
229 response pad as fast as possible. Wrong answers were followed by a negative feedback. This test was designed  
230 to test whether reaction times (RTs) would be modulated by both 1<sup>st</sup>- and 2<sup>nd</sup>-order conditional probability of the  
231 occurring object.

232 The second post-test was a paper-pencil test. Eight video clips were presented in randomized order.  
233 Videos ended after the actor had used one object and reached for another. The participants' task was to mark  
234 those objects out of the set of eight that they expected to be used next and to weight them according to their  
235 respective probability. They made this judgment in the form of eight crosses, which they could assign among the

236 eight objects. For instance, if participants saw a clip in which the long screw had been used and they expected  
237 the board and the screw nut afterwards with equal probabilities, they assigned four crosses to each of them. The  
238 number of eight crosses allowed participants to select up to all eight possible objects and to weigh them  
239 accurately (each cross corresponded to  $p = .125$ ).

240

#### 241 2.4 Data Acquisition

242 A 3T Siemens Magnetom Trio (Siemens, Erlangen, Germany) system equipped with a standard  
243 birdcage head coil was used in the functional imaging session. Participants lay supine in the scanner and their  
244 right hand was placed on a four-button response-box. Index and middle finger were placed on the response  
245 buttons and response contingencies were the same as in the training sessions. Participants' heads and arms were  
246 stabilized using form-fitting cushions, and earplugs were provided to attenuate scanner noise. The experiment  
247 was presented via a mirror that was built into the head coil and adjusted individually to provide a good view of  
248 the entire screen.

249 During the functional imaging, 28 axial slices (128.8 mm field of view, 4 mm thickness, 0.6 mm  
250 spacing; in-plane resolution of 3x3 mm) parallel to the bi-commissural line (AC-PC) were collected using a  
251 single-shot gradient echo-planar (EPI) sequence (2000 ms repetition time; echo time 30 ms, flip angle 90°, serial  
252 recording, 1260 repetitions) blood-oxygenation level-dependent (BOLD) contrast. After the functional imaging,  
253 28 slices of anatomical T1-weighted MDEFT images (4 mm thickness, 0.6 mm spacing) were acquired.

254 High-resolution 3D T1-weighted whole brain MDEFT sequences (128 sagittal slices, 1 mm thickness)  
255 were recorded for each participant in a separate session for improved localization of activation foci. Functional  
256 data were offline motion-corrected using the Siemens motion protocol PACE (Siemens, Erlangen, Germany).  
257 Further processing was conducted with the LIPSIA software package, version 2.1 (Lohmann et al., 2001). To  
258 correct for temporal offsets between the slices acquired in one scan, a cubic-spline interpolation was used. To  
259 remove low-frequency signal changes and baseline drifts from the BOLD signal, we applied a high-pass filter of  
260 1/89 – 1/70 Hz, defined by an algorithm implemented in the Lipsia software package. Functional data slices  
261 were aligned with a 3D stereotactic coordinate system. The matching parameters (six degrees of freedom, three  
262 rotational, three translational) of the T1-weighted 2D-MDEFT data onto the individual 3D-MDEFT reference  
263 set were calculated. These parameters were used in a transformation matrix for a rigid spatial registration,  
264 normalized to a standardized Talairach brain size ( $x = 135$ ,  $y = 175$ ,  $z = 120$  mm; Talairach & Tournoux, 1988)  
265 by linear scaling. Thereafter the normalized transformation matrices were applied to the functional slices in  
266 order to transform them using trilinear interpolation and align them with the 3D-reference set in the stereotactic

267 coordinate system. The spatial resolution of the resulting functional data was 3 mm \* 3 mm \* 3 mm (27 mm<sup>3</sup>).

268 A spatial Gaussian filter of 8 mm full width at half maximum (FWHM) was applied to the data.

269

## 270 2.5 Data Analyses

### 271 2.5.1 Information Theoretical Modeling

272 To operationalize the amount of information provided by the 1<sup>st</sup> and 2<sup>nd</sup> order, respectively, we used  
273 measures derived from information theory and an *ideal observer model* to estimate conditional probabilities of  
274 action steps (cf. Ahlheim et al. 2014; Bornstein & Daw 2012; Harrison et al. 2006; Strange, Duggins, Penny,  
275 Dolan, & Friston, 2005). Therefore, simulated probabilities were calculated across the training session, and  
276 continued through the scanning session. The base probabilities ( $p$ ) of single items were calculated as the number  
277 of occurrences  $n$  of item  $x_i$  divided by the sum of all items  $x_i$  that have appeared so far (see equation 1).  
278 Conditional probabilities were calculated by dividing the probability of co-occurrence of two items by the  
279 preceding item's base probability (see equation 1b); this formula was extended for the case of 2<sup>nd</sup>-order  
280 conditional probabilities.

281

$$282 \quad p(x_i) = \frac{n(x_i) + 1}{\sum_i x_i + 1}$$

283 Equation 1a. Calculation of base probabilities.

$$284 \quad p(x_t | x_{t-1}) = \frac{p(x_t \cap x_{t-1})}{p(x_{t-1})}$$

285 Equation 1. Calculation of 1<sup>st</sup>-order conditional probabilities.

286

287 The amount of information provided by an event can be quantified as the degree to which uncertainty  
288 about an upcoming event is reduced. Uncertainty can be represented as entropy ( $H$ ) (Equation 2), which is  
289 higher when unexpected events are probable (Cover & Thomas, 1991; Shannon, 1948). Entropy is therefore also  
290 referred to as expected surprise. The surprise of an event is defined as the negative logarithm of its probability,  
291 i.e. the surprise of an event is higher if the event was less likely. Formally, entropy is maximal if all possible  
292 events are equally likely to occur, so that  $p_{\text{event}} = 1/n_{\text{events}}$ . On the 1<sup>st</sup> order, the entropy about possible upcoming  
293 events (members of  $X$ ) after occurrence of one other event (member  $x_{t-1}$  of all  $X$ ) can be quantified as forward  
294 entropy (Ahlheim et al. 2014; Bornstein & Daw 2012, see Equation 3). If the forward entropy  $H(X|x_{t-1})$  is  
295 smaller than the general entropy  $H(X)$ , occurrence of  $x_{t-1}$  provided information about the occurrence of  $X$ . This



296 information  $I_1$  can be quantified as the difference between the general entropy  $H(X)$  and the forward entropy  
297 (taking the preceding event into account, i.e.,  $H(X|x_{t-1})$ ). The same logic applies to information provided by the  
298 2<sup>nd</sup> order  $I_2$ , which can be quantified as the difference between the 1<sup>st</sup>-order forward entropy  $H(X|x_{t-1})$  and the  
299 2<sup>nd</sup>-order forward entropy  $H(X|x_{t-1}, x_{t-2})$  (Equation 4). To ensure that differences between 1<sup>st</sup>- and 2<sup>nd</sup>-order  
300 forward entropy were not driven by different 1<sup>st</sup>-order conditional probabilities, we normalized the forward  
301 entropy by the 1<sup>st</sup>-order probability of co-occurrence.

$$303 \quad H(X) = \sum_i p(x_t^i) * -\log p(x_t^i)$$

304 Equation 2. Calculation of the general entropy.

$$306 \quad H(X|x_{t-1}) = p(x_{t-1}) \sum_i p(x_t^i | x_{t-1}) * -\log p(x_t^i | x_{t-1})$$

307 Equation 3. Calculation of the 1<sup>st</sup>-order forward entropy.

$$309 \quad H(X|x_{t-1}, x_{t-2}) = p(x_{t-1}, x_{t-2}) \sum_i p(x_t^i | x_{t-1}, x_{t-2}) * -\log p(x_t^i | x_{t-1}, x_{t-2})$$

310 Equation 4. Calculation of the 2<sup>nd</sup>-order forward entropy.

## 312 2.5.2 Behavioral Analysis of post-fMRI Tests

313 The behavioral analysis was conducted with the statistic software package R, version 3.1 (R  
314 Foundation for Statistical Computing, Vienna, Austria) and SPSS statistics version 22 (SPSS Inc. Chicago,  
315 Illinois, USA). If not indicated otherwise, all inferential decisions were based on an alpha level of .05.

### 317 3) SRTT Analysis

318 The first post-fMRI test, the SRTT, was designed to measure whether RTs were modulated by 1<sup>st</sup>- and  
319 2<sup>nd</sup>-order conditional probability. This would provide evidence for implicit learning of the respective orders. To  
320 test for this, we conducted a multiple regression analysis separately for each participant, which included the  
321 predictors of 1<sup>st</sup>-order conditional probability and 2<sup>nd</sup>-order conditional probability (see Equation 1) as well as  
322 the trial number to control for general learning effects. Using multiple regressions enables us to identify how  
323 much each predictor contributes to the observed data in the context of the simultaneously available predictors.



324 Only correct trials with an RT between 100 ms and 2000 ms were included in the analysis. On average, 7 % (45  
325 of 651 trials) were excluded per participant. One participant had to be excluded due to excessively prolonged  
326 RTs ( $z > 2$ ), resulting in 18 participants in the final analysis of the SRTT. To account for the non-normal  
327 distribution of the RT data, all RTs were logarithmized prior to analysis. For each participant, we obtained one  
328 standardized regression coefficient that reflected how strongly their RTs were modulated by the 1<sup>st</sup>-order  
329 conditional probabilities, and one that reflected how strongly RTs were modulated by 2<sup>nd</sup>-order conditional  
330 probabilities, while controlling for effects of the respective other predictor. Those standardized regression  
331 coefficients were tested for significant deviation from zero, using separate one-sample *t*-tests (cf. Bornstein &  
332 Daw, 2012 for a similar approach).

333

## 334 2) Paper-Pencil Analysis

335 The second post-fMRI test was a paper-pencil test where we assessed participants' explicit knowledge  
336 of the 1<sup>st</sup>-order structure. One participant failed to complete the post-test and was thus excluded from the  
337 analysis. We aggregated the number of crosses for the underlying true probability level (0, 12.5, 25, 37.5), for  
338 instance, how many crosses a participant distributed on average for a 0.25 conditional probability between  
339 action steps. This data was entered into a univariate ANOVA with the factor PROBABILITY (0, 0.125, 0.25,  
340 0.375) to test for significant differences between the levels. To test for the expected increase of probability  
341 ratings with implemented probabilities, planned paired *t*-tests between the successive probability levels were  
342 conducted.

343

### 344 2.5.3 fMRI Data analysis

345 For the statistical evaluation of the BOLD signal, a design matrix was generated modeling events with  
346 a delta (stick) function, convolved with the hemodynamic response function (gamma function; Glover 1999).  
347 All modeled actions had a minimal inter-stimulus-interval of 2 seconds. The first two actions of each sequence  
348 were discarded, as 2<sup>nd</sup>-order information was not available for those. The general linear model included five  
349 regressors, which were modeled time-locked to the onset of the action steps and with a duration of 1 s. Onset of  
350 action steps was defined as the moment the hand started to reach towards the next object. The first regressor  
351 served as a baseline and was modeled with an amplitude of 1.

352 To model information provided by the 1<sup>st</sup> order, we included a parametric regressor in which entries in  
353 the amplitude vector corresponded to the amount of information provided by the 1<sup>st</sup> order ( $I_1$ ). Paralleling this  
354 account, we included another parametric regressor in which entries in the amplitude vector corresponded to the

355 amount of information provided by the 2<sup>nd</sup> order ( $I_2$ ). To assess whether exploitation of the 2<sup>nd</sup>-order information  
356 depended on whether the 1<sup>st</sup>-order structure provided more or less information, we constructed an additional  
357 parametric regressor which modeled only those events for which the amount of information provided by the 1<sup>st</sup>  
358 order fell within the 1<sup>st</sup> or 4<sup>th</sup> quartile of the distribution of information provided by the 1<sup>st</sup> order (lowest and  
359 highest 25%). The amplitude entries on this regressor corresponded to the interaction term of 1<sup>st</sup>- and 2<sup>nd</sup>-order  
360 information, calculated as their mean-centered product (see Figure 2a for an illustration for the course of the  
361 parametric regressors during an excerpt of the experiment).

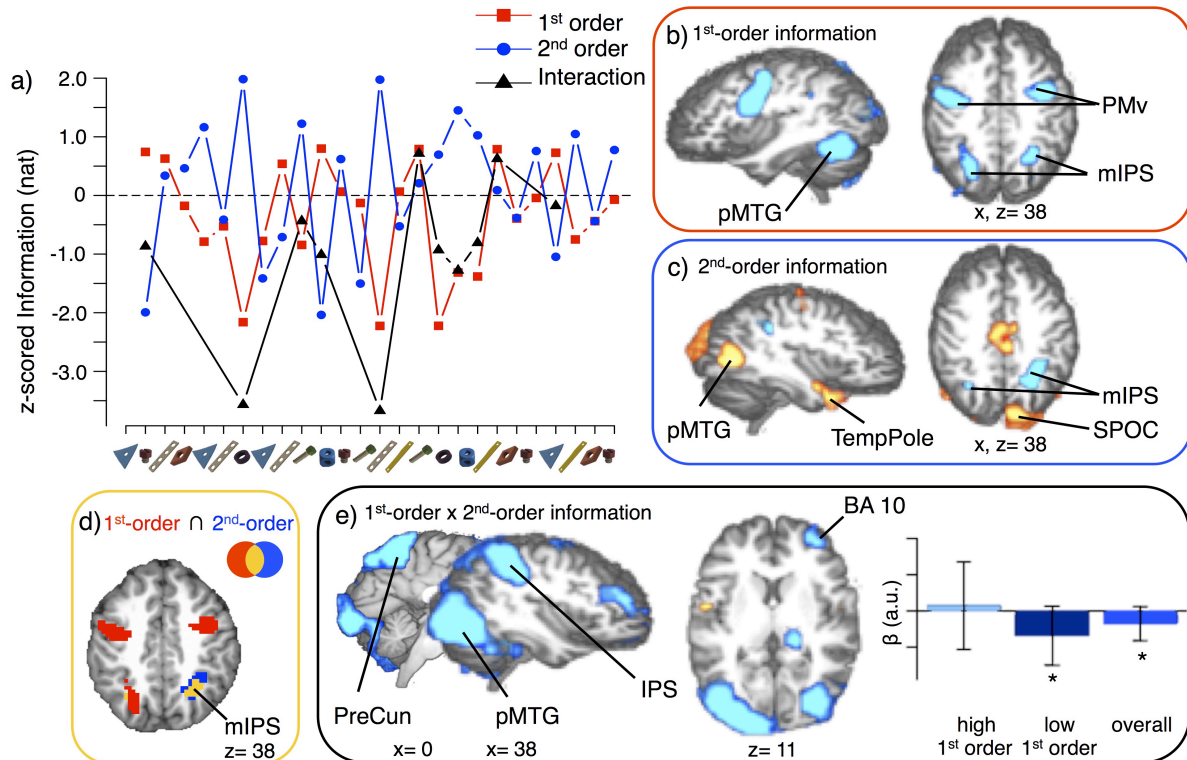
362 In addition to the parameters modeling amount of provided information, we included the 1<sup>st</sup>-order  
363 conditional surprise, i.e., the negative logarithm of each action step's conditional probability, as a nuisance  
364 regressor. Amplitudes of all parametric regressors were separately z-scored for each participant.

365 To account for question trials and general effects of observing actions, we included question trials with  
366 a duration of 3 s and video clips with a duration according to the duration of the video, both with an amplitude  
367 of 1.

368 We corrected for multiple comparisons by applying a two-step correction approach, resulting in a  
369 correction at  $p < .05$  at the cluster level. In the first step, an initial  $z$ -threshold of 2.57 ( $p < .01$ , two-tailed) was  
370 defined. All voxels showing activation above this threshold entered the second step of the correction. Here, a  
371 Monte Carlo simulation was used to define thresholds for cluster-size and cluster-value at a significance level of  
372  $p < .05$ . The combination of cluster size and cluster value decreases the risk of neglecting true activations in  
373 small structures. Thus, all reported activations were significant at  $p < .05$ , corrected for multiple comparisons at  
374 the cluster level.

375

Figure 2



a) Example course of the parametric regressors for 1<sup>st</sup>-order information (red), 2<sup>nd</sup>-order information (blue), and their interaction term (black) during an excerpt of the experiment. b) Parametric effects of the amount of information provided by the 1<sup>st</sup>-order statistical structure. PMv: ventral premotor cortex, mIPS: midposterior intraparietal sulcus, pMTG: posterior middle temporal gyrus. c) Parametric effects of the amount of information provided by the 2<sup>nd</sup>-order statistical structure. mIPS: midposterior intraparietal sulcus, pMTG: posterior middle temporal gyrus, SPOC: superior parieto-occipital cortex, TempPole: temporal pole. d) Overlay of the parametric effects of the 1<sup>st</sup>- and 2<sup>nd</sup>-order statistical structure in observed action videos. Effects of 1<sup>st</sup>-order information are displayed in red, 2<sup>nd</sup>-order in blue. Effects of both parameters overlapped in the midposterior intraparietal sulcus (yellow) and comprised 1188mm<sup>3</sup> (59.46% of the activation cluster revealed in the 1<sup>st</sup>-order contrast) in the right and 432mm<sup>3</sup> (5.05%) in the left hemisphere. e) Interaction of parametric effects of the amount of information provided by the 2<sup>nd</sup>-order statistical structure and the amount of information provided by the 1<sup>st</sup>-order structure. The bar chart depicts beta-values in the BA 10 when the interaction term modeled only events with high 1<sup>st</sup>-order information (light blue,  $t(18) = -0.18$ ,  $p = .855$ ), low 1<sup>st</sup>-order information (dark blue,  $t(18) = -3.12$ ,  $p = .006$ ), and the interaction effect when events with high or low 1<sup>st</sup>-order information were modeled (middle blue,  $t(18) = -3.41$ ,  $p = .003$ ). Error bars depict  $\pm 1$  standard deviation. pMTG: posterior middle temporal gyrus, IPS: intraparietal sulcus, preCun: precuneus.

377

378

379 3. Results

380 Participants answered on average 26.4 out of 32 question trials correctly ( $SD = 3.27$ ), indicating a high  
381 attentiveness during the fMRI session.

382

383 3.1 Behavioral Results

384 3.1.1 Results of the post-fMRI SRTT

385 The multiple regression testing for effects of the 1<sup>st</sup>-order and 2<sup>nd</sup>-order conditional probabilities on the  
386 logarithmized RTs revealed a significant negative relationship between 1<sup>st</sup>-order conditional probability and  
387 RTs, showing that higher 1<sup>st</sup>-order probabilities led to faster RTs ( $t(17) = -6.92, p < .001$ , two-tailed,  $M = -0.12$ ,  
388  $SD = 0.07$  of the standardized coefficients). This effect was consistent across all participants. The effect of the  
389 2<sup>nd</sup>-order conditional probability was also significant ( $t(17) = 2.37, p = .030$ , two-tailed,  $M = 0.03, SD = 0.06$ ),  
390 indicating slower RTs with higher 2<sup>nd</sup>-order probabilities (see Figure 3). Thirteen out of the 18 tested  
391 participants showed a positive correlation between 2<sup>nd</sup>-order conditional probabilities and RTs. As we conducted  
392 multiple regressions, those results show that RTs were slower for higher 2<sup>nd</sup>-order conditional probabilities  
393 whilst controlling for an effect of 1<sup>st</sup>-order conditional probabilities.

394 We furthermore wanted to test whether the effect of 2<sup>nd</sup>-order conditional probabilities depended on the  
395 degree to which expectations based on 1<sup>st</sup>-order conditional probabilities had been modulated by these 2<sup>nd</sup>-order  
396 conditional probabilities. To that end, we conducted a median split of the data for each participant, dividing  
397 trials by whether the 2<sup>nd</sup> order modulated the 1<sup>st</sup> order to a greater or lesser extent. We performed two multiple  
398 regressions parallel to the multiple regression described above, with 1<sup>st</sup>-order and 2<sup>nd</sup>-order conditional  
399 probability, as well as trial number, as predictors. The resulting standardized coefficients for the 2<sup>nd</sup>-order  
400 conditional probability depending on how strongly the 2<sup>nd</sup> order changed the expectations based on the 1<sup>st</sup>-order  
401 conditional probabilities were tested against each other using a paired  $t$ -test. A marginally significant difference  
402 was revealed ( $t(17) = 2.04, p = .057$ , two-sided). Thus, RTs showed a trend for being more strongly modulated by  
403 2<sup>nd</sup>-order probabilities if those modulated the expectations based on 1<sup>st</sup>-order probabilities strongly ( $M = 0.11$ ,  
404  $SD = 0.15$ ) compared to if the modulation was weak ( $M = 0.03, SD = 0.11$ ; see Figure 3).

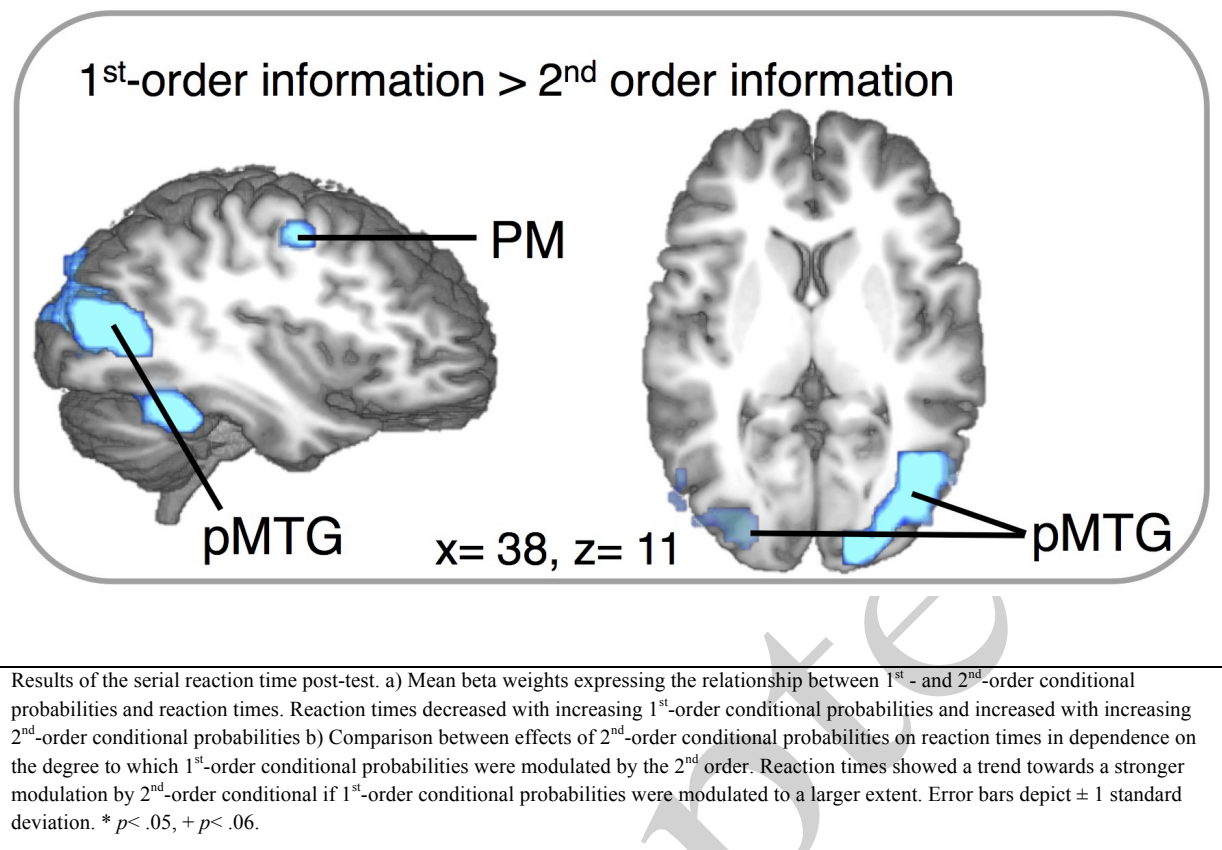
405

406

407

408

Figure 3



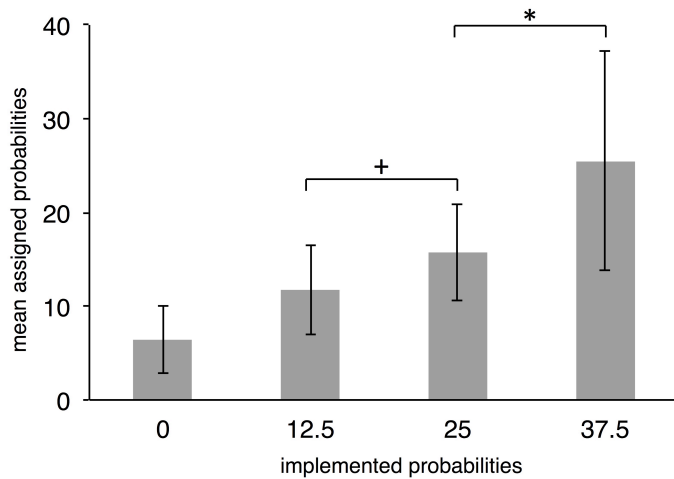
409

### 410 3.1.2 Results of the post-fMRI paper-pencil test

411 The results of the paper-pencil post-test, which assessed knowledge of the 1<sup>st</sup>-order structure, further  
412 corroborated the significant effect of 1<sup>st</sup>-order conditional probabilities on RTs. The repeated-measures ANOVA  
413 testing for an overall effect of the factor PROBABILITY on the assigned weight turned out significant ( $F(3,$   
414  $51) = 18.17, p < .001$ , partial  $\eta = .52$ ). As we expected rated probabilities to reflect actually implemented  
415 probabilities, planned paired  $t$ -tests were conducted between the single successive levels. We found no  
416 difference between probabilities of 0 and 0.125 ( $t(17) = 1.61, p = .063$ , one-tailed,  $d = 0.38$ ), a marginally  
417 significant difference between probabilities of 0.125 and 0.25 ( $t(17) = 2.09, p = .026$ , one-tailed,  $d = 0.49$ ) and a  
418 significant difference between 0.25 and 0.375 ( $t(17) = 3.48, p = .002$ , one-tailed,  $d = 0.82$ ), with an alpha-level of  
419  $.017$ , adjusted for the three comparisons (see Figure 4; note that the mean assigned values were scaled by the  
420 factor 12.5 to match the scaling of the implemented probabilities). This indicates that participants formed  
421 predictions based on the 1<sup>st</sup>-order conditional probabilities, and that their representation of 1<sup>st</sup>-order conditional  
422 probabilities was more precise for higher probability values. None of the participants claimed conscious  
423 knowledge of the structure when interviewed after the experiment.

424

Figure 4



Results of the paper-pencil post-test, showing that assigned probabilities increased as implemented probabilities increased. Number of assigned crosses was multiplied by 12.5 to achieve same scaling as underlying probabilities. Error bars depict  $\pm 1$  standard deviation. \*  $p < .017$ , +  $p < .03$ .

425

### 426 3.2 fMRI Results

427 Manipulating the amount of information provided by the 1<sup>st</sup> and 2<sup>nd</sup> order of the statistical structure  
428 independently of each other allowed us to assess functional correlates of the exploitation of each of the levels  
429 independently. Furthermore, it enabled us to investigate how the amount of information provided by the 1<sup>st</sup>  
430 order affects exploitation of further information provided by the 2<sup>nd</sup> order.

#### 431 1) Effects of 1<sup>st</sup>-order information

432 The contrast testing for a modulation of the BOLD response by the amount of information provided by  
433 the 1<sup>st</sup>-order structure yielded an attenuation of activation in the predicted network of ventral premotor cortex  
434 (PMv), the midposterior part of the intraparietal sulcus (mIPS), and the fusiform gyrus and posterior middle  
435 temporal gyrus (pMTG), which is classically reported for action observation (see Table 1 for a list of all  
436 activations, Figure 2b). Since information provided by the 1<sup>st</sup>-order structure and information provided by the  
437 2<sup>nd</sup> order were modeled simultaneously, this finding shows that increased predictability based on information  
438 provided by the 1<sup>st</sup>-order structure can reduce activation even when information from the 2<sup>nd</sup>-order structure is  
439 also available.

440

441

442

443

444 Table 1: MNI coordinates and maximal z-scores of significantly activated clusters following correction for  
 445 multiple comparison for the parametric contrast of information provided by the 1<sup>st</sup>-order structure

Localization	MNI coordinates			z-values, local maxima	Cluster size (mm <sup>3</sup> )
	x	y	z		
ventral premotor cortex	-41	1	33	-4.39	11691
	37	4	33	-4.22	9855
midposterior intraparietal sulcus	-17	-62	48	-3.99	8559
	25	-53	42	-3.38	1998
midposterior intraparietal sulcus/ Precuneus (BA 19)	13	-65	54	-2.87	567
	28	-71	22	-2.97	810
Fusiform gyrus / posterior middle temporal gyrus	-50	-59	0	-3.96	6939
	40	-50	-21	-3.06	1107

446

447 2) Effects of 2<sup>nd</sup>-order information

448 We expected 2<sup>nd</sup>-order information to draw onto activation in the AON as well, though we considered either a  
 449 positive or a negative correlation as possible.

450 Higher 2<sup>nd</sup>-order information was associated with a decrease of activation in mIPS, which overlapped  
 451 with the cluster observed in the 1<sup>st</sup>-order contrast (1188 mm<sup>3</sup> in the left, 432 mm<sup>3</sup> in the right hemisphere; see  
 452 Figure 2d for a conjunction of the two contrasts). The mIPS was the only area for which an overlap was  
 453 revealed. We found an increase in activation with higher 2<sup>nd</sup>-order information in pMTG and superior parieto-  
 454 occipital cortex (SPOC). An unhypothesized positive correlation between BOLD activation and 2<sup>nd</sup>-order  
 455 information was furthermore revealed in the right temporal pole (see Table 2 for a list of all activations, Figure  
 456 2c). Those findings show that 2<sup>nd</sup>-order information is spontaneously integrated, independent of 1<sup>st</sup>-order  
 457 information. To additionally test which areas are more sensitive towards 1<sup>st</sup>- than towards 2<sup>nd</sup>-order information,  
 458 we calculated the direct contrast between the two parametric regressors. This contrast revealed significantly  
 459 higher activation for the 2<sup>nd</sup>-order in the premotor cortex and the pMTG, showing that activation there was more  
 460 strongly attenuated by 1<sup>st</sup>-order information (see supplementary Table 1 and supplementary Figure 1).

461 To test for the hypothesized correlation between 2<sup>nd</sup>-order information and activation in the  
 462 hippocampal formation reflecting effects for retrieval of 2<sup>nd</sup>-order information, we additionally conducted an  
 463 ROI analysis in the anterior hippocampus. ROI coordinates were taken from a previous publication of our group

464 (Ahlheim et al., 2014) and were based on reported effects of sensitivity of the hippocampus to entropy  
 465 (Bornstein & Daw, 2012; Harrison et al., 2006; Strange et al., 2005). The center of the ROI in the left  
 466 hippocampus was at  $x = -25$ ,  $y = -16$ ,  $z = -18$ , and the center of the ROI in the right anterior hippocampus was  
 467 at  $x = 31$ ,  $y = -17$ ,  $z = -19$ . Both ROIs had a sphere with a radius of two adjacent voxels (6 mm). Unexpectedly,  
 468 neither ROI showed a significant modulation by 2<sup>nd</sup>-order information (all  $p > .09$ , Bonferroni-corrected alpha-  
 469 level of .025; see Table 3 for inferential statistics).

470

471 Table 2: MNI coordinates and maximal z-scores of significantly activated clusters following correction for  
 472 multiple comparison for the parametric contrast of information provided by the 2<sup>nd</sup>-order structure.

Localization	MNI coordinates			z-values, local maxima	Cluster size (mm <sup>3</sup> )
	x	y	z		
dorsal premotor cortex	28	-11	54	3.82	4725
local maximum in pCC	7	-12	39	3.58	
midposterior intraparietal sulcus	-29	-59	30	-2.91	594
posterior middle temporal gyrus	25	-50	36	-3.31	3294
	-50	-68	18	3.11	405
superior parieto-occipital cortex (BA 18)	37	-62	9	4.23	4455
	-20	-89	15	3.00	648
Temporal pole	16	-92	21	4.56	13851
	52	4	-30	3.50	4401

473

474 Table 3: Inferential statistics of hippocampal ROI analyses.

		<i>t</i> (18)	<i>p</i>
Parametric effect of 1 <sup>st</sup> -order information	left hippocampus	1.75	.097
	right hippocampus	0.42	.683
Parametric effect of 2 <sup>nd</sup> -order information	left hippocampus	1.29	.212
	right hippocampus	1.75	.096
Parametric effect of interaction term	left hippocampus	-0.25	.806
	right hippocampus	-0.07	.943

475



476 3) 1<sup>st</sup>-order dependent exploitation of 2<sup>nd</sup>-order information

477 We hypothesized that exploitation of the 2<sup>nd</sup>-order information depends on the amount of information  
478 provided by the 1<sup>st</sup>-order structure. To test this, we included an interaction term modeling only those events for  
479 which the 1<sup>st</sup>-order structure provided least information (lowest 25% of the distribution) or the most information  
480 (uppermost 25% of the distribution). The interaction therefore reveals areas that were significantly more  
481 strongly modulated by information provided by the 2<sup>nd</sup>-order structure if the 1<sup>st</sup>-order structure provided only  
482 little information about the upcoming event. We found that activation in the PMd, the IPS, the precuneus, and  
483 the occipito-temporal lobe were more strongly modulated by information provided by the 2<sup>nd</sup> order of the  
484 statistical structure when less information was provided by the 1<sup>st</sup>-order structure.

485 Additionally, the interaction contrast yielded the predicted modulation of activity in lateral BA 10. BA  
486 10 did not show a significant modulation by 2<sup>nd</sup>-order information or 1<sup>st</sup>-order information alone, which  
487 indicates that it is more strongly modulated by information provided by the 2<sup>nd</sup> order if integration of this  
488 information was actually beneficial, i.e. when the 1<sup>st</sup>-order provided less information (see Table 4 for a list of all  
489 activations, Figure 2e). As can be seen from the bar chart in Figure 2e, this interaction effect was indeed driven  
490 by the cases in which 1<sup>st</sup>-order information was low.

491 Notably, the pattern of this revealed interaction effect also held when modeling all instead of only the  
492 most (un-) informative 25% of trials (data not shown).

493

494 Table 4: MNI coordinates and maximal z-scores of significantly activated clusters following correction for  
 495 multiple comparison for the interaction contrast of information provided by the 2<sup>nd</sup>-order structure, depending  
 496 on the amount of information provided by the 1<sup>st</sup>-order structure.

Localization		MNI coordinates			z-values, local maxima	Cluster size (mm <sup>3</sup> )
		x	y	z		
anterior prefrontal cortex:	BA 10	32	52	9	-3.23	5481
	BA 11	14	50	-15	-3.82	
dorsal premotor cortex		-23	-8	60	-4.27	5076
		22	-2	57	-3.72	4428
Parietal and occipital lobe	intraparietal sulcus	-29	-44	57	-5.49	201285
		33	-40	56	-4.68	
	Precuneus	-9	-62	68	-4.90	
		13	-65	46	-4.56	
	superior parieto-occipital cortex	-15	-101	-6	-5.20	
		posterior middle	-38	-87	-13	
temporal gyrus	39	-70	-17	-4.47		
Thalamus		16	-26	12	-4.00	1080
Cerebellum		10	-71	-33	-3.03	621
Temporal pole		52	4	-30	3.50	4401

497

498

499 4. Discussion

500 While it is well established that humans use predictive information in their environment to prepare for  
501 upcoming events, it is still unclear to what extent and under which conditions they do so. It is one of the  
502 currently most urgent questions how the brain selects the sources of information to generate predictions  
503 (Blokpoel, Kwisthout, & van Rooij, 2012; Phillips, 2013). The present study investigated whether information  
504 from an action's 2<sup>nd</sup>-order statistical structure is exploited in dependence on the information provided on the 1<sup>st</sup>  
505 level; in other words, whether the brain predicts upcoming actions in a cost-benefit sensitive manner.

506 Our results show that the brain exploits 1<sup>st</sup>- as well as 2<sup>nd</sup>-order statistical information, and that it does  
507 so in a cost-benefit effective manner. Our findings are threefold: first, the information derived from the action at  
508  $t-1$  saves processing costs of the upcoming action. Second, at the same point in time, information from the  $t-2$   
509 action is additionally exploited and facilitates the observer's predictions further. And finally, information  
510 derived from the  $t-2$  action is exploited more when the last action alone is less useful in shaping expectations.

511

512 Attenuation in the action observation network based on 1<sup>st</sup>-order statistical information

513 The first aim of the present study was to replicate and expand previous findings concerning the neural  
514 correlates of an increase in predictability by the 1<sup>st</sup>-order structure in action sequences (Ahlheim et al., 2014).  
515 We established in our behavioral post-tests that human observers learned 1<sup>st</sup>-order conditional probabilities and  
516 were particularly good at discriminating between action pairs of high conditional probability, even though no  
517 participant reported noticing those regularities in a post-experimental survey.

518 Previous studies reported that valid prediction of upcoming events leads to decreased activity levels in  
519 brain areas that code for these events, and that predictive information facilitates perception (Bar, 2004; den  
520 Ouden, Kok, & de Lange, 2012; Kok, Jehee, & de Lange, 2012; Summerfield et al., 2008). We extended these  
521 findings to the case of action observation and found that an increase in the amount of 1<sup>st</sup>-order information led to  
522 the predicted attenuation of activity in the action observation network, composed of PMv, mIPS, and posterior  
523 temporal cortex (Caspers et al., 2010; Jeannerod, 2001). This shows that prediction of the upcoming action step  
524 was facilitated by information provided by the 1<sup>st</sup>-order structure. The established attenuation in this network  
525 adds to previous findings, showing that prediction-facilitating effects of 1<sup>st</sup>-order structure also occur in the  
526 presence of a 2<sup>nd</sup>-order structure.

527

528 Integration of 2<sup>nd</sup>-order statistical information

529 To test whether human observers are capable of processing the 2<sup>nd</sup>-order conditional probabilities in  
530 our paradigm, we modeled the amount of information provided by the 2<sup>nd</sup>-order structure. We found that  
531 activation of the mIPS decreased with the additional information provided by the 2<sup>nd</sup> order, on top of the  
532 decrease that mIPS showed as a function of 1<sup>st</sup>-order information. The mIPS was the only component of the  
533 AON that showed this pattern. The mIPS has been found to be a central focus of execution as well as  
534 observation of reaching movements (Vingerhoets, 2014). It is particularly interesting here that the mIPS area  
535 that we found is suggested to underlie the coupling of reaching and eye movements that is needed when we  
536 pursue visual hand input during reaching (Vesia & Crawford, 2012). Using temporally occluded targets during  
537 smooth pursuit eye movements, Lencer and co-workers (2004) found that this area bridges target occlusion,  
538 pointing to a role in anticipatory saccade tuning. In our paradigm, using 2<sup>nd</sup>-order information increases the  
539 predictability of the upcoming action step further, which allows for a more precise prediction of which object is  
540 going to be grasped next, and where this object can be found in the scene. This interpretation is in line with a  
541 recent finding showing that separable subregions of the intraparietal sulcus are modulated by processing  
542 unexpected events as well as events that require an adaptation of a currently valid predictive model (O'Reilly et  
543 al., 2013). The further attenuation of mIPS activation with 2<sup>nd</sup>-order information here reflects the further  
544 reduced processing costs of the upcoming reaching of the object, as target and direction of the reaching can be  
545 better predicted.

546 Contrary to 1<sup>st</sup>-order information, 2<sup>nd</sup>-order information could not be accessed directly through a  
547 pairwise association between action  $t-2$  and  $t$ . Instead, it was necessary to retrieve information about the action  
548 step  $t-2$  from memory and furthermore integrate this information with the information provided by the action  $t-1$   
549 on the 1<sup>st</sup> order, as the action step at  $t-2$  alone was not informative of  $t$ . Potentially, these additional processing  
550 costs could further account for the unpredicted finding of increased RTs with 2<sup>nd</sup>-order conditional probabilities  
551 in our post-fMRI SRTT: here, RTs increased with higher 2<sup>nd</sup>-order conditional probabilities whilst controlling  
552 for an effect of 1<sup>st</sup>-order conditional probabilities. Further, a trend-level effect ( $p = .057$ ) tentatively suggests that  
553 these processing costs, reflected in RT increase, is higher when 2<sup>nd</sup>-order information changed expectations  
554 based on the 1<sup>st</sup>-order conditional probabilities to a larger extent. Studies on learning of 2<sup>nd</sup>-order statistical  
555 regularities using a SRTT reported a decrease of RTs as reflection of statistical learning (Curran, 1989;  
556 Remillard, 2008). Speculating on possible reasons for the diverging results, it should be noted that our SRTT  
557 differed in a critical point from a standard SRTT: Statistical regularities among the action steps were already  
558 established at the beginning of the testing, whereas the association between observed object and button press

559 was not. How and when the processing costs of higher-order information begin to turn into a behavioral benefit  
560 thus needs to be explored further.

561 On the neural level, we expected that the retrieval of information about the action step  $t-2$ , which is  
562 necessary to assess 2<sup>nd</sup>-order information, would be reflected in an increased hippocampal activation with more  
563 2<sup>nd</sup>-order information. Yet, using an ROI analysis, we did not find evidence for an increase of activation ( $p > .09$ )  
564 with increasing information provided by the 2<sup>nd</sup>-order structure in the hippocampus. We found, however, an  
565 unhypothesized increase of activation in the right temporal pole, the more information was provided by the 2<sup>nd</sup>-  
566 order, as well as in the pMTG and the SPOC. The temporal pole is considered as “semantic hub” where  
567 semantic information about entities is processed, irrespective of their modality (Patterson, Nestor, & Rogers,  
568 2007). In particular, it decodes conceptual object properties that go beyond the object’s properties, as for  
569 instance the associated manipulation or the usual location of the object (Peelen & Caramazza, 2012).  
570 Furthermore, the temporal pole has been found to show a higher activation for initially biased perceptual  
571 decisions, and to pass this perceptual bias to visual areas (Summerfield & Koechlin, 2008). In the present study,  
572 higher 2<sup>nd</sup>-order information led to an increase in predictability of the upcoming action step and its associated  
573 object – in other words, the expectation of the upcoming action became more biased. This allows for a retrieval  
574 of semantic knowledge about the object – for instance, its shape or how it will be grasped and manipulated. We  
575 suggest that this retrieval of conceptual knowledge also drove the activation in the temporal pole in our study.  
576 Conceptual information is then passed to visual areas, i.e. the SPOC and pMTG. Area SPOC, at the mesial  
577 boundary between IPS and occipital lobe, is proposed to store internal representations of reach-to-grasp goals  
578 (Vesia & Crawford, 2012). We propose that here enhanced activation in SPOC reflects the maintenance of  
579 likely reach targets and their locations, which informs monitoring of the reaching movement in more parietal  
580 sites. Processing of this target, which is an object, is additionally enhanced in pMTG, which is a key-site of the  
581 processing man-made tools (Beauchamp & Martin, 2007). It should be noted though that we did not distinguish  
582 between different aspects of an action, that is the involved object and its manipulation. However, the amount of  
583 information provided by a certain object or action step varied depending on its position in the sequence,  
584 ensuring that the identity of the object itself could not be the cause of the effects revealed here.

585

586 Evidence for information-state dependent use of 2<sup>nd</sup>-order information

587 To test the hypothesis that exploitation of the 2<sup>nd</sup>-order statistical structure depends on the amount of  
588 information provided by the 1<sup>st</sup> order, we conducted a parametric analysis for those events on which the 1<sup>st</sup> order  
589 was of very high or low informative value and tested for an interaction effect of 1<sup>st</sup>- and 2<sup>nd</sup>-order information.

590 We found that activation in the PMd, the IPS, the pMTG, and the SPOC was more strongly modulated by the  
591 interaction term. Those areas, which have been described as the core areas of the AON (Caspers et al., 2010),  
592 were thus modulated more strongly by 2<sup>nd</sup>-order information when 1<sup>st</sup>-order information was low. This provides  
593 evidence for our hypothesis that higher-order information is preferentially used if 1<sup>st</sup>-order information is  
594 insufficient to generate precise predictions. Exploitation of 2<sup>nd</sup>-order information causes higher processing costs,  
595 as a retrieval of the action at  $t-2$  is necessary and 2<sup>nd</sup>-order information needs to be integrated with 1<sup>st</sup>-order  
596 information. Thus, we hypothesized exploitation of 2<sup>nd</sup>-order information to depend on a cost-benefit criterion:  
597 we expected 2<sup>nd</sup>-information to be used the more, the less information was provided by the 1<sup>st</sup> order. Areas  
598 implementing this cost-benefit criterion should show a correlation with the interaction term of 1<sup>st</sup>-and 2<sup>nd</sup>-order  
599 information, rather than with either main effect. We hypothesized that BA 10 implements this cost-benefit trade-  
600 off by bolstering search for additional information from the action at  $t-2$  if action  $t-1$  was of only little  
601 informative value. With the current paradigm and methods, some uncertainty remains as to whether BA 10  
602 activation reflects the cost-benefit optimized exploitation of 2<sup>nd</sup>-order information or the increased search for  
603 additional information from preceding actions.

604 In line with our hypothesis, we found that activity in the lateral BA 10 was correlated with the  
605 interaction term. This correlation resulted from a stronger correlation of activity in the BA 10 with 2<sup>nd</sup>-order  
606 information if the 1<sup>st</sup> order provided only little information, i.e. if the action step  $t-1$  did not allow for a  
607 sufficiently precise prediction of action  $t$ . Notably, significant activation of the BA 10 was only revealed in the  
608 interaction contrast. This corroborates our hypothesis that BA 10 recruitment increases under low 1<sup>st</sup>-order  
609 predictability and enhances the exploitation of 2<sup>nd</sup>-order information. Across a variety of different paradigms,  
610 BA 10 has been reported to be activated when several relations among tasks or rules have to be integrated or  
611 organized (Golde, Cramon, & Schubotz, 2010; Koechlin & Hyafil, 2007; Nee et al., 2013; Ramnani & Owen,  
612 2004; Schubotz, 2011). Here, and in line with findings from Golde et al. (2010), we showed that the BA 10 is  
613 also engaged when information derived from actions needs to be integrated. A particularly interesting parallel to  
614 our paradigm is the engagement of BA 10 in uncertainty-driven search for information, when available cues  
615 provide insufficient information (Badre et al., 2012). Whereas information in the study by Badre and co-workers  
616 (2012) could be gained by searching the environment, in the present study information was gained through  
617 retrieval of the action at  $t-2$ . Our results suggest that BA 10 may particularly contribute to a strategic retrieval of  
618 associations if these associations provide a clear gain in information. In other words, BA 10 may implement an  
619 efficiency criterion for the exploitation of higher-order information, presumably both in actions as well as in  
620 abstract stimuli.

621

622 Conclusion

623 The present findings provide several novel insights about the neurofunctional mechanisms underlying  
624 the prediction of observed action sequences. It shows that human observers spontaneously use both 1<sup>st</sup>- and 2<sup>nd</sup>-  
625 order statistical structure to predict upcoming actions, especially when little information is provided by the 1<sup>st</sup>  
626 order. In particular, 1<sup>st</sup>-order statistical information in action sequences is automatically exploited and results in  
627 a faster and more efficient processing of the upcoming action step, manifesting in smaller RTs and a significant  
628 attenuation in the action observation network, respectively. Furthermore, information provided by the 2<sup>nd</sup>-order  
629 structure is retrieved and integrated to sharpen expectations, as indicated by activation increase in the temporal  
630 pole, and by attenuation in the IPS. Findings suggest that frontolateral BA 10 moderates the retrieval and  
631 integration of 2<sup>nd</sup>-order information, in line with the emerging understanding of this brain area as a hub for  
632 strategic integration of information from various sources.

633

634 Notes

635 Supplemental material for this article is available at <https://dx.doi.org/10.6084/m9.figshare.3443633.v3>.

636

637

638

639

640 Acknowledgments

641 The authors wish to thank Christian Jessen and Anna Schubotz for experimental assistance as well as  
642 Jan Balaguer, Hannah Tickle, Sarah Weigelt, and Moritz Wurm for valuable suggestions on the manuscript.

643 This research was supported by the University of Münster, Germany, and the Max-Planck-Society.

644

- 646 Ahlheim, C., Stadler, W., & Schubotz, R. I. (2014). Dissociating dynamic probability and predictability in  
647 observed actions - an fMRI study. *Frontiers in Human Neuroscience*, 8(May), 1–13.  
648 doi:10.3389/fnhum.2014.00273
- 649 Badre, D., Doll, B. B., Long, N. M., & Frank, M. J. (2012). Rostrolateral prefrontal cortex and individual  
650 differences in uncertainty-driven exploration. *Neuron*, 73(3), 595–607. doi:10.1016/j.neuron.2011.12.025
- 651 Baldwin, D. A., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via  
652 statistical structure. *Cognition*, 106(3), 1382–1407. doi:10.1016/j.cognition.2007.07.005
- 653 Bar, M. (2004). Visual objects in context. *Nature Reviews. Neuroscience*, 5(8), 617–629. doi:10.1038/nrn1476
- 654 Beauchamp, M. S., & Martin, A. (2007). Grounding object concepts in perception and action: evidence from  
655 fMRI studies of tools. *Cortex*, (43), 461–468.
- 656 Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, 58, 47–73.  
657 doi:10.1146/annurev.psych.57.102904.190152
- 658 Blokpoel, M., Kwisthout, J., & van Rooij, I. (2012). When Can Predictive Brains be Truly Bayesian? *Frontiers*  
659 *in Psychology*, 3(November), 406. doi:10.3389/fpsyg.2012.00406
- 660 Bornstein, A. M., & Daw, N. D. (2012). Dissociating hippocampal and striatal contributions to sequential  
661 prediction learning. *European Journal of Neuroscience*, 35(7), 1011–23. doi:10.1111/j.1460-  
662 9568.2011.07920.x
- 663 Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in*  
664 *Human Neuroscience*, 4, 1–15. doi:10.3389/fnhum.2010.00025
- 665 Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and  
666 imitation in the human brain. *NeuroImage*, 50(3), 1148–1167. doi:10.1016/j.neuroimage.2009.12.112
- 667 Cover, T., & Thomas, J. (1991). Entropy, relative entropy and mutual information. In *Elements of Information*  
668 *Theory* (2nd ed., pp. 12–49). Hoboken, New Jersey: John Wiley & Sons, Inc.
- 669 Curran, T. (1989). Higher-Order Associative Learning in Amnesia Evidence from the Serial Reaction Time  
670 Task. *Journal of Cognitive Neuroscience*, 522–533.
- 671 Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory  
672 decisions in humans. *Nature*, 441(7095), 876–9. doi:10.1038/nature04766
- 673 den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and  
674 motivation. *Frontiers in Psychology*, 3(December), 548. doi:10.3389/fpsyg.2012.00548
- 675 Domenech, P., & Dreher, J.-C. (2010). Decision threshold modulation in the human brain. *The Journal of*  
676 *Neuroscience*, 30(43), 14305–17. doi:10.1523/JNEUROSCI.2371-10.2010
- 677 Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape  
678 sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 458–467.  
679 doi:10.1037//0278-7393.28.3.458
- 680 Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for  
681 sequences of events. *Nature Neuroscience*, 5(5), 458–562. doi:10.1038/nn834
- 682 Friston, K. J., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical*  
683 *Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1521), 1211–1211.  
684 doi:10.1098/rstb.2008.0300
- 685 Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality.  
686 *Psychological Review*, 103(4), 650–669. doi:10.1037/0033-295X.103.4.650
- 687 Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, 9(4),  
688 416–429.
- 689 Golde, M., Cramon, D. von, & Schubotz, R. I. (2010). Differential role of anterior prefrontal and premotor  
690 cortex in the processing of relational information. *NeuroImage*, 49(3), 2890–2900.  
691 doi:10.1016/j.neuroimage.2009.09.009



- 692 Gureckis, T. M., & Love, B. C. (2010). Direct Associations or Internal Transformations? Exploring the  
693 Mechanisms Underlying Sequential Learning Behavior. *Cognitive Science*, 34(1), 10–50.  
694 doi:10.1111/j.1551-6709.2009.01076.x
- 695 Harrison, L. M., Duggins, A., & Friston, K. J. (2006). Encoding uncertainty in the hippocampus. *Neural*  
696 *Networks*, 19(5), 535–546. doi:10.1016/j.neunet.2005.11.002
- 697 Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *NeuroImage*, 14,  
698 103–109. doi:10.1006/nimg.2001.0832
- 699 Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making.  
700 *Science (New York, N.Y.)*, 318(5850), 594–8. doi:10.1126/science.1142995
- 701 Kok, P., Brouwer, G. J., van Gerven, M. a. J., & de Lange, F. P. (2013). Prior Expectations Bias Sensory  
702 Representations in Visual Cortex. *The Journal of Neuroscience*, 33(41), 16275–16284.  
703 doi:10.1523/JNEUROSCI.0742-13.2013
- 704 Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: expectation sharpens representations in the  
705 primary visual cortex. *Neuron*, 75(2), 265–70. doi:10.1016/j.neuron.2012.04.034
- 706 Kumaran, D., & Maguire, E. A. (2009). Novelty signals: a window into hippocampal information processing.  
707 *Trends in Cognitive Sciences*, 13(2), 47–54. doi:10.1016/j.tics.2008.11.004
- 708 Lencer, R., Nagel, M., Sprenger, A., Zapf, S., Erdmann, C., Heide, W., & Binkofski, F. (2004). Cortical  
709 mechanisms of smooth pursuit eye movements with target blanking. An fMRI study. *European Journal of*  
710 *Neuroscience*, 19(5), 1430–1436. doi:10.1111/j.1460-9568.2004.03229.x
- 711 Lohmann, G., Müller, K., Bosch, V., Mentzel, H.-J., Hessler, S., Chen, L., Zysset, S., von Cramon, D. Y.  
712 (2001). LIPSIA--a new software system for the evaluation of functional magnetic resonance images of the  
713 human brain. *Computerized Medical Imaging and Graphics*, 25(6), 449–457.
- 714 Nee, D., Jahn, A., & Brown, J. (2013). Prefrontal Cortex Organization: Dissociating Effects of Temporal  
715 Abstraction, Relational Abstraction, and Integration with fMRI. *Cerebral Cortex*.  
716 doi:10.1093/cercor/bht091
- 717 Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance  
718 measures. *Cognitive Psychology*, 19, 1–32. doi:10.1016/0010-0285(87)90002-8
- 719 O'Reilly, J. X., Schüfflgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013).  
720 Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of*  
721 *the National Academy of Sciences of the United States of America*, 110(38), 3660–3669.  
722 <http://doi.org/10.1073/pnas.1305373110>
- 723 Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of  
724 semantic knowledge in the human brain. *Nature Reviews. Neuroscience*, 8(12), 976–987.  
725 doi:10.1038/nrn2277
- 726 Peelen, M. V., & Caramazza, A. (2012). Conceptual object representations in human anterior temporal cortex.  
727 *The Journal of Neuroscience*, 32(45), 15728–36. doi:10.1523/JNEUROSCI.1953-12.2012
- 728 Phillips, W. A. (2013). Neuronal inference must be local, selective, and coordinated. *Behavioral and Brain*  
729 *Sciences*, 36(03), 222–223. doi:10.1017/S0140525X12002257
- 730 Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: insights into function from anatomy and  
731 neuroimaging. *Nature Reviews. Neuroscience*, 5(3), 184–94. doi:10.1038/nrn1343
- 732 Remillard, G. (2008). Implicit learning of second-, third-, and fourth-order adjacent and nonadjacent sequential  
733 dependencies. *Quarterly Journal of Experimental Psychology (2006)*, 61(3), 400–24.  
734 doi:10.1080/17470210701210999
- 735 Schiffer, A.-M., Ahlheim, C., Ulrichs, K., & Schubotz, R. I. (2013). Neural changes when actions change:  
736 adaptation of strong and weak expectations. *Human Brain ...*, 34, 1713–1727. doi:10.1002/hbm.22023
- 737 Schubotz, R. I. (2011). Long-Term Planning and Prediction: Visiting a Construction Site in the Human Brain. In  
738 W. Welsch, W. J. Singer, & A. Wunder (Eds.), *Interdisciplinary Anthropology* (pp. 79–104). Berlin,  
739 Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-11668-1
- 740 Schubotz, R. I., Wurm, M. F., Wittmann, M. K., & von Cramon, D. Y. (2014). Objects tell us what action we  
741 can expect: Dissociating brain areas for retrieval and exploitation of action knowledge during action

- 742 observation in fMRI. *Frontiers in Psychology*, 5(JUN), 1–15. doi:10.3389/fpsyg.2014.00636
- 743 Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–  
744 423.
- 745 Stadler, W., Schubotz, R. I., von Cramon, D. Y., Springer, A., Graf, M., & Prinz, W. (2011). Predicting and  
746 memorizing observed action: differential premotor cortex involvement. *Human Brain Mapping*, 32(5),  
747 677–87. doi:10.1002/hbm.20949
- 748 Strange, B. A., & Dolan, R. J. (2001). Adaptive anterior hippocampal responses to oddball stimuli.  
749 *Hippocampus*, 11(6), 690–698. doi:10.1002/hipo.1084
- 750 Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., & Friston, K. J. (2005). Information theory, novelty and  
751 hippocampal responses: unpredicted or unpredictable? *Neural Networks*, 18(3), 225–230.  
752 doi:10.1016/j.neunet.2004.12.004
- 753 Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: neural and  
754 computational mechanisms. *Nature Reviews Neuroscience*, (October). doi:10.1038/nrn3838
- 755 Summerfield, C., & Koechlin, E. (2008). A Neural Representation of Prior Information during Perceptual  
756 Inference. *Neuron*, 59(2), 336–347. doi:10.1016/j.neuron.2008.05.021
- 757 Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Eger, T. (2008). Neural repetition  
758 suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, 11(9), 1004–6.  
759 doi:10.1038/nn.2163
- 760 Swallow, K. M., & Zacks, J. M. (2008). Sequences learned without awareness can orient attention during the  
761 perception of human activity. *Psychonomic Bulletin & Review*, 15(1), 116–122.  
762 doi:10.3758/PBR.15.1.116
- 763 Talairach, J., & Tournoux, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain*. New York: Thieme.
- 764 Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical  
765 learning: efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*,  
766 21(10), 1934–1945. doi:10.1162/jocn.2009.21131
- 767 Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit Perceptual Anticipation  
768 Triggered by Statistical Learning. *The Journal of Neuroscience*, 30(33), 11177–11187.  
769 doi:10.1523/JNEUROSCI.0858-10.2010
- 770 Vesia, M., & Crawford, J. D. (2012). Specialization of reach function in human posterior parietal cortex.  
771 *Experimental Brain Research*, 221(1), 1–18. doi:10.1007/s00221-012-3158-9
- 772 Vingerhoets, G. (2014). Contribution of the posterior parietal cortex in reaching, grasping, and using objects and  
773 tools. *Frontiers in Psychology*, 5(MAR), 1–17. doi:10.3389/fpsyg.2014.00151
- 774 Wurm, M. F., Hrkač, M., Morikawa, Y., & Schubotz, R. I. (2014). Predicting goals in action episodes attenuates  
775 BOLD response in inferior frontal and occipitotemporal cortex. *Behavioural Brain Research*, 274C, 108–  
776 117. doi:10.1016/j.bbr.2014.07.053
- 777 Zacks, J. M., Kurby, C. a, Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the  
778 perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience*, 23(12), 4057–66.  
779 doi:10.1162/jocn\_a\_00078
- 780