NeuroImage 296 (2024) 120687

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Updating predictions in a complex repertoire of actions and its neural representation

Rosari Naveena Selvan [a,b,c,*], Minghao Cheng [c], Sophie Siestrup [a,b], Falko Mecklenbrauck [a,b], Benjamin Jainta [a], Jennifer Pomp [a,b], Anoushiravan Zahedi [a,b], Minija Tamosiunaite [c,d], Florentin Wörgötter [c], Ricarda I. Schubotz [a,b]

[a] Department of Psychology, University of Münster, Münster, Germany
[b] Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany
[c] Department for Computational Neuroscience, Third Institute of Physics – Biophysics, University of Göttingen, Göttingen, Germany
[d] Faculty of Informatics, Vytautas Magnus University, Kaunas, Lithuania

ARTICLE INFO

Keywords:
Action observation
Branching structure
Action perception
Action prediction hierarchy
Internal models
Violation of expectation

ABSTRACT

Even though actions we observe in everyday life seem to unfold in a continuous manner, they are automatically divided into meaningful chunks, that are single actions or segments, which provide information for the formation and updating of internal predictive models. Specifically, boundaries between actions constitute a hub for predictive processing since the prediction of the current action comes to an end and calls for updating of predictions for the next action. In the current study, we investigated neural processes which characterize such boundaries using a repertoire of complex action sequences with a predefined probabilistic structure. Action sequences consisted of actions that started with the hand touching an object (T) and ended with the hand releasing the object (U). These action boundaries were determined using an automatic computer vision algorithm. Participants trained all action sequences by imitating demo videos. Subsequently, they returned for an fMRI session during which the original action sequences were presented in addition to slightly modified versions thereof. Participants completed a post-fMRI memory test to assess the retention of original action sequences. The exchange of individual actions, and thus a violation of action prediction, resulted in increased activation of the action observation network and the anterior insula. At U events, marking the end of an action, increased brain activation in supplementary motor area, striatum, and lingual gyrus was indicative of the retrieval of the previously encoded action repertoire. As expected, brain activation at U events also reflected the predefined probabilistic branching structure of the action repertoire. At T events, marking the beginning of the next action, midline and hippocampal regions were recruited, reflecting the selected prediction of the unfolding action segment. In conclusion, our findings contribute to a better understanding of the various cerebral processes characterizing prediction during the observation of complex action repertoires.

1. Introduction

When we observe actions, our brain generates predictions about how this complex, dynamic stimulus will evolve (Botvinick and Plaut, 2004; Colder, 2011; Csibra and Gergely, 2007; Kilner et al., 2004, 2007; Schiffer et al., 2013; Stadler et al., 2011). Expected and actual stimulations are compared, and only the difference between the two, the prediction error, is propagated to update future predictions about the probabilistic structure of our reality (Friston, 2005). It is generally assumed that these computations, as complex perception in general, are based on hierarchical predictive processing. According to this hierarchical framework, prediction errors are conveyed in a bottom-up manner from lower cortical areas to higher ones via forward connections, while predictions are delivered top-down by backward connections (Friston, 2010; Friston and Kiebel, 2009). As of today, we still know little about the number of layers and their computational characteristics in this postulated action prediction hierarchy, and in which cerebral networks they are specifically implemented.

One way to conceptualize action prediction is along a part-whole hierarchy, meaning that a goal-directed action sequence consists of

* Corresponding author at. University of Münster, Fliednerstraße 21, Münster 48149, Germany.
E-mail address: rselvan@uni-muenster.de (R.N. Selvan).

https://doi.org/10.1016/j.neuroimage.2024.120687
Received 21 February 2024; Received in revised form 3 May 2024; Accepted 11 June 2024
Available online 12 June 2024

multiple segments or action steps (Uithol et al., 2012). Such segmental action structures are not deterministic, since at boundaries between action segments, we can often choose between multiple plausible steps that can be performed next. Studies have shown that subjects can reliably detect segment boundaries (Newtson, 1973) and that the brain's event prediction is also oriented toward these segments (Zacks et al., 2011). This has been observed when watching movies or reading stories, but also when observing actions of others (Kurby and Zacks, 2008). The segment boundaries are of particular interest, because here the prediction comes to a point where a new segment is pending but not yet observable. Accordingly, the prediction error reaches a maximum (Zacks et al., 2011) and triggers prediction updating (Schubotz et al., 2012; Wahlheim et al., 2022; Zacks et al., 2007; Zacks and Sargent, 2010). At the same time, it is reasonable to assume that the predictive model holds several options of connectable segments, which now form the new expectation. These can be compared to the next observed action, resulting in an update of the currently valid prediction. In a recent study (Pomp et al., 2021), it became clear that specific action boundaries, namely start of hand-object contact (T, Touching) and end of hand-object contact (U, Untouching), are important features for the segmentation of observed actions. Additionally, we could not only confirm that prediction error signals peak at the segmental boundaries of actions but also disentangle prediction error and prediction updating as two temporally and neuronally distinct stages of processing (Pomp et al., 2021).

In the current study, we built on these findings to investigate action prediction processes at boundaries between actions of a structured action repertoire. In a previous study (Pomp et al., 2021), we investigated updating at action boundaries within single, disconnected object manipulations. In contrast, everyday actions mostly consist of action sequences which are hierarchically structured in a way that some action boundaries, namely those with different connection options, involve a change in the level of the action hierarchy (Ondobaka and Bekkering, 2012). In addition, different connection options create a variable degree of uncertainty and predictability. In the present study, we took the view that in an ecologically valid action repertoire, updating should comprise two distinct stages: first, opening of the option space by retrieving all expectable (probable) options, and then, upon exploiting evidence from the ongoing stimulus, restricting to one of them as the most probable option. To investigate this, we generated an action repertoire with a well-defined segmental branching structure; in particular, actions could be followed by one, two, three, or four possible other actions. We videotaped an actress performing all 82 actions of this repertoire and determined action boundaries (i.e., segments) in these action videos using a computer vision-based model. Boundaries were defined as touching (T), i.e., the hand touching an object, and untouching (U), i.e., the hand releasing/ untouching an object. In the present stimulus material, T events defined the beginning of an action segment, whereas U events marked their ending (Aksoy et al., 2011; Wörgötter et al., 2013; Ziaeetabar et al., 2021).

Participants first practiced the action repertoire themselves before entering the MRI scanner. There, they were presented with videos of actions that corresponded to the trained repertoire, but also action sequences with a single action that did not match the learned action repertoire to induce prediction errors and, consequently, updating of event models. Specifically, we modeled the brain activation at hand-object T/U events (see Fig. 4 for the representation) to disentangle the two suggested stages of model updating in suprasegmental actions: retrieval of all upcoming options at the end of the previous action (hand untouching last object) and selecting the option at the beginning of the next action (hand touching next object). Moreover, we modeled the brain activation time-locked to untouching at the event boundaries by the parametric degree of probabilistic structure (i.e., branching). Since we used branches of equal probability, the level of uncertainty was proportional to the number of branches, meaning that more possible branches led to higher uncertainty, and fewer possible branches resulted in lower uncertainty.

We expected the updating of predictions to activate areas upstream of the action observation network (AON), which is reliably engaged by viewing videos showing individual (separate) object manipulation actions. The AON has been suggested to feature an intrinsic hierarchical processing architecture (Sasaki et al., 2018; Urgen et al., 2019; Urgen and Saygin, 2020). In addition to the goal level represented in the AON, additional layers have been proposed that propagate and prioritize goals of actions, supposedly involving dorsolateral prefrontal cortex and ventromedial networks, respectively (Pezzulo et al., 2018). However, as these latter supra-goal layers have so far been proposed for predicting one's own actions, and, being related to control and motivation, they may not be relevant for passive observers. Therefore, it remains to be seen whether they are also used in predicting observed actions.

Another strong candidate for higher layers of the prediction hierarchy of observed actions is a network of the medial prefrontal cortex, posterior cingulate cortex/precuneus, and temporo-parietal junction. Depending on the research context of prediction hierarchy, this network has been referred to as the default mode network (DMN) whose computations are relevant to the most complex integrative functions of the brain (Pesquita et al., 2018). Studies where participants were asked to infer others' intentions or beliefs from their actions found an increased DMN-AON connectivity (Koster-Hale and Saxe, 2013; Wurm et al., 2011). Moreover, recently proposed hierarchical prediction and prediction error processing models (Alexander and Brown, 2015, 2018) suggest that the medial prefrontal cortex (a DMN area) provides high-level predictive input to the lateral prefrontal cortex (an AON area). More generally, the DMN including the medial prefrontal cortex is involved in processing complex and hierarchically structured schemas (Baldassano et al., 2018; Masís-Obando et al., 2022; Reagh and Ranganath, 2023; Sommer et al., 2022). Yet, the specific role of these areas in potentially providing higher-level input to action observation, reflecting a repertoire-informed updating of predictions, awaits further testing.

2. Methods

2.1. Participants

Forty-five subjects (30 women, 15 men) between the age of 18 and 30 years ($M = 23.92$; $SD = 4.32$) were recruited for the study. Data from four participants were not included in the final analyses due to excessive movement (> 5 mm) during the fMRI session. Consequently, the final sample included 41 datasets. Participants had normal or corrected to normal vision with no color blindness and were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) with handedness scores varying between 50 and 100 ($M = 87.02$, $SD = 15.31$). All participants gave written informed consent to participate in this study and were compensated with money or course credits. The study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee of the University of Münster.

2.2. Stimuli

As stimuli material, we used short video clips that showed multistep actions, presented in the third person perspective. Four objects, two cups with green bottom and two-colored cubes (red and blue; Fig. 1), were used to execute action sequences, with predefined actions (e.g., put the cup upside down) and probabilistic transitions between them (see Fig. 2 for all possible object configurations). Video clips of 41 action sequences were generated, hereafter termed *legal* sequences (Fig. 3A). Each video started with the same initial frame showing two hands of an actress wearing white gloves placed on a table and the four objects, each on a fixed table location. The action sequences varied in length, ranging from four to six action steps and six to twelve seconds in duration.

In addition to the 41 complete action sequences, the individual steps of each action sequence were also recorded separately for the imitation

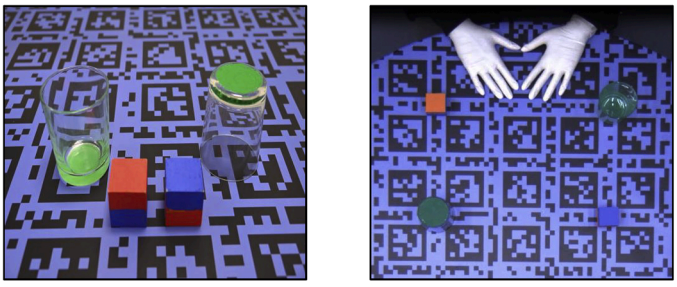

Fig. 1. A. The objects used for the stimuli preparation: two cups and two cubes; B. Starting frame of the videos, showing the actress from the 3pp and the starting distribution of all four objects on the table.

training. The table on which the actions took place was covered with a table cover printed with a QR code-like pattern known as the "AprilTag" markers (Fig. 1). These markings are described in more detail below. The recorded videos were cropped to show only the hands of the actress with the objects (see Section 2.3 for detailed information).

Furthermore, another set of 41 action sequences (termed *illegal* hereafter) was recorded, each of which corresponded to a legal sequence, but a single action was modified (Fig. 3B). This modification consisted of doing a different manipulation with the same object as in the legal ones or doing a different manipulation with a different object. The modification happened in one of the actions of the entire action sequence, however, the first step always remained unchanged. Importantly, after the modification, action sequences continued the same way as the legal ones so that only one action differed between legal and illegal sequences. In the present study, illegal sequences were included to (1) allow for a discrimination task during the fMRI session (see Section 2.5) and (2) to distinguish well-known prediction error effects due to mismatch of expectancies from those that accumulate at event boundaries, which were the main interest of this study.

2.3. Action segmentation and T/U determination

For the creation of stimulus material and the subsequent segmentation of actions based on touching (T) and untouching (U) events, a computer vision system was used. T was defined as the point in time when the actor's hand made contact with one of the objects, U was when the actor's hand released the object. T and U marked the beginning and the end of an action step, respectively (Fig. 4).

The computer vision system consists of four parts: (1) The multi-camera setup, (2) the touching events detector, (3) the eye-tracker and the corresponding transformation system and (4) the synchronization scheme. These parts will be described in detail in the following.

To record the visual content and track the hand and objects in the video, a five-camera setup was installed around the experiment table. The camera used here was Grasshopper GS3-U3–32S4C—C from Teledyne FLIR (resolution 1024×768; 60.0 frames per second (fps); color space: RGB8; synchronization mode: hardware synchronization via cables). The camera setup was calibrated using a multi-camera calibration tool (Michels, 2022). Images captured from the cameras were saved as video files which consequently went through a pipeline for detection and tracking using YoloV5 (Jocher, 2020). In addition, the "Apriltag" markers (Olson, 2011) were employed to calculate both the eye tracker's

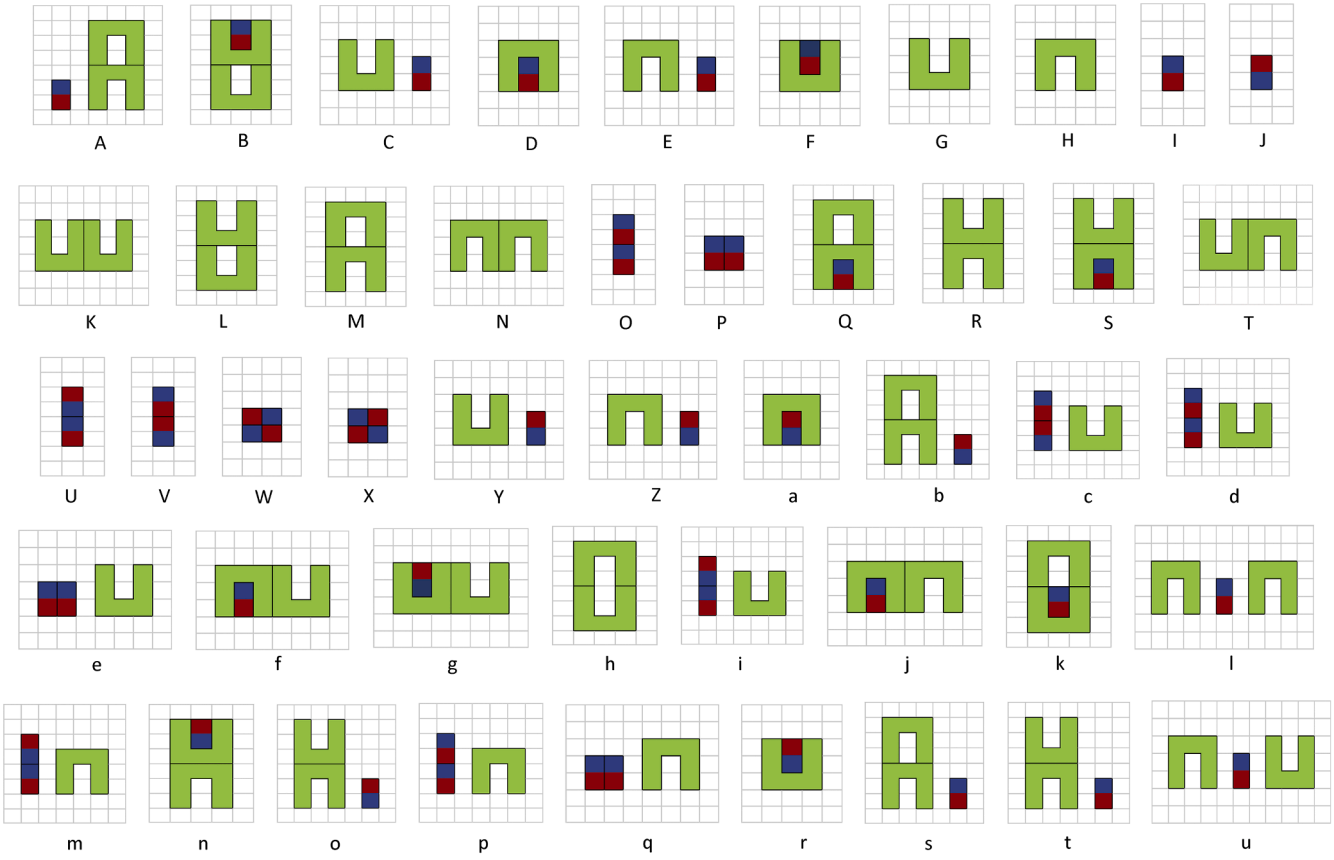

Fig. 2. Schema of the object configurations achieved by actions. Two cups and two cubes were manipulated by stacking up, placing next to, putting in, taking out, turning over, placing over, and taking down. The trained action repertoire consisted only of actions generating object configurations labeled by capital letters A to Z. Lower case letters show configurations only achieved by illegal actions. Note that some actions were also illegal when generating a legal object configuration but in the wrong action sequence.

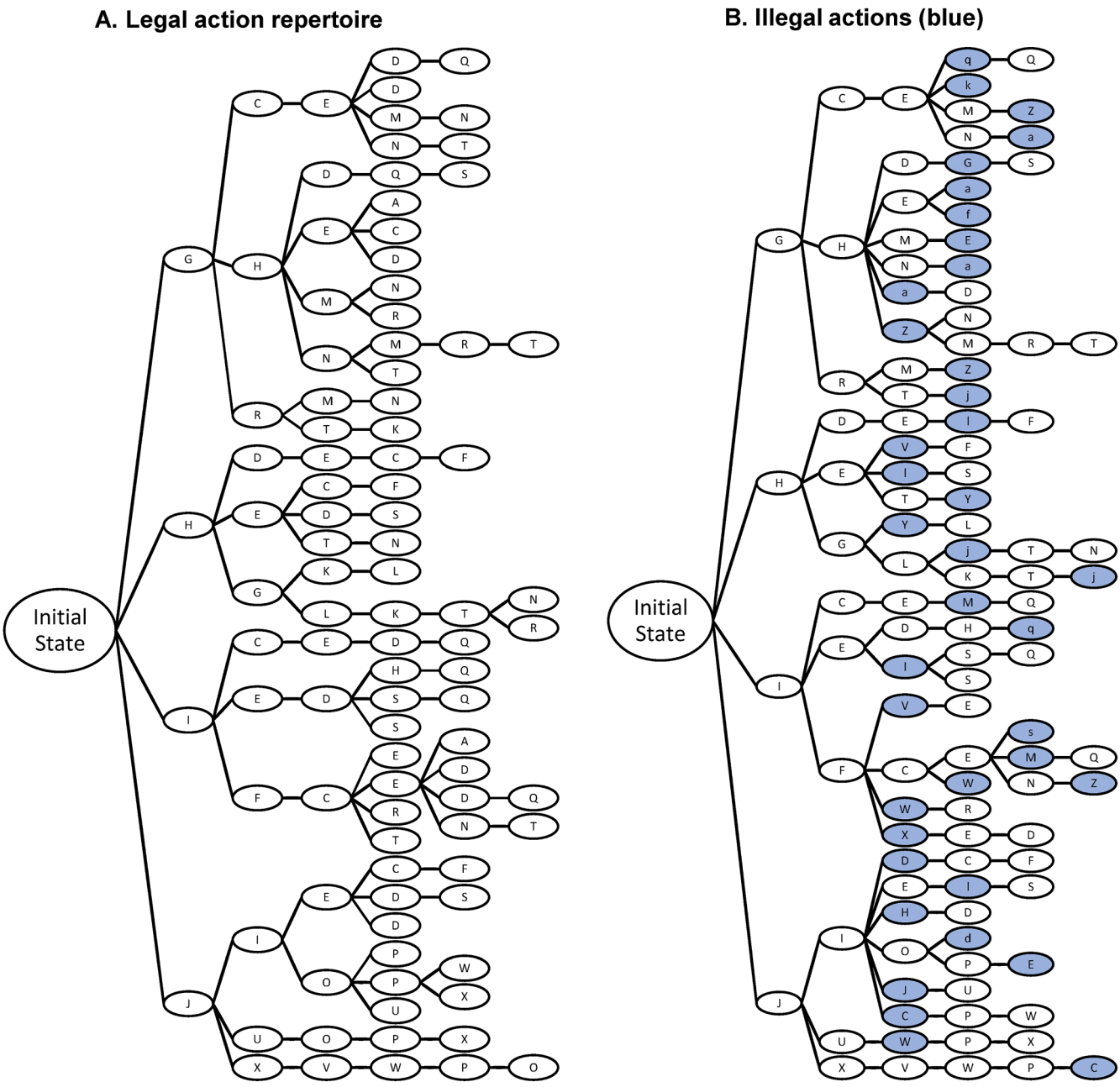

Fig. 3. A. Decision tree of the legal action repertoire; B. Illegal actions; actions that violate the learned repertoire are marked in blue. Letters indicate different actions as defined by their outcome, i.e., a certain configuration of objects (for a detailed description of object configurations, refer to Fig. 2). The branching level can be recognized at the transitions between the individual actions; for example, it is three for action D after the sequence I-E-D, but four for action C after the sequence I-F-C.

pose (extrinsic parameters of the camera) and the transformation matrix between the eye tracker and the five-camera setup. A tablecloth, with 7 by 7 "Apriltag36h11" markers printed on it, was fixed on the experiment table. Hands and objects were tracked by the multi-camera setup. During the stimuli preparation, eye movements of the actress were also tracked, but these data were not relevant to the current study and are therefore not mentioned further below.

For T/U event detection, a touching sensor was attached to the right index finger. The touching sensor consisted of a force-sensing resistor, a bias resistor with a 5-volt power supply, an OPAMP for buffering and an Arduino Uno for A/D conversion with a sampling rate of 1KHz and resolution of 8-bit. The sampled signal was compared with a threshold value of 127 (half of the max 8-bit binary number) and then denoised with manual human help to exclude false touching and untouching events, such as accidentally touching the table. Importantly, T and U events that were analyzed in the present study were restricted to the touching and untouching of the experimental objects (cups and cubes). Notably, due to the multistep nature of action sequences, untouching events signal transitions between actions and are, therefore, indicative of event boundaries.

2.4. Training procedure

The experimental protocol included four consecutive days. On the first two days, participants went through imitation training sessions. The imitation training consisted of three phases for each action sequence. During the first phase, participants were asked to observe the action sequence presented to them on a monitor (27 inches) using Presentation

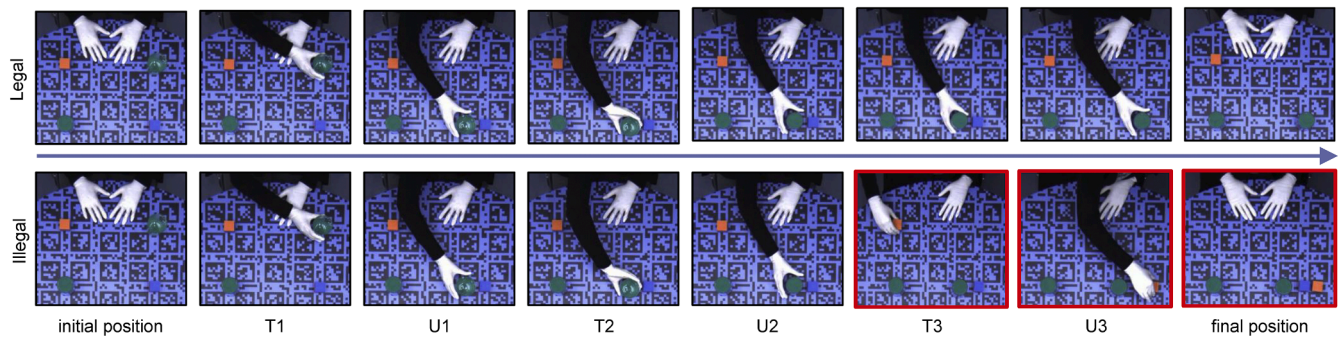

Fig. 4. An example of a legal and the corresponding illegal action sequence with all Touching (T) and Untouching (U) events showing the moment of modification at T3. In the legal action sequence, at T3, the cup is taken to be placed over the blue-side-up cube next to it. In the illegal action sequence, instead of the cup, the red-side-up cube is grasped.

software 20.3 (Neurobehavioral Systems Inc., Berkley, CA, USA). In the second phase, participants imitated the actions of action sequences while the recorded videos were presented to them. They were instructed to imitate with the same speed and hand movements. In the last phase, participants imitated the complete action sequences along with the video. The distance between the seated participant and the screen measured approximately 80 cm. The same table cover and objects that were used for stimuli preparation were employed for the imitation training as well.

The presentation of the 41 legal videos was organized in four blocks. Videos were assigned to a block based on the object that was manipulated first during the action sequence (upright cup, inverted cup, blue-side-up cube, and red-side-up cube) independent of its concrete manipulation. Within each block, videos were presented in the order of the length of the sequences of four, five, and six steps action sequences. Each training session included a short pause after two blocks, and the whole session lasted for about 75 to 90 min. During imitation training, participants were observed by an experimenter, and if the participants made a mistake, they were asked to imitate the action sequence again. The same procedure was followed on the second day of imitation training.

2.5. fMRI session

During the fMRI session, on the third day, participants were presented with the legal and illegal action sequences in randomized order. The session consisted of three blocks, with all 41 legal action sequences presented in three blocks. Forty-one illegal action sequences were distributed across blocks with 13 presented in the first block and 14 each in the second and third block. Each block always started with a legal action sequence. An illegal sequence and its corresponding legal one were never presented consecutively and not more than three illegal sequences were presented in a row. After each block, there was a short break of approximately two to four minutes for eye-tracking calibration (see Section 2.7). Additionally, after the presentation of every action sequence, there was a question trial during which participants responded by pressing a button to determine whether they had seen the specific action sequence during the imitation training. The question trial lasted till a button was pressed or a maximum of two seconds (Fig. 5A). Moreover, there were 15 null events, five in each block, during which only a fixation cross was presented for 1500, 2000, or 2500 ms. The entire experiment consisted of 123 trials of legal and 41 trials of illegal action sequences, 164 question trials, and 15 null events, and lasted for about 60 min, with each block 16 to 20 min long.

2.6. Post-fMRI memory test

One day after the fMRI session, participants performed a post-fMRI memory test on the legal sequences that they had observed and imitated during the first two days of the experiment. The task comprised

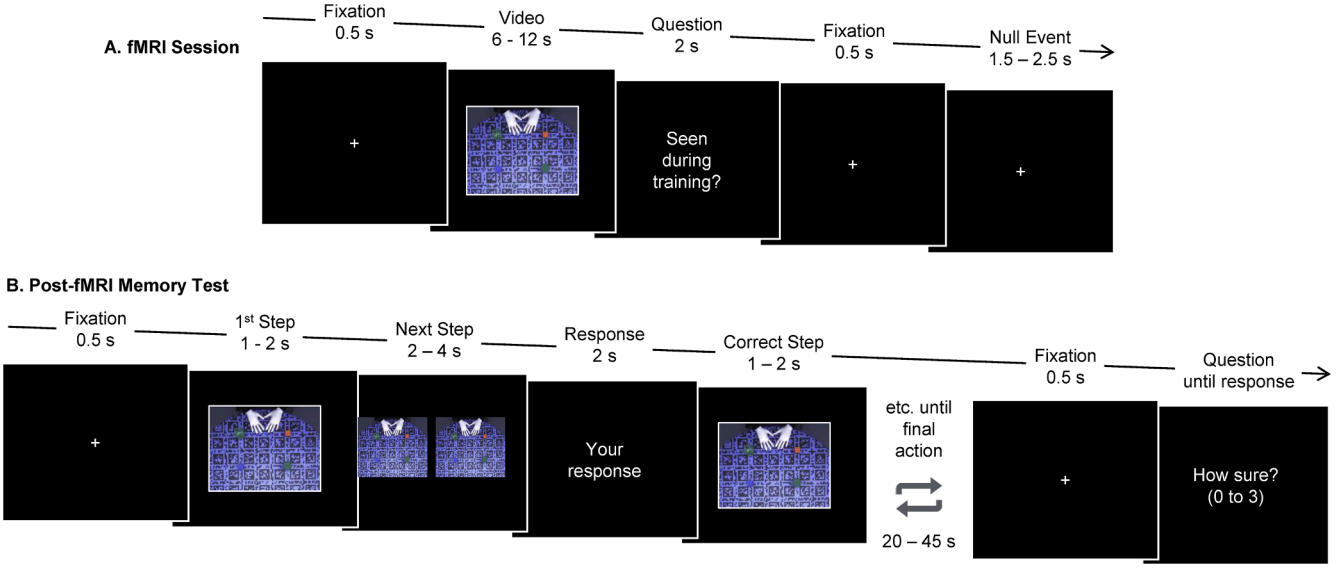

Fig. 5. Schema of the experimental paradigm A. Trial schema for the fMRI session with legal and illegal action sequences B. Trial schema for the post-fMRI memory test.

all the 41 legal sequences presented on the monitor step by step. The video of the action sequence paused after each action and two options (one correct and one incorrect) appeared on the screen. The options, those are the possible next individual actions, were presented as videos one after the other, first on the left of the screen and then on the right. The distribution of the correct options was balanced between the right and left side of presentation. The incorrect options consisted of some of the illegal events of the fMRI session and some additionally new illegal events that had not been presented before during the imitation training sessions. The participants were instructed to decide which of the two presented actions would be the correct next action and give the response using two keys on their keyboard. The response time window was two seconds after the presentation of second video option (Fig. 5B). Once the participant gave a response or timed out, the video continued to play from where it paused and thus provided real-time feedback on the response. At the end of each action sequence, the participants were asked to rate how sure they had been to predict the upcoming steps on a Likert scale of 0 to 3 (0 – no idea, 1 – rather a guess, 2 – rather know, 3 – know for sure) and there was no time-out for this answer. The session lasted for approximately 35 to 45 min.

2.7. fMRI data acquisition

Neuroimaging data were acquired using a 3-Tesla Siemens Magnetom Prisma MR tomograph with a 20-channel head coil. Participants lay supine on the scanner bed and were instructed to remain still. Additionally, the fixation of the head, arms, and hands with form-fitting cushions minimized movements. During fMRI, we also recorded eye-tracking data using the Eyelink 1000 Plus eye tracker, which is not the focus of this manuscript and will not be reported further. Participants were provided with a response box with the right index and middle finger positioned on two buttons. Additionally, participants wore headphones and earplugs to attenuate scanner noise. Stimuli were presented using Presentation software 20.3 and participants watched the videos on the screen in 640×512 pixels resolution through a mirror mounted on the head coil. The structural T1 weighted images were acquired using a 3-D magnetization prepared rapid gradient echo sequence (MPRAGE) sequence with 192 slices, repetition time (TR)= 2130 ms, echo time (TE)= 2.28 ms, slice thickness= 1 mm, field of view (FoV)= 256×256 mm^2 and flip angle= 8° Following that, the functional data were obtained in an interleaved order, with the scanning parameters: 33 slices, TR= 2000 ms, TE= 30 ms, slice thickness= 3 mm, slice spacing= 1 mm, FoV= 192×192 mm^2, flip angle= 90°

2.8. fMRI data preprocessing

Structural and functional images were processed using Statistical Parametric Mapping 12 (SPM12; The Wellcome Centre for Human Neuroimaging, London, UK) implemented in MATLAB (MathWorks R2022a). Functional images were slice time corrected to the middle slice, then movement correction was applied, and images were realigned to the mean image. The structural image was co-registered with the mean functional image. Structural and functional images were then normalized to MNI space (Montreal Neurological Institute, Montreal, QC, Canada). Spatial smoothing was performed using a Gaussian kernel of Full-Width at Half Maximum (FWHM) 8 × 8 × 8 mm^3, and a temporal high-pass filter of 128 s was applied.

2.9. fMRI design specification

Statistical analyses were performed using SPM12. We used a general linear model (GLM) for serially autocorrelated observations (Friston et al., 1995; Worsley and Friston, 1995) and convolved regressors with the canonical hemodynamic response function. On the first level of the analysis, we used a grey matter mask that was comprised of smoothed individual normalized grey matter images (8 × 8 × 8 mm^3 FWHM) with a threshold of 0.2 implemented in ImCalc in SPM12.

A total of 12 regressors were included in the GLM. There were four regressors for the experimental conditions, namely one regressor for the legal action sequences (Legal), one for the illegal action sequences (Illegal), one for hand-object touching events (T), and another one for the hand-object untouching events (U). The fMRI experiment comprised three blocks of trials. Per block, all 41 legal action sequences were presented once, summing up to 123 legal action trials in total. The 13 illegal action sequences were presented in the first block, 14 in the second, and 14 in the third block. Out of 123 trials of legal action sequences presented across three blocks, 14 from the first block, 13 from the second, and 13 from the third that corresponded to the illegal action sequences were selected for the GLM. This was done to balance the number of trials in each regressors, namely, legal and illegal action sequences. Regressors Legal and Illegal included 41 trials each, while there were 450 trials each for T and U events. Additionally, two regressors for question events (164 trials) and null events (15 trials) were added. The onsets of the illegal action sequences were the point of expectation (or action repertoire) violation, and corresponding time points were determined in the legal sequences. The activity was modeled from the point of expectation violation to the end of the action sequences, T and U of the legal action sequences were modeled as events. Furthermore, a parametric regressor for U events with the number of branches (1, 2, 3, or 4) in decreasing order (i.e., the lowest number for the highest number of branches) was modeled. For the question trials and null events, the onsets were the beginning of each trial and modeled with a duration of zero. All parametric modulators were mean-centered (Mumford et al., 2015). Finally, the six subject-specific rigid-body transformations obtained from realignment (three translations and three rotations) were included as regressors of nuisance.

2.10. Whole brain fMRI analysis

At the first level, the following t-contrasts were computed: (a) To test the effect of violation of expectation, a contrast between Illegal and Legal events of the action sequences *(Illegal > Legal)* was calculated; (b) To investigate the brain activity involved in the retrieval of information from the internal model formed during the imitation training, the direct contrast *(U > T)* was calculated; (c) To examine the role of the probabilistic structure of the actions in legal action sequences, parametric modeling of number of branches was performed *(Branching)*; (d) To investigate neural changes during the update of the internal model when an action is initiated, the direct contrast *(T > U)* was generated.

For the second level analysis, one sample *t*-tests were performed across participants. A false discovery rate (FDR) threshold of $p < .05$ or lower was applied. For the FDR of $p < .05$, the results are reported with an additional moderate cluster threshold of 20 voxels. However, when there was no significant activation at FDR of $p < .05$, we inspected *t*-maps at $p < .001$ (uncorrected) for completeness. Brain activation was visualized with MRIcroGL (version 1.2.20211006, McCausland Center for Brain Imaging, University of South Carolina, USA).

2.11. Behavioral data analysis

Accuracy and reaction time (RT) data from the fMRI session and the post-fMRI memory test were analyzed using generalized linear modeling conducted in the R programming language (R Core Team, 2019; Version 4.2.1). For further RT analyses, only the trials with correct responses and RTs above 100 milliseconds were included. As the empirical distributions of RTs and accuracy rates were not normal (assessed by the Shapiro-Wilk-Test and the Cullen and Frey graph), for analyses of RTs and accuracy, a log-normal family and a binomial family were used, respectively. Notably, as the output is not normally distributed, it is not possible to use ANOVA or an equivalent general linear model (GLM) using a Gaussian family. For the behavioral results of the fMRI session, the GLM model (Eq. 1) contained the Stimulus Type as a fixed effect;

further, two random intercepts were assumed for subjects and trials. For the post-fMRI memory test, the GLM (Eq. 2) consisted of the Stimulus History of the to-be-rejected option (Legal, Illegal in the fMRI session) and Branching Structure (1, 2, 3, 4) as the fixed effects. For the post-hoc analyses, *p*-values are adjusted using Tukey's correction for multiple comparisons. fMRI Session:

$$Outcome \sim Stimulus\_Type + (1|ID) + (1|Trial) \quad (1)$$

Post-fMRI memory test:

$$Outcome \sim Branching\_Structure + Stimulus\_History + (1|ID) + (1|Trial) \quad (2)$$

3. Results

3.1. Behavioral results

3.1.1. fMRI session

During the fMRI session, participants made 76.5 % correct responses, 20.02 % incorrect responses, and 3.48 % of responses were missed due to time out. The main effects of the Stimulus Type with respect to accuracy $\chi 2(1) = 5.6824$, $p = 0.0171$, *Cramer's V* $= 0.029$ and RTs $\chi 2(1) = 80.07$, $p < 0.001$, *Cramer's V* $= 0.125$ were significant. Tukey-corrected post hoc contrasts showed that the accuracy was significantly higher for illegal events than the legal events ($z.ratio = 2.384, p = 0.0171$) and RT for legal events was significantly longer than that of the illegal events ($z.ratio = 8.948$, $p < 0.001$).

3.1.2. Post-fMRI memory test

In the post-fMRI memory test, participants' responses were correct in 89.31 % of trials, while 8.33 % of incorrect responses were recorded, and 2.36 % of responses were missed due to time out. There was no main effect of Stimulus History, meaning that actions which were part of other legal sequences and actions which were not derived from legal sequences were rejected with the same probability and equally fast.

The data from the post-fMRI memory test showed that there was a significant main effect of Branching Structure observed in accuracy $\chi 2(3) = 41.591$, $p < 0.001$, *Cramer's V* $= 0.055$ and RTs $\chi 2(3) = 36.05$, $p < 0.001$, *Cramer's V* $= 0.052$. Post hoc contrasts showed that the accuracy of one branch (95.7 % ± 0.203) was significantly higher compared to two banches (90 % ± 0.3) ($z.ratio = 3.332$, $p = 0.005$) and three branches (85.96 % ± 0.347) ($z.ratio = 6.205$, $p < 0.001$). There were high accuracy for four branches (93.87 % ± 0.24) than three branches (85.96 % ± 0.347) ($z.ratio = -3.265$, $p = 0.006$).

Further, regarding the RTs, the post hoc tests showed that actions with one branch ($M = 429 \pm 265.52$ ms) were significantly shorter than for those with three branches ($M = 442 \pm 292.96$ ms) ($z.ratio = -2.734, p = 0.031$), and longer than for actions with four branches ($M = 406.5 \pm 245.08$ ms) ($z.ratio = 3.116$, $p = 0.009$). Additionally, the RT for actions with two branches ($M = 413 \pm 240.84$ ms) were significantly shorter than for those with three branches ($M = 442 \pm 292.96$ ms) ($z.ratio = 1.658, p = 0.005$) and RTs for actions with three branches ($M = 442 \pm 292.96$ ms) were significantly longer than for those with four branches ($M = 406.5 \pm 245.08$ ms) ($z.ratio = 5.654$, $p < 0.001$). The actions with three branches had significantly longer RT than one, two and four branches and the actions with one branch had longer RT than four branches.

3.2. fMRI results

To investigate the effect of violation of expectation generated by the mismatch between predicted and perceived actions, the contrast between illegal and legal events of the action sequences *(Illegal > Legal)* was calculated. The contrast revealed activation in the bilateral supplementary motor area (SMA), supramarginal gyrus (SMG), and middle anterior insula (INS). In the right hemisphere, there was significant activation in the posterior superior temporal sulcus (pSTS) and thalamus (THA) (Fig. 6A, Table 1).

To determine the effects of the event boundary U of the legal action sequences, the first step of the two-step predictive process, where up to four possible courses of action could be expected depending on branching, a direct contrast *(U > T)* was calculated. This contrast yielded bilateral activation in SMA, INS/PUT, and lingual gyrus (LIN) (Fig. 6B, Table 2). The parametric modulator *(Branching)* at U events, reflecting the effects of the probabilistic branching structure of the action repertoire, revealed activation in the left fusiform gyrus (FFG), bilateral INS, and right anterior cingulate cortex (ACC) (Fig. 6C, Table 2). However, it is important to note that activation found for this effect did not survive correction for multiple comparisons and is therefore described at $p < .001$, uncorrected.

Additionally, to investigate the role of event boundary T of the legal action sequences, the time point when the action begins and which allows the restriction of the prediction to now only one option, the contrast *(T > U)* was calculated. This contrast showed significant activation in the left caudate (CAU), medial prefrontal cortex (MPFC), posterior cingulate cortex (PCC), and hippocampus (HIP), as well as the bilateral medial cingulate cortex (MCC) (Fig. 6D, Table 2).

As a control, we explored the effect of event boundaries U and T of the illegal action sequences. To this end, two contrasts *($T_i > U_i$)* and *($U_i > T_i$)* were calculated in a second GLM model. The *($U_i > T_i$)* showed activation of the bilateral superior temporal gyrus (STG) and bilateral lingual gyrus (LIN). Furthermore, the contrast *($T_i > U_i$)* did not survive multiple comparison corrections. However, the event boundaries, U and T, between the legal and illegal sequences could not be compared with the caveat of a highly unequal number of trials for the latter.

4. Discussion

Although actions appear to be a continuous stream of activities, they consist of meaningful segments. Studies have shown that the segmentation of actions takes place automatically and plays an important role in learning and memory through the formation of internal models (Eisenberg et al., 2018; Kurby and Zacks, 2008; Richmond and Zacks, 2018). Specifically, the boundaries where one action comes to an end and the next action begins are the essential time points for predictive processing (Schubotz et al., 2012; Swallow et al., 2009). The current study aimed to better understand the processes that happen at these boundaries, using a probabilistic branching structure of actions in a defined action repertoire. The results confirm the assumption of a two-stage updating process, which is also reflected at the neurofunctional level. These two updating steps involve, first, a set of frontal areas that seem to retrieve learned options of upcoming actions, and second, midline and hippocampal regions that reflect the finally selected episodic prediction of the emerging action. We also find that this updating is distinct at the neurofunctional level from, and likely upstream of, the response to a prediction error, caused by an unexpected action at the single action segment level that informs the so-called action observation network. Our findings contribute to a better understanding of action prediction in the context of complex action repertoires that are typical of our everyday lives.

The present study was designed to investigate prediction-updating processes at action boundaries by differentiating U and T events and examining the effects of the statistical properties of the action repertoire. In the first step, however, it was important to present the AON, which is well-reported to be underlying the processing of observed actions as a sanity check. To this end, we modified a single action in some of the presented action sequences, hypothesizing that such modifications induce a prediction error (Exton-McGuinness et al., 2015; Fernández et al., 2016; Jainta et al., 2022; Siestrup et al., 2023; Sinclair and Barense, 2019). In this case, the prediction had to quickly switch from complex expectation models of the suprasegmental structure to simpler expectation models that only call up the set of possible manipulations.

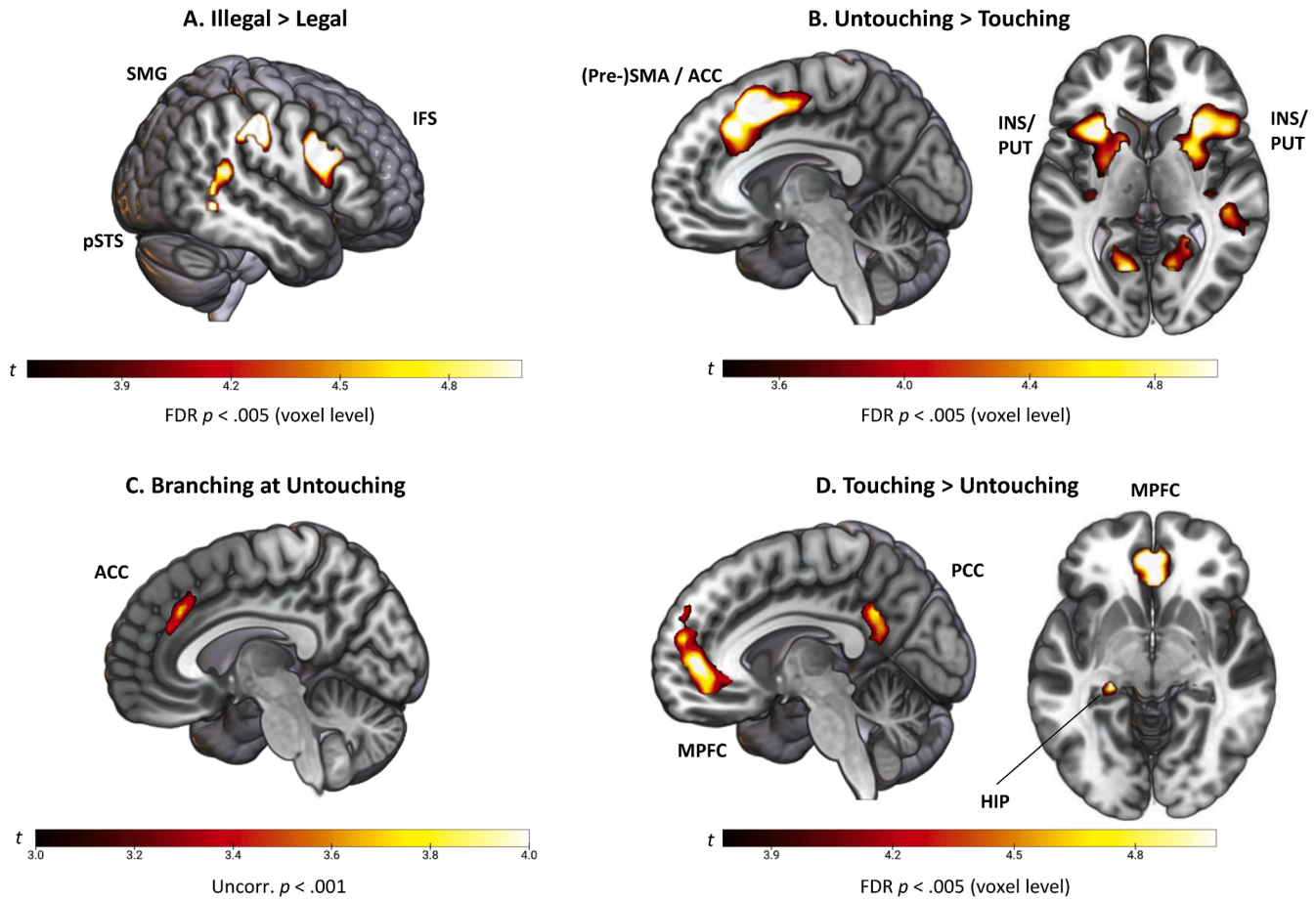

Fig. 6. Brain Activation A. Effect of Violation of Expectation – IFS, inferior frontal sulcus; SMG, supramarginal gyrus; pSTS, posterior superior temporal sulcus; B. Effect of Untouching events – Pre - SMA, Pre - supplementary motor area; INS/PUT, insula/putamen; C. Effect of Branching – ACC, anterior cingulate cortex; D. Effect of Touching events – PCC, posterior cingulate cortex; MPFC, medial prefrontal cortex; HIP, hippocampus.

Table 1
Whole-brain activation for contrasts of *Illegal* > *Legal* action sequences at FDR $p < 0.005$.

Anatomy	H	Cluster extent (voxels)	MNI Coordinates			t Value
			x {mm}	y {mm}	z {mm}	
Pre-Supplementary Motor Area	L	50	−3	17	53	5.14
	R	35	3	20	50	5.46
Supramarginal Gyrus/ anterior Inferior Parietal Sulcus	L	103	−42	−31	41	7.39
	R	265	54	−22	41	9.80
Anterior Dorsal Insula	L	75	−33	20	−4	6.26
	R	142	33	23	−1	8.87
Inferior Precentral Sulcus	L	124	−54	5	35	6.16
	R	562	51	11	29	7.80
Inferior Frontal Sulcus	R	l.m.	48	23	23	7.31
Inferior Temporal Gyrus	R	216	48	−55	−7	8.90
Posterior Superior Temporal Sulcus	R	l.m.	54	−40	14	5.71
Thalamus	R	39	9	−16	8	4.59

Only clusters with a minimum extent of 20 voxels are reported. *H*= Hemisphere; MNI= Montreal Neurological Institute; *L*= Left; *R*= Right; l.m.= local maximum.

As expected, these prediction errors at the level of single actions induced a BOLD response in the AON (Caspers et al., 2010). In addition, activation of the bilateral anterior insula increased in response to altered actions, indicating that participants recognized the modification as an action error (Bossaerts, 2018; Klein et al., 2007, 2013; Ullsperger et al., 2010). This neurophysiological response to error detection also corresponded to the significantly faster classification times for sequences with an altered action. It should be noted here that the prediction error only occurred at the level of a single action segment, while the sequence remained repertoire-compliant up to this point and thereafter. Consequently, no structures further upstream were involved in the processing of the prediction error.

At event boundaries, prediction errors accumulate (Kurby and Zacks, 2008), since the current event model no longer allows for precise prediction, especially when several next actions are possible. In the present study, this situation was implemented by U events in our multistep actions, where the learned action repertoire had to be accessed to form expectations with regard to the next action. At these U events, we found increased activation in SMA, lingual gyrus, putamen, and caudate nucleus. The striatum and the SMA are known to be involved in arranging sequences of action steps in the order in which they are to be executed (Bednark et al., 2015; Graybiel, 1998; Shima and Tanji, 2000; Tanji, 2001). The striatum plays a key role in the coding of action sequences; in particular, it serves to integrate individual actions into coherent, organized behavioral units, a process known as chunking (Favila et al., 2023; Graybiel, 1998). Studies in animals suggest that frontal input to the striatum, in particular the dorsomedial prefrontal cortex including SMA, shapes the representation of action sequences (Ostlund et al., 2009).

Table 2
Whole-brain activation for contrasts of Event boundaries *(Touching & Untouching)* and parametric modulator *Branching* at FDR $p < .005$/ uncorrected $p < .001$.

Anatomy	H	Cluster extent (voxels)	MNI Coordinates			t Value
			x {mm}	y {mm}	z {mm}	
Untouching > Touching (FDR at $p < .005$)						
(Pre-)Supplementary Motor Area/ Anterior Cingulate Cortex	L	1871	−3	14	50	7.79
	R	l.m.	12	8	62	7.86
Insula/ Putamen	L	414	−33	23	−4	6.23
	R	76	36	−22	8	4.78
Lingual Gyrus	L	117	−15	−67	−1	6.04
	R	218	24	−49	−1	5.63
Posterior Inferior Temporal Sulcus	R	133	54	−49	−7	5.71
Posterior Superior Temporal Sulcus	R	l.m.	48	−31	−1	5.08
Primary Motor Cortex	L	54	−36	−22	53	5.70
Anterior Middle Frontal Gyrus	L	36	−33	47	17	4.51
Branching (uncorrected at $p < .001$)						
Fusiform Gyrus	L	18	−30	−61	−10	4.99
Insula	L	29	−30	17	−10	4.04
	R	52	42	20	−7	4.76
Anterior Inferior Parietal Sulcus	R	59	48	−25	44	4.21
Middle Frontal Gyrus	R	11	39	50	20	4.00
Anterior Cingulate Cortex	R	40	6	38	23	3.97
Touching > Untouching (FDR at $p < .005$)						
Medial Prefrontal Cortex/ BA10/ Pregenual Anterior Cingulate Cortex	L	598	−3	50	−1	7.33
Dorsal Premotor Cortex	L	80	−24	−7	56	5.68
	R	121	30	−10	59	5.85
Mid Cingulate Sulcus	L	22	−15	−22	44	5.44
	R	23	9	−19	47	4.93
Posterior Cingulate Cortex	L	108	−9	−52	26	6.12
Fusiform Gyrus	L	41	−30	−55	−19	5.54
Hippocampus	L	25	−24	−28	−7	6.09

Only clusters with a minimum extent of 20 voxels for FDR at $p < .005$ and 10 voxels for uncorrected at $p < .001$ are reported. BA= Brodmann Area; *H*= Hemisphere; MNI= Montreal Neurological Institute; *L*= Left; *R*= Right; l.m.= local maximum.

Interestingly, rats with SMA lesions were still able to learn action sequences but failed when sequences had to be re-organized. Their behavior suggested that they represented the elements of the sequence as distinct behavioral units, corroborating a critical role of the SMA in sequence-level representations. Although the execution of highly trained action sequences probably depends more on the striatum than on the SMA, both areas seem to remain relevant, especially for the initiation of action chunks (Favila et al., 2023). Against the background of these findings, the joint activation of SMA and striatum at the beginning of each action boundary is highly plausible.

Moreover, increased activation detected in the lingual gyrus replicates previous findings where we investigated brain activity at boundaries within single-segment action (Pomp et al., 2021). At this point in time, there is an increased need to visually evaluate the current positions of the objects (i.e., the local scene), to employ this information during the prediction of the subsequent action in a specific action sequence (Kamps et al., 2016; Ruotolo et al., 2019; Sulpizio et al., 2013). More generally, increased exploratory vision and visual gain (Shipp, 2016) reflect processes that are typically associated with a prediction error in terms of the predictive coding framework.

In summary, these findings suggest that at the end of an action, when the object is released (U) to enable the initiation of the next action (T), the minimum of predictability is reached and counteracted by the retrieval of previously encoded event models, while at the same time, an intensive use of visual information already begins.

As described above, U events constituted the first of a two-step predictive process in the viewing of multistep actions. At these event boundaries, the brain uses the before-trained probabilistic structure of actions to update predictions of upcoming actions. Our rationale was that this process should depend on the number of possible following actions, as operationalized as the number of branches in the present experiment. The number of branches ranged from one to four, while a situation with only one possible branch represents the lowest level of uncertainty or strongest prediction. With regard to the BOLD effects, the parametric analysis of the branching structure suggested that as the strength of the prediction increased, activity increased in the ACC and the anterior insula. Since this contrast did not survive the correction for multiple comparisons, we refrain from discussing this observation in detail. Also, although the accuracy and RT measures from the post-fMRI test were significantly influenced by the branching structure, the directions of these effects were rather mixed. However, we would like to point out that this observation fits very well with the previous findings that ACC was more active when the internal model was biased towards a specific prediction (Akam et al., 2021; Kennerley et al., 2006; Klein--Flügge et al., 2022; Schiffer et al., 2013) and when participants were able to make predictions on the basis of previously trained sequences compared to observing completely new ones (Jainta et al., 2022; Siestrup et al., 2022, 2023; Siestrup and Schubotz, 2023). Thus, in the present study, higher ACC activation for more restricted expectations might reflect a more straightforward retrieval of available predictive models at U events.

After retrieving information from the internal model at the end of an action (U event), new information that supplies predictive processes is provided when a new action is initiated upon touching the next object (T event). In the case of our task, we assume that the expected options retrieved at U are matched with the now observed hand movements of the actress, leading to a final selection of the best matching option, which is broadcast as the newly expected action. This second step in the prediction updating process was characterized by increased BOLD responses in cortical midline structures MPFC and PCC, fitting our hypothesis regarding the role of DMN areas in updating predictions at the suprasegmental levels of action organization. Notably, the observed brain activation could be related to comparing own predictive models with the unfolding sequence at T events to aid further prediction (Koster-Hale and Saxe, 2013; Wurm et al., 2011). Moreover, recently proposed hierarchical prediction and prediction error processing models (Alexander and Brown, 2015, 2018) suggest that the MPFC provides high-level predictive input to the lateral frontal cortex, thus, DMN areas may provide top-down modulatory input to the AON to update and shape predictions of upcoming object transport and manipulation. Future studies should test this further by addressing causal network interactions using for instance dynamic causal modeling (Stephan et al., 2010). Interestingly, we also found hippocampal activation at T events. The hippocampus is known to be sensitive to event boundaries (Ben-Yakov and Henson, 2018; Cooper and Ritchey, 2020). It has an important role in the generation of internal models and predictions (Barron et al., 2020), and operates as a match/mismatch detector between situational input and predictions (Duncan et al., 2009, 2012; Sinclair et al., 2021) Thus, hippocampal activation may reflect the comparison of the previously formed internal model and new information concerning the relevant object.

In summary, these findings suggest that at the beginning of the next action (T events), predictions that have been prepared during finalizing the previous action (U events) are now evaluated in reference to the unfolding observed action to exclude all but one prediction, involving a

higher-level network of cortical midline and hippocampal areas.

5. Conclusion

Actions we observe in our daily life are organized in sequences of individual segments, and this segmental structure is highly relevant for prediction processes. In the present study, we showed that at U events, when the current event model comes to an end, the brain must retrieve previously formed internal models to generate predictions about the upcoming events. The observed neural responses also reflect the level of prediction strength. When new information about the continuation of an action sequence becomes available at T events, the brain evaluates the match between selected predictions and evidence from the observed action. In summary, this study provides valuable new insights into the stepwise predictive processes during the observation of complex manual action repertoires.

CRediT authorship contribution statement

Rosari Naveena Selvan: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Minghao Cheng:** Writing – original draft, Software, Methodology, Data curation. **Sophie Siestrup:** Writing – review & editing, Writing – original draft, Software, Resources. **Falko Mecklenbrauck:** Software, Methodology. **Benjamin Jainta:** Software, Resources, Formal analysis. **Jennifer Pomp:** Visualization, Validation, Resources. **Anoushiravan Zahedi:** Writing – review & editing, Formal analysis. **Minija Tamosiunaite:** Writing – review & editing, Methodology, Conceptualization. **Florentin Wörgötter:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Ricarda I. Schubotz:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare no competing interest.

Data availability

Unthresholded statistical maps of all reported and visualized fMRI contrasts in the manuscript have been deposited on NeuroVault (https://www.neurovault.org/collections/CGYMMSBZ/). The entire stimulus material are made available via the Action Video Corpus Muenster (AVICOM, https://www.uni-muenster.de/IVV5PSY/AvicomSrv/). The raw fMRI data and the raw SEC time point extraction data that support the findings of this study are available from the corresponding author upon reasonable request.

Funding

This work was supported by the German Research Foundation (DFG) [WO 388/13–1]. The funder had no role in study design, data collection, analysis and interpretation, decision to publish, or writing of the report.

Acknowledgments

We would like to thank Monika Mertens, Christin Schwarzer, Lena Puder, Maria Leonie Hofmann, Simon Wieczorek, Kristin Stroop, Lars Schlüter, Leon Exeler and Niklas Dielitzsch for helping to recruit participants and/or assisting during data collection for the pilot and final studies.

References

Akam, T., Rodrigues-Vaz, I., Marcelo, I., Zhang, X., Pereira, M., Oliveira, R.F., Dayan, P., Costa, R.M., 2021. The anterior cingulate cortex predicts future states to mediate model-based action selection. Neuron 109, 149–163. https://doi.org/10.1016/j.neuron.2020.10.013.

Aksoy, E.E., Abramov, A., Dörr, J., Ning, K., Dellen, B., Wörgötter, F., 2011. Learning the semantics of object-action relations by observation. Int. J. Rob. Res. 30 (10), 1229–1249. https://doi.org/10.1177/0278364911410459.

Alexander, W.H., Brown, J.W., 2015. Hierarchical error representation: a computational model of anterior cingulate and dorsolateral prefrontal cortex. Neural Comput. 27, 2354–2410. https://doi.org/10.1162/NECO_a_00779.

Alexander, W.H., Brown, J.W., 2018. Frontal cortex function as derived from hierarchical predictive coding. Sci. Rep. 8 (1), 1–11. https://doi.org/10.1038/s41598-018-21407-9.

Baldassano, C., Hasson, U., Norman, K.A., 2018. Representation of real-world event schemas during narrative perception. J. Neurosci. 38 (45), 9689–9699. https://doi.org/10.1523/JNEUROSCI.0251-18.2018.

Barron, H.C., Auksztulewicz, R., Friston, K., 2020. Prediction and memory: a predictive coding account. Prog. Neurobiol. 192, 101821 https://doi.org/10.1016/j.pneurobio.2020.101821.

Bednark, J.G., Campbell, M.E.J., Cunnington, R., 2015. Basal ganglia and cortical networks for sequential ordering and rhythm of complex movements. Front. Hum. Neurosci. 9, 1–13. https://doi.org/10.3389/fnhum.2015.00421 (JULY).

Ben-Yakov, A., Henson, R.N., 2018. The hippocampal film editor: sensitivity and specificity to event boundaries in continuous experience. J. Neurosci. 38 (47), 10057–10068. https://doi.org/10.1523/JNEUROSCI.0524-18.2018.

Bossaerts, P., 2018. Formalizing the function of anterior insula in rapid adaptation. Front. Integr. Neurosci. 12, 1–9. https://doi.org/10.3389/fnint.2018.00061.

Botvinick, M., Plaut, D.C., 2004. Doing without schema hierarchies: a recurrent connectionist approach to normal and impaired routine sequential action. Psychol. Rev. 111 (2), 395–429. https://doi.org/10.1037/0033-295X.111.2.395.

Caspers, S., Zillers, K., Laird, A.R., Eickhoff, S.B., 2010. ALE meta-analysis of action observation and imitation in the human brain. Neuroimage 50 (3), 1148–1167. https://doi.org/10.1016/j.neuroimage.2009.12.112.ALE.

Colder, B., 2011. Emulation as an integrating principle for cognition. Front. Hum. Neurosci. 5 (54), 1–12. https://doi.org/10.3389/fnhum.2011.00054.

Cooper, R.A., Ritchey, M., 2020. Progression from feature-specific brain activity to hippocampal binding during episodic encoding. J. Neurosci. 40 (8), 1701–1709. https://doi.org/10.1523/JNEUROSCI.1971-19.2019.

Csibra, G., Gergely, G., 2007. "Obsessed with goals": functions and mechanisms of teleological interpretation of actions in humans. Acta Psychol. 124, 60–78. https://doi.org/10.1016/j.actpsy.2006.09.007.

Duncan, K., Curtis, C., Davachi, L., 2009. Distinct memory signatures in the hippocampus: intentional states distinguish match and mismatch enhancement signals. J. Neurosci. 29 (1), 131–139. https://doi.org/10.1523/JNEUROSCI.2998-08.2009.

Duncan, K., Ketz, N., Inati, S.J., Davachi, L., 2012. Evidence for area CA1 as a match/mismatch detector: a high-resolution fMRI study of the human hippocampus. Hippocampus 22 (3), 389–398. https://doi.org/10.1002/hipo.20933.

Eisenberg, M.L., Zacks, J.M., Flores, S., 2018. Dynamic prediction during perception of everyday events. Cogn. Res. 3 (53), 1–12. https://doi.org/10.1186/s41235-018-0146-z.

Exton-McGuinness, M.T.J., Lee, J.L.C., Reichelt, A.C., 2015. Updating memories-the role of prediction errors in memory reconsolidation. Behav. Brain Res. 278, 375–384. https://doi.org/10.1016/j.bbr.2014.10.011.

Favila, N., Gurney, K., Overton, P.G., 2023. Role of the basal ganglia in innate and learned behavioural sequences. Rev. Neurosci. 35 (1), 35–55. https://doi.org/10.1515/revneuro-2023-0038.

Fernández, R.S., Boccia, M.M., Pedreira, M.E., 2016. The fate of memory: reconsolidation and the case of prediction error. Neurosci. Biobehav. Rev. 68, 423–441. https://doi.org/10.1016/j.neubiorev.2016.06.004.

Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. B 360 (1456), 815–836. https://doi.org/10.1098/rstb.2005.1622.

Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11 (2), 127–138. https://doi.org/10.1038/nrn2787.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P, Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2 (4), 189–210. https://doi.org/10.1002/hbm.460020402.

Friston, K., Kiebel, S., 2009. Predictive coding under the free-energy principle. Philos. Trans. R. Soc. B 364 (1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300.

Graybiel, A.M., 1998. The basal ganglia and chunking of action repertoires. Neurobiol. Learn. Mem. 70, 119–136. https://doi.org/10.1006/nlme.1998.3843.

Jainta, B., Siestrup, S., El-Sourani, N., Trempler, I., Wurm, M.F., Werning, M., Cheng, S., Schubotz, R.I., 2022. Seeing what i did (not): cerebral and behavioral effects of agency and perspective on episodic memory re-activation. Front. Behav. Neurosci. 15, 1–18. https://doi.org/10.3389/fnbeh.2021.793115.

Jocher, G., 2020. YOLOv5 By Ultralytics. Github. https://github.com/ultralytics/yolov5.

Kamps, F.S., Julian, J.B., Kubilius, J., Kanwisher, N., Dilks, D.D., 2016. The occipital place area represents the local elements of scenes. Neuroimage 132, 417–424. https://doi.org/10.1016/j.neuroimage.2016.02.062.

Kennerley, S.W., Walton, M.E., Behrens, T.E.J., Buckley, M.J., Rushworth, M.F.S., 2006. Optimal decision making and the anterior cingulate cortex. Nat. Neurosci. 9 (7), 940–947. https://doi.org/10.1038/nn1724.

Kilner, J.M., Friston, K.J., Frith, C.D., 2007. Predictive coding: an account of the mirror neuron system. Cogn. Process 8 (3), 159–166. https://doi.org/10.1007/s10339-007-0170-2.
Kilner, J.M., Vargas, C., Duval, S., Blakemore, S.J., Sirigu, A., 2004. Motor activation prior to observation of a predicted movement. Nat. Neurosci. 7 (12), 1299–1301. https://doi.org/10.1038/nn1355.
Klein, T.A., Endrass, T., Kathmann, N., Neumann, J., von Cramon, D.Y., Ullsperger, M., 2007. Neural correlates of error awareness. Neuroimage 34 (4), 1774–1781. https://doi.org/10.1016/j.neuroimage.2006.11.014.
Klein, T.A., Ullsperger, M., Danielmeier, C., 2013. Error awareness and the insula: links to neurological and psychiatric diseases. Front. Hum. Neurosci. 7, 1–14. https://doi.org/10.3389/fnhum.2013.00014.
Klein-Flügge, M.C., Bongiosnni, A., Rushworth, M.F.S., 2022. Medial and orbital frontal cortex in decision-making and flexible behavior. Neuron 220 (17), 2743–2770. https://doi.org/10.1016/j.neuron.2022.05.022.
Koster-Hale, J., Saxe, R., 2013. Theory of mind: a neural prediction problem. Neuron 79 (5), 836–848. https://doi.org/10.1016/j.neuron.2013.08.020.
Kurby, C.A., Zacks, J.M., 2008. Segmentation in the perception and memory of events. Trends Cogn. Sci. 12 (2), 72–79. https://doi.org/10.1016/j.tics.2007.11.004.
Masís-Obando, R., Norman, K.A., Baldassano, C., 2022. Schema representations in distinct brain networks support narrative memory during encoding and retrieval. eLife 11, 1–25. https://doi.org/10.7554/eLife.70445.
Michels, T., 2022. Multi_Camera_Calibration. Gitlab. https://gitlab.com/ungetym/Multi_Camera_Calibration.
Mumford, J.A., Poline, J.B., Poldrack, R.A., 2015. Orthogonalization of regressors in FMRI models. PloS one 10 (4), e0126255. https://doi.org/10.1371/journal.pone.0126255.
Newtson, D., 1973. Attribution and the unit of perception of ongoing behavior. J. Pers. Soc. Psychol. 28 (1), 28–38. https://doi.org/10.1037/h0035584.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4.
Olson, E. (2011). AprilTag: a robust and flexible visual fiducial system. Proceedings - IEEE International Conference on Robotics and Automation, 3400–3407. https://doi.org/10.1109/ICRA.2011.5979561.
Ondobaka, S., Bekkering, H., 2012. Hierarchy of idea-guided action and perception-guided movement. Front. Psychol. 3, 1–5. https://doi.org/10.3389/fpsyg.2012.00579.
Ostlund, S.B., Winterbauer, N.E., Balleine, B.W., 2009. Evidence of action sequence chunking in goal-directed instrumental conditioning and its dependence on the dorsomedial prefrontal cortex. J. Neurosci. 29 (25), 8280–8287. https://doi.org/10.1523/JNEUROSCI.1176-09.2009.
Pesquita, A., Whitwell, R.L., Enns, J.T., 2018. Predictive joint-action model: a hierarchical predictive approach to human cooperation. Psychon. Bull. Rev. 25 (5), 1751–1769. https://doi.org/10.3758/s13423-017-1393-6.
Pezzulo, G., Rigoli, F., Friston, K.J., 2018. Hierarchical active inference : a theory of motivated control. Trends Cogn. Sci. 22 (4), 294–306. https://doi.org/10.1016/j.tics.2018.01.009.
Pomp, J., Heins, N., Trempler, I., Kulvicius, T., Tamosiunaite, M., Mecklenbrauck, F., Wurm, M.F., Wörgötter, F., Schubotz, R.I., 2021. Touching events predict human action segmentation in brain and behavior. Neuroimage 243 (March), 118534. https://doi.org/10.1016/j.neuroimage.2021.118534.
Reagh, Z.M., Ranganath, C., 2023. Flexible reuse of cortico-hippocampal representations during encoding and recall of naturalistic events. Nat. Commun. 14 (1279), 1–15. https://doi.org/10.1038/s41467-023-36805-5.
Richmond, L.L., Zacks, J.M., 2018. Constructing experience. Event models from perception to action 21 (12), 962–980. https://doi.org/10.1016/j.tics.2017.08.005.
Ruotolo, F., Ruggiero, G., Raemaekers, M., Iachini, T., van der Ham, I.J.M., Fracasso, A., Postma, A., 2019. Neural correlates of egocentric and allocentric frames of reference combined with metric and non-metric spatial relations. Neuroscience 409, 235–252. https://doi.org/10.1016/j.neuroscience.2019.04.021.
Sasaki, A.T., Okamoto, Y., Kochiyama, T., Kitada, R., Sadato, N., 2018. Distinct sensitivities of the lateral prefrontal cortex and extrastriate body area to contingency between executed and observed actions. Cortex 108, 234–251. https://doi.org/10.1016/j.cortex.2018.08.003.
Schiffer, A.M., Ahlheim, C., Ulrichs, K., Schubotz, R.I., 2013. Neural changes when actions change: adaptation of strong and weak expectations. Hum. Brain Mapp. 34 (7), 1713–1727. https://doi.org/10.1002/hbm.22023.
Schubotz, R.I., Korb, F.M., Schiffer, A.M., Stadler, W., von Cramon, D.Y., 2012. The fraction of an action is more than a movement: neural signatures of event segmentation in fMRI. Neuroimage 61 (4), 1195–1205. https://doi.org/10.1016/j.neuroimage.2012.04.008.
Shima, K., Tanji, J., 2000. Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. J. Neurophysiol. 84 (4), 2148–2160. https://doi.org/10.1152/jn.2000.84.4.2148.
Shipp, S., 2016. Neural elements for predictive coding. Front Psychol 7, 1–21. https://doi.org/10.3389/fpsyg.2016.01792.
Siestrup, S., Jainta, B., Cheng, S., Schubotz, R.I., 2023. Solidity meets surprise: cerebral and behavioral effects of learning from episodic prediction errors. J. Cogn. Neurosci. 35 (2), 291–313. https://doi.org/10.1162/jocn_a_01948.
Siestrup, S., Jainta, B., El-Sourani, N., Trempler, I., Wurm, M.F., Wolf, O.T., Cheng, S., Schubotz, R.I., 2022. What happened when? Cerebral processing of modified structure and content in episodic cueing. J. Cogn. Neurosci. 34 (7), 1287–1305. https://doi.org/10.1162/jocn_a_01862.
Siestrup, S., Schubotz, R.I., 2023. Minor changes change memories: functional magnetic resonance imaging and behavioral reflections of episodic prediction errors. J. Cogn. Neurosci. 35 (11), 1823–1845. https://doi.org/10.1162/jocn_a_02047.
Sinclair, A.H., Barense, M.D., 2019. Prediction error and memory reactivation: how incomplete reminders drive reconsolidation. Trends Neurosci. 42 (10), 727–739. https://doi.org/10.1016/j.tins.2019.08.007.
Sinclair, A.H., Manalili, G.M., Brunec, I.K., Alison Adcock, R., Barense, M.D, 2021. Prediction errors disrupt hippocampal representations and update episodic memories. Proc. Natl. Acad. Sci. USA 118 (51), 1–12. https://doi.org/10.1073/pnas.2117625118.
Sommer, T., Hennies, N., Lewis, P.A., Alink, A., 2022. The assimilation of novel information into schemata and its efficient consolidation. J. Neurosci. 42 (30), 5916–5929. https://doi.org/10.1523/JNEUROSCI.2373-21.2022.
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. Hum. Brain Mapp. 32 (5), 677–687. https://doi.org/10.1002/hbm.20949.
Stephan, K.E., Penny, W.D., Moran, R.J., den Ouden, H.E.M., Daunizeau, J., Friston, K.J., 2010. Ten simple rules for dynamic causal modeling. Neuroimage 49 (4), 3099–3109. https://doi.org/10.1016/j.neuroimage.2009.11.015.
Sulpizio, V., Committeri, G., Lambrey, S., Berthoz, A., Galati, G., 2013. Selective role of lingual/parahippocampal gyrus and retrosplenial complex in spatial memory across viewpoint changes relative to the environmental reference frame. Behav. Brain Res. 242 (1), 62–75. https://doi.org/10.1016/j.bbr.2012.12.031.
Swallow, K.M., Zacks, J.M., Abrams, R.A., 2009. Event boundaries in perception affect memory encoding and updating. J. Experim. Psychol. 138 (2), 236–237. https://doi.org/10.1037/a0015631.
Tanji, J., 2001. Sequential organization of multiple movements: involvement of cortical. Annu. Rev. Neurosci. 24, 631–651.
Uithol, S., van Rooij, I., Bekkering, H., Haselager, P., 2012. Hierarchies in action and motor control. J. Cogn. Neurosci. 24 (5), 1077–1086. https://doi.org/10.1162/jocn_a_00204.
Ullsperger, M., Harsay, H.A., Wessel, J.R., Ridderinkhof, K.R., 2010. Conscious perception of errors and its relation to the anterior insula. Brain Struct Funct 214, 629–643. https://doi.org/10.1007/s00429-010-0261-1.
Urgen, B.A., Pehlivan, S., Saygin, A.P., 2019. Distinct representations in occipito-temporal, parietal, and premotor cortex during action perception revealed by fMRI and computational modeling. Neuropsychologia 127, 35–47. https://doi.org/10.1016/j.neuropsychologia.2019.02.006.
Urgen, B.A., Saygin, A.P., 2020. Predictive processing account of action perception: evidence from effective connectivity in the action observation network. Cortex 128, 132–142. https://doi.org/10.1016/j.cortex.2020.03.014.
Wahlheim, C.N., Eisenberg, M.L., Stawarczyk, D., Zacks, J.M., 2022. Understanding everyday events: predictive-looking errors drive memory updating. Psychol. Sci. 33 (5), 765–781. https://doi.org/10.1177/09567976211053596.
Wörgötter, F., Aksoy, E.E., Krüger, N., Piater, J., Ude, A., Tamosiunaite, M., 2013. A simple ontology of manipulation actions based on hand-object relations. IEEE Trans. Auton. Ment. Dev. 5 (2), 117–134. https://doi.org/10.1109/TAMD.2012.2232291.
Worsley, K.J., & Friston, K.J. (1995). Analysis of fMRI time-series revisited - again. In NeuroImage (Issue 2, pp. 173–181). https://doi.org/10.1006/nimg.1995.1023.
Wurm, M.F., von Cramon, D.Y., Schubotz, R.I., 2011. Do we mind other minds when we mind other minds' actions? A functional magnetic resonance imaging study. Hum. Brain Mapp. 32 (12), 2141–2150. https://doi.org/10.1002/hbm.21176.
Zacks, J.M., Kurby, C.A., Eisenberg, M.L., Haroutunian, N., 2011. Prediction error associated with the perceptual segmentation of naturalistic events. J. Cogn. Neurosci. 23 (12), 4057–4066. https://doi.org/10.1162/jocn_a_00078.
Zacks, J.M., & Sargent, J.Q. (2010). Event perception: a theory and its application to clinical neuroscience. In Psychology of Learning and Motivation - Advances in Research and Theory (1st ed.). Elsevier Inc. https://doi.org/10.1016/S0079-7421(10)53007-X.
Zacks, J.M., Speer, N.K., Swallow, K.M., Braver, T.S., Reynolds, J.R., 2007. Event perception: a mind/brain perspective. Psychol. Bull. 133 (2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273.Event.
Ziaeetabar, F., Pomp, J., Pfeiffer, S., El-Sourani, N., Schubotz, R.I., Tamosiunaite, M., Wörgötter, F., 2021. Using enriched semantic event chains to model human action prediction based on (minimal) spatial information. PLoS One 15, 1–25. https://doi.org/10.1371/journal.pone.0243829.