

Prefrontal Cortex Activation Reflects Efficient Exploitation of Higher-order Statistical Structure

Christiane Ahlheim^{1,2}, Anne-Marike Schiffer^{3,4}, and Ricarda I. Schubotz^{1,5}

Abstract

■ Because everyday actions are statistically structured, knowing which action a person has just completed allows predicting the most likely next action step. Taking even more than the preceding action into account improves this predictability but also causes higher processing costs. Using fMRI, we investigated whether observers exploit second-order statistical regularities preferentially if information on possible upcoming actions provided by first-order regularities is insufficient. We hypothesized that anterior pFC balances whether or not second-order information should be exploited. Participants watched videos of actions that were structured by first- and second-order conditional probabilities. Information provided by the first and by the second

order was manipulated independently. BOLD activity in the action observation network was more attenuated the more information on upcoming actions was provided by first-order structure, reflecting expectation suppression for more predictable actions. Activation in posterior parietal sites decreased further with second-order information but increased in temporal areas. As expected, second-order information was integrated more when less first-order information was provided, and this interaction was mediated by anterior pFC (BA 10). Observers spontaneously used both the present and the preceding action to predict the upcoming action, and integration of the preceding action was enhanced when the present action was uninformative. ■

INTRODUCTION

Humans use knowledge about structural regularities to shape their expectations about upcoming events (Kok, Brouwer, van Gerven, & de Lange, 2013; Bubic, von Cramon, & Schubotz, 2010; Turk-Browne, Scholl, Johnson, & Chun, 2010; Friston & Kiebel, 2009; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008). A good example of this ability is action observation: Actions provide a conditional structure of sequential action steps, so that knowing about a preceding action step improves predictability of the upcoming action (Zacks, Kurby, Eisenberg, & Haroutunian, 2011). Therefore, it appears that the more preceding action steps an observer takes into account, the more accurate the prediction will be. For instance, we do expect that a person will put a tea bag into a mug after switching on a kettle, but we do not if we observed that person putting a descaler into the kettle right before. Here, the first-order conditional probability of “putting a tea bag in a mug” after observing “switching on a kettle” is modulated by taking one additional previous action step into account, which constitutes a second-order conditional probability. However, retrieving this second-order information comes with processing costs and may thus not always

be worth the investment. This leads to the question: Do observers always consider as many preceding action steps as possible to optimize their predictions, or do they only do so if their expectation is hardly informed by the directly preceding action? We know that humans do not take into account all available sources of information to make optimal decisions, but often jump to conclusions, taking heuristic shortcuts (Gigerenzer & Goldstein, 1996). A basic question in human cognition concerns this cost–benefit ratio: How much information processing is invested (as a cost) to optimize expectations and behavior (as a benefit)?

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

¹Westfälische Wilhelms-Universität, Münster, Germany, ²Max Planck Institute for Neurological Research, Cologne, Germany, ³University of Oxford, ⁴Brunel University, Uxbridge, Middlesex, UK, ⁵University Hospital of Cologne, Germany
Supplemental material for this article is available at <https://dx.doi.org/10.6084/m9.figshare.3443633.v3>.

for upcoming action steps. However, it is so far unknown how many previous action steps are considered to improve predictability, and whether this occurs as a function of the uncertainty regarding the next action step.

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

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

Dolan, 2006). Badre et al. (2012) showed that activation in the BA 10 increases with relative uncertainty about a potential action outcome, but only in participants that showed a so-called explorative behavior, that is, participants that were searching for additional information from unknown choices. This links the BA 10 to explorative choice. In a similar vein, Daw et al. (2006) showed that activation in the lateral BA 10 is higher for explorative or information-gathering choices. Exploration can be understood as search for information, and higher activation in the BA 10 is also frequently observed during episodic or source memory retrieval tasks (Ramnani & Owen, 2004), that is, when information needs to be gathered from memory. Furthermore, the BA 10 has been associated with the integration of different sources of information (Nee, Jahn, & Brown, 2013).

In the present fMRI study, we tested the hypothesis that observers' exploitation of second-order statistical information in action sequences depends on how much information was already provided by the first order. We used fMRI to test whether information from an observed action's second-order statistical structure is used the more the less informative the action's first-order statistical structure is and whether this cost-efficient integration of information would be signified by BA 10 activity.

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

1. First, we expected to replicate findings from our previous studies (Ahlheim et al., 2014; Wurm et al., 2014), showing that facilitating the prediction of the upcoming action step leads to attenuation of activity in the AON. The more informative action $t - 1$, the better the prediction of the upcoming action t . Accordingly, we expected the BOLD response in the AON to decrease with the amount of information provided by action $t - 1$.
2. At the same point in time, integrating information from action $t - 2$ with information from action $t - 1$ can effectively modulate expectations based on the relation between the actions $t - 1$ and t and thereby increase predictability of action t . Unlike first-order information, second-order information cannot be accessed through direct associations between stimuli but requires action $t - 2$ to be retrieved from working memory and integrated with action $t - 1$. Moreover, previous encounters of a particular combination of preceding action

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

3. Lastly, we were particularly interested in the question as to how exploitation of second-order information depends on the amount of information already provided by the first-order—that is, which brain areas show a stronger modulation by second-order information when first-order is low compared with when it is high. We hypothesized that integration of second-order information should be especially enhanced when action $t - 1$ alone was less informative about the upcoming action t and the need for further information is high. Thus, we expected a stronger modulation of the BOLD signal by second-order information for trials with low compared with trials with high first-order information. We expected BA 10 at the frontal pole to show this interaction effect, as it has not only been reported to be activated by integration of information (Nee et al., 2013) but also to orchestrate uncertainty-driven search for information (Badre et al., 2012; Daw et al., 2006).

METHODS

Participants

Twenty-two healthy, right-handed participants volunteered for the study and were paid €80 for their participation. The

local ethics committee of the University of Münster approved the experimental protocol, and written informed consent was obtained from each participant. Three participants had to be excluded after completing the experiment, one because of poor performance in the control task (score below 2 *SD* from mean) and two because of self-reported inattentiveness and sleep during the fMRI session. All following analyses are based on the data of the remaining 19 participants (mean age = 25.35 ± 2.13 years, 14 women).

Stimuli and Task

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

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

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

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

Markov Matrix

The succession of action steps within the sequences followed predefined first- and second-order conditional

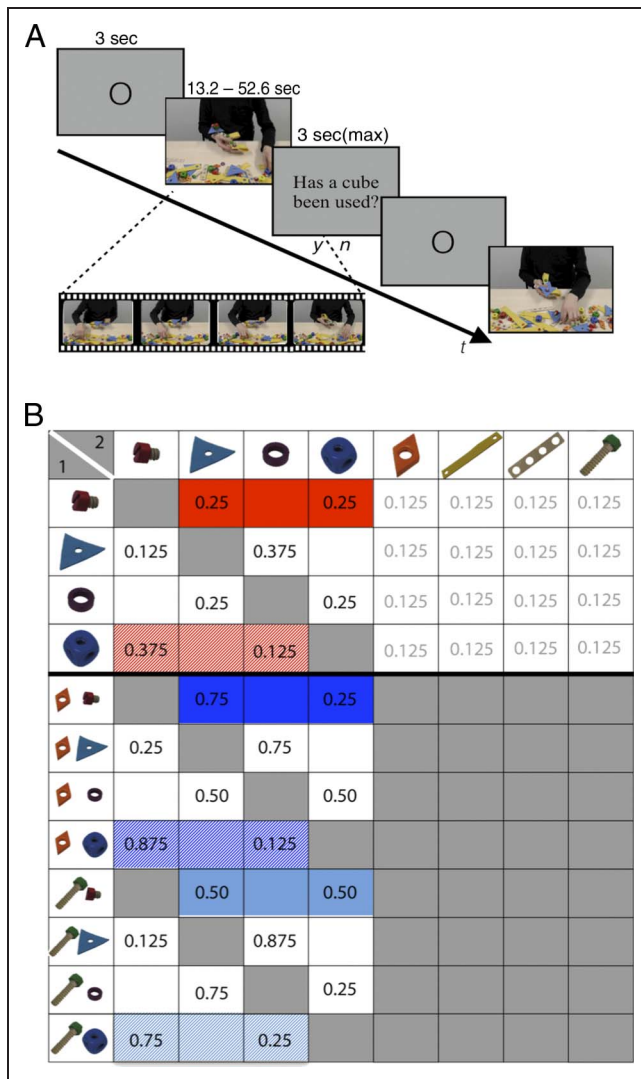


Figure 1. (A) Illustration of the trial course. A fixation circle preceded each video, and 46% of the videos were followed by a two-alternative forced-choice question. Feedback on correctness of responses was only given during the training sessions. (B) Excerpt of the employed transition matrix. Rows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

probabilities (see Figure 1B for an excerpt of the transition matrix). First-order conditional probability refers to the probability of each action step based on the immediately preceding action, ranging from 12.5% to 37.5% (rows 1–4 in the transition matrix; Figure 1B). The larger the difference between probabilities of the possible upcoming actions, the more information about the upcoming action was provided by the first-order structure. For instance, the blue cube provided more first-order infor-

mation than the short screw, as it allowed for a better prediction of the upcoming action. Paralleling the first-order, the second-order conditional probability refers to the probability of each action step based on the combination of the two preceding actions, ranging from 12.5% to 87.5% (Rows 5–12 in the transition matrix; Figure 1B). Here, the larger the difference between probabilities of the possible upcoming actions, that is, between all actions within one row of the matrix, the more information was provided by the second-order structure. For instance, if a screw nut preceded the short screw, it provided much information on the upcoming action: The previously balanced probabilities on the first-order structure would become biased, and putting the triangle would become the most likely action step. Contrary to that, a long screw preceding the short screw provided little information, as the probability ratio between the next possible actions stays the same. As can be seen from the matrix, the amount of information provided by the second-order structure varied independently of the information provided by the first-order structure. This feature of the statistical structure is important as it allowed us to test if the amount of information provided by the first order affects exploitation of the second order as an additional source of information.

Experimental Procedure

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

The course of the fMRI session was identical to the training session, but no feedback was provided after question trials. To account for the limits in maximal duration of fMRI sessions, only 70 of the 140 action sequences were presented, resulting in approximately 45 min of fMRI scan. The selected 70 sequences were a representative sample of the total set of sequences, while ensuring that rare action combinations (i.e., with low first- or second-order conditional probabilities) occurred with sufficient frequency.

To test our prediction that participants would be capable of learning both first- and second-order conditional probabilities, we implemented two postscanner tests to assess participants' knowledge of the action syntax.

The first computer-based posttest was a serial RT task (SRTT; Nissen & Bullemer, 1987) wherein pictures of the eight Baufox objects occurred at different locations on the screen. Unknown to the participants, the succession of the objects was defined by the same statistical structure

as in the main experiment. Participants had to press a button, specifically assigned to each of the objects, on an eight-button response pad as fast as possible. Wrong answers were followed by a negative feedback. This test was designed to test whether RTs would be modulated by both first- and second-order conditional probability of the occurring object.

The second posttest was a paper and pencil test. Eight video clips were presented in randomized order. Videos ended after the actor had used one object and reached for another. The participants' task was to mark those objects out of the set of eight that they expected to be used next and to weight them according to their respective probability. They made this judgment in the form of eight crosses, which they could assign among the eight objects. For instance, if participants saw a clip in which the long screw had been used and they expected the board and the screw nut afterwards with equal probabilities, they assigned four crosses to each of them. The number of eight crosses allowed participants to select up to all eight possible objects and to weigh them accurately (each cross corresponded to $p = .125$).

Data Acquisition

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

During the functional imaging, 28 axial slices (128.8 mm field of view, 4 mm thickness, 0.6 mm spacing; in-plane resolution of 3×3 mm) parallel to the bicommissural line (AC-PC) were collected using a single-shot gradient-echo planar (EPI) sequence (repetition time = 2000 msec; echo time = 30 msec, flip angle = 90° , serial recording, 1260 repetitions) BOLD contrast. After the functional imaging, 28 slices of anatomical T1-weighted MDEFT images (4 mm thickness, 0.6 mm spacing) were acquired.

High-resolution 3-D T1-weighted whole-brain MDEFT sequences (128 sagittal slices, 1 mm thickness) were recorded for each participant in a separate session for improved localization of activation foci. Functional data were offline motion-corrected using the Siemens motion protocol PACE (Siemens, Erlangen, Germany). Further processing was conducted with the LIPSIA software package, version 2.1 (Lohmann et al., 2001). To correct for temporal offsets between the slices acquired in one scan, a cubic-spline interpolation was used. To remove low-frequency signal changes and baseline drifts from the BOLD signal,

we applied a high-pass filter of 1/89–1/70 Hz, defined by an algorithm implemented in the Lipsia software package. Functional data slices were aligned with a 3-D stereotactic coordinate system. The matching parameters (six degrees of freedom, three rotational, three translational) of the T1-weighted 2D-MDEFT data onto the individual 3D-MDEFT reference set were calculated. These parameters were used in a transformation matrix for a rigid spatial registration, normalized to a standardized Talairach brain size ($x = 135$, $y = 175$, $z = 120$ mm; Talairach & Tournoux, 1988) by linear scaling. Thereafter, the normalized transformation matrices were applied to the functional slices to transform them using trilinear interpolation and align them with the 3-D reference set in the stereotactic coordinate system. The spatial resolution of the resulting functional data was $3 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm}$ (27 mm^3). A spatial Gaussian filter of 8 mm FWHM was applied to the data.

Data Analyses

Information Theoretical Modeling

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

Calculation of base probabilities:

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

Calculation of first-order conditional probabilities:

$$p(x_i|x_{i-1}) = \frac{p(x_i \cap x_{i-1})}{p(x_{i-1})} \quad (1b)$$

The amount of information provided by an event can be quantified as the degree to which uncertainty about an upcoming event is reduced. Uncertainty can be represented as entropy (H ; Equation 2), which is higher when unexpected events are probable (Cover & Thomas, 1991; Shannon, 1948). Entropy is therefore also referred to as expected surprise. The surprise of an event is defined as the negative logarithm of its probability, that is, the surprise of an event is higher if the event was less likely. Formally, entropy is maximal if all possible events are equally

likely to occur, so that $p_{\text{event}} = 1/n_{\text{events}}$. On the first order, the entropy about possible upcoming events (members of X) after occurrence of one other event (member x_{t-1} of all X) can be quantified as forward entropy (Ahlheim et al., 2014; Bornstein & Daw, 2012; see Equation 3). If the forward entropy $H(X|x_{t-1})$ is smaller than the general entropy $H(X)$, occurrence of x_{t-1} provided information about the occurrence of X . This information I_1 can be quantified as the difference between the general entropy $H(X)$ and the forward entropy (taking the preceding event into account, i.e., $H(X|x_{t-1})$). The same logic applies to information provided by the second order I_2 , which can be quantified as the difference between the first-order forward entropy $H(X|x_{t-1})$ and the second-order forward entropy $H(X|x_{t-1}, x_{t-2})$ (Equation 4). To ensure that differences between first- and second-order forward entropy were not driven by different first-order conditional probabilities, we normalized the forward entropy by the first-order probability of cooccurrence.

Calculation of the general entropy:

$$H(X) = \sum_i p(x_i^i) \times -\log p(x_i^i) \quad (2)$$

Calculation of the first-order forward entropy:

$$H(X|x_{t-1}) = p(x_{t-1}) \sum_i p(x_i^i|x_{t-1}) \times -\log p(x_i^i|x_{t-1}) \quad (3)$$

Calculation of the second-order forward entropy:

$$H(X|x_{t-1}, x_{t-2}) = p(x_{t-1}, x_{t-2}) \sum_i p(x_i^i|x_{t-1}, x_{t-2}) \times -\log p(x_i^i|x_{t-1}, x_{t-2}) \quad (4)$$

Behavioral Analysis of Post-fMRI Tests

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

SRTT analysis. The first post-fMRI test, the SRTT, was designed to measure whether RTs were modulated by first- and second-order conditional probability. This would provide evidence for implicit learning of the respective orders. To test for this, we conducted a multiple regression analysis separately for each participant, which included the predictors of first-order conditional probability and second-order conditional probability (see Equation 1b) as well as the trial number to control for general learning effects. Using multiple regressions enables us to identify how much each predictor contributes to the observed data in the context of the simultaneously available predictors. Only correct trials with an RT between 100 and 2000 msec were included in the analysis. On average, 7% (45 of 651 trials)

were excluded per participant. One participant had to be excluded because of excessively prolonged RTs ($z > 2$), resulting in 18 participants in the final analysis of the SRTT. To account for the non-normal distribution of the RT data, all RTs were logarithmized before analysis. For each participant, we obtained one standardized regression coefficient that reflected how strongly their RTs were modulated by the first-order conditional probabilities and one that reflected how strongly RTs were modulated by second-order conditional probabilities, while controlling for effects of the respective other predictor. Those standardized regression coefficients were tested for significant deviation from zero, using separate one-sample t tests (cf. Bornstein & Daw, 2012, for a similar approach).

Paper and pencil analysis. The second post-fMRI test was a paper–pencil test where we assessed participants' explicit knowledge of the first-order structure. One participant failed to complete the posttest and was thus excluded from the analysis. We aggregated the number of crosses for the underlying true probability level (0, 12.5, 25, 37.5), for instance, how many crosses a participant distributed on average for a .25 conditional probability between action steps. These data were entered into a univariate ANOVA with the factor Probability (0, .125, .25, .375) to test for significant differences between the levels. To test for the expected increase of probability ratings with implemented probabilities, planned paired t tests between the successive probability levels were conducted.

fMRI Data Analysis

For the statistical evaluation of the BOLD signal, a design matrix was generated modeling events with a delta (stick) function, convolved with the hemodynamic response function (gamma function; Glover, 1999). All modeled actions had a minimal interstimulus interval of 2 sec. The first two actions of each sequence were discarded, as second-order information was not available for those. The general linear model included five regressors, which were modeled time-locked to the onset of the action steps and with a duration of 1 sec. Onset of action steps was defined as the moment the hand started to reach toward the next object. The first regressor served as a baseline and was modeled with an amplitude of 1.

To model information provided by the first order, we included a parametric regressor in which entries in the amplitude vector corresponded to the amount of information provided by the first order (I_1). Paralleling this account, we included another parametric regressor in which entries in the amplitude vector corresponded to the amount of information provided by the second order (I_2). To assess whether exploitation of the second-order information depended on whether the first-order structure provided more or less information, we constructed an additional parametric regressor, which modeled only those events for which the amount of information provided

by the first order fell within the first or fourth quartile of the distribution of information provided by the first order (lowest and highest 25%). The amplitude entries on this regressor corresponded to the interaction term of first- and second-order information, calculated as their mean-centered product (see Figure 2A for an illustration for the course of the parametric regressors during an excerpt of the experiment).

In addition to the parameters modeling amount of provided information, we included the first-order conditional surprise, that is, the negative logarithm of each action step's conditional probability, as a nuisance regressor. Amplitudes of all parametric regressors were separately z-scored for each participant.

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

We corrected for multiple comparisons by applying a two-step correction approach, resulting in a correction at $p < .05$ at the cluster level. In the first step, an initial z threshold of 2.57 ($p < .01$, two-tailed) was defined. All

voxels showing activation above this threshold entered the second step of the correction. Here, a Monte Carlo simulation was used to define thresholds for cluster size and cluster value at a significance level of $p < .05$. The combination of cluster size and cluster value decreases the risk of neglecting true activations in small structures. Thus, all reported activations were significant at $p < .05$, corrected for multiple comparisons at the cluster level.

RESULTS

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

Behavioral Results

Results of the Post-fMRI SRTT

The multiple regression testing for effects of the first-order and second-order conditional probabilities on the logarithmized RTs revealed a significant negative relationship

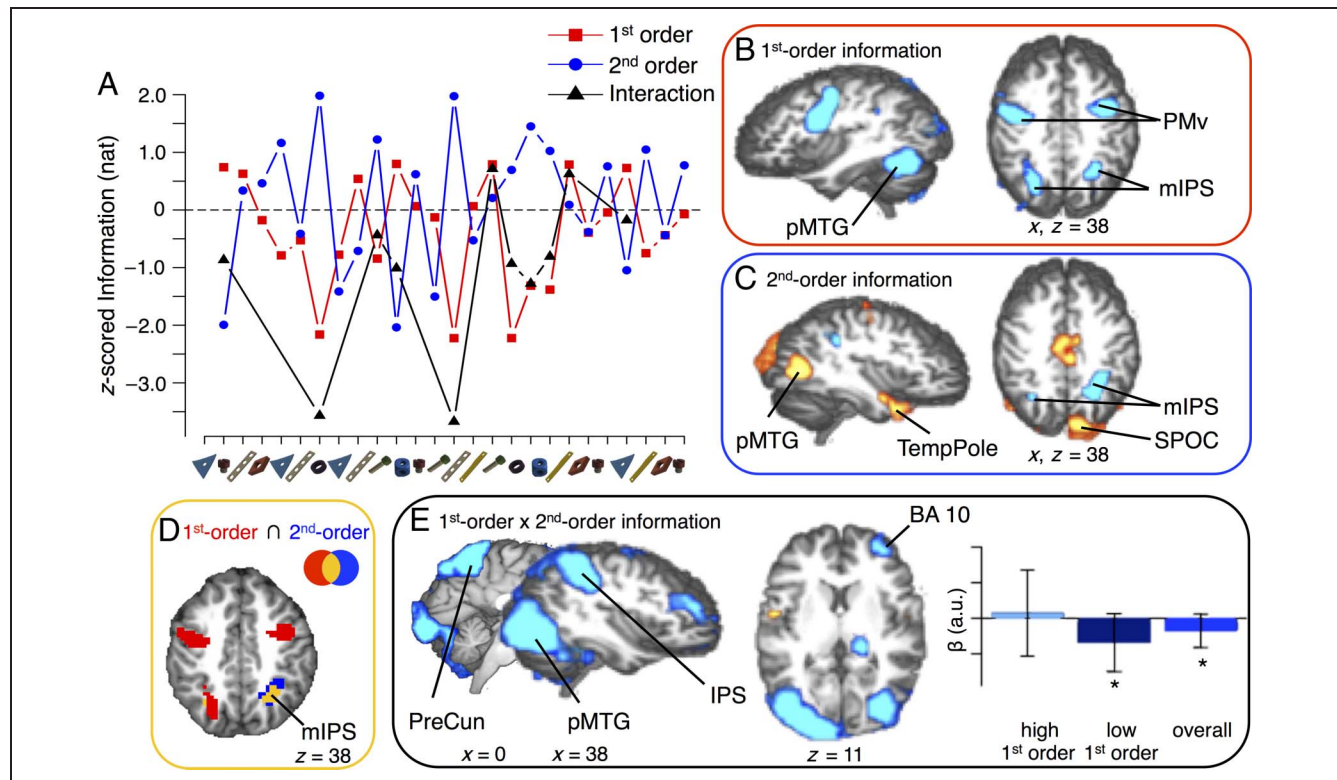


Figure 2. (A) Example course of the parametric regressors for first-order information (red), second-order information (blue), and their interaction term (black) during an excerpt of the experiment. (B) Parametric effects of the amount of information provided by the first-order statistical structure. (C) Parametric effects of the amount of information provided by the second-order statistical structure. TempPole = temporal pole. (D) Overlay of the parametric effects of the first- and second-order statistical structure in observed action videos. Effects of first-order information are displayed in red, second-order in blue. Effects of both parameters overlapped in the mIPS (yellow) and comprised 1188 mm³ (59.46% of the activation cluster revealed in the first-order contrast) in the right and 432 mm³ (5.05%) in the left hemisphere. (E) Interaction of parametric effects of the amount of information provided by the second-order statistical structure and the amount of information provided by the first-order structure. The bar chart depicts beta values in the BA 10 when the interaction term modeled only events with high first-order information (light blue, $t(18) = -0.18$, $p = .855$), low first-order information (dark blue, $t(18) = -3.12$, $p = .006$), and the interaction effect when events with high or low first-order information were modeled (middle blue, $t(18) = -3.41$, $p = .003$). Error bars depict $\pm 1 SD$. preCun = precuneus.

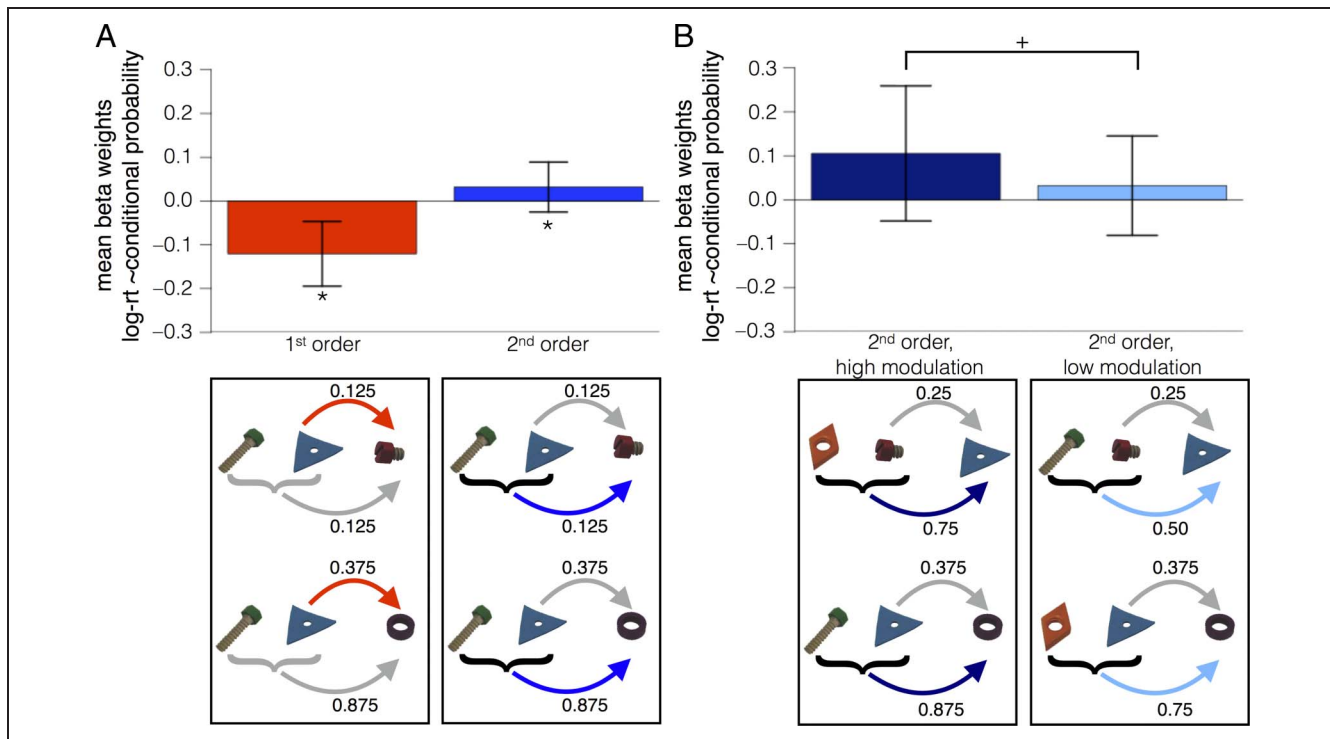


Figure 3. Results of the serial RT posttest. (A) Mean beta weights expressing the relationship between first- and second-order conditional probabilities and RTs. RTs decreased with increasing first-order conditional probabilities and increased with increasing second-order conditional probabilities. (B) Comparison between effects of second-order conditional probabilities on RTs in dependence on the degree to which first-order conditional probabilities were modulated by the second order. RTs showed a trend toward a stronger modulation by second-order conditional if first-order conditional probabilities were modulated to a larger extent. Error bars depict ± 1 *SD*. * $p < .05$, + $p < .06$.

between first-order conditional probability and RTs, showing that higher first-order probabilities led to faster RTs ($t(17) = -6.92$, $p < .001$, two-tailed, $M = -0.12$, $SD = 0.07$ of the standardized coefficients). This effect was consistent across all participants. The effect of the second-order conditional probability was also significant ($t(17) = 2.37$, $p = .030$, two-tailed, $M = 0.03$, $SD = 0.06$), indicating slower RTs with higher second-order probabilities (see Figure 3). Thirteen of the 18 tested participants showed a positive correlation between second-order conditional probabilities and RTs. As we conducted multiple regressions, those results show that RTs were slower for higher second-order conditional probabilities while controlling for an effect of first-order conditional probabilities.

We furthermore wanted to test whether the effect of second-order conditional probabilities depended on the degree to which expectations based on first-order conditional probabilities had been modulated by these second-order conditional probabilities. To that end, we conducted a median split of the data for each participant, dividing trials by whether the second order modulated the first order to a greater or lesser extent. We performed two multiple regressions parallel to the multiple regression described above, with first-order and second-order conditional probability, as well as trial number, as predictors. The resulting standardized coefficients for the second-order conditional probability depending on how strongly the second order

changed the expectations based on the first-order conditional probabilities were tested against each other using a paired t test. A marginally significant difference was revealed ($t(17) = 2.04$, $p = .057$, two-sided). Thus, RTs showed a trend for being more strongly modulated by second-order probabilities if those modulated the expectations based on first-order probabilities strongly ($M = 0.11$, $SD = 0.15$) compared with if the modulation was weak ($M = 0.03$, $SD = 0.11$; see Figure 3).

Results of the Post-fMRI Paper and Pencil Test

The results of the paper and pencil posttest, which assessed knowledge of the first-order structure, further corroborated the significant effect of first-order conditional probabilities on RTs. The repeated-measures ANOVA testing for an overall effect of the factor probability on the assigned weight turned out significant ($F(3, 51) = 18.17$, $p < .001$, partial $\eta = .52$). As we expected rated probabilities to reflect actually implemented probabilities, planned paired t tests were conducted between the single successive levels. We found no difference between probabilities of 0 and .125 ($t(17) = 1.61$, $p = .063$, one-tailed, $d = 0.38$), a marginally significant difference between probabilities of .125 and .25 ($t(17) = 2.09$, $p = .026$, one-tailed, $d = 0.49$) and a significant difference between .25 and .375 ($t(17) = 3.48$, $p = .002$, one-tailed, $d = 0.82$), with an alpha level of .017, adjusted for the three comparisons

(see Figure 4; note that the mean assigned values were scaled by the factor 12.5 to match the scaling of the implemented probabilities). This indicates that participants formed predictions based on the first-order conditional probabilities and that their representation of first-order conditional probabilities was more precise for higher probability values. None of the participants claimed conscious knowledge of the structure when interviewed after the experiment.

fMRI Results

Manipulating the amount of information provided by the first and second order of the statistical structure independently of each other allowed us to assess functional correlates of the exploitation of each of the levels independently. Furthermore, it enabled us to investigate how the amount of information provided by the first order affects exploitation of further information provided by the second order.

Effects of First-order Information

The contrast testing for a modulation of the BOLD response by the amount of information provided by the first-order structure yielded an attenuation of activation in the predicted network of ventral premotor cortex (PMv), the midposterior part of the intraparietal sulcus (mIPS), and the fusiform gyrus and posterior middle temporal gyrus (pMTG), which is classically reported for action observation (see Table 1 for a list of all activations; Figure 2B). Because information provided by the first-order structure and information provided by the second order were modeled simultaneously, this finding shows that increased predictability based on information provided by the first-order structure can reduce activation even when information from the second-order structure is also available.

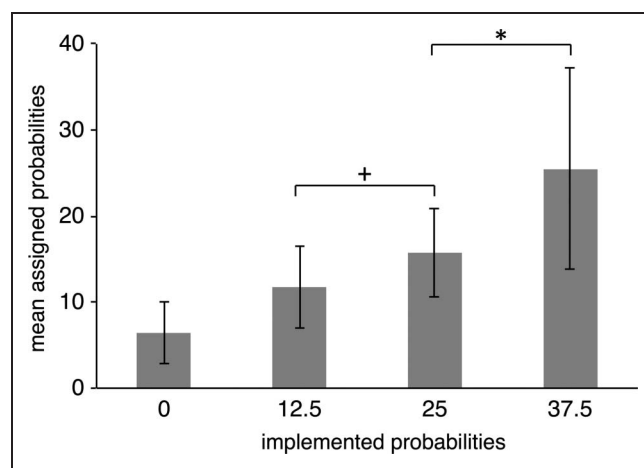


Figure 4. Results of the paper and pencil posttest, showing that assigned probabilities increased as implemented probabilities increased. Number of assigned crosses was multiplied by 12.5 to achieve same scaling as underlying probabilities. Error bars depict ± 1 SD. * $p < .017$, + $p < .03$.

Table 1. MNI Coordinates and Maximal z Scores of Significantly Activated Clusters following Correction for Multiple Comparison for the Parametric Contrast of Information Provided by the First-order Structure

| Localization | MNI Coordinates | | | z Values, Local Maxima | Cluster Size (mm^3) |
|------------------------|-----------------|-----|-----|--------------------------|--------------------------------|
| | x | y | z | | |
| PMv | -41 | 1 | 33 | -4.39 | 11691 |
| | 37 | 4 | 33 | -4.22 | 9855 |
| mIPS | -17 | -62 | 48 | -3.99 | 8559 |
| | 25 | -53 | 42 | -3.38 | 1998 |
| mIPS/precuneus (BA 19) | 13 | -65 | 54 | -2.87 | 567 |
| | 28 | -71 | 22 | -2.97 | 810 |
| Fusiform gyrus/pMTG | -50 | -59 | 0 | -3.96 | 6939 |
| | 40 | -50 | -21 | -3.06 | 1107 |

Effects of Second-order Information

We expected second-order information to draw on activation of the AON as well, although we considered either a positive or a negative correlation as possible.

Higher second-order information was associated with a decrease of activation in mIPS, which overlapped with the cluster observed in the first-order contrast (1188 mm^3 in the left, 432 mm^3 in the right hemisphere; see Figure 2D for a conjunction of the two contrasts). The mIPS was the only area for which an overlap was revealed. We found an increase in activation with higher second-order information in pMTG and superior parieto-occipital cortex (SPOC). An unhypothesized positive correlation between BOLD activation and second-order information was furthermore revealed in the right temporal pole (see Table 2 for a list of all activations; Figure 2C). Those findings show that second-order information is spontaneously integrated, independent of first-order information. To additionally test which areas are more sensitive toward first- than toward second-order information, we calculated the direct contrast between the two parametric regressors. This contrast revealed significantly higher activation for the second-order in the premotor cortex and the pMTG, showing that activation there was more strongly attenuated by first-order information (see Supplementary Table 1 and Supplementary Figure 1).

To test for the hypothesized correlation between second-order information and activation in the hippocampal formation reflecting effects for retrieval of second-order information, we additionally conducted an ROI analysis in the anterior hippocampus. ROI coordinates were taken from a previous publication of our group (Ahlheim et al., 2014) and were based on reported effects of sensitivity of the hippocampus to entropy (Bornstein & Daw, 2012; Harrison et al., 2006; Strange et al., 2005). The center of the ROI in the left hippocampus was at $x = -25$, $y = -16$, $z = -18$, and the center of the ROI in the right

Table 2. MNI Coordinates and Maximal z Scores of Significantly Activated Clusters following Correction for Multiple Comparison for the Parametric Contrast of Information Provided by the Second-order Structure

| Localization | MNI Coordinates | | | z Values, Local Maxima | Cluster Size (mm^3) |
|---------------------------------------------|-----------------|-----|-----|--------------------------------|-----------------------------------|
| | x | y | z | | |
| Dorsal premotor cortex | 28 | -11 | 54 | 3.82 | 4725 |
| Local maximum in posterior cingulate cortex | 7 | -12 | 39 | 3.58 | |
| mIPS | -29 | -59 | 30 | -2.91 | 594 |
| | 25 | -50 | 36 | -3.31 | 3294 |
| pMTG | -50 | -68 | 18 | 3.11 | 405 |
| | 37 | -62 | 9 | 4.23 | 4455 |
| SPOC (BA 18) | -20 | -89 | 15 | 3.00 | 648 |
| | 16 | -92 | 21 | 4.56 | 13851 |
| Temporal pole | 52 | 4 | -30 | 3.50 | 4401 |

anterior hippocampus was at $x = 31, y = -17, z = -19$. Both ROIs had a sphere with a radius of two adjacent voxels (6 mm). Unexpectedly, neither ROI showed a significant modulation by second-order information (all $p > .09$, Bonferroni-corrected alpha level of .025; see Table 3 for inferential statistics).

First-order Dependent Exploitation of Second-order Information

We hypothesized that exploitation of the second-order information depends on the amount of information provided by the first-order structure. To test this, we included an interaction term modeling only those events for

Table 3. Inferential Statistics of Hippocampal ROI Analyses

| | $t(18)$ | p |
|-----------------------------------------------|---------|------|
| Parametric effect of first-order information | | |
| Left hippocampus | 1.75 | .097 |
| Right hippocampus | 0.42 | .683 |
| Parametric effect of second-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 |

which the first-order structure provided least information (lowest 25% of the distribution) or the most information (uppermost 25% of the distribution). The interaction therefore reveals areas that were significantly more strongly modulated by information provided by the second-order structure if the first-order structure provided only little information about the upcoming event. We found that activation in the premotor cortex (PMd), the intraparietal sulcus (IPS), the precuneus, and the occipitotemporal lobe were more strongly modulated by information provided by the second order of the statistical structure when less information was provided by the first-order structure.

Additionally, the interaction contrast yielded the predicted modulation of activity in lateral BA 10. BA 10 did not show a significant modulation by second-order information or first-order information alone, which indicates that it is more strongly modulated by information provided by the second order if integration of this information was actually beneficial, that is, when the first-order provided less information (see Table 4 for a list of all activations; Figure 2E). As can be seen from the bar chart in Figure 2E, this interaction effect was indeed driven by the cases in which first-order information was low.

Table 4. MNI Coordinates and Maximal z Scores of Significantly Activated Clusters following Correction for Multiple Comparison for the Interaction Contrast of Information Provided by the Second-order Structure, Depending on the Amount of Information Provided by the First-order Structure

| Localization | MNI Coordinates | | | z Values, Local Maxima | Cluster Size (mm^3) |
|-----------------------------|-----------------|------|-----|--------------------------------|-----------------------------------|
| | x | y | z | | |
| Anterior pFC | | | | | |
| 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 | | | | | |
| IPS | -29 | -44 | 57 | -5.49 | 201285 |
| | 33 | -40 | 56 | -4.68 | |
| Precuneus | -9 | -62 | 68 | -4.90 | |
| | 13 | -65 | 46 | -4.56 | |
| SPOC | -15 | -101 | -6 | -5.20 | |
| pMTG | -38 | -87 | -13 | -5.13 | |
| | 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 |

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

DISCUSSION

Although it is well established that humans use predictive information in their environment to prepare for upcoming events, it is still unclear to what extent and under which conditions they do so. It is one of the currently most urgent questions how the brain selects the sources of information to generate predictions (Phillips, 2013; Blokpoel, Kwisthout, & van Rooij, 2012). This study investigated whether information from an action's second-order statistical structure is exploited in dependence on the information provided on the first level; in other words, whether the brain predicts upcoming actions in a cost-benefit sensitive manner.

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

Attenuation in the AON Based on First-order Statistical Information

The first aim to this study was to replicate and expand previous findings concerning the neural correlates of an increase in predictability by the first-order structure in action sequences (Ahlheim et al., 2014). We established in our behavioral posttests that human observers learned first-order conditional probabilities and were particularly good at discriminating between action pairs of high conditional probability, even though no participant reported noticing those regularities in a postexperimental survey.

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

Integration of Second-order Statistical Information

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

Contrary to first-order information, second-order information could not be accessed directly through a pairwise association between action $t - 2$ and t . Instead, it was necessary to retrieve information about the action step $t - 2$ from memory and furthermore integrate this information with the information provided by the action $t - 1$ on the first order, as the action step at $t - 2$ alone was not informative of t . Potentially, these additional processing costs could further account for the unpredicted finding of increased RTs with second-order conditional probabilities in our post-fMRI SRTT: Here, RTs increased with higher second-order conditional probabilities while controlling for an effect of first-order conditional probabilities. Furthermore, a trend level effect ($p = .057$) tentatively suggests that these processing costs, reflected in RT increase, is higher when second-order information changed expectations based on the first-order conditional probabilities to a larger extent. Studies on learning of second-order statistical regularities using a SRTT reported a decrease of RTs as reflection of statistical learning (Remillard, 2008; Curran, 1989). Speculating on possible reasons for the diverging results, it should be noted that our SRTT differed in a critical point from a standard SRTT: Statistical regularities among the action steps were already established at the beginning of the testing, whereas the association between

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

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

Evidence for Information State-dependent Use of Second-order Information

To test the hypothesis that exploitation of the second-order statistical structure depends on the amount of information provided by the first order, we conducted a parametric

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

In line with our hypothesis, we found that activity in the lateral BA 10 was correlated with the interaction term. This correlation resulted from a stronger correlation of activity in the BA 10 with second-order information if the first order provided only little information, that is, if the action step $t - 1$ did not allow for a sufficiently precise prediction of action t . Notably, significant activation of the BA 10 was only revealed in the interaction contrast. This corroborates our hypothesis that BA 10 recruitment increases under low first-order predictability and enhances the exploitation of second-order information. Across a variety of different paradigms, BA 10 has been reported to be activated when several relations among tasks or rules have to be integrated or organized (Nee et al., 2013; Schubotz, 2011; Golde, von Cramon, & Schubotz, 2010; Koechlin & Hyafil, 2007; Ramnani & Owen, 2004). Here, and in line with findings from Golde et al. (2010), we showed that the BA 10 is also engaged when information derived from actions needs to be integrated. A particularly interesting parallel to our paradigm is the engagement of BA 10 in uncertainty-driven search for information, when available cues provide insufficient information (Badre et al., 2012). Whereas information in the study by Badre and co-workers (2012) could be gained by searching the environment, in this study information was gained through retrieval of the action at $t - 2$. Our results suggest that BA 10 may particularly contribute to a strategic retrieval of associations if

these associations provide a clear gain in information. In other words, BA 10 may implement an efficiency criterion for the exploitation of higher-order information, presumably both in actions as well as in abstract stimuli.

Conclusion

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

Acknowledgments

The authors thank Christian Jessen and Anna Schubotz for experimental assistance as well as Jan Balaguer, Hannah Tickle, Sarah Weigelt, and Moritz Wurm for valuable suggestions on the manuscript. This research was supported by the University of Münster, Germany, and the Max Planck Society.

Reprint requests should be sent to Christiane Ahlheim, Fliednerstr. 21, 48149 Münster, Germany, or via e-mail: c.ahlheim@ucl.ac.uk.

REFERENCES

- Ahlheim, C., Stadler, W., & Schubotz, R. I. (2014). Dissociating dynamic probability and predictability in observed actions—An fMRI study. *Frontiers in Human Neuroscience*, *8*, 1–13.
- Badre, D., Doll, B. B., Long, N. M., & Frank, M. J. (2012). Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron*, *73*, 595–607.
- Baldwin, D. A., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*, 1382–1407.
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*, 617–629.
- Beauchamp, M. S., & Martin, A. (2007). Grounding object concepts in perception and action: Evidence from fMRI studies of tools. *Cortex*, *43*, 461–468.
- Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, *58*, 47–73.
- Blokpoel, M., Kwisthout, J., & van Rooij, I. (2012). When can predictive brains be truly Bayesian? *Frontiers in Psychology*, *3*, 406.
- Bornstein, A. M., & Daw, N. D. (2012). Dissociating hippocampal and striatal contributions to sequential prediction learning. *European Journal of Neuroscience*, *35*, 1011–1023.
- Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, *4*, 1–15.
- Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage*, *50*, 1148–1167.
- Cover, T., & Thomas, J. (1991). Entropy, relative entropy and mutual information. In *Elements of information theory* (2nd ed., pp. 12–49). Hoboken, NJ: Wiley.
- Curran, T. (1989). Higher-order associative learning in amnesia: evidence from the serial reaction time task. *Journal of Cognitive Neuroscience*, *9*, 522–533.
- Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*, 876–879.
- den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*, 548.
- Domenech, P., & Dreher, J.-C. (2010). Decision threshold modulation in the human brain. *Journal of Neuroscience*, *30*, 14305–14317.
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 458–467.
- Fortin, N. J., Agster, K. L., & Eichenbaum, H. B. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience*, *5*, 458–562.
- Friston, K. J., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *364*, 1211–1221.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.
- Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage*, *9*, 416–429.
- Golde, M., von Cramon, D., & Schubotz, R. I. (2010). Differential role of anterior prefrontal and premotor cortex in the processing of relational information. *Neuroimage*, *49*, 2890–2900.
- Gureckis, T. M., & Love, B. C. (2010). Direct associations or internal transformations? Exploring the mechanisms underlying sequential learning behavior. *Cognitive Science*, *34*, 10–50.
- Harrison, L. M., Duggins, A., & Friston, K. J. (2006). Encoding uncertainty in the hippocampus. *Neural Networks*, *19*, 535–546.
- Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *Neuroimage*, *14*, 103–109.
- Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science*, *318*, 594–598.
- Kok, P., Brouwer, G. J., van Gerven, M. A. J., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience*, *33*, 16275–16284.
- Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron*, *75*, 265–270.
- Kumaran, D., & Maguire, E. A. (2009). Novelty signals: A window into hippocampal information processing. *Trends in Cognitive Sciences*, *13*, 47–54.
- Lencer, R., Nagel, M., Sprenger, A., Zapf, S., Erdmann, C., Heide, W., et al. (2004). Cortical mechanisms of smooth pursuit eye movements with target blanking. An fMRI study. *European Journal of Neuroscience*, *19*, 1430–1436.

- Lohmann, G., Müller, K., Bosch, V., Mentzel, H.-J., Hessler, S., Chen, L., et al. (2001). LIPSI—A new software system for the evaluation of functional magnetic resonance images of the human brain. *Computerized Medical Imaging and Graphics*, *25*, 449–457.
- Nee, D., Jahn, A., & Brown, J. (2013). Prefrontal cortex organization: Dissociating effects of temporal abstraction, relational abstraction, and integration with fMRI. *Cerebral Cortex*, *24*, 2377–2387.
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1–32.
- O'Reilly, J. X., Schüffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *110*, 3660–3669.
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976–987.
- Peelen, M. V., & Caramazza, A. (2012). Conceptual object representations in human anterior temporal cortex. *Journal of Neuroscience*, *32*, 15728–15736.
- Phillips, W. A. (2013). Neuronal inference must be local, selective, and coordinated. *Behavioral and Brain Sciences*, *36*, 222–223.
- Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience*, *5*, 184–194.
- Remillard, G. (2008). Implicit learning of second-, third-, and fourth-order adjacent and nonadjacent sequential dependencies. *Quarterly Journal of Experimental Psychology (2006)*, *61*, 400–424.
- Schiffer, A.-M., Ahlheim, C., Ulrichs, K., & Schubotz, R. I. (2013). Neural changes when actions change: Adaptation of strong and weak expectations. *Human Brain Mapping*, *34*, 1713–1727.
- Schubotz, R. I. (2011). Long-term planning and prediction: Visiting a construction site in the human brain. In W. Welsch, W. J. Singer, & A. Wunder (Eds.), *Interdisciplinary anthropology* (pp. 79–104). Berlin, Heidelberg: Springer.
- Schubotz, R. I., Wurm, M. F., Wittmann, M. K., & von Cramon, D. Y. (2014). Objects tell us what action we can expect: Dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. *Frontiers in Psychology*, *5*, 1–15.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.
- Stadler, W., Schubotz, R. I., von Cramon, D. Y., Springer, A., Graf, M., & Prinz, W. (2011). Predicting and memorizing observed action: Differential premotor cortex involvement. *Human Brain Mapping*, *32*, 677–687.
- Strange, B. A., & Dolan, R. J. (2001). Adaptive anterior hippocampal responses to oddball stimuli. *Hippocampus*, *11*, 690–698.
- Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., & Friston, K. J. (2005). Information theory, novelty and hippocampal responses: Unpredicted or unpredictable? *Neural Networks*, *18*, 225–230.
- Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*, 745–756.
- Summerfield, C., & Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron*, *59*, 336–347.
- Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*, 1004–1006.
- Swallow, K. M., & Zacks, J. M. (2008). Sequences learned without awareness can orient attention during the perception of human activity. *Psychonomic Bulletin & Review*, *15*, 116–122.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
- Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*, 1934–1945.
- Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*, *30*, 11177–11187.
- Vesia, M., & Crawford, J. D. (2012). Specialization of reach function in human posterior parietal cortex. *Experimental Brain Research*, *221*, 1–18.
- Vingerhoets, G. (2014). Contribution of the posterior parietal cortex in reaching, grasping, and using objects and tools. *Frontiers in Psychology*, *5*, 1–17.
- Wurm, M. F., Hrkač, M., Morikawa, Y., & Schubotz, R. I. (2014). Predicting goals in action episodes attenuates BOLD response in inferior frontal and occipitotemporal cortex. *Behavioural Brain Research*, *274C*, 108–117.
- Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience*, *23*, 4057–4066.