Cognitive Psychology 63 (2011) 141–171

Contents lists available at SciVerse ScienceDirect

Cognitive Psychology

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

From movements to actions: Two mechanisms for learning action sequences

Ansgar D. Endress [a,*], Justin N. Wood [b]

[a] Massachusetts Institute of Technology, Cambridge, MA, United States
[b] University of Southern California, Los Angeles, CA, United States

ARTICLE INFO

Article history:
Accepted 7 July 2011

Keywords:
Statistical learning
Positional memory
Sequence learning
Actions
Goals
Intentions

ABSTRACT

When other individuals move, we interpret their movements as discrete, hierarchically-organized, goal-directed actions. However, the mechanisms that integrate visible movement features into actions are poorly understood. Here, we consider two sequence learning mechanisms – transitional probability-based (TP) and position-based encoding computations – that have been studied extensively in the domain of language learning, and investigate their potential for integrating movements into actions. If these learning mechanisms integrate movements into actions, then they should create memory units that contain (i) movement information, (ii) information about the order in which movements occurred, and (iii) information allowing actions to be recognized from different viewpoints. We show that both mechanisms retain movement information. However, only the position-based mechanism creates movement representations that are view-invariant and contain order information. The TP-based mechanism creates movement representations that are view-dependent and contain no order information. We therefore suggest that the TP-based mechanism is unlikely to play an important role for integrating movements into actions. In contrast, the position-based mechanism retains some of the types of information needed to represent goal-directed actions, which makes it an attractive target for further research to explore what, if any, role it plays in the perception of goal-directed actions.

* Corresponding author. Address: Massachusetts Institute of Technology, Department of Brain & Cognitive Sciences, Room 46-4127, 43 Vassar Street, Cambridge, MA 02139, United States.
E-mail address: ansgar.endress@m4x.org (A.D. Endress).

0010-0285/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2011.07.001

Humans are an extremely social species. Accordingly, we spend a considerable amount of time and energy thinking about other individuals' behavior, in contexts that range from cooperative action to resource competition. How do individuals quickly and accurately predict and interpret the intentions of others' actions?

To date, considerable progress has been made toward understanding the origins and nature of the capacity to infer the goals of others' actions. Studies of human infants show that core components of action understanding are present and functional early in development (e.g., Gergely & Csibra, 2003; Woodward, 1998); studies of nonhuman primates show that some of these same components are shared with closely and distantly related animal species (e.g., Call, Hare, Carpenter, & Tomasello, 2004; Range, Viranyi, & Huber, 2007; Wood, Glynn, Phillips, & Hauser, 2007); and studies employing cellular recordings and functional neuroimaging methods have begun to unite these psychological mechanisms with their neural substrates (e.g., Brass, Schmitt, Spengler, & Gergely, 2007; Perrett, Harries, Mistlin, & Chitty, 1990; Rizzolatti, Fogassi, & Gallese, 2001).

In order to interpret actions as either goal-directed or accidental, the observer must first segment the visual input into discrete actions. This process is complicated by the fact that actions can be embedded within a continuous flow of dynamic motion. In everyday action, human behavior tends to flow continuously, with few pauses to mark meaningful boundaries between distinct movements (Asch, 1952; Heider, 1958; Newtson & Engquist, 1976). Nevertheless, human behavior is hierarchically organized, consisting of distinct hierarchically-organized actions (see, among many others, e.g. Cooper & Shallice, 2006; Lashley, 1951; Norman & Shallice, 1986; Zacks & Swallow, 2007). Our perception of others' behavior seems to mirror this fact: we perceive hierarchically-organized actions rather than a continuous stream of movement. For example, if we observe an individual stand, walk to a table, and grasp a bottle of soda without pausing between each act, we represent three distinct acts, each of which accomplished a particular sub-goal needed to achieve the broader goal of obtaining the soda. Critically, how does the visual system segment the continuous flow of retinal information into discrete hierarchically-organized actions? Below, we refer to 'action segmentation' as the process by which continuous movement is segmented into discrete acts. We refer to 'action integration' as the process by which discrete acts are integrated into hierarchically-organized actions. All of our experiments test action segmentation, because, as we show below, our movement stimuli are not perceived as goal-directed actions. However, the goal of these studies is to provide a basis for future, more targeted studies on action integration, by evaluating two candidate mechanisms in terms of necessary properties that any mechanism of action integration must have.

Prior research on action segmentation shows that adults readily segment others' behavior into distinct actions. The majority of this research has used the procedure developed by Newtson (1973). Observers watch a movie and press a button to indicate their judgment about when one meaningful event ends and another event begins. This paradigm has produced a number of important findings. First, observers tend to identify boundaries that are readily namable "chunks" that correspond to sub-goals that an actor performs in order to fulfill the larger goal of the activity. Consequently, event segmentation is hierarchically structured, with fine-grained events clustered into larger course-grained events (see also Zacks & Swallow, 2007). For example, the larger event of "washing a car" might consist of several smaller fine-grained events such as "spraying the car with water," "scrubbing the car," and "drying the car." Second, event segmentation is automatic, in the sense that it occurs even when observers are not aware that they need to segment visual input into discrete events (Zacks et al., 2001). For example, Zacks and colleagues showed that during passive viewing of events, regions in the posterior and frontal cortex increase in activity several seconds before an event boundary and peak several seconds after the boundary. Third, adults generally agree about the boundaries separating distinct events (e.g., Zacks & Swallow, 2007).

These studies suggest that top-down, higher-level information about the sub-goals of an event facilitate action segmentation. For example, if an observer sees an individual washing a car, they may use prior knowledge about possible sub-goals to segment this event into smaller action units (e.g., "spraying the car with water," "scrubbing the car," and "drying the car"). In these cases, causal knowledge about the motions needed to achieve the goals plays a role in identifying the boundaries separating actions.

Other studies suggest that action segmentation also draws on bottom-up cues. For example, Newtson (1973) suggested several low-level cues that might indicate action boundaries, such as changes in direction, acceleration or the relative position of objects (see also Zacks, 2004; Zacks & Swallow, 2007). Other researchers investigated a different kind of bottom-up mechanism that detects and stores information about the structural regularities within motion (e.g., Baldwin, Andersson, Saffran, & Meyer, 2008). In human behavior, there are structured patterns to the flow of movements produced by actors (e.g., Newtson, Engquist, & Bois, 1977), and observers can use this information to track the probabilities of certain movements following other movements (Baldwin et al., 2008). In particular, Baldwin and colleagues (2008) presented observers with a novel string of small-scale intentional acts (e.g., *pour*, *poke*, *scrub*). Within the string of acts, four three-motion-element combinations were created and randomly intermixed within the string. After observing a 20-min string of continuous motion, observers were able to distinguish three-motion-element actions composed of individual acts that were likely to follow each other from other three-motion-element actions composed of acts less likely to follow each other. Baldwin et al. (2008) conclude that observers can use statistical learning mechanisms for action integration.

Building on Baldwin et al.'s (2008) work, we ask whether there are other bottom-up, sequence-learning mechanisms that might be used to parse continuous movement into goal-directed actions. Specifically, we derive three necessary functions that any mechanism used for parsing movements into actions must accomplish, and ask whether the statistical mechanism investigated by Baldwin et al. (2008), and a second mechanism that encodes the positions of elements in sequences (see below for more details) fulfill these requirements. We will now turn to these functions.

First, the mechanism must operate over movement sequences; if it does not, it cannot be used for either action segmentation or action integration. While this criterion seems obvious, it is more constraining than it might seem. As mentioned above, according to most models of action planning and perception (e.g., Cooper & Shallice, 2006; Lashley, 1951; Norman & Shallice, 1986; Zacks & Swallow, 2007), actions are represented hierarchically. The top level consists of higher-level goals, which have sub-goals at the next level, which can have subgoals in turn; the end points of this hierarchy (i.e., the leafs of the hierarchical tree) are commands to motor effectors (when actions are produced) or raw movements (when actions are perceived). Thus, a mechanism that integrates movements into actions must operate on this lowest level, that is, at the level of raw movements. Of course, it is possible that such a mechanism also combines different sub-goals into a higher-level goal, but given previous claims of the bottom-up nature of at least one of the mechanisms investigated here (Baldwin et al., 2008), we focus on their viability for combining raw movements into goal-directed actions.

Second, a sequence learning mechanism that supports action segmentation must contain information about the order in which movements occurred. In particular, the learning mechanism must sustain information about the causality of the movements, reflecting the fact that an effect *follows* its cause. For example, to move a lamp and turn it on, one would grasp the lamp, move it to the table, and then push the lamp's power button; to turn off the lamp and put it away, one would perform these same movements in the reverse order: one would push the lamp's power button, move it from the table, and contract the arm. A mechanism that is not sensitive to the temporal order of movements could not distinguish between these two actions, and thus would be unsuitable for action segmentation and integration.

That said, once goals are computed, there might be some flexibility in the order in which different sub-goals of a goal can be accomplished. For example, to make a phone call, one can (i) first get out the phone book, look up a number in the book, pick up the phone, and then dial the number; (ii) first pick up the phone, then get out the phone book, look up the number in the book, and then dial the number; or (iii) first get out the phone book, pick up the phone, look up the number in book, and then dial the number. Importantly, however, getting out a phone book, picking up a phone, etc. are all goal-directed actions; their goals are the sub-goals of the action "making a phone call." Critically, each of the sub-goals requires that movements be performed in specific orders. To get the phone book out of a desk, we cannot first grasp the phone book and then open the drawer where it is placed; similarly, to pick up the phone, we cannot first make a grasping movement and then extend the arm to reach the phone, and so on.

Third, a sequence learning mechanism that supports action segmentation must construct representations that allow actions to be recognized from different viewpoints. A sequence learning mechanism that could not recognize actions from different viewpoints would, of course, fail to support action recognition under all circumstances where the observer had not previously seen the action from that particular viewpoint.

It is important to note that these are just *necessary* conditions for integrating movements into goal-oriented actions. Here, we are just concerned with identifying potential candidate mechanisms; if any of the mechanisms investigated here fulfills all three necessary conditions, it is an important topic for further research to determine whether it actually extracts goal information. Conversely, if one of these mechanisms fails to fulfill these conditions, it is unlikely to be useful for integrating movements into actions.

In what follows, we review some important analogies between the domains of action perception and language, and then draw on recent artificial language learning experiments to ask whether two sequence learning mechanisms uncovered in these experiments may also be used for integrating movement sequences into actions. Specifically, it has been shown that humans can analyze continuous signals such as speech both according to the "transitional probabilities" between the elements in the signal, and according to certain cues in the signal that are used to encode the positions of the elements within the "units" they extract. Finally, we present the results of 14 new experiments that examine directly whether the sequence learning mechanisms that encode transitional probabilities and the positions of elements in a sequence satisfy the three necessary conditions for mechanisms used for action segmentation and integration.

1. Lessons from language

Action perception and word segmentation parallel one another in at least two ways, raising the possibility that research on word segmentation might inform our understanding of the mechanisms that support action perception. First, language consists of a continuous speech signal with no explicit pauses between words (e.g., Aslin, Saffran, & Newport, 1998; Saffran, Newport, & Aslin, 1996; Saffran, Aslin, & Newport, 1996); similarly actions consist of continuous movement information with no explicit pauses between movements (e.g., Asch, 1952; Heider, 1958; Newtson & Engquist, 1976). However, the continuous speech signal contains other, prosodic cues to word boundaries that can be perceived across languages (e.g., Brentari, González, Seidl, & Wilbur, 2011; Christophe, Mehler, & Sebastian-Galles, 2001; Endress & Hauser, 2010; Fenlon, Denmark, Campbell, & Woll, 2008; Pilon, 1981); likewise, there are other cues to action boundaries present in movements as well (e.g., Newtson & Engquist, 1976; Zacks, 2004; Zacks & Swallow, 2007). Hence, just as listeners need to individuate words in fluent speech, observers need to individuate others' actions in a continuous movement sequence, and it is an open question which cues and mechanisms can be used for this purpose.

Second, both language and action are organized hierarchically. While actions are observed in the form of movement sequences from which observers need to derive hierarchically organized action plans (e.g., Cooper & Shallice, 2006; Lashley, 1951; Norman & Shallice, 1986), language is heard in the form of sound sequences from which listeners need to derive (or perceive) the underlying hierarchical linguistic structure (e.g., Chomsky, 1980, 1981; Feldman, Goldin-Meadow, & Gleitman, 1978; Pinker, 1984). In both cases, the representations of words and actions thus need to be compatible with the hierarchical representations in their respective domains, which may, in turn, constrain the mechanisms that can be used to learn words and segment actions.

We will now discuss the first of these similarities between language and action in more detail, and explain how different mechanisms that track units in continuous input can be isolated from one another. In the general discussion, we return to the issue of the hierarchical nature of action representation, and discuss how this might constrain the mechanisms used to segment movements into actions.

1.1. Two mechanisms for segmenting units from fluent speech

It has long been recognized that words in fluent speech are not separated by silences akin to white space in written text. Hence, to learn the words of one's native language, one needs mechanisms that recover word boundaries. A well accepted candidate mechanism relies on "transitional probabilites" (TPs). The basic idea is that syllables that are part of the same word tend to co-occur more frequently than syllables that do not occur in the same word; TPs thus indicate how likely it is that two syllables will follow each other. More formally, TPs are conditional probabilities of encountering a syllable after having encountered another syllable. Conditional probabilities like $P(\sigma_{i+1} = \text{pet} \mid \sigma_i = \text{trum})$ (in the word trumpet) are high within words, and low between words (σ denotes syllables in a speech stream). Dips in TPs may give cues to word boundaries, while high-TP transitions may indicate that words continue. That is, learners may postulate word boundaries between syllables that rarely follow each other.

There is extensive evidence that very young infants can deploy TP computations over fluent speech (e.g., Aslin et al., 1998; Saffran et al., 1996; Saffran, 2001). The availability of TP computations is typically assessed by exposing participants to continuous speech streams where TPs are the only cues to word-boundaries. Subsequently, they have to choose between items with stronger TPs and items with weaker TPs; typically, participants choose items with stronger TPs. (In experiments with infants, participants' "choices" are measured using looking-time methods.) Moreover, TP computations have been observed with a wide range of non-linguistic stimuli in humans (e.g., Baldwin et al., 2008; Fiser & Aslin, 2001, 2002a; Saffran, Johnson, Aslin, & Newport, 1999; Saffran & Griepentrog, 2001; Turk-Browne, Jungé, & Scholl, 2005; Tillmann & McAdams, 2004; Toro, Sinnett, & Soto-Faraco, 2005; Turk-Browne & Scholl, 2009), and with speech stimuli in non-human animals (Hauser, Newport, & Aslin, 2001; Toro & Trobalón, 2005).

The wide availability of such computations raises the possibility that they can also be deployed on movement sequences, and, in fact, previous experiments seem to suggest so (Baldwin et al., 2008). However, in these experiments the movement stimuli were actions performed on objects (e.g., drinking, poking, stacking). Given that humans (and many other animals) are equipped with early developing mechanisms that infer other agents' goals when observing object-directed actions (e.g., Gergely & Csibra, 2003; Wood et al., 2007), it is unclear whether observers in Baldwin et al.'s (2008) experiments tracked the TPs of movements or the TPs of object-directed actions. In the experiments presented below, we avoid this problem by using animated actions with no obvious goals that, as we will show empirically, are *not* perceived as goal-directed.

While the aforementioned results suggest that TPs can be computed in a variety of domains, it is less clear whether TP-based mechanisms allow observers to perceive syllable sequences as integrated units. This is important because words are integrated units. Take Endress and Mehler's (2009b) experiments as an example. As in standard TP experiments, these authors familiarized participants with continuous speech streams where TPs were the only cues to word-boundaries. However, the streams were constructed such that (statistically defined) "words" had identical TPs to "phantom-words" that, in contrast to words, never occurred in the stream. Surprisingly, participants were unable to decide whether they had heard words or phantom-words even after hearing each word 600 times. Moreover, they were more familiar with phantom-words (which they had not heard) than with items that did occur in the speech stream but had weaker TPs. These results suggest that, while individuals deploy TP computations over continuous speech streams, they do not use these computations to extract integrated units (such as words). If similar problems arise when TPs are deployed on movement sequences, then this would significantly limit the utility of TP computations for reconstructing actions from movement sequences, because actions are integrated "units" composed of movements.

If TP-based computations fail to learn integrated units, how do listeners learn to extract words from continuous speech streams? Results from artificial grammar learning experiments suggest that listeners can use, in parallel with TP-computations, a second sequence-learning mechanism to segment and organize fluent speech, and this mechanism might be capable of learning integrated units. Similar to the operating principles of certain short-term memory mechanisms (e.g., Henson, 1998; Hitch, Burgess, Towse, & Culpin, 1996; Ng & Maybery, 2002), this mechanism represents units in terms of the positions of the elements within the unit (Endress & Bonatti, 2007; Endress & Mehler, 2009a). For example, if the non-word *puliki* is represented as a unit, this mechanism would encode that *pu* occurred in the first position, *ki* in the last position, and that *li* occurs between the first and the last position. That is, this mechanism represents the positions of the elements of the units *relative* to the first and the last position (that is, the unit edges).

There is an important difference between TP-based mechanisms and the encoding of positions relative to edges, however: the positional encoding requires units to be delimited by some cues. At first sight, this difference seems to render the positional encoding useless for extracting words from fluent speech, since there are no silences between words (e.g., Aslin et al., 1998; Saffran et al., 1996; Saffran et al., 1996). However, as mentioned above, even though words are not separated by silences, some boundary cues come from the prosodic organization of language. In fact, listeners are sensitive to cues to boundaries of prosodic constituents, even in languages they have never heard (e.g., Brentari et al., 2011; Christophe et al., 2001; Endress & Hauser, 2010; Pilon, 1981). Hence, there seem to be at least some cues to word boundaries in the speech signal. Interestingly, prosodic information reestablished a sensitivity to the items people had actually heard in Endress and Mehler's (2009b) aforementioned experiments with phantom-words: even a minimal prosody-like cue allowed them to discriminate actual items from phantom-words, suggesting that explicit boundary information such as that provided by prosody might allow listeners to perceive words as integrated units.

Given that observers can detect action boundaries, movement sequences might also present cues to action boundaries. For example, when observers segment movies into events, the boundaries are predicted by changes in movement parameters, including the acceleration of the objects and their location relative to one another (e.g., Hard, Tversky, & Lang, 2006; Zacks, 2004). In this paper, we therefore make no assumptions about the kinds of cues that are available to extract units from fluent movement signals, though we note that some cues clearly are available. Rather, we focus on the informational content of the segmented units produced by TP-based and position-based mechanisms.

1.2. A situation for testing positional and TP-based mechanisms

If both TP-based and position-based computations operate over continuous signals, how can the characteristics of these systems be isolated from one another? Crucially, in the domain of language processing, researchers have been able to isolate these mechanisms (Endress & Bonatti, 2007; Endress & Mehler, 2009a; see also Peña, Bonatti, Nespor, & Mehler, 2002). The current experiments are based on these prior experiments; thus we will now review them in some detail.

In these studies, participants were told that they would listen to an artificial language, and were then presented with a continuous speech stream. Following this speech stream, they were presented with pairs of speech items, and had to decide which of the items was more likely to be part of the language they had listened to. The speech stream was composed of tri-syllabic words. The first syllable predicted the last syllable with certainty, while the middle syllable was variable. That is, words had the form A_iXB_i. The first and the last syllable were chosen from one of three frames of the form $A_i...B_i$; the middle syllable (represented by X) was then chosen from a different syllable set. By construction, these words therefore implemented a positional regularity: certain syllables had to occur in the first and in the last position, respectively. These words also implemented a TP-based regularity in two different ways. First, the deterministic dependency between the first and the last syllable implied that the TP between the first and the last syllable was always 1.0. Second, because there were only three different frames and three different X syllables, participants could also track the TPs between adjacent syllables in the word.

To investigate both TP-based and position-based computations, these authors used three types of test items: rule-words, class-words and part-words. (These labels were motivated by Peña et al.'s (2002) experiments, but these motivations are irrelevant for the current purposes. However, we will keep these labels for consistency with the previous literature.) The main item types are shown in Table 1. Rule-words conserved the $A_i\ldots B_i$ frames; however, the middle syllable in rule-words had never occurred in this position during the familiarization stream, and was in reality an A or a B syllable. Rule-words thus had the form $A_iX'B_i$; they conformed to the positional regularity (because the "correct" syllables occurred initially and finally, respectively), they implemented the TP-relation between the first and the last syllable, but the TPs between their adjacent syllables was zero.

Class-words had the form $A_iX'B_j$. That is, their first and their last syllable had occurred in these positions during familiarization, but never in the same word because they came from different frames; as in rule-words, the middle syllable was an A or a B syllable. Class-words thus implemented the positional regularity, but TPs between all of their syllables were zero.

Table 1
Summary of the main test item types used by Peña et al. (2002), Endress and Bonatti (2007) and Endress and Mehler (2009a).

Words	Part-words	Rule-words	Class-words
Items used by Peña et al. (2002) and Endress and Bonatti (2007)			
$\mathbf{A}_i X \mathbf{C}_i$	$\mathbf{C}_i \vert \mathbf{A}_j X$ or $X \mathbf{C}_i \vert \mathbf{A}_j$	$\mathbf{A}_i X' \mathbf{C}_i$	$\mathbf{A}_i X' \mathbf{C}_j$
Items used by Endress and Mehler (2009a)			
Edge Condition			
$A_i XYZ \mathbf{C}_i$	$YZ \mathbf{C}_i \vert \mathbf{A}_j X$ or $Z \mathbf{C}_i \vert \mathbf{A}_j XY$	$\mathbf{A}_i X' Y' Z' \mathbf{C}_i$	$\mathbf{A}_i X' Y' Z' \mathbf{C}_j$
Middle condition			
$X \mathbf{A}_i Y \mathbf{C}_i Z$	$\mathbf{C}_i Z \vert X \mathbf{A}_j Y$ or $Y \mathbf{C}_i Z \vert X \mathbf{A}_j$	$X' \mathbf{A}_i Y' \mathbf{C}_i Z'$	$X' \mathbf{A}_i Y' \mathbf{C}_j Z'$
Explanation			
Appear in the stream $TP(A_i \rightarrow C_i) = 1$	Appear in the stream but straddle a word boundary[a]	As words, but with new X, Y and Z syllables	As rule-words, but with first and last syllable from different families
	Violate dep. betw. 1st and last syll.	Respect dep. betw. 1st and last syll.	Violate dep. betw. 1st and last syll.
	Violate positional regularity	Respect positional regularity	Respect positional regularity

[a] The location where the word boundaries fell during familiarization is shown by |; no boundaries were present in the test items.

Part-words were trisyllabic items straddling a word boundary. That is, they were constructed either by taking the last two syllables from one word and the first syllable from the next word, or by taking the last syllable from the first word and the first two syllables from the next word. Hence, they had either the form XB_iA_j or the form B_iA_jX. As a result, they violated the positional regularity, but the TPs between their syllables were positive.

Pitting rule-words against class-words thus assesses how well participants learned the TP-relation between the first and the last syllable in words. Pitting class-words against part-words pits positional information against TPs. Endress and Bonatti (2007) showed that TP-based computations and position-based computations have fundamentally different properties. First, while participants could track TPs among syllables on fluent speech, the position-based computations operated only when words were separated by imperceptible 25 ms silences. Second, once the silences were available, the position-based computations operated on a much faster time scale than the TP-based mechanism. For example, participants preferred class-words to part-words after a 2-min exposure to a speech stream; in contrast, after a 60-min exposure to the speech stream participants preferred part-words to class-words. These results suggest that the positional information in class-words "wins" over the TP-information in part-words after shorter familiarization durations, but that the relative weight of these cues reverses after longer familiarization durations. Third, when using longer, five-syllable words, Endress and Mehler (2009a) showed that syllable positions can be tracked for word-initial and word-final syllables, while it was much harder to track word-medial positions; TPs, in contrast, operated fairly well also on word-medial syllables. Together, these results thus suggest that position-based computations and TP-based computations create different types of representations, and have different properties.

2. The current experiment: three requirements for reconstructing actions from movement sequences

As discussed above, segmenting words in fluent speech and segmenting actions when observing dynamic movement present the learner with similar problems. In the experiments described below, we will therefore investigate two mechanisms that have been linked to finding words in fluent speech – those tracking TPs and positional information, asking whether these mechanisms would be suitable for learning actions from movement sequences. Specifically, we will asses three criteria that, we believe, are necessary for integrating movements into actions: such mechanisms must operate over movement sequences, respect causality (i.e., the temporal order of movements), and create units that can be recognized from different viewpoints. There is a fourth important criterion that we do not address here: both words and actions need to be compatible with the hierarchical representations in their respective domains. We will come back to this issue in the general discussion, evaluating the potential of these mechanisms in the light of previous experimental work and theoretical considerations, especially from formal linguistics. As mentioned above, these criteria are just *necessary* criteria. If either of the mechanisms fulfill all three necessary conditions, then further research would be needed to determine whether the mechanism actually extracts goal information.

The experiments are shown in Table 2. In all experiments, participants were told that they would see a dancer training for a ceremony, repeating ceremonial movement sequences. Then, they were presented with a stream of movements performed by an animated agent. Unbeknownst to the participants, the stream was a concatenation of triplets of movements, analogous to the syllable triplets used in the speech experiments reviewed above. Following this familiarization, they were presented with pairs of test triplets, and had to decide which triplet was more likely to be a ceremonial movement sequence.

2.1. Tracking movement sequences

Experiments 1–5 investigate a basic requirement for a mechanism that integrates movement sequences into actions: such a mechanism needs to operate on movement sequences. In these experiments, we will therefore replicate some key experiments by Endress and Bonatti (2007) and

Table 2
Summary of the experiments.

Experiments	Familiarization sequence	Test items	Tests for[a]	Preference for
1	Continuous	Rule-triplets vs. class-triplets	TPs	Rule-triplets
2	Segmented	Class-triplets vs. part-triplets	PRs vs. TPs	Class-triplets
3	Continuous	Class-triplets vs. part-triplets	PRs vs. TPs	Part-triplets
4	Segmented	Class-triplets vs. part-triplets (Critical mvts. at unit-edges)	PRs vs. TPs	Class-triplets
5	Segmented	Class-triplets vs. part-triplets (Critical mvts. in unit-middles)	PRs vs. TPs	None
6	Continuous	Rule-triplets vs. class-triplets (both items reversed)	Reverse TPs	Rule-triplets
7	Segmented	Class-triplets vs. part-triplets (both items reversed)	Reverse PRs vs. reverse TPs	Part-triplets
8	Segmented	Class-triplets vs. part-triplets (only class-triplets reversed)	Reverse PRs vs. forward TPs	None
9	Continuous	Rule-triplets vs. class-triplets (both items rotated)	TPs across viewpoint change	None
10a	Continuous	Rule-triplets vs. class-triplets (Replication of Exp. 1 with longer familiarization)	TPs	Rule-triplets
10b	Continuous	Rule-triplets vs. class-triplets (Replication of Exp. 9 with longer familiarization)	TPs across viewpoint	None
11	Segmented	Class-triplets vs. part-triplets (both items rotated)	PRs vs.TPs across viewpoint change	Class-triplets
12a	Segmented	Class-triplets vs. part-triplets (both items rotated)	PRs vs. TPs across viewpoint change	Class-triplets
12b	Continuous	Class-triplets vs. part-triplets (both items rotated; replication of Exp. 11)	PRs vs. TPs across viewpoint change	None

[a] TP: transitional probability; PR: positional regularity.

Endress and Mehler (2009a), asking whether the TP-based mechanism and the position-based mechanism exhibit the same characteristics as in the speech domain when deployed on movement sequences. In Experiment 1, we ask whether participants can deploy TP-based computations over movement sequences. Experiments 2 and 3 investigate whether participants can track the positions of movements in sequences and, if so, whether they require segmentation markers analogous to the 25 ms needed for language processing (Endress & Bonatti, 2007). Experiments 4 and 5 examined whether position-based computations are observed preferentially in edge-positions (that is, in the first and the last position), or also in middle positions.

In contrast to the position-based mechanism, the TP-based mechanism was studied only after familiarization to movies without segmentation markers. From Saffran et al.'s (1996) pioneering work onwards, TP-based computations have been assumed to depend on a mechanism that extracts units from continuous input. Indeed, this is one of this learning mechanism's most attractive properties and it is the main motivation for using TPs in the context of speech segmentation. Hence, a TP-based mechanism of the sort described by Saffran et al. (1996) and others must fulfill all of our necessary conditions in the absence of explicit segmentation markers. Conversely, position-based computations have been hypothesized to rely on explicit boundary cues such as prosodic information or, in the present experiments, gaps in motion. Hence, to ensure that learning depended on the position-based mechanism, it was necessary to demonstrate that it operates when gaps in motion are present in the input stream, but not in the face of unsegmented input.

Further, TP-based mechanisms are typically assessed using TPs between adjacent items (e.g., Saffran et al., 1996). Here, however, we test TPs between non-adjacent movements, because, as we will show below, they provide a similar baseline to the position-based computations, and because this allowed us to follow the design by Endress and Bonatti (2007) and Endress and Mehler (2009a) more closely.

2.2. Respecting causality

In Experiments 6–8, we asked whether TP-based and position-based computations respect causality. For (goal-directed) actions, temporal order is of utmost importance, since the order in which movements are performed can lead to entirely different outcomes. In Experiments 6 and 7, we familiarized participants with streams of movements under conditions in which it was possible to use either TP-based or position-based computations, as determined in Experiments 1–5. In these experiments, however, the test movements were presented in reverse order. Experiment 8 controls for an alternative interpretation of Experiment 7.

2.3. Invariance under rotation

Finally, in Experiments 9–12, we investigate whether TP-based and position-based computations produce action representations that can be recognized from different viewpoints. We familiarized participants with the streams of movements under conditions in which it was possible to use either TP-based or position-based computations, as determined in Experiments 1–5. In these experiments, however, the test movements were observed from a different viewpoint (the agent performing the movements was rotated 90°).

3. Tracking movement sequences

Experiments 1–5 ask whether participants can track (i) the TPs between movements and (ii) the positions of movements in actions. These experiments replicate studies performed with speech sequences by Endress and Bonatti (2007) and Endress and Mehler (2009a). Experiment 1 asks whether participants can track TPs among movements in a continuous movement sequence; Experiments 2 and 3 ask the same question about positional regularities, and whether such regularities can be tracked only when explicit segmentation cues are given. Finally, Experiments 4 and 5 ask whether positional regularities are tracked predominantly at sequence edges.

3.1. Experiment 1: tracking TPs over movements

Experiment 1 asked whether participants can track TPs over movements in a continuous movement sequence. Participants were familiarized with movement streams composed of movement triplets (see Supplementary Movies for sample clips of the familiarization sequences and test items). As in the prior experiments with speech stimuli, triplets had the form A_iXB_i, where the first and the last movement was taken from one of three frames, and the middle movement was taken from a different set of movements. Following this familiarization, participants were presented with the analogues of rule-words and class-words (hereafter rule-triplets and class-triplets), and were instructed to decide which of these items was more likely to be part of the familiarization stream. As mentioned above, the only difference between rule-triplets and class-triplets is that rule-triplets implement the TP-relation between the first and the last movement in a triplet; hence, if participants choose rule-triplets over class-triplets, they must have tracked the TPs among movements in the continuous movement stream.

3.1.1. Materials and method

3.1.1.1. Participants. Twenty participants (14 females, 6 males, mean age 19.5 range 18–28) took part in this experiment for course credit or monetary compensation. Unless otherwise stated, participants were recruited through the Harvard University study pool, were native speakers of American English, and reported normal or corrected-to-normal vision with no known auditory impairment. Each participant took part in only one experiment.

3.1.1.2. Apparatus. The experiment was run using Psyscope X (http://psy.ck.sissa.it). Responses were collected on pre-marked keys on a keyboard.

3.1.1.3. Materials. Individual movements were prepared using Poser 6 software (Smith Micro Software, Inc., Aliso Viejo, CA). From the front viewpoint, the figure subtended 10.5° (height) × 4° (width) in the center of a video monitor. We used nine highly discriminable movements (see Fig. 1 for a static depiction of each movement). All movements were dynamic (i.e., they involved fluid, continuous movement, rather than being presented as static pictures). Previous studies used these same types of movements to characterize the storage capacity of visual working memory for actions (Wood, 2007, 2008), showing that observers can maintain 2–3 integrated action representations at once.

Care was taken that individual movements had no obvious verbal labels or goals (see Supplementary Movies). To ensure that participants would not perceive the movements as goal-directed, we evaluated our stimuli in the following way. Participants ($N = 10$) were told that the movements were part of a dance. For each movement, they were then asked to rate on a scale from 1 to 7 whether the movement was intended to accomplish a goal, '1' indicating that it definitely was not, '7' indicating that it definitely was, and '4' indicating no preference. In addition, we also recreated the actions used by Baldwin et al. (2008) using video recordings of a live actor, and asked participants to rate these movies. Results showed that participants did *not* perceive our movements as goal-directed, with an average rating significantly below 4, ($M = 3.28$, $SD = 0.8$), $t(9) = 2.7$, $p = 0.03$, Cohen's $d = 0.85$, $CI_{.95} = 2.68$, 3.89. In contrast, in line with the finding that actions on objects are likely to be seen as goal directed (e.g., Rizzolatti et al., 2001), Baldwin et al.'s (2008) actions were seen as significantly more goal-directed than ours, $F(1,9) = 7.2$, $p = .03$, $\eta_p^2 = .444$, although, at least on our scale, the ratings did not differ significantly from 4 ($M = 4.2$, $SD = 1.4$), $t(9) = .5$, $p = .599$, Cohen's $d = .17$, $CI_{.95} = 3.3$, 5.2.

Each movement started and ended in the same body posture (hereafter the "neutral" position); this allowed us to concatenate the individual movements. Each movement had a duration of 0.58 s. Movements were concatenated using the catmovie utility from the QTCoffee package (http://www.3am.-pair.com/QTCoffee.html). The concatenation was saved using the H.264 codec and the mov container format with a frame rate of 29.97 frames/s. Each sample had a size of 320 × 240 pixels. The resulting movie was faded in and out over a duration of 5 s using iMovie HD (Apple, Inc., Cupertino, CA).

3.1.1.4. Familiarization. Participants were told that they would see a dancer training for a ceremony. Following this, they were presented with a concatenation of the movements described above.

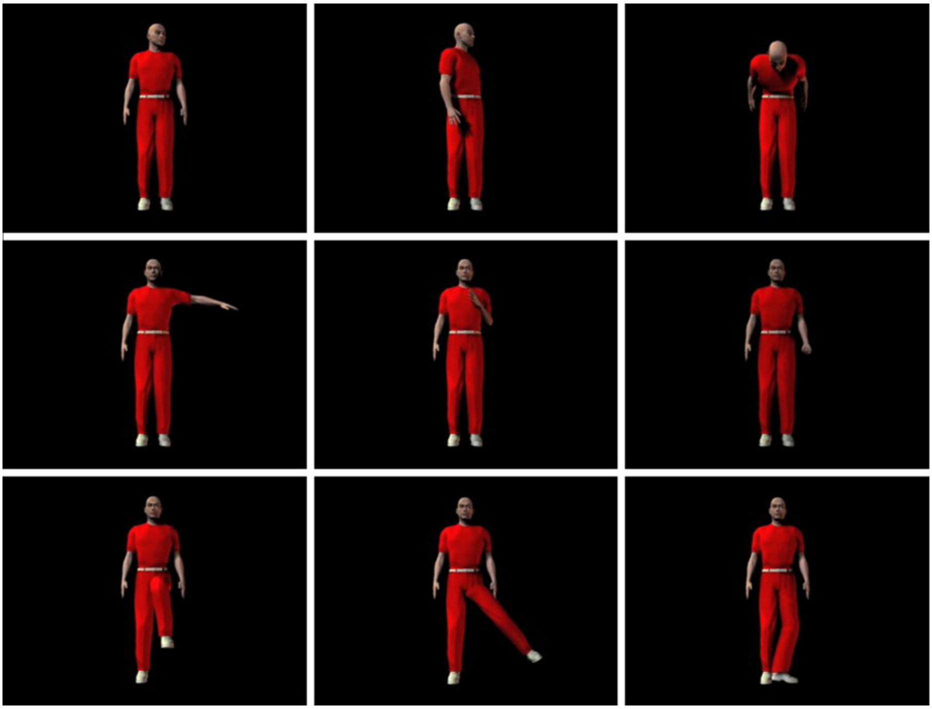

Fig. 1. Static depiction of the movements used in Experiments 1–3 and 6 and 7. The images show the movements' maximal deviation from the neutral position. Sample clips are available as Supplementary Material.

Individual movements were combined into triplets. As in Endress and Bonatti's (2007) experiments, triplets had the form A_iXB_i. That is, the first and the last movement came from one of three frames; the middle movement was chosen from another three movements, yielding a total of nine triplets. Depictions of the individual movements are shown in Fig. 1. Care was taken that one movement in each position (first, second or third) involved the arms, one the legs, and one the rest of the body.

Triplets were then randomly arranged in a sequence with the constraint that two consecutive triplets could not share a movement. Each triplet was presented 60 times, for a total familiarization duration of 5 min 10 s. TPs between adjacent movements within triplets were 0.333, and 0.5 between triplets. (Recall that consecutive triplets could not share any movements; hence, each frame could be followed only by one of the other two frames, leading to a TP of 0.5.) Second-order TPs were 1.0 within triplets and 0.333 between triplets.

3.1.1.5. Test. Following familiarization, participants were informed that they would see pairs of movement sequences, and that they had to decide which sequence in each pair was more likely to be a ceremonial sequence.

Following this, they were presented with pairs of movement triplets. One triplet in each pair was always a rule-triplet, while the other was a class-triplet. As mentioned above, rule-triplets had the form $A_iX'B_i$. That is, their first and their last movement came from one of the frames observed during familiarization, but their middle movement had never occurred in that position and was in reality an A or B movement. Care was taken that the middle movement did not come from the same frame as the frame of the rule-triplet. Class-triplets, in contrast, had the form $A_iX'B_j$; that is, they were identical to rule-triplets except that their first and last movements came from different frames. Hence, the crucial difference between rule-triplets and class-triplets was that rule-triplets but not class-triplets respected the TP-relation between the first and the last movement of the frames.

Participants were presented with eight test pairs twice, once with the rule-triplet first and once with the class-triplet first. Half of the triplets overlapped in their first two movements, and half overlapped in their last two movements. Test pairs were presented in random order, with the constraints that (i) no more than three pairs in row could start with a rule-triplet or a class-triplet, and (ii) no more than three pairs in a row could overlap in their first two movements or their last two movements.

3.1.2. Results and discussion

As shown in Fig. 2, participants preferred rule-triplets to class-triplets (preference for rule-triplets: $M = 57.5\%$, $SD = 11.4\%$), $t(19) = 2.94$, $p < 0.01$, Cohen's $d = 0.66$, $CI_{.95} = 52.2\%, 62.83\%$. (All statistical tests reported in the manuscript are two-tailed with a chance level of 50% and a significance threshold of 0.05.) Hence, participants were sensitive to the TP-relation between the first and the last movement in a triplet.

3.2. Experiment 2: tracking positional regularities in movement sequences with segmented input

While Experiment 1 established that observers can track TPs with movement sequences, the goal of Experiment 2 was to establish whether a similar ability exists for tracking positional information. Experiment 2 was similar to Experiment 1 with two exceptions. First, the familiarization movie was segmented, that is, after each triplet, the actor remained in the neutral position for 1.2 s, corresponding to the silences between words in Endress and Bonatti's (2007) and Endress and Mehler's (2009a) experiments.[1] Second, during test, participants had to choose between class-triplets and part-triplets, directly pitting the positional regularity against TPs.

3.2.1. Materials and method

3.2.1.1. Participants. Twenty new participants (9 females, 11 males, mean age 21.6 range 18–32) took part in this experiment for course credit or monetary compensation.

3.2.1.2. Familiarization. The familiarization was identical to that in Experiment 1, except that the actor remained in the neutral position for 1.2 s after each triplet.

3.2.1.3. Test. Participants had to choose between class-triplets and part-triplets. As mentioned above, class-triplets have "legal" initial and final movements, but the TPs between their movements are zero. Hence, they follow the positional regularity, but not the TP-based regularity. Part-triplets, in contrast, are sequences that occurred during the familiarization movie, but straddled a triplet boundary. That is, they have either the form XB_iA_j, taking the last two movements from one triplet and the first movement from the next triplet, or the form B_iA_jX, taking the last movement from the first triplet and the first two movements from the next triplets. Part-triplets thus have positive TPs between their movements, but they violate the positional regularity.

Participants were presented with 12 test pairs twice, once with the class-triplet first and once with the part-triplet first. The two part-triplet types (BAX and XBA) were equally represented in the test pairs. Test pairs were presented in random order with the constraints that (i) no more than three pairs in a row could start with a class-triplet or a part-triplet, and that (ii) no more than three pairs in a row could have the same part-triplet type.

[1] Pilot experiments showed that having the actor remain in the neutral position for 25 ms was not sufficient for triggering the positional computations. There are several possible reasons for this. First, the movements are much longer than the syllables used in previous experiments. Second, the temporal precision in audition is much greater than in vision. Third, all movements had the same starting and end point, so the agent's stationary posture in the neutral position might have been partially perceived as part of the movements. Fourth, we did not use other, naturally occurring cues to action boundaries such as changes in the acceleration or location of the movements (Zacks, 2004). However, we are not aware of empirical data on the distribution of pauses in naturalistic movements, making it difficult to decide whether or not our segmentation markers were long relative to natural marker durations.

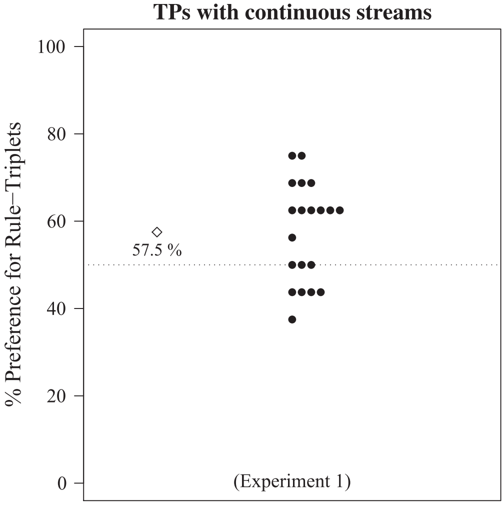

Fig. 2. Results of Experiment 1. Dots represent the means of individual participants, the diamond the sample average, and the dotted line the chance level of 50%. After familiarization with a continuous movement sequence, participants were sensitive to TPs between movements, preferring rule-triplets to class-triplets.

3.2.2. Results and discussion

As shown in Fig. 3, participants preferred class-triplets to part-triplets ($M = 57.7\%$, $SD = 11.7\%$), $t(19) = 2.94$, $p < 0.01$, Cohen's $d = 0.66$, $CI_{.95} = 52.2\%, 63.2\%$. Hence, participants tracked positional information; they noticed that certain movements had to occur triplet-initially, and others triplet-finally.

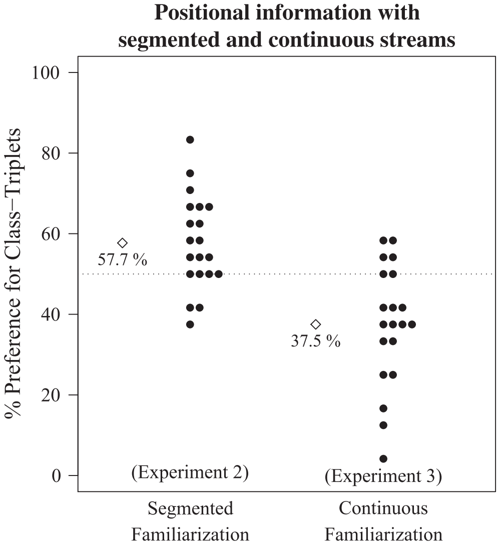

Fig. 3. Results of Experiments 2 and 3. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. Participants preferred class-triplets to part-triplets, showing a sensitivity to positional information, when familiarized with a movie where triplets were separated by stasis (no motion); in contrast, they preferred part-triplets to class-triplets when familiarized with a continuous movies, suggesting that participants require explicit segmentation marks for tracking positional information.

3.3. Experiment 3: tracking positional regularities of movement sequences with continuous input

In Experiment 2, participants tracked positional information after familiarization with a segmented familiarization movie. If the mechanism that tracks positional information for movement sequences is similar to the mechanism that tracks positional information for speech sequences, then we would expect this ability to disappear when participants are familiarized with a continuous movement stream. Experiment 3 tested this prediction by replicating Experiment 2, but with a continuous rather than segmented familiarization movie.

3.3.1. Materials and method

Experiment 3 was identical to Experiment 2 except that the familiarization movie was continuous. That is, triplets directly followed each other, without the intervening 1.2 s of stasis. Twenty new participants (9 females, 11 males, mean age 21.6 range 18–35) took part in this experiment for course credit or monetary compensation.

3.3.2. Results and discussion

As shown in Fig. 3, participants preferred part-triplets to class-triplets (preference for class-triplets: $M = 37.5\%$, $SD = 15.1\%$), $t(19) = 3.71$, $p = 0.001$, Cohen's $d = 0.83$, $CI_{.95} = 30.5\%, 44.6\%$. The preference for class-triplets was higher in Experiment 2 than in Experiment 3, $F(1,38) = 22.4$, $p < 0.0001$, $\eta^2 = 0.371$.

In contrast to Experiment 2, where participants preferred class-triplets over part-triplets, participants in Experiment 3 preferred part-triplets over class-triplets. These results suggest that, when segmentation cues are available, the positional regularity implemented in class-triplets is stronger than the TP-based regularities carried by part-triplets; in contrast, when no segmentation cues are available the positional information does not seem to be encoded, and, accordingly, participants choose part-triplets because they are favored by TPs.[2]

3.4. Experiment 4: tracking positional regularities at edges

In line with research on positional memory in short-term memory (see, among many others, Henson, 1998; Hitch et al., 1996; Ng & Maybery, 2002), Endress and Mehler (2009a) found that positional information is tracked predominantly in the first and the last position of words (that is, at the word edges) as opposed to word-medial positions. Experiments 4 and 5 ask whether the mechanism responsible for tracking positional information within movement sequences behaves similarly.

Experiments 4 and 5 are similar to Experiment 2 except that the movements were presented in movement quintuplets instead of triplets. This allowed us to ask whether participants track positional information for the first and the last movement in a quintuplet as well as for the second and the fourth movement in a quintuplet. If the position-based mechanism predominantly tracks positions at unit edges, we would expect participants to perform better in Experiment 4 (where they have to track positional information at the edges of quintuplets) than in Experiment 5 (where they have to track positional information in the middles of quintuplets).

Specifically, participants were familiarized with a concatenation of quintuplets of the form $\mathbf{A}_i$XYZ-$\mathbf{B}_i$; quintuplets in the familiarization movie were separated by 1.2 s of stasis, during which the agent remained in the neutral position. In each quintuplet, the first movement predicted the last movement with certainty, while the middle movements were variable. The structure of the items was thus analogous to that used in Experiment 2, except that they had three instead of one middle movements. As in Experiment 2, the crucial 'A' and 'B' movements were located at item edges. Following this familiarization, participants had to choose between "class-quintuplets" and "part-quintuplets".

[2] While participants in Endress and Bonatti's (2007) experiments did not prefer part-words to class-words after a familiarization with a continuous speech stream, those results are in line with these obtained here. Since the individual movements were much longer than individual syllables, the TP-based associations among movements might be stronger than the associations among syllables. In line with this interpretation, participants prefer part-words to class-words after familiarizations with continuous streams when longer words are used (Endress & Mehler, 2009a).

3.4.1. Materials and method

3.4.1.1. Participants. Twenty new participants (14 females, 6 males, mean age 19.3 range 18–22) took part in this experiment for course credit or monetary compensation.

3.4.1.2. Familiarization. This design required increasing the number of movement elements from 9 to 17 (see Fig. 4 for a static depiction of each movement).

Participants were familiarized with a concatenation of movement quintuplets with the structure $\mathbf{A}_i$XYZ$\mathbf{B}_i$, where the ‘A’ and ‘B’ movements belonged to four different $A_i \cdots B_i$ frames. The ‘X’, ‘Y’ and ‘Z’ movements were filler movements (similar to the ‘X’ movement in Experiment 2). Three movements could appear in each of the ‘X’, ‘Y’ and ‘Z’ positions. To limit the number of possible quintuplets, each ‘X’ movement could only precede *two* of the three ‘Y’ movements; likewise, each ‘Y’ movement could only be followed by two ‘Z’ movements. The TPs between ‘X’ and ‘Y’ movements and between ‘Y’ and ‘Z’ movements were 0.5, the TPs between ‘A’ and ‘X’ movements were 0.33, and the TPs between ‘Z’ and ‘B’ movements were 0.25; TPs across quintuplet boundaries were 0.33. Concerning the higher order TPs, the first movement predicted the last movement with certainty (i.e., the corresponding TP was 1.0) while the other TPs were much smaller. The familiarization movie contained two repetitions of each of the 48 quintuplets; no movement could appear in two consecutive quintuplets. Quintuplets were separated by 1.2 s of stasis.

3.4.1.3. Test. Following this familiarization, participants had to choose between class-quintuplets and part-quintuplets. Class-quintuplets had the form $A_iX'YZ'B_j$. As A_i and B_j belonged to different frames, the TPs between the first and the last movement were zero (instead of 1.0 during the familiarization); still, these movements appeared in the positions in which they had been encountered during familiarization (that is, in the first and the last position in quintuplets, respectively), and thus conformed to the positional regularity. The movements X′ and Z′ had never appeared in their respective positions during familiarization but were ‘A’ or ‘B’ movements; hence, class-quintuplets could have one of the following structures: $A_iA_kYA_lB_j$, $A_iA_kYB_lB_j$ and $A_iB_kYA_lB_j$, $A_iB_kYB_lB_j$, which were equally represented in the test pairs. All ‘A’ and ‘B’ movements in a class-quintuplet came from different frames.

Part-quintuplets could have one of the following structures: $XYZB_i|A_j$, $YZB_i|A_jX$, $ZB_i|A_jXY$ and $B_i|A_jXYZ$, where the vertical bars indicate the positions of quintuplet boundaries during the familiarization (although no boundaries were present in part-quintuplets during test). We used only two

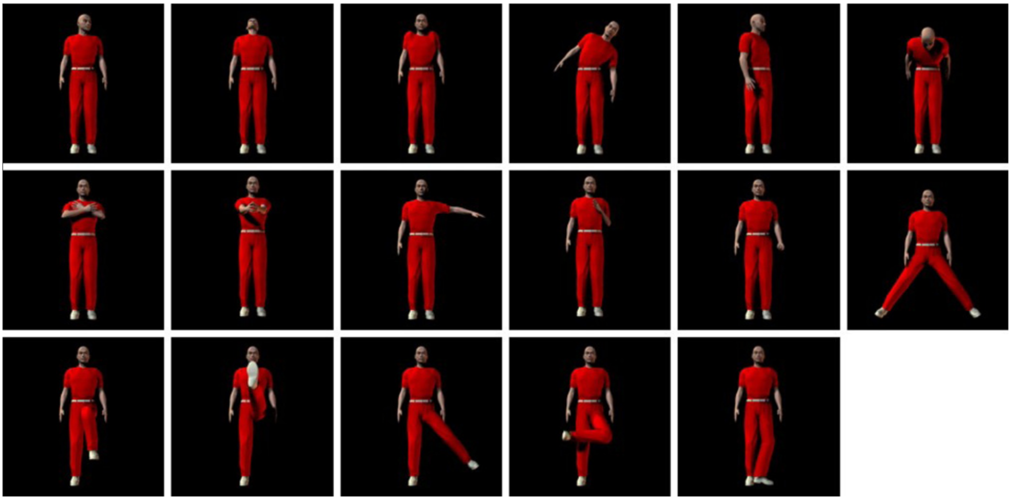

Fig. 4. Static depiction of the movements used in Experiments 4 and 5. The images show the movements’ maximal deviation from the neutral position. Sample clips are available as Supplementary Material.

types of part-quintuplets, namely YZB$_i$|A$_j$X and ZB$_i$|A$_j$XY, because 'A' and 'B' movements would appear at the edges of the other part-quintuplet types. We used 24 test pairs presented once in random order. In each pair, the class-quintuplet shared the 'A', 'B' and 'Y' movement with the part-quintuplet. Half of the trials started with a class-quintuplet and the other half with a part-quintuplet.

3.4.2. Results and discussion

As shown in Fig. 5, participants preferred class-quintuplets to part-quintuplets ($M = 62.3\%$, $SD = 17.7\%$), $t(19) = 3.11$, $p < 0.01$, Cohen's $d = 0.7$, $CI_{.95} = 54.0\%, 70.6\%$, suggesting that they tracked the positional information when the crucial movements were in edge positions.

3.5. Experiment 5: tracking positional regularities in middles

Experiment 5 was identical to Experiment 4 except that participants had to track positional information about 'A' and 'B' movements in quintuplets with the form X**A**$_i$Y**B**$_i$Z (rather than **A**$_i$XYZ**B**$_i$ as in Experiment 4); in other words, the critical movements were now quintuplet-medial rather than at the edges of the quintuplets. Again, there was a TP of 1.0 between each A$_i$ and its B$_i$.

3.5.1. Materials and method

3.5.1.1. Participants. Twenty new participants (13 females, 7 males, mean age 21.1 range 18–34) took part in this experiment for course credit or monetary compensation.

3.5.1.2. Familiarization. The familiarization movie was constructed as in Experiment 4, except that quintuplets had the structure X**A**$_i$Y**B**$_i$Z such that the critical movements were no longer at the edges. The assignment of the movements to the different positions (X, A, Y, B and Z) was the same as in Experiment 4. The TPs between adjacent movements were 0.25 or 0.33 within quintuplets, and 0.33 between quintuplets; also the higher order TPs were much lower than 1.0 (that is, the TP between 'A' and 'B' quintuplets). With only two repetitions of each quintuplet, it turned out to be impossible to generate a familiarization movie that controlled all TPs exactly, in particular between 'Z' and 'X' movements (that is, between the last movement of one quintuplet and the first movement of the next quintuplet); as these transitions occurred in part-quintuplets, we included their frequency in the data analysis to assess their influence.

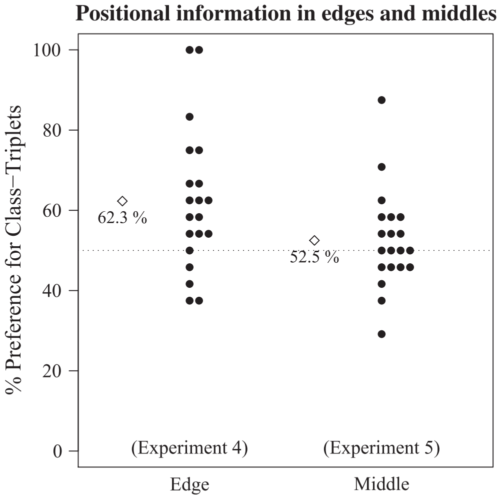

Fig. 5. Results of Experiments 4 and 5. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. Participants showed a sensitivity to positional regularities when the crucial movements occurred at the edges of quintuplets (Experiment 4), but not when they occurred in the middles of quintuplets (Experiment 5).

3.5.1.3. Test. In this test phase, participants had to choose between class-quintuplets and part-quintuplets. Class-quintuplets had the form $X'A_iYB_jZ'$, where X' and Z' were in reality 'A' and 'B' movements and had never appeared in initial or final positions during the familiarization movie. As in Experiment 4, A_i and B_j belonged to distinct frames; the TPs between the 'A' and 'B' movements were thus broken. Still, since these movements appeared in the positions in which they were encountered during familiarization, class-quintuplets respected the positional regularity from the familiarization movie. Importantly, however, in contrast to Experiment 2 and 4, the crucial movements occurred in middle positions rather than at the edges.

Class-quintuplets could have one of the following structures: $A_kA_iYB_jA_l$, $A_kA_iYB_jB_l$, $B_kA_iYB_jA_l$ or $B_kA_iYB_jB_l$; each of these structures appeared equally often in the test pairs. Part-quintuplets could have the structures $A_iYB_iZ|X$, $YB_iZ|XA_j$, $B_iZ|XA_jY$ and $Z—XA_jYB_j$,[3] but we used only the structures $B_iZ|XA_jY$ and $YB_iZ|XA_j$, because these are the only structures where the 'A' and 'B' movements come from different frames, and where the 'X' and 'Z' movements do not occur at edges, and because the quintuplet boundaries in these part-quintuplets were at the same positions as in Experiment 4 (that is, between the second and the third movement, or between the third and the fourth movement). Each part-quintuplet type was represented equally in the test pairs. In each test pair, both test items shared the 'A', 'B' and 'Y' movements.

Some transitions between 'Z' and 'X' movements (that is, between the last movement of one quintuplet and the first movement of the next one) in the part-quintuplets were more frequent than others, since we could not find a uniform randomization for the familiarization movie (see above); we thus included the frequency of these transitions in the data analysis by forming a "frequent" group with an average TP of 0.416 between 'Z' and 'X', and a "rare" group with an average TP of 0.168 between 'Z' and 'X'. Test pairs were presented in random order.

3.5.2. Results and discussion

As shown in Fig. 5, participants did not prefer class-quintuplets to part-quintuplets ($M = 52.5\%$, $SD = 12.3\%$), $t(19) = 0.91$, $p = 0.374$, Cohen's $d = 0.20$, $CI_{.95} = 46.8\%, 58.2\%$, ns. Performance did not differ depending on the frequency of the movement transitions within part-quintuplets, $F(1,19) = 3.16$, $p = 0.092$, $\eta_p^2 = 0.143$, ns. Participants' preference for class-quintuplets was higher in Experiment 4 than in Experiment 5, $F(1,38) = 4.1$, $p = 0.049$, $\eta^2 = 0.10$, suggesting that they tracked positional information predominantly at the quintuplet edges.

Before accepting this conclusion, it was necessary to test an alternative interpretation. In Experiment 5, participants had no preference for either class-triplets or part-triplets; these results contrast with those of Experiment 3, where participants preferred part-triplets to class-triplets. Do these results suggest that participants were sensitive to positional information in Experiment 5 because we did not observe a preference for part-triplets? While we cannot rule out this possibility, an alternative interpretation is that it was harder for participants to track TPs within part-triplets in Experiment 5 because, during familiarization, the transitions in these part-triplets were disrupted by long, 1.2 s breaks (see Shukla, Nespor, & Mehler (2007) for similar results with speech items). Hence, participants might have preferred part-triplets in Experiment 3 (where they were familiarized with continuous movies), but not in Experiment 5, where quintuplets were separated by 1.2 s breaks during familiarization. Importantly, however, even if participants tracked positional information in Experiment 5, their sensitivity to this information was much weaker than in Experiment 4. That is, we do not claim that it is impossible to track positional information in middle positions, and, in fact, work on positional memory has shown this to be possible to some limited extent (e.g., Henson, 1998; Hitch et al., 1996; Ng & Maybery, 2002). In line with much work on positional memory, we conclude that positional information is much stronger at edge positions than in other positions because all positions are encoded *relative* to the edges.

3.6. Discussion of Experiments 1–5

Together, the results of Experiments 1–5 replicate those obtained by Endress and Bonatti (2007) and Endress and Mehler (2009a) with speech items: participants are sensitive to TPs computed over

[3] Again, the vertical bars signal the positions of quintuplet boundaries during familiarization.

movements when presented with a continuous movement sequence (see also Baldwin et al., 2008). Further, observers can track positional information about movements: they learn which movements occur initially and which movements occur finally in movement sequences. However, as in the prior experiments with speech items, the positional computations can be performed only when movement sequences are delimited by explicit segmentation cues; moreover, positions are tracked predominantly at unit-edges as opposed to unit-medial positions. The main difference between the experiments reported here and the aforementioned speech experiments seems to be that TPs between adjacent movements are somewhat stronger than TPs between adjacent syllables. This might occur because the individual movements themselves are much longer than individual syllables, which, in turn, might facilitate the establishment of associations between them. Importantly, however, the combined results of Experiments 1–5 suggest that participants can track both TPs and positional information from movement sequences, a basic requirement if these mechanisms are to be used for integrating movement sequences into actions.

Next, we further investigate the potential of these mechanisms for integrating movements into actions. For this purpose, a useful outcome of Experiments 1 and 2 is that the preference for rule-triplets over class-triplets is very similar to the preference for class-triplets over part-triplets. Hence, Experiments 1 and 2 establish comparable baseline performances for tracking TPs and for tracking positional information, making any manipulation of these experiments directly comparable.[4]

4. Respecting causality

Experiments 1–5 establish that participants track both TPs and positional information over movement sequences. In Experiments 6–8, we further investigate the potential of these mechanisms for integrating movements into (goal-directed) actions. Specifically, a mechanism that integrates movements into actions should respect causality; that is, it should be sensitive to the temporal order of the movements in a sequence.

In Experiments 6 and 7, we investigated whether both TP-based and position-based computations encode the temporal order of movements. Specifically, we familiarized participants with the movies from Experiments 1 and 2. Then, all test items were presented in a reverse order. That is, the test items consisted of the same movements as in Experiments 1 and 2, but the order of the movements was reversed (while the movements themselves remained unchanged). Based on previous experiments that presented sequences of visual shapes (Turk-Browne & Scholl, 2009), we expected TP-based computations to recognize co-occurring movements regardless of whether the test items are presented in the same order or in the reverse order. While we are not aware of any previous studies examining the sensitivity of position-based computations to temporal order, we would not expect positional information to be reversible. After all, the first position is the first position, and the last position is the last position; if these two positions were interchangeable, it would be hard to see how positions could be encoded in a meaningful way. Hence, when presented with backward test items, we would expect participants to prefer backward rule-triplets to backward class-triplets, showing a sensitivity to backward TPs. In contrast, they should not prefer backward class-triplets to backward part-triplets because this would require the first and the last position to be interchangeable. In fact, participants might even prefer backward part-triplets to backward class-triplets, because backward part-triplets contain (potentially reversible) TP information, while class-triplets only contain positional information.

[4] While the comparison between rule-triplets and class-triplets in Experiment 1 provides a pure test of TPs between non-adjacent movements, the comparison between class-triplets and part-triplets in Experiment 2 pits positional information against TP information. Thus, at first glance, these two experiments do not appear to be comparable baselines for transitional probability and positional learning, respectively. However, Experiment 8 shows that the preference for part-triplets is not significantly above chance when the input consists of a segmented familiarization movie and the comparison items are backward class-triplets. Thus, TPs in part-triplets do not appear to be tracked reliably in segmented movies. Since our crucial comparisons between class-triplets and part-triplets all involve segmented familiarization movies, the comparisons between rule-triplets and class-triplets in continuous movies, and between class-triplets and part-triplets in segmented movies, provide roughly comparable baselines.

4.1. Experiment 6: tracking TPs in reversed test items

Experiment 6 asks whether participants can recognize TPs in movement sequences when the test items are played backward. Participants were familiarized with the same (continuous) movie as in Experiment 1. Then, they had to choose between backward rule-triplets and backward class-triplets. Recall that the only difference between rule-triplets and class-triplets is that the TP between the first and the last syllable is 1.0 in rule-triplets, and zero in class-triplets. Hence, if participants can track TPs even when the test items are reversed, they should prefer backward rule-triplets to backward class-triplets.

4.1.1. Materials and method

Experiment 6 was identical to Experiment 1, except that the test items were played backwards. That is, the test items contained the same movements as in Experiment 1, but the order of the movements was reversed (without reversing the actual movements). Twenty new participants (11 females, 9 males, mean age 19.4 range 17–22) took part in this experiment for course credit or monetary compensation.

4.1.2. Results and discussion

As shown in Fig. 6, participants preferred backward rule-triplets to backward class-triplets ($M = 60.0\%$, $SD = 13.7\%$), $t(19) = 3.27$, $p = 0.004$, Cohen's $d = 0.73$, $CI_{.95} = 53.6\%, 66.4\%$. Performance in Experiment 6 did not differ from that observed in Experiment 1, $F(1,38) = 0.4$, $p = 0.534$, $\eta^2 = 0.01$. Hence, participants are as good at recognizing TPs between movements when the test items are played forward as when they are played backward.

4.2. Experiment 7: tracking positional information in reversed test items (1)

Experiment 7 asks whether participants can recognize positional information when the test items are reversed. Given that positions are encoded relative to the first and the last position, one would not expect participants to exhibit this ability, since the first and the last position are unlikely to be interchangeable.

Experiment 7 tested this hypothesis by familiarizing participants with the (segmented) movie from Experiment 2. Following this, they had to choose between backward class-triplets and backward part-triplets. If participants cannot recognize positional information after reversal of the items, then they should either show no preference, or they should prefer reverse part-triplets, since the (reversible) TPs between their movements are non-zero.

4.2.1. Materials and method

Experiment 7 was identical to Experiment 2, except that the test items were played backwards. That is, the test items contained the same movements as in Experiment 2, but the order of the movements was reversed (without reversing the actual movements). Twenty new participants (9 females, 11 males, mean age 19.6 range 17–24) took part in this experiment for course credit or monetary compensation.

4.2.2. Results and discussion

As shown in Fig. 7, participants preferred backward part-triplets to backward class-triplets (preference for class-triplets: $M = 43.8\%$, $SD = 11.7\%$), $t(19) = 2.40$, $p = 0.027$, Cohen's $d = 0.54$, $CI_{.95} = 38.3\%, 49.2\%$. Their preference for class-triplets differed from that observed in Experiment 2, $F(1,38) = 14.2$, $p < 0.001$, $\eta^2 = 0.273$.

An ANOVA with the factors contrast type (class-triplet/part-triplet vs. rule-triplet/class-triplet) and inversion (original vs. reversed test items) yielded main effects of contrast type, $F(1,76) = 8.72$, $p < 0.005$, $\eta^2 = 0.09$, and inversion, $F(1,76) = 4.45$, $p = 0.038$, $\eta^2 = 0.05$, and, crucially, an interaction between these factors, $F(1,76) = 9.18$, $p < 0.005$, $\eta^2 = 0.09$. Hence, in contrast to Experiment 2, participants preferred backward part-triplets to backward class-triplets.

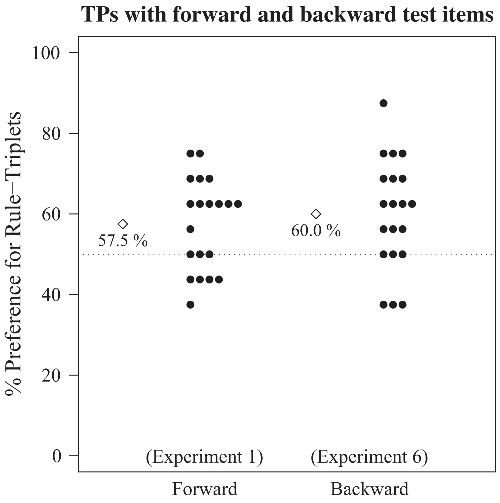

Fig. 6. Results of Experiments 1 and 6. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. Participants are as good at recognizing TPs between movements when the test items are played forward (Experiment 1) as when they are played backwards (Experiment 6), preferring rule-triplets to class-triplets to the same extent.

While these results show that positional information is not reversible, the preference for backward part-triplets might have occurred for two reasons. First, participants might have preferred backward part-triplets to backward class-triplets due to the TPs in part-triplets. Second, this preference might arise because backward part-triplets have "correct" initial or final movements (but not both) and, therefore, contain partially correct positional information. In Experiment 8, we address this issue by replicating Experiment 7, but asking participants to choose between backward class-triplets and forward part-triplets; forward part-triplets contain TP information, but all positional information is obliterated.

4.3. Experiment 8: tracking positional information in reversed test items (2)

Experiment 8 asked why participants in Experiment 7 preferred backward part-triplets to backward class-triplets. On the one hand, they might have recognized the backward TPs contained in backward part-triplets; on the other hand, they might have relied on the backward part-triplets' partially correct positional information. We tested this issue by replicating Experiment 7, but asking participants to choose between backward class-triplets and *forward* part-triplets.

4.3.1. Materials and method

Experiment 8 was identical to Experiment 7, except that the test part-triplets were played forward. Twenty new participants (10 females, 10 males, mean age 26.2 range 20-34) from the MIT community took part in this experiment for monetary compensation.

4.3.2. Results and discussion

As shown in Fig. 7, participants did not significantly prefer forward part-triplets to backward class-triplets (preference for class-triplets: $M = 46.5\%$, $SD = 11.8\%$), $t(19) = 1.34$, $p = 0.196$, Cohen's $d = 0.3$, $CI_{.95} = 40.9\%$, 51.98%, ns, although removing an outlier at 2.42 standard deviation from the mean would yield a significant preference for forward part-triplets (preference for class-triplets: $M = 45.0\%$, $SD = 10.0\%$), $t(18) = 2.2$, $p = 0.041$, Cohen's $d = 0.51$, $CI_{.95} = 40.2\%$, 49.8%.

The preference for class-triplets differed from the preference in Experiment 2, $F(1,38) = 9.1$, $p = 0.0045$, $\eta^2 = 0.194$, as well as from the preference in Experiments 3, $F(1,38) = 4.4$, $p = 0.043$,

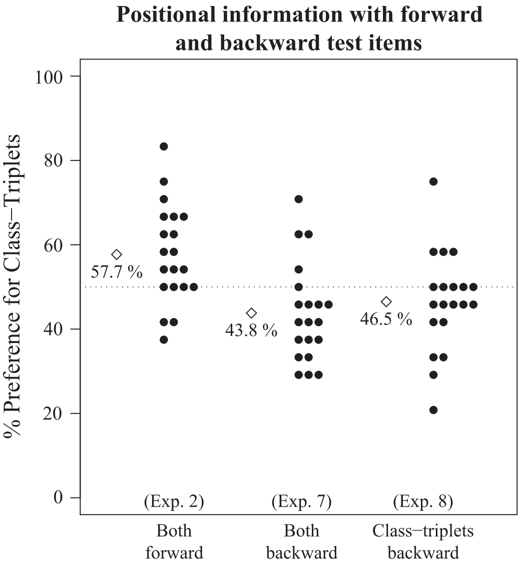

Fig. 7. Results of Experiments 2, 7 and 8. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. When both test items were played forward, participants preferred class-triplets to part-triplets, showing a sensitivity to positional information (Experiment 2). When both test items were played backwards, however, participants preferred part-triplets to class-triplets (Experiment 7), because the TPs between movements in part-triplets were reversible while the positional information in class-triplets was not (or because backward part-triplets contain partial positional information). When only class-triplets are played backward and part-triplets are played forward, participants had no significant preference for either item (Experiment 8).

$\eta^2 = 0.103$, suggesting that the preference in Expeiment 8 was in-between the preferences observed in Experiments 2 and 3. However, the preference for class-triplets in Experiment 8 did not differ from that in Experiment 7, $F(1,38) = 0.5$, $p = 0.47$, $\eta^2 = 0.0138$, ns. Likelihood ratio analysis (Glover & Dixon, 2004) shows that the hypothesis that the results of Experiment 7 and 8 do not differ is 4.8 or 2.4 more likely than the hypothesis that they do differ after correction with the Bayesian Information Criterion and the Akaike Information Criterion, respectively. Hence, while Experiment 8 replicated the result of Experiment 7, showing that positional information is not reversible, the combined results of Experiments 7 and 8 are unclear as to whether the preference for part-triplets in Experiment 7 was driven by partial positional information, or rather by TPs that participants tracked across triplet boundaries despite the intervening gaps.[5]

4.4. Discussion of Experiments 6–8

Experiments 6–8 investigated one crucial requirement for a mechanism that integrates movement sequences into (goal-directed) actions: to respect causality. Such a mechanism must create representations that contain order information because the order of movements often determines the goals that the movements accomplish.

Results showed that participants recognized TPs just as well whether the test items were played forward or backward; positional information, in contrast, was recognized only when the test items were played forward.

[5] While participants in Experiment 3 preferred part-triplets to class-triplets after a familiarization with a continuous movie, these results are not inconsistent with those of Experiment 8, where participants failed to show such a preference. At least in the speech domain, there appears to be a diminished sensitivity to TPs when the TPs span explicit segmentation markers (Shukla et al., 2007). Given that the familiarization movie was segmented in Experiment 8 but not in Experiment 3, we should expect diminished sensitivity to part-triplets in Experiment 8 (see also the discussion of our Experiment 5 above).

While these results suggest that the position-based mechanism respects causality, the implications are less clear for the TP-based mechanism. First, the results of Experiment 6 do not imply that TP-based computations are completely insensitive to temporal order. Indeed, at least with shape sequences, participants have a moderate sensitivity to temporal order, as they successfully discriminate forward sequences from backward sequences (Turk-Browne & Scholl, 2009), even though the effect size for discriminating backward sequences from backward foils was almost twice as large as the effect size for discriminating forward sequences from backward sequences.

The ability of TP-based computations to track both forward and backward TPs might lead to problems identifying integrated actions. For example, the order of movements in an action partially characterizes the action because, as shown in the introduction in the example about moving a lamp and turning it on, different actions sometimes contain the same set of movements — but in different temporal orders. Thus, it is unclear how a TP-based mechanism could efficiently encode goal-directed actions without also encoding robust information about the temporal order of movements. This problem is compounded if, as in the domain of speech, TP-based computations do not allow the observer to extract integrated units (Endress & Mehler, 2009b). At minimum, it thus seems necessary to investigate how well TP-based computations encode causality in more ecologically-relevant situations (rather than with the simplified artificial stimuli and two-alternative forced choice tasks used here). Given the current data, we thus conclude that the position-based mechanism respects causality, while it is unclear whether this is true for the TP-based mechanism.

5. Invariance under viewpoint changes

The experiments presented so far suggest that both the TP-based and positional-based mechanisms can encode movement information, and that the positional-based mechanism contains temporal order information. While our results do not rule out the possibility that the TP-based mechanism might also encode order information, additional research, under more ecologically-appropriate conditions, is needed to provide support for this possibility.

Experiments 9–12 investigate another likely property for a mechanism that integrates movement sequences into (goal-directed) actions: such a mechanism should create representations that allow actions to be recognized from different viewpoints. Indeed, the goal of an action is identical irrespectively of the viewpoint from which the action is observed; hence, actions should be recognizable from different viewpoints.

Experiments 9–12 test for this property by familiarizing participants with the movies from Experiments 1 and 2. While the animated agent faced the participant during these familiarization movies, participants saw the agent from the side (90° in-depth rotation) during the test phase of the experiment.

5.1. Experiment 9: tracking TPs across a viewpoint change

Experiment 9 asks whether participants can track TP information across a viewpoint change. Experiment 9 was a replication of Experiment 1 (where participants had to choose between rule-triplets and class-triplets after a familiarization with a continuous movie), except that, during test, the agent was rotated by 90°. That is, while participants faced the agent during familiarization, they saw him from the side during test. A 90° viewpoint change was used for two reasons. First, as illustrated in Fig. 8, a 90° rotation produces substantial changes in the visible features of the movements while still allowing observers to recognize the movements. Second, observers can recognize movements retained in visual working memory across a 90° viewpoint change (Wood, 2010). Thus, in the present study we examined whether TP-based and position-based mechanisms create movement representations that are analogous to those maintained in working memory (i.e., representations that can be recognized from different viewpoints).

5.1.1. Materials and method

Experiment 9 was identical to Experiment 1, except that the agent was rotated by 90° during test. Twenty new participants (12 females, 8 males, mean age 21.6 range 19–31) took part in this experiment for course credit or monetary compensation.

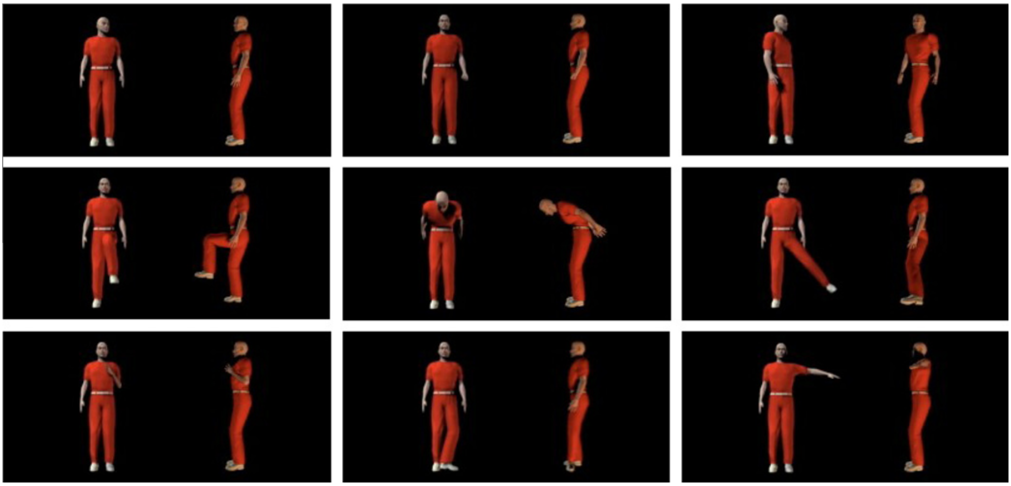

Fig. 8. Static depiction of the movements used in Experiments 9–12. The images show the movements' maximal deviation from the neutral position. In each square, the view on the left corresponds to the familiarization sequence, while participants saw the view on the right in the test items. Sample clips are available as Supplementary Material.

5.1.2. Results and discussion

As shown in Fig. 9, participants had no preference for rule-triplets over class-triplets when the movements were observed from a different viewpoint ($M = 54.1\%$, $SD = 12.5\%$), $t(19) = 1.45$, $p = 0.164$, Cohen's $d = 0.32$, $CI_{.95} = 48.2\%, 59.9\%$, ns. However, their performance did not differ significantly from that observed in Experiment 1 either, $F(1,38) = 0.8$, $p = 0.37$, $\eta^2 = 0.021$, ns, and power analysis revealed that we would need at least 8,372 participants in each experiment to achieve a power of 80%. Further, likelihood ratio analysis (Glover & Dixon, 2004) suggested that a linear model postulating no difference between the experiments was 2.1 times more likely than a model postulating a difference between the experiments after AIC correction, and 4.1 times more likely after BIC correction. Hence, the results of Experiment 9 are difficult to interpret. On the one hand, participants failed to track TP information when the movements were observed from a different viewpoint, and the effect size was one-half of that observed in Experiment 1; on the other hand, participants' performance did not differ significantly from that in Experiment 1.

In Experiments 10a and 10b, we attempt to provide more clarity to the issue of whether TPs can be tracked across viewpoint changes. Since TPs become more robust with increased exposure to a familiarization stream (Endress & Bonatti, 2007), we replicated Experiment 9 but doubled the length of the familiarization movie. To provide an equivalent baseline for performance with non-rotated test items, we also replicated Experiment 1, but again doubled the length of the familiarization movie.

5.2. Experiment 10: tracking TPs with and without a viewpoint change after increased exposure

The goal of Experiments 10a and 10b was to clarify whether participants can track TPs across viewpoint changes after prolonged exposure to the familiarization stream. Experiment 10a was a replication of Experiment 1, but with doubled exposure; that is, participants were familiarized with continuous movies, and were then tested on items that had *not* been rotated. Experiment 10b was identical to Experiment 10a, except that, as in Experiment 9, the agent was rotated by 90° during test.

5.2.1. Materials and method

Experiments 10a and 10b were replications of Experiments 1 and 9, respectively, except that the familiarization movie was played twice. Forty new participants (32 females, 8 males, mean age 20.2

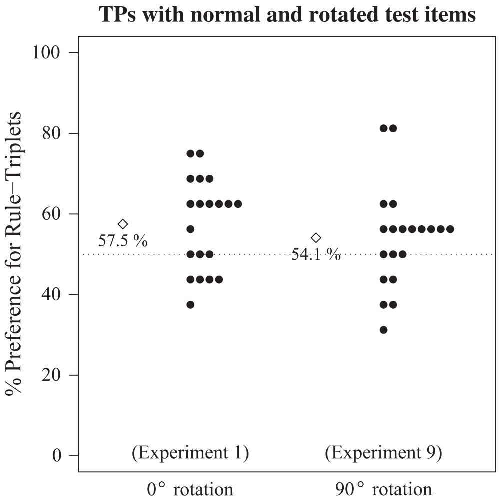

Fig. 9. Results of Experiments 1 and 9. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. Participants preferred rule-triplets to part-triplets when the agent was presented from the same viewpoint during familiarization and test (Experiment 1); in contrast, when the agent was rotated by 90° during test relative to the familiarization movie participants had no such preference (Experiment 9).

range 18–26) from the University of Southern California Study Pool took part in this experiment for course credit. They were randomly assigned to Experiments 10a and 10b, respectively.

5.2.2. Results and discussion

As shown in Fig. 10, participants in Experiment 10a (where test items were not rotated) preferred rule-triplets to class-triplets ($M = 59.7\%$, $SD = 13.8\%$), $t(19) = 3.1$, $p = 0.005$, Cohen's $d = 0.7$, $CI_{.95} = 53.2\%, 66.2\%$. In contrast, participants in Experiment 10b had no preference for either test item type ($M = 51.6\%$, $SD = 13.9\%$), $t(19) = 0.5$, $p = 0.62$, Cohen's $d = 0.11$, $CI_{.95} = 45.1\%, 58.1\%$, ns. The preference for rule-triplets differed marginally between the two experiments, $F(1,38) = 3.4$, $p = 0.071$, $\eta^2 = 0.083$, ns. Hence, even when the duration of the familiarization movie was doubled, participants failed to track TPs in rotated items.

5.3. Experiment 11: tracking positional information across a viewpoint change (1)

Experiment 11 asks whether participants can track positional information across a viewpoint change. Experiment 11 was a replication of Experiment 2 (where participants had to choose between class-triplets and part-triplets after a familiarization with a segmented movie), except that, during test, the agent was rotated by 90°. That is, while participants faced the agent during familiarization, they saw him from the side during test.

5.3.1. Materials and method

Experiment 11 was identical to Experiment 2, except that the agent was rotated by 90° during test. Twenty new participants (11 females, 9 males, mean age 20.3 range 18–24) took part in this experiment for course credit or monetary compensation.

5.3.2. Results and discussion

As shown in Fig. 11, participants preferred class-triplets to part-triplets ($M = 58.8\%$, $SD = 16.7\%$), $t(19) = 2.34$, $p = 0.03$, Cohen's $d = 0.52$, $CI_{.95} = 50.9\%, 66.6\%$. Performance in Experiment 11 did not differ from that observed in Experiment 2, $F(1,38) = 0.05$, $p = 0.82$, $\eta^2 = 0.001$, ns. Participants thus seem capable of tracking positional information across a viewpoint change.

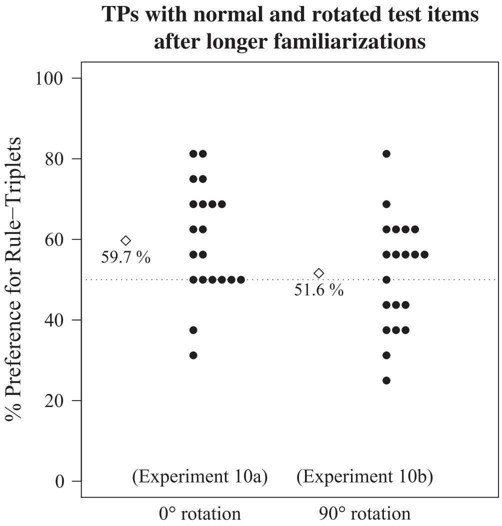

Fig. 10. Results of Experiments 10a and 10b. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. Participants preferred rule-triplets to class-triplets when the agent was presented from the same viewpoint during familiarization and during test (Experiment 10a); in contrast, when the agent was rotated by 90° during test relative to the familiarization movie participants had no such preference (Experiment 10b). In both experiments, the duration of the familiarization movie was twice the length of the other experiments.

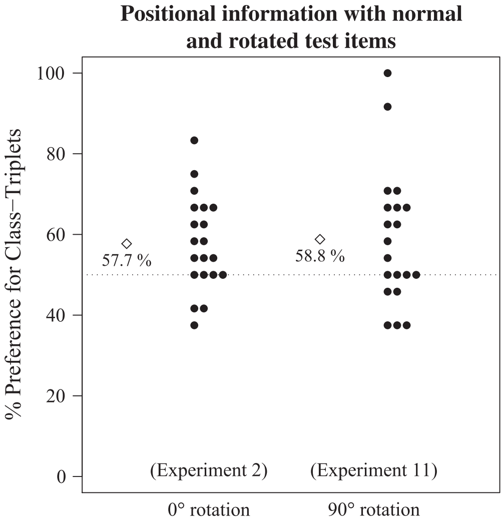

Fig. 11. Results of Experiment 2 and 11. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. Participants preferred class-triplets to part-triplets both when the agent was observed from the same viewpoint during familiarization and test (Experiment 2), and when, during the test phase, the agent was rotated by 90° relative to the familiarization movie, suggesting that the position-based mechanism creates representations that can be recognized from different viewpoints.

An ANOVA with the factors contrast type (class-triplet/part-triplet vs. rule-triplet/class-triplet) and rotation (original vs. rotated test items) yielded no main effects or interactions (all *F*'s < 1).

However, as shown in Fig. 11, there was one participant in Experiment 11 whose performance deviated from the mean by 2.47 standard deviations; when this participant is removed from the analyses, there is only a marginal preference for rotated class-triplets, ($M = 56.6\%$, $SD = 14.0\%$), $t(18) = 2.1$, $p = 0.055$, Cohen's $d = 0.47$, $CI_{.95} = 49.8\%$, 63.3%, no difference between Experiments 2 and 11, $F(1,37) = 0.07$, $p = 0.786$, $\eta^2 = 0.002$, ns, but a significant difference between Experiments 3 and 11, $F(1,37) = 16.8$, $p = 0.0002$, $\eta^2 = 0.312$. In Experiment 12, we therefore confirm this result by replicating Experiment 11 both with continuous and with segmented familiarization movies.

5.4. Experiment 12: tracking positional information across a viewpoint change (2)

The purpose of Experiment 12 was to replicate the results of Experiment 11. Experiment 12a was a replication of Experiment 11; Experiment 12b was identical to Experiment 12a except that the familiarization movie was continuous.

5.4.1. Materials and method

Forty new participants (29 females, 11 males, mean age 19.9 range 18–26) from the University of Southern California Study Pool took part in this experiment for course credit. They were randomly assigned to Experiments 12a and 12b, respectively.

5.4.2. Results and discussion

As shown in Fig. 12, participants in Experiment 12a preferred rotated class-triplets to rotated part-triplets ($M = 59.6\%$, $SD = 15.2\%$), $t(19) = 2.82$, $p = 0.011$, Cohen's $d = 0.63$, $CI_{.95} = 52.5\%$, 66.7%. In Experiment 12b, in contrast, participants had no preference for either item type ($M = 47.5\%$, $SD = 14.9\%$), $t(19) = 0.8$, $p = 0.462$, Cohen's $d = 0.17$, $CI_{.95} = 40.5\%$, 54.5%, ns. The preference for class-triplets differed significantly between Experiments 12a and 12b, $F(1,38) = 6.5$, $p = 0.015$, $\eta^2 = 0.145$.

The results of Experiment 12a and 12b thus replicate those of Experiment 11. After observing a segmented familiarization movie, participants preferred class-triplets to part-triplets even after a viewpoint change. After observing a continuous familiarization movie, in contrast, participants showed no such preference. While participants in Experiment 3 (who observed a continuous movie and non-rotated test items) preferred part-triplets to class-triplets, such a preference should not be

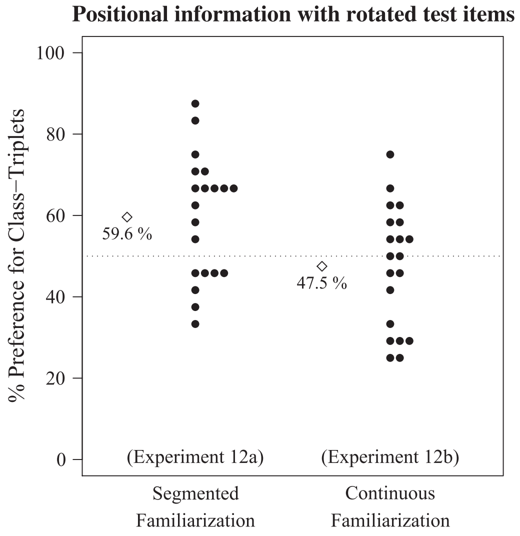

Fig. 12. Results of Experiment 12a and 12b. Dots represent the means of individual participants, diamonds sample averages, and the dotted line the chance level of 50%. When the agent was rotated by 90° relative to the familiarization movie, participants preferred class-triplets to part-triplets after observing a segmented familiarization movie (Experiment 12a) but not after observing a continuous familiarization movie (Experiment 12b), suggesting that the position-based mechanism tolerates changes in viewpoint.

expected in Experiment 12b; after all, the preference for part-triplets is driven by TPs, and the results of Experiment 9 and 10 show that TPs are not tracked across changes in viewpoint.

5.5. Discussion of Experiments 9–12

In Experiments 9–12, we asked whether participants could recognize positional information and TP-based information when the movements were observed from a different viewpoint. If these mechanisms are used for integrating movement sequences into (goal-oriented) actions, then they should tolerate viewpoint changes, as goals are invariant under such changes.

Participants were familiarized with the same movies as in Experiments 1 and 2. In the test phase, however, the agent performing the movements had been rotated by 90°. Results showed that participants were sensitive to positional information when they viewed the movements from this different viewpoint; in contrast, they did not track TP-information under these conditions, even when the duration of the familiarization movie was doubled. Importantly, it was not the case that the viewpoint change simply made the task harder. In fact, participants tracked positional information equally well with and without a viewpoint change, and, without the viewpoint change, they tracked positional information and TPs to similar extents. Hence, it seems reasonable to conclude that, while positional information can be used to recognize encoded actions from different viewpoints, further research is needed to determine whether this is also true for the TP-based mechanism.

6. General discussion

When we observe others act, we see only the surface appearance of their behavior. Nevertheless, we readily perceive such movements as goal-directed actions (e.g., Gergely & Csibra, 2003; Woodward, 1998; Wood & Hauser, 2008), which raises the question of how we integrate (observable) movement sequences into (goal-directed) actions. In the experiments presented here, we investigated this question by noting important analogies between how, in the domain of language, words are extracted from fluent speech, and how, in the domain of action perception, goal-directed actions are extracted from dynamic movement sequences. We studied two mechanisms that have been shown to operate on fluent speech, and asked whether they fulfill three requirements that seem necessary for a mechanism to successfully integrate movement sequences into actions. First, such mechanisms must operate over movement sequences. Second, such mechanisms must respect causality; that is, they should contain information about the order in which movements occurred. Third, the units created by such mechanisms should be recognized across changes in viewpoint. Of course, these are just necessary conditions, and such mechanisms must also have other properties to successfully integrate movements into goal-directed actions.

We evaluated these criteria for two sequence learning mechanisms that operate over fluent speech. In all experiments, participants were familiarized with a sequence of movements performed by an animated actor. The sequence was a concatenation of movement triplets that implemented both a positional regularity and a regularity based on transitional probabilities (TPs). Following this familiarization, participants were presented with pairs of test triplets, and had to decide which of the test triplets was more like what they had seen during familiarization.

In Experiments 1–5, we showed that participants can track both transitional probabilities (TPs) and positional information from movement sequences. As in prior experiments with speech stimuli, positional information was tracked only when triplets were separated by explicit segmentation cues (that is, stasis). Moreover, positional information was tracked predominantly at the edges of units as opposed to unit-medial positions, presumably because positions are encoded relative to the sequence edges.

In Experiments 6–8, we asked whether the TP-based mechanism and the position-based mechanism respect causality. Specifically, we examined whether participants could recognize movement patterns played backwards. If these mechanisms respect causality, it should be more difficult to recognize backward items compared to forward items. Participants did not show any position-based memory for backward items; in contrast, they were able to recognize TPs whether the test items were presented in the learned order or in the reverse order. Thus, these results suggest that the position-based learning mechanism respects causality, whereas we observed no evidence that the TP-based mechanism respects causality.

In Experiments 9–12, we asked whether TP-based and position-based mechanisms create action representations that can be recognized from different viewpoints. After being familiarized to the sequences from Experiments 1 and 2, participants were presented with test items in which the agent was rotated by 90°. Thus, the test movements were observed from a different viewpoint than the movements in the familiarization sequence. Remarkably, participants tracked position information equally well whether the agent performed the actions from the same viewpoint or from a different viewpoint. This suggests that position-based computations create view-invariant action representations. In contrast, participants failed to track TP information across a viewpoint change, even when the duration of the familiarization movie was doubled. This suggests that TP-based computations create representations that are closely tied to the surface appearance of behavior (i.e., what actions look like when they are being observed), as opposed to tracking the TPs of three-dimensional movements.

Together, our results suggest that the position-based learning mechanism fulfills all the necessary conditions for a mechanism used for integrating movement sequences into (goal-oriented) actions: it operates on movement sequences, it respects causality, and its units can be recognized from different viewpoints. While TP-based mechanisms seem to operate on movement sequences, our results do not provide any evidence that they fulfill the other two necessary criteria. It is worth emphasizing, however, that we assessed TP-based mechanisms by testing TPs between non-adjacent movements. Additional studies are needed to determine whether these same patterns obtain when TP-based mechanisms are assessed using TPs between adjacent items.

In the remainder of the discussion, we will ask whether these mechanisms fulfill two further necessary conditions for a mechanism that integrates movements into actions: (1) the action representations must be hierarchical representations, and (2) the representations must contain goal information.

6.1. Compatibility with hierarchical organization

As mentioned in the introduction, both language and action are organized hierarchically. Mechanisms that extract words from fluent speech, or that integrate movements into actions must therefore be compatible with hierarchical representations. Unfortunately, however, there is no conclusive evidence about whether the two aforementioned mechanisms – TP-based and position-based computations – are compatible with hierarchical representations. While several authors have attempted to provide evidence for hierarchical TP computations, the results could also be explained by appealing to non-hierarchical representations. On the other hand, there are important theoretical considerations, especially from formal linguistics, which suggest that edge-based positional codes can be used hierarchically; but, to our knowledge, there is no experimental work testing this possibility empirically. We will now discuss these issues in turn.

To our knowledge, there have been two attempts to observe hierarchically organized TP processes. First, Saffran and Wilson (2003) proposed that infants can use TPs between syllables to learn words from fluent speech, and, on a higher level, to use TPs between words to learn syntax-like structures. However, in their experiments, the second-order TPs between syllables showed large differences between legal test items and foils; as adults (e.g., Endress & Bonatti, 2007; Onnis, Monaghan, Richmond, & Chater, 2005; Peña et al., 2002), but probably also infants (Gómez, 2002) are sensitive to such TPs, it is possible that infants in Saffran and Wilson's (2003) experiments detected second-order TPs rather than processing TPs hierarchically.

Second, Fiser and Aslin (2005) proposed that adults can use TPs to learn hierarchical combinations of visual shapes (see also Fiser & Aslin, 2001, 2002a, 2002b). Their argument was based on the observation that, once participants had learned that a set of shapes formed a unit, they did not recognize any sub-units. To use an analogy with words, listeners should be unable to recognize the word "ham" when listening to the word "hamster", because "ham" is a subunit of "hamster". While some of their experiments were consistent with this prediction, others were not (e.g., their Experiment 5), and other research revealed further departures from such predictions (Endress & Vickery, in preparation). Hence, these results do not seem to offer strong support for the capacity of TP-based mechanisms to learn hierarchical structures.

In contrast, although we are not aware of any experimental work investigating whether positional information can be acquired hierarchically, there is a substantial body of theoretical work, especially from formal linguistics, suggesting that edge-positions are crucial for hierarchical processing. For example, different components of language have different hierarchical representations that do not always coincide, such as prosodic and syntactic hierarchies. For example, the plural [s] (in the syntactic hierarchy) is not a syllable (in the prosodic hierarchy); still the last edge of a (prosodic) syllable in which the plural [s] might occur (e.g., in the syllable of the word "dogs") coincides with last edge of the (syntactic) [s] morpheme. While this example might seem somewhat simplistic, there are numerous more complex linguistic regularities that can be explained if the edges of hierarchical constituents are aligned (e.g., Hayes, 1989; McCarthy & Prince, 1993; Nespor & Vogel, 1986).

Similar to language, event perception is also hierarchically structured (see Zacks & Swallow, 2007), and may thus require integrating information from different levels of representation. For example, accurate perception of someone's behavior requires integrating information about the larger goal of the activity (e.g., washing the car) with the sub-goals used to fulfill that goal (e.g., spraying the car with water, scrubbing the car, drying the car). Since one of the edges of the first and the last fine-grained event will align with the edges of a course-grained event, positional information might be an important reference point to integrate these different levels of action representation. It is, therefore, a promising direction for future research to find out whether the role of edges in action hierarchies is similar to that in linguistic hierarchies.

6.2. Where are the goals?

The present results suggest that the position-based mechanism fulfills several necessary conditions for integrating movements into actions: it operates over movement sequences, it encodes the temporal order of movements as needed for causality, its units can be recognized from different viewpoints, and theoretical considerations suggest that position information plays an important role in hierarchical representation. Moreover, since positions are encoded relative to the first and the last position of a sequence, they might be well suited for encoding the goals of actions. That is, position-based computations would encode both the first and last movements of a sequence, which may correspond to the starting point and the goal of an action.

In contrast, while a TP-based mechanism seems to operate on movement sequences, the results presented here do not provide any evidence that such a mechanism fulfills the other necessary conditions for a mechanism to integrate movements into actions.

However, although the conditions investigated here are necessary conditions for integrating movements into actions, they are by no means sufficient. Indeed, both position-based and TP-based mechanisms can operate over sequences that do not involve any goals at all, such as speech streams and the movement sequences used in the present experiments (which had no obvious goals). Thus, how are goals then linked to movement sequences encoded by TP-based or position-based mechanisms?

While our results do not directly speak to this issue, we speculate, in contrast to previous authors (e.g., Baldwin et al., 2008), that TPs or positional mechanisms are not sufficient for linking movements to goals. After all, these mechanisms are fundamentally memory mechanisms for sequences (Endress & Mehler, 2009a, 2009b). Movements, however, are not fully predictive of goals; for example, as mentioned in the introduction, extending one's arm is compatible with many different actions, from grasping a bottle to conducting an orchestra. A mechanism that simply associates a specific movement to an intended outcome will therefore fail to successfully link movements to goals. Furthermore, one of the defining properties of human and nonhuman social cognition is that individuals make inferences about others' intentions and goals by evaluating their actions in relation to the constraints imposed by the environment (e.g., Brass et al., 2007; Gergely & Csibra, 2003; Wood et al., 2007). This capacity enables individuals to go beyond the surface appearance of behavior to draw inferences about an individual's mental states. Thus, in order for a sequence learning mechanism to support action representation, the informational content of its representations must not be strictly tied to the surface appearance of movement. Rather, the representations should include information about how actions unfold within the constraints of the environment. The current results provide evidence that the position-based mechanism creates action representations that are not strictly tied to the surface appearance of behavior because its units can be recognized from different viewpoints. It will be interesting for future studies to examine whether the position-based mechanism can analyze movements in relation to the environmental constraints that guide rational action, thereby integrating information about goals and mental states.

Acknowledgments

This research was supported by the Mind, Brain and Behavior Initiative at Harvard University and the University of Southern California.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cogpsych.2011.07.001.

References

Asch, S. (1952). *Social psychology*. Englewood Cliffs, NJ: Prentice Hall.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321–324.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition, 106*(3), 1382–1407.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: Inferential processes versus action simulation. *Current Biology, 17*(24), 2117–2121.

Brentari, D., González, C., Seidl, A., & Wilbur, R. (2011). Sensitivity to visual prosodic cues in signers and nonsigners. *Language and Speech, 54*(1), 49–72.

Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). 'Unwilling' versus 'unable': Chimpanzees' understanding of human intentional action. *Developmental Science, 7*(4), 488–498.

Chomsky, N. (1980). *Rules and representations*. Oxford: Basil Blackwell.

Chomsky, N. (1981). *Lectures in government and binding*. Dordrecht: Foris.

Christophe, A., Mehler, J., & Sebastian-Galles, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy, 2*(3), 385–394.

Cooper, R. P., & Shallice, T. (2006). Hierarchical schemas and goals in the control of sequential behavior. *Psychological Review, 113*(4), 887–916 (discussion 917–931).

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition, 105*(2), 247–299.

Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology, 61*(2), 177–199.

Endress, A. D., & Mehler, J. (2009a). Primitive computations in speech processing. *The Quarterly Journal of Experimental Psychology, 62*(11), 2187–2209.

Endress, A. D., & Mehler, J. (2009b). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language, 60*(3), 351–367.

Feldman, H., Goldin-Meadow, S., & Gleitman, L. R. (1978). Beyond Herodotus: The creation of language by linguistically deprived deaf children. In A. Lock (Ed.), *Action, symbol, and gesture: The emergence of language* (pp. 351–414). New York: Academic Press.

Fenlon, J., Denmark, T., Campbell, R., & Woll, B. (2008). Seeing sentence boundaries. *Sign Language & Linguistics, 10*(2), 177–200.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*(6), 499–504.

Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 458–467.

Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America, 99*(24), 15822–15826.

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General, 134*(4), 521–537.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends in Cognitive Sciences, 7*(7), 287–292.

Glover, S., & Dixon, P. (2004). Likelihood ratios: a simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin and Review, 11*(5), 791–806.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431–436.

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

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition, 78*(3), B53–B64.

Hayes, B. (1989). The prosodic hierarchy in meter. In *Phonetics and phonology*. In P. Kiparsky & G. Youmans (Eds.). *Rhythm and meter* (Vol. 1, pp. 201–260). Orlando, FL: Academic Press.

Heider, F. (1958). *The psychology of interpersonal relations*. NY: Wiley.

Henson, R. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*(2), 73–137.

Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49*, 116–139.

Lashley, K. (1951). The problem of serial order in behavior. In L. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley.

McCarthy, J. J., & Prince, A. (1993). Generalized alignment. In G. Booij J. van Marle (Ed.), *Yearbook of morphology 1993* (pp. 79–153). Boston, MA: Kluwer.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology, 28*(1), 28–38.

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology, 12*, 436–450.

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

Ng, H. L., & Maybery, M. T. (2002). Grouping in short-term verbal memory: Is position coded temporally? *Quarterly Journal of Experimental Psychology: Section A, 55*(2), 391–424.

Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. Davidson, G. Schwartz, & D. Shapiro (Eds.). *Consciousness and self regulation: Advances in research and theory* (Vol. 4, pp. 1–18). New York, NY: Plenum Press.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in speech processing. *Journal of Memory and Language, 53*(2), 225–237.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*(5593), 604–607.

Perrett, D., Harries, M., Mistlin, A., & Chitty, A. (1990). Three stages in the classification of body movements by visual neurons. In H. Barlow, C. Blakemore, & M. Weston-Smith (Eds.), *Images and understanding* (pp. 94–107). New York, NY: Cambridge University Press.

Pilon, R. (1981). Segmentation of speech in a foreign language. *Journal of Psycholinguistic Research, 10*(2), 113–122.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: MIT Press.

Range, F., Viranyi, Z., & Huber, L. (2007). Selective imitation in domestic dogs. *Current Biology, 17*(10), 868–872.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience, 2*(9), 661–670.

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition, 81*(2), 149–169.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: evidence for developmental reorganization. *Developmental Psychology, 37*(1), 74–85.

Saffran, J. R., Johnson, E., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4*(2), 273–284.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology, 54*(1), 1–32.

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(5), 1131–1142.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97*(2), B25–B34.

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception and Psychophysics, 67*(5), 867–875.

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134*(4), 552–564.

Turk-Browne, N. B., & Scholl, B. J. (2009). Flexible visual statistical learning: Transfer across space and time. *Journal of Experimental Psychology: Human Perception and Performance, 35*(1), 195–202.

Wood, J. N. (2007). Visual working memory for observed actions. *Journal of Experimental Psychology: General, 136*(4), 639–652.

Wood, J. N. (2008). Visual memory for agents and their actions. *Cognition, 108*(2), 522–532.

Wood, J. N. (2010). Visual working memory retains movement information within an allocentric reference frame. *Visual Cognition, 18*(10), 1464–1485.

Wood, J. N., Glynn, D. D., Phillips, B. C., & Hauser, M. D. (2007). The perception of rational, goal-directed action in nonhuman primates. *Science, 317*(5843), 1402–1405.

Wood, J. N., & Hauser, M. D. (2008). Action comprehension in non-human primates: Motor simulation or inferential reasoning? *Trends in Cognitive Sciences, 12*(12), 461–465.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*(1), 1–34.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science, 28*(6), 979–1008.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., et al (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience, 4*(6), 651–655.

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science, 16*(2), 80–84.