Action Segmentation in the Brain: The Role of Object–Action Associations

Jennifer Pomp[1], Annika Garlichs[1], Tomas Kulvicius[2,3], Minija Tamosiunaite[2,4], Moritz F. Wurm[5], Anoushiravan Zahedi[1], Florentin Wörgötter[2], and Ricarda I. Schubotz[1]

Abstract

■ Motion information has been argued to be central to the subjective segmentation of observed actions. Concerning object-directed actions, object-associated action information might as well inform efficient action segmentation and prediction. The present study compared the segmentation and neural processing of object manipulations and equivalent dough ball manipulations to elucidate the effect of object–action associations. Behavioral data corroborated that objective relational changes in the form of (un-)touchings of objects, hand, and ground represent meaningful anchor points in subjective action segmentation rendering them objective marks of meaningful event boundaries. As expected, segmentation behavior became even more systematic for the weakly informative dough. fMRI data were modeled by critical subjective, and computer-vision-derived objective event boundaries. Whole-brain as well as planned ROI analyses showed that object information had significant effects on how the brain processes these boundaries. This was especially pronounced at untouchings, that is, events that announced the beginning of the upcoming action and might be the point where competing predictions are aligned with perceptual input to update the current action model. As expected, weak object–action associations at untouching events were accompanied by increased biological motion processing, whereas strong object–action associations came with an increased contextual associative information processing, as indicated by increased parahippocampal activity. Interestingly, anterior inferior parietal lobule activity increased for weak object–action associations at untouching events, presumably because of an unrestricted number of candidate actions for dough manipulation. Our findings offer new insights into the significance of objects for the segmentation of action. ■

INTRODUCTION

Everyday actions consist of smoothly concatenated action steps. The segmental structure of actions is reflected in the way that we teach, learn, and execute actions ourselves (Braun, Mehring, & Wolpert, 2010), and also in how we perceive actions performed by others (Newtson, Hairfield, Bloomingdale, & Cutino, 1987). Behavioral studies in children (Buchsbaum, Griffiths, Plunkett, Gopnik, & Baldwin, 2015; Baldwin, Baird, Saylor, & Clark, 2001) and adults (Hard, Recchia, & Tversky, 2011; Newtson & Engquist, 1976) show that action segmentation arises spontaneously (see also Zacks, Speer, Swallow, Braver, & Reynolds, 2007) and helps us process and remember dynamic events efficiently (Kurby & Zacks, 2018; Zacks & Swallow, 2007).

To measure subjective segmentation behavior, researchers ask participants to indicate when they perceive event boundaries, that is, those points in time when one action segment ends and the next begins (Newtson, 1973). This procedure has been shown to yield intra-individually consistent action segments (for a review: Sargent, Zacks, & Bailey, 2015), but the question remains which stimulus properties drive the segmentation behavior. A number of studies have specifically addressed the role of motion as a cue for updating action models at event boundaries (Zacks, Kumar, Abrams, & Mehta, 2009; Hard, Tversky, & Lang, 2006; Newtson, Engquist, & Bois, 1977). Typical measures to quantify motion include binary time interval coding for separate movement types (Hard et al., 2006) or motion tracking through speed and acceleration of hands and head (Zacks et al., 2009). Correspondingly, the activity of the motion-selective area MT was found to increase during the perception of event boundaries in actions (Schubotz, Korb, Schiffer, Stadler, & von Cramon, 2012; Speer, Swallow, & Zacks, 2003; Zacks et al., 2001), pointing to change in motion as an efficient cue that announces event boundaries and triggers updating processes in frontal networks.

However, having a life-long experience with manipulable objects, the movements one expects when observing object-directed actions certainly also depend on the involved object and might influence spatial attention and processing. Objects are an important source of information that individuals use to understand an observed action because we have learned how to act with or on an object and thereby build object–action associations (Borghi,

[1] University of Münster, [2] University of Göttingen, [3] University Medical Center Göttingen, [4] Vytautas Magnus University, Kaunas, Lithuania, [5] University of Trento

https://doi.org/10.1162/jocn_a_02210

2021; Zhao, 2019). In a former study (Schubotz, Wurm, Wittmann, & von Cramon, 2014), we built on the idea that objects are reminiscent of actions often performed with them. For instance, the combination of a knife and an apple remind us of peeling the apple or cutting it. Findings confirmed that the BOLD response in action-related inferior parietal and posterior temporal areas varied with the number of object-implicated actions. This impact of objects has been shown to influence the processing of observed action, even when these objects are not actually used (El-Sourani, Trempler, Wurm, Fink, & Schubotz, 2019; El-Sourani, Wurm, Trempler, Fink, & Schubotz, 2018; Hrkać, Wurm, Kühn, & Schubotz, 2015). However, because action segmentation appears to be highly dependent on movement-related information and may develop in early infant action observation when functional or semantic knowledge about objects is still rudimentary, object information may not be essential for action segmentation. One may ask how action structures are processed before having experience-based knowledge of object-associated actions, for instance, when encountering actions with novel objects, which is common in young infancy (cf. Hunnius & Bekkering, 2010).

In the present study, we aimed to investigate the effect of object–action knowledge on action segmentation and underlying brain processes. We built on a previous study (Pomp et al., 2021), which examined action segmentation in everyday object manipulations. To this end, we recreated the movies of the object manipulation actions, but this time using formed pieces of play dough as objects. This replacement of common objects by formed dough minimized object–action associations, that is, individuals did not strongly associate the formed dough with specific actions (except for kneading, if at all). The actions themselves were kept as similar as possible to the actions performed on the everyday objects to balance the movement patterns between the current and the previous study. After a passive action observation session in the MRI scanner, individual behavioral action segmentations of these actions were gained using the unit marking procedure (Newtson, 1973). Although subjective reports are important and can be informative, we do not necessarily have explicit access to all event boundaries that our brain registers and exploits to make sense of the world. Moreover, subjective reports may be focused on behaviorally relevant events and have been shown to be highly dependent on the exact task, for example, with regard to the instruction of detecting "meaningful" boundaries or selecting a specific "fine" or "coarse" grain of the segmentation (Zacks et al., 2007). Manual action segmentation is therefore a possible, but not necessarily a reliable, approximation for the way in which the brain segments events.

An exciting complement to research into action segmentation is therefore a more objectifiable stimulus-based approach to action segmentation (Pomp et al., 2021). We extracted objective stimulus characteristics based on the notion of *semantic event chains* (Wörgötter et al., 2013; Aksoy et al., 2011). In an object-directed action, this approach describes actions as a sequence of relational changes in the form of *touchings* (T) and *untouchings* (U) of objects, hands, and ground (TUs, hereafter). For instance, when a hand grasps an object, motion velocity usually reaches zero whereas the hand and object touch. In case of a subsequent object transport, the object then untouches the ground, and velocity increases again until it decreases before the object touches its destination. In case of a subsequent object manipulation, for example, turning, velocity increases while the object is turned and decreases before the hand untouches the object after manipulation. Thus, the binary coding of touching relations (touch, untouch) between each pair of objects, hands, and ground in an action scene can be used to describe the course of action without the need to analyze velocity and trajectory patterns and was used in the current study to model brain activity. Note that the above-explained underlying computer-vision algorithm that we used is model-free and stimulus-driven (Aksoy et al., 2011). Therefore, it does not require functional or semantic knowledge about objects (or hands or ground), which might imitate the simple model of early infant action observation. The use of objective event boundaries, which can be extracted directly from the stimulus material, offers promising opportunities to understand the neural processes underlying ongoing action segmentation.

Using the touching–untouching approach in the present study, we examined the impact of object–action knowledge on action segmentation and underlying brain processes. If object–action associations play a role in action segmentation, we expected significant differences between our previous study on object manipulation and our current study on dough manipulation in terms of segmentation behavior and time-point-specific brain activity. To this end, we compared the neural processing of object and dough videos at different types of event boundaries, including group-consistent behavioral segmentations (unit marks, Ms hereafter) and objective TU events as relevant points in time. We refer to the boundaries assessed by the participants as unit marks (conceptually based on the unit marking procedure) and not as event boundaries, as we assume that they are only one type of event boundary of interest. For object manipulations, TU events were found to be meaningful anchor points for action segmentation behavior (Pomp et al., 2021), and we expected TUs to gain even more importance when object–action associations are weak. Specifically, we expected participants' action segmentation behavior to be even more dependent on TU events, that is, temporally less spread and closer to TUs. We refer to the temporal relation between participant-judged event boundaries and TU events as being *systematic* if their occurrence coincided more than randomly often, which we examined on single subject and group level. For object manipulations, this systematic relation had been shown (Pomp et al., 2021) and we expected that this systematicity in behavior would increase for dough manipulations. Thus, we expected that participant-judged event boundaries would reliably coincide with TU events, but not necessarily vice versa.

With regard to brain activity, we examined at which of the critical time points T, U, and M activity would differ between object and dough manipulation in one of three ROIs derived from previous findings: the anterior inferior parietal lobule (aIPL), the parahippocampal cortex (PHC), and the biological motion-sensitive area (BMA, hereafter) in the lateral temporo-occipital cortex. Concerning the first ROI, as mentioned above, Schubotz and colleagues (2014) showed inferior parietal regions' activity to vary with the number of object-implicated actions at the mere sight of the object, independent of its usage. This activity was located in aIPL, and therefore, we expected increased aIPL activation for actions performed on objects versus dough pieces. The aIPL, as part of the ventrodorsal visual processing route (Binkofski & Buxbaum, 2013), is engaged in the representation of pragmatic object properties (Bosch et al., 2023) and hand–object interactions (Pelgrims, Olivier, & Andres, 2011; Vingerhoets, 2008) when we perform, plan, or observe object manipulations. Correspondingly, aIPL is known to be an important anatomic substrate underlying ideomotor apraxia (O'Neal et al., 2021), and it has been suggested to resolve competition between possible actions (Watson & Buxbaum, 2015). Concerning the second ROI, as for aIPL, we hypothesized an increased PHC activation for actions performed on objects versus actions performed on dough. The PHC is generally involved in processing contextual associations (Li, Lu, & Zhong, 2016; Aminoff, Kveraga, & Bar, 2013; Bar, Aminoff, & Schacter, 2008), which is the principal element underlying many cognitive processes, including spatial processing in scenes and episodic memory. In previous studies, we found PHC activity to specifically increase at action boundaries, possibly signaling the memory-driven updating of expectations of the next action associated with the object (Pomp et al., 2021; Schubotz et al., 2012). We here expected that familiar objects would trigger more contextual action associations than formed pieces of play dough accompanied by higher PHC activity. Finally, regarding the third ROI, we expected motion information to gain importance for play dough compared with object videos, which we hypothesized to detect in BMA. We reasoned that detailed motion analysis might be less critical when objects provide clues about which actions are about to be performed, whereas detailed motion analysis might be especially important, when pieces of dough are manipulated, to constrain the observer's predictions efficiently.

METHODS

For the current study, we used the experimental design of a previous study (Pomp et al., 2021), employed new videos, and tested a new group of participants comparable in size. The current study was kept as similar as possible to the previous one to allow direct statistical comparisons. This includes that the participants were recruited through the same channels, the study took place at the same institute, participants were scanned in the same MRI scanner, behavioral sessions were in the same laboratory rooms, and all sessions followed the exact same experimental protocols with similar equipment and materials (except for the stimulus videos). The results of the previously published study will not be shown here again, but only new analyses relating to statistical between-studies comparisons. Regarding brain activity contrasts, only interaction effects are reported, to make the results resistant to any differences between groups. With regard to the interpretation of direct comparisons between the two studies, we statistically compared the sample characteristics to rule out that differences between the samples could account for differences observed between the video types. We used the demographic details on age, sex, and profession, as well as participants' answers to the short surveys about their physical and mental condition, and experimental task features that concluded each of the separate sessions (for details on the survey, see section Experimental Procedure) to predict the participant's affiliation to either study. In separate analyses for the continuous, ordinal, and binary data types, no significant differences between groups were found using Bayesian modeling. To be precise, these analyses yielded support for the null hypothesis in all but one case, where the evidence ratio was inconclusive—giving neither evidence for the null nor for the alternative hypothesis. We uploaded the corresponding data, the R script of the analyses, and the results to the OSF repository (DOI 10.17605/OSF.IO/MGQSF).

Participants

Thirty-three right-handed participants (M_{age} = 23.03 years, *SD* = 3.06, age range = 18–29 years, 28 women, 5 men) took part in this study. This sample size was based on previous work (Pomp et al., 2021) that showed robust results with a similar sample size. All participants reported intact color perception, and none of the participants reported any history of neurological or psychiatric disorders. The participants had not taken part in related precursor studies. In the course of the experiment, it became apparent that one participant had not understood the instructions of the behavioral segmentation task correctly; hence, this participant's data set was excluded from the behavioral model construction but was included in the fMRI data set (as the fMRI session was before and independent of the behavioral categorization task). Therefore, in the behavioral analysis, the data of 32 participants (f = 27, m = 5) aged between 18 and 29 years (*M* = 22.88, *SD* = 3.13) were considered. Participants gave written consent to voluntarily participate in the experiment and were self-reportedly suitable for fMRI measures. They either received course credits or were paid for their participation. The current study is in accordance to the Declaration of Helsinki and was approved by the local ethics committee of the Faculty of Psychology at the University of Münster (Germany).

Stimulus Material

The transitive actions employed in this study were designed based on the Semantic Event Chain (SEC) framework described by Wörgötter and colleagues (2013). Only transitive actions involving one active hand and one or two objects are included in this framework whereof 12 actions were selected for the current study that belonged to six action categories. The 12 selected actions were: turn, pull, rip off, uncover, take down, take away, out on top, put together, cut, scoop, hide, and put into. The execution of these transitive actions was recorded using an industrial camera (BASLER acA 1300-75gc) with a TV zoom lens (11.5–69 mm, 1:1.4) as well as an ASUS Xtion Live RGB-D sensor (ASUS TeK Computer Inc.) recording color as well as depth images. The video material presented in this study showed an actress from the front (BASLER camera) up to the shoulders performing the action with formed pieces of blue play dough on a white table. The ASUS Xtion Live recorded the actions from above, and its recordings were utilized for SEC time point extraction. For each object manipulation, 24–25 unique video takes were chosen for the final stimulus set (to account for the natural variation usually observed in human action performances), meaning that no video was repeatedly presented. In total, 294 action videos were shown to the participants. The videos had a frame rate of 23 fps. Each video started 10 frames before the hand lifts from the table to act and finished five frames after the hand lies back on the table with a video duration ranging from 68 frames to 165 frames ($M = 112.35, SD = 18.13$), that is, 2957 msec to 7174 msec ($M = 4885, SD = 788$). To increase the perceptual variability, all videos were vertically mirrored so that actions seemed to be performed by the left hand. Each participant saw 50% of the actions mirrored.

Adopted from our previous study (Pomp et al., 2021), the stimulus sequence was designed as a second-level counterbalanced De Bruijn sequence with seven conditions (six action categories + null condition) created using the De Bruijn cycle generator (Aguirre, Mattar, & Magis-Weinberg, 2011). Subsequently, condition labels of the six action categories were permuted to create 20 different stimulus lists. Per list, half of the stimuli were shown mirrored, and a second list contained the complement of these, which gave 40 different stimulus lists in total. For the second and third experimental sessions, the start of the individual stimulus sequence was shifted by one third and two thirds, respectively, to prevent recognition of the stimulus sequence as well as to prevent time-dependent effects. For the fMRI session, the stimulus sequence was subdivided into seven runs, and at the start of each run, the last two videos of the preceding run were repeated and then discarded from analyses to presume a continuous stimulus sequence (the first run started with the last two videos of the last run).

Video Segmentation and SEC Determination

As previously described (Pomp et al., 2021), we used an automated extraction of time points of TU events. Extracting these TU events automatically had the advantage that human bias could be avoided in the objective segmentation process. A flow diagram for the automated extraction of time points at which touching/untouching relations between object pairs change is shown in Figure 1. Here, we used the frame number to define the time points. The input to the algorithm is a sequence of RGB-D frames f_i (i = 1…n, n is the number of frames), and the output is a sequence of time events t_i (i = 1…m, m is the number of TU events, which was predefined manually). In the following subsections, we provide details for the four main steps of the algorithm.

Point Cloud Extraction and Preprocessing

Point clouds for each frame f_i were generated from depth images, which were acquired using ASUS Xtion Live sensor. ROI on the left side of the frame was cut as shown in Figure 1, because always only one hand was involved in the analyzed actions. Furthermore, point clouds were subsampled by a factor of four to reduce the number of points, this way speeding up the clustering procedure. Before clustering, ground plain subtraction was performed. Ground plain subtraction, that is, removing points

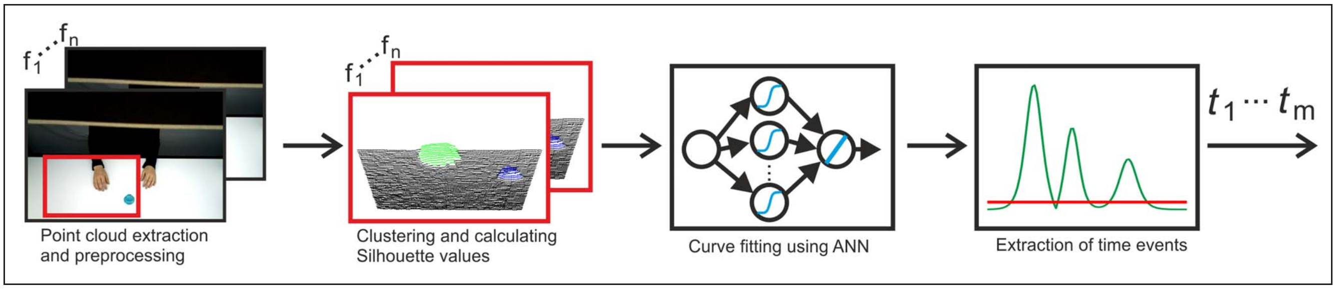
$f_1 \cdots f_n$
Point cloud extraction and preprocessing
$f_1 \cdots f_n$
Clustering and calculating Silhouette values
Curve fitting using ANN
Extraction of time events
$t_1 \cdots t_m$

Figure 1. Flow diagram for the automated extraction of time points of TU events (see Methods section for details). ANN = artificial neural network.

corresponding to the table, was done as follows. First, we fitted a flat 2-D surface and then removed all points from the 3-D point cloud data, which were above the fitted plane, that is, we first removed points $p_i = \{x_i, y_i, z_i\}$, if $z_i - Z1_i > th1$, where $Z1_i = P1(x_i, y_i)$ are corresponding points of the fitted plane $P1$, and $th1 = 0.02$ is the manually set threshold. Afterward, we fitted the plane one more time to the remaining background points $bg\_p_i = \{bg\_x_i bg\_y_i, bg\_z_i\}$ and we removed points that were below the fitted plane (see black points in Figure 2A, bottom row), that is, $p_i = \{x_i, y_i, z_i\}$, if $z_i - Z2_i < th2$, where $Z2_i = P2(bg\_x_i bg\_y_i)$ are corresponding points of the fitted plane $P2$, and $th2 = 0.01$ is the manually set threshold. The removed points p_i were not included to further cluster analysis. Thus, for the clustering step, we only used point clouds of the hand and objects.

Clustering and Calculation of Silhouette Scores

Clustering of points (objects) was performed based on 3-D point coordinates $p_i = \{x_i, y_i, z_i\}$ by using hierarchical clustering with Euclidean distance as a similarity measure and nearest distance as a linkage method. The clustering procedure was repeated K-1 times for each frame f_i (i = 1…n) with a predefined number of clusters $k = 2 \ldots K$, where K is the number of objects including the hand (but excluding the table). For each frame f_i, we computed a maximal Silhouette score as follows:

$$S(f_i) = max(S_k), k = 1 \ldots K), \text{with} \tag{1}$$

$$S_k(j) = sum[(min(D_{between}(j,l)) - D_{within}(j)) \\ /max(D_{within}(j), min(D_{between}(j,l)))]/N, \tag{2}$$

where $D_{within}(j)$ is the average distance from the j-th point to the other points in its own cluster, and $D_{between}(j, l)$ is the average distance from the j-th point to points in another cluster l. Here, N is the total number of points. The Silhouette score for each point j measures how similar that point is to points in its own cluster in comparison to points in other clusters. The values of the Silhouette score are between −1 and 1. Thus, when two clusters are getting closer, then the score $S(f_i)$ decreases, whereas it increases when clusters are moving apart (see Figure 2B).

Fitting of Silhouette Curve Using Artificial Neural Network

The time points of TU events can be extracted from the Silhouette curve; however, Silhouette scores are noisy because of noise present in the point cloud data obtained from the RGB-D sensor. Thus, we first filtered the Silhouette scores

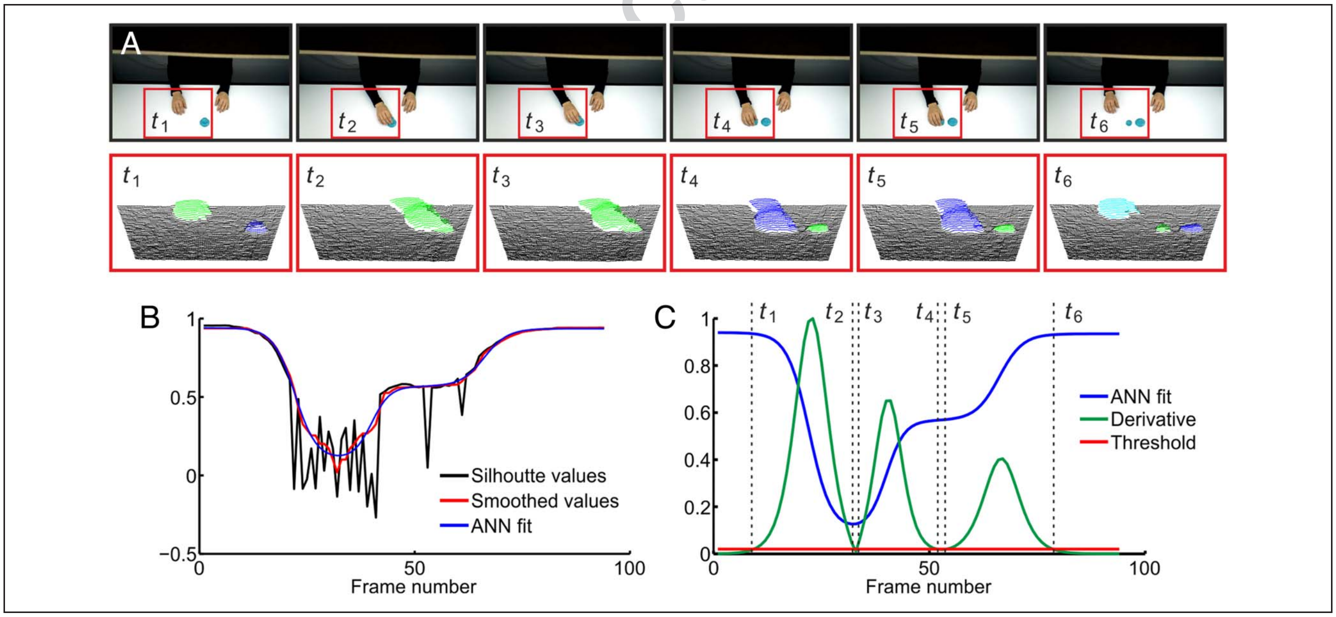

Figure 2. Schema of the procedure for automatically extracting the time points for touching and untouching events from an exemplary action, here "take down." (A) RGB images (top) from the above-scene installed ASUS Xtion Live RGB-D sensor and corresponding clustered point clouds (bottom). Clustered point clouds (objects) are color-coded and when two objects touch, they become one cluster with a shared color. When these objects untouch, the point clouds separate and one cloud changes to an individual color. (B) Raw silhouette values (black), smoothed silhouette values using a median filter (red), and fitted silhouette curve using an artificial neural network (ANN; blue). (C) Derivative of the ANN fit (green) and obtained time points of TU events after thresholding: t_1 = hand detaches from the table (i.e., first untouching); t_2 = hand touches the upper play dough object (i.e., first touching); t_3 = hand lifts the upper play dough object from the bottom play dough object (i.e., second untouching); t_4 = hand places the play dough object on the table (i.e., second touching); t_5 = hand detaches from the play dough object (i.e., third untouching); and t_6 = hand touches the table (i.e., third touching). Thus, in this example, a U-T-U-T-U-T event sequence is extracted. A demo source code of automated extraction that corresponds to the shown example can be downloaded from the OSF repository (DOI 10.17605/OSF.IO/MGQSF).

$S(f_i)$ using a median filter with a time window of 20 frames and then fitted filtered scores with an artificial neural network (ANN). This leads to a smooth curve with descending and raising slopes that allows extracting of time points in the next step. For fitting $S(f_i)$, we used a fully connected feed-forward network with one hidden layer where, in the hidden layer, we used a *tansig* transfer function and, in the output layer, a *linear* transfer function was used. The number of neurons in the hidden layer corresponded to the number of sigmoid functions needed to fit the Silhouette value function S (see Figure 2B), which corresponded to changes in cluster configuration, that is, if two clusters are merging, then objects are touching each other (T) and, if two clusters are getting apart, then objects are detaching from each other (U). In the given example in Figure 2 for a "take down" action, we have six TU events (hand lifts up from the table, hand touches upper play dough object, hand lifts the upper play dough object from the lower play dough object, hand places the play dough object on the table, hand leaves the play dough object, and hand touches the table). Thus, the TU events follow an irregular pattern of Ts and Us, and to represent two TU events, one sigmoid function is needed as demonstrated by an example shown in Figure 2C (see t_1, t_2; t_3, t_4; and t_5, t_6). The number of neurons h in the hidden layer was set based on the number of TU events m, that is, h = round(m/2). In this case, we used three neurons in the hidden layer. The network was fitted 10 times, and then the best outcome with respect to the minimal mean squared error between $S(f_i)$ and network's prediction $S_{ANN}(f_i)$ was used for the next step.

Extraction of Time Points

Finally, time points of TU events were extracted by applying dynamic thresholding to the derivative of the $S_{ANN}(f_i)$. We started with some initial threshold value $TH_{ini} = 0.01$ and increased it by 0.005 until the predefined number of TU time points was obtained. The time points were extracted at the frame numbers where the derivative of the $S_{ANN}(f_i)$ crossed the threshold value TH (see Figure 2C).

Whenever the algorithm misinterpreted the scene, which gave an error message, the extracted time points were checked against manual TU segmentation results and time points. Deviation from human TU segmentation, on average, was 4.14 frames (SD = 3.42), and in 93.02% of the cases, deviation was less than 10 frames (i.e., approx. mean value +2*SD). Thus, we corrected outliers in 6.98% of the cases, where TU event segmentation differences were larger than nine frames, by setting values of automated segmentation to corresponding values of human TU segmentation. The framework was implemented using MATLAB (https://www.mathworks.com) where standard MATLAB functions for clustering and ANN fitting were used. Extracted TU events were taken as machine-determined objective events (TUs) and the middle frames between two TU events were taken as corresponding non-events (nTU) to be maximally far away from an event.

Experimental Procedure

Congruent with our previous study (Pomp et al., 2021), participants completed three sessions. The MRI session was, on average, 4 days (range = 3–6) before the behavioral test–retest sessions, which were, on average, 14 days apart from one another (range = 14–18). In the first session, participants paid attention to the action videos while being in the MRI scanner. Action videos were back-projected onto a screen and displayed centrally with a screen resolution of 640 × 512 pixels by Presentation 20.3 (Neurobehavioral Systems Inc.). Participants viewed the screen binocularly through a mirror above the head coil. Attention-capturing questions followed 14% of the videos, asking whether an action description was appropriate for the preceding action video (see Figure 3A for the experimental trial design). Participants responded by pressing one of the two response keys with their right index and middle finger. Including anatomical scans and six short breaks during the task, the scanning time amounted to approximately 60 min. The overall duration of the first session was between 90 and 120 min including consent forms, instructions, preparation, scanning, and a short survey at the end.

The second experimental session comprised the unit marking task (Newtson, 1973). Participants saw the same videos as in the first session. Stimuli were presented on a 23-in. monitor by Presentation 18.1 (Neurobehavioral Systems Inc.), and participants were instructed to press a button with their right index finger whenever they think an action step is finished, that is, an event boundary occurred. Training trials were offered at the beginning, and two self-paced breaks were provided after one third and two thirds of the trials. This task took approximately 45 min. See Figure 3B for the experimental trial design. In the third session, this task was repeated to retest the unit marking behavior.

At the end of each of the three sessions, participants filled in a two-paged survey about their current state (mood, subjective health, tiredness before and after the task); the amount of sleep in the last night and whether this was more, less, or as much as usual; drug consumption (the day before and in general) ; their feeling of hunger before and after the task; and task-related questions about difficulty, monotony, task fatigue, inattentive phases, handedness of the shown actor, subjective guessing rate for the answered questions during the task, recognizability of the objects and actions, change in individual segmentation strategy within-session and between-sessions, and their segmentation strategy.

Behavioral Reliability Measures

Intra-individual Retest Reliability of Unit Marking Responses

As the unit marking procedure is a subjective judgment task and, therefore, responses cannot be right or wrong, retest reliability was assessed on the single-subject as well as on the group level to ensure that responses were consistent and meaningful. Details regarding these reliability

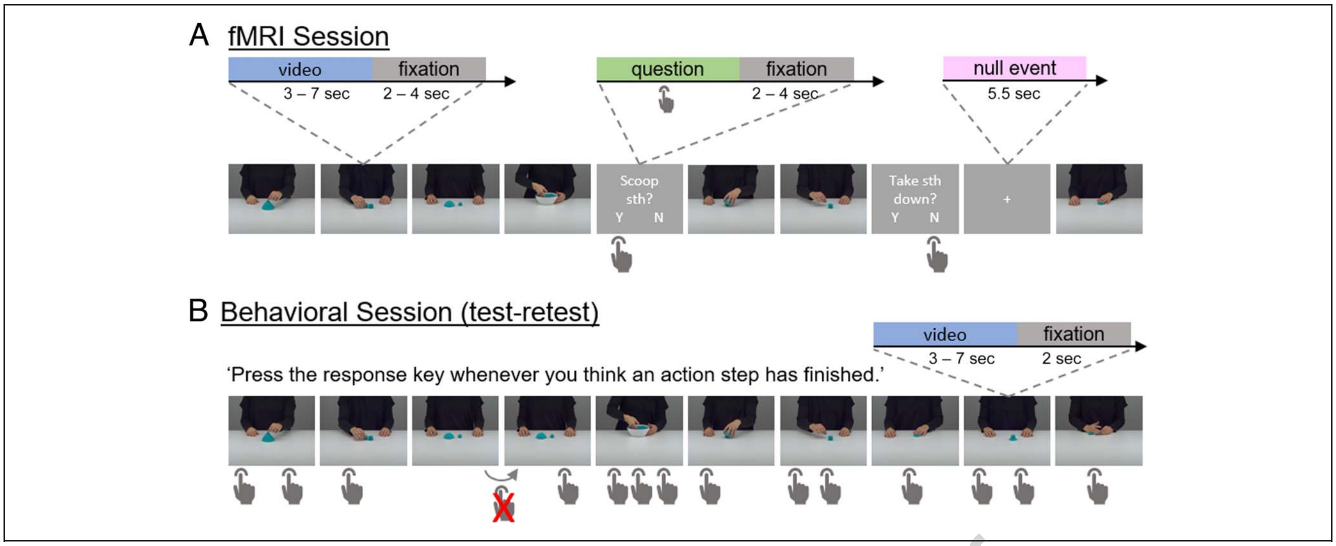

Figure 3. Experimental task design. (A) In the fMRI session, video trials (action video followed by a jittered ISI that showed a white fixation cross) and null event trials (showing a white fixation cross) were passively attended to but question trials (question followed by a jittered ISI that showed a white fixation cross) required participants to confirm or reject an action description with regard to the preceding action video by button press. The question disappeared only after button press and followed 14% of the action videos. For the video trials, here, each single frame represents a full action video plus ISI as indicated by the dotted lines. In total, 308 videos, 42 questions, and 49 null events were presented to each participant, separated in seven blocks with short breaks in between. (B) In the two subsequent behavioral sessions (test–retest), each participant saw the same videos in the same sequence as during fMRI and indicated by button press (hand icon) when they thought an action step had finished. In case no response was given (hand icon crossed out in red), the video at hand was repeated. Participants were instructed to use this mechanism in case they wanted to rewatch the video before indicating action steps. Thus, minimally one button press was necessary per action video but no instruction was given about the expected total number of button presses per action video. Each single frame in the figure represents a full action video plus an ISI that showed a white fixation cross, as indicated by the dotted lines. Example videos are provided in an OSF repository (DOI 10.17605/OSF.IO/MGQSF). The entire stimulus material is available via the Action Video Corpus Münster (AVICOM, https://www.uni-muenster.de/IVV5PSY/AvicomSrv/).

measurements have been previously described (Pomp et al., 2021). As the first step, responses were converted from milliseconds to frames (one frame amounting to a 1000/23 msec segment) to allocate each response to the correspondingly presented frame of the video. Note that we did not subtract any motor RT as participants were highly familiar with the kind of simple everyday actions that we employed, which they saw for the second and third time in the behavioral sessions. Hence, we adopted the premise that responses were delivered in clear anticipation of critical events in the videos, not in a reactive manner.

On the single-subject level, we examined whether test session responses matched retest session responses consistently. To this end, trials with an equal number of responses in the test and retest session were selectively used to define an individual temporal consistency criterion c_{i}, which was then applied to all trials independent of the number of responses. For each response in each of these equal-number-of-responses-trials, the absolute difference $d_{|t-t'|}$ in frames between test button press t and retest button press t' was determined and then averaged over all responses per participant. The upper bound of the 95% confidence interval (CI) of this mean difference score per participant was taken as individual criterion c_{i} for consistent button presses in the test and retest sessions. In summary, for each retest response t', it was determined whether a test response t appeared within the individual time window around the retest response ($t' \pm c_{\mathrm{i}}$). If this was the case, it was considered a consistent unit marking response. That is, the participant pressed the response key at the same time during the action video in the test and retest session. Subsequently, as a measure of intra-individual retest reliability, the percentage of consistent responses per participant was identified. These consistency rates were statistically compared with the corresponding object study's values using independent-samples t tests and the corresponding Bayesian test with JASP (JASP Team, 2024), and JZS Bayes factors are reported (Rouder, Speckman, Sun, Morey, & Iverson, 2009).

To ensure the validity of our intra-individual retest reliability results, we compared the intra-individual retest reliability results to random button presses. To this end, we extracted the time intervals between button presses (for the first button press in a video, we used the distance to the start of the video) of the test session per participant. From this distribution, we randomly drew and cumulated intervals to simulate random test session data while preserving the stochastic characteristics of the individual behavior. By this procedure, we generated 10 simulated test session data sets, calculated the percentage of consistent responses per participant based on the real retest session data (applying the identical protocol as for the actual behavioral data), and averaged this percentage per participant over the 10 simulations. To test whether the participants performed more reliably than randomly, we calculated a paired-samples t test between the actual percentage of consistent responses per participant and the percentages based on the simulated data sets.

Retest Reliability of Unit Marking Responses at the Group Level

To examine the unit marking responses on the group level, we smoothed the frame-by-frame data of all participants with a rectangular kernel of a width of three frames (3*(1000/23) ≈ 130.4 msec, referred to as *bin* hereafter). This means, for each video, we aggregated the number of responses for each frame f_t plus those from adjacent frames f_{t-1} and f_{t+1}. Thereby, we pooled the data of all participants. Maximally, one response per participant was taken into one bin of three frames so that the total number of participants was the maximum value a bin could reach. The bin value was then allocated to the middle frame f_t of the bin and will be referred to as *frame value* hereafter. Consequently, the frame value was set to zero if no response had occurred within the bin. To determine the group-level retest reliability, we correlated the time series of frame values per video between the test and the retest sessions (Pearson *r*). The *r* values per video were then Fisher *z*-transformed, averaged, and retransformed to *r* to give a mean correlation indicating group-level retest reliability. Furthermore, the *r* values per video were statistically compared with the corresponding object study's values using independent-samples *t* tests and the corresponding Bayesian test reporting JZS Bayes factors (Rouder et al., 2009) with JASP.

Group-consistent Unit Mark (M) Determination and Their Relation to TU Events

Determination of Group-consistent Unit Marks

The maximum frame value per video was taken to indicate a group-consistent unit mark (M) as it reflects the point of maximum group agreement. To assure the meaningfulness of these values, we utilized the 10 simulated test session data sets that were generated to evaluate intra-individual retest reliability. We applied the same protocol to these 10 simulated data sets as we did to the original data. Thereby, we determined simulated group-consistent unit marks and then compared their maximum frame values to the actual one per video.

To determine the non-unit-mark (nM) as relevant points in time for the fMRI analyses, one of the frames with the minimum frame value of zero was randomly chosen, excluding the first 12 and last 12 frames of each video. Ms and nMs were then used to model brain responses.

Temporal Convergence of Participant-determined Unit Marks and Objective Events

We investigated the temporal relation of Ms to TUs by evaluating whether the majority of Ms coincides with TUs. We examined how often an M was not further than two frames (i.e., maximally ~130 msec) away from a TU. Subsequently, we compared this result to randomly distributed unit marks to validate the systematics of the relationship.

Equal to the protocol for the test–retest performance of individual participants, we shuffled the time intervals generated by the unit marks, randomly drew from this shuffled distribution, and cumulated intervals to simulate random unit marks while preserving the stochastic characteristics of the group behavior. This way, we generated 10 simulated unit mark data sets, examined per data set the proportion of simulated Ms being not further than two frames away from a TU, and then calculated a one-sample *t* test to compare simulated and actual coincidence rates.

Effects of Object–Action Associations on the Temporal Relation of Ms to TUs

Building on our previous study that showed that Ms were systematically delivered in relation to TU events (Pomp et al., 2021), we hypothesized weak object–action associations to increase the temporal proximity of M to TU. As the first analytic step, we tested whether the M-TU difference distributions differed between studies using the Mann–Whitney *U* test, and tested for equality of variances using Levene's test. Following our hypothesis that Ms are closer to TU events in dough actions (independent of whether they appear before or after the TU), absolute difference values were further analyzed. As these absolute temporal differences between Ms and its closest TUs in both studies had a negative binomial distribution, we fitted a generalized linear (negative binomial) model using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in the R programming language (URL https://www.R-project.org/). In the model, the absolute temporal differences between Ms and TUs, measured in frames, were predicted by *Study* (i.e., Dough vs. Object) and *Event Type* (i.e., Touch vs. Untouch). In the model, the action categories of the videos were used as a random intercept:

$$\begin{aligned} absolute(M\text{-}TU) \sim\ & Study \times EventType \\ & +(1|\,ActionCategory). \end{aligned} \tag{3}$$

fMRI Data Acquisition

Structural and fMRI data were acquired using a 3-Tesla Siemens Magnetom Prisma MR tomograph with a 20-channel head coil at the Translational Research Imaging Center of the University Hospital Münster. High-resolution, T1-weighted images were obtained by a 3-D-multiplanar rapidly acquired gradient-echo sequence (scanning parameters: 192 slices, repetition time = 2130 msec, echo time = 2.28 msec, slice thickness = 1 mm, field of view = 256 × 256 mm^2, flip angle = 8°). For the functional images, a BOLD contrast was measured by gradient-EPI. Seven EPI sequences were used to measure the seven experimental blocks (scanning parameters: 33 slices, TR = 2000 msec, echo time = 30 msec, slice thickness = 3 mm, field of view = 192 × 192 mm2, flip angle = 90°).

fMRI Data Analysis

Preprocessing

Anatomical and functional images were preprocessed using the Statistical Parametric Mapping software (SPM12; The Wellcome Centre for Human Neuroimaging) implemented in MATLAB R2019a. Preprocessing included slice time correction to the first slice, realignment to the mean image, co-registration of the individual structural scan to the mean functional image, normalization into the standard anatomical MNI (Montreal Neurological Institute) space on the basis of segmentation parameters, as well as spatial smoothing using an isotropic 8-mm FWHM Gaussian kernel. To remove low-frequency noise, a 128-sec temporal high-pass filter was applied to the time-series of functional images.

fMRI Design Specification and Whole-brain Statistics

The statistical analyses of the functional images were done using SPM12, implementing a general linear model for serially autocorrelated observations (Worsley & Friston, 1995; Friston et al., 1994) and a convolution with the canonical hemodynamic response function. As regressors of no interest, the six subject-specific rigid-body transformations obtained from realignment were included. The volumes of the first two video presentations of each EPI were discarded to allow for T1-equilibrium effects. To investigate functional areas specialized in the processing of subjective action boundaries, as well as objective T and U events, a general linear model was constructed including eight regressors of interest coding for onsets and durations of the specific event types: video trial, group-consistent unit mark of the test–retest session (M), no unit mark in the test–retest session (nM), objective touching event (T), objective untouching event (U), no touching or untouching event (nTU), null event, and question trial. For each of the 340 Ms, an nM was determined ($n = 340$; see Determination of Group-consistent Unit Marks section) and included in the design. Likewise, all 735 touching and all 808 untouching events were included and correspondingly 735 nTUs (see Video Segmentation and SEC Determination section). Both types of noncritical events (nTU and nM) appeared distributed over the video duration and were chosen to be maximally far away from their corresponding events (TU and M, respectively). The rapid succession of Ms and TUs with naturally jittered interevent intervals made it possible to differentiate associated BOLD responses, and the difference in frequency of occurrence ensured the overall low overlap between M and TU events. Moreover, we applied the post hoc variance inflation factor (VIF) method using the CANlab imaging analysis tools (https://canlab.gi http://canlab.github.io/ thub.io/) to rule out multicollinearity issues and this yielded VIFs below 10 (object study VIFs < 7.2, dough study VIFs < 7.9), speaking against a severe issue of collinearity.

On the first level, t-contrasts for Ms versus nMs were calculated and submitted to a second-level t test to detect functional areas specialized in the processing of group-determined event boundaries. Analogously, t-contrasts for T versus nTU, U versus nTU, and the complete video trials versus null events were conducted on the first level and then passed to a second-level t test. To elucidate the central question of the object–action association effect, we contrasted activity patterns for play dough actions of this study to activity patterns for object actions of our previous study in a second-level two-sample t test. We did this for all full-length videos as well as time-point specifically at M, T, and U events. Importantly, because we only considered interactions, all contrasts controlled for the main effects of group, action type, and so forth.

To identify brain areas where neural activity was significantly explained by both object and play dough actions' events, we performed conjunction analyses testing against the conjunction null hypothesis, p(false discovery rate [FDR]) < .005 (Nichols, Brett, Andersson, Wager, & Poline, 2005) using a second-level, one-way ANOVA on individual statistical maps derived from the M > nM, T > nTU, U > nTU, and video > null contrasts.

For the second-level, whole-brain analyses, we applied FDR correction at $p < .005$ peak level and a cluster extent threshold of 15 voxels. Activity patterns were visualized using bspmview (DOI 10.5281/zenodo.595175) in MATLAB R2022a, and graphs for visualization were generated using the *ggplot2* library (Wickham, 2016) in RStudio (R Core Team, 2022). We uploaded the unthresholded statistical maps to NeuroVault.org (Gorgolewski et al., 2015), which are available at https://neurovault.org/collections/16065/.

ROI Analyses

To inspect the effects of object–action associations more specifically in the hypothesized regions, we additionally performed planned ROI analyses. Addressing aIPL, we used area PFt (Caspers et al., 2006, 2008) of the Julich-Brain Cytoarchitectonic Atlas (Amunts, Mohlberg, Bludau, & Zilles, 2020; Eickhoff et al., 2005), noting that also relevant peak MNI coordinates of our previous study all fell into this field (Schubotz et al., 2014). The aIPL Julich-Brain ROI was created using the SPM anatomy toolbox (www.fz-juelich.de/inm/inm-7/JuelichAnatomyToolbox). As second ROI, we used the PHC. We defined the extend of the PHC ROI using the Harvard-Oxford anatomical atlas (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases) and the software MRIcron (https://www.mccauslandcenter.sc.edu/mricro/mricron), including voxels if the atlas labeled them as "Parahippocampal Gyrus, posterior division" or "Parahippocampal Gyrus, anterior division" with a probability of > 25% (Li et al., 2016; Ward, Chun, & Kuhl, 2013). As third ROI, we employed the temporo-occipital area sensitive to biological motion (BMA), which we gratefully adopted from a recent meta-analysis on the functional organization of the posterior lateral temporal cortex (Hodgson, Lambon Ralph, & Jackson, 2022). We extracted the mean contrast estimates of our main contrasts for each ROI using the Marsbar toolbox (Brett, Anton, Valabregue, & Poline, 2002), which were then compared between studies by a two-sample *t* test (unequal variances, $\alpha = .05$, *two-sided*) per region using MATLAB R2022a.

RESULTS

Behavioral Reliability Measures

Intra-individual Retest Reliability of Unit Marking Responses

Concerning single-subject retest reliability, on average, 63.27% were consistent responses (i.e., the test response matched the retest response in time) ranging between the participants from minimally 48.18% to maximally 71.22% ($SD = 6.34$). The individual consistency criterion c_i, which defined the width of the time window around the retest response separately for each participant, was minimum 3.9 frames (i.e., ~170 msec), median 5.7 frames (i.e., ~248 msec), and maximum 11.3 frames (i.e., ~491 msec). Importantly, the consistency of the participants' unit marking behavior was significantly higher than the consistency of simulated random button presses, $t(31) = 17.81$, 95% CI [28.65, 36.07], $p < .001$, $d = 3.15$, *two-sided*. Thus, participants' unit marking behavior followed a specific nonrandom pattern and was intra-individually consistent across the test–retest sessions. Compared with the object manipulation study (Pomp et al., 2021), the intra-individual retest reliability was similar regarding the individual percentages of consistent responses as indicated by a Bayesian independent-samples *t* test that showed evidence for the null hypothesis and its classical counterpart yielding nonsignificant results, $BF_{01} = 3.502$), $t(61) = 0.139$, $p = .89$, $d = 0.035$, *two-sided*. With regard to the respective comparison to random button presses, a greater Cohen's *d* of 3.15 in dough study's individual retest reliability versus 1.91 in the object study, indicated that individual participants' segmentations were even more systematic for dough videos.

Retest Reliability of Unit Marking Responses at the Group Level

Corresponding to the single-subject retest reliability results, between-subject unit marking behavior was consistent, as revealed by a highly significant correlation between group-based test–retest segmentation performance. That is, correlations testing the group level retest reliability gave a mean correlation of test and retest smoothed time series of frame values per video of $r_z(292) = .72$ ($r_{\min} = .40$, $r_{\max} = .90$; each individual correlation per video being significant, all $p \leq .0001$). Compared with the object manipulation study (Pomp et al., 2021), group-level retest reliability was significantly higher for dough manipulations, $t(586) = 17.153$, $p < .001$, $d = 1.415$, *two-sided* ($BF_{10} = 1.365 \times 10^{+50}$).

Group-consistent Unit Mark (M) Determination and Their Relation to TU Events

Determination of Group-consistent Unit Marks

The frame with the maximum frame value in a video that represents the maximum agreement between participants was taken as group-consistent M. On average, this maximum frame value was 9.93 ($SD = 2.00$), ranging from 6 to 18. All maximum frame values were at least 2 *SD*s above the mean frame value of the respective video, following previous approaches (Pomp et al., 2021; Schubotz et al., 2012). In contrast, the maximum frame values resulting from simulated random unit markings ranged, on average, between 6.11 and 6.37 (i.e., < 9.93). In none of these simulated data sets all maximum frame values passed the criterion of being at least 2 *SD*s above the respective video mean frame value. Taken together, this finding suggests that the participants did not segment the action videos randomly, and overall, the group showed a specific nonrandom segmentation behavior.

Furthermore, we inspected the relation between the number of Ms and the number of TUs per video: The number of Ms per video on group level ranged from one to four ($M = 1.2$, $SD = 0.36$, $n = 294$) and was significantly lower than the number of TUs per video that ranged from three to six ($M = 5.2$, $SD = 1.01$, $n = 294$; $t(586) = 64.97$, 95% CI [3.97, 4.22], $p < .001$, $d = 5.36$, *two-sided*). On the single-subject level, the average number of individual test–retest consistent unit marking responses per video ranged from 0.6 to 1.9 with a mean of 1.4 ($SD = 0.26$, $n = 294$). Crucially, the number of individually consistent unit marking responses per action significantly correlated with the number of TUs per action video, $r(292) = .55$, $p < .0001$, as well as the number of group-level Ms that positively correlated with the number of TUs, $r(292) = .17$, $p = .003$, both pointing to a systematic relationship between the number of Ms and TUs.

Temporal Convergence of Participant-determined Unit Marks and Objective Events

With regard to the temporal relation of Ms to TUs, for more than one third (39.1%) of the Ms, the time lag to the next TU was maximally two frames, that is, ±130 msec. This coincidence rate was significantly higher than the coincidence rates obtained from the 10 sets of simulated random unit marks, $t(9) = -9.46$, 95% CI [24.32, 30.03], $p < .0001$, $d = 2.99$, *two-sided*, underpinning our expectation that Ms were systematically delivered in relation to TUs. Compared with the object manipulation study (Pomp et al., 2021), this significant coincidence rate's difference to simulated random unit marks was more pronounced in the dough study with a Cohen's *d* of 2.99 compared with a

Cohen's *d* of 1.27 in the object manipulation study, indicating a stronger systematicity on the group level.

Effects of Object–Action Associations on the Temporal Relation of Ms to TUs

As shown above and in Pomp and colleagues (2021), single subject as well as group behavior was consistent across test and retest sessions in both studies, and descriptive behavioral values regarding the number of Ms per video were comparable between both studies. Still, and as hypothesized, our current results showed a higher coincidence rate between Ms and TUs, with 39.12% for play dough actions compared with 28.3% for object actions. Furthermore, inspecting the temporal distances between Ms and their closest TUs, Levene's test for equality of variances indicated unequal variances, $F(1, 688) = 5.71$, $p = .017$, with dough action M-TU distances having a significantly lower variance ($Var = 23.34$) than object M-TU distances ($Var = 37.12$) and thus, as hypothesized, a smaller spread of data. The distributions of M-TU differences differed significantly between studies ($W = 48885.50$, $p < .001$, $r = -.18$, $n = 690$). To test our hypothesis that Ms are temporally closer to TUs when only weak object–action association is present, we compared the absolute temporal delay between the occurrence of M and TU for object and play dough actions. Generalized linear (i.e., negative binomial) modeling showed that although the two studies were not significantly different, *Wald* $X^2(1) = 0.02$, z test $= -0.02$, $p = .89$, $d = 0.02$; dough: *mean* $= 4.1 \pm 3.2$, *median* $= 3.5$; object: *mean* $= 5.5 \pm 4.4$, *median* $= 5$, generally, the M-T differences differed significantly from the M-U differences, *Wald* $X^2(1) = 13.87$, z test $= 0.30$, $p < .001$, $d = 0.30$; M-T: *mean* $= 4.84 \pm 4.0$, *median* $= 4$; m-u: *mean* $= 4.78 \pm 3.7$, *median* $= 4$. Furthermore, a significant interaction between *Event Type* (touch, untouch) and *Study* (dough, object) was observed, *Wald* $X^2(1) = 15.98$, z test $= -0.48$, $p < .001$, $d = -0.48$. To elucidate this interaction, we conducted Bonferroni-adjusted post hoc contrasts, which revealed that although the M-T differences were significantly different between the two studies, z-ratio(object/dough) $= -3.41$, $p < .001$ (Figure 4B), the M-U differences were not significantly different, z-ratio(object/dough) $= 0.13$, $p = .89$ (Figure 4D). These results indicate that actions were segmented closer to T events in case of weak object–action associations. For signed and unsigned M-T and M-U differences, see Figure 4. The signed temporal differences in Figure 4A and Figure 4C illustrate when participant-judged Ms appear in relation to T and U events. Moreover, the unsigned differences shown in Figure 4B and Figure 4D address the question whether Ms were temporally closer to T or U events independent of the sign.

fMRI Results

To investigate the whole-brain and ROI effect of object–action associations, we compared brain activity patterns of the two studies for the full video length (video > null)

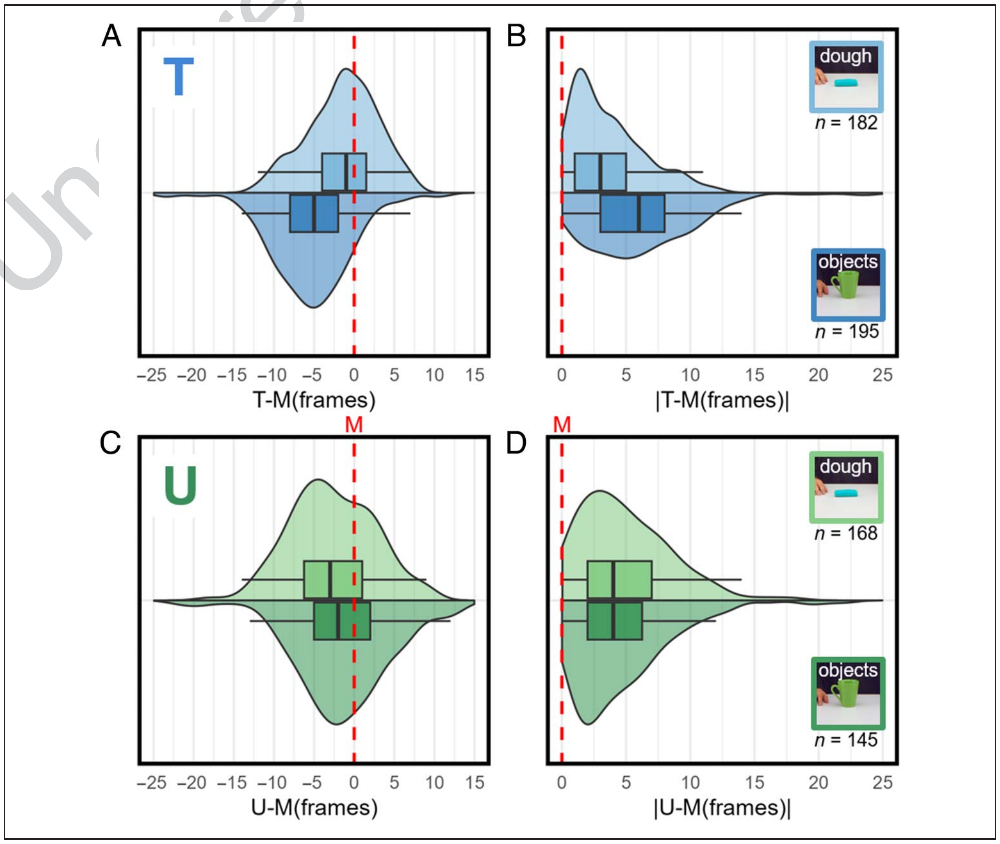

Figure 4. The temporal relation of unit marks (M) to touch (T) and untouch (U) events. The distribution of M to T differences (blue) shown as signed values (A) and unsigned values (B) given in frames and grouped by study (top, light: dough study; bottom, dark: object study). Similarly, the distribution of M to U differences (green) shown as signed values (C) and unsigned values (D) grouped by study. The red dashed line at $x = 0$ indicates when participants behaviorally segmented actions, that is, a unit mark (M) was determined. The action videos had a frame rate of 23 frames per second (1 f ≙ 43.5 msec).

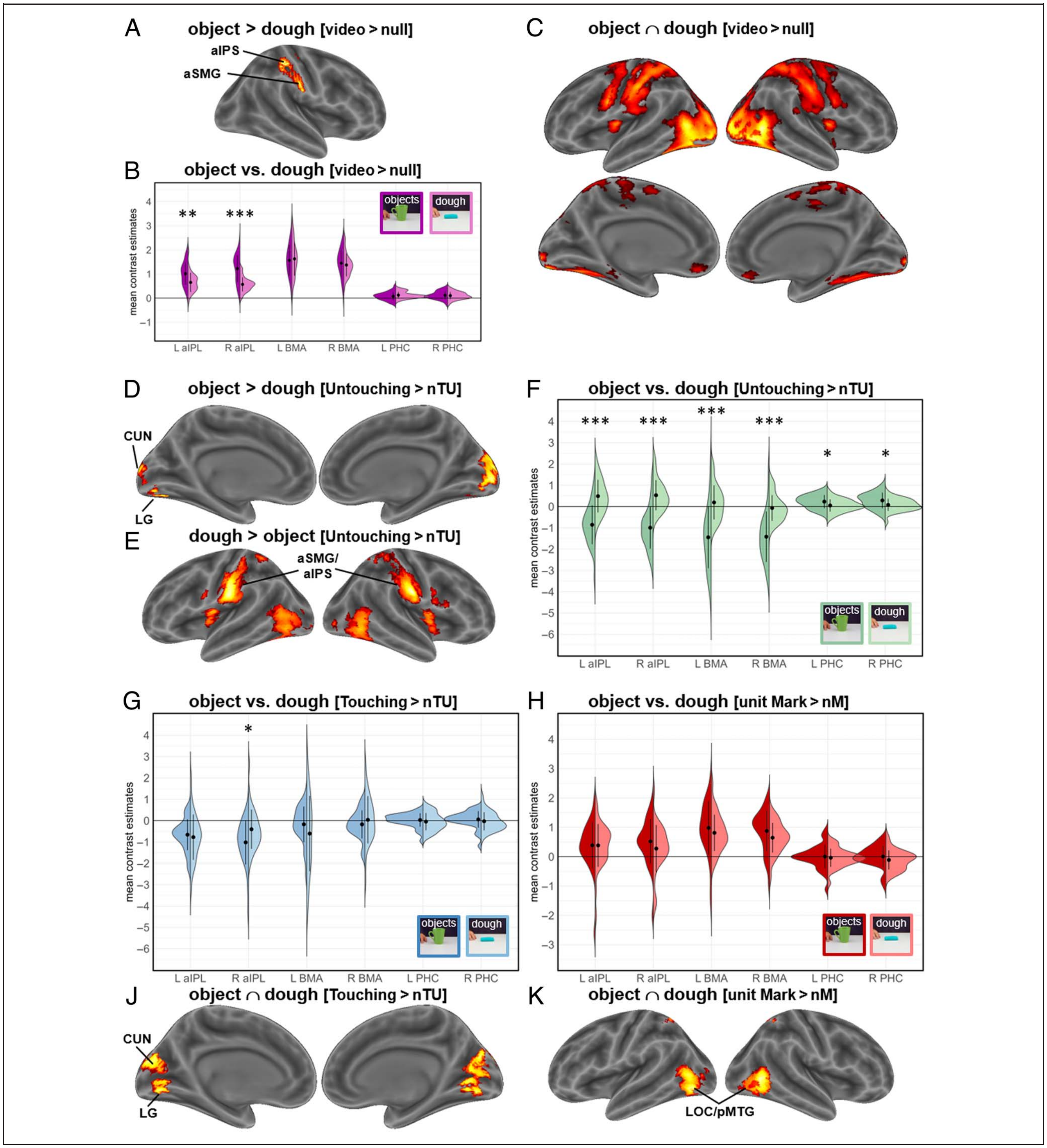

Figure 5. fMRI activation in contrasts and conjunctions between object and dough data at $p < .005$, peak-level FDR-corrected, and ROI analyses of left (L) and right (R) aIPL, biological motion area (BMA), and PHC. A, B, and C illustrate the between-studies' effects for the full video length (video > null; purple). D and E show the whole-brain effects, and F shows the ROI analyses for the between-study comparison at untouching events (U > nTU; green). ROI analyses for the between-study comparison at touching events (T > nTU; blue) are illustrated in G and for unit marks (M > nM; red) in H. Finally, between-study conjunction results are depicted in J for touching events and K for unit marks. For ROI analyses: Mean contrast estimates were extracted from the contrasts video > null, U > nTU, T > nTU, and M > nM of the object (dark shade) and dough (light shade) study. Note that all comparisons show Group × Event interaction effects. For objects $n = 31$, for dough $n = 33$. Statistics: two-sample t tests (*two-tailed*). $*p < .05$, $**p < .01$, $***p < .001$. Unthresholded statistical maps of the whole-brain analyses have been uploaded to NeuroVault.org and are available at https://neurovault.org/collections/16065/.

Table 1. Maxima of Activation from the Contrasts and Conjunctions of Dough Study and Object Study Contrasts at $p < .005$ Peak-level FDR-Corrected

					MNI Coordinates		
Macroanatomical Location	*Abbreviation*	*H*	*Cluster Extent*	*t Value*	*x*	*y*	*z*
U > nTU							
Dough > object							
Anterior supramarginal gyrus/ ventral postcentral sulcus	aSMG/ vPoS	R	672	9.48	60	−16	26
Anterior supramarginal gyrus	aSMG	R		9.38	66	−16	29
Anterior intraparietal sulcus	aIPS	R		7.19	57	−25	53
Anterior supramarginal gyrus/ ventral postcentral sulcus	aSMG/ vPoS	L	742	9.17	−60	−19	23
Anterior intraparietal sulcus	aIPS	L		7.27	−57	−25	50
Superior parietal lobule	SPL	L		5.82	−18	−49	71
Mid-insula	MIC	L	152	7.94	−39	−4	14
		R	162	7.08	39	−1	14
Insula	IC	R		6.48	39	−1	−1
Lateral occipito-temporal cortex	LOTC	R	341	7.17	51	−70	−7
Posterior inferior temporal gyrus	pITG	R		6.36	51	−58	−19
Lateral occipito-temporal cortex	LOTC	L	379	7.16	−48	−73	−1
Ventral precentral gyrus	preCG	R	127	5.70	57	11	35
		L	23	4.30	−57	8	29
Cerebellum	CER	L	25	5.39	−15	−67	−46
Object > dough							
Lingual gyrus	LG	R	445	7.18	15	−88	2
		L		6.29	−24	−76	−4
Cuneus	Cun	R		6.52	9	−94	11
		L		6.22	−9	−100	14
Object ∩ dough							
Parahippocampal cortex	PHC	R	18	6.37	33	−55	−7
		L	7	5.37	−33	−55	−7
Dorsal premotor cortex	PMd	L	29	5.48	−21	−10	56
Video > null							
Object > dough							
Anterior intraparietal sulcus/ postcentral sulcus	aIPS/PoS	R	174	6.49	42	−31	47
Anterior supramarginal gyrus	aSMG	R		6.00	60	−19	32
Object ∩ dough							
Posterior middle temporal gyrus	pMTG	R	1834	15.98	48	−64	2
Inferior occipital gyrus	IOG	R		15.09	42	−73	−7
Middle occipital gyrus	MOG	R		14.30	30	−91	5

Table 1. (*continued*)

					MNI Coordinates		
Macroanatomical Location	*Abbreviation*	*H*	*Cluster Extent*	*t Value*	*x*	*y*	*z*
Hippocampus	HC	R		5.01	24	−13	−16
Posterior middle temporal gyrus	pMTG	L	1665	15.24	−45	−67	5
Lingual gyrus	LG	L		13.12	−27	−91	−10
Inferior occipital gyrus	IOG	L		12.60	−39	−76	−7
Fusiform gyrus	FG	L		11.20	−39	−61	−13
Parahippocampal gyrus	PHG	L		3.90	−24	−28	−16
Insula	IC	L	4379	10.79	−36	−7	14
Ventral postcentral sulcus	vPoS	L		10.79	−51	−25	41
Ventral premotor cortex	PMv	L		10.70	−57	5	32
Insula	IC	R	4379	10.37	36	−4	14
Postcentral gyrus	PoG	R		10.16	54	−19	41
Anterior intraparietal sulcus	aIPS	L		9.94	−42	−31	47
Cerebellum	CER	L	117	7.97	−9	−73	−43
		R	89	6.44	12	−73	−43
Rectal gyrus	RG	L	105	6.02	0	29	−22
Mid cingulum	MCC	R	22	4.72	15	−16	44
SMA	SMA	L	81	4.63	−9	−1	56
		R		4.44	9	2	56
Amygdala	AMY	R	29	4.34	36	−1	−16
M > nM							
Object ∩ dough							
Lateral occipital cortex	LOC	L	252	8.41	−48	−73	−7
		L		8.02	−45	−70	2
Posterior middle temporal gyrus	pMTG	R	274	8.29	48	−64	2
Superior parietal lobule	SPL	R	40	5.55	18	−58	68
		L	49	5.24	−21	−58	65
T > nTU							
Object ∩ dough							
Cuneus	CUN	L	657	6.75	−6	−82	23
		R		5.99	9	−79	26
Lingual gyrus	LG	L		6.17	−9	−76	−1
		R		5.87	12	−76	−4

H = hemisphere; L = left; R = right; U = untouching events; nTU = non-(un-)touching events; M = unit marks; T = touching events.

as well as for the time-point-specific activation contrasts at M (M > nM), T (T > nTU), and U (U > nTU) events. Please note that we always refer to the just enumerated contrasts when we refer to M, T, and U as events. This means that all reported between-study time-point-specific effects are interaction effects (e.g., object study M > nM vs. dough study M > nM), ruling out group effects.

The Entire-video Effects

Comparing object versus dough videos for the full video length (Figure 5A), we found a single cluster in the right aIPL to be significant, including the posterior bank of the ventral postcentral sulcus, the anterior supramarginal gyrus (aSMG), and the anterior intraparietal sulcus (aIPS). ROI analyses affirmed and extended this result by yielding significant activation increases not only in the right, $t(44.53) = 5.77, p < .001, d = 1.45$, *two-tailed*, but also in the left, $t(58.62) = 3.30, p = .002, d = 0.82$, *two-tailed*, aIPL for strong object–action associations in object manipulations (Figure 5B). The reverse contrast did not yield significant results. For the common activity between studies during the entire action videos, see the corresponding conjunction results as illustrated in Figure 5C and Table 1.

Interaction Effects at Specific Time Points in the Video (M, T, U)

Contrasting object versus dough videos at critical time points' contrasts, there were no significant differences at M or T events but at U events (Figure 5D) in bilateral cuneus and lingual gyrus. Moreover, ROI analyses (Figure 5F) showed increased activity in bilateral PHC, left PHC: $t(60.96) = 2.43, p = .018, d = 0.60$, *two-tailed*; right PHC: $t(56.86) = 2.53, p = .014, d = 0.63$, *two-tailed*.

The opposite contrast of dough versus object videos (Figure 5E) led to higher BOLD responses at U events in bilateral aIPS extending into the SMG in the right hemisphere and to the superior parietal lobule (SPL) in the left hemisphere; furthermore, bilateral insula, bilateral lateral occipital cortex (LOC), and bilateral ventral precentral gyrus activations were detected. The ROI analyses (Figure 5F) showed dough versus object effects at U events in the bilateral aIPS (left: $t(58.75) = 6.35, p < .001, d = 1.58$, *two-tailed*; right: $t(54.27) = 7.06, p < .001, d = 1.76$, *two-tailed*) and bilateral BMA, left: $t(45.87) = 5.51, p < .001, d = 1.39$, *two-tailed*; right: $t(44.01) = 5.68, p < .001, d = 1.43$, *two-tailed*. Moreover, comparing dough versus object yielded a significant increase in the right aIPS ROI for T events, $t(60.00) = 2.51, p = .015, d = 0.62$, *two-tailed* (Figure 5G), whereas whole-brain contrasts at T events were nonsignificant. Neither whole-brain nor ROI analyses revealed significant differences between dough and object videos' time-point-specific M activity (for ROI analyses, see Figure 5H).

Conjunction Effects at Specific Time Points in the Video (M, T, U)

To examine whether dough video effects at M, T, and U resemble corresponding object video effects, conjunctions between studies were calculated, and they generally replicated T- and M-specific activity patterns. For T, the conjunction yielded bilateral cuneus as well as bilateral lingual gyrus activity (Figure 5J), and for M, the conjunction revealed bilateral LOC and bilateral SPL activation (Figure 5K). Notably, the U-specific activity was partially replicated. The conjunction showed overlapping activity in bilateral PHC and left lateralized dorsal premotor area. See Table 1 for the peak maxima of the described contrasts and conjunctions.

DISCUSSION

Previous studies have shown that motion information is of central importance for the brain segmentation of observed actions. Accordingly, we recently showed that touching–untouching events indicating maximal motion changes are an efficient cue for participant-judged event boundaries and are associated with specific processing steps at the neural level (Pomp et al., 2021). In the current fMRI study, we hypothesized that objects also have a significant influence on action segmentation because they are associated with specific manipulations. Extending the previous study, we replaced objects with formed pieces of dough to weaken the object–action associations and compared the behavioral and neural processes of action segmentation between the two fMRI studies. Findings show that, indeed, objects influence action segmentation behavior and the neural processing at specific events.

Behavioral findings showed that touching–untouching information was used for action segmentation, no matter whether object-associated action knowledge was strong or weak. Moreover, intra-individual and group retest reliability measures corroborated reliable segmentation behavior for both studies, as tested via a unit marking procedure (Newtson, 1973). In both object and dough videos, participants reported event boundaries systematically in relation to (un-)touchings. However, as expected, the variance in segmentation behavior was significantly smaller when object–action associations were weak. In addition, when compared with random button presses, participants' segmentations were even more systematic for dough videos. Accordingly, the retest reliability on the group level was higher. In summary, this suggests a lower dispersion of data values in the absence of strongly learned object–action associations. Besides, behavioral measures of reliability and consistency, as well as event frequencies and systematicity in dough action segmentation, resembled those in object actions, corroborating the interpretability of subjective event boundaries and their systematic relationship to objective touching and untouching events.

Inspecting the temporal relationship between participant-judged event boundaries and (un-)touchings, we observed that, as hypothesized, the coincidence rate between unit marks and (un-)touchings was higher for dough actions. Furthermore, here again, the specific response pattern's coincidence rate differed from simulated random unit marks' coincidence rate more pronounced when object–action associations were weak. This result indicated higher behavioral systematicity in the absence of strong object–action associations. In addition, actions were segmented temporally closer to touching events when object–action associations were weak, indicating increased reliance on objective touching events. This is in line with our previous findings suggesting especially touching events announcing an untouching event to be important anchor points of behavioral action segmentation (Pomp et al., 2021). Note that the systematic relation of Ms to TU does not imply a generally high overlap of events in time, as there were considerably more TU events (735 touching and 808 untouching events) than participant-judged unit marks (340). Thus, consistent with our first study (Pomp et al., 2021), we also found in the dough manipulation study that participant-judged event boundaries very frequently coincided with TU events, but the majority of TU events did not coincide with a participant-judged event boundary. Future studies need to investigate exactly which TU events are used as anchor points triggering subjective boundary detection.

Taken together, the smaller spread of data and the larger behavioral systematicity in the responses to dough videos showed that the subjective event boundaries relied even more on touching events when strong object–action associations are absent. Thus, before having experience-based knowledge of object-associated actions, the individual presumably relies particularly strongly on objective (un-)touchings. In general, our behavioral findings corroborated that relational changes in the form of touchings and untouchings of objects, hands, and ground represent meaningful anchor points in subjective action segmentation. This finding is critical for creating objective event boundaries that can be used for meaningful action segments. Hard and colleagues (2006) underpinned that goal-based event schemas are not required to detect event structure and concluded that physical changes in the actions subserve event segmentation, measured as bursts of change in movement features. Zacks and colleagues (2009) came to a similar conclusion that movement variables play an important role in action segmentation using a motion tracking system and transcribing movement as a set of 15 variables. Notably, both studies agreed that event structure can be extracted from movement parameters but used complex and costly methods to quantify movement. This is not required in our current approach, which illustrates its practical advantage in this area of research.

Extending the picture arising from the behavioral analyses, fMRI data revealed that object information had significant effects on how the brain processes different types of event boundaries. Importantly, based on interaction contrasts from within-study main effects, our approach controlled for mere perceptual differences arising from the sight of objects or dough pieces. We expected that aIPL and PHC processing might be more relevant for the segmentation of object-directed actions than dough-directed actions, whereas the opposite might be true for an area sensitive to biological motion (BMA). Our findings partly confirmed these hypotheses and also revealed that, among the three types of event boundaries, untouchings were associated with prominent differences between object and dough videos. By contrast, modeling brain data with touching events and participant-judged unit marks replicated the effects that we found for object-directed action segmentation largely (see Appendix). We will, therefore, focus our discussions on untouching events. As shown in our previous study (Pomp et al., 2021), participants reported event boundaries in response to a subset of touching–untouching motifs, that is, the point in time where the observed movement increased significantly from null (touching) to positive change (untouching) and thus became highly informative in respect of the upcoming manipulation. We suggest that object–action associations made the biggest difference at untouching events because participants had to rely much more on movement information when observing dough videos as compared with object videos.

At untouching events, activity increased for dough versus object manipulations in the prespecified ROIs aIPL and BMA, along with bilateral insula and bilateral ventral precentral gyrus activity. Conversely, object versus dough manipulations led to increased bilateral activity in the PHC ROI along with bilateral lingual and cuneal activity. These findings corroborated our hypotheses (a) regarding the increased impact of biological motion for action segmentation in the absence of strong object–action associations and (b) regarding the particular role of long-term mnemonic associations of object and context as reflected by parahippocampal sites for action segmentation in the presence of strong object–action associations.

In light of the fact that (un)touching events provide abstracted dynamic information, the BOLD difference in the BMA at untouchings is a strong indication that participants rely heavily on hand movements to meaningfully process action segments in the absence of strong object–action associations. The employed BMA ROI was functionally defined in a recent meta-analysis (Hodgson et al., 2022) for biological motion. Importantly, the reported effect in our study cannot be because of an increase in motion in the stimuli per se because videos differed only with regard to the target of manipulation, dough, or everyday objects. BMA forms part of the ventro-dorsal route for visual input (Binkofski & Buxbaum, 2013), which has been argued to process information aconceptually (Mahon, 2023), that is, without "knowing" what the moving object is. Concerning the analysis of critical events in studies on action observation, participant-judged event boundaries have been found to activate BMAs (Pomp et al., 2021; Schubotz et al., 2012; Speer et al., 2003). Similarly, dough manipulation data showed BMAs to be active at participant-judged unit marks. This unit-mark-related increase was found for both dough and object-directed manipulations but the untouching-related increase was more prominent for dough-related actions. Therefore, the current approach extends our understanding of motion as playing a key role in event structure perception. Because activity in BMA at untouchings was particularly prominent when objects were weakly informative with regard to associated actions, one may speculate that infants' brains at an age when they do not yet have a mature knowledge of object–action associations can already segment actions into meaningful units based on movement information and may even begin to categorize object manipulation types using this structure (Wörgötter et al., 2013, 2020). A similar principle is used to allow robots to gain some kind of "action understanding." These machines are also, without programming them with additional knowledge, agnostic with respect to the action semantics of objects (Ziaeetabar et al., 2021), and (un-)touching sequences (SECs; Aksoy et al., 2011) can be used by them to recognize actions of humans with whom a robot has to cooperate.

Object manipulations that offered associated action options (and thus assumingly an informed predictive action model) showed the hypothesized increase in PHC activity at untouchings. PHC engagement is reliably seen in tasks where contextual associative information is encoded or retrieved from memory (Li et al., 2016; Aminoff et al., 2013) and is sensitive to the stochastic structure of observed events (Schiffer, Ahlheim, Wurm, & Schubotz, 2012; Turk-Browne, Scholl, Johnson, & Chun, 2010; Amso, Davidson, Johnson, Glover, & Casey, 2005). We take the stronger PHC engagement for object versus dough at untouching to reflect a stronger top–down signal of action prediction, as objects contained more information about possible upcoming actions than pieces of dough. This information about possible upcoming actions possibly provided a restriction on the matching process between the observed and the expected action based on object–action association knowledge. In the absolutely reduced scenery we used in our videos, which consisted only of the table surface, one or two objects, and the actress's upper body up to the shoulders (without head/face), contextual-associative information consisted solely in the combination of the respective object(s) and the manipulation performed on it.

Unexpectedly, aIPL activity did not increase for object versus dough videos, but on the contrary, dominated for dough compared with object videos when we modeled brain activity at untouching events. In our view, this result can only be interpreted if we also consider two other conditions in which the same area was also significantly activated: for object versus dough videos when we modeled the entire video length, and for the conjunction of both, object and dough videos in their full length. Thus, the aIPL was not specifically associated with the processing of only object-related information, and its engagement precisely increased at untouchings when weak object–action associations were available. Notably, in our study, untouching is the phase where updating of the current expectation occurs, as reflected by the engagement of frontal, parahippocampal, and insula regions (Pomp et al., 2021). Note that, although this finding was replicated in the present study (see Appendix), here we focus only on the specific modulations of these responses by the strength of object information. Updating expectations would normally mean that object information is used to select a restricted number of possible manipulations, which can be (or are typically) associated with the presented object. Thus, expectations could be restricted based on this kind of long-term memory, as reflected by the dominance of parahippocampal activity for modeling the BOLD response at untouching events for object versus dough videos. However, in the case of dough videos, this restriction was not provided by the piece of dough, and aIPL activity increase must be related to this unrestricted search for expectable manipulations. The aIPL is generally engaged in tasks highlighting object–hand interactions (Pelgrims et al., 2011; Vingerhoets, 2008). The activated cluster in the inferior parietal lobule that we observed included closely co-localized activation maxima in aSMG and aIPS, which have been assigned distinct but synergetic functions underlying the usage of tools. The aSMG was proposed to integrate semantic and technical information about objects, whereas aIPS rather selects the object-appropriate grasp based on object affordances (Bosch et al., 2023). Moreover, the aSMG may be particularly challenged by unfamiliar tools or conflicting alternative object-directed actions, whereas aIPS modulates this competition by structure-based and skilled use knowledge (Bosch et al., 2023; Buxbaum, 2017; Watson & Buxbaum, 2015). In a previous study, we found that activity in the aIPL varied as a function of the number of actions that participants associated with objects or object sets, even when these actions were not observed (Schubotz et al., 2014). Against this background, we suggest that aIPL was observed time-locked to untouchings when object–action associations were weak because of an unrestricted number of candidate actions in the case of dough manipulation videos, reflecting the matching of the beginning manipulation to the large repertoire of possible manipulations unrestricted by object–action associations. In line with this suggestion, Sacheli, Candidi, Era, and Aglioti (2015) demonstrated that the inhibition of aIPL selectively impaired participants' performance during complementary interactions and suggested aIPL to predictively code other people's actions. In addition, Benedek and colleagues (2018) reported that generating new object uses compared with the generation of known object uses was associated with increased left aIPL activation.

Limitations

Although we made every effort to achieve equal experimental conditions for both experiments' samples, we cannot completely rule out that some behavioral differences at the group level are an artifact. To avoid this limitation, future studies need to randomly assign participants to either group. Importantly, this limitation only concerns the comparison of behavioral data as fMRI analyses consisted only of interaction effects that rule out group effects. However, the overall similarity between the behavior of the two experimental groups concerning segmentation frequencies, intra-individual retest reliability, and group coherence as well as the systematic relationship to TU events gives us confidence in the authenticity of our behavioral results.

Furthermore, objects and dough pieces did not only differ with respect to the associated actions and there might be alternative interpretations of the differences in segmentation behavior. Future research is needed to address the degree of object–action associations in dependence on affordance, functional knowledge, object familiarity, and object complexity. It might be promising to parametrically vary the strength of object–action associations and other dimensions of relevance, and assess their impact on the corresponding segmentation behavior.

Conclusion

Having a life-long experience with manipulable objects provides individuals with a huge repertoire of object–action associations, which is used to efficiently predict object-directed actions. In the present study, modeling brain activity with objective and subjective event boundaries, we showed that object information had, indeed, significant effects on how the brain processes these events. In the absence of strong object–action associations, the increased impact of biological motion processing at objective untouching events, as well as the increased impact of contextual associative information when strong object–action associations were present, confirmed our hypotheses. At the same time, aIPL activity increased for weak object–action associations, presumably because of an unrestricted number of candidate actions. Furthermore, when objects were only weakly informative with regard to associated actions, segmentation behavior became even more systematic and tied to touching events. The present study confirms that objective relational changes in the form of touchings and untouchings of objects, hand, and ground represent meaningful anchor points in subjective action segmentation, rendering them objective marks of meaningful event boundaries. Our findings offer interesting insights into the neural segmentation of object-directed action and the significant influence objects have on the processing of different types of event boundaries because of their association with specific manipulations.

APPENDIX

Figure A1 shows the event-related main effects of touching and untouching events as well as of the participant-judged unit marks for object-directed and dough-directed actions. See Table A1 for the activation peaks of the dough manipulation study and Pomp and colleagues (2021) for the activation peaks of the object manipulation study. It gets obvious that object-directed action activation patterns are largely replicated by dough-directed activation, which means that event processing is mostly not modulated by object–action associations. One striking difference is the activation in aIPS/SMG, which was found for unit marks in object-directed actions and for untouching events in dough-directed actions. Direct whole-brain comparison of the contrasts though yielded no significant differences between action types at unit marks just as the

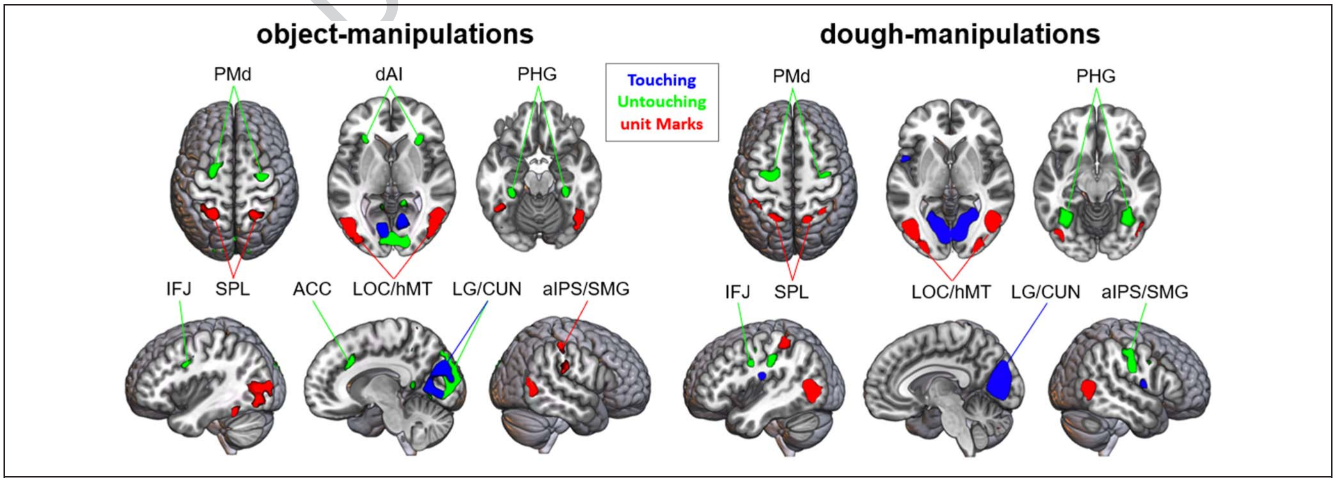

Figure A1. fMRI activation at $p < .005$, peak-level FDR-corrected, for the main contrasts of post-fMRI, participant-judged unit marks (M > nM, red), objective touching events (T > nTU, blue), and objective untouching events (U > nTU, green) of the object-directed action study (left), and the dough-directed action study (right). PMd = dorsal premotor cortex; dAI = dorsal anterior insula; PHG = parahippocampal gyrus; IFJ = inferior frontal junction; LG = lingual gyrus; CUN = cuneus; hMT = motion area; ACC = anterior cingulate cortex.

Table A1. Maxima of Activation from the Main Contrasts of the Second-level Whole-brain Analyses of the Dough Manipulation Study at $p < .005$ Peak-level FDR-Corrected

				t Value	*MNI Coordinates*		
Macroanatomical Location	*Abbreviation*	*H*	*Cluster Extent*		*x*	*y*	*z*
M > nM							
Posterior middle temporal gyrus	pMTG	R	381	10.04	48	−64	2
Inferior occipital gyrus	IOG	R		6.38	30	−97	−1
Lateral occipital cortex	LOC	L	371	9.46	−48	−73	−7
Inferior occipital gyrus	IOG	L		7.01	−27	−97	−1
Superior parietal lobule	SPL	R	101	6.33	33	−52	65
Anterior intraparietal sulcus	aIPS	L	178	6.01	−51	−34	56
Superior parietal lobule	SPL	L		5.74	−24	−58	68
Intraparietal sulcus	IPS	L		5.58	−45	−40	62
Cerebellum	CER	R	19	5.04	9	−73	−43
T > nTU							
Cuneus	CUN	L	1313	9.09	−12	−82	23
Lingual gyrus	LG	L		8.20	−12	−79	8
Calcarine gyrus	CG	L		7.57	−18	−73	14
Lingual gyrus	LG	R		7.47	15	−73	5
Calcarine gyrus	CG	R		7.41	15	−79	17
Cuneus	CUN	R		7.08	15	−79	26
Insula	IC	R	80	7.03	39	−16	23
Rolandic operculum	ROL	L	47	6.25	−42	−16	17
Rolandic operculum (lateral)	ROL	L	33	5.28	−57	5	5
		R	36	5.04	54	−1	8
U > nTU							
Postcentral gyrus / anterior intraparietal sulcus	PoG/aIPS	L	195	7.62	−63	−16	35
Anterior intraparietal sulcus	aIPS	L		4.99	−45	−22	38
		L		4.91	−42	−28	41
Postcentral gyrus	PoG	R	219	7.22	66	−10	29
Anterior intraparietal sulcus	aIPS	R		5.75	51	−16	44
		R		5.05	60	−16	44
		R		4.78	51	−25	44
Mid-insula	mIC	R	62	7.01	36	−4	20
Parahippocampal cortex	PHC	R	114	7.00	36	−58	−7
		L	127	6.66	−36	−55	−10
		L		5.46	−33	−43	−16

Table A1. (*continued*)

				t Value	*MNI Coordinates*		
Macroanatomical Location	*Abbreviation*	*H*	*Cluster Extent*		*x*	*y*	*z*
Cuneus	CUN	R	26	6.44	12	−94	29
Middle intraparietal sulcus	mIPS	L	80	6.13	−27	−43	50
Dorsal premotor cortex	PMd	L	156	6.13	−30	−13	50
Mid-insula	mIC	L	85	6.11	−36	−7	20
Inferior frontal junction	IFJ	L		5.72	−54	2	32
Dorsal premotor cortex	PMd	R	35	5.91	36	−10	56
Middle intraparietal sulcus	mIPS	R	49	5.34	27	−40	50
Posterior intraparietal sulcus	pIPS	L	24	5.02	−21	−73	35

H = hemisphere; L = left; R = right; M = unit mark; nM = non-unit mark; T = touching event; U = untouching event; nTU = non-(un-)touching event.

corresponding ROI analyses (see Results section). At untouching events, however, significant differences were found for aIPS/SMG as well as in other regions as discussed in the Discussion section.

Acknowledgments

The authors thank Monika Mertens and Yuyi Xu for their assistance during data collection. Furthermore, we thank Mina-Lilly Shibata and Simon Reich for their help during the creation of stimulus material, and Ima Trempler for advice regarding data analysis. Finally, we thank Rosari Naveena Selvan for valuable discussions and for proofreading the article.

Corresponding author: Jennifer Pomp, Department of Psychology, University of Münster, Germany, or via e-mail: jennifer.pomp@uni-muenster.de.

Data Availability Statement

The data of the behavioral analyses as well as of the ROI analyses of this article have been deposited in an OSF repository (DOI 10.17605/OSF.IO/MGQSF). Unthresholded statistical maps of all reported and visualized fMRI contrasts in the article have been deposited on NeuroVault (https://neurovault.org/collections/16065/). The entire stimulus material is available via the Action Video Corpus Muenster (AVICOM, https://www.uni-muenster.de/IVV5PSY/AvicomSrv/). The raw fMRI data and the raw SEC time point extraction data that support the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions

Jennifer Pomp: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Validation; Visualization; Writing—Original draft; Writing—Review & editing. Annika Garlichs: Investigation. Tomas Kulvicius: Formal analysis; Software; Visualization; Writing—Original draft. Minija Tamosiunaite: Conceptualization; Formal analysis; Methodology; Software. Moritz F. Wurm: Methodology; Writing—Review & editing. Anoushiravan Zahedi: Formal analysis; Writing—Review & editing. Florentin Wörgötter: Conceptualization; Funding acquisition; Methodology; Resources; Supervision; Writing—Review & editing. Ricarda I. Schubotz: Conceptualization; Funding acquisition; Methodology; Resources; Supervision; Visualization; Writing—Original draft; Writing—Review & editing.

Funding Information

This research was funded by the German Research Foundation (https://dx.doi.org/10.13039/501100001659; DFG), grant numbers: SCHU 1439/8–1 and WO 388/13–1.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this paper report its proportions of citations by gender category to be: M/M = .373; W/M = .275; M/W = .157; W/W = .196.

REFERENCES

Aguirre, G. K., Mattar, M. G., & Magis-Weinberg, L. (2011). De Bruijn cycles for neural decoding. *Neuroimage*, *56*, 1293–1300. https://doi.org/10.1016/j.neuroimage.2011.02.005, PubMed: 21315160

Aksoy, E. E., Abramov, A., Dörr, J., Ning, K., Dellen, B., & Wörgötter, F. (2011). Learning the semantics of object–action relations by observation. *International Journal of Robotics Research*, *30*, 1229–1249. https://doi.org/10.1177/0278364911410459

Aminoff, E. M., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in Cognitive Sciences*, *17*, 379–390. https://doi.org/10.1016/j.tics.2013.06.009, PubMed: 23850264

Amso, D., Davidson, M. C., Johnson, S. P., Glover, G., & Casey, B. J. (2005). Contributions of the hippocampus and the striatum to simple association and frequency-based learning. *Neuroimage*, *27*, 291–298. https://doi.org/10.1016/j.neuroimage.2005.02.035, PubMed: 16061152

Amunts, K., Mohlberg, H., Bludau, S., & Zilles, K. (2020). Julich-brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. *Science*, *369*, 988–992. https://doi.org/10.1126/science.abb4588, PubMed: 32732281

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, *72*, 708–717. https://doi.org/10.1111/1467-8624.00310, PubMed: 11405577

Bar, M., Aminoff, E., & Schacter, D. L. (2008). Scenes unseen: The parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *Journal of Neuroscience*, *28*, 8539–8544. https://doi.org/10.1523/JNEUROSCI.0987-08.2008, PubMed: 18716212

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Benedek, M., Schües, T., Beaty, R. E., Jauk, E., Koschutnig, K., Fink, A., et al. (2018). To create or to recall original ideas: Brain processes associated with the imagination of novel object uses. *Cortex*, *99*, 93–102. https://doi.org/10.1016/j.cortex.2017.10.024, PubMed: 29197665

Binkofski, F., & Buxbaum, L. J. (2013). Two action systems in the human brain. *Brain and Language*, *127*, 222–229. https://doi.org/10.1016/j.bandl.2012.07.007, PubMed: 22889467

Borghi, A. M. (2021). Affordances, context and sociality. *Synthese*, *199*, 12485–12515. https://doi.org/10.1007/s11229-018-02044-1

Bosch, T. J., Fercho, K. A., Hanna, R., Scholl, J. L., Rallis, A., & Baugh, L. A. (2023). Left anterior supramarginal gyrus activity during tool use action observation after extensive tool use training. *Experimental Brain Research*, *241*, 1959–1971. https://doi.org/10.1007/s00221-023-06646-1, PubMed: 37365345

Braun, D. A., Mehring, C., & Wolpert, D. M. (2010). Structure learning in action. *Behavioural Brain Research*, *206*, 157–165. https://doi.org/10.1016/j.bbr.2009.08.031, PubMed: 19720086

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox [abstract]. In *Paper Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002*. Sendai: Japan.

Buchsbaum, D., Griffiths, T. L., Plunkett, D., Gopnik, A., & Baldwin, D. (2015). Inferring action structure and causal relationships in continuous sequences of human action. *Cognitive Psychology*, *76*, 30–77. https://doi.org/10.1016/j.cogpsych.2014.10.001

Buxbaum, L. J. (2017). Distributed neurocognitive mechanisms. *Psychological Review*, *124*, 346–360. https://doi.org/10.1037/rev0000051, PubMed: 28358565

Caspers, S., Eickhoff, S. B., Geyer, S., Scheperjans, F., Mohlberg, H., Zilles, K., et al. (2008). The human inferior parietal lobule in stereotaxic space. *Brain Structure and Function*, *212*, 481–495. https://doi.org/10.1007/s00429-008-0195-z, PubMed: 18651173

Caspers, S., Geyer, S., Schleicher, A., Mohlberg, H., Amunts, K., & Zilles, K. (2006). The human inferior parietal cortex: Cytoarchitectonic parcellation and interindividual variability. *Neuroimage*, *33*, 430–448. https://doi.org/10.1016/j.neuroimage.2006.06.054, PubMed: 16949304

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage*, *25*, 1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034, PubMed: 15850749

El-Sourani, N., Trempler, I., Wurm, M. F., Fink, G. R., & Schubotz, R. I. (2019). Predictive impact of contextual objects during action observation: Evidence from functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, *32*, 326–337. https://doi.org/10.1162/jocn_a_01480, PubMed: 31617822

El-Sourani, N., Wurm, M. F., Trempler, I., Fink, G. R., & Schubotz, R. I. (2018). Making sense of objects lying around: How contextual objects shape brain activity during action observation. *Neuroimage*, *167*, 429–437. https://doi.org/10.1016/j.neuroimage.2017.11.047, PubMed: 29175612

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. -P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*, 189–210. https://doi.org/10.1002/hbm.460020402

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., et al. (2015). NeuroVault.Org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics*, *9*, 8. https://doi.org/10.3389/fninf.2015.00008, PubMed: 25914639

Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General*, *140*, 586–604. https://doi.org/10.1037/a0024310, PubMed: 21806308

Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. *Memory and Cognition*, *34*, 1221–1235. https://doi.org/10.3758/BF03193267

Hodgson, V. J., Lambon Ralph, M. A., & Jackson, R. L. (2023). The cross-domain functional organization of posterior lateral temporal cortex: Insights from ALE meta-analyses of 7 cognitive domains spanning 12,000 participants. *Cerebral Cortex*, *33*, 4990–5006. https://doi.org/10.1093/cercor/bhac394, PubMed: 36269034

Hrkać, M., Wurm, M. F., Kühn, A. B., & Schubotz, R. I. (2015). Objects mediate goal integration in ventrolateral prefrontal cortex during action observation. *PLoS One*, *10*, e0134316. https://doi.org/10.1371/journal.pone.0134316, PubMed: 26218102

Hunnius, S., & Bekkering, H. (2010). The early development of object knowledge: A study of infants' visual anticipations during action observation. *Developmental Psychology*, *46*, 446–454. https://doi.org/10.1037/a0016543, PubMed: 20210504

JASP Team. (2024). *JASP* (Version 0.18.3). [Computer software]. https://jasp-stats.org/

Kurby, C. A., & Zacks, J. M. (2018). Preserved neural event segmentation in healthy older adults. *Psychology and Aging*, *33*, 232–245. https://doi.org/10.1037/pag0000226, PubMed: 29446971

Li, M., Lu, S., & Zhong, N. (2016). The parahippocampal cortex mediates contextual associative memory: Evidence from an fMRI study. *BioMed Research International*, *2016*, 9860604. https://doi.org/10.1155/2016/9860604, PubMed: 27247946

Mahon, B. Z. (2023). Higher order visual object representations: A functional analysis of their role in perception and action. In G. G. Brown, B. Crosson, K. Y. Haaland, & T. Z. King (Eds.), *APA handbook of neuropsychology: Neuroscience and neuromethods* (pp. 113–138). American Psychological Association. https://doi.org/10.1037/0000308-006

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*, 28–38. https://doi.org/10.1037/h0035584

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, *12*, 436–450. https://doi.org/10.1016/0022-1031(76)90076-7

Newtson, D., Engquist, G. A., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology*, *35*, 847–862. https://doi.org/10.1037/0022-3514.35.12.847

Newtson, D., Hairfield, J., Bloomingdale, J., & Cutino, S. (1987). The structure of action and interaction. *Social Cognition*, *5*, 191–238. https://doi.org/10.1521/soco.1987.5.3.191

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, *25*, 653–660. https://doi.org/10.1016/j.neuroimage.2004.12.005, PubMed: 15808966

O'Neal, C. M., Ahsan, S. A., Dadario, N. B., Fonseka, R. D., Young, I. M., Parker, A., et al. (2021). A connectivity model of the anatomic substrates underlying ideomotor apraxia: A meta-analysis of functional neuroimaging studies. *Clinical Neurology and Neurosurgery*, *207*, 106765. https://doi.org/10.1016/j.clineuro.2021.106765, PubMed: 34237682

Pelgrims, B., Olivier, E., & Andres, M. (2011). Dissociation between manipulation and conceptual knowledge of object use in the supramarginalis gyrus. *Human Brain Mapping*, *32*, 1802–1810. https://doi.org/10.1002/hbm.21149, PubMed: 21140435

Pomp, J., Heins, N., Trempler, I., Kulvicius, T., Tamosiunaite, M., Mecklenbrauck, F., et al. (2021). Touching events predict human action segmentation in brain and behavior. *Neuroimage*, *243*, 118534. https://doi.org/10.1016/j.neuroimage.2021.118534, PubMed: 34469813

R Core Team. (2022). *R: A language and environment for statistical computing* (Version 2022.07.1). [Computer software] R Foundation for Statistical Computing. https://www.r-project.org/

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, *16*, 225–237. https://doi.org/10.3758/PBR.16.2.225

Sacheli, L. M., Candidi, M., Era, V., & Aglioti, S. M. (2015). Causative role of left aIPS in coding shared goals during human–avatar complementary joint actions. *Nature Communications*, *6*, 7544. https://doi.org/10.1038/ncomms8544, PubMed: 26154706

Sargent, J. Q., Zacks, J. M., & Bailey, H. R. (2015). Perceptual segmentation of natural events: Theory, methods, and applications. In R. R. Hoffman, P. A. Hancock, M. W. Scerbo, R. Parasuraman, & J. L. Szalma (Eds.), *The Cambridge handbook of applied perception research* (Vol. 1, pp. 443–465). Cambridge University Press. https://doi.org/10.1017/CBO9780511973017.029

Schiffer, A. M., Ahlheim, C., Wurm, M. F., & Schubotz, R. I. (2012). Surprised at all the entropy: Hippocampal, caudate and midbrain contributions to learning from prediction errors. *PLoS One*, *7*, e36445. https://doi.org/10.1371/journal.pone.0036445, PubMed: 22570715

Schubotz, R. I., Korb, F. M., Schiffer, A. M., Stadler, W., & von Cramon, D. Y. (2012). The fraction of an action is more than a movement: Neural signatures of event segmentation in fMRI. *Neuroimage*, *61*, 1195–1205. https://doi.org/10.1016/j.neuroimage.2012.04.008, PubMed: 22521252

Schubotz, R. I., Wurm, M. F., Wittmann, M. K., & von Cramon, D. Y. (2014). Objects tell us what action we can expect: Dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. *Frontiers in Psychology*, *5*, 636. https://doi.org/10.3389/fpsyg.2014.00636, PubMed: 25009519

Speer, N. K., Swallow, K. M., & Zacks, J. M. (2003). Activation of human motion processing areas during event perception. *Cognitive, Affective, & Behavioral Neuroscience*, *3*, 335–345. https://doi.org/10.3758/CABN.3.4.335, PubMed: 15040553

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*, *30*, 11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010, PubMed: 20720125

Vingerhoets, G. (2008). Knowing about tools: Neural correlates of tool familiarity and experience. *Neuroimage*, *40*, 1380–1391. https://doi.org/10.1016/j.neuroimage.2007.12.058, PubMed: 18280753

Ward, E. J., Chun, M. M., & Kuhl, B. A. (2013). Repetition suppression and multi-voxel pattern similarity differentially track implicit and explicit visual memory. *Journal of Neuroscience*, *33*, 14749–14757. https://doi.org/10.1523/JNEUROSCI.4889-12.2013, PubMed: 24027275

Watson, C. E., & Buxbaum, L. J. (2015). A distributed network critical for selecting among tool-directed actions. *Cortex*, *65*, 65–82. https://doi.org/10.1016/j.cortex.2015.01.007, PubMed: 25681649

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (Version 3.4.0). [Computer software] New York: Springer-Verlag. https://ggplot2.tidyverse.org

Wörgötter, F., Aksoy, E. E., Krüger, N., Piater, J., Ude, A., & Tamosiunaite, M. (2013). A simple ontology of manipulation actions based on hand–object relations. *IEEE Transactions on Autonomous Mental Development*, *5*, 117–134. https://doi.org/10.1109/TAMD.2012.2232291

Wörgötter, F., Ziaeetabar, F., Pfeiffer, S., Kaya, O., Kulvicius, T., & Tamosiunaite, M. (2020). Humans predict action using grammar-like structures. *Scientific Reports*, *10*, 3999. https://doi.org/10.1038/s41598-020-60923-5, PubMed: 32132602

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited—Again. *Neuroimage*, *2*, 173–181. https://doi.org/10.1006/nimg.1995.1023, PubMed: 9343600

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., et al. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*, 651–655. https://doi.org/10.1038/88486, PubMed: 17576286

Zacks, J. M., Kumar, S., Abrams, R. A., & Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition*, *112*, 201–216. https://doi.org/10.1016/j.cognition.2009.03.007

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind–brain perspective. *Psychological Bulletin*, *133*, 273–293. https://doi.org/10.1037/0033-2909.133.2.273, PubMed: 17338600

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, *16*, 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480.x, PubMed: 22468032

Zhao, L. (2019). The role of the action context in object affordance. *Psychological Research*, *83*, 227–234. https://doi.org/10.1007/s00426-018-1002-y

Ziaeetabar, F., Pomp, J., Pfeiffer, S., El-Sourani, N., Schubotz, R. I., Tamosiunaite, M., et al. (2021). Using enriched semantic event chains to model human action prediction based on (minimal) spatial information. *PLoS One*, *15*, e0243829. https://doi.org/10.1371/journal.pone.0243829, PubMed: 33370343