Memory & Cognition
https://doi.org/10.3758/s13421-024-01525-8

People can reliably detect action changes and goal changes during naturalistic perception

Xing Su[1] · Khena M. Swallow[2]

Accepted: 17 January 2024

Abstract

As a part of ongoing perception, the human cognitive system segments others' activities into discrete episodes (*event segmentation*). Although prior research has shown that this process is likely related to changes in an actor's actions and goals, it has not yet been determined whether untrained observers can reliably identify action and goal changes as naturalistic activities unfold, or whether the changes they identify are tied to visual features of the activity (e.g., the beginnings and ends of object interactions). This study addressed these questions by examining untrained participants' identification of action changes, goal changes, and event boundaries while watching videos of everyday activities that were presented in both first-person and third-person perspectives. We found that untrained observers can identify goal changes and action changes consistently, and these changes are not explained by visual change and the onsets or offsets of contact with objects. Moreover, the action and goal changes identified by untrained observers were associated with event boundaries, even after accounting for objective visual features of the videos. These findings suggest that people can identify action and goal changes consistently and with high agreement, that they do so by using sensory information flexibly, and that the action and goal changes they identify may contribute to event segmentation.

Keywords Event segmentation · Goal processing · Action perception · Perspective · Event cognition

Introduction

The human cognitive system divides ongoing experience into discrete events as it occurs (*event segmentation*; Richmond & Zacks, 2017). For example, while watching someone else prepare breakfast, one might divide their activity into collecting all the needed items (cereal, milk, and a bowl and spoon), pouring cereal into the bowl, and pouring milk into the bowl. Empirically, changes in both perceptual information, such as object movements or an actor's body position, and higher-level content, such as an actor's goals, are associated with event segmentation (Baldwin et al., 2001; Hard et al., 2011; Kopatich et al., 2019; Newtson et al., 1977; Zacks, 2004). However, the contributions of content changes to event segmentation, independent of low-level perceptual features, have been difficult to isolate because these two sources of information are correlated (Baldwin et al., 2001; Cutting, 2014; Swallow et al., 2018). For example, when an actor makes coffee, changes in motion (the actor's arm is accelerating), the spatial relationship between the actor and the object (the actor gets closer to the grinder), and the actor's posture (the actor leans forward to pick up the grinder) could signal a change in the actor's goal (to grind coffee beans). Compounding the issue, for actions and goal changes to contribute to segmentation, people must be able to detect them as a part of ongoing, naturalistic perception (i.e., without explicit training). This issue has not, to our knowledge, previously been addressed. It is therefore unclear whether changes in more abstract, or inferred, content information like goals contribute to event segmentation beyond their association with perceptual change. This study addressed these questions by examining whether untrained participants can consistently identify action and goal changes in ongoing, naturalistic activities, the perceptual features that contribute to their identification, and how

✉ Khena M. Swallow
kms424@cornell.edu

[1] Department of Psychological and Brain Sciences, Washington University in Saint Louis, Saint Louis, MO, USA

[2] Department of Psychology and Cognitive Science Program, Cornell University, 211 Uris Hall, Ithaca, NY 14853, USA

Published online: 05 February 2024

the action and goal changes identified by untrained observers relate to event segmentation.

Event segmentation may reflect predictive processing

Just as activities are hierarchically organized by actions and goals (Vallacher & Wegner, 1989), event segmentation organizes the stream of continuous experience into hierarchical part-whole structures (Zacks et al., 2010, Zacks, Braver et al., 2001). This process is usually measured by instructing participants to press a button to mark the *boundaries* between events as they watch a video or read a narrative text (Newtson et al., 1977; Zacks et al., 2009). Despite being given minimal guidance on how to define events, observers tend to press the button at similar moments in the video, indicating high within- and across-observer agreement on when the boundaries between events in a specific video occur (Newtson et al., 1977; Sasmita & Swallow, 2022; Speer et al., 2003). Further, when asked to identify events of different grains, people identify short, *fine* grained, events (often lasting seconds) that are nested within longer, *coarse* grained events (often lasting dozens of seconds to minutes; Newtson et al., 1977; Vallacher & Wegner, 1989; Zacks, Braver et al., 2001, Zacks, Tversky et al., 2001). These behavioral findings, coupled with converging data indicating that neural processing changes around event boundaries even during naturalistic (task-free) viewing (Baldassano et al., 2017; Zacks et al., 2010, Zacks, Braver et al., 2001), suggest that event segmentation is an integral part of normal perception.

According to Event Segmentation Theory (EST; Kurby & Zacks, 2008; Zacks et al., 2007, Zacks, Braver et al., 2001, Zacks, Tversky et al. 2001), this mechanism reflects both knowledge about how events typically unfold and perceptual information about the current event. EST proposes that people predict upcoming perceptual input based on a model of the current event (*event model*). Event models reflect the combination of abstract semantic knowledge about events and activities (e.g., scripts, goals, and the hierarchical relationship between actions and goals) and information about the current state of the environment (e.g., the relative locations of objects and people). When an event model fails to sufficiently predict perceptual input, an event boundary is perceived, and a new event model is constructed. In support of this account, several studies have found that the flow of information within events is more predictable than it is between events (Baldwin et al., 2008; Zacks et al., 2011), and that event boundaries are associated with transitions in the information that is actively maintained in memory (Ezzyat & Davachi, 2011; Kurby & Zacks, 2022; Speer & Zacks, 2005; Swallow et al., 2009).

Changes in what an observer infers about the content of an event, particularly an actor's actions and goals, thus could play a central role in event segmentation. In the context of movie watching, we define *actions* to be intentional movements of an actor that result in a change to an object, the environment, or the actor's location (thus excluding movements caused by something else, such as an object contacting the actor, or tangential movements that failed to effect some change in the environment). We define a *goal* as the desired outcome of an action or a sequence of actions (for clarity, we distinguish goals from *intentions*, which may refer to more abstract levels of description, including the motivation to perform a behavior, e.g., one may drink coffee to regain energy). The goal construct thus captures aspects of the actor's unobservable mental state that organize their observable actions (Newtson et al., 1977). From the observer's perspective, correctly inferring an actor's goals creates an additional source of knowledge that could be used to form better predictions about the sequence of actions they are likely to observe (Bach & Schenke, 2017; El-Sourani et al., 2018), and when those actions are likely to start and stop. Better predictions about actions may in turn lead to better predictions about the low-level perceptual features of the activity such as motion generated from biological movements and interactions with objects.

Indeed, a large body of work associations event segmentation with changes in an actor's actions and goals (Baldwin et al., 2001; Levine et al., 2017; Magliano et al., 2005). Fine and coarse event boundaries often coincide with action or goal changes that have been identified by experts or experimenters (Swallow & Wang, 2020; Zacks & Tversky, 2001), and this may be especially true when the observers are also experts (Bläsing, 2015; Levine et al., 2017; Newberry et al., 2021). Similar findings have been observed for narrative text and picture sequences (Kopatich et al., 2019; Magliano et al., 2005; Zacks et al., 2009). Moreover, during task-free viewing, fine and coarse boundaries align with changes in the activity and representational content of brain regions associated with the formation of explicit inferences about others' mental states and goals (Baldassano et al., 2017; Barrett & Satpute, 2013; Wurm & Lingnau, 2015; Zacks et al., 2010).

These findings are correlational but provide convergent evidence of a role for action and goal changes in event segmentation. Yet the claim that event segmentation is driven, in part, by action and goal changes relies on at least three assumptions that need to be tested: (1) that untrained observers are reliably sensitive to action and goal changes during naturalistic perception of another person's activities, (2) that the action and goal changes they identify are not reducible to changes in lower level, more concrete information, and (3) if both of the former are true, that action and goal changes identified by untrained observers are associated with event segmentation above and beyond their association with other, more concrete visual changes.

Action identification and goal inferences during naturalistic perception

People are adept at inferring the mental states that underlie other agents' actions, including their goals, beliefs, desires, and emotions (Baker et al., 2009; Blakemore & Decety, 2001). Sensitivity to the presumed targets of another person's actions, such as their expected interaction with an object (sometimes referred to as the goal of an action, e.g., grabbing a cup that is being reached for), are evident in brain activity (Ziaeetabar et al., 2020), observer's estimates of where people are reaching (Hudson et al., 2016), and in people as young as 12 months of age (Olofson & Baldwin, 2011; Woodward & Sommerville, 2000). There is also evidence that observers can rapidly access goal information, inferring action types and actor roles from pictures viewed for as little as 37 ms (Decroix et al., 2020; Hafri et al., 2013). The extensive literature on people's ability to infer or describe an actor's goals is consistent with the possibility that observers can detect changes in them as they happen.

However, there are challenges in applying this research to event segmentation in everyday perception. One concern is that goals are often investigated as the immediate target or outcome of a single, simple action with well-defined start and end points (e.g., Decroix et al., 2020; Hudson et al., 2016; Olofson & Baldwin, 2011). Yet, goals in everyday activities are often achieved by sequences of movements with fluid (but not always smooth) transitions between actions. Further, event segmentation occurs online, in response to changes in perceptually rich, dynamic experiences of naturalistic situations (Zacks et al., 2007). If action and goal changes are to contribute to such a process, then observers must be sensitive to *changes* in an actor's actions and goals, not just be able to identify the actions and goals themselves. They must further be sensitive to action and goal changes as they occur, in ongoing naturalistic activities and without explicit training.

Yet, much of the work on action and goal identification evaluates this process offline, after the goal or action has ended (Barsalou, 2008; Gallese & Goldman, 1998; Vallacher & Wegner, 1989) or as part of an explicit task that requires participants to label or evaluate the actor's goals as the activity is viewed (Decroix et al., 2020; Spunt et al., 2010). While there is evidence that neural systems involved in action understanding are engaged online (e.g., Mukamel et al., 2010; Spunt et al., 2010) and are active around event boundaries (Zacks et al., 2010), it is not yet clear to what extent these systems are involved in goal understanding (Heyes & Catmur, 2022). Furthermore, identifying changes in an actor's goals might involve cognitive processes that may be inconsistently engaged across observers (Catmur, 2015; Koul et al., 2016; Naish et al., 2013).

Another concern is that untrained observers may identify the start or end of a goal at different times. For example, when watching someone prepare breakfast, some observers might detect a goal change when the actor begins to walk to the box of cereal, while others may detect it when the actor touches the box of cereal. Both observers are tracking the actor's goal and would likely identify it as "preparing breakfast," but they would disagree about when the goal started, and therefore when a goal change occurred. Consistent with this possibility, even when given the ability to review footage, the time windows within which different experts identify the start or end of a sequence of goal directed activity can last several seconds (Levine et al., 2017), which is on the longer end of the range of window sizes that are typically used to evaluate online event segmentation (cf. Sasmita & Swallow, 2022). Though similar concerns about actions are less pronounced, whether observers identify the starts and ends of actions (e.g., reaches) at similar times during the viewing of other's naturalistic activities is unknown. Because prior research has investigated the role of action and goal changes in event segmentation using materials in which these factors were either manipulated with text (Kopatich et al., 2019) or that were coded by trained researchers or experts (Levine et al., 2017; Magliano et al., 2005; Swallow & Wang, 2020; Zacks et al., 2010), it remains an open question whether action and goal changes identified by untrained observers correlate with event boundaries.

Finally, even if untrained observers identify action and goal changes consistently and as part of naturalistic perception, how these changes influence event segmentation independent of their visual correlates is unclear. Theories of action understanding often focus on information that is derived from visual input, but that does not closely correspond to specific, objective visual features. For example, the simulation theory of action understanding proposes that mirror-neuron areas support a direct match between observed actions and a motor simulation of the same action (Gallese & Goldman, 1998). Similarly, Baker and colleagues (Baker et al., 2009) suggest that observers rationally infer an actor's goals by applying their knowledge of the current situation and of how goals structure activities. Other accounts consider how object affordances and typical uses guide predictions about an actor's likely goals and behaviors (Bach et al., 2014). These accounts acknowledge the role of contextual and perceptual information in specifying an actor's actions and goals but are focused more on identification than change detection and the cues that may facilitate it. They rarely directly consider the role of *changes* in visual motion, body posture, and contact between an actor and an object in action and goal change detection. This likely reflects the reasonable assumption that objective visual features, actions, and goals have a many-to-many relationship: many low-level visual features correspond to many actions, which in turn correspond to many goals, and vice versa. However, given the relationship between changes in objective visual features and event segmentation (e.g., Cutting, 2014; Newtson et al., 1977; Swallow et al., 2018; Zacks, 2004), it is prudent to evaluate whether low-level perceptual changes are sufficient to account for the relationship between event segmentation, actions, and goals. We explore this issue in more detail in the next section.

Invariance and event segmentation

Changes in actions and goals are likely to be coupled with changes in objective level visual features in naturalistic perception (Baldwin et al., 2001; Cutting, 2014; Smith & Anderson, 2004; Wurm & Lingnau, 2015). Such a possibility raises the question of whether observers rely on inferences about an actor's goals or mental state to guide event segmentation, or instead use changes in low-level perceptual information (see Heyes & Catmur, 2022, for related concerns about mentalizing). Indeed, people appear to identify similar event boundaries when videos are played forwards or backward (which could influence action and goal perception; Hard et al., 2006) or regardless of whether they are given additional information about an actor's goals when viewing conceptually impoverished stimuli (Zacks et al., 2009). The results cast doubt on the centrality of action and goal identification for event segmentation, instead highlighting potential contributions from spatio-temporal discontinuities.

However, event segmentation may be best understood as being driven by features that are dynamically smooth, noise tolerant, and invariant across viewpoints or viewing conditions, like actions and goals (Richmond & Zacks, 2017). In one study (Swallow et al., 2018), participants viewed and segmented a set of activities that were simultaneously recorded from both a first-person perspective (with a head-mounted camera) and a third-person perspective (with a tripod-mounted camera several feet away). Changing the perspective from which an activity is viewed changes objective visual features with low tolerance to variable perspectives, such as motion generated by movements of the body, hands, and head, that have been previously associated with event segmentation (Zacks et al., 2009). However, low-level differences in the first- and third-person perspectives videos did not result in consistent changes in when observers segmented the activities. It is therefore possible that viewpoint dependent features of videos, like motion, are associated with event segmentation because they are also associated with viewpoint independent features of the activity itself, including, potentially, action changes and goal changes.

The study reported in Swallow et al. (2018) does not address the possibility that online action and goal processing are themselves reducible to objective visual features and would therefore be influenced by the viewer's perspective. Additionally, previous research shows that first-person perspectives lead to more embodied, concrete processing focused on how an action was performed, and third-person perspectives lead to more abstract, goal-oriented processing focused on the action's purpose (Libby et al., 2009). As a result, the observer's viewpoint could change how observers identify actions and goals, as well as changes in those aspects of an actor's activity. It could also influence their contributions to event segmentation.

The current study

The current study addresses several questions about the role of action and goal changes in event segmentation. First, it tests whether untrained observers agree with each other about the moment that an actor's action or goal has changed. Participants marked action changes, goal changes, or event boundaries as they watched videos of everyday activities. The data were analyzed to characterize the degree to which participants pressed the button at similar times within and across tasks using methods commonly employed in the event segmentation literature (cf. Sasmita & Swallow, 2022). Given the apparent ease with which participants can label goals (e.g., Decroix et al., 2020; Olofson & Baldwin, 2011; Woodward & Sommerville, 2000), we hypothesized that people can reliably track action and goal changes as they observe other's activities, and that they do so at a level of precision that would support event segmentation.

Second, this study examines the relationship between action and goal changes and objective visual features of an activity. It does so by asking whether action and goal changes are identified at similar times when an activity is presented from different perspectives, and by directly estimating the relationship between action and goal changes and video features. We expected action and goal changes to be associated with changes in observable visual features. However, we hypothesized that changes in actions and, especially, goals (which should be more abstract), are at least somewhat invariant across perspectives.

Finally, this study examines whether untrained observers identify action and goal changes that are associated with event segmentation, even when accounting for visual feature changes in the videos. Because event segmentation may be driven by noise-tolerant features of an experience (Richmond & Zacks, 2017; Swallow et al., 2018), we hypothesized that the relationship between event boundaries and action and goal changes is not reducible to their shared association with visual change.

If these hypotheses are affirmed, this study would support the assumptions that the human cognitive system tracks changes in an actor's actions and goals during normal, everyday perception. They would further indicate that these changes independently contribute to the division of experience into meaningful events.

Method

Participants

All volunteers were recruited from Cornell University's undergraduate and graduate population. The Cornell Institutional Review Board approved all methods and procedures, which were in accord with the standards set forth by the Declaration of Helsinki. All participants provided informed consent and were compensated with course credits.

A target sample size of 70 for each group (for a total of 140 participants) was decided before data collection to ensure adequate power to detect small effects of perspective on task performance (see Swallow et al., 2018). Based on a power analysis with G∗Power (Faul et al., 2007), an experiment with N = 140, $\alpha = .05$, and power ($1-\beta$) = .95 has the sensitivity to detect an effect of $f \geq .178$ of a within-between group interaction in a 2 × 4 analysis of variance, with perspective as a two-level within-participants factor and task as a four-level between-participants factor (see Procedure and design).

Data from five participants were excluded because they either did not complete the task (n = 2), were observed to be using an electronic device during the experiment (n = 2), or held the space bar down for many seconds at a time resulting in more than 93% of their recorded button presses being within 500 ms of an earlier one (n = 1). Of the remaining 135 participants, 67 (39 female, 28 male: age M = 19.46 years, SD = 1.33) completed the event segmentation task second, after they completed the action change-detection task (n = 33) or the goal change-detection task (n = 34), and 68 (45 female, 23 male, age M = 20.12 years, SD = 1.44) completed the event segmentation task first, before the action change-detection task (n = 33) or the goal change-detection task (n = 35). Data from the group that performed the change-detection task first were collected before data from the other group.

Equipment

Data were acquired with a Dell PC, with a standard keyboard and mouse, and a 24 in. LCD display (1,024 × 760-pixel resolution, 144-Hz refresh rate). The tasks were programmed and run in MATLAB (Mathworks, Inc) using Psychtoolbox (Brainard, 1997). Testing was performed in an interior, normally lit room. Participants sat approximately 50 cm from the display and were free to move around during breaks.

Materials

Two activities were selected from a set of four that were used in a prior study (Swallow et al., 2018) because of their relatively brief duration. This allowed participants to view the activities several times during a single experimental session. The first activity involved a female actor organizing a desk and bookcase (*office*, 212 s long). The second activity involved a male actor washing laundry (*wash*, 231 s long).

Both activities were simultaneously recorded from two vantage points to produce two videos of the same activity from different perspectives (recorded with 1,910 × 1,080 pixel resolution; 29.098 fps). For the *third-person perspective* video, a camera (GoPro Hero 4, Silver Edition) was positioned on a stationary tripod. The camera was positioned at about the eye level of a typical adult, 5–6 feet above the floor. The camera was placed as closely as possible to the action while ensuring that the actor and objects that the actor interacted with were visible throughout the video. For the *first-person perspective* video, the actor wore a camera (GoPro Hero 3+. Black Edition) on his or her forehead using an elastic head strap. The camera was positioned to capture the region of space directly in front of the actor, including the space in which they were interacting with objects. The head-mounted camera was visible in the third-person videos. The third-person camera was rarely visible in the first-person videos. Prior to the recording the actors were given a rough script of what they should do, with which objects, and in what order. For example, in the office video, the actor was asked to dust the bookcase before writing a note. Actors were asked to ensure that their actions would be visible in both videos.

A video of a third activity depicting a man relaxing outside, reading, and using his phone recorded in third-person perspective was used as a practice video (99 s long).

Procedure and design

Participants were asked to perform tasks as they watched videos of a single actor engaged in a common, everyday activity. The video was presented at the center of the display (75% of the horizontal dimension of the screen; aspect ratio was preserved) over a black background. Each video was preceded by a pre-stimulus fixation period during which a red circle appeared in the center of a black background for 1,000 ms. The video then appeared on the screen and played continuously until it was over. After the video ended, participants were prompted to take a break and press the space bar when they were ready to start the next video.

Three tasks were used in this experiment: *action change detection*, *goal change detection*, and *event segmentation*. For all tasks, participants were instructed to press the space bar on the keyboard to mark the moments when they believed a particular kind of change occurred. They were told to keep their preferred hand near the space bar throughout the videos. The tasks differed in the types of changes participants marked and in their instructions.

For the action change-detection task, the participants were instructed to press the space bar whenever they believed that the action the actor was performing changed. Participants were provided with a specific definition of actions and were given a concrete example. The exact instruction was "For this experiment you will see several movies of everyday activities. As you watch the movies, we would like you to tell us whenever the actor is performing a new action. Actions are intentional movements that change an object, the environment, or the actor's location. For example, when unlocking a door, putting the key in the lock is one action, and turning the key is another. For this task, you would press the space bar at the beginning of the turning action to mark the beginning of the new action. For this task you will need to mark the boundaries between actions by pressing the space bar. You should press the space bar every time you believe the action has changed."

For the goal change-detection task, participants were told to press the space bar whenever they believed that the actor's goal changed. As with the action change-detection task, participants were provided with a specific definition of goals and a concrete example. The exact instruction was "For this experiment you will see several movies of everyday activities. As you watch the movies, we would like you to tell us whenever the actor's goal has changed. Goals are the reasons that an action is performed. Goals can vary in how abstract they are, but for this task, we would like you to identify goals that capture sequences of behavior. For example, when a person walks up to their door, unlocks it, and pushes it open, the goal is to enter their home. For this task you will need to mark the boundaries between goals by pressing the space bar. You should press the space bar every time you believe the actor's goal has changed."

For the event segmentation task, participants were instructed to mark the boundaries between events by pressing the space bar whenever they believed that one event had ended, and another event had begun. Unlike the action and goal change-detection tasks, participants were not provided with a precise definition of events or concrete examples. The instructions were also *neutral* with respect to the size, or grain, of the events that participants were asked to identify. This makes the instructions consistent with earlier approaches. The exact instruction for the segmentation task was "For this experiment, you will see several movies of everyday activities. As you watch those movies, we would like you to press the space bar on the keyboard when one event ends and another begins. We will not provide you with any definition of event. It's entirely up to you to define it."

Following instruction, participants practiced the task while watching the practice video. Participants were only provided with feedback for the segmentation task and were asked to repeat the practice run if they identified less than three event boundaries or more than 16 event boundaries for the practice movie (corresponding to 1.8–9.7 button presses per minute). Once the practice trial was complete, each participant performed the instructed task while watching four videos depicting each activity from each perspective. The order of activity and perspective was counterbalanced across all participants, such that both activities were viewed from one perspective before they were watched again from the other perspective (Fig. 1a). Once participants had completed their first task on all four videos, they were offered a break before the process was repeated for the second task.

This procedure resulted in four groups that differed in whether participants performed the action change-detection task or the goal change-detection task and in the order in which they performed the change detection and segmentation tasks (i.e., each task was performed first or second; Fig. 1). For analysis, the event segmentation task was treated differently for those participants who performed the action change-detection task versus those who performed the goal change-detection task. This simplified the design and ensured comparable numbers of participants within each task condition. It also provided a baseline comparison for evaluating agreement between two groups when instructions and conditions are equal (i.e., those who performed segmentation before the action change-detection task were performing the same task under the same conditions as those who performed segmentation task before the goal change-detection task). Within each group, all participants viewed first- and third-person perspective versions of each activity, with the order of perspectives counterbalanced across participants. All participants also viewed the book and wash activities, with their order counterbalanced across participants. This procedure thus resulted in a 4 (task: Action, Goal, SegA, SegG) × 2 (task number: first vs. second) × 2 (perspective: first- vs. third- person perspective) × 2 (perspective order: first-person first vs. third-person first) × 2 (activity: book vs. wash) × 2 (activity order: book first vs. wash first) design, where perspective, activity, and task number were manipulated within participants.

Analysis

Video coding As in a previous study (Swallow et al., 2018), we examined the relationship between button presses and objective visual features of the videos. For each second of the video, we used the *visual activity index (VAI)*, a measure

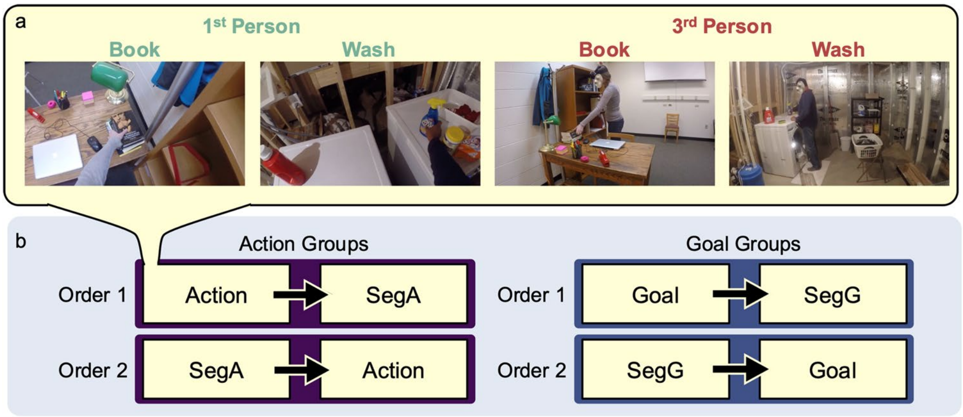

Fig. 1 Illustration of study design. (**a**) Trial order. For each task, participants viewed each of two activities from the first- and third-person perspective for a total of four videos per task. Both activities were viewed from one perspective before switching to the other. In this example, a participant viewed the book and then the wash activity from the first-person perspective, followed by the third-person perspective. Perspective order and activity order were counterbalanced across participants. (**b**) Task orders. Participants were assigned to four groups based on whether they performed the action or goal change-detection task, and whether they performed that task before or after the segmentation task (labeled as SegA for the action group or SegG for the goal group, though the tasks were identical). The order of the tasks was counterbalanced across groups. Trial order was maintained within participants across the two tasks

of objective pixel to pixel changes over time (Cutting, 2014), *touch onset*, a measure of whether the actor began an object touch, and *touch offset*, a measure of whether the actor ended an object touch, codes from that prior study. The VAI was z-scored per video by subtracting the within video mean VAI and dividing by the within video SD before use in analyses.

Button press data processing Minimal steps were taken to prepare the times that participants pressed the button in each video viewing (trial) for analysis. To exclude button presses that reflect a failure to release the button quickly enough, any button press that occurred within 500 ms of the prior button press was excluded from analyses (% excluded button presses, range across participants = 0-63%, M = 19.5%, SD = 17.6). The remaining button presses were then used to calculate several metrics that characterize task performance.

Button presses per minute For each viewing of a video the number of button presses was divided by the duration of the video in minutes.

Hierarchical organization The degree to which goals consist of actions was evaluated by adapting metrics of hierarchical event segmentation described in previous studies. Because participants performed either the action task or the goal task, but not both, we compared the button presses of individuals identifying the larger grained units (goals) to the normative changes identified by the group identifying the smaller grained units (actions). Normative action changes were obtained using a procedure described in prior work (Sasmita & Swallow, 2022). In brief, the *group density time series* was created by pooling the button presses for all participants that watched a video in the same condition, and then using a Gaussian smoothing kernel to estimate the density of button presses over time. The bandwidth of the smoothing kernel was determined using the Sheather-Jones algorithm and an adjustment factor of .04 to produce a time series with clear peaks and valleys in all conditions (see Online Supplementary Material (OSM)). Normative action changes were then defined as the times of the *j* highest peaks in the group density time series, where *j* was the mean number of times participants pressed the button in that condition. The observed distance was then defined as the mean temporal distance between participants' button presses when performing the goal task and the nearest normative action change. If goals consist of actions, then participants should identify goal changes that are nearer to normative action boundaries than expected by chance. Therefore, to obtain a measure of *alignment* between action and goal changes, the observed distances were subtracted from the expected distances if the same number of button presses were randomly distributed in the video (Zacks & Tversky, 2001). If goals consist of actions, then button presses in the goal task may be more likely to follow, rather than precede, the nearest normative action boundary (Hard et al., 2011). The *enclosure* score captures this expectation by computing the proportion of button presses in the goal task that followed the nearest normative action change.

Agreement metrics Some analyses relied on metrics capturing button press agreement. To capture agreement *within groups*, we used the *peakiness* metric described by Sasmita and Swallow (2022). This metric quantifies the amount of agreement within a group without having to compare the group to another group or individual. First, the rugosity (a measure of variability) of the observed group density time series is calculated for each condition. Peakiness was then defined as the ratio of that observed value to the rugosity of a group density time series that consisted of the same number of evenly spaced button presses. We evaluated whether the observed peakiness value was greater than expected by chance by comparing it to a bootstrapped distribution of peakiness values (N = 1000) when the same number of button presses were randomly timed.

Though there are different ways to capture agreement between an individual and a group (cf. Sasmita & Swallow, 2022), we adopted the *agreement index* (Kurby & Zacks, 2011) because it was used in a related study (Swallow et al., 2018) and would therefore allow for easier comparison of this study to those earlier results. To calculate the agreement index, *binned individual time series* were generated for each viewing of a video by coding whether an individual pressed a button or not for each second of the video. The individual time series were then correlated with a *group binned time series* generated by averaging the binned individual time series for a comparison group of participants. The comparison group varied across analyses and will be specified in the results. Importantly, this group never contained the individual being examined, even when they were in identical conditions. The minimum and maximum correlations possible (given the number of times that individual pressed the button and the comparison group time series) were then calculated and used to scale the observed correlation using the following equation (max correlation – observed correlation)/(max correlation – min correlation).

Statistical analyses

Data were analyzed in R (v.4.1.1 R Core Team 2021) using linear or generalized linear mixed-effects models (lmerTest; Kuznetsova et al., 2017), emmeans (Lenth, 2023), and lab produced analytical tools (Sasmita & Swallow, 2022; available here: https://github.com/ksasmita/esMethods). The mixed-effects models will be described with the results, but all models included activity as a fixed effect rather than a random effect, due to there being only two activities in this study (Oberpriller et al., 2022). Because activity was not of interest, significant effects involving activity will only be described if they show that an effect differed in presence/absence or sign across the two activities. Models also included random intercept terms for participants (adding random slopes for participants resulted in singular fits). For contrast tests, linear effects were used for the VAI, treatment contrasts were used for onsets, offsets, and task (with action change detection as the baseline), and deviation contrasts were used for perspective, task number, and activity. The Holm (1979) correction was used to avoid inflating type 1 error rates.

Results

Data from the action change, goal change, and segmentation tasks were analyzed to address four main questions. (1) Did participants perform the tasks as instructed? If they did, then participants should have identified more action changes than goal changes. We also evaluated these in relation to segmentation. Three additional analyses addressed questions that were central to the study's goals of evaluating whether untrained observers could use action and goal changes to segment events. (2) Did participants agree with each other about when action and goal changes occurred and was this comparable to agreement for event segmentation? (3) Were objective visual features predictive of action and goal changes as well as event boundaries? Finally, (4) Were event boundaries associated with the action and goal changes that were identified by untrained observers? We addressed these questions using a variety of approaches standard to the event segmentation literature and describe each in the following sections.

Did participants perform the tasks as instructed?

We first asked whether the action and goal task instructions led participants to identify more action changes than goal changes, and whether the action and goal changes they identified were finer or coarser than events. The number of button presses per minute (*bpm*) was calculated for each viewing of a video. Button press rates were analyzed in a linear mixed-effects model with task, perspective, task number, activity and their interactions as fixed effects and participant as a random effect (bpm~ task*perspective*number*activity + (1 | participant)).

If participants performed the tasks as instructed, then button presses should be more frequent in the action task than in the goal task. The data (Fig. 2) were consistent with this expectation. Button presses were more than four times as frequent in the action task, M = 18.116, SD = 8.909, 95% CI: [15.926, 20.306], than in the goal task, M = 3.944, SD = 3.134, 95% CI: [3.191, 4.697], significant main effect of task, $F(3, 150.74) = 481.669$, $p < .001$. The difference in button press rates in the action and goal tasks was larger when the tasks were performed first, $t(151) = 11.507$, $p < .001$, $d = 3.809$, rather than second, $t(151) = 10.684$, $p < .001$, $d = 3.512$. This resulted in significant interactions between task and task number, $F(3, 131) = 5.571$, $p = .001$, main effect of activity, $F(1, 917) = 52.985$, $p < .001$, main effect of task number, $F(1, 917) = 34.568$, $p < .001$, and task × activity interaction, $F(3, 917) = 10.926$, $p < .001$.

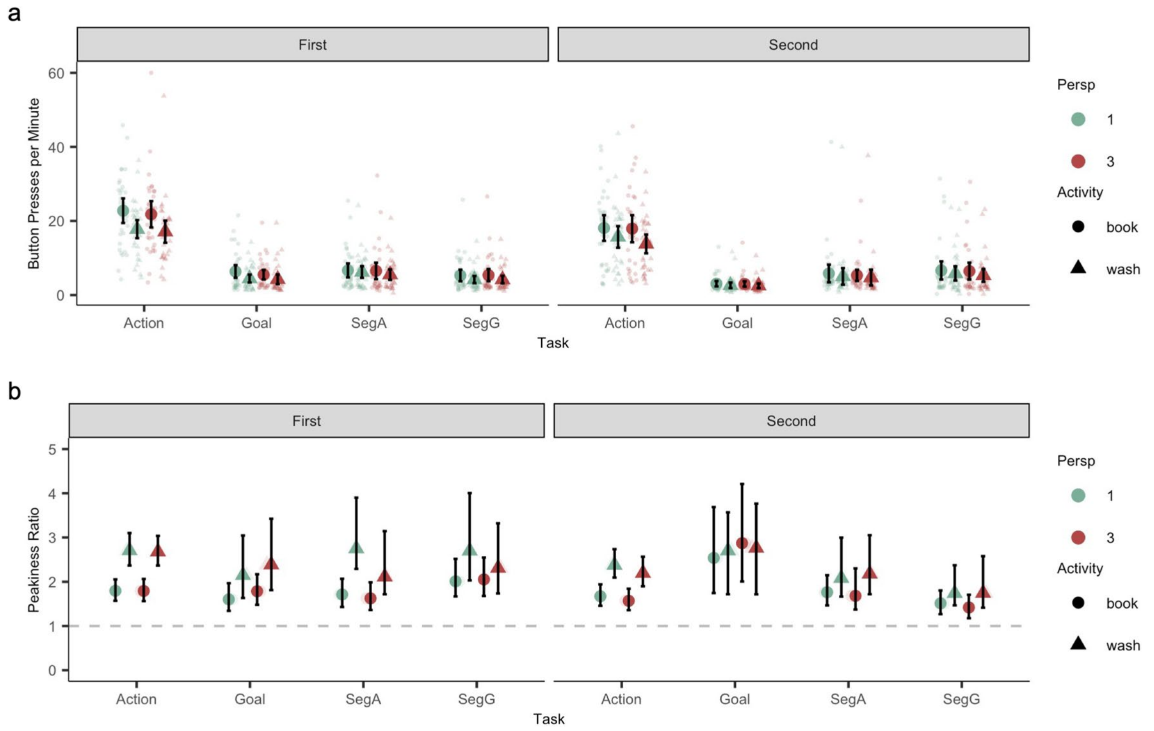

Fig. 2 (**a**) Button presses rates (button presses per minute) in each task, perspective, and task number, and activity. Small shapes indicate rates for each participant. Large shapes indicate the sample mean. Error bars reflect the 95% confidence interval around the mean. In some cases, the error bars do not extend beyond the symbol indicating the mean value. (**b**) The ratio of observed to expected peakiness values in each task, perspective, task number, and activity. Error bars reflect the 95% confidence interval around the mean based on the bootstrapped distribution for expected peakiness. The dashed line indicates the expectation under the null hypothesis

Button press rates were similar across the two groups performing the segmentation task, segA, M = 5.717, SD = 5.502, 95% CI: [4.364, 7.07], and segG, M = 5.39, SD = 5.097, 95% CI: [4.166, 6.615], $t(151) = 0.352$, $p = .726$. This was true for both activities, $t(193) < 0.516$, $p > .607$, as well as when the tasks were performed first or second, |$t(151)$s| < 1.156, *ps* .> .515. Button press rates were higher in the segmentation tasks than in the goal task, but this difference reach significance only for the segG group, $t(917) = -4.372$, $p < .001$, $d = 0.372$, and $t(151) = -1.947$, $p = .107$ for segA. Button press rates were lower in segA and segG than in the action task, $t(151)$s > 14.098, *ps* < .001, *d*s > 3.206. No other effects or interactions were significant, largest $F(1, 917) = 3.655$, $p = .056$ for the main effect of perspective.

Thus, participants identified more actions than goals, as instructed, for both activities and task orders. Further, the events that participants identified were larger than actions and tended to be slightly smaller than goals. This pattern was present, though not always significantly so, in all but one comparison (goal vs. SegG when segmentation was performed first; Fig. 2a). Overall participants identified fewer changes in the second half of the experiment. The effect of task number could have multiple sources, including fatigue and increased familiarity with the videos.

Do untrained observers agree with each other about action and goal changes?

If action and goal changes contribute to event segmentation, then untrained observers should be able to consistently identify them in an ongoing activity. Consequently, untrained viewers should agree with each other about when an actor's actions and goals change. Visual inspection of the group density time series (OSM Figs. S1 and S2) suggested that participants in each task pressed the button at similar times. To quantitatively evaluate whether they were more likely to do so than chance, we used a measure of within-group agreement, peakiness (Sasmita & Swallow, 2022). Because goals should be related to actions, we also asked whether participants identified goals that contained the actions that were identified by a separate group of individuals, reflecting the hierarchical organization of these two aspects of human activity (Vallacher & Wegner, 1989).

Peakiness We evaluated whether button presses in the group time series clustered together more than expected by chance. To do this we computed the ratio of the observed peakiness of the group time series to the mean and confidence intervals of expected values estimated from randomly generated data. These values indicated that, across all tasks and conditions, peakiness was 1.422–2.854 times greater than expected by chance (range of the 95% CI lower limits: 1.2–2.397; Fig. 2b). Participants therefore agreed with each other about when action changes, goal changes, and event boundaries occurred in the videos.

Alignment of goal changes with action changes To assess participants' sensitivity to the hierarchical structure of goal-directed activity, the proximity of an individual's button presses during the goal task to normative changes in action (see methods) was calculated and compared to the expected distance if they were randomly placed (cf. Zacks, Braver et al. 2001, Zacks, Tversky et al. 2001). The differences between the expected and observed distances (alignment) are plotted in Fig. 3 (larger values indicate better alignment of goal changes with action changes) and were analyzed in a linear mixed-effects model that included perspective, task number, and activity as fixed effects and participant as a random effect (alignment ~ perspective*number*activity + (1 | participant)). All fixed effects were contrast coded to ensure that the intercept is the mean observed difference across all conditions.

On average, participants placed goal changes closer to normative action boundaries than would be expected by chance, as indicated by the model intercept being significantly greater than 0, $\beta_0 = 0.316$, $SE = .022$, $t(67) = 14.176$, $p < .001$. The degree to which they did so, however, varied across conditions such that those conditions that increased the amount of information available to observers resulted in greater alignment (third- rather than first-person perspective, $M_{Diff} = 0.15$, $SE_{Diff} = 0.026$, $d = 0.694$, 95%CI = [0.447-0.941], $F(1, 201) = 33.217$, $p < .001$ and goal task performed second rather than first, $M_{Diff} = 0.237$, $SE_{Diff} = 0.045$, $d = 1.1$, 95%CI = [0.645-1.55], $F(1, 67) = 28.305$, $p < .001$). However, a three-way interaction, $F(1, 201) = 17.791$, $p < .001$, indicated that the effect of task number was present in all conditions except when the book activity was viewed from the first-person perspective, $M_{Diff} = 0.07$, $SE_{Diff} = 0.063$, 95%CI = [-0.259-0.903], $t(203) = 1.098$, $p = .274$, $d = 0.322$.

If goals contain actions, then goal changes may be more likely to be identified after the nearest action change than before it. However, we did not observe reliable evidence of this (OSM). This could be because action boundaries were defined by a separate group of individuals, and thus may not be the best reflection of precisely when observers in the goal task identified the ends of actions.

These data suggest that participants tended to align goal changes with a subset of action changes, and this tendency differed across perspectives and task numbers. This pattern could reflect participant's sensitivity to the hierarchical structure of activity as it unfolds, and that this sensitivity may increase when more information about the activity's spatial context (the activity is presented from the third-person perspective) or temporal context (the activity has already been viewed) is available.

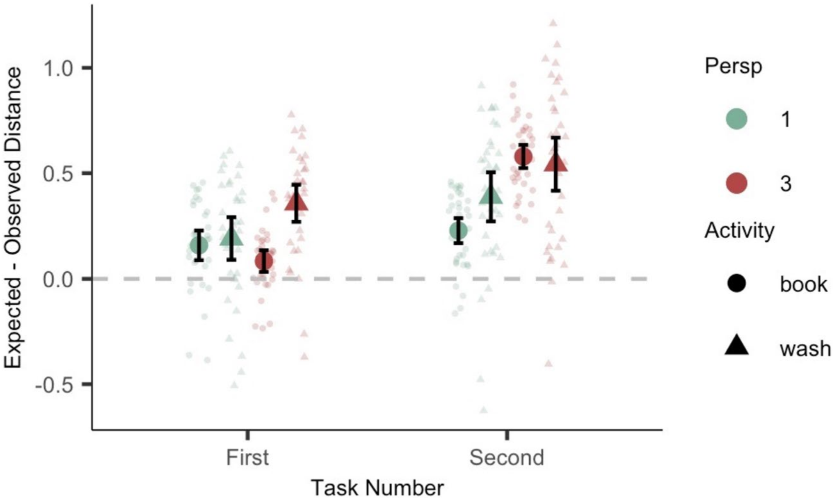

Fig. 3 The difference between expected and observed distances between button presses in the goal task and normative action boundaries for each perspective, task number, and activity. Small shapes indicate differences for each participant. Large shapes indicate the sample mean. Error bars reflect the 95% confidence interval around the mean. In some cases the error bars are small relative to the symbol indicating the mean value

Are action and goal changes invariant across stimulus features?

Even if untrained observers are able to identify action and goal changes consistently, the changes they identify may be reducible to low-level stimulus features rather than to the content of the activity itself. Because videos of the same activity recorded from different perspectives had different objective visual features (see Swallow et al., 2018 for additional characterization of these differences), we addressed this question by examining the effect of perspective on when participants pressed the button under different conditions. First, we asked whether the action or goal changes that participants identified occurred at similar times when the same activity was viewed from the *same* or a *different* perspective. Lower agreement across different perspectives relative to within the same perspective would suggest that action and goal changes depend, at least to some degree, on what participants can see in the videos. Second, we asked whether action and goal changes are associated with objective stimulus features, and whether this relationship depends on which perspective the video was recorded from. A difference in the relationship between button presses and stimulus features across perspectives suggests flexibility in how the changes are identified and is consequently consistent with a less concrete basis for identifying those changes.

The Effect of Perspective on Change Identification To evaluate whether action and goal changes were tied to the activity rather than to stimulus features we used the agreement index to compare segmentation patterns of the same activity when viewed from different perspectives. Individual time series were compared to the group time series generated by participants who viewed an activity from the same or a different perspective (*comparison group*). If perspective influences when participants press the button in any of these tasks, then individuals should agree more with the group that viewed the activity from the same perspective than with the group that viewed it from the other, different perspective (Swallow et al., 2018). A preliminary analysis indicated that the pattern of results differed for participants performing the task first versus second (see OSM). To simplify the analysis and avoid carry-over effects, only the data from the first task are presented here. These were fit with a linear mixed-effects model that included task, perspective, comparison group, activity, and their interactions as fixed effects (agreement ~ task*perspective*comparison group*activity + (1 | participant)). We report only those results that are relevant to the effect of perspective in this section (see OSM for additional effects).

Overall, agreement was greater between individuals and groups that viewed the activity from the same perspective, M = .594, SD = .122, 95%CI = [.573–.615] rather than from a different perspective, M = .569, SD = .124, 95%CI = [.548–.589], $F(1, 917) = 25.791$, $p < .001$, $d = 0.309$, and this effect did not significantly interact with any other factors, Fs < 1.777, ps > .150 (Fig. 4). Although we hypothesized that perspective would influence button press patterns more for the action task than for the other tasks, the results were inconsistent with this possibility. Therefore, the effect of perspective on when participants pressed the button was similar across tasks.

Contribution of objective stimulus features to change identification

If action and goal changes are tied to the content of an activity rather than to stimulus features, then any relationship between stimulus features and action and goal changes should differ across perspectives (which show the same

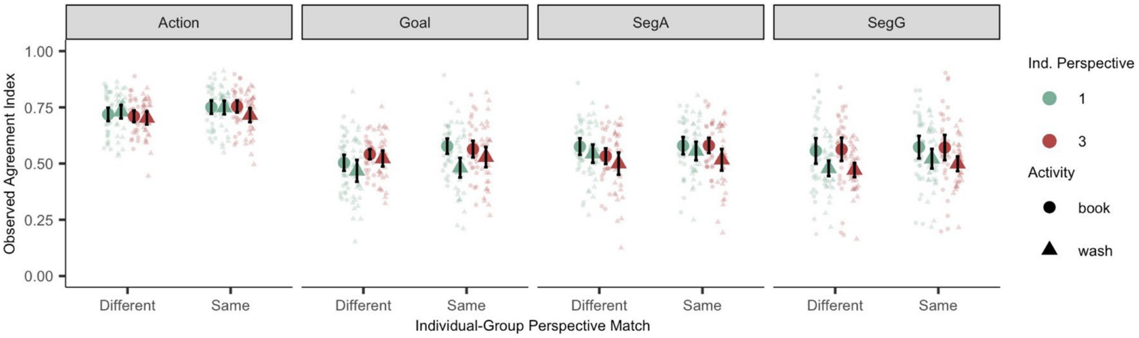

Fig. 4 Agreement between individuals and groups viewing an activity from the different perspective or the same perspective, for each task, perspective, and activity. Small shapes indicate agreement for each participant in each condition. Large shapes indicate the sample mean. Error bars reflect the 95% confidence interval around the mean. In some cases, the error bars do not extend beyond the symbol indicating the mean value

content but with different observable features). To establish whether stimulus features are consistently associated with action and goal changes across perspectives, we analyzed the data with a generalized linear mixed-effects model (binomial link function). The model was fit to the individual time series and included perspective, task, the z-scored VAI, object touch onsets, object touch offsets, and all of their interactions as fixed effects (bp ~ task*perspective*VAI*Onset*Offset + activity + (1 | participant)). To reduce the complexity of the model and the results, only first-task data were analyzed and no interactions involving activity were included. Because the full model revealed many high-level interactions, including one involving all fixed effects, we additionally fit separate, feature only models (bp ~ VAI*Onset*Offset + activity + (1 | participant)) for each task and each perspective to better visualize how the relationship between stimulus features and button presses changed across these conditions. Regression coefficients from these models are plotted in Fig. 5. All coefficients for the full model and for the corresponding model on second task data are reported in the OSM. Our presentation focuses on three issues: whether a relationship between button presses and stimulus features exists, whether that relationship differs across tasks, and whether it depends on perspective.

We first established whether task performance was associated with objective stimulus features at all (a precondition for asking whether these relationships are modulated by task and perspective), and whether this relationship differed across tasks. As can be seen in Fig. 5, button presses were associated with the VAI, odds ratio (OR) for 0.5 versus -0.5 z-scored VAI = 1.24, SE = 0.017, $F(1, \text{inf}) = 255.146$, $p < .001$, onsets, OR for present versus absent = 1.3, SE = 0.03, $F(1, \text{inf}) = 128.139$, $p < .001$, and offsets, OR for present versus absent = 1.12, SE = 0.026, $F(1, \text{inf}) = 23.732$, $p < .001$. These relationships were moderated by several higher-order interactions among the features, including both super-additive (VAI × offset interaction, interaction contrast OR = 1.12, SE = 0.031, $F(1, \text{inf}) = 18.738$, $p < .001$), and sub-additive (e.g., VAI × onset × offset interaction contrast OR = 0.669, SE = 0.037, $F(1, \text{inf}) = 54.382$, $p < .001$) effects. Thus, the likelihood that a participant would identify a change was influenced by stimulus features.

Importantly, the relationship between the coded features and button presses differed across tasks, resulting in several significant two- and three-way interactions involving task, smallest $F(1, \text{inf}) = 4.474$, $p = .004$ for onset × offset × task. For the action task, the effect of VAI was weaker than in the other tasks (see Fig. 5), smallest interaction OR for other tasks versus action task = 1.15, SE = 0.043, $z = 3.800$, p < .001, and the onset main effect and the onset × offset interactions were stronger for the action task than for two of the three other tasks (smallest OR = 1.35, SE = 0.160, z

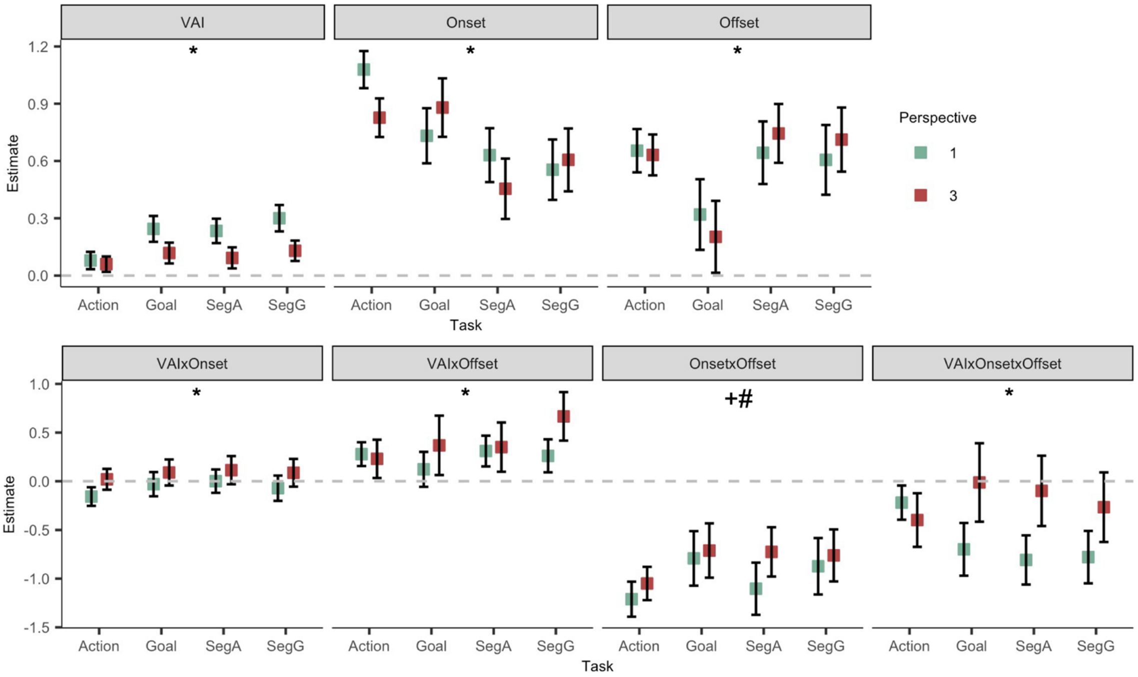

Fig. 5 Estimates (log odds) of the relationship between stimulus features and button presses derived from generalized linear mixed-effects models that were separately fit to each perspective within each task, using only the first-task data. Squares indicate the model estimate. Error bars reflect the 95% confidence interval around the estimate. Dashed line indicates the expectation under the null hypothesis. +: effect interacted with perspective in the full model. #: effect interacted with task in the full model. *: effect interacted with task × perspective in the full model (supersedes any significant two-way interactions)

= 2.55, p = .022). In contrast, for the goal task the effect of offset was weaker than in the other tasks, smallest interaction OR effect = 1.207, SE = 0.072, z = 3.172, p = .006. This pattern suggests that action changes were more likely to be identified at the beginning of contact with an object (as long as it did not coincide with the end of object contact), while goal changes and event boundaries were more likely to be associated with a broad measure of visual change. In contrast to action changes and event boundaries, goal changes were relatively unlikely to be identified at the end of object contact.

Variability in the degree to which different features contribute to action change, goal change, or event boundary identification when an activity is viewed from different perspectives could indicate flexibility in how these features are used (Swallow et al., 2018). We therefore expected greater effects of perspective on the association between button presses and stimulus features for goal changes (which should be based on more abstract information) than for action changes (which should be based on more concrete information). Consistent with this possibility, the model suggested differential effects of perspective across tasks. The effect of perspective on the relationship between button presses and video features was evident in a number of interactions, including a five-way interaction between task, VAI, onset, offset, and perspective, $F(1, \text{inf}) = 5.304$, p = .001. Post hoc analyses indicated that for the goal change and event segmentation tasks, combining more feature changes (onsets, offsets, and VAI) resulted in greater suppression of button presses (relative to lower-order effects) only when participants viewed first-person videos (Fig. 5), weakest contrast for first-person videos OR = 0.513, SE = 0.071, z = -4.84, p < .001 for goal, strongest contrast for third-person videos OR = 0.776, SE = 0.141, z = -1.390, p = .163 for segG. For the action task, perspective interacted only with onsets, whose effect was greater for first-person videos, OR = 1.62, SE = 0.074, than for third-person videos, OR = 1.36, SE = 0.059, contrast OR = 0.842, SE = 0.053, z = -2.726, p = .006.

Thus, the relationship between button presses and stimulus features changed across perspectives in all tasks. For action change detection, perspective modulated the influence of touch onsets. In contrast, perspective influenced the effects of all stimulus features for goal change detection and event segmentation. This pattern suggests greater flexibility in the use of stimulus features when identifying goal changes and event boundaries than when identifying action changes. Consistent with this possibility, adding perspective and its interactions to generalized linear effects models that were separately fit to each task increased the variance explained by the fixed effects (R^2; Nakagawa & Schielzeth, 2013) five to seven times more for the goal and segmentation tasks than it did for the action task (.005-.007 vs. .001; see OSM).

Do untrained observers identify action and goal changes that coincide with event boundaries?

A final question addressed in this study was whether action and goal changes identified by untrained observers are associated with event boundaries. We examined whether button presses in the segmentation task were predicted by button presses in the action and goal change-detection tasks, and whether this relationship was present after accounting for changes in the visual features of the videos. The proportion of participants who identified an action change or a goal change in each second of the videos was z-scored (mean centered and divided by the standard deviation) within each activity and condition. The z-scored values were then entered as fixed effects in generalized linear mixed effect models of button presses in the event segmentation task only (see Table 1 for details). In contrast to approaches that code action and goal changes as either present or absent, this

Table 1 Fit metrics and likelihood ratio tests (LRTs) comparing models characterizing the relationship between stimulus features, the proportion of participants that identified action and goal changes, and segmentation task performance

Model[a]	npar	AIC	BIC	Deviance	Chi sq	df	p	R^2_m	R^2_c	Inc. R^2_m
1	18	32864	33026	32828				0.040	0.181	
1 * actCh[b]	34	32047	32353	31979	848.44	16	< .001	0.094	0.231	0.054
1 * goalCh[b]	34	31519	31825	31451	1376.60	16	< .001	0.105	0.244	0.064
1 * actCh * goalCh[c]	66	31136	31730	31004	447.57	32	< .001	0.150	0.283	0.045

a Model 1 is bp ~ task * perspective * VAI * onset * offset + activity + (1|participant). Models with actCh and goalCh included their main effects and interactions with all fixed effects except activity. Models were fit to data from the segA and segG tasks, and only when they were the first task performed

b Chi square test is in relation to the base model that does not include action or goal changes

c Chi square test is in relation to the model that includes goal changes

AIC Akaike Information Criterion, *BIC* Bayesian Information Criterion, R^2_m proportion of variance explained by fixed effects, R^2_c proportion of variance explained by fixed and random effects, *Inc.* R^2_m increment in proportion of variance explained by fixed effects in relation to comparison model approach thus treated action and goal changes as continuous variables. We used model comparisons to evaluate whether adding action changes, goal changes, and their interactions to the models explained more variance than including only the VAI, onsets, offsets, and their interactions (Table 1). Both action and goal changes explained significantly more variance in event boundary identification than the coded stimulus features on their own, particularly when combined. An examination of the model including both action and goal changes indicated significant, high-order interactions between all fixed effects (including a six-way interaction among all fixed effects, $F(1, \text{inf}) = 5.773$, $p = .001$). Action changes and goal changes increased the likelihood that a boundary would be reported, action change OR for -0.5 versus 0.5 = 1.23, SE = 0.032, $z = 7.931$, $p < .001$, goal change OR for -0.5 versus 0.5 = 1.72, SE = 0.046, $z = 20.069$, $p < .001$. However, their effects were often subadditive with each other and with other effects (see OSM for all fixed effect estimates and a visualization of the higher order interactions).

These findings illustrate that the action and goal changes identified by untrained observers capture aspects of the videos that are related to segmentation, above and beyond those captured by the VAI, touch onsets, or touch offsets. They also affirm the relationship between segmentation and action and goal changes described in previous reports, showing that this relationship exists even when using changes identified by untrained observers.

Discussion

Many accounts of event perception propose that observers identify event boundaries by tracking changes in content such as an actor's actions and goals (Baldwin et al., 2001; Richmond & Zacks, 2017; Zacks et al., 2007). For this to be the case, however, observers must also at least partially agree about when the actors' actions and goals themselves change. In support of this possibility, the current study revealed three critical but underexplored phenomena related to event segmentation: First, untrained observers can reliably and consistently identify goal changes and action changes in ongoing naturalistic activities. Second, they identify action changes and goal changes that are related, but not reducible, to several visual features of activities. Third, independent of these visual features, action and goal changes identified by untrained observers are correlated with event boundaries. These findings are consistent with theories suggesting the importance of content information in event perception, specifically, a role of action and goal inferences in processes that structure the perception of others' activities over time.

People can identify action and goal changes as activities unfold

This study directly demonstrated that untrained observers can reliably identify action and goal changes in ongoing, naturalistic activities. Agreement was highest for action changes and comparable for goal changes and event boundaries (Figs. 2 and 4; OSM). Thus, the data show that people perform the action and goal change-detection tasks with similar, if not greater, levels of agreement than when segmenting events. This is despite the fact that, like events, action and goal changes lack an objective cue for when they have changed, and observers may vary in when they detect a change and when they press a button to mark it (e.g., some might anticipate the change, whereas others may wait to respond until it occurs; see also Levine et al., 2017). Moreover, the data also showed that untrained participants marked goal and action changes that reflected the hierarchical relationship between actions and goals (Vallacher & Wegner, 1989). Reminiscent of the relationship between coarse and fine boundaries (Zacks & Tversky, 2001), participants identified roughly four times as many action changes as goal changes, and goal changes were statistically closer to normative action boundaries than expected by chance (though we did not observe evidence that goal changes followed the nearest action change). This finding supports the proposal that observers use information about an actor's actions and goals when segmenting continuous streams of naturalistic activities into events.

Action and goal changes are associated with, but not reducible to, objective visual features

This study demonstrated that, like event boundaries, observers identified action and goal changes that were associated with objective visual features in the movies, though the precise relationship differed across detection tasks and perspectives. In general, the more abstract or content level changes (goal changes and event boundaries, in contrast to action changes) were more strongly dependent on complex interactions among several objective visual features. This was particularly true for activities presented in the first- rather than the third-person perspective. One implication of these findings is that objective visual features and their interactions, not just high-level knowledge and inferences, could contribute to the online perception of actions and goals.

Yet, the relationship between action and goal change detection and visual features was not straightforward because the presence of multiple types of visual changes in a brief time window suppressed button presses. Combining feature changes within a 1-s bin, particularly onsets and offsets, often decreased the likelihood of a button press relative to the presence of one of the feature changes on its own. This is consistent with prior suggestions that goal-directed behavior may be defined by the cohesiveness of action, rather than the mere presence of change (Levine et al., 2017). Notably, this pattern differs from prior observations that more changes in a situation (e.g., changes in space, time, character goals, character interactions, etc.) increase the likelihood that an event boundary will be identified (Zacks et al., 2009, 2010). These factors and their interactions should be further considered in future work.

Importantly, the data are also consistent with claims that visual features are not sufficient to account for action and goal understanding (Catmur, 2015; Koul et al., 2016; Naish et al., 2013). Like event segmentation, action and goal change detection utilized these features flexibly, being more strongly driven by some features when an activity was viewed from one perspective rather than the other. Further, the flexible use of these features was more evident for goal change detection and event segmentation than for action change detection. This aligns with the hypothesis that more abstract changes are less deterministically tied to objective visual features than are more concrete changes.

Untrained observers identify action and goal changes that correspond to event boundaries

While prior research demonstrates that experimenter-defined action and goal changes are associated with segmentation (e.g., Kopatich et al., 2019; Swallow & Wang, 2020), it does not address whether those doing the segmenting (untrained observers) detect action and goal changes at event boundaries. In this study, model comparisons indicated that both action and goal changes identified by untrained observers were positive and significant predictors of event segmentation, above and beyond their relationship with other objective visual features of the activity and for both perspectives. To our knowledge, this is the first demonstration that untrained observers identify changes in actions and goals in everyday activities that correspond to event boundaries and that are not reducible to simple features of the activity. An important consideration for future research will be to examine whether action and goal changes identified by untrained observers contribute differently to event segmentation at different granularities, as implied by prior work (e.g., Zacks & Tversky, 2001).

Implications for understanding online goal processing

Goals are defined in the literature with varying degrees of abstraction. On one end of the spectrum, goals and intentions are grouped together (Hamilton & Grafton, 2007). On the other end, goal completion is defined as the endpoint of a movement or series of movements (the outcome of actions; e.g., Levine et al., 2017; Olofson & Baldwin, 2011; Woodworth, 1899) rather than the reason that the outcome is desired (the intention, or "why" of the actions; Catmur, 2015). In this study, goals were characterized as the reason that a sequence of behaviors is performed, and were tied to the outcome of those behaviors rather than the motive driving the behavior. Whether people detect changes in an actor's intentions as a naturalistic activity unfolds remains an open question.

A related issue is the interplay between perceptual processing and higher-level inference about an actor's goals or intentions. Whereas some accounts emphasize the importance of mental state representations in interpreting an agent's actions as they unfold (i.e., actions are identified within the context of a hypothesized goal or intention; inverse planning theory; Baker et al., 2009; Ullman et al., 2009), others suggest that online goal inferences are supported by a direct mapping between observed actions and an observer's action production system (simulation theory; Fogassi et al., 2005). Still other research raises questions about the degree to which action production systems could support inferences about an agent's mental state on their own (Catmur, 2015). Although this study was not designed to distinguish between these perspectives, the results are relevant to work on goal processing, as we describe next.

First, this study demonstrates that observers process an agent's actions and goals continuously, in relation to each other, and in the moment, at least to the point of being able to identify a change in these features of an activity as it unfolds. This work thus demonstrates a critical pre-condition for any account of goal processing (including inverse planning and simulation theory) that suggests that it is done as a part of normal perception.

Second, our observation that the relationship between goal changes and observable features of the videos depended on perspective offers constraints on the mechanisms that process goals. In particular, goal processing appears to (1) succeed in the absence of some relevant information (e.g., body posture in first-person perspectives; precise information about hand grip and trajectories in third-person perspectives), (2) be based on information that has at least been partially abstracted from the stimulus, and (3) be context dependent. Thus, it appears unlikely that goal changes in continuous, naturalistic activities can be boiled down to the simple detection of a few features or combinations of features of the activity, such as the precise movements one is making, interactions with new objects, or large changes in an actor's position or posture.

Finally, the importance of context and knowledge for goal processing is underscored by our observation that the alignment of goal changes with action changes was greater with wider views of the scene (e.g., in the third- rather than the first-person perspective) or with greater familiarity with the activity (e.g., when goal changes were detected during the second viewing). These findings argue against the possibility that goal processing is tightly coupled to specific motor productions, and favor accounts that include a role for more integrative and knowledge-based processing. Given these findings, much like events (Newberry et al., 2021; Richmond & Zacks, 2017; Swallow et al., 2018), goal processing may be best considered to rely on mid-level (e.g., relating to spatial configurations, object shapes, and biological motion) to high-level (e.g., relating to schemas, scripts, and inferred motivations or desires) information that is consistent across variable sensory inputs and is informed by prior knowledge.

Implications for event segmentation

Richmond and Zacks (2017) argued that event models should tolerate variability in perceptual input, smoothing out perceptual processing streams with input from more stable sources of information such as the environment, objects, and intentional actions. The finding that visual features are not sufficient to account for action and goal change detection, along with the observation that action and goal changes are associated with event boundaries, even after considering objective stimulus features, indicates that they could serve as a source of stability.

Although knowledge about what an actor is doing has been found to contribute to segmentation (Newberry, et al., 2021), these effects are not always detectable in studies examining everyday activities (Hard et al., 2006, Swallow & Wang, 2020; Zacks et al., 2009). Furthermore, changes in the content of an experience are correlated with changes in observable visual features (Cutting, 2014). The observation that action and goal changes contribute to event segmentation even after accounting for low-level visual change addresses this concern and provides additional evidence suggesting that an observer's knowledge may play a role in how they segment everyday activities into meaningful events.

The current study does suggest some differences in event segmentation and the detection of action and goal changes across perspectives. Though there was weak and inconsistent evidence of a similar effect in prior work (Swallow, et al., 2018), the current study used a larger sample size to increase the ability to detect a smaller effect. Indeed, the magnitude of the effect of comparison group on agreement (same vs. different perspective mean d' across segmentation groups = .271) was not much larger than the effect of segmentation group on agreement (SegA vs SegG d' = .232), or that of task order (segmentation first vs. second mean |d'| across segmentation groups = .221). This general pattern was further confirmed in correlational and cluster analyses of the group time series data, which suggested that segmentation pattern similarity was greatest when comparing first- and third-person perspective segmentation by the same group (see OSM). Thus, while the detection of action changes, goal changes, and event boundaries appears to be at least partially dependent on the information presented in the video, they are also similar across perspectives. Differences in segmentation across perspectives could reflect the reliance of mid-level information (e.g., an object interaction) on low-level information (e.g., an object contact or movement) that varies in its presence across perspectives (e.g., the object may be occluded by the actor's body).

Although this study shows that online action and goal changes are correlated with event boundaries, it does not provide insight into the directionality of this relationship. Even if event segmentation occurs as a consequence of detecting action and goal changes (e.g., Richmond & Zacks, 2017; Zacks et al., 2007), the relationship between event segmentation and action and goal processing may be complex and interactive (cf. Newtson, 1980). Furthermore, this relationship could change with development (Zheng et al., 2020), as children bootstrap inferences about goals from physical markers of action boundaries (Baldwin et al., 2001) and learned action sequences (Buchsbaum et al., 2015; Kosie & Baldwin, 2018; Levine et al., 2017). It is also plausible that under some circumstances event segmentation could facilitate the detection of changes in an actor's actions or goals, rather than be caused by them. This type of relationship could occur if prediction errors generated by unexpected changes in sensory information (e.g., generated by a discontinuity in an actor's arm trajectory) are sufficiently large to generate an event boundary, prompting a new assessment of the actor's behavior. Alternatively, such prediction errors could be suppressed when the situation continues to be consistent with the inferred goal of the actor, leading to the continued maintenance of the event model (Kuperberg, 2021).

In experimental settings, one way to operationalize event boundaries is to define them as changes in an actors' goals. A notable finding from our data is the similarity between participants' identification of goal changes and event boundaries, as well as similarities in how participants performed these two tasks (e.g., the relationship between visual features and button presses). This finding partially supports previous work suggesting a strong relationship between everyday events and an actor's goals. However, our data also indicates that goals and everyday events are not interchangeable. Untrained observers tended to identify fewer goal changes compared to event boundaries, and goal changes were less likely to be identified at touch offsets. Additionally, including action changes alongside goal changes in models predicting event boundaries yielded the best fit. Although the emphasis is different, this finding aligns with prior work on the relationship between event boundaries and goals (Levine et al., 2017). That study found that novices identified event boundaries during the starts and ends of expert identified goals. However, consistent with our findings, Levine, et al. (2017) also identified boundaries at other times in the activity. The differences between goals and events that we observed here likely reflect the nuance afforded by the temporal precision, use of the entire time series, and inclusion of actions and visual features in our analytical methods. Hence, untrained observers do not treat goal changes and events identically, despite their evident relationship. Event segmentation and goal change detection do not appear to be the same thing.

Limitations and future directions

There are additional limitations to this study that suggest the need for more research on this topic. First, this study examined the VAI, a global measure of visual change, and touch onsets and offsets, leaving open the possibility that a more refined measure of visual change or the inclusion of additional features might fully account for action and goal changes. We believe this is unlikely, however, as we used relatively simple videos and earlier work investigating optic flow (Swallow et al., 2018) or movement trajectories (Olofson & Baldwin, 2011) in event or goal processing suggest limited contributions to segmentation. Second, participants in the current study watched each activity several times, sometimes from different perspectives, and sometimes while performing different tasks. This procedure could induce fatigue, increase familiarity with the stimulus, or produce carry-over effects from one task to another. For these reasons, many of the analyses presented in this paper utilized data from the first task. Still, the effects of each of these factors on segmentation are worth investigating. Additionally, to encourage participants to adopt understandings of actions and goals that were clearly distinguished from each other, we provided them with examples that characterized actions at a fine grain and goals as the reasons behind sequences of behavior. It is possible that different instructions or examples would have produced different results. Finally, as with most research on event segmentation, this study utilized a small number of activities. This design choice kept the experiment to a reasonable length but raises questions about the generalizability of the results. Indeed, although the results were consistent across the two activities, some effects were stronger in one activity than in the other. This is itself an interesting finding, as it suggests that the effects of objective visual features, action changes, and goal changes on segmentation depend on as yet unspecified factors that vary across activities and stimuli.

Conclusion

This study provides direct evidence that people can consistently identify action and goal changes while watching everyday activities. Aligned with Event Segmentation Theory (Zacks et al., 2007), goal changes and action changes influence event segmentation while accounting for the influence of low-level objective stimulus features.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13421-024-01525-8.

Acknowledgements The authors would like to thank Michael Goldstein for his input on an earlier version of this manuscript and Joel Cruz for help with data collection. This work was funded by the Cornell University College of Arts and Sciences.

References

Bach, P., Nicholson, T., & Hudson, M. (2014). The affordance-matching hypothesis: How objects guide action understanding and prediction. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00254

Bach, P., & Schenke, K. C. (2017). Predictive social perception: Towards a unifying framework from action observation to person knowledge. *Social and Personality Psychology Compass, 11*(7), e12312. https://doi.org/10.1111/spc3.12312

Bailey, H. R., Kurby, C. A., Giovannetti, T., & Zacks, J. M. (2013). Action perception predicts action performance. *Neuropsychologia, 51*(11), 2294–2304. https://doi.org/10.1016/j.neuropsychologia.2013.06.022

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition, 113*(3), 329–349. https://doi.org/10.1016/j.cognition.2009.07.005

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron, 95*(3), 709-721. e5. https://doi.org/10.1016/j.neuron.2017.06.041

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition, 106*(3), 1382–1407. https://doi.org/10.1016/j.cognition.2007.07.005

Baldwin, D., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development, 72*(3), 708–717. https://doi.org/10.1111/1467-8624.00310

Barrett, L. F., & Satpute, A. B. (2013). Large-scale brain networks in affective and social neuroscience: Towards an integrative functional architecture of the brain. *Current Opinion in Neurobiology, 23*(3), 361–372. https://doi.org/10.1016/j.conb.2012.12.012

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Bläsing, B. E. (2015). Segmentation of dance movement: Effects of expertise, visual familiarity, motor experience and music. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01500

Blakemore, S.-J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience, 2*(8), 561–567. https://doi.org/10.1038/35086023

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436. https://doi.org/10.1163/156856897X00357

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

Catmur, C. (2015). Understanding intentions from actions: Direct perception, inference, and the roles of mirror and mentalizing systems. *Consciousness and Cognition, 36*, 426–433. https://doi.org/10.1016/j.concog.2015.03.012

Cutting, J. E. (2014). Event segmentation and seven types of narrative discontinuity in popular movies. *Acta Psychologica, 149*, 69–77. https://doi.org/10.1016/j.actpsy.2014.03.003

Decroix, J., Roger, C., & Kalénine, S. (2020). Neural dynamics of grip and goal integration during the processing of others' actions with objects: An ERP study. *Scientific Reports, 10*(1), 5065. https://doi.org/10.1038/s41598-020-61963-7

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

Ezzyat, Y., & Davachi, L. (2011). What Constitutes an Episode in Episodic Memory? *Psychological Science, 22*(2), 243–252. https://doi.org/10.1177/0956797610393742

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal Lobe: From Action Organization to Intention Understanding. *Science, 308*(5722), 662–667. https://doi.org/10.1126/science.1106138

Friend, M., & Pace, A. (2011). Beyond event segmentation: Spatial- and social-cognitive processes in verb-to-action mapping. *Developmental Psychology, 47*(3), 867–876. https://doi.org/10.1037/a0021107

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493–501. https://doi.org/10.1016/S1364-6613(98)01262-5

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General, 142*(3), 880–905. https://doi.org/10.1037/a0030045

Hamilton, A. F. D. C., & Grafton, S. (2007). The motor hierarchy: From kinematics to goals and intentions. In P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Attention and Performance* (xxii ed.). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199231447.001.0001

Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General, 140*(4), 586–604. https://doi.org/10.1037/a0024310

Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. *Memory & cognition, 34*(6), 1221–1235.

Heyes, C., & Catmur, C. (2022). What Happened to Mirror Neurons? *Perspectives on Psychological Science, 17*(1), 153–168. https://doi.org/10.1177/1745691621990638

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Hudson, M., Nicholson, T., Ellis, R., & Bach, P. (2016). I see what you say: Prior knowledge of other's goals automatically biases the perception of their actions. *Cognition, 146*, 245–250. https://doi.org/10.1016/j.cognition.2015.09.021

Kopatich, R. D., Feller, D. P., Kurby, C. A., & Magliano, J. P. (2019). The role of character goals and changes in body position in the processing of events in visual narratives. *Cognitive Research: Principles and Implications, 4*(1), 22. https://doi.org/10.1186/s41235-019-0176-1

Kosie, J. E., & Baldwin, D. A. (2018). Tuning to the Task at Hand: Processing Goals Shape Adults' Attention to Unfolding Activity. In *CogSci*.

Koul, A., Cavallo, A., Ansuini, C., & Becchio, C. (2016). Doing It Your Way: How Individual Movement Styles Affect Action Prediction. *PLOS ONE, 11*(10), e0165297. https://doi.org/10.1371/journal.pone.0165297

Kuperberg, G. R. (2021). Tea With Milk? A Hierarchical Generative Framework of Sequential Event Comprehension. *Topics in Cognitive Science, 13*(1), 256–298. https://doi.org/10.1111/tops.12518

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12*(2), 72–79. https://doi.org/10.1016/j.tics.2007.11.004

Kurby, C. A., & Zacks, J. M. (2011). Age differences in the perception of hierarchical structure in events. *Memory & Cognition, 39*(1), 75–91. https://doi.org/10.3758/s13421-010-0027-2

Kurby, C. A., & Zacks, J. M. (2022). Priming of movie content is modulated by event boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(11), 1559–1570. https://doi.org/10.1037/xlm0001085

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lenth, R. V. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means (Version 1.9.0)* [Computer software]. https://CRAN.R-project.org/package=emmeans

Levine, D., Hirsh-Pasek, K., Pace, A., & MichnickGolinkoff, R. (2017). A goal bias in action: The boundaries adults perceive in events align with sites of actor intent. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(6), 916–927. https://doi.org/10.1037/xlm0000364

Libby, L. K., Shaeffer, E. M., & Eibach, R. P. (2009). Seeing meaning in action: A bidirectional link between visual perspective and action identification level. *Journal of Experimental Psychology: General, 138*(4), 503–516. https://doi.org/10.1037/a0016795

Magliano, J. P., Todaro, S., Millis, K., Wiemer-Hastings, K., Kim, H. J., & McNamara, D. S. (2005). Changes in Reading Strategies as a Function of Reading Training: A Comparison of Live and Computerized Training. *Journal of Educational Computing Research, 32*(2), 185–208. https://doi.org/10.2190/1LN8-7BQE-8TN0-M91L

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-Neuron Responses in Humans during Execution and Observation of Actions. *Current Biology, 20*(8), 750–756. https://doi.org/10.1016/j.cub.2010.02.045

Naish, K. R., Reader, A. T., Houston-Price, C., Bremner, A. J., & Holmes, N. P. (2013). To eat or not to eat? Kinematics and muscle activity of reach-to-grasp movements are influenced by the action goal, but observers do not detect these differences. *Experimental Brain Research, 225*(2), 261–275. https://doi.org/10.1007/s00221-012-3367-2

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Newberry, K. M., Feller, D. P., & Bailey, H. R. (2021). Influences of domain knowledge on segmentation and memory. *Memory & Cognition, 49*(4), 660–674. https://doi.org/10.3758/s13421-020-01118-1

Newtson, D. (1980). An Interactionist Perspective on Social Knowing. *Personality and Social Psychology Bulletin, 6*(4), 520–531. https://doi.org/10.1177/014616728064004

Newtson, D., Engquist, G., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology, 35*(12), 847–862.

Oberpriller, J., De Souza Leite, M., & Pichler, M. (2022). Fixed or random? On the reliability of mixed-effects models for a small number of levels in grouping variables. *Ecology and Evolution*, *12*(7). https://doi.org/10.1002/ece3.9062

Olofson, E. L., & Baldwin, D. (2011). Infants recognize similar goals across dissimilar actions involving object manipulation. *Cognition, 118*(2), 258–264. https://doi.org/10.1016/j.cognition.2010.11.012

R Core Team. (2021). R: A Language and Environment for Statistical Computing [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Richmond, L. L., & Zacks, J. M. (2017). Constructing Experience: Event Models from Perception to Action. *Trends in Cognitive Sciences, 21*(12), 962–980. https://doi.org/10.1016/j.tics.2017.08.005

Sasmita, K., & Swallow, K. M. (2022). Measuring event segmentation: An investigation into the stability of event boundary agreement across groups. *Behavior Research Methods, 55*(1), 428–447. https://doi.org/10.3758/s13428-022-01832-5

Smith, M. A., & Anderson, B. D. (2004). A Window on Reality? *Journal of Clinical Oncology, 22*(8), 1360–1362. https://doi.org/10.1200/JCO.2004.01.946

Speer, N. K., Swallow, K. M., & Zacks, J. M. (2003). Activation of human motion processing areas during event perception. *Cognitive, Affective, & Behavioral Neuroscience, 3*(4), 335–345. https://doi.org/10.3758/CABN.3.4.335

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts☆. *Journal of Memory and Language, 53*(1), 125–140. https://doi.org/10.1016/j.jml.2005.02.009

Spunt, R. P., Falk, E. B., & Lieberman, M. D. (2010). Dissociable Neural Systems Support Retrieval of *How* and *Why* Action Knowledge. *Psychological Science, 21*(11), 1593–1598. https://doi.org/10.1177/0956797610386618

Swallow, K. M., Kemp, J. T., & Candan Simsek, A. (2018). The role of perspective in event segmentation. *Cognition, 177*, 249–262. https://doi.org/10.1016/j.cognition.2018.04.019

Swallow, K. M., & Wang, Q. (2020). Culture influences how people divide continuous sensory experience into events. *Cognition, 205*, 104450. https://doi.org/10.1016/j.cognition.2020.104450

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General, 138*(2), 236–257. https://doi.org/10.1037/a0015631

Ullman, T., Baker, C., Macindoe, O., Evans, O., Goodman, N., & Tenenbaum, J. (2009). Help or Hinder: Bayesian Models of Social Goal Inference. *Advances in Neural Information Processing Systems, 22*. https://proceedings.neurips.cc/paper_files/paper/2009/hash/52292e0c763fd027c6eba6b8f494d2eb-Abstract.html

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review, 94*(1), 3–15. https://doi.org/10.1037/0033-295X.94.1.3

Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology, 57*(4), 660–671. https://doi.org/10.1037/0022-3514.57.4.660

Woodward, A. L., & Sommerville, J. A. (2000). Twelve-Month-Old Infants Interpret Action in Context. *Psychological Science, 11*(1), 73–77. https://doi.org/10.1111/1467-9280.00218

Woodworth, R. S. (1899). Accuracy of voluntary movement. *The Psychological Review: Monograph Supplements, 3*(3), i–114. https://doi.org/10.1037/h0092992

Wurm, M. F., & Lingnau, A. (2015). Decoding Actions at Different Levels of Abstraction. *The Journal of Neuroscience, 35*(20), 7727–7735. https://doi.org/10.1523/JNEUROSCI.0188-15.2015

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science, 28*(6), 979–1008. https://doi.org/10.1207/s15516709cog2806_5

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*(1), 3–21. https://doi.org/10.1037/0033-2909.127.1.3

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience, 4*(6), 651–655. https://doi.org/10.1038/88486

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General, 130*(1), 29–58. https://doi.org/10.1037/0096-3445.130.1.29

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*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General, 138*(2), 307–327. https://doi.org/10.1037/a0015305

Zacks, J. M., Speer, Nicole K., Swallow, Khena M., & Maley, Corey J. (2010). The brain's cutting-room floor: Segmentation of narrative cinema. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00168

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction Error Associated with the Perceptual Segmentation of Naturalistic Events. *Journal of Cognitive Neuroscience, 23*(12), 4057–4066. https://doi.org/10.1162/jocn_a_00078

Zheng, Y., Zacks, J. M., & Markson, L. (2020). The development of event perception and memory. *Cognitive Development, 54*, 100848. https://doi.org/10.1016/j.cogdev.2020.100848

Ziaeetabar, F., Pomp, J., Pfeiffer, S., El-Sourani, N., Schubotz, R. I., Tamosiunaite, M., & Wörgötter, F. (2020). Using enriched semantic event chains to model human action prediction based on (minimal) spatial information. *PLOS ONE, 15*(12), e0243829. https://doi.org/10.1371/journal.pone.0243829

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Practices Statement The data and materials for this experiment will be made available upon reasonable request (participants did not consent to having their data made publicly available). Statistical analyses were performed using publicly available tools, including programs available here: https://github.com/ksasmita/esMethods. The experiment was not pre-registered.