What happened when? Brain and behavioral responses to violated expectations about the structure and content of episodic memories

Sophie Siestrup[1,2*], Benjamin Jainta[1], Nadiya El-Sourani[1], Ima Trempler[1,2], Oliver T. Wolf[c], Sen Cheng[d], Ricarda I. Schubotz[1,2]

[1] Department of Psychology, University of Münster, Münster, Germany,

[2] Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany

[3] Department of Cognitive Psychology, Faculty of Psychology, Ruhr University Bochum, Bochum, Germany

[4] Institute for Neural Computation, Ruhr University Bochum, Bochum, Germany

*Corresponding author

E-mail: s.siestrup@uni-muenster.de (SS)

Abstract

Episodic memories are not static but can be modified on the basis of new experiences, potentially allowing us to make valid predictions in the face of an ever-changing environment. Recent research has identified mnemonic prediction errors as a possible trigger for such modifications. In the present study, we investigated the influence of different types of mnemonic prediction errors on brain activity and subsequent memory performance using a novel paradigm for episodic modification. Participants encoded different episodes which consisted of short toy stories. During a subsequent functional magnetic resonance imaging (fMRI) session, episodic retrieval was cued by presenting videos showing the original episodes, or modified versions thereof. In modified videos either the order of two subsequent action steps was changed (violating structure expectancy) or an object was exchanged for another (violating content expectancy). While brain responses to structure expectancy violations were only subtle, content expectancy violations recruited brain areas relevant for processing of new object information. In a post-fMRI memory test, the participants' tendency to accept modified episodes as originally encoded increased significantly when they had experienced expectancy violations during the fMRI session. Our study provides valuable initial insights into the neural processing of different types of mnemonic prediction errors and their influence on subsequent memory.

Introduction

Episodic memories are our window to the past. They enable us to vividly re-live events that we experienced at some point in our personal life, allowing us to re-locate them in space and time [1]. But there is evidence that episodic memories are not always an exact, veridical reconstruction of our past. On the contrary, memories are flexible and prone to change [2, 3, 4, 5]. It is often believed that the modification of episodic memories is undesirable and occurs due to erroneous processes during memory encoding, consolidation or retrieval [6, 7]. However, a valuable new perspective is that such modifications serve an important function for our predictive capabilities [8, 9, 10]. Prediction is thought to be based on internal models [11] derived from memories. What is crucial for predictive success is not only that we establish and recall appropriate internal models of the world, but also attune them to persistent change to maintain valid predictions in the long run. This is accomplished via prediction errors that serve as bottom-up learning signals for model updating [12, 13, 14]. Accordingly, it is likely adaptive that largely faithful memories are modified in favor of valid internal models informed and updated by later experiences [10, 15]. Such prediction errors could arise in everyday life if the episodic memory recalled in a situation only partially resembles this new situation, which is probably the normal case. Accordingly, these differences could become effective as prediction errors in the process of memory reconstruction. Although some have suggested that prediction errors can also lead to an improvement of memories under certain circumstances [16, 17, 18], potentially as they can trigger one-shot learning [17], evidence is also accumulating that violated expectations are an important driver of memory change, consistent with the theoretical framework mentioned above [19]. In addition to behavioral studies, researchers have begun to investigate how the human brain processes mnemonic prediction errors. For example, some have found evidence that the hippocampus responds to unexpected outcomes that contradict previously established memories, and thus may play an important role in memory updating based on new relevant information [20, 21].

Memory traces include different types of information about an experienced episode, i.e., “what” happened “when” and “where” [21]. To our knowledge, previous studies have not yet specifically targeted different types of mnemonic prediction errors based on separate components of the memory trace. Moreover, reported memory tasks are often not very naturalistic, employing simple stimuli (e.g. word or picture lists) and encoding periods with little participant involvement (e.g. learning a list or watching a video; e.g. [22, 23]). In the present study, we therefore aimed to extend the understanding of mnemonic prediction errors at the neural and behavioral level by selectively targeting predictions about two different aspects of episodic memories, either their content (“what”) or their structure (“when”). To do so, we employed a new naturalistic paradigm for encoding episodic memories in the laboratory, then violated expectations based on reactivated episodes in two different ways and measured neural responses using functional magnetic resonance imaging (fMRI). Episodes were reactivated by videos reminiscent of episodes that each participant had encoded in a training session one week before entering the fMRI session. A subset of these videos was selectively manipulated with regard to either the occurrence of an object (content modification) or the order of two consecutive action steps (structure modification). Importantly, these modifications were rather subliminal as they neither challenged general world knowledge nor the overall outcome of the episodes. Following the paradigm implemented by Schiffer et al. [24, 25], modified episodes were repeatedly presented to simulate the natural circumstances which potentially foster memory modification, i.e. updating of internal models due to increasing evidence for the validity of an alternative. In the present study, episodes were comprised of short stories played with *PLAYMOBIL®* toys. We chose this approach to allow for complex but controlled, laboratory-encoded episodes which are largely unbiased by previous everyday experiences.

In a post-fMRI memory test, participants were presented videos with episodes reminiscent of the original training session or modified versions thereof. Participants were asked to indicate on a four-point Likert scale whether the presented video shows an episode from the training session or not. We expected that mnemonic prediction errors experienced during scanning should trigger internal

model updating [24, 25] and thus lead to a decrease in memory performance afterwards (H1), which we examined using the discrimination index P_r. More specifically, we might find that repeated expectancy violation during the fMRI session leads to an increased acceptance of modified videos as original episodes (H1a), similar to previous behavioral findings reporting misattribution of new items to old memories [20, 22]. Moreover, we might even see a reduced acceptance of unmodified videos as a result of updating (H1b) [26, 27]. We tested these assumptions by analyzing rating scores from the memory test. Since we expected to find differential brain activation between the two types of episodic modifications (as described below), we might also see that the type of modification further modulates memory performance which we also wanted to investigate in a more exploratory manner.

Concerning brain activation, we expected that some regions would react to both types of expectancy violations (H2). Common activation was investigated by first calculating the separate contrasts of the two types of expectancy violation against original videos, i.e. the no-expectancy violation condition (structure expectancy violation vs. original, content expectancy violation vs. original), and combining them through conjunction analysis. Several brain networks have been described to underlie the retrieval of declarative memories regardless of their specific content [28]. Particularly, the frontomedian cortex (mPFC) is a candidate area that may serve more general control over consolidation and retrieval of long-term memories [29, 30]. Furthermore, the hippocampus is regarded a core structure of episodic memory [31, 32, 33] and memory modification is believed to be accomplished by the hippocampus linking reactivated content to new experience during reactivation, and delivering updated memories to the frontomedian cortex [34]. Additionally, as outlined above, the hippocampus has already been identified as a region that responds to mnemonic prediction errors [20, 21].

The episodes employed in the present study could be classified in two different ways, as stories showing events experienced by toy figure actors/actresses, but also as playing actions performed by an actor/actress. As to the latter interpretation, a hypothesis was that brain areas related to the observation of actions may also be engaged for videos showing modified episodes, including lateral

premotor cortex, the intraparietal sulcus (IPS), the ventrolateral prefrontal cortex, and the lateral occipitotemporal cortex (OTC), together coined 'AON' for "action-observation network" [35]. Accordingly, we found previously that activity in all of these four areas attenuated when subjects adapted to an expectancy violation affecting both objects and sequential structure *combined* [24, 25]. Taken together, the AON, the mPFC and the hippocampal formation were expected to respond to both types of expectancy violations.

Additionally, we expected to also find specific activations for structure and content episodic expectancy violations, which we investigated by calculating the above-mentioned separate contrasts.

First, we hypothesized that structure-expectancy violation leads to elevated activity in the lateral premotor cortex [36] and mesial premotor cortex (called supplementary motor area (SMA) [37]) due to their role in sequential association of single manipulations. We further expected to find elevated brain activity in the area of dorsal premotor cortex and adjacent prefrontal sites along the superior frontal sulcus (PMd/SFS) for structure expectancy violation. This region is characteristically active in stepwise ordinal linking of individual action steps, as required in quite different, but especially predictive, tasks [38, 39, 40, 41, 42](H3).

Content expectancy violation was expected to lead to a particularly strong response in the lateral OTC [43], in the anterior IPS [44, 45], and, in addition, in the fusiform gyrus (FG [46]), due to their complex object-action-focused functional profile (H4).

Apart from the expected common activation in the ventrolateral prefrontal cortex, we might find different response patterns in this area for the two types of episodic expectancy violation. For example, it has been suggested that within the inferior frontal gyrus (IFG), Brodmann area 44 (BA 44) may be especially involved in processing the sequential structure of actions [47, 48] whereas BA 45 may be more relevant for semantic content [49, 50].

Materials and Methods

Participants

Forty female participants took part in the study. Only females participated to guarantee a high level of self-identification with our stimulus material which showed female hands. Two participants did not return for the fMRI session after training. Additionally, data from one participant had to be excluded from analysis due to nausea and dizziness during the fMRI session. In another case, an unknown technical problem led to the incomplete acquisition of functional data. Thus, the final sample for fMRI analyses consisted of 36 participants, age 18 – 28 years (*M* = 22.67 years, *SD* = 2.40 years). Participants of the final sample were all right-handed as assessed by the Edinburgh Handedness Inventory [51]. Handedness scores varied from +60 to +100 (*M* = 92.17, *SD* = 10.95). Data from one participant was not used for the analysis of the post-fMRI memory test due to misunderstood instructions. All participants had normal or corrected-to-normal vision, were native German speakers, and reported no history of neurological or psychiatric disorders or substance abuse. Participants signed a written informed consent and received monetary compensation or course credits for their participation. The study was conducted in accordance with the Declaration of Helsinki and approved by the Local Ethics Committee of the University of Münster.

Stimuli

Even though episodic memory can potentially never be completely independent of semantic knowledge [52] we wanted to reduce influences of everyday experiences as far as possible in our study. To this end we chose to use abstract but complex toy-based stories so that we could ensure controlled encoding of unique episodes in the laboratory. At the end of the study, participants were asked whether the stories used in the study reminded them of experiences they made in everyday life and 61.1 % answered no.

Stimuli were 78 short videos (duration = 8.80 – 17.88 s, *M* = 12.71 s) of stories that were played with *PLAYMOBIL®* toys, showing only the toys and hands and underarms of an actress. Stories comprised

six to nine action steps (*M* = 7.4 steps) and four to 14 separable objects (*M* = 6.93 objects), such as characters, animals, vehicles and tools. The exact same object appeared in only one of the stories. Most participants were familiar with *PLAYMOBIL®* toys, as 78 % of them indicated in the first session that they had played with these toys during their childhood. Those that had not played with *PLAYMOBIL®* toys had all played with similar toys.

Stories were filmed from above with a digital single-lens reflex camera (Nikon D5300) which was centrally mounted above the table and faced straight down. Matt white paper served as a base. A frame of 47.5 cm x 28 cm was taped on the paper, congruent with the section captured by the camera (in the following referred to as camera frame). Objects that were needed for a particular story were positioned next to the camera frame and were only moved into view in the moment at which they appeared in the story. During filming, the actress wore a black pullover and black rubber gloves to mask individual physical characteristics. To facilitate future imitation from demo videos, the back of the right hand was marked with a yellow dot [53]. For a schematic overview of the filming setup, see Fig 1A. Video material was edited using Adobe Premiere Pro CC (Adobe Systems Software, Dublin, Ireland, Version 12.1.2). All videos had a frame of size 1920 x 1080 pixels and a frame rate of 25 frames per second. Videos started with seven frames showing only background and ended after seven frames showing the final toy constellation. The original perspective of the videos was the third-person (observer) perspective (3pp) and they were rotated 180° to create their first-person (field) perspective (1pp) counterparts using Adobe Premiere Pro CC [54] (Fig 1B). Throughout the experiment, videos were presented at a visual angle of approximately 7.3° x 13° using the stimulus presentation software Presentation (version 20.3 02.01.19, NeuroBehavioral Systems, Berkeley, CA, USA).

Fig 1. Creation of stimulus material. (A) An actress wearing a black pullover and black gloves (right hand marked with a yellow dot) was filmed from above while playing stories with toy objects within a

marked frame. (B) Videos were rotated by 180° to create videos in the first-person perspective (1pp) using Adobe Premiere Pro CC.

On the basis of two pilot studies, we chose 24 out of originally 30 stories for our stimulus set. Stories were excluded when they were particularly difficult to imitate or describe. We made sure that all chosen stories were comparably memorable. We did this by presenting a questionnaire that included the correct description of each story and additionally one lure description (similar objects and context, but different story) to a separate sample of 16 participants and checked for outliers in the discrimination index P_r for each story. One of the 30 stories was excluded due to low memorability.

The 24 final stories existed in three different versions each: (1) an original version (*ori*) as encoded by the participants, (2) a version in which two adjacent action steps were switched (structure modification) to elicit structure-based expectancy violations (*str*) and (3) another variation of the original video in which one object was exchanged (content modification) to elicit content-based expectancy violations (*con*). Story scripts were created by five experimenters who all had to agree that the original story as well as modifications thereof were semantically valid (within a toy world). This was important since we did not want to violate general semantic expectations. For creating videos with modifications, the respective stories were played and filmed again exactly the way as for the original video. The only aspect that differed between original and modified versions was a single change of either the order of two action steps (i.e., one transition out of 7.33 transitions on average, for structure modifications) or one object (i.e., one object out of 6.95 objects on average, for content modifications). Modifications were never introduced in the first two action steps to allow for expectation build-up prior to modifying the videos and an associated prediction error. Furthermore, no modifications were introduced in the last two action steps either. For examples of different story versions, see Fig 2 and S1 Table. The exact time point of the modification in each video (in the following referred to as time of violation) was determined by identifying the video frame which diverged from the original version.

Fig 2. Different story versions used in the study. Twenty-four stories existed in three different versions each: an original, a structure modification, and a content modification. For the structure modifications, two adjacent action steps were switched compared to the original (here: step 4 and 5). For the content modifications, an object was exchanged compared to the original (here: tortoise instead of guinea pig on the sales counter in step 5). Note that in the fMRI experiment, each participant was only presented with one of the three versions of a story.

Six additional stories existed in only one version. Four of those served as so-called novels in the fMRI session to provide an additional control condition, which will be discussed in the paper by Jainta et al. (in preparation). The two remaining videos were used for practice trials and did not appear in the fMRI experiment. The stimulus material will be made available at the Action Video Corpus Muenster (https://www.uni-muenster.de/IVV5PSY/AvicomSrv/). Some example videos are available in an OSF repository (https://osf.io/34knw/?view_only=3194f4ad2fa54a0fa47f4189b2aa1f77).

Procedure

Training

Training sessions were conducted in a computer lab at the Department of Psychology at the University of Münster. The training consisted of two sessions that took place on two consecutive days and lasted about 2 h and 1.5 h, respectively. During each of the two sessions, participants encoded half of the episodes. We chose to split the training over two days to avoid fatigue or a decrease in motivation due to the relatively long duration of the task. For encoding episodes, participants imitated half of the 24 stories from demo videos (i.e., the original versions) and only observed the other half. Please note that this experimental factor (*agency*) will not be further addressed in this paper, but in a companion paper (Jainta et al., in preparation).

The 24 demo videos were organized in four subsets, containing six videos each, balanced for the number of action steps (*A1*, *A2*, *B1*, *B2*). Participants either imitated videos belonging to *A* subsets and observed videos belonging to *B* subsets or vice versa. In each training session, participants went through one imitation block and one observation block, each consisting of one subset of videos. This means that each participant trained each video either during session one or during session two, the same video was not trained on both days. The order of videos within these blocks was randomized for each participant. All stories were imitated and only observed equally often over the course of the study. The first training session started with a short practice phase of the task. Then, half of the participants started with imitation, the other half with observation. The same applied to the second training session, except that participants who had started with imitation on the first day now started with observation, and vice versa. The order of blocks and the order in which video subsets were trained was balanced between participants. Participants always went through each story eight times, including four times in 3pp and another four times in 1pp. If the story was to be imitated, participants first watched the story four times in 3pp, then once in 1pp and then performed the last three repetitions of the story themselves. If the story was only observed, the video was presented three more times in 1pp. The optimal number of video presentations for successful imitation was determined in a pilot study. Fig 3 depicts the structure of the training.

Fig 3. Training procedure. Training consisted of two sessions on two consecutive days (middle row). The 24 videos to be encoded were organized in four subsets with 6 videos each (*A1*, *A2*, *B1*, *B2*). Participants either imitated videos belonging to *A* subsets and observed videos belonging to *B* subsets or vice versa. In each training session, participants went through one imitation block and one observation block, each consisting of one subset of videos. Hence, Participants imitated and observed half of the stories each (balanced between participants), equally divided between the two sessions. In an imitation block (top row), each demo video was shown four times in 3pp and once in 1pp, followed by three correct imitations by the participant. In an observation block (bottom row),

participants watched each demo video four times in both perspectives. After each video, the participant had to give a detailed description of the story (not depicted here).

During training, the participant sat at the same setup which had been used for filming the stimulus material and likewise wore a black pullover and gloves with a yellow dot on the right hand. The experimenter sat opposite of the participant, supervising the performance (Fig 4). For each story, the toys were positioned next to the camera frame, following the same arrangement as used while creating the stimulus material. After imitation or observation, participants had to deliver a detailed description of the story to ensure that they understood it correctly and had paid attention to all objects involved. If participants made a mistake during an imitation or description trial, they were immediately interrupted by the experimenter to avoid encoding of incorrect scripts. They would then start over with a new imitation/description attempt.

Fig 4. Training setup. During training participants observed and imitated toy stories, while sitting at the filming setup. Their performance was monitored by the experimenter.

Cover-story

Participants were told that videos of themselves playing the stories would be presented in the fMRI session. Even though participants were actually filmed, these videos were not used during the fMRI experiment. This was only a cover story to elevate personal identification with the videos to benefit episode reactivation. In a previous pilot study, we examined whether participants could be convinced that they are shown in the videos one week after training. Participants were asked for all videos (presented in pseudo-random order) whether they believed to be shown themselves, rated on a scale from 0 – 4 (0 = no, 1 = rather no, 2 = rather yes, 4 = yes). We found a significant main effect of

agency (repeated measures analysis of variance, $F_{(1,15)} = 33.130$, $p < 0.001$, $\eta p^2 = 0.688$, $n = 16$), driven by higher identification ratings for imitated than only observed videos ($M_{imi} = 2.89$, $SD_{imi} = 0.44$; $M_{obs} = 2.12$, $SD_{obs} = 0.59$), which showed that participants were convinced by our cover story. In our current study, participants were asked to indicate in percent in how many videos they had been depicted themselves during the fMRI session. On average, they believed that 40.5 % percent of the videos showed themselves (*SD* = 21.8 %). Please note that this cover story was most relevant for the experimental factor *agency* which is the subject of a different paper. Participants were fully debriefed after completion of the study.

fMRI session

The fMRI session took place approximately one week after the second training session (range = 7 - 12 days; *M* = 7.42 days, *SD* = 0.90 days). Participants were presented with original and modified videos reminiscent of the previously encoded episodes. Importantly, each video was only shown in the original or one divergent version. Thus, eight videos were repeatedly presented in the original version, eight included a structure modification and eight a content modification. Which stories belonged to which conditions was counterbalanced between participants, so that over the course of the study, all stories equally often belonged to all experimental conditions. Half of the stories in the original and modification conditions were shown in 1pp, the other half in 3pp. The perspective of videos was counterbalanced between participants. This experimental factor (*perspective*) is not the subject of the present paper, but of another paper by Jainta et al. (in preparation). Additionally, four novel stories that had not been part of the training were included in the fMRI session to serve as an additional control.

The fMRI experiment consisted of six blocks, each containing the 24 videos reminiscent of the previously encoded episodes. Consequently, each video was presented six times over the course of the session. Within blocks, videos were presented in pseudo-random order. Additionally, each block contained three null events during which only a fixation cross was presented (duration: 7 – 10 s). Furthermore, each novel video was presented once per block. Therefore, the whole experiment

contained 18 null events and 24 novel video trials. Participants were not informed about the block-structure of the experiment.

Participants were instructed to attentively watch the presented videos. They were told that after some videos a short description would be presented (e.g. “Rescuing princess”) that either matched or did not match the story shown in the video (question trials). The task was to either accept or reject the description by pressing one of two buttons on a response box with the right index or middle finger, respectively. This type of task has been used successfully before to focus participants’ attention on complex video stimuli [55]. The descriptions represented the stories on a gist-level, so specific details were irrelevant to the correct answer (Fig 5). Throughout the entire experiment, each story was once followed by a matching description and once by a non-matching description, resulting in a total number of 56 question trials in the experiment. Each block contained nine to ten question trials and per block, approximately 50 % of descriptions were to be accepted, 50 % were to be rejected. The question was presented for a maximum of 3 s or until participants responded. Upon response delivery, participants received 1 s written feedback whether they answered correctly, incorrectly, or too late, in case no response was given. Participants were naïve with regard to this distribution of question trials.

Between trials, a fixation cross was presented for a duration of 2 s (1 s after question trials) to serve as an interstimulus interval. Before each trial, a variable jitter of 0, 0.5, 1 or 1.5 s of fixation was added for enhancement of the temporal resolution of the BOLD response (Fig 5).

Fig 5. Schematic depiction of task during fMRI session. Video trials consisted of a variable jitter (0, 0.5, 1, or 1.5 s of fixation), a video showing a toy story (ca. 9 – 18 s) and a 2 s interstimulus interval (fixation). Question trials included a variable jitter, a question regarding the story shown in the preceding video (maximally 3 s long or terminated by response), and a 2 s interstimulus interval. The interstimulus interval after question trials was divided into 1 s feedback (“correct”, “incorrect”, “too

late”) and 1 s fixation. Aside from the question, it was depicted which button should be pressed to accept (left, green) or reject (right, red) the description.

In total, the fMRI task had a duration of approximately 48 min. The task had previously been practiced at the end of the second training session (four video trials, four question trials, one null event). Practice videos were shown twice each and not re-used in the fMRI session.

Memory test

Immediately after the fMRI session, participants completed an explicit memory test. Importantly, encoding occurred incidentally, as participants were not informed beforehand that their memory for episodes would be tested.

Participants were seated in a separate room in front of a laptop and instructed to remember their training sessions one week prior. They were presented all stories that they had seen in the fMRI session in two different versions. More precisely, when expectations had been violated during the fMRI session by the presentation of episode videos containing a structure modification, these modified videos were presented again during the memory test and *additionally* each story was shown in the original version. The same applied when expectations had been violated by content modifications. When expectations had not been violated, i.e., when original episode videos had been presented during the fMRI scan, these original videos were presented again in the memory test and *additionally*, each story was shown in a modified version, *either* containing a structure modification in half of the cases *or* a content modification.

The participants’ task was to rate after each video whether they knew this exact episode from the training, using a Likert scale including 1 (“*yes*”), 2 (“*rather yes*”), 3 (“*rather no*”) and 4 (“*no*”), by pressing one out of four marked keys on the laptop’s keyboard. Response time was not restricted but participants were instructed to respond quickly and intuitively. Videos were always presented in the same perspective (1pp or 3pp) as during the fMRI session and in a pseudo-randomized order, so that

half of the stories (of each experimental condition) were first presented in their original version followed by a modified version and vice versa. Novels were shown twice in the same version, so that, in total, the memory test comprised 56 video trials. The completion of the task took approximately 15 min.

MRI data acquisition and preprocessing

MRI scans were conducted with a 3-Tesla Siemens Magnetom Prisma MR tomograph (Siemens, Erlangen, Germany) using a 20-channel head coil. Participants lay supine on the scanner bed with their right index and middle finger positioned on the two appropriate buttons on a response box. Head, arm and hand movements were minimized by tight fixation with form-fitting cushions. Participants were provided with earplugs and headphones to attenuate scanner noise. Stimuli were projected on a screen that the participants saw via an individually adjusted mirror which was mounted on the head coil.

High resolution T1 weighted anatomical images were obtained with a 3D-multiplanar rapidly acquired gradient-echo (MPRAGE) sequence prior to functional imaging. 192 slices with a thickness of 1 mm were acquired, using a repetition time (TR) of 2130 ms, an echo time (TE) of 2.28 ms, a flip angle of 8° and a field of view (FoV) of 256 x 256 mm^2. Functional images of the whole brain were acquired in interleaved order along the AC–PC plane using a gradient-echo echoplanar imaging (EPI) sequence to measure blood-oxygen-level-dependent (BOLD) contrast. Thirty-three axial slices with a thickness of 3 mm were obtained in an interleaved order, using a TR of 2000 ms, a TE of 30 ms, a FoV of 192 x 192 mm^2 and a flip angle of 90°.

Processing of imaging data was conducted with SPM12 (Wellcome Trust, London, England) implemented in MATLAB (version R2018b, The MathWorks Inc., Natick, MA, USA). Data were preprocessed by slice time correction to the middle slice, movement correction and realignment to the mean image, co-registration of the functional data to individual structural scans, normalization of functional and structural images into the standard MNI space (Montreal Neurological Institute,

Montreal, QC, Canada) on the basis of segmentation parameters and spatial smoothing using a Gaussian kernel of full-width at half maximum (FWHM) of 8 mm. Further, a 128 s high-pass temporal filter was applied.

Statistical data analysis

Behavioral data analysis

The behavioral data analysis was conducted using RStudio ([56]; version 1.3.1073).

After completion of the memory test task, one participant reported that she had not understood it correctly, so data from this participant was excluded from the analysis (n = 35).

For the analysis of P_r values (general memory performance; unmodified videos = targets, modified videos = distractors) and ratings/reaction times for modified videos in the memory test, we applied a 2 x 2 within-subject factorial design with the factors *violation* (*yes* = expectancy violation during fMRI session; *no* = no expectancy violation) and *alternative* (*str* = structure modification included in memory test; *con* = content modification included), giving rise to the four factorial combinations *no,str* (1), *no,con* (2), *yes,str* (3) and *yes,con* (4). Please note that due to the described allocation of videos to factorial combinations, there were always twice as many trials belonging to factor level *yes* (8 trials) than *no* (4 trials). For each factorial combination, we calculated the discrimination index P_r (hit rate minus false alarm rate [57]), the mean rating score, and the mean reaction time for correct responses. In cued memory tasks, reaction time has traditionally been used as a measure for how long it takes to retrieve information from memory [58]. Longer reaction times are thought to indicate increased difficulty of retrieval due to higher cognitive processing demands [59, 60].

For analyzing responses to original (unmodified) videos in the memory test, videos were allocated to the three different conditions *ori* (no expectancy violation during fMRI session), *str* (structure expectancy violation during fMRI session) and *con* (content expectancy violation during fMRI session). Mean rating scores as well as mean reaction times were calculated. Additionally, we

416 calculated the mean rating scores for novel videos. Please note that low mean rating scores indicate
417 high acceptance, while high mean rating scores mean low acceptance.

418 As a next step, we conducted an explorative analysis on behavioral data from the fMRI session. We
419 calculated the error rate and mean reaction time for each condition (*ori, str, con*) per participant. In
420 only 0.7 % of all question trials no answer was given, so that such trials were not further considered
421 in the analysis.

422 Data were inspected for normal distribution using the Shapiro Wilk Test. Further, data were checked
423 for extreme outliers as defined as values above quartile 3 + 3 x interquartile range (*IQR*) or lower
424 than quartile 1 - 3 x *IQR*. When data were normally distributed or could be transformed to fit normal
425 distribution (reaction times for original videos, reaction times during fMRI task; logarithmic
426 transformation) and showed no extreme outliers, we used parametric analysis approaches (one-way
427 repeated measures analyses of variance (rmANOVA)). Post-hoc pair-wise comparisons were
428 conducted with paired *t*-tests (one-tailed when comparing *ori* and *str* and *ori* and *con* conditions,
429 two-tailed when comparing *str* and *con* conditions). When the prerequisites for parametric analysis
430 were not met, we used a non-parametric two-way rmANOVA based on aligned rank data (package
431 *ARTool* [61]; P_r values, ratings of modified videos) or the Friedman test for one-way approaches
432 (ratings of original videos, error rates in fMRI task). Several participants did not give any correct
433 answers (i.e. rejection) in response to modified videos in the post-fMRI memory test for one or more
434 factorial combinations. To avoid the listwise exclusion of many datapoints for reaction times when
435 using a rmANOVA approach, we decided to use one-tailed Wilcoxon signed rank tests when
436 comparing reaction time between levels of the factor *violation*, separately for each modification
437 type. This way, we could include 16 datapoints for structure and 29 datapoints for content modified
438 videos. We additionally excluded another datapoint for content modifications due to a very long
439 mean reaction time in the *yes* condition (8.85 s) which likely does not represent purely mnemonic
440 processes. Please note that this did not influence the overall outcome of the analysis.

When parametric statistical approaches were used, we report (untransformed) mean values and standard errors of the mean as descriptive statistics, when non-parametric approaches were applied, we report median values and *IQR*. For all behavioral analyses, we applied a significance level of α = 0.05. *P*-values were adjusted according to the Bonferroni correction for multiple comparisons [62]. Raw and processed behavioral data have been deposited in an OSF repository (https://osf.io/34knw/?view_only=3194f4ad2fa54a0fa47f4189b2aa1f77).

fMRI design specifications

Statistical analyses of the fMRI data were conducted with SPM12. We used a general linear model (GLM) for serially autocorrelated observations [63, 64] and convolved regressors with the canonical hemodynamic response function. Regressors were original videos (*ori*), videos containing a structure modification (*str*) and videos containing a content modification (*con*), each comprising 48 trials. For *str* and *con* trials, the onsets of events were time-locked to the point in the video at which the modification occurred (time of violation). For *ori* trials, we calculated a hypothetical time of violation (mean of times that corresponded to points of structure and content modification in the non-modified video) to serve as a comparable onset. These conditions were modeled as events as we were interested in the phasic effect of the expectancy violation at the precise moment it occurred. To each of those regressors, we added a parametric modulator to model the repeated presentation of each video. Two additional regressors modeled the 18 null events and the 56 question trials. The modeled activation of null events and questions was time-locked to their respective onsets. Null events were modeled as epochs, containing their full presentation time (7 – 10 s), while questions were modeled as events. The six subject-specific rigid-body transformations obtained from realignment were included as regressors of no interest. Therefore, the GLM comprised 14 regressors in total. Please note that factors which were not considered in the present paper (*agency*, *perspective*) were perfectly balanced over the factors of interest so that we can rule out any confounding effects (fully-crossed design).

466 We calculated first-level-*t*-contrasts for *str* > *ori* and *con* > *ori* to analyze brain activity in response to
467 the specific expectation violation types. To exploratively examine whether the BOLD response to the
468 violations attenuated over repetitions, we additionally considered *t*-contrasts for the parametric
469 regressors. Grey matter masking was applied on the first level of the analysis. For masking, we used
470 the smoothed individual normalized grey matter image (8 mm FWHM) which was thresholded at 0.2
471 using ImCalc in SPM12 to create a binary mask. Second-level group analyses were performed by
472 using one-sample *t*-tests across participants. A conjunction of *str* > *ori* and *con* > *ori* contrasts was
473 calculated to detect shared effects of both expectancy violations [65]. We applied false discovery
474 rate (FDR) correction and a threshold of $p < 0.05$ (peak level). In case no significant activation was
475 detected using this threshold, we applied no correction for multiple comparison with $p < 0.001$. This
476 is a common procedure when specific neuroanatomical hypotheses are examined [66, 67, 68, 69].
477 Brain activation patterns were visualized with the software MRIcroGL (version 1.2.20200331,
478 McCausland Center for Brain Imaging, University of South Carolina, USA). Unthresholded statistical
479 maps have been uploaded to NeuroVault.org ([70], https://neurovault.org/collections/SKPCADYC/).

480 To explore whether IFG activity is modulated by the two types of expectancy violation, we performed
481 region of interest (ROI) analyses for two different IFG areas (pars opercularis, BA 44; pars triangularis,
482 BA 45). Separate structural ROIs were generated per IFG area and hemisphere using the automated
483 anatomical labeling atlas with the WFU PickAtlas toolbox [71] in SPM12. Mean beta values for the *str*
484 > *ori* and *con* > *ori* contrasts were extracted from each ROI using the MarsBaR toolbox [72]. The
485 conditions for parametric analysis were confirmed as described above for behavioral data. One-tailed
486 one-sample *t*-tests were used to test whether mean beta values per ROI and contrast were greater
487 than zero. For each ROI, mean beta values were compared between *str* > *ori* and *con* > *ori* using two-
488 tailed paired *t*-tests. There were no specific hypotheses regarding differences between left and right
489 hemispheres, so we did not compare them directly. Because we analyzed four ROIs per contrast and
490 performed multiple pair-wise comparisons, we applied the Bonferroni correction [62] and report
491 adjusted *p*-values, for both one-sample and paired *t*-tests. The extracted mean beta values were

deposited in an OSF repository (https://osf.io/34knw/?view_only=3194f4ad2fa54a0fa47f4189b2aa1f77).

Results

Post-fMRI memory test

Discrimination index P_r

We found a significant main effect of *violation* ($F_{(1,34)} = 20.220$, $p < 0.001$, $\eta p^2 = 0.373$) on P_r values which was driven by higher P_r values for *no* (*Median* = 0.50, *IQR* = 0.50) compared to *yes* (*Median* = 0.19, *IQR* = 0.50). There was also a significant main effect of *alternative* ($F_{(1,34)} = 85.828$, $p < 0.001$, $\eta p^2 = 0.716$), which was explained by higher P_r values for *con* (*Median* = 0.63, *IQR* = 0.38) than for *str* (*Median* = 0.06, *IQR* = 0.25). There was no significant interaction of *violation* and *alternative* ($F_{(1,34)} = 2.074$, $p = 0.159$, $\eta p^2 = 0.058$) (Fig 6A).

Ratings and reaction times for modified videos

We found a significant main effect of *violation* on mean rating scores for modified videos ($F_{(1,34)} = 25.554$, $p < 0.001$, $\eta p^2 = 0.429$), which was explained by higher ratings for *no* (*Median* = 2.5, *IQR* = 1.75) than for *yes* (*Median* = 2, *IQR* = 1.75). Thus, when an expectancy violation had occurred during the fMRI session, participants were more likely to accept modified versions afterwards. Additionally, there was a significant main effect of *alternative* on mean rating scores for modified videos ($F_{(1,34)} = 122.647$, $p < 0.001$, $\eta p^2 = 0.783$). Higher ratings were detected for *con* (*Median* = 3.25, *IQR* = 1.375) than for *str* (*Median* = 1.563, *IQR* = 1.094), which means that participants were generally more prone to accept structurally modified videos than those with modified content. There was no significant interaction of the two factors *violation* and *alternative* ($F_{(1,34)} = 2.010$, $p = 0.165$, $\eta p^2 = 0.056$; Fig 6B).

We found a near-significant trend that reaction times were longer for structurally modified videos when this version had already been presented in the fMRI session ($Z = -1.80$, $p = 0.072$; Fig 6C). No

significant difference was found concerning reaction times for videos with content modification (Z = -1.089, p = 0.276; Fig 6D).

Ratings and reaction times for unmodified videos

Mean rating scores for unmodified videos did not differ significantly between conditions ($\chi 2_{(2)}$ = 2.124, p = 0.346, *Kendall's W* = 0.030). Descriptively, rating scores were slightly higher when repeated expectancy violation had occurred during the fMRI session ($Median_{ori}$ = 1.0, IQR_{ori} = 0.313; $Median_{str}$ = 1.25, IQR_{str} = 0.375; $Median_{con}$ = 1.125, IQR_{con} = 0.5), which might indicate a tendency to accept original videos slightly less after repeated expectancy violation (Fig 6E).

There was a significant main effect of condition on mean reaction times for correctly classifying unmodified videos ($F_{(2,68)}$ = 3.725, p = 0.029, ηp^2 = 0.099). Pairwise comparisons revealed that reaction times were significantly longer after structure expectancy violation in the scanner than when no expectancy violation had occurred ($t_{(34)}$ = 3.287, p = 0.004). No significant differences were detected between *con* and *ori* ($t_{(34)}$ = 1.779, p = 0.126) and *str* and *con* conditions ($t_{(34)}$ = 0.725, p = 1; Fig 6F).

Fig 6: Behavioral results from post-fMRI memory test. (A) Discrimination index P_r from post-fMRI memory test. Original (unmodified) videos were the targets while modified videos (*str* = structure modification, *con* = content modification) were distractors. Statistics: non-parametric rmANOVA based on aligned rank data. a vs. b = p < 0.001 (main effect *violation*); *** = p < 0.001 (main effect *alternative*). (B) Mean rating scores for modified videos for each factorial combination. Statistics: non-parametric rmANOVA based on aligned rank data. a vs. b = p < 0.001 (main effect *violation*); *** = p < 0.001 (main effect *alternative*. (C) Mean reaction times for structure modified videos, n = 16. Statistics: Wilcoxon signed rank tests. (D) Mean reaction times for content modified videos, n = 28. (E) Mean rating scores for unmodified videos for each condition (*ori* = no expectation violation during fMRI session, *str* = structure expectancy violation, *con* = content expectancy violation). Statistics:

Friedman test. (F) Mean reaction times for unmodified videos for each condition. Statistics: rmANOVA, post-hoc pairwise comparisons: paired *t*-tests, *p*-values were corrected for multiple comparisons. ** = $p < 0.01$. Box plots show medians, quartiles, minima and maxima. Bar plots show means and standard errors.

Ratings for novel videos

High ratings for novel videos, i.e., videos which were new during the fMRI session (*Median* = 4, *IQR* = 1), indicate that participants correctly remembered that these videos had not been part of the training.

Behavioral performance during fMRI session

Error rates of the task during the fMRI session were generally very low ($Median_{ori} = 0$, $IQR_{ori} = 0$; $Median_{con} = 0$, $IQR_{con} = 0.02$; $Median_{str} = 0$, $IQR_{str} = 0$) and did not differ significantly between conditions ($\chi 2_{(2)} = 0.520$, $p = 0.771$, *Kendall's W* = 0.007). We also did not find a significant main effect of condition on mean reaction times during the fMRI session ($F_{(2,70)} = 0.983$, $p = 0.380$, $\eta p^2 = 0.027$). These findings indicate that the participants' attention to (and understanding of) the stories was very good and not deteriorated by the switch of single action steps or objects. Importantly, this shows that any neural differences between these conditions are not simply based on confusion or generally elevated cognitive effort in respect to expectancy violations. All questions in the fMRI session asked for the gist of the stories. We can thus assume that neither the structure, nor the content modification changed this gist compared to the original story.

fMRI results

First, we investigated whether structure and content expectation violations elicit common brain activation patterns compared to original episodes. The conjunction of the whole brain contrast *str* > *ori* and *con* > *ori* did not yield significant activation after correction for multiple comparisons.

564 Without correction and at a threshold of $p < 0.001$, there was significant activation in the opercular
565 part of the right IFG which corresponds to BA 44 ($x = 51$, $y = 17$, $z = 20$, 44 voxels).

566 To identify which brain regions are active for structure expectancy violation, we calculated the whole
567 brain contrast *str* > *ori*. This contrast did not yield significant activation after correction for multiple
568 comparisons. Without correction and at a threshold of $p < 0.001$, activation clusters were detected in
569 the right hemisphere in SFS, PMd and the opercular part of the IFG (BA 44) (Fig 7 and Table 1).

570 To detect brain regions which respond to content expectancy violations, we calculated the whole
571 brain contrast *con* > *ori*. Significant bilateral activation was found in pre-SMA and the middle IPS.
572 Additionally, there were activation clusters in left and right inferior frontal sulcus (IFS) which
573 expanded into the triangular part of the IFG (BA 45) and the inferior frontal junction (IFJ). There was
574 significant bilateral activation in the anterior insula. Finally, elevated brain activity was detected in
575 the occipitotemporal cortex (OTC), bilaterally including FG and extending into parahippocampal gyrus
576 (PHG) in the right hemisphere (Fig 7 and Table 1).

577

578 **Fig 7. Whole brain activation for expectancy violation effects.** (A) Uncorrected *t*-map ($p < 0.001$) for
579 the *str* > *ori* contrast. (B) FDR-corrected *t*-map ($p < 0.05$) for the *con* > *ori* contrast. IFG: inferior
580 frontal gyrus; SFS: superior frontal sulcus; PMd: dorsal premotor cortex; IFS: inferior frontal sulcus;
581 pre-SMA: pre-supplementary motor area; IPS: intraparietal sulcus; AI: anterior insula; FG: fusiform
582 gyrus; PHG: parahippocampal gyrus.

583

584 **Table 1. Peak activations from second-level whole-brain analyses of expectancy violation effects.**

Localization	H	Cluster extent	MNI Coordinates			*t*-value
			x	y	z	
str* > *ori (uncorrected at $p < 0.001$)						

Superior frontal sulcus	R	16	27	23	59	3.72
Dorsal premotor cortex	R	48	39	2	56	4.74
Inferior frontal gyrus (pars opercularis, BA 44)	R	70	54	14	20	4.63
con* > *ori (FDR-corrected at $p < 0.05$)						
Pre-supplementary motor area	L	49	-6	14	53	4.16
	R	64	9	17	47	4.11
Intraparietal sulcus	R	294	30	-61	50	6.18
	L	241	-27	-64	38	5.41
Inferior frontal junction	L	320	-39	2	32	6.38
Inferior frontal sulcus, extending into inferior frontal gyrus (pars triangularis, BA 45)	L		-42	20	26	5.51
	R	505	45	29	20	6.99
Anterior insula	R	134	33	20	-1	5.58
	L	143	-30	20	-4	5.70
Occipitotemporal cortex	L	236	-45	-55	-13	7.48
Fusiform gyrus	L		-27	-40	-22	3.87
Occipitotemporal cortex, including fusiform gyrus and extending into parahippocampal gyrus	R	259	33	-52	-13	5.10

Note: H = Hemisphere, MNI = Montreal Neurological Institute, L = Left, R = Right, *str* = structure expectancy violation, *con* = content expectancy violation, *ori* = no expectancy violation.

For an exploratory, more detailed analysis of IFG activity, we extracted beta values from different IFG areas (pars opercularis, BA 44; pars triangularis, BA 45) for the *str* > *ori* and *con* > *ori* contrasts. As we did not have specific hypotheses concerning differences between left and right hemisphere, we extracted beta values separately for both hemispheres but did not directly compare them. One-sample *t*-tests against zero revealed significant activation for content expectancy violations in left and right BA 45 as well as in right BA 44. For structure expectancy violations, significant activation was only found for right BA 44. There were near-significant trends for activation ($p < 0.01$) in left BA 44 for content expectancy violations and in right BA 45 for structure expectancy violations (Fig 8). For

a detailed overview of all one-sample *t*-tests, see S3 Table. Paired *t*-tests showed that mean beta values were significantly higher for content expectancy violations compared to structure expectancy violations in left and right BA 45 (left: $t_{(35)} = 4.545$, $p < 0.001$; right: $t_{(35)} = 4.005$, $p < 0.001$). We did not detect significant differences between conditions in BA 44 (left: $t_{(35)} = -2.053$, $p = 0.191$; right: $t_{(35)} = -0.771$, $p = 1$; Fig 8).

Fig 8. Region of interest analysis of left (A) and right (B) opercular (BA 44) and triangular (BA 45) part of the IFG. Mean beta values for the contrasts *str* > *ori* and *con* > *ori* were extracted from subregions of the IFG as defined by the automated anatomical labeling atlas (depicted on the right). Bar plots show means and standard errors. One-tailed one-sample *t*-test were used to determine whether values were greater than zero (symbols above bars). Two-tailed paired *t*-tests were used to determine differences between conditions (symbols above lines). *P*-value were corrected for multiple comparisons. ** = $p < 0.01$; *** = $p < 0.001$.

Regarding the explorative analysis of attenuation of the BOLD response to structure expectancy violations, we observed only sub-threshold attenuation in the IFS, occipitotemporal areas (including FG), IPS and thalamus. In contrast, several regions showed significant attenuation in for content expectancy violations, including posterior precuneus, OTC (including FG and PHG), pIPS, thalamus and the right inferior frontal junction (S1 Fig and S2 Table)

Discussion

Episodic memories allow us to mentally relive our personal past. Against what we might believe, they are not exact imprints of previous experiences, as they are prone to later modification, potentially triggered by prediction errors. In the present study, we investigated brain and behavioral responses to expectancy violations based on two properties which characterize episodic memories, namely

structure and content information. Participants encoded episodes in the form of toy stories. By presenting modified videos in a subsequent fMRI session, we aimed at violating the brain's expectations based on episode structure or content. While brain responses were largely as expected for content expectancy violations, only subtle activation was detected following structure expectancy violations, potentially because structure modifications were less salient. An explicit memory test revealed that the acceptance of original videos slightly decreased after expectancy violation. Further, we found an increased tendency to accept modified versions when the same had been presented during the fMRI scan. Together, findings provide first evidence that different types of mnemonic prediction errors are processed differently by the brain and may contribute to memory changes.

Influence of expectancy violation on post-fMRI memory performance

As expected, episodes that had been reactivated in a modified version during the fMRI experiment were generally remembered worse, as reflected by lower discrimination indexes (P_r). More specifically, modified videos were more likely to be perceived as original stories thereafter, as evidenced by significantly decreased rating scores. Similarly, others have reported that mnemonic prediction errors lead to an intrusion of new information into an established memory repertoire [20, 22]. In our study, participants were only descriptively less prone to accept original stories following expectation violation in the scanner. This finding is in line with previous reports that the inclusion of new information into an old memory is not necessarily accompanied by a deterioration of the original memory [73, 22]. Thus, it is likely that the prediction errors in the current study left the original memory traces largely intact but led to the encoding of an additional, alternative representation of the same episode. This finding contrasts with the results of a study by Kim et al. [27], who showed that prediction violations led to decreased recognition of original memory content. However, in their study, original memories were encoded by showing the to-be-predicted sequence only once, whereas participants in our study had experienced and actively encoded each story multiple times. Thus, it may be that memory traces were particularly stable in the current paradigm,

646 such that a greater number of repeated alternative evidences would be required to significantly
647 reduce memory performance for the initially encoded episodes. All memory changes we observed in
648 the present study were rather subtle, as median values indicated a change in the degree of certainty
649 rather than a complete transition from acceptance to rejection or vice versa. Participants also took
650 longer to correctly classify unmodified videos when they had previously experienced structure
651 expectancy violations. This tendency was also observed for modified videos. Longer response times
652 in cued-memory paradigms are interpreted as indicative of increased difficulty of retrieval due to
653 higher cognitive processing demands [59, 60]. Thus, it is likely that it became more difficult to
654 differentiate between alternative competing versions of episodes when versions diverging from the
655 original experience had been already encountered in the fMRI session.

656 Taken together, our results suggest that violations of both, structure and content expectations,
657 influenced subsequent memory performance for affected episodes. These findings are in line with
658 the idea that mnemonic prediction errors can trigger episodic memory modification [19]. We can,
659 however, not rule out that the repeated presentation of modified versions during the fMRI scan led
660 to a confusion of different versions, which could be an alternative explanation for our findings. On
661 the other hand, participants remembered correctly that novel videos had not been part of the
662 original episode repertoire even though they had repeatedly been presented during the fMRI
663 experiment as well. This finding suggests that our post-test results are at least not a product of
664 source confusion in its simplest form. Even if our results are to be attributed to confusion-effects,
665 this would not per se contradict the idea of memory modification. In fact, it is not yet clear exactly by
666 what mechanism memory modifications are brought about, and it has been suggested that they may
667 result, at least in part, from interference between old and new memory traces [74, 75, 22].

668 Another interesting behavioral finding was that participants generally had a strong tendency to
669 accept structurally modified versions as originals in the memory test. It is likely that structural
670 modifications were generally less salient than content modifications, although both types of
671 modifications resulted from a single change in the story (see S1 Table for examples). This result also

672 fits the fact that after the fMRI session, two-thirds of the participants self-reported noticing at least
673 one object swap, while less than a quarter of the participants had noticed a change regarding the
674 order of action steps. Recently, it was reported that memory performance after an expectancy
675 violation differed depending on whether changes were detected (and remembered) by participants
676 or not [18]. While undetected changes led to reduced memory performance, detected changes had
677 the opposite effect. Although our results suggest that structural changes were probably significantly
678 more difficult to detect, content expectancy violations led to an increase in the acceptance of
679 modified version, i.e. reduced memory performance, as well.

680 The current study provides evidence that different types of mnemonic prediction errors can modify
681 memory. Interestingly, others have found that prediction errors can also have to opposite effect, that
682 is improve subsequent memory [16, 17, 18]. Thus, it seems that the effect of expectation violation on
683 memory underlies a complex dynamic which might depend on the strength of prediction errors. It
684 has been suggested that strong expectancy violations lead to better memory by promoting learning
685 of new associations [17]. Additionally, no/weak prediction errors were also found to yield better
686 memory performance [76], and it has been suggested that expectancy violations need to overcome a
687 certain threshold to initiate memory updating [10]. Overall, it seems like the strength of prediction
688 errors modulates memory in a U-shaped fashion [76]. These findings are in accordance with the
689 theoretical framework by Ritvo et al. [77], which would predict a decrease in memory performance
690 following moderate memory reactivation. In conclusion, it is likely that mnemonic prediction errors
691 in the current study were medium-sized, i.e. large enough to decrease, but too small to increase,
692 memory performance. Within this range, prediction errors caused by structural deviance might have
693 been smaller than those elicited by content deviance, as our behavioral data suggests that the latter
694 were more salient. However, one must keep in mind that much of the previous research used one-
695 shot learning to study the effects of prediction errors on memory [17, 22]. In our paradigm, we
696 intentionally took a different approach, namely, repetition of the same divergent information that
697 often gives rise to learning processes in daily life. The aim of the current study was to characterize

brain and behavioral responses to different types of mnemonic prediction errors for this first time with a novel paradigm. In contrast, we did not specifically aim to investigate prediction error strengths, so this aspect is not discussed in detail here.

Neural responses to different episodic expectancy violations

Content and structure violation of expectation were expected to draw on distinct brain areas, but also to share some common violation-induced activation in mPFC, AON, and the hippocampal formation. Without correction for multiple comparisons, common activation was detected in right IFG, which reflects our previous finding that IFG increases for inconsistent or highly informative detail in observed actions [78, 79, 55]. After correction for multiple comparisons, no common violation-related activity could be observed. At this point, it is not possible to distinguish whether this was due to a large functional disparity between content and structural information in episodic memories or simply to the brain's overly subtle response to violations of structural expectancy.

On the one hand, we had expected that structure expectation violations lead to activation in brain regions involved in the temporal organization of episodes. We did not find significant activation patterns after correction for multiple comparisons. Without such correction, we detected activation in a hypothesized area, right PMd/SFS. The activity of a region comprising PMd/SFS is related to linking successive action steps [38, 39, 41, 42], and could contribute to updating the current event or action model with respect to each next segment [80, 40, 39, 42].

On the other hand, we found content expectancy violations to specifically recruit pIPS and OTC, including FG, which were hypothesized on the basis of their role for processing of object properties in the context of actions [81, 46, 82, 43, 55]. More specifically, pIPS encodes basic visual features of graspable objects [44, 83], reflecting the interaction with toy objects in our paradigm. Additionally, content expectancy violations elicited activity in the hippocampal formation (right PHG). In general, hippocampus and PHG are important in learning contexts [84, 85, 86, 87] and there is evidence that the PHG is involved in processing of competing memories [88]. Moreover, the hippocampal

formation is believed to generate mismatch signals when predictions do not fit perceptual inputs [69, 89, 20]. Because our post-fMRI memory test data imply that content changes were more salient than structural changes, one could speculate that the discrepancy between prediction and perception was strong enough to be reflected in the activation of the hippocampal formation for content expectancy violations. This would also fit with our finding that content expectancy violation triggered activity in the anterior insula, which is relevant for error detection [90] and salience processing [91].

Similar to our conjunction analysis, the ROI analysis in different IFG regions showed that the right BA 44 responded to both types of episodic violations. These results confirm the assumption that BA 45 is more responsive to content than to structural expectancy violations. We take this as a first indication that within IFG, different subregions might process different types of episodic prediction errors. However, our study cannot rule out the possibility that this effect is merely due to quantitative (rather than qualitative) differences between structural and content expectancy violation, as structure modifications might have elicited weaker prediction errors in our paradigm.

In order to interpret brain activity triggered by violating expectations considering retrieved episodes, it is informative to consider attenuation of this response when participants repeatedly encountered modified stories, as this gives insight into adaptation processes during model updating [24, 25]. In fact, such neuronal repetition suppression has been suggested to occur due to a decrease of prediction error strength [92], which would be in line with idea that memory modification occurs due to gradual updating of internal models triggered by prediction errors. We thus additionally explored possible attenuation effects concerning modified episodes. Brain activation was found to attenuate with repeated exposure to content modifications, indicating ongoing adaptation processes. Some regions that showed a decrease in brain activation were also identified in the general violation effect contrast, including PHG. Correspondingly, Schiffer et al. [24] found hippocampal attenuation with repeated presentation of divergent episodes. They concluded that this activity decrease is a sign of increasing validity of internal models due to updating. The attenuation effects we found for structure expectancy violations were quite subtle, so one could speculate that for this type of expectancy

violations, adaptation might be still at an early stage. It is also possible that, at least because of the substantial attenuation effect on content violation, some brain areas that responded to expectation violation could no longer be detected in the main effect. For example, attenuation was found bilaterally in FG, extending into PHG for content expectancy violations. While widespread activation in FG was also found in response to the violation itself, only a small portion of right PHG showed significant activation.

Limitations and implications for future research

One factor that might limit the generalizability of our findings is that, for practical reasons, only women participated in the study. However, because the processing of episodic memory in the brain seems to be broadly similar between women and men [93], we are confident that our findings are applicable to a more general population. In the future, our paradigm could be adapted to circumvent such practical limitations, for example, by applying virtual reality techniques so that encoding could be detached from the true physical appearance of participants' hands.

One potential limitation of the current paradigm is that in our modified episodes, the occurrence of objects and temporal structures were new. Thus, we cannot completely rule out that aside from prediction errors, novelty influenced our findings. It would be valuable to find a way to keep novelty constant and have participants make active predictions which are then either violated or not. Aside from that, it would be interesting to compare different prediction strengths which we will target in an upcoming study where we will introduce manipulations of different salience.

Another limitation of the current study is that we only found very subtle brain responses to structure-expectancy violations, which may be explained by their lower salience. As discussed above, structure modification might have caused weaker prediction errors than content modifications. The hypothesized (subthreshold) activation in the right PMd/SFS in response to structure-expectancy violations could be indicative for qualitative difference in the processing of the two types of expectancy violation. Nevertheless, an important question arises: could prediction errors be less informative for complex episode structures in general, or is this finding specific to certain, more

variable circumstances? Our results therefore represent exiting starting points for future research. We plan to investigate whether additional consolidation of encoded memories through active retrieval training [94] could enhance the brain response to structure violations. In addition, we will analyze whether structural modifications of higher salience (e.g., rearrangement of multiple steps, or rearrangements more significant to the episode) could produce a similar effect.

It is possible that by repeating the same modified versions, attenuation effects may have limited the detectability of brain regions that process certain types of mnemonic prediction errors. To clarify this, we will conduct a follow-up study in which we do both, repeating the same modified versions of an episode, but also including the presentations of constantly changing versions, similar to the paradigm of Schiffer et al. [24]. Finally, it will be interesting to reanalyze our dataset to link fMRI and behavioral data to analyze whether specific brain activations predict good or bad memory performance.

Conclusion

When recalling episodes, our memory can change, for instance due to prediction errors in the re-activation process. Our results suggest that structural and content prediction errors in episode retrieval differ in their neural processing. Future studies need to further validate whether the observed differences are indeed qualitative, and not just quantitative, in nature. This is also suggested by the fact that the tendency to falsely classify modified episodes as originally experienced ones increased comparably after experiencing repeated structural as well as content-related prediction errors. Thus, memory change seems to be able to occur as a result of different types of prediction errors. In summary, the present findings and our newly developed paradigm provide very good starting points for further research into the temporal dynamics, limits and conditions of episodic memory modification.

Acknowledgments

The authors thank Monika Mertens for her help during data collection. Furthermore, we thank Annika Garlichs, Christin Schwarzer, Helena Sydlik and Yuyi Xu for their assistance during the creation of stimulus material, data collection and data management. Lastly, we thank Jennifer Pomp, Lena Schliephake, Falko Mecklenbrauck and Nina Heins for advice regarding data analysis and the members of research unit FOR 2812 for valuable discussions. This work was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft) – project numbers 419037023, 419039274, 419037518. The funders had no role in study design, data collection, analysis and interpretation, decision to publish, or writing of the report.

References

1. Tulving E. Episodic Memory: From Mind to Brain. Annu Rev Psychol. 2002;53:1–25. doi: 0084-6570/02/0201-0001.

2. Nader K, Einarsson EÖ. Memory reconsolidation: An update. Ann N Y Acad Sci. 2010;1191:27–41. doi: 10.1111/j.1749-6632.2010.05443.x.

3. Nader K. Reconsolidation and the Dynamic Nature of Memory. Nov Mech Mem. 2015;7(10):1–20. doi: 10.1007/978-3-319-24364-1_1.

4. Lee JLC, Nader K, Schiller D. An Update on Memory Reconsolidation Updating. Trends Cogn Sci. 2017;21(7):531–45. doi: 10.1016/j.tics.2017.04.006.

5. Scully ID, Napper LE, Hupbach A. Does reactivation trigger episodic memory change? A meta-analysis. Neurobiol Learn Mem. 2017;142(Pt A):99–107. doi: 10.1016/j.nlm.2016.12.012.

6. Schacter DL. The cognitive neuroscience of false memories: Introduction. Cogn Neuropsychol. 1999;16(3/4/5):193–5. doi: 10.1080/026432999380753.

7. Straube B. An overview of the neuro-cognitive processes involved in the encoding, consolidation, and retrieval of true and false memories. Behav Brain Funct. 2012;8(35):1–10. doi: 10.1186/1744-9081-8-35.

8. Suddendorf T, Corballis MC. The evolution of foresight: What is mental time travel, and is it unique to humans? Behav Brain Sci. 2007;30(3):299–351. doi: 10.1017/S0140525X07001975.

9. Schacter DL, Guerin SA, St. Jacques PL. Memory distortion: An adaptive perspective. Trends Cogn Sci. 2011;15(10):467–74. doi: 10.1016/j.tics.2011.08.004.

10. Exton-McGuinness MTJ, Lee JLC, Reichelt AC. Updating memories-The role of prediction errors in memory reconsolidation. Behav Brain Res. 2015;278:375–84. doi: 10.1016/j.bbr.2014.10.011.

11. Rosen R. Anticipatory systems: Philosophical, mathematical, and methodological foundations. 2nd ed. New York, Springer; 2012. doi: 10.1007/978-1-4614-1269-4.

12. Friston K. A theory of cortical responses. Philos Trans R Soc B Biol Sci. 2005;360(1456):815–36. doi: 10.1098/rstb.2005.1622.

13. Friston K, Kiebel S. Predictive coding under the free-energy principle. Philos Trans R Soc B Biol Sci. 2009;364(1521):1211–21. doi: 10.1098/rstb.2008.0300.

14. Schubotz RI. Prediction and expectation. In: Toga AW, editor. Brain Mapping: An Encyclopedic

Reference. Cambridge: Academic Press: Elsevier; 2015. pp. 295–302. doi: 10.1016/B978-0-12-397025-1.00205-0.

15. Fernández RS, Boccia MM, Pedreira ME. The fate of memory: Reconsolidation and the case of Prediction Error. Neurosci Biobehav Rev. 2016;68:423–41. doi: 10.1016/j.neubiorev.2016.06.004.

16. Smith TA, Hasinski AE, Sederberg PB. The context repetition effect: Predicted events are remembered better, even when they don't happen. J Exp Psychol Gen. 2013;142(4):1298–308. doi: 10.1037/a0034067.

17. Greve A, Cooper E, Kaula A, Anderson MC, Henson R. Does prediction error drive one-shot declarative learning? J Mem Lang. 2017;94:149–65. doi: 10.1016/j.jml.2016.11.001.

18. Wahlheim CN, Zacks JM. Memory guides the processing of event changes for older and younger adults. J Exp Psychol Gen. 2019;148(1):30–50. doi: 10.1037/xge0000458.

19. Sinclair AH, Barense MD. Prediction Error and Memory Reactivation: How Incomplete Reminders Drive Reconsolidation. Trends Neurosci. 2019;42(10):727–39. doi: 10.1016/j.tins.2019.08.007.

20. Long NM, Lee H, Kuhl BA. Hippocampal Mismatch Signals Are Modulated by the Strength of Neural Predictions and Their Similarity to Outcomes. J Neurosci. 2016;36(50):12677–87. doi: 10.1523/JNEUROSCI.1850-16.2016.

21. Bein O, Duncan K, Davachi L. Mnemonic prediction errors bias hippocampal states. Nat Commun. 2020;11(1). doi: 10.1038/s41467-020-17287-1.

22. Sinclair AH, Barense MD. Surprise and destabilize: Prediction error influences episodic memory reconsolidation. Learn Mem. 2018;25(8):369–81. doi: 10.1101/lm.046912.117.

23. Scully ID, Hupbach A. Different reactivation procedures enable or prevent episodic memory updating. Hippocampus. 2020;30(8):806–14. doi: 10.1002/hipo.23159.

24. Schiffer AM, Ahlheim C, Wurm MF, Schubotz RI. Surprised at all the entropy: Hippocampal, caudate and midbrain contributions to learning from prediction errors. PLoS One. 2012;7(5):1–11. doi: 10.1371/journal.pone.0036445.

25. Schiffer AM, Ahlheim C, Ulrichs K, Schubotz RI. Neural Changes when Actions Change: Adaptation of Strong and Weak Expectations. Hum Brain Mapp. 2013;34(7):1713–27. doi: 10.1002/hbm.22023.

26. Wichert S, Wolf OT, Schwabe L. Updating of episodic memories depends on the strength of

870 new learning after memory reactivation. Behav Neurosci. 2013;127(3):331–8. doi:
871 10.1037/a0032028.

872 27. Kim G, Lewis-Peacock JA, Norman KA, Turk-Browne NB. Pruning of memories by context-
873 based prediction error. Proc Natl Acad Sci U S A. 2014;111(24):8997–9002. doi:
874 10.1073/pnas.1319438111.

875 28. Burianova H, McIntosh AR, Grady CL. A common functional brain network for
876 autobiographical, episodic, and semantic memory retrieval. Neuroimage. 2010;49(1):865–74.
877 doi: 10.1016/j.neuroimage.2009.08.066.

878 29. Euston DR, Gruber AJ, McNaughton BL. The Role of Medial Prefrontal Cortex in Memory and
879 Decision Making. Neuron. 2012;76(6):1057–70. doi: 10.1016/j.neuron.2012.12.002.

880 30. Peters GJ, David CN, Marcus MD, Smith DM. The medial prefrontal cortex is critical for
881 memory retrieval and resolving interference. Learn Mem. 2013;20(4):201–9. doi:
882 10.1101/lm.029249.112.

883 31. Maguire EA, Intraub H, Mullally SL. Scenes, Spaces, and Memory Traces: What Does the
884 Hippocampus Do? Neuroscientist. 2016;22(5):432–9. doi: 10.1177/1073858415600389.

885 32. Horner AJ, Doeller CF. Plasticity of hippocampal memories in humans. Curr Opin Neurobiol.
886 2017;43:102–9. doi: 10.1016/j.conb.2017.02.004.

887 33. Stachenfeld KL, Botvinick MM, Gershman SJ. The hippocampus as a predictive map. Nat
888 Neurosci. 2017;20(11):1643–53. doi: 10.1038/nn.4650.

889 34. Van Kesteren MTR, Ruiter DJ, Fernández G, Henson RN. How schema and novelty augment
890 memory formation. Trends Neurosci. 2012;35(4):211–9. doi: 10.1016/j.tins.2012.02.001.

891 35. Caspers S, Zilles K, Laird AR, Eickhoff SB. ALE meta-analysis of action observation and imitation
892 in the human brain. Neuroimage. 2010;50(3):1148–67. doi:
893 10.1016/j.neuroimage.2009.12.112.

894 36. Schubotz RI. Human premotor cortex: Beyond motor performance. MPI Series in Human
895 Cognitive and Brain Sciences, vol. 50. Max Planck Institute for Human Cognitive and Brain
896 Sciences, Leipzig. 2004.

897 37. Shima K, Tanji J. Neuronal Activity in the Supplementary and Presupplementary Motor Areas
898 for Temporal Organization of Multiple Movements. J Neurophysiol. 2000;84(4):2148–60. doi:
899 10.1152/jn.2000.84.4.2148.

900 38. Stadler W, Schubotz RI, von Cramon DY, Springer A, Graf M, Prinz W. Predicting and

Memorizing Observed Action: Differential Premotor Cortex Involvement. Hum Brain Mapp. 2011;32(5):677–87. doi: 10.1002/hbm.20949.

39. Schubotz RI, Korb FM, Schiffer AM, Stadler W, von Cramon DY. The fraction of an action is more than a movement: Neural signatures of event segmentation in fMRI. Neuroimage. 2012;61(4):1195–205. doi: 10.1016/j.neuroimage.2012.04.008.

40. Tamber-Rosenau BJ, Esterman M, Chiu YC, Yantis S. Cortical Mechanisms of Cognitive Control for Shifting Attention in Vision and Working Memory. J Cogn Neurosci. 2011;23(10):2905–19. doi: 10.1162/jocn.2011.21608.

41. Hrkać M, Wurm MF, Schubotz RI. Action Observers Implicitly Expect Actors to Act Goal-Coherently, Even if they Do Not: An fMRI Study. Hum Brain Mapp. 2014;35(5):2178–90. doi: 10.1002/hbm.22319.

42. Pomp J, Heins N, Trempler I, Kulvicius T, Tamosiunaite M, Mecklenbrauck F, et al. Touching events predict human action segmentation in brain and behavior. Neuroimage. 2021;243:118534. doi: 10.1016/j.neuroimage.2021.118534.

43. Lingnau A, Downing PE. The lateral occipitotemporal cortex in action. Trends Cogn Sci. 2015;19(5):268–77. doi: 10.1016/j.tics.2015.03.006.

44. Creem-Regehr SH. Sensory-motor and cognitive functions of the human posterior parietal cortex involved in manual actions. Neurobiol Learn Mem. 2009;91(2):166–71. doi: 10.1016/j.nlm.2008.10.004.

45. Schubotz RI, Wurm MF, Wittmann MK, von Cramon DY. Objects tell us what action we can expect: Dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. Front Psychol. 2014;5(JUN):1–15. doi: 10.3389/fpsyg.2014.00636.

46. Reber PJ, Gitelman DR, Parrish TB, Mesulam MM. Priming Effects in the Fusiform Gyrus: Changes in Neural Activity beyond the Second Presentation. Cereb Cortex. 2005;15(6):787–95. doi: 10.1093/cercor/bhh179.

47. Fiebach CJ, Schubotz RI. Dynamic anticipatory processing of hierarchical sequential events: A common role for Broca's area and ventral premotor cortex across domains? Cortex. 2006;42(4):499–502. doi: 10.1016/S0010-9452(08)70386-1.

48. Clerget E, Winderickx A, Fadiga L, Olivier E. Role of Broca's area in encoding sequential human actions: A virtual lesion study. Neuroreport. 2009;20(16):1496–9. doi: 10.1097/WNR.0b013e3283329be8.

932 49. Badre D, Wagner AD. Left ventrolateral prefrontal cortex and the cognitive control of
933 memory. Neuropsychologia. 2007;45(13):2883–901. doi:
934 10.1016/j.neuropsychologia.2007.06.015.

935 50. Heim S, Eickhoff SB, Amunts K. Different roles of cytoarchitectonic BA 44 and BA 45 in
936 phonological and semantic verbal fluency as revealed by dynamic causal modelling.
937 Neuroimage. 2009;48(3):616–24. doi: 10.1016/j.neuroimage.2009.06.044.

938 51. Oldfield RC. The assessment and analysis of handedness: The Edinburgh inventory.
939 Neuropsychologia. 1971;9(1):97–113. doi: 10.1016/0028-3932(71)90067-4.

940 52. Cheng S, Werning M, Suddendorf T. Dissociating memory traces and scenario construction in
941 mental time travel. Neurosci Biobehav Rev. 2016;60:82–9. doi:
942 10.1016/j.neubiorev.2015.11.011.

943 53. Franz EA, Ford S, Werner S. Brain and cognitive processes of imitation in bimanual situations:
944 Making inferences about mirror neuron systems. Brain Res. 2007;1145:138–49. doi:
945 10.1016/j.brainres.2007.01.136.

946 54. Wurm MF, Schubotz RI. The role of the temporoparietal junction (TPJ) in action observation:
947 Agent detection rather than visuospatial transformation. Neuroimage. 2018;165(2017):48–55.
948 doi: 10.1016/j.neuroimage.2017.09.064.

949 55. El-Sourani N, Trempler I, Wurm MF, Fink GR, Schubotz RI. Predictive Impact of Contextual
950 Objects during Action Observation: Evidence from Functional Magnetic Resonance Imaging. J
951 Cogn Neurosci. 2019;32:326–337. doi: 10.1162/jocn_a_01480.

952 56. R Core Team. R: A language and environment for statistical computing. R Foundation for
953 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. 2020.

954 57. Snodgrass JG, Corwin J. Pragmatics of measuring recognition memory: Applications to
955 dementia and amnesia. J Exp Psychol Gen. 1988;117(1):34–50. doi: 10.1037/0096-
956 3445.117.1.34.

957 58. Collins AM, Quillian MR. Retrieval Time from Semantic Memory. J Verbal Learning Verbal
958 Behav. 1969;8(2):240–7. doi: 10.1016/S0022-5371(69)80069-1.

959 59. Larsen SF, Plunkett K. Remembering experienced and reported events. Appl Cogn Psychol.
960 1987;1(1):15–26. doi: 10.1002/acp.2350010104.

961 60. Noppeney U, Price CJ. Retrieval of abstract semantics. Neuroimage. 2004;22(1):164–70. doi:
962 10.1016/j.neuroimage.2003.12.010.

963 61. Wobbrock JO, Findlater L, Gergle D, Higgins JJ. The Aligned Rank Transform for Nonparametric
964 Factorial Analyses Using Only ANOVA Procedures. CHI '11 CHI Conf Hum Factors Comput Syst.
965 2011;143–6. doi: 10.1145/1978942.1978963.

966 62. Bonferroni CE. Teoria Statistica Delle Classi e Calcolo Delle Probabilità, Vol. 8. Pubblicazioni del
967 R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 3–62. 1936.

968 63. Friston KJ, Holmes AP, Worsley KJ, Poline J -P, Frith CD, Frackowiak RSJ. Statistical Parametric
969 Maps in Functional Imaging: A General Linear Approach. Hum Brain Mapp. 1994;2(4):189–
970 210. doi: 10.1002/hbm.460020402.

971 64. Worsley KJ, Friston KJ. Analysis of fMRI Time-Series Revesited - Again. Neuroimage.
972 1995;2(3):173–81. doi: 10.1006/nimg.1995.1023.

973 65. Nichols T, Brett M, Andersson J, Wager T, Poline JB. Valid conjunction inference with the
974 minimum statistic. Neuroimage. 2005;25(3):653–60. doi: 10.1016/j.neuroimage.2004.12.005.

975 66. Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach
976 for determining significant signals in images of cerebral activation. Hum Brain Mapp.
977 1996;4(1):58–73. doi: 10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O.

978 67. Farrer C, Frith CD. Experiencing Oneself vs Another Person as Being the Cause of an Action:
979 The Neural Correlates of the Experience of Agency. Neuroimage. 2002;15(3):596–603. doi:
980 10.1006/nimg.2001.1009.

981 68. Mechelli A, Sartori G, Orlandi P, Price CJ. Semantic relevance explains category effects in
982 medial fusiform gyri. Neuroimage. 2006;30(3):992–1002. doi:
983 10.1016/j.neuroimage.2005.10.017.

984 69. Kumaran D, Maguire EA. Match–Mismatch Processes Underlie Human Hippocampal
985 Responses to Associative Novelty. J Neurosci. 2007;27(32):8517–24. doi:
986 10.1523/JNEUROSCI.1677-07.2007.

987 70. Gorgolewski KJ, Varoquaux G, Rivera G, Schwarz Y, Ghosh SS, Maumet C, et al.
988 NeuroVault.Org: A web-based repository for collecting and sharing unthresholded statistical
989 maps of the human brain. Front Neuroinform. 2015;9(8):1–9. doi: 10.3389/fninf.2015.00008.

990 71. Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and
991 cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage. 2003;19(3):1233–
992 9. doi: 10.1016/S1053-8119(03)00169-1.

993 72. Brett M, Anton JL, Valabregue R, Poline J-B. Region of interest analysis using an SPM toolbox.

994 Present 8th Int Conferance Funct Mapp Hum Brain, June 2-6, 2002, Sendai, Japan.

995 73. St. Jacques PL, Olm C, Schacter DL. Neural mechanisms of reactivation-Induced updating that
996 enhance and distort memory. Proc Natl Acad Sci U S A. 2013;110(49):19671–8. doi:
997 10.1073/pnas.1319630110.

998 74. Sederberg PB, Gershman SJ, Polyn SM, Norman KA. Human memory reconsolidation can be
999 explained using the temporal context model. Psychon Bull Rev. 2011;18(3):455–68. doi:
1000 10.3758/s13423-011-0086-9.

1001 75. Klingmüller A, Caplan JB, Sommer T. Intrusions in episodic memory: reconsolidation or
1002 interference? Learn Mem. 2017;24(5):216–24. doi: 10.1101/lm.045047.117.

1003 76. Greve A, Cooper E, Tibon R, Henson RN. Knowledge Is Power: Prior Knowledge Aids Memory
1004 for Both Congruent and Incongruent Events, but in Different Ways. J Exp Psychol Gen.
1005 2019;148(2):325–41. doi: 10.1037/xge0000498.

1006 77. Ritvo VJH, Turk-Browne NB, Norman KA. Nonmonotonic Plasticity: How Memory Retrieval
1007 Drives Learning. Trends Cogn Sci. 2019;23(9):726–42. doi: 10.1016/j.tics.2019.06.007.

1008 78. Wurm MF, Schubotz RI. Squeezing lemons in the bathroom: Contextual information
1009 modulates action recognition. Neuroimage. 2012;59(2):1551–9. doi:
1010 10.1016/j.neuroimage.2011.08.038.

1011 79. Hrkać M, Wurm MF, Kühn AB, Schubotz RI. Objects Mediate Goal Integration in Ventrolateral
1012 Prefrontal Cortex during Action Observation. PLoS One. 2015;10(7):1–12. doi:
1013 10.1371/journal.pone.0134316.

1014 80. Kurby CA, Zacks JM. Segmentation in the perception and memory of events. Trends Cogn Sci.
1015 2008;12(2):72–9. doi: 10.1016/j.tics.2007.11.004.

1016 81. Grill-Spector K, Kourtzi Z, Kanwisher N. The lateral occipital complex and its role in object
1017 recognition. Vision Res. 2001;41(10–11):1409–22. doi: 10.1016/S0042-6989(01)00073-6.

1018 82. Wiggett AJ, Downing PE. Representation of Action in Occipito-temporal Cortex. J Cogn
1019 Neurosci. 2011;23(7):1765–80. doi: 10.1162/jocn.2010.21552.

1020 83. Mruczek REB, von Loga IS, Kastner S. The representation of tool and non-tool object
1021 information in the human intraparietal sulcus. J Neurophysiol. 2013;109(12):2883–96. doi:
1022 10.1152/jn.00658.2012.

1023 84. Aguirre GK, Detre JA, Alsop DC, D'Esposito M. The Parahippocampus Subserves Topographical
1024 Learning in Man. Cereb Cortex. 1996;6(6):823–9. doi: 10.1093/cercor/6.6.823.

1025 85. O'Reilly RC, Rudy JW. Computational Principles of Learning in the Neocortex and
1026 Hippocampus. Hippocampus. 2000;10(4):389–97. doi: 10.1002/1098-
1027 1063(2000)10:4<389::AID-HIPO5>3.0.CO;2-P.

1028 86. Davachi L, Wagner AD. Hippocampal Contributions to Episodic Encoding: Insights From
1029 Relational and Item-Based Learning. J Neurophysiol. 2002;88(2):982–90. doi:
1030 10.1152/jn.2002.88.2.982.

1031 87. Köhler S, Crane J, Milner B. Differential Contributions of the Parahippocampal Place Area and
1032 the Anterior Hippocampus to Human Memory for Scenes. Hippocampus. 2002;12(6):718–23.
1033 doi: 10.1002/hipo.10077.

1034 88. Kuhl BA, Bainbridge WA, Chun MM. Neural Reactivation Reveals Mechanisms for Updating
1035 Memory. J Neurosci. 2012;32(10):3453–61. doi: 10.1523/JNEUROSCI.5846-11.2012.

1036 89. Duncan K, Curtis C, Davachi L. Distinct Memory Signatures in the Hippocampus: Intentional
1037 States Distinguish Match and Mismatch Enhancement Signals. J Neurosci. 2009;29(1):131–9.
1038 doi: 10.1523/JNEUROSCI.2998-08.2009.

1039 90. Klein TA, Ullsperger M, Danielmeier C. Error awareness and the insula: links to neurological
1040 and psychiatric diseases. Front Hum Neurosci. 2013;7(JAN):1–14. doi:
1041 10.3389/fnhum.2013.00014.

1042 91. Menon V, Uddin LQ. Saliency, switching, attention and control: a network model of insula
1043 function. Brain Struct Funct. 2010;214(5–6):655–67. doi: 10.1007/s00429-010-0262-0.

1044 92. Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T. Neural repetition
1045 suppression reflects fulfilled perceptual expectations. Nat Neurosci. 2008;11(9):1004–6. doi:
1046 10.1038/nn.2163.

1047 93. Nyberg L, Habib R, Herlitz A. Brain activation during episodic memory retrieval: Sex
1048 differences. Acta Psychol (Amst). 2000;105(2–3):181–94. doi: 10.1016/s0001-6918(00)00060-
1049 3.

1050 94. Antony JW, Ferreira CS, Norman KA, Wimber M. Retrieval as a Fast Route to Memory
1051 Consolidation. Trends Cogn Sci. 2017;21(8):573–6. doi: 10.1016/j.tics.2017.05.001.

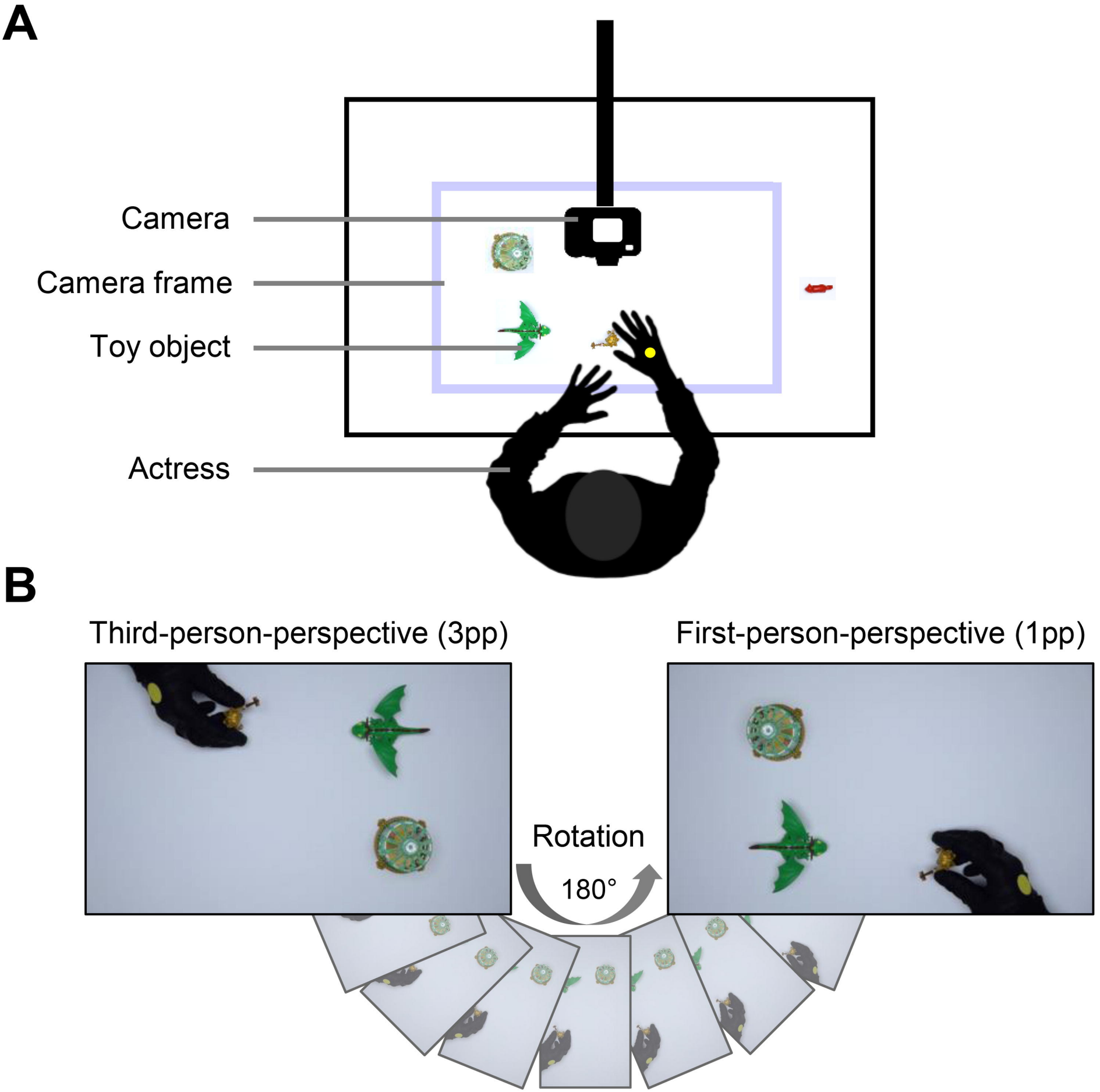
Supporting information

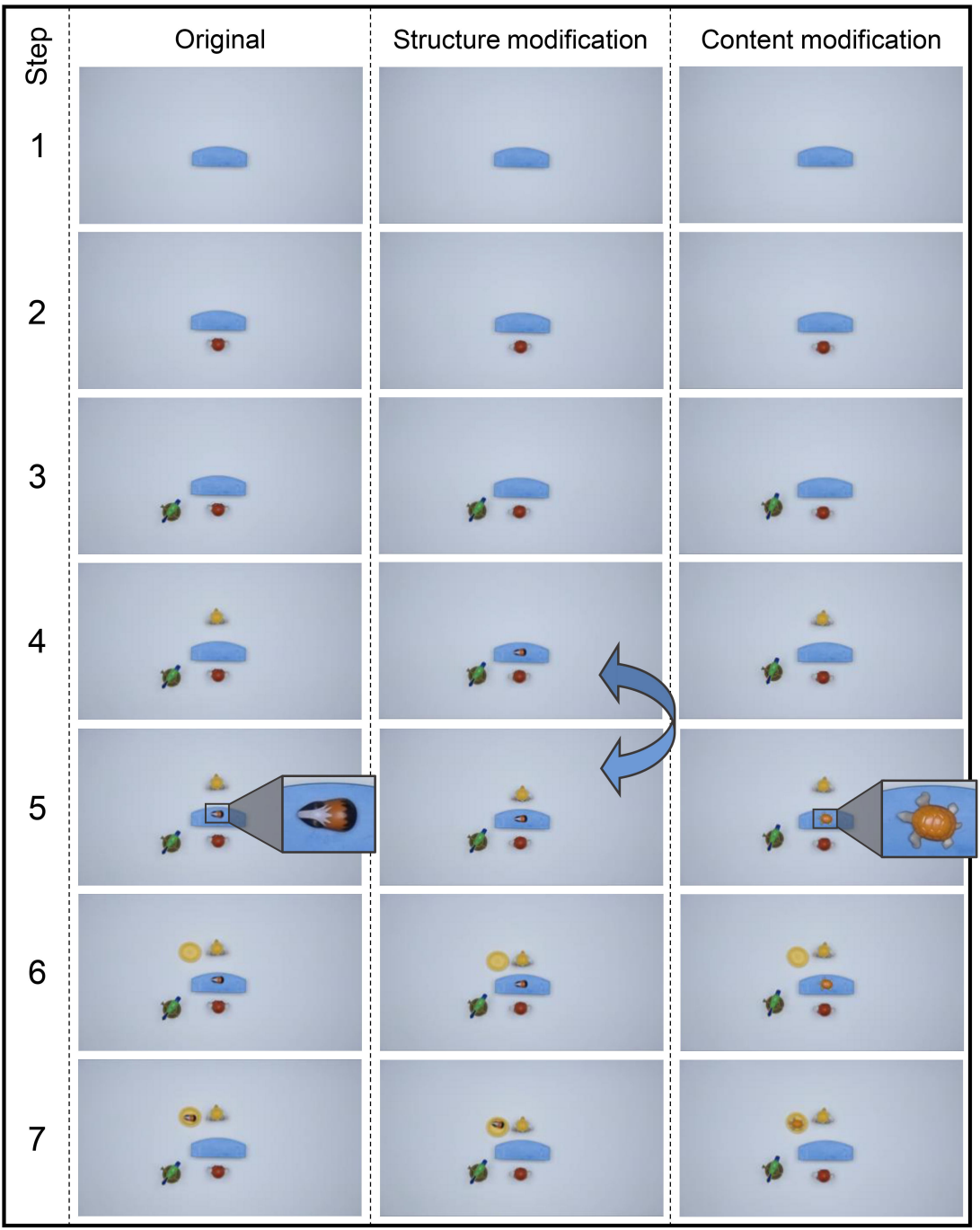
S1 Fig. Whole brain activation for attenuation effects. (A) Uncorrected *t*-map ($p < 0.001$) for structure violation attenuation effect. (B) FDR-corrected *t*-map ($p < 0.05$) for content expectancy violation attenuation effect. IFJ: inferior frontal junction; IFS: inferior frontal sulcus; IPS: intraparietal sulcus; FG: fusiform gyrus; PCUN: precuneus; THA: thalamus; PHG: parahippocampal gyrus.

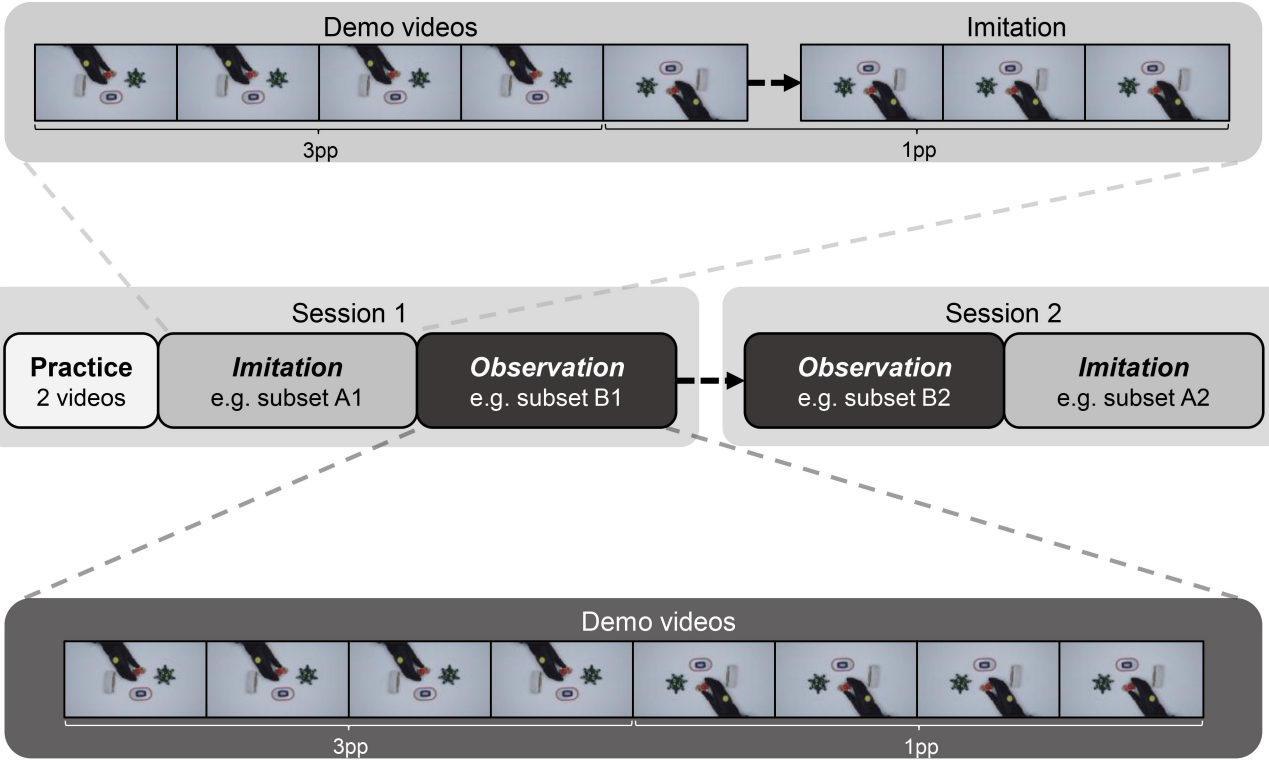
S1 Table. Three examples of the 24 toy stories used in the study.

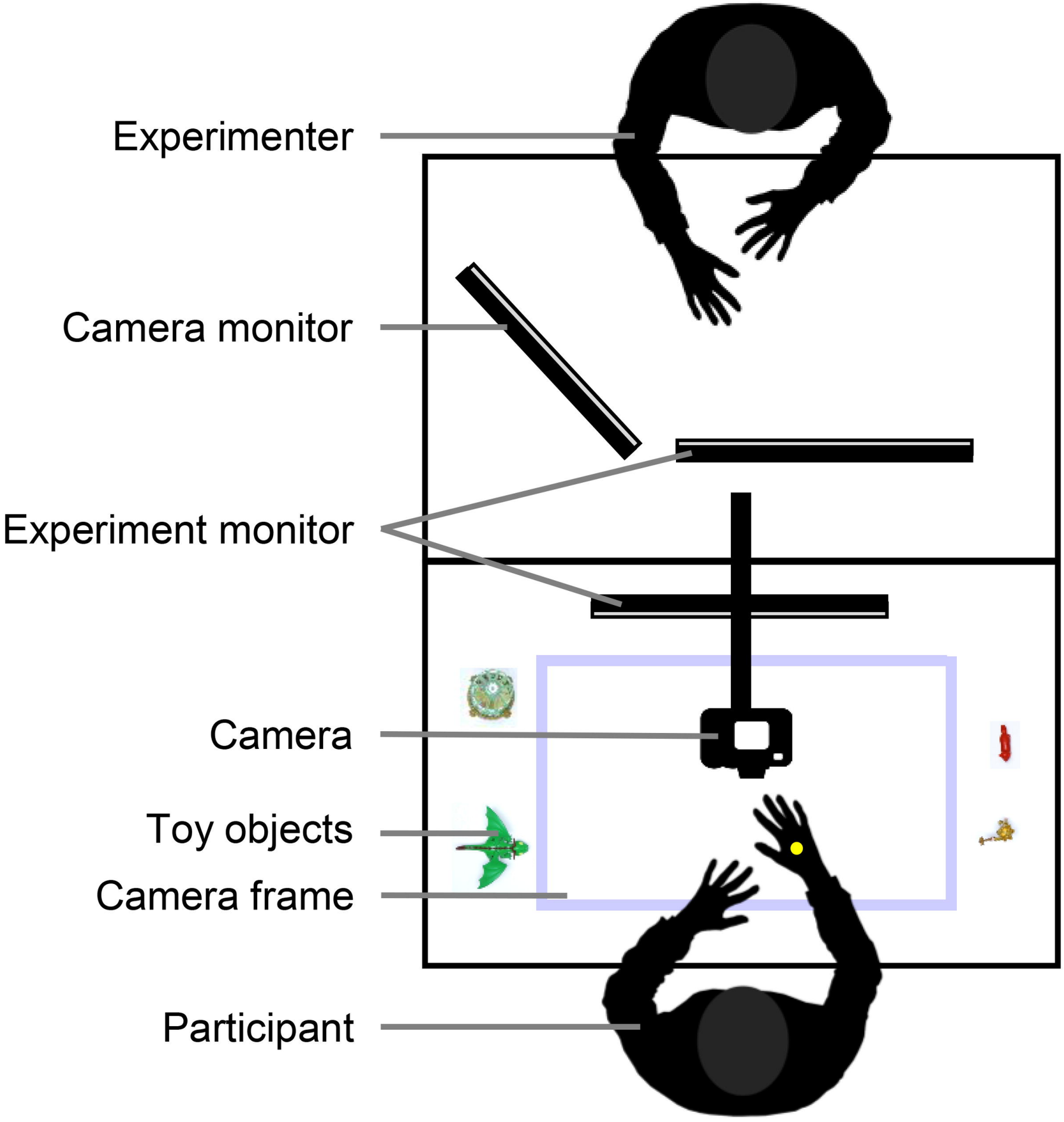
S2 Table. Peak activations from second-level whole-brain analyses of attenuation effects at the time of violation.

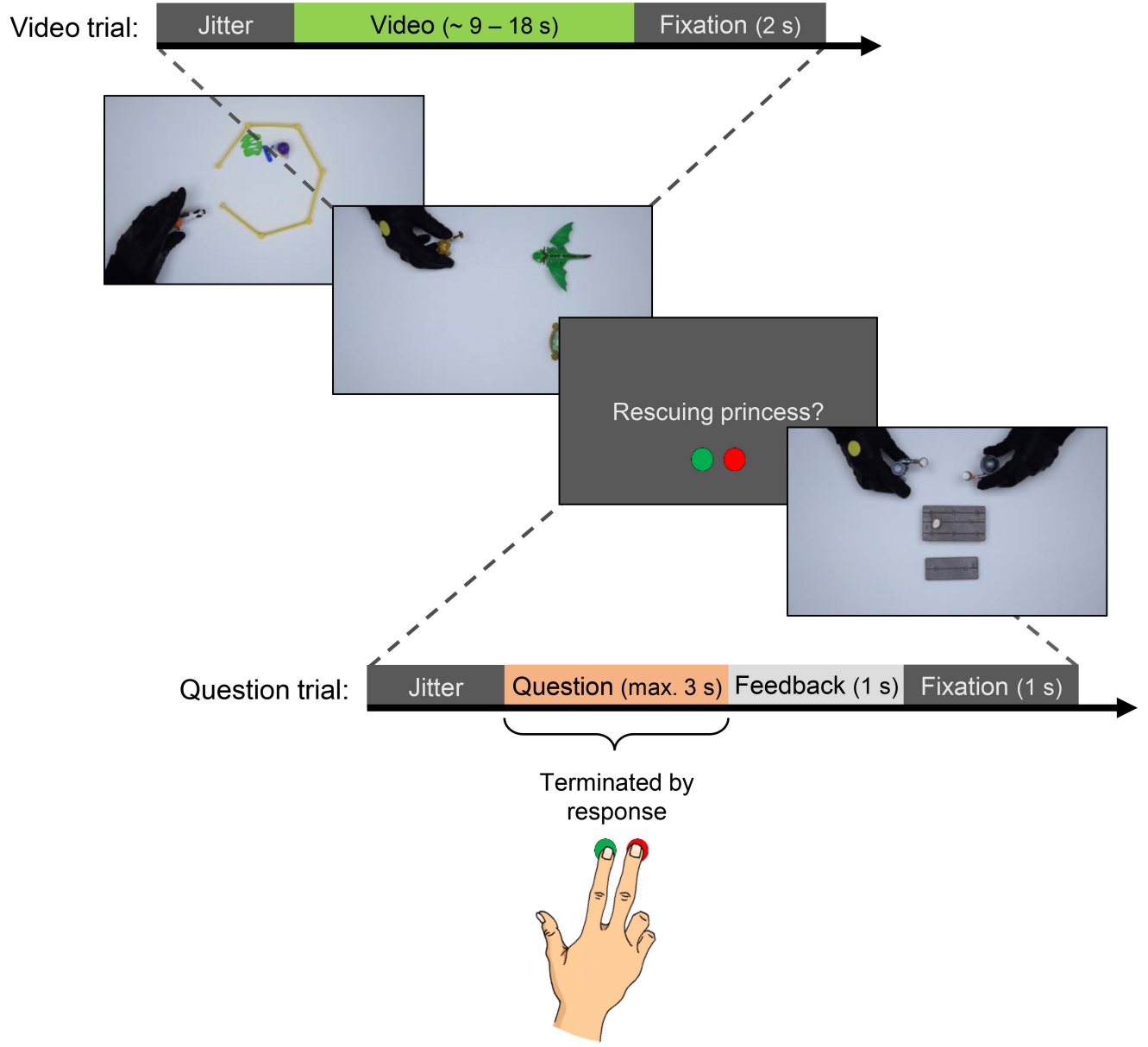
S3 Table. One-sample *t*-tests (one-tailed) for mean beta values compared to zero.

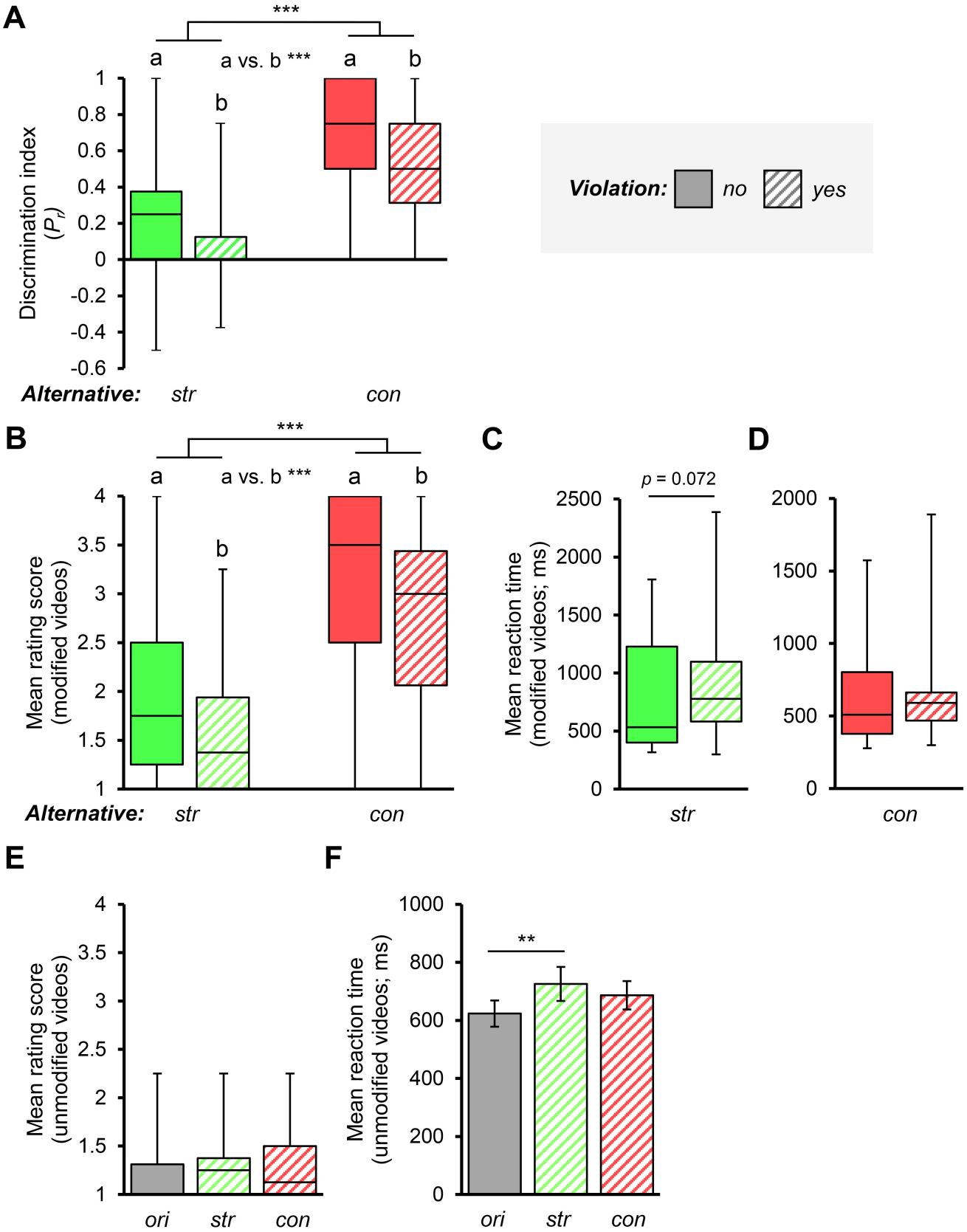

A
Camera
Camera frame
Toy object
Actress
B
Third-person-perspective (3pp)
First-person-perspective (1pp)
Rotation
180°

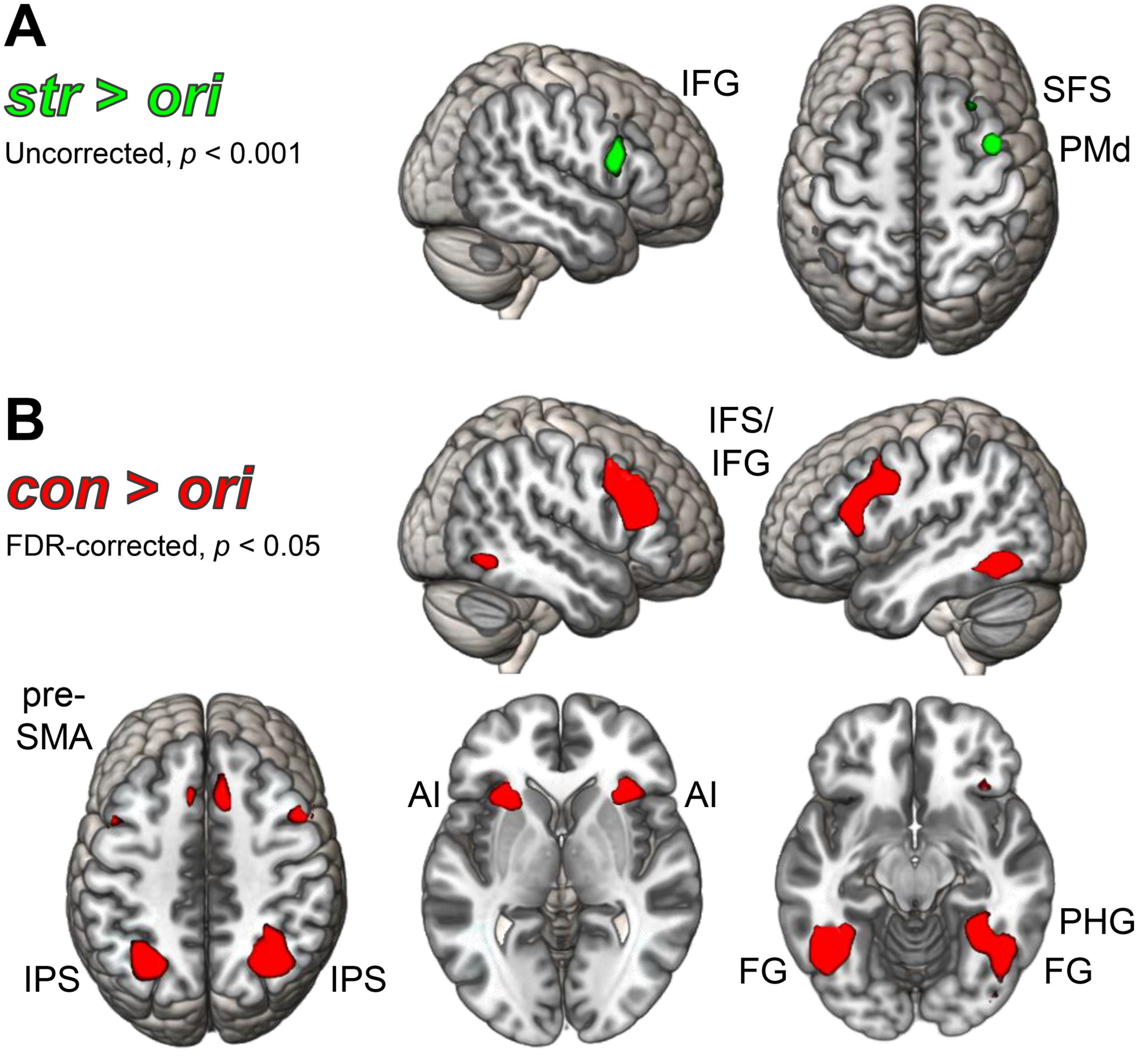

Step
Original
Structure modification
Content modification
1
2
3
4
5
6
7

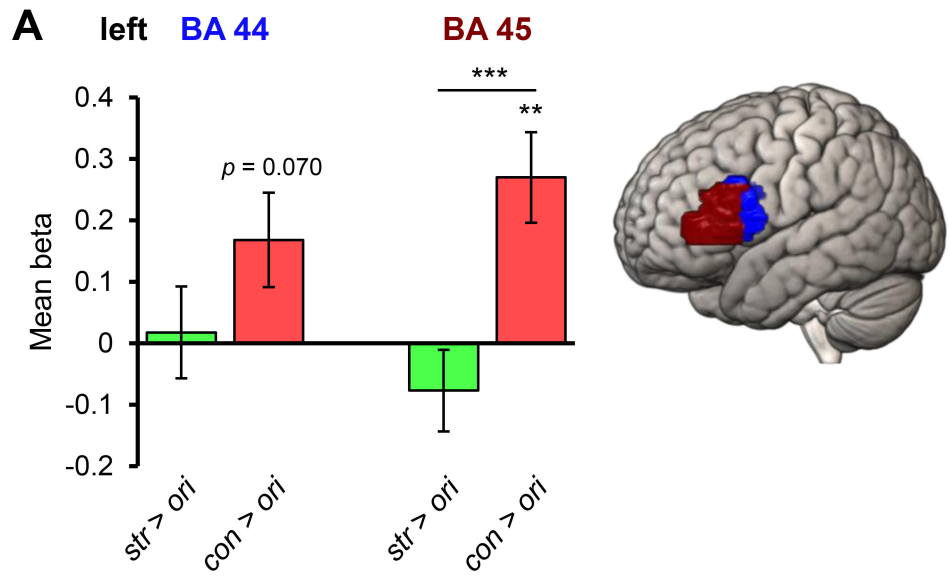

Demo videos
Imitation
3pp
1pp
Session 1
Practice
2 videos
Imitation
e.g. subset A1
Observation
e.g. subset B1
Session 2
Observation
e.g. subset B2
Imitation
e.g. subset A2
Demo videos
3pp
1pp

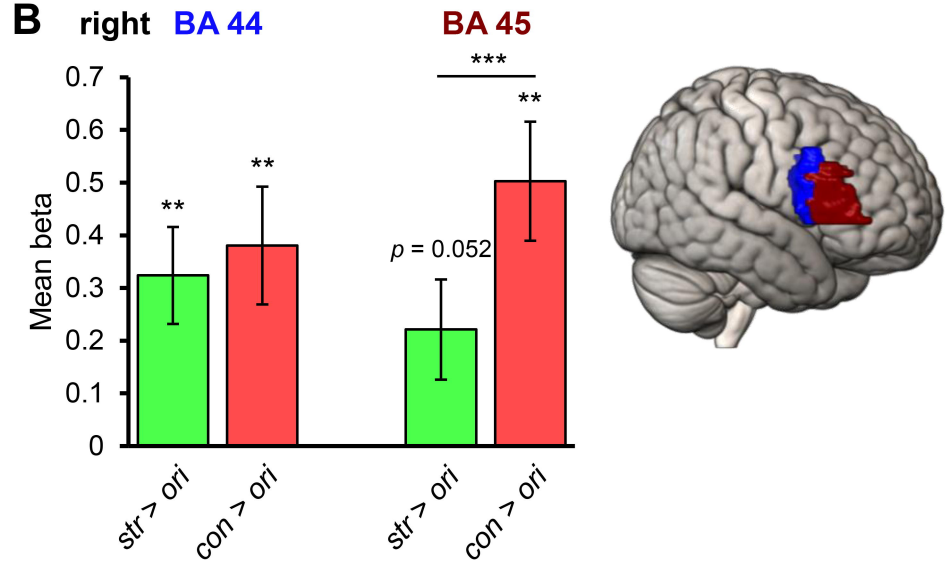

Experimenter
Camera monitor
Experiment monitor
Camera
Toy objects
Camera frame
Participant


Video trial:
Jitter
Video (~ 9 – 18 s)
Fixation (2 s)
Rescuing princess?
Question trial:
Jitter
Question (max. 3 s)
Feedback (1 s)
Fixation (1 s)
Terminated by response


A
***
a
a vs. b ***
a
b
b
Discrimination index (P_r)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
Alternative: str con
Violation: no yes
B
***
a
a vs. b ***
a
b
b
Mean rating score (modified videos)
4
3.5
3
2.5
2
1.5
1
Alternative: str con
C
p = 0.072
Mean reaction time (modified videos; ms)
2500
2000
1500
1000
500
0
str
D
2000
1500
1000
500
0
con
E
Mean rating score (unmodified videos)
4
3.5
3
2.5
2
1.5
1
ori str con
F
**
Mean reaction time (unmodified videos; ms)
1000
800
600
400
200
0
ori str con


A
str > ori
Uncorrected, p < 0.001
IFG
SFS
PMd
B
con > ori
FDR-corrected, p < 0.05
IFS/
IFG
pre-
SMA
AI
AI
PHG
IPS
IPS
FG
FG


A
left
BA 44
BA 45
***
**
p = 0.070
Mean beta
0.4
0.3
0.2
0.1
0
-0.1
-0.2
str > ori
con > ori
str > ori
con > ori


B
right
BA 44
BA 45
***
**
**
**
p = 0.052
Mean beta
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
str > ori
con > ori
str > ori
con > ori