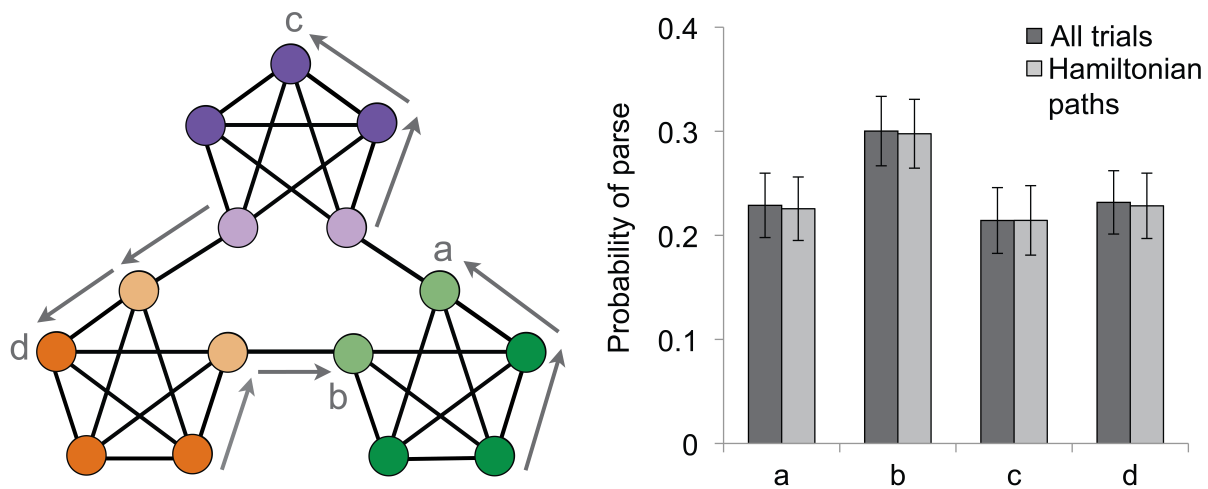


Supplementary Material for:

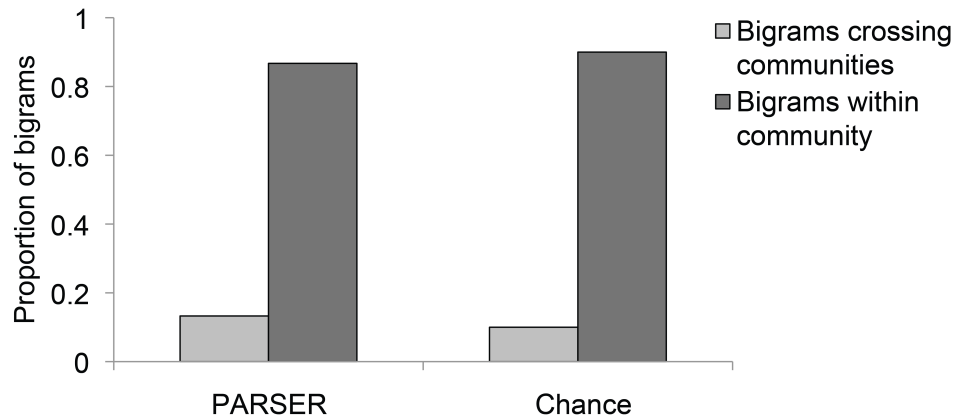
Neural representations of events arise from temporal community structure

Anna C. Schapiro, Timothy T. Rogers, Natalia I. Cordova, Nicholas B. Turk-Browne,
and Matthew M. Botvinick

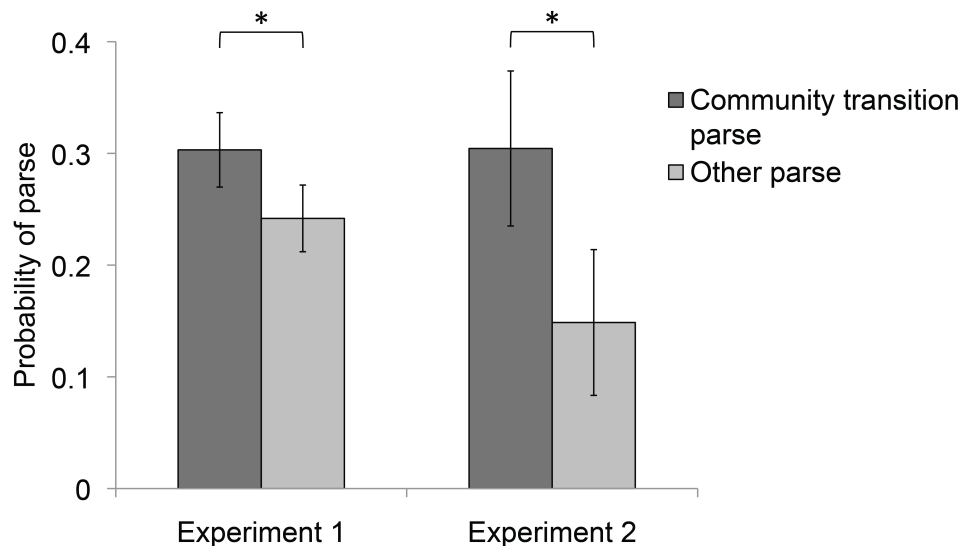


Supplementary Figure 1. Markov property of the sequences and the role of non-adjacent dependencies.

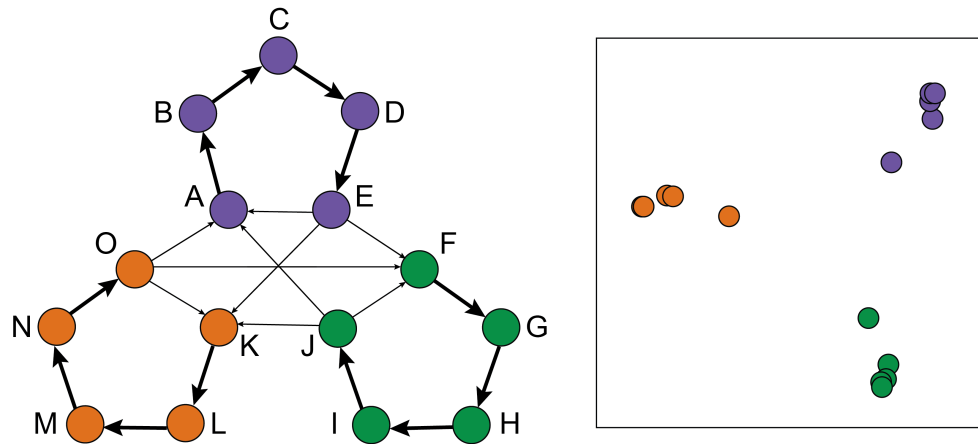
The sequences in our experiments were designed to ensure that the number and probability of immediate successors was uniform across all items, so that sequences could not be parsed on the basis of “surprising” adjacent transitions. It is possible, however, that the sequences could be parsed on the basis of surprising non-adjacent dependencies. Because the sequences were Markov, items two steps (lag-2) or further in the past provided no additional information about the upcoming item beyond that provided by the immediately preceding (lag-1) item. Thus n^{th} -order predictions were identical to first-order predictions as long as participants paid attention to every item, as their cover task performance indicates they did. Of course, the fact that lag-2 items provided no additional information given the lag-1 item does not necessarily mean that participants were insensitive to lag-2 correlations, or that such correlations were not used in parsing decisions. It is also true that, if participants ignored lag-1 items and used only lag-2 correlations to make predictions, then they would be more surprised by a transition into a new community than by a within-community transition. While this lag-2 hypothesis is logically coherent, it is untenable for three reasons. First, previous research suggests that non-adjacent learning is unlikely in our task, since such learning requires either perceptual similarity between the non-adjacent elements (Creel, Newport, & Aslin, 2004, *J Exp Psychol Learn Mem Cogn*) or that the task involves processing the non-adjacency (Pacton & Perruchet, 2008, *J Exp Psychol Learn Mem Cogn*), neither of which applies here. Second, there was no difference in cover task reaction time (RT) during Hamiltonian paths at community boundaries compared to elsewhere in the sequence (see Supplementary Table 1). Thus, non-adjacent dependencies (of any lag) do not appear to cause surprise at the transition into new communities (insofar as surprise would slow responses on the cover task). Third, although the lag-2 hypothesis could account for the parsing behavior we observed, it also predicts — with equal strength — parsing behavior that did *not* occur. In the labeled graph *left*, our theory and the lag-2 hypothesis both predict parsing at item *b* (boundary node coming two steps after an internal node from a different community) more often than item *a* (boundary node coming two steps after an internal node from the same community). The lag-2 hypothesis, however, also predicts parsing at item *d* (internal node coming two steps after a boundary node from a different community) more often than *c* (internal node coming two steps after a boundary node from the same community), since the probability of observing item *c* two steps after the boundary node in the same community is 2/16 and the probability of observing item *d* two steps after the boundary node in the previous community is 1/16 (more surprising). The graph, *right*, shows the probability of parses at items *a*, *b*, *c*, and *d* after traversing the corresponding paths illustrated on the left (averaged in each case over all isomorphic node-path combinations), collapsed across Experiments 1 and 2. There was no difference in parsing between *d* and *c*, both for all trials ($t[39] = 1.24$, $p = 0.224$) and for Hamiltonian paths in isolation ($t[39] = 0.870$, $p = 0.390$), and there was a significant interaction ($b - a > d - c$), both for all trials ($t[39] = 2.49$, $p = 0.017$) and for Hamiltonian paths in isolation ($t[39] = 2.47$, $p = 0.018$). Thus, we did not find evidence in favor of the lag-2 hypothesis. Error bars denote ± 1 s.e.m.



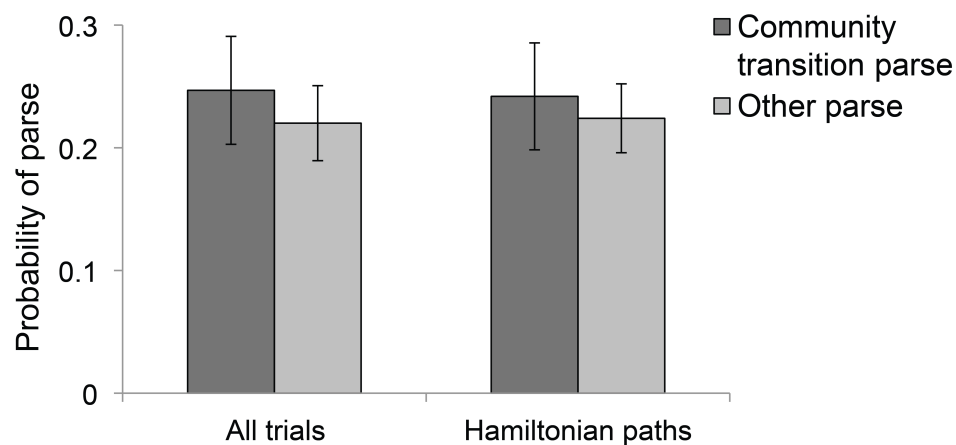
Supplementary Figure 2. Chunking models. The artificial grammar learning (AGL) literature has given rise to the idea that sequential structure is learned through chunking, i.e., the discovery of recurring subsequences (Perruchet & Vinter, 1998, *J Mem Lang*; Servan-Schreiber & Anderson, 1990, *J Exp Psychol Learn Mem Cogn*). Although chunking has typically been applied to findings concerning judgments of grammaticality, some work has applied it to sequence segmentation (Perruchet, Vinter, Pacteau, & Gallego, 2002, *Q J Exp Psychol A*). It is therefore of interest to consider whether a chunking account might explain the parsing behavior in our experiments. An important aspect of the sequential regime we studied is that, within it, all subsequences (bigrams, trigrams, etc.) occur with equal expected frequency. Based on this, we reasoned that any model that depends on the discovery of high-frequency subsequences should fail to account for our parsing results. To test this, we applied PARSER (Perruchet & Vinter, 1998, *J Mem Lang*) — widely regarded as the most advanced chunking model from the AGL literature — to sequences matching those occurring in our experiments. We used the latest implementation of PARSER (<http://leadserv.u-bourgogne.fr/~perruchet/>) with the default parameters (decay 0.05, interference 0.005). As in our experiments, the sequences presented to PARSER consisted of 15 elements ordered via a random walk on our graph. The model was trained in 100 independent runs, each involving a randomly generated sequence of the same length as those in Experiments 1 and 2. In 99 simulations, PARSER identified 15 single-item ‘chunks.’ In the one remaining simulation, 14 of the 15 items were identified as single-item chunks. In 78 simulations, at least one bigram chunk was identified (one chunk in 39 simulations; two in 28 simulations; and three in 11 simulations). Four simulations also identified one trigram chunk. Importantly, however, the chunks discovered did not reflect the community structure of the graph: Among the bigram chunks identified, 111 involved nodes lying within the same graph community, and 17 involved nodes spanning two communities. The resulting ratio (17/128) does not differ significantly from the ratio expected by chance (6/60). Indeed, as the bar graph shows, the trend runs opposite to that expected if chunks tended to form within, rather than between, communities. These simulation results validate the assertion that this chunking mechanism does not discover the structure in our stimulus sequences, and that chunking in this sense cannot explain our experimental results.



Supplementary Figure 3. Ruling out an alternative chunking-based account. The results illustrated in Supplementary Figure 2 showed that a standard chunking model, PARSER, failed to discover the structure in our stimulus sequences. This simulation also suggested a more subtle alternative explanation of the parsing data, however. Although the PARSER simulation runs in aggregate did not discover the community structure in our graphs, a larger number of simulation runs did find more within-community bigram chunks than between-community chunks. That is, the *proportion* of within-community chunks discovered was no higher than expected by chance, but because there are more within-community bigrams overall, PARSER more often identified within-community chunks purely by chance. If participants in our experiments were the same in this regard, this might explain their parsing behavior. Imagine that participants parsed randomly, but with slightly *lower* probability upon presentation of the second item in a memorized bigram ‘chunk.’ Under such circumstances, the tendency for randomly formed chunks to lie within communities would yield an overall tendency to parse more often following transitions from one community to another. To test this random chunking hypothesis, we reanalyzed the data from Experiments 1 and 2 in a way that equalizes the number of potential within- and between-community bigram chunks: We repeated the main *t*-test from each experiment, comparing the proportion of parses occurring after community transitions versus parses occurring at other points. Rather than considering all other points, however, we randomly selected six within-community stimulus bigrams, and calculated the proportion of parse responses at points in the experiment where the second member of any of these bigrams was presented. This approach equates the number of within- and between-community bigrams under consideration, since there are six potential between-community bigrams. If the parsing results were driven by ‘random chunking,’ the difference in parse proportions should therefore disappear under this analysis. In contradiction to this prediction, the difference in proportions remained significant in both experiments (Experiment 1: $t[29] = 2.14$, $p < 0.05$; Experiment 2: $t[9] = 2.27$, $p < 0.05$). Error bars denote ± 1 s.e.m.



Supplementary Figure 4. Extension to non-uniform and asymmetric transition probabilities. To allow for a clean test between alternative accounts, the experiments described in the main text focused on a sequential regime where transition probabilities were uniform and temporally symmetric. These properties clearly do not apply to many sequences that learners are likely to observe in naturalistic environments. Does the community-structure account extend to situations where transition probabilities are non-uniform and/or asymmetric? A useful setting for evaluating this question comes from work on statistical learning. A number of studies have provided evidence for segmentation or boundary-detection in sequences made up of randomly interleaved subsequences, each containing a specific set of items (e.g., syllables or images) in a fixed order (Fiser & Aslin, 2002, *J Exp Psychol Learn Mem Cogn*; Saffran, Aslin, & Newport, 1996, *Science*; Turk-Browne, Scholl, Johnson, & Chun, 2010, *J Neurosci*). Abstracting from such studies, consider a stimulus stream involving 15 distinct items (designated by the letters *A* through *O*) that randomly interleaves the subsequences *ABCDE*, *FGHIJ*, and *KLMNO*. The graph, *left*, captures the transition probabilities in such a sequence. Note that, unlike the graph discussed in the main text, this graph contains directed edges with different weights, representing non-uniform and asymmetric transition probabilities. Heavy edges represent a transition probability of 1, lighter edges a probability of 1/3. Despite these differences it is clear that the graph displays community structure: It decomposes naturally into three subgraphs whose internal connections are collectively stronger than the connections between them. As in the undirected case, complex network research (Fortunato, 2010, *Physics Reports*) provides formal definitions that explicitly identify communities in directed, weighted graphs like this one, and application of these definitions yields the three communities identified by color in the figure. The neural network simulation presented in the main text showed how community structure might be detected as a natural byproduct of learning to predict future events. Preliminary work (Diuk & Botvinick, *in prep.*) indicates how a generalization of the same computational account may apply to the current scenario. Rather than focusing exclusively on the immediate successor to the current item, prediction yields a *successor representation*, a vector-format representation of predicted future events, developed in the reinforcement learning literature (Dayan, 1993, *Neural Comput*) and recently proposed to have psychological relevance (Gershman, Moore, Todd, Norman, & Sederberg, 2012, *Neural Comput*). The right panel presents results from a neural network simulation, in which the network from the main text was applied to the *A–O* domain, with localist inputs representing the 15 individual items, and with the corresponding successor representations as output targets during training. The figure shows a multi-dimensional scaling plot based on the hidden layer representations for the 15 items after training. As in the simulation reported in the main text, the community structure of the graph is reflected in the similarity structure of the network’s internal representations. Thus, the neural network model detects the community structure in both the symmetric graph with uniform transitions and in the asymmetric graph with non-uniform transitions. Other candidate mechanisms for sequence parsing (such as chunk-detection, discussed in Supplementary Figs. 2 and 3) may not succeed in both cases, so this contrast provides a useful means of discriminating alternative hypotheses. This computational result is presented here simply as an indicator of possible directions for further research. Its relevance to the cognitive and neural mechanisms underlying event segmentation is of course a matter for empirical evaluation.



Supplementary Figure 5. Parsing results from Experiment 3. Due to time restrictions, the parsing task in Experiment 3 was run during anatomical scanning, at the end of the scanning session. Debriefing forms indicated that the parsing strategy adopted by numerous participants was strongly influenced by the pulsatile auditory noise generated by the scanner (noise that is particularly salient during anatomical imaging). The results may thus be difficult to interpret. We report the data here for completeness. Data for one participant were lost due to button-box malfunction during the parsing task, and the following results are thus based on the remaining pool of 19 subjects. The mean difference between parsing frequency at boundary traversals versus other points was 3%. This difference was not statistically significant (all trials: $t[18] = 0.968$, $p = 0.346$; Hamiltonian paths: $t[18] = 0.614$, $p = 0.547$), though it also was not significantly smaller than the differences obtained in Experiments 1 or 2 (Wilcoxon signed-rank test). Additionally, 20% of fMRI participants showed a significant effect in the predicted direction (by two-tailed chi square with Bonferroni correction), while none showed a significant effect in the opposite direction.

Exposure phase	Mean RT at boundary cross (ms)	Mean RT elsewhere (ms)	<i>t</i>	<i>p</i>
Experiment 1	671	652	1.79	0.084
Experiment 2	743	695	2.01	0.076
Experiment 3 pre-scan	798	757	4.43	0.0003
Experiment 3 scan non-Hamiltonian paths	900	820	3.82	0.001
Experiment 3 scan Hamiltonian paths	865	857	0.886	0.387

Supplementary Table 1. Reaction times on cover task during exposure phases. In the random walk exposure phases of Experiments 1, 2, and 3 (pre-scan and scan non-Hamiltonian paths), the orientation judgment at a boundary cross was marginally or significantly slower than at other locations in the sequence. This result may, however, reflect a form of repetition priming: During the random walks, items were often repeated in close temporal proximity many times within a community, potentially facilitating the orientation judgment. On this account, an item from a new community does not benefit from priming, leading to slower responses. The Hamiltonian paths in the Experiment 3 scan phase provide a test of this account, since items were not repeated in these sequences, but the transitions between communities were preserved. Within Hamiltonian paths, the RT for orientation judgments at boundary crosses was no different than at other locations in the sequence. Insofar as RT indexes surprise, this result supports our claim that participants were not *surprised* to be passing into a new community, i.e., that they did not experience a prediction error at event boundaries. In all exposure phases, there were no differences in the accuracy of orientation judgments at boundary crosses compared to other locations in the sequence.

Criterion	Proportion of community transition parses	Community transition parse vs. other parse	
		t	p
Experiment 1			
At least one item	0.288	1.93	0.063
At least two items	0.295	2.06	0.049
At least three items	0.295	2.06	0.049
At least four items	0.303	2.27	0.031
At least five items	0.302	2.21	0.035
Experiment 2			
At least one item	0.304	2.29	0.048
At least two items	0.315	2.31	0.046
At least three items	0.315	2.31	0.046
At least four items	0.319	2.30	0.047
At least five items	0.315	2.28	0.049

Supplementary Table 2. Parsing criteria. For parsing analyses, we operationalized passage into a new community as involving arrival into any community following at least four consecutive steps in another single community. The imposition of this four-step restriction was based on an *a priori* prediction — independent of our central theory — that participants might be reluctant to press the 'parse' button twice in close temporal succession. The specific choice of four steps was based on the fact that this criterion was met by two thirds of all boundary-traversal events. Very similar results were obtained, however, from re-running our parsing analyses for Experiments 1 and 2 using dwell-time criteria of one through five items before a boundary cross.