

# A lexical approach for identifying behavioral action sequences

Gautam Reddy<sup>a,1</sup>, Laura Desban<sup>b</sup>, Hidenori Tanaka<sup>c</sup>, Julian Rousset<sup>b</sup>, Olivier Mirat<sup>b</sup>, and Claire Wyart<sup>b,1</sup>

<sup>a</sup>NSF-Simons Center for Mathematical & Statistical Analysis of Biology, Harvard University, Cambridge, MA 02138; <sup>b</sup>Institut du Cerveau et de la Moelle épinière (ICM), Inserm U 1127, CNRS UMR 7225, Sorbonne Université, Paris, France; <sup>c</sup>Department of Applied Physics, Stanford University, Stanford, CA

This manuscript was compiled on January 27, 2020

**Comparative analyses in ethology across stimulus environments or genetic variants often require identifying subtle variations in behavioral action sequences ('motifs'). A challenging inferential problem lies in finding such motifs, which represent recurring sequences of shorter, stereotyped elementary maneuvers and are manifested as few copies of noisy patterns interspersed with other unknown sequences and erratic movements. Here, we propose a lexical model of animal behavior, where we view behavior as being composed of noisy instantiations of motif templates from an unknown dictionary. We develop a novel, statistical physics-inspired, unsupervised algorithm "BASS" to identify and segment motifs from high-throughput behavioral data. When applied to zebrafish larvae, our lexical model better explains than a Markov model the basic exploratory behavior and reveals a dictionary of unusually long motifs consisting of repeats and mixtures of slow forward and turn bouts. We further investigated a novel aversive chemotaxis assay where fish chemotax yet display no major differences in kinematic parameters. BASS revealed that fish avoid aversive cues by implementing a conserved transient chemotactic response consisting of sequences of fast large-angle turns and burst swims. Our approach allows us to characterize the functional significance of specific action sequences for solving a behavioral task. BASS can be easily incorporated into existing behavioral analysis pipelines and also be used as a generic algorithm for motif discovery in any sequential data that has a low-dimensional embedding.**

ethology | behavioral action sequences | motif discovery | chemotaxis

A major challenge in ethology is to infer the behavioral algorithms used by animals to perform tasks necessary for their survival. Inference of behavioral responses in natural environments is non-trivial when the input stimulus is unknown and uncontrolled, leading to an issue of unspecified context, and is further encumbered by stochasticity at various levels in the animal's execution of the response. A statistical analysis of coarse-grained observables from tracked behavioral data may exhibit statistically significant changes induced by the stimulus, but inferring behavioral responses often requires a precisely-controlled setting. One increasingly common computational approach is to leverage recent developments in the automated tracking of postural dynamics (1–5). These methods exploit clusters in low-dimensional embeddings of postural dynamics to describe behavior as a sequence of elementary maneuvers or behavioral 'syllables' drawn from a probabilistic model. The resulting descriptions parallel language models, containing information about local dynamics in the form of a probabilistic syntax over individual syllables (6–15).

While such descriptions provide useful insight into an animal's behavioral repertoire, the differences in behavior across different environments or genetic variants are often quite subtle, making comparative analyses difficult and uninterpretable.

The key difficulty lies in that the majority of behavioral responses are transient and occur only a few times in the dataset. By focusing on capturing short time-scale dynamics, dynamical models miss low-copy-number, behaviorally relevant patterns. Such long-correlated stretches are lost in the noise and are difficult to pick out from a large dataset. To give a simple example, consider a scenario where one is presented with a control 'behavioral' dataset consisting of a sequence of 100,000 fair coin tosses, and a treatment dataset which is otherwise statistically identical except for 25 sequences of 20 consecutive tails placed at random locations within the sequence. By eye, the sequences in the treatment dataset clearly stand out as abnormal, relevant stretches. On the other hand, a Markov model on heads and tails, say, when fit to the treatment data may indeed show a statistically significant deviation in its transition matrix from the control, but does not point to the nature of the abnormal stretches or where to find them.

We adopt an alternative, overlapping viewpoint, where we view behavior as being composed of recurring action sequences, which we call motifs, resulting in a lexical description of behavior as a chain of 'words' independently drawn from a dictionary with no attention paid to dynamics, i.e., there is no syntax. Note that motifs and syllables are sometimes used synonymously. To fix terminology, we define motifs as recurring action sequences of shorter, well-defined syllables. Motifs arguably contain more meaning in the context of a behavioral algorithm, much like words in English (16). While inferring the syntax of a hierarchical model is not impossible using syllable-based models such as those used in machine transla-

## Significance Statement

Animals in the wild perform characteristic motor sequences during a task, for example, the surge-and-cast of a male moth while it searches for a female or that of a soaring bird spiraling up a thermal. How can we find such conserved yet transient action sequences from noisy behavioral data? To address this question, we develop an unsupervised algorithm to extract an animal's action sequence repertoire in a manner analogous to how young children learn language from speech. Applying this approach on larval zebrafish, we uncover a sequence of fast large-angle turns and burst swims that fish use to escape from an aversive environment. The algorithm is broadly applicable.

GR performed theoretical research. GR and CW conceived the project and wrote the manuscript with inputs from all authors. LD, JR, HT, GR collected data. GR & HT analyzed the behavioral data. OM optimized the tracking algorithm (Zebrazoom).

The authors declare no conflict of interest.

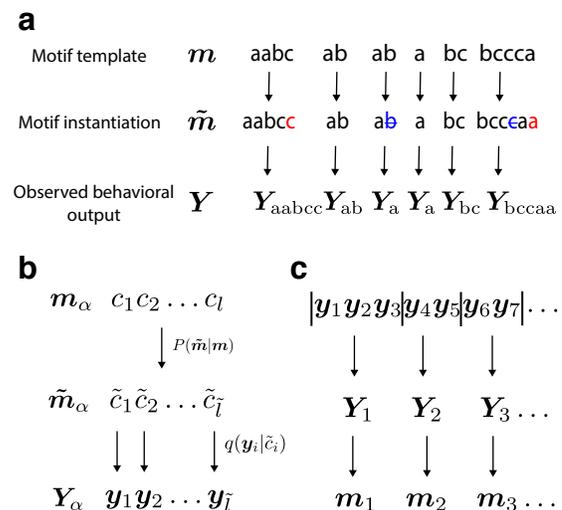
<sup>1</sup> To whom correspondence should be addressed. E-mail: gautam\_nallamala@fas.harvard.edu or claire.wyart@icm-institute.org.

tion (17), such models require orders of magnitude more data than existing experimental paradigms can offer. Importantly, inferring a hierarchical structure over a set of motifs is much more data-efficient, particularly for smaller dictionaries. On the other hand, a dictionary of motifs is not usually available; it is unclear what fraction of an animal's behavioral repertoire is composed of motifs and how this fraction changes across freely-behaving animals and those solving a specific task. To begin, motifs have to be first segmented out from data and fluctuations quantified before they can be used to construct higher-order word-based models.

Here, we develop a novel lexical model of behavior and an unsupervised method (Behavioral Action Sequence Segmentation or BASS) to discover and construct dictionaries of motifs from behavioral data. We assume, as is common for state-space models, that short time-scale postural dynamics can be mapped onto a set of elementary maneuvers that appear as clusters in a lower dimensional space, which has indeed been shown in a variety of systems including rodents, flies, worms and zebrafish larvae (8, 13, 18–20). These elementary maneuvers then form an 'alphabet'. The clustered behavioral time series yields a soft symbolic representation as probability vectors over the alphabet, which is then amenable to statistical segmentation methods that further break it up into identified motifs (of arbitrary length), while taking various sources of noise into account.

If elementary maneuvers are represented by symbols, one straightforward approach to motif discovery is to enumerate over-represented sequences of  $n$  symbols ( $n$ -grams). However, the memory and computation time required for this approach increases exponentially with  $n$ . Often-used compression methods (21–23) optimize an altogether different "coding" objective, which do not necessarily lead to meaningful motifs; for example, the two-symbol word  $ab$  could be identified as a motif simply because  $a$  and  $b$  occur often, even if  $a$  and  $b$  occur next to each other purely by chance. An alternative approach, similar to ours, is to maintain a set of possible sequences (in the form of a dictionary (24) or a suffix tree (25–27)) and add a new motif  $m_1 m_2$  to this set by concatenating two existing motifs  $m_1$  and  $m_2$  only if they are juxtaposed more often than chance. The resulting model can be viewed as an infinite-order Markov model, where only the paths through state space that show non-trivial temporal dependencies are stored.

However, the complexity of behavioral data prevents the direct application of the latter class of methods developed for bioinformatics and text processing. In these applications, one is presented a well-defined sequence of letters (AGTC or the English alphabet) and with little variability in instantiations of a particular word (words are rarely misspelled). We identify three sources of variability that impair typical methods of motif discovery in behavioral data: (1) Action pattern noise, which is the variability in instantiations of a particular motif template, (2) Syllable noise, i.e., the variations in observed output, which may lead to a syllable appearing as a similar one, and (3) Background variability due to rare behaviors and erratic movements. To make an analogy with speech learning (28), our task is similar to learning new words from *spoken* language (with no distinctive pauses separating the words) and given prior knowledge of phonology. Action pattern noise, in this analogy, corresponds (not exclusively) to stutters in speech, syllable noise to substitutions of similar phonemes (for

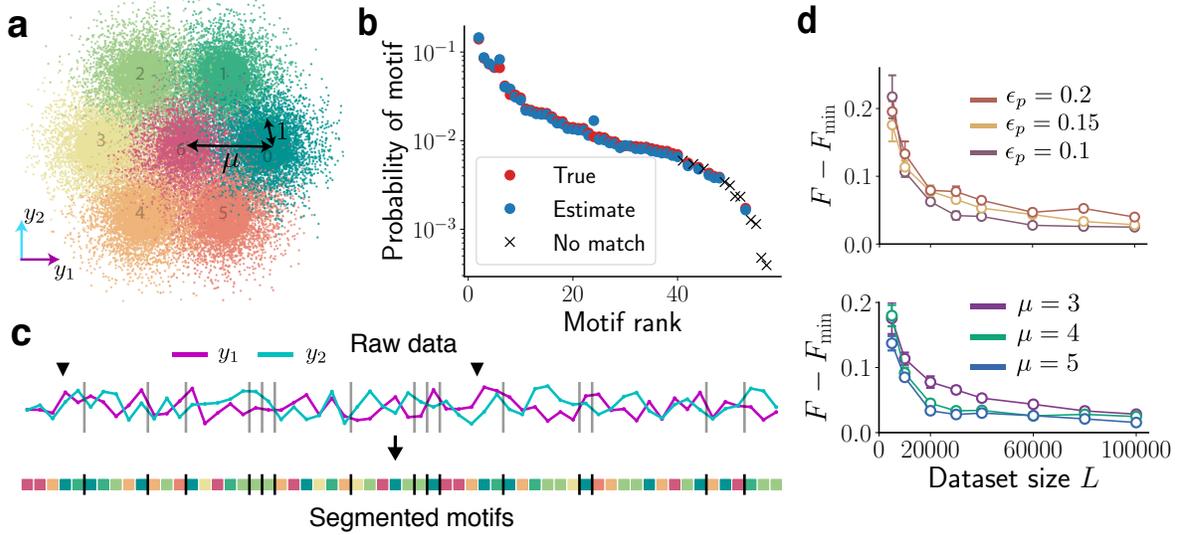


**Fig. 1.** The generative model from motifs to behavioral output. (a) Motif templates are fixed sequences of behavioral syllables (labeled a,b and c in this example). The observed behavioral output is generated from motif templates drawn sequentially from a dictionary. An instantiation of a template may "mutate" by insertions (red) or deletions (blue), which then generates the observed output as shown in panel (b). (b) The generative process from a motif template  $c_1 c_2 \dots c_l$  to instantiation  $\tilde{c}_1 \tilde{c}_2 \dots \tilde{c}_l$  to observed output  $y_1, y_2, \dots, y_l$ . (c) The unsupervised inference procedure (BASS) first learns a dictionary of motifs and then segments (vertical bars) the observed behavioral output  $y_1, y_2, \dots$  into the most likely sequence of motifs  $m_1, m_2, \dots$  from the dictionary that generated it.

example, the aspirated  $/p^h/$  and the unaspirated  $/p/$ ), and background noise to the utterance of unique proper nouns or unusual sounds. To overcome these challenges, we generalize the modeling framework in ref. (24) by introducing an additional two-level hierarchical model, the lower level mapping observed behavioral data to a latent state space and the second level introducing a model for noisy instantiations of motifs (Figure 1a). Despite the model's complexity, we show that inference is tractable and motifs efficiently extracted.

Zebrafish is an interesting vertebrate model organism to investigate the emergence of behavioral action sequences. Starting from five days post fertilization, in order to survive, zebrafish larvae actively explore their environment for food using stereotypical maneuvers consisting of bouts of activity lasting few hundreds of milliseconds separated by distinct pauses (13, 14, 20, 29). The small size enables the recording of numerous larvae in parallel, leading to the collection of thousands of swim bouts in a few minutes. Using our lexical approach, we first investigate the behavioral action sequences, i.e., the stereotyped sequences of bout types that zebrafish larvae use to spontaneously explore their environment. Next, we take advantage of a novel chemotaxis assay in which larvae navigate in arenas with gradients of acidic pH and effectively avoid acidic regions. The behavioral response that results in aversive chemotaxis is unknown. Moreover, the classical examination of global kinematic parameters reveals only minor differences, which makes identifying the chemotactic response challenging and thus makes for an appropriate benchmark for our approach.

We first develop the lexical model and the motif identification algorithm, BASS. We apply the algorithm to synthetic data and to datasets obtained from freely exploring and chemotactic zebrafish larvae. A comparison of the dictionaries in



**Fig. 2.** BASS accurately identifies and segments motifs in noisy, synthetic data: (a) The seven clusters from which the two-dimensional data (along  $y_1, y_2$ ) is drawn. (b) The true probabilities of the motifs (red dots) and probabilities estimated (blue dots) by our algorithm showing successful reconstruction of the dictionary. The crosses are low-probability motifs not identified by the algorithm (see main text). (c) A snippet of the raw data sequence and the most likely partitioning into motifs found by the algorithm. The vertical bars delineate two successive motifs. The black arrows mark two instantiations of the same length-five motif. (d) The difference in the negative log-likelihood per symbol after convergence when the true dictionary is unknown ( $F$ ) and known ( $F_{\min}$ ). Action pattern noise  $\epsilon_p$  and syllable noise  $\mu$  are successfully integrated out with larger datasets. Top:  $\mu = 3, p_d = 0.5$ , Bottom:  $\epsilon_p = 0.15, p_d = 0.5$ . Errors bars are s.e.m.

147 the two environments is then made to identify the sequences  
148 that larvae use to chemotax.

## 149 Results

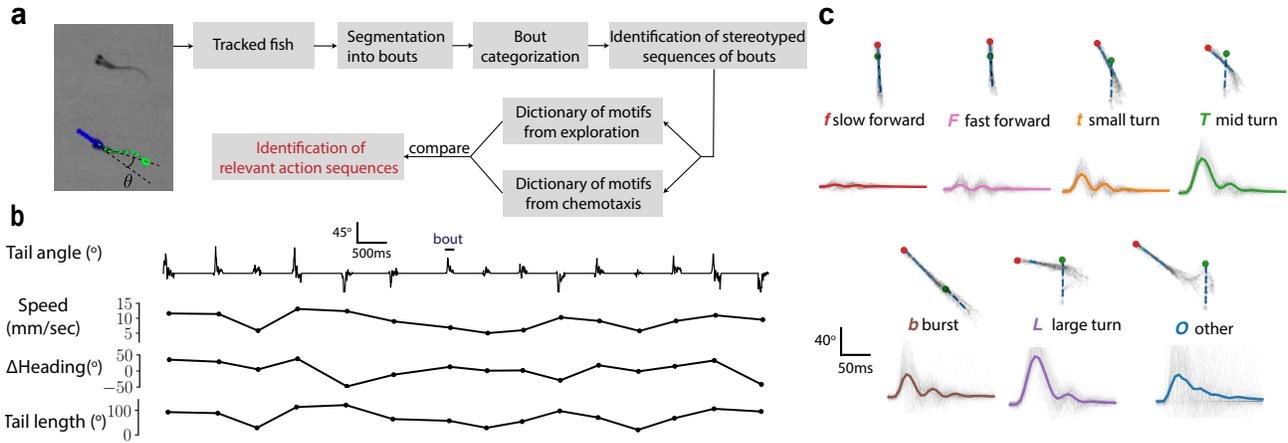
150 **A lexical model of animal behavior.** Much like language, we  
151 assume the behavior of an animal in a particular environment  
152 can be described by a sequence of motifs drawn from a dictio-  
153 nary  $\mathcal{D}$ , where each motif is a string of arbitrary length  
154 containing characters from an alphabet. Motifs are to be  
155 considered as *templates* for the generation of action sequences.  
156 Each of the  $K$  characters (which represent behavioral syllables)  
157 in the alphabet corresponds to the unique label of one of the  
158  $K$  soft clusters that define the elementary maneuvers, usually  
159 defined in a lower dimensional embedding of postural space.  
160 The character likelihood function  $q(\mathbf{y}|c)$  specifies the proba-  
161 bility of observing a maneuver  $\mathbf{y}$  corresponding to the label  $c$ .  
162 The implicit assumption here is the existence of well-defined  
163 elementary maneuvers; we may relax this assumption and  
164 instead consider clustering as a tiling of postural space, which  
165 would manifest as additional noise and a larger alphabet. We  
166 do not address the details of finding an appropriate clustering  
167 scheme, which is often non-trivial; we refer to reviews on the  
168 topic (3–5).

169 Behavior is generated from motif templates, which are  
170 sequentially sampled independently and identically from a dis-  
171 tribution  $\{p_m\}$  over the motifs in the dictionary and individual  
172 characters (Figure 1a). The independence of successive motifs  
173 arises from the lack of syntax in our model. The inclusion  
174 of individual characters accounts for movements that are not  
175 part of any motif, for instance, rare behaviors and erratic  
176 movements. These movements constitute background noise  
177 that impair motif identification since a motif  $\mathbf{m} = c_1 c_2 \dots c_l$  is  
178 detectable only if its likelihood is comparable to its constituent  
179 characters,  $p_m \gtrsim \prod_i p_{c_i}$ . Given a sequence of motifs, the data  
180 is generated from each template  $\mathbf{m}$  according to the motif

likelihood function  $Q(\cdot|\mathbf{m})$  defined below, which is a central  
181 element of the model. 182

183 The probability,  $Q(\mathbf{Y}_\alpha|\mathbf{m}_\alpha)$ , of an observed output pat-  
184 tern  $\mathbf{Y}_\alpha = \mathbf{y}_1 \mathbf{y}_2 \dots \mathbf{y}_l$  given a motif template  $\mathbf{m}_\alpha = c_1 c_2 \dots c_l$   
185 (Figure 1b) defines the behavioral output generated by  $\mathbf{m}_\alpha$ .  
186 A motif template can be viewed as the averaged trajectory  
187 of a stochastic dynamical system traversing through a state  
188 space. We then introduce a model for ‘pattern noise’, which  
189 corresponds to one where in a particular realization, the tra-  
190 jectory spends a longer or shorter duration at certain regions  
191 of state space, but does not deviate into distant regions of  
192 state space. In particular, in each instantiation,  $\mathbf{m}_\alpha$  ‘mutates’  
193 to  $\tilde{\mathbf{m}} = \tilde{c}_1 \tilde{c}_2 \dots \tilde{c}_l$  with probability  $P(\tilde{\mathbf{m}}|\mathbf{m}_\alpha)$ . The output  
194  $\mathbf{y}_i$  is drawn independently for each character in the mutated  
195 sequence from  $q(\mathbf{y}_i|\tilde{c}_i)$ . To quantify pattern noise, we fix the  
196 probability of error per character that results either in the  
197 deletion or duplication of that symbol. Note that syllable noise  
198 is implicitly incorporated in the character likelihood,  $q(\mathbf{y}|c)$ ,  
199 and is determined by the discriminability of neighboring states.  
200 We derive a recursive equation for the efficient calculation of  
201  $Q(\mathbf{Y}_\alpha|\mathbf{m}_\alpha)$  (see SI Appendix).

202 Performing inference on this model requires constructing  
203 the dictionary  $\mathcal{D}$  as well as estimating the motif probabilities  
204  $\{p_m\}$ . To build our dictionary, we use an iterative procedure  
205 generalized from ref. (24) to our latent space model, where  
206 we start from a dictionary with only single characters and  
207 progressively add words based on how often smaller sub-words  
208 occur next to each other. In particular, we cycle between:  
209 (1) estimating  $\{p_m\}$  using maximum likelihood estimation  
210 (MLE), (2) expanding  $\mathcal{D}$  if certain pairs of motifs occur next  
211 to each other more often than you would expect from  $\{p_m\}$ ,  
212 (3) truncate shorter motifs from  $\mathcal{D}$  that are “explained away”  
213 by the addition of the longer motifs into the dictionary. We  
214 briefly expand on these three steps; see SI Appendix for further  
215 details.



**Fig. 3.** Analysis of larval zebrafish behavior in exploratory and aversive environments. (a) Overview of the analysis pipeline.  $\theta$  is the tail angle. (b) A time series of the tail angle showing the discrete nature of bouts. The corresponding speed, change in heading and the tail length (summed absolute amplitude of the tail angle) for each of the bouts are shown. (c) Samples of the seven bout types identified using a Gaussian Mixture Model. The green and red dots correspond to the head position at the beginning and end of the bout. Below each sample, the average tail angle is also shown in solid color with 200 trajectories shown in grey.

Given a behavioral dataset  $\mathbf{Y} = \mathbf{y}_1\mathbf{y}_2 \dots \mathbf{y}_L$ , the sequence of motif templates that generate it are unknown. For example, if  $L = 3$ , we have  $\mathbf{Y} = \mathbf{y}_1\mathbf{y}_2\mathbf{y}_3$ , whose likelihood is obtained by summing over all possible ways the dataset can be partitioned:  $Q(\mathbf{y}_1)Q(\mathbf{y}_2)Q(\mathbf{y}_3) + Q(\mathbf{y}_1)Q(\mathbf{y}_2\mathbf{y}_3) + Q(\mathbf{y}_1\mathbf{y}_2)Q(\mathbf{y}_3) + Q(\mathbf{y}_1\mathbf{y}_2\mathbf{y}_3)$ , where each marginal probability factor in each term is from an instantiation of a particular motif template. In general, the likelihood of  $\mathbf{Y}$  under our generative model is the sum over all possible partitionings  $\{\pi\}$  of the dataset (of which there are  $2^{L-1}$ ) into observed data sequences  $\{\mathbf{Y}_\alpha^\pi\}$ , weighted by the likelihood of each partitioning:

$$P(\mathbf{Y}; \{p_m\}) = \sum_{\pi} \prod_{\alpha=1}^{N(\pi)} Q(\mathbf{Y}_\alpha^\pi), \quad [1]$$

where the marginal probability is  $Q(\mathbf{Y}_\alpha^\pi) = \sum_{\mathbf{m}} Q(\mathbf{Y}_\alpha^\pi | \mathbf{m}) p_m$  and  $N(\pi)$  is the total number of templates in partition  $\pi$ . We show (SI Appendix) that the MLE for  $p_m$  satisfies the implicit equation

$$p_m^* \propto \sum_{\pi} \sum_{\alpha'=1}^{N(\pi)} p(\mathbf{m} | \mathbf{Y}_{\alpha'}^\pi) \prod_{\alpha=1}^{N(\pi)} Q(\mathbf{Y}_\alpha^\pi), \quad [2]$$

where  $p(\mathbf{m} | \mathbf{Y}_\alpha^\pi)$  is the posterior probability of  $\mathbf{m}$  given the data and the pre-factor is determined from normalization. The sum over the posterior probabilities can be interpreted as an effective number of counts of  $\mathbf{m}$  in the partition  $\pi$ ; Eq. (2) can then be re-cast as  $p_m^* = \langle N_m \rangle / \bar{N}$ , where  $\langle N_m \rangle$  is the expected number of counts of  $\mathbf{m}$  over the ensemble of partitions and  $\bar{N} = \sum_{\mathbf{m}'} \langle N_{\mathbf{m}'} \rangle$  is the average number of partitions.

Given the large sum in Eq. (2) and the hierarchical structure of the model, it is rather surprising that the MLE can be performed efficiently. To compute  $p_m^*$ , it is useful to define the free energy,  $F \equiv -\ln P(\mathbf{Y}; \{p_m\})$ , which is to be minimized. The gradients of  $F$  can be efficiently calculated using dynamic programming methods (SI Appendix), which allows for computation of  $p_m^*$  using standard gradient descent methods. Note that the number of counts is then  $\langle N_m \rangle = -p_m \partial_{p_m} F$ . New motifs are added to the dictionary if they occur more often than expected by random concatenations of motifs already in

the dictionary. The probability of a new motif  $\mathbf{m}$  being generated through all possible concatenations of smaller motifs in the dictionary,  $\zeta(\mathbf{m})$ , is compared to the empirical probability of  $\mathbf{m}$ ,  $-\zeta(\mathbf{m}) \partial_{p_m} F / \bar{N}$ . A standard likelihood ratio test yields a  $p$ -value and pairs below a  $p$  threshold ( $10^{-3}$ ) are added to the dictionary. Motifs which have low counts or which are similar to other motifs are discarded.

An implementation of BASS is publicly available (30).

**An illustration on synthetic data.** To illustrate the generative process and the effectiveness of the method in identifying and segmenting motifs, we first apply it to a synthetically generated dataset. We assume individual data points are two-dimensional (representing a lower-dimensional embedding of postural dynamics) and are drawn from 7 distinct states (which make up the characters in our alphabet) with a Gaussian emission function as shown in Figure 2a. A dictionary of 50 motifs is constructed such that each motif has a mean length of five. Given the generated dictionary, the probability of each motif,  $p_m$ , is drawn and scaled with a parameter  $1 - \epsilon_b$ , where  $\epsilon_b$  is the fraction of the dataset that is made up of individual characters. We use  $\epsilon_b$  as a measure of ‘background noise’. Sequential data is sampled according to the lexical model, with  $\epsilon_p$  as a measure of action pattern noise and syllable noise  $\mu$ , defined as the distance between neighboring clusters (Figure 2a). In the sample shown, we use  $L = 40000$ ,  $\epsilon_p = 0$ ,  $\epsilon_b = 0.5$ ,  $\mu = 3$ .

On this dataset, the algorithm builds a dictionary containing 44 motifs with 11 false negatives and 6 false positives. Of the 11 false negatives (crosses in Figure 2b), 8 occur fewer than 25 times in the entire dataset. The three other false negatives (632, 631, 421325, see Figure 2a for cluster labels) were in fact closely related to the three highest probability false positives (32,31,421335). The estimated probabilities of the true positive motifs match very well with their true probabilities (Figure 2b) despite significant background and syllable noise. A snippet of the raw data is shown in Figure 2c along with the most likely partitioning into motifs from the learned dictionary (SI Appendix). With larger datasets, the method robustly integrates out fluctuations due to significant

272 action pattern and syllable noise  $\epsilon_p$  and  $\mu$  (Figure 2d). BASS  
 273 found no motifs in shuffled data, as expected.

274 We now apply BASS to larval zebrafish behavior in ex-  
 275 ploratory (pH neutral) and aversive (acidic) chemotaxis assays.  
 276 An outline of our analysis pipeline is shown in Figure 3a.

277 **A dictionary for freely exploring zebrafish larvae.** Zebrafish  
 278 larvae swim in short punctuated bouts (duration mean  $\pm$  s.d  
 279 =  $150 \pm 50$  ms) separated by longer periods of rest (mean  
 280  $\pm$  s.d =  $700 \pm 500$  ms) (Figure 3b). Larvae explore their  
 281 environment by performing mainly slow bouts occurring as  
 282 forward swims and routine turns, often by repeating turns in  
 283 the same direction (31), and rarely exhibit fast bouts such as  
 284 burst swims or escapes (13, 20). We collected a dataset of  
 285  $\approx 85000$  bouts from exploring fish ( $\approx 180$  fish) swimming in  
 286 an elongated well geometry.

287 A single bout is well-characterized by the fish’s tail move-  
 288 ment and other kinematic variables such as average speed and  
 289 change in heading. From raw tracking data (SI Appendix),  
 290 we use a six-dimensional parameterization  $\mathbf{y}$  for each bout,  
 291 which includes the speed, the change in heading, the tail length  
 292 (summed absolute amplitude of the tail angle) and the first  
 293 three principal components of the tail angle over time (SI  
 294 Appendix). Based on this parameterization, bouts were cate-  
 295 gorized into different bout types using a Gaussian Mixture  
 296 Model (GMM). A GMM yields the likelihood function,  $q(\mathbf{y}|c)$ ,  
 297 for each category  $c$ , which serves as a statistical description  
 298 of each category in terms of the means and covariances of the  
 299 six variables. We clustered bouts into seven categories (Figure  
 300 3c, SI Movie S1,S2), which correspond to two forward swims  
 301 of different speeds ( $f$ , slow and  $F$ , fast), three turns based on  
 302 the magnitude of change in heading ( $t, T$  and  $L$ , increasing  
 303 angle), bursts ( $b$ ) and an other ( $O$ ) category. The  $O$  category  
 304 contained a variety of different bouts that did not clearly fall  
 305 into one class; these included O-bends, long turns and bursts,  
 306 and improperly tracked bouts. The categories are not sharply  
 307 delineated; this is not an issue for the BASS algorithm since  
 308 variability in  $\mathbf{y}$  is implicitly taken into account via  $q(\mathbf{y}|c)$  as  
 309 noted before.

310 Typical bout types are displayed in Figure 3c. Compared to  
 311 previous categorizations performed on spontaneous exploration  
 312 (13, 14, 29), our categories (except  $O$ ) likely correspond to  
 313 sub-divisions of forward swims, routine turns and burst swims.

314 The seven categories make up the alphabet of our generative  
 315 model. Sequences of consecutive bouts for each fish ( $\sim 600$   
 316 bouts per fish) served as input to BASS. A coarse exploration of  
 317 the pattern noise parameter  $\epsilon_p$  and the probability of insertion  
 318  $p_d$  using a held-out dataset yielded  $\epsilon_p = 0.1$  and  $p_d = 0.2$ ,  
 319 which were used for the rest of our analysis. While these  
 320 numbers suggest noisy motif instantiation and a bias towards  
 321 insertions (i.e., repeats), precise estimates of these parameters  
 322 require cross-validation from independent methods.

323 The algorithm converged to a dictionary consisting of 66  
 324 motifs (with similar results across trials and subsamples). A  
 325 subset of these motifs are shown in Table 1 (see also Table  
 326 S1). In Figure 4a,b, a sample sequence of bouts segmented  
 327 into a sequence of motifs is presented.

328 The dictionary has several surprising features. A significant  
 329 fraction ( $\sim 74\%$ ) of the dataset was made of motifs. Motifs  
 330 as large as 14 bouts were found (which may further expand  
 331 in a particular realization due to insertions). In particular,  
 332  $f$  repeated 14 times occurred more than 500 times. While

**Table 1. Motifs over-represented in the exploratory dataset.**

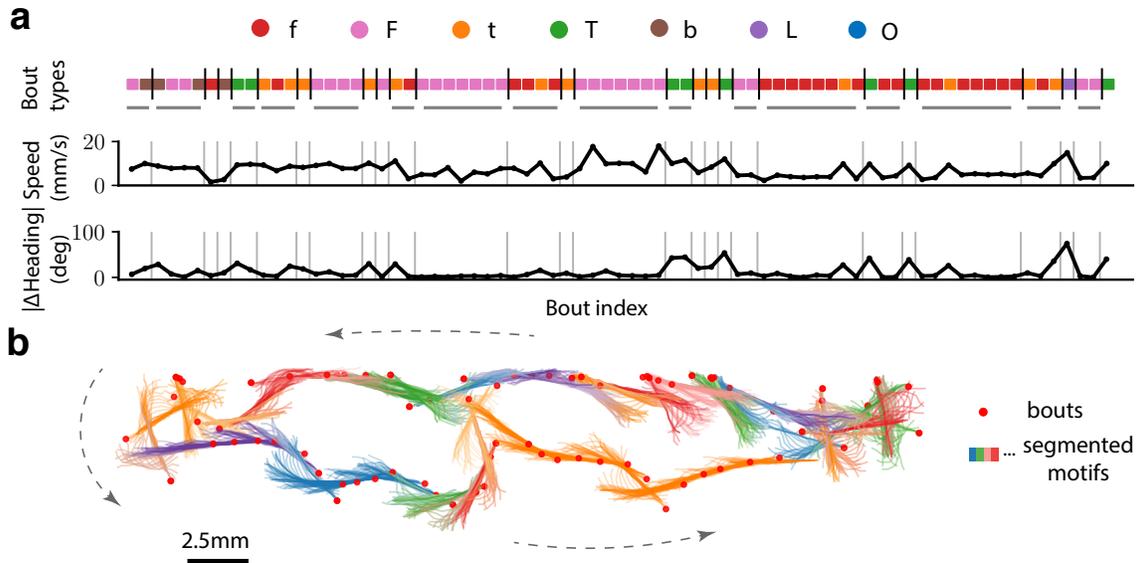
Motifs	$-\log_{10} p$	Observed	Expected
ffffffff	>300	1366	387
fffffffffff	>300	510	50
ffffff	208.01	3234	1797
ffff	42.33	9544	8327
FFFFFFF	28.23	311	153
fffffft	27.64	497	290
ftffff	25.07	495	297
ftfff	22.72	1125	824
ftff	21.12	1745	1377
ftf	18.5	2724	2289
ft	13.96	4337	3859
TfT	11.12	722	554
FFFF	7.94	1428	1224
TfTf	7.28	346	254
tttt	6.7	256	181
TTTT	5.06	160	110
bb	3.87	924	1044
bbbb	3.21	115	82
FbFb	2.19	99	74

A subset of motifs occur (‘Observed’ column) more often than  
 predicted by a first-order Markov model (the ‘Expected’ column). The  
 $p$ -value is obtained using a likelihood ratio test. See also Table S1.

this may be explained by the large fraction of  $f$ , repeats were  
 also found for  $T$ ,  $F$  and  $b$ . Overall, the most enriched and  
 common motifs correspond to repetitions of the same bout  
 type, and typically occur 2-14 times in a row. Motifs containing  
 mixtures of bouts included typically 2 different bout types.  
 The  $f$ ,  $t$ ,  $T$  bout types typically correspond to the slow regime  
 of locomotion,  $F$ ,  $b$  and  $O$  belong to the fast regime (13, 20).  
 Notably, throughout the list of enriched motifs, all bouts  
 forming a particular motif belonged either to low speed or to  
 high speed, but not a mixture of the two.

To quantify how unusual these sequences were under a  
 Markov model, we compared the observed occurrence of the  
 identified motifs to those predicted from the best-fit Hidden  
 Markov Model (HMM). Our lexical model yielded a better  
 fit compared to an HMM (difference in held-out free energy  
 per bout of 0.12), and a significant portion of motifs deviated  
 from Markovianity (Tables 1,S1). The non-Markovianity likely  
 arises from two sources: First, while long repeats of the same  
 bout type occur often, the distribution of the number of repeats  
 has a heavy tail and decays much slower than a geometric  
 distribution. Second, sequences with mixtures of two bout  
 types such as  $TfTf$  and  $ftf$  are common; while the repeats  
 emphasize (say)  $f \rightarrow f$  transitions, the motifs with mixtures  
 of bout types on the other hand emphasize  $f \rightarrow t$  transitions,  
 creating a tension between the two in a purely Markovian  
 picture.

To verify that the long chain of repeats were not an arti-  
 fact due to our elongated well geometry, we applied a similar  
 pipeline of bout categorization and motif identification on a  
 previously published dataset (13) (see SI Appendix, Figure  
 S2). The dataset consists of  $\approx 120,000$  bouts (23 fish) obtained  
 from fish freely swimming in a square well (of side  $\sim 25$ mm)  
 under varying light intensities. Notably, the resulting diction-  
 ary also displays long chains of repeats and significant  
 non-Markovianity albeit with a heavier emphasis on turns  
 compared to forward swims (Table S2). Mixtures of turns and



**Fig. 4.** Motifs identified by BASS make up a significant fraction of the dataset. (a) A sample sequence of 75 bouts from the exploratory data segmented (separated by vertical bars) into the most likely sequence of motifs from the learned dictionary. The corresponding speed and absolute change in heading are shown. Motifs longer than one character are underlined in gray. (b) A sample trajectory consisting of 80 bouts (head position at beginning of bout in red dots) are segmented into motifs (head and tail at each frame are shown), where successive bouts from the same motif have the same color.

369 slow forward swims ( $ffTf$ ,  $TfTf$ ), as well as fast forward and  
 370 burst swims ( $FbFb$ ,  $bbFb$ ) are also present in this dictionary,  
 371 while mixtures of *slow* bouts (forward swims or turns) and  
 372 *fast* bouts (forward or burst swims) are conspicuously absent  
 373 in both dictionaries.

374 **Fish chemotax away from acidic pH using conserved se-**  
 375 **quences of fast bursts and large avoidance turns.** Recent  
 376 studies have investigated how zebrafish respond to the acute ap-  
 377 plication of aversive or appetitive chemicals in the surrounding  
 378 water (32–34). However, the behavioral responses of freely-  
 379 swimming zebrafish larvae navigating in chemical gradients  
 380 of aversive or appetitive cues have not yet been investigated.  
 381 Acid was applied to the two ends of the arena forming a sharp  
 382 gradient (SI Appendix); diffusive transport at the time scale of  
 383 the experiment (ten minutes) is at most 1cm and therefore is  
 384 confined to the ends. Zebrafish larvae successfully performed  
 385 chemotaxis and avoided the two extremities (Figure 5a), yet  
 386 displayed only minor differences in kinematic parameters typ-  
 387 ically used in analyses (Figure 5b). The over-representation  
 388 of certain bout categories shown in Figure 5c suggests that  
 389 fish perform more burst  $b$ , and fast turns  $T$  and  $O$  bouts in  
 390 response to the aversive gradient. However, the sequence of  
 391 actions the fish takes in order to chemotax is unknown.

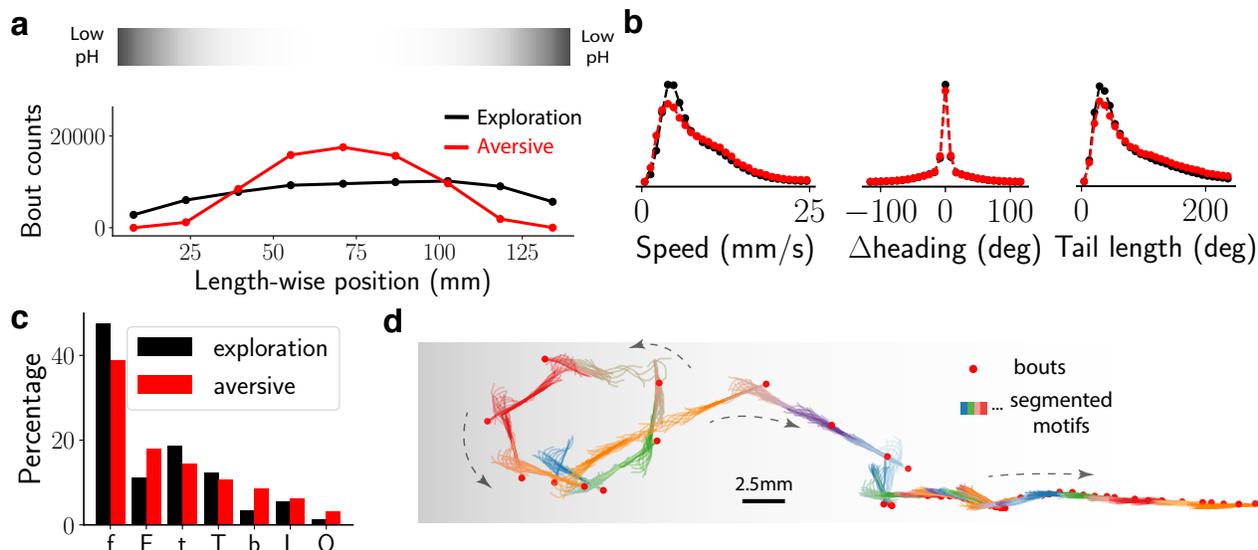
392 We implemented a comparative approach of finding con-  
 393 served *sequences* of actions that are highly over-represented  
 394 in the aversive environment compared to exploration. We  
 395 applied the BASS algorithm to the dataset from fish in the  
 396 aversive environment (~66,000 bouts from ~100 fish). The  
 397 resulting dictionary of motifs contained a total of 81 motifs,  
 398 slightly larger than the one obtained from exploration (Table  
 399 S3). The two dictionaries contain broad similarities: both  
 400 contain long repeats of the same bout type and mixtures of  $t$   
 401 and  $f$ , yet contain important differences, particularly in the  
 402 over-representation of mixtures of  $b, F$  and  $O, b$  in the aversive  
 403 environment.

**Table 2. Motifs consistently over-represented in the aversive chemo-  
 taxis assay.**

Motifs	$-\log_{10} p$	$\langle N_m \rangle_{aver}$	$\langle N_m \rangle_{expl}$
fTff	90.84	101	5
OO	66.63	111	11
bb	55.84	210	55
bOOb	46.06	63	5
Ob	35.74	135	36
bO	34.6	140	39
bFbb	24.96	49	6
Obbb	24.07	56	9

404 We examined over-represented sequences by comparing the  
 405 relative occurrences of motifs in the exploratory and aversive  
 406 environments. The two dictionaries were combined to obtain  
 407 a total of 103 unique motifs and the expected number of  
 408 occurrences of each motif,  $\langle N_m \rangle$ , for the two environments.  
 409 Intuitively,  $\langle N_m \rangle$  is the number of times a motif occurs after  
 410 appropriately discounting its occurrences within a longer motif.  
 411 We use  $-\log_{10} p$  as a measure of over-representation, where  
 412  $p$  is obtained from a likelihood ratio test on  $\langle N_m \rangle$  in the  
 413 two environments. To calibrate this score, we first split the  
 414 exploratory dataset into halves and computed the  $-\log_{10} p$   
 415 for each motif; the threshold 15 was chosen for a false positive  
 416 rate of 10%. To ensure that sequences from a few abnormal  
 417 fish did not dominate our comparison, we sub-sampled our  
 418 dataset from the aversive environment to 80% its size ten times,  
 419 performed the comparison for each sub-sample, and chose only  
 420 those motifs above threshold in *all* ten sub-samplings.

421 Table 2 shows the eight over-represented motifs conserved  
 422 across fish. All motifs except one ( $fTff$ ) are mixtures of  $b$   
 423 and  $O$ . The bouts from the selected motifs (except  $fTff$ ) are  
 424 significantly over-represented close to the aversive gradient  
 425 compared to the rest of the bouts in the aversive environment  
 426 (Figure 6a), though no such selection was explicitly imposed



**Fig. 5.** Fish avoid an aversive environment using a transient chemotactic response. (a) Histogram of larva positions along the well with and without the aversive (acidic) gradient, located at the ends of the well. An illustration of the aversive gradient is shown above. (b) The distribution of speed, change in heading and the tail length of all bouts during exploration (black) and in aversive environment (red). The difference in global kinematic parameters between the two environments is small. (c) The fraction of each bout type in exploratory and aversive environments, where a total of  $\approx 85000$  and  $\approx 66000$  bouts were collected respectively. (d) BASS segments a sequence of bouts from the aversive environment into a sequence of motifs. Shown here is a sample trajectory (as in Figure 4b) where the fish escapes from the aversive environment.

427 *a priori* in our analysis. To verify that these motifs were  
 428 indeed the ones involved in aversive chemotaxis, we computed  
 429 the distance per bout the fish travels in the direction down  
 430 the aversive gradient for bouts within motifs flagged as over-  
 431 represented and the rest of the bouts. Figure 6b shows a highly  
 432 significant bias for the over-represented bouts for swimming  
 433 down the gradient. Strikingly, when bouts from the flagged  
 434 sequences that begin at the two extreme quarters of the well  
 435 excluding *fTff* are considered, the bias greatly increases, even  
 436 beyond the typical length-wise distance travelled per bout,  
 437 indicating that these bouts are triggered by the sharp acidic  
 438 gradient and are almost exclusively aimed down the gradient.  
 439 In contrast, the motif *fTff*, which was also reliably observed  
 440 across fish, was not directed away from the aversive gradients  
 441 nor did it preferentially occur close to the gradient, suggesting  
 442 that *fTff* is induced by the global acidification of the swim  
 443 arena but not as a direct response to the acidic gradient (see  
 444 SI Movie S3 for a sample of the motif *fTff*).

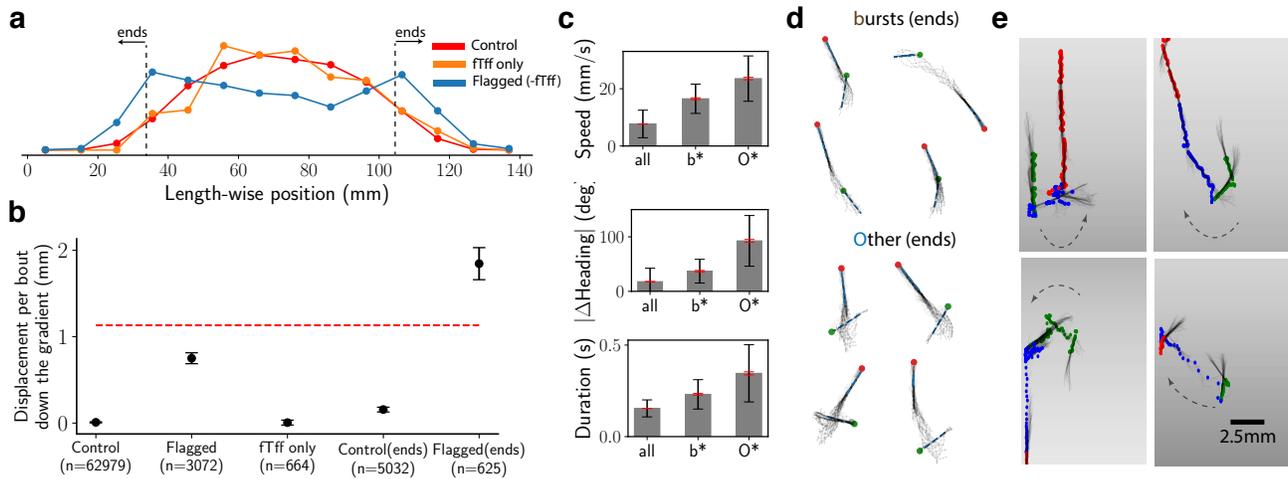
445 Further examination of the bouts from the *b* and *O*  
 446 categories implicated in chemotaxis showed that bouts from both  
 447 categories were significantly longer and the fish swam faster  
 448 compared to the unflagged bouts (Figure 6c). Inspection of  
 449 the tail movements from this subset of *O* bouts showed that  
 450 these bouts typically consisted of a large-angle avoidance turn  
 451 followed by a long burst swim (see Figure 6d for examples).  
 452 The induced sequences of fast bouts composed of *b* and *O* near  
 453 the edges of gradient are distinct from the recently described  
 454 slow avoidance response to  $\text{CO}_2$  (32). The previous report of  
 455 a lack of behavioral response to  $\text{HCl}$   $\text{pH}=4.5$  ((32)) suggests  
 456 that *b* and *O* may be elicited for more acidic pH.

## 457 Discussion

458 In this study, we present a lexical model of animal behavior,  
 459 where we view observed behavior as a composition of recur-  
 460 ring motif templates drawn from a dictionary. We develop

the BASS algorithm for performing inference on this model, which ultimately yields a dictionary of motifs that the animal performs in its designated environment. Applying the method on data from exploring zebrafish larvae revealed a long time-scale organization of bout sequences that cannot be explained in a Markovian model on single bouts. In an aversive chemotaxis task, we identified conserved sequences of bouts that the fish employ to escape an aversive environment. We argue that our model yields complementary insight to traditional dynamical models, and is better suited for comparative behavioral analyses, particularly for comparisons across animals in similar environments and closely related genetic variants. While the generative model of independent motifs is perhaps the simplest one, it is rather notable that it fits better to our behavioral data than a syllable-based Markov model with a similar number of parameters, the latter often used to depict quantitative ethograms. Indeed, in his seminal paper, Lashley (16, 35) rejects the *reflex chain* theory, which posits, in modern terminology, a first-order Markov model for the dynamics of movements in favor of a model based on noisy motifs.

Our model generalizes past work on motif discovery from bioinformatics, machine learning and time series analysis by incorporating two important generative processes crucial for behavioral modeling that were not modeled previously: elevating motif templates to latent variables and introducing a data-generating process for elementary maneuvers. This generalization is necessary to take into account the significant variability in behavioral data, where clusters in postural space are not always well-defined and erratic movements are not uncommon. It should be noted that the generative process for *motif templates* can also be viewed as a particular hidden Markov model, where the  $|\mathcal{D}| + 1$  hidden states at the topmost level are the motifs in the dictionary and the ‘background’. In this picture, the full generative model has three levels of hierarchy, which we have not explicitly represented for con-



**Fig. 6.** Fish chemotaxis via sequences of fast bursts and large-angle avoidance turns. (a) Distributions of length-wise positions for the Control bouts (all bouts from aversive environment except the ones from sequences flagged as over-represented in Table 2), *fTff* only and Flagged bouts (from the sequences in Table 2) except *fTff* in red, orange and blue respectively. (b) The length-wise displacement travelled in a bout down the gradient for bouts tagged as Control, Flagged (as defined in (a)), *fTff* only, Control(ends) (all bouts from the two ends of well shown in (a)), Flagged(ends) (flagged but with *fTff* removed and in the ends of the well). For scale, the red, dashed line shows the mean length-wise distance per bout for unflagged bouts. Error bars are s.e.m. (c) The mean speed, change in heading and duration of the bouts (black, red error bars for s.d., s.e.m. respectively) from *b* and *O* categories part of the Flagged(ends) sequences from (b). (d) Four random samples of each bout type from the flagged sequences in (a). The green and red are the head positions at the beginning and end of the bout respectively. (e) Four random samples from the Flagged(ends) sequences, which include the three bouts before and after the flagged sequence. The green, blue and red dots show head position for the bouts before, during and after the flagged sequence respectively. Note that the depicted gradient is illustrative.

497 ciseness. Several extensions of the model are possible. Prior  
 498 knowledge can be easily applied. For instance, priors on the  
 499 distributions of motif lengths or the distribution of frequencies  
 500 can be introduced by weighting different partitions or a Dirichlet  
 501 prior on the probabilities, respectively. In both cases, an  
 502 MLE equation (or a MAP estimate in the latter case) similar  
 503 to Eq. (2) can be derived. More complex, hierarchical models  
 504 over motifs may be learned by noticing that the Markovian  
 505 structure of the partitioning is compatible with structured  
 506 variational approximations (36). Further structure can be  
 507 introduced into the ‘background’, which we have assumed is  
 508 made of independently drawn characters, similar to those used  
 509 in bioinformatics (37).

510 Importantly, BASS receives no explicit information about  
 511 the stimulus experienced by the animal; the extracted motifs  
 512 therefore contain no information about the precise stimulus-  
 513 response map of the animal, but may reveal relevant qualitative  
 514 aspects of the animal’s behavior. For example, BASS may  
 515 discover the surge-and-cast motion of a male moth searching  
 516 for a female (38) or the spiraling of a soaring bird (39) with  
 517 no reference to what stimulus triggers those responses.

518 In the same spirit, we show here that during chemotaxis,  
 519 freely-swimming zebrafish larvae in a gradient of aversive  
 520 cues exhibit conserved sequences of 2-4 fast bouts mixing  
 521 burst and large-angle avoidance turns to swim away from  
 522 the aversive environment. Notably, these sequences make  
 523 up only a small fraction of the dataset (~ 0.2%), yet are  
 524 successfully captured in our analysis. Further, we found that  
 525 zebrafish larvae exploring either a rectangular arena (this  
 526 study) or confined square arenas (13) repeat specific bout  
 527 types, particularly forward and turn swims respectively, for  
 528 ~3-8 iterations. This observation is consistent with prior  
 529 observations of repetition of turns in freely swimming larvae  
 530 (31). The discovery of highly-specific sequences of either slow  
 531 forward or turn bouts suggests that the descending command

signals sent to the spinal cord underlying forward bouts (40)  
 or turns (41) is maintained over tens of seconds, possibly  
 via sustained inputs to reticulospinal neurons or via sensory  
 feedback, and may have important implications for foraging.

BASS can be easily incorporated into existing behavioral  
 analyses pipelines alongside the expanding repertoire of meth-  
 ods for unsupervised behavioral clustering (1-5). In its current  
 form, our implementation can handle datasets of size ~ 300,000  
 bouts and dictionaries of size ~ 500, beyond which approxima-  
 tions for scalable inference have to be developed. A discussion  
 on the statistical power lent by this method for comparative  
 analyses is necessary, but is beyond the scope of this work; such  
 an analysis is non-trivial due to individual-to-individual vari-  
 ability and other environmental factors. We further highlight  
 the connection between behavioral modeling and genomics,  
 where a wealth of algorithms have been developed. Exploiting  
 this connection may lead to a fruitful exchange of techniques  
 between the two seemingly disparate fields. Finally, we re-  
 mark that our method is an addition to a rapidly enlarging  
 computational toolkit for extracting mechanistic answers to  
 behavioral questions.

## Materials and Methods

The code for BASS and the scripts used to reproduce the figures are  
 publicly available (30). The repository includes the six-dimensional  
 data used for bout categorization and BASS analysis. Due to its  
 large size, the raw tracking data has not been uploaded and is  
 available at request.

**ACKNOWLEDGMENTS.** We thank Monica Dicu and Antoine  
 Arneau for fish care in the Phenoparc animal core facility of ICM,  
 Maxime Kermarquer for use of the ICM cluster, Kristen D’Elia,  
 Sophia Horowitz and Clara Besserer for trouble-shooting at the early  
 stages of the chemotaxis experiments, Bing Brunton, Elena Rivas  
 and Massimo Vergassola for useful comments, and Joao Marques

- 565 and Michael Orger for sharing their zebrafish larvae dataset. This  
566 research was initiated during the summer school “Neural computa-  
567 tion for sensory navigation” in the Kavli Institute of Theoretical  
568 Physics (University of California in Santa Barbara, USA) and was  
569 supported in part by the National Science Foundation under Grant  
570 No. NSF PHY-1748958, NIH Grant No. R25GM067110, and the  
571 Gordon and Betty Moore Foundation Grant No. 2919.01. This work  
572 was also supported by a New York Stem Cell Foundation (NYSCF)  
573 Robertson Award 2016 Grant #NYSCF-R-NI39, the HFSP Pro-  
574 gram Grants #RGP0063/2018, the Fondation Schlumberger pour  
575 l’Education et la Recherche (FSER/2017) for C.W. The research  
576 leading to these results has received funding from the program “In-  
577 vestissements d’avenir” ANR-10-IAIHU-06 (Big Brain Theory ICM  
578 Program) and ANR-11-INBS-0011 – NeurATRIS: Translational Re-  
579 search Infrastructure for Biotherapies in Neurosciences. GR was  
580 partly supported by the NSF-Simons Center for Mathematical &  
581 Statistical Analysis of Biology at Harvard (award number #1764269)  
582 and the Harvard Quantitative Biology Initiative.
- 583 1. S. R. Datta, D. J. Anderson, K. Branson, P. Perona, A. Leifer, Computational Neuroethology:  
584 A Call to Action, *Neuron*, 104(1):11-24, 2019.
  - 585 2. D. J. Anderson & P. Perona, Toward a Science of Computational Ethology, *Neuron*, 84:18-31,  
586 2014.
  - 587 3. A. Gomez-Marín, J. J. Paton, A. R. Kampff, R. M. Costa & Z. F. Mainen, Big behavioral  
588 data: psychology, ethology and the foundations of neuroscience, *Nature Neuroscience*,  
589 17(11):1455-1462, 2014.
  - 590 4. A. E. X. Brown & B. de Bivort, Ethology as a physical science, *Nature Physics*, 14(7):653-657,  
591 2018.
  - 592 5. G. J. Berman, Measuring behavior across scales, *BMC Biology*, 16(23), 2018.
  - 593 6. G. J. Berman, W. Bialek & J. W. Shaevitz, Predictability and hierarchy in *Drosophila* behavior,  
594 *Proc. Natl. Acad. Sci.*, 113(42):11943-11948, 2016.
  - 595 7. A. Sharma, R. E. Johnson, F. Engert & S. W. Linderman, Point process latent variable models  
596 of larval zebrafish behavior, *Advances in Neural Information Processing Systems* 32, 2018.
  - 597 8. A. B. Wiltschko et al, Mapping Sub-Second Structure in Mouse Behavior, *Neuron*, 88:1121-  
598 1135, 2015.
  - 599 9. B. C. Daniels, W. S. Ryu & I. Nemenman, Automated, predictive, and interpretable inference  
600 of *Caenorhabditis elegans* escape dynamics, *Proc. Natl. Acad. Sci.*, 116 (15), 7226-7231,  
601 2019.
  - 602 10. G. J. Stephens, B. Johnson-Kerner, W. Bialek, W. S. Ryu, Dimensionality and Dynamics in  
603 the Behavior of *C. elegans*, *PLoS Computational Biology*, 4(4):1-10, 2008.
  - 604 11. G. J. Stephens, M. B. de Mesquita, W. S. Ryu & W. Bialek, Emergence of long timescales and  
605 stereotyped behaviors in *Caenorhabditis elegans*, *Proc. Natl. Acad. Sci.*, 108(18):7286-7289,  
606 2011.
  - 607 12. A. J. Calhoun, J. W. Pillow & M. Murthy, Unsupervised identification of the internal states that  
608 shape natural behavior, *Nature Neuroscience*, 22, 2040–2049, 2019.
  - 609 13. J. C. Marques, S. Lackner, R. Felix & M. B. Orger, Structure of the Zebrafish Locomotor  
610 Repertoire Revealed with Unsupervised Behavioral Clustering, *Current Biology*, 28:181-195,  
611 2018.
  - 612 14. D. S. Mearns, J. C. Donovan, A. M. Fernandes, J. L. Semmelhack & H. Baier, Deconstructing  
613 Hunting Behavior Reveals a Tightly Coupled Stimulus-Response Loop, *Current Biology*, 30:1-  
614 6, 2020.
  - 615 15. A. Costa, T. Ahamed & G. J. Stephens, Adaptive, locally linear models of complex dynamics,  
616 *Proc. Natl. Acad. Sci.*, 116(5): 1501-1510, 2019.
  - 617 16. K. S. Lashley, The problem of serial order in behavior. In *Cerebral mechanisms in behavior*,  
618 L. A. Jeffress Eds., (New York: Wiley, 1951), pp. 112–131.
  - 619 17. I. Sutskever, O. Vinyals & Q. V. Le, Sequence to Sequence Learning with Neural Networks,  
620 *Advances in Neural Information Processing Systems* 27, 2014.
  - 621 18. G. J. Berman, D. M. Choi, W. Bialek & J. W. Shaevitz, Mapping the stereotyped behavior of  
622 freely moving fruit flies, *J. R. Soc. Interface*, 11, 2014.
  - 623 19. A. E. X. Brown, E. I. Yemini, L. J. Grundy, T. Jucikas & W. Schafer, A dictionary of behav-  
624 ioral motifs reveals clusters of genes affecting *C. elegans* locomotion, *Proc. Natl. Acad. Sci.*,  
625 110(2):791-796, 2013.
  - 626 20. O. Mirat, J. R. Stemberg, K. E. Severi & C. Wyart, ZebraZoom: an automated program for  
627 high-throughput behavioral analysis and categorization, *Frontiers in Neural Circuits*, vol.7, art.  
628 104:1-12, 2013.
  - 629 21. C. G. Nevill-Manning & I. H. Witten, On-line and off-line heuristics for inferring hierarchies of  
630 repetitions in sequences. *Proc. IEEE*, 88, 1745-1755, 2000.
  - 631 22. M. Ghosh & J. Rihel, Hierarchical Compression Reveals Sub-Second to Day-Long Structure  
632 in Larval Zebrafish Behaviour, *bioRxiv*, 2019.
  - 633 23. A. Gomez-Marín, G. J. Stephens & A. E. X. Brown, Hierarchical compression of *Caenorhab-*  
634 *ditis elegans* locomotion reveals phenotypic differences in the organization of behaviour, *J. R.*  
635 *Soc. Interface*, 13, 2016.
  - 636 24. H. J. Bussemaker, H. Li & E. D. Siggia, Building a dictionary for genomes: Identification  
637 of presumptive regulatory sites by statistical analysis, *Proc. Natl. Acad. Sci.*, 97(18):10096-  
638 10100, 2000.
  - 639 25. J. E. Markowitz, E. Ivie, L. Kligler & T. J. Gardner, Long-range Order in Canary Song, *PLoS*  
640 *Computational Biology*, 9(5): e1003052, 2013.
  - 641 26. D. Ron, Y. Singer & N. Tishby, The Power of Amnesia: Learning Probabilistic Automata with  
642 Variable Memory Length, *Machine Learning*, 25:117-149, 1996.
  - 643 27. G. Bejerano & G. Yona, Variations on probabilistic suffix trees: statistical modeling and pre-  
644 diction of protein families, *Bioinformatics*, 17(1):23-43, 2001.
  - 645 28. S. Goldwater, T. L. Griffiths, L. Johnson, A Bayesian framework for word segmentation: Ex-  
646 ploring the effects of context, *Cognition*, 112(1):21-54, 2009.
  - 647 29. R. E. Johnson, S. Linderman, T. Panier, K. J. Herrera, A. Miller & F. Engert, Probabilistic  
648 Models of Larval Zebrafish Behavior Reveal Structure on Many Scales, *Current Biology*, 30,  
649 1-13, 2020.
  - 650 30. <https://github.com/greddy992/BASS>. Implementation of BASS in Cython and the scripts re-  
651 quired to reproduce the figures in the paper are uploaded here.
  - 652 31. Dunn et al, Brain-wide mapping of neural activity controlling zebrafish exploratory locomotion,  
653 *eLife*, 5:e12741, 2016.
  - 654 32. T. Koide, Y. Yabuki & Y. Yoshihara, Terminal Nerve GnRH3 Neurons Mediate Slow Avoidance  
655 of Carbon Dioxide in Larval Zebrafish, *Cell Reports*, 22, 1115–1123, 2018.
  - 656 33. Chen W. Y., Peng W. Y., Deng Q. S., Chen M. J. Du J. L. & Zhang B. B., Role of Olfactorily  
657 Responsive Neurons in the Right Dorsal Habenula-Ventral Interpeduncular Nucleus Pathway  
658 in Food-Seeking Behaviors of Larval Zebrafish, *J. Neuroscience*, 404:259-267, 2019.
  - 659 34. Candelier R., Murmu M. S., Romano S. A., Jouary A., Debrégeas G. & Sumbre G., A microfluidic  
660 device to study neuronal and motor responses to acute chemical stimuli in zebrafish, *Sci*  
661 *Rep.*, July 21:5:12196, 2015.
  - 662 35. D. A. Rosenbaum, R. G. Cohen, S. A. Jax, D. J. Weiss & R. van der Wel, The problem of  
663 serial order: Lashley’s legacy, *Human Movement Science*, 26:525-554, 2007.
  - 664 36. M. I. Jordan, Z. Ghahramani, T. S. Jaakkola & L. K. Saul, An Introduction to Variational  
665 Methods for Graphical Models, *Machine Learning*, 37:183–233, 1999.
  - 666 37. N. Rajewsky, M. Vergassola, U. Gaul & E. D. Siggia, Computational detection of genomic cis-  
667 regulatory modules applied to body patterning in the early *Drosophila* embryo, *BMC Bioinform-*  
668 *atics*, 3(30), 2002.
  - 669 38. C. T. David, J. S. Kennedy & A. R. Ludlow, Finding of a sex pheromone source by gypsy  
670 moths released in the field, *Nature*, 303, 804–806, 1983.
  - 671 39. Z. Akos, M. Nagy & T. Vicsek, Comparing bird and human soaring strategies, *Proc. Natl.*  
672 *Acad. Sci.*, 105(11):4139-4143, 2008.
  - 673 40. A. Pujala & M. Koyama, Chronology-based architecture of descending circuits that underlie  
674 the development of locomotor repertoire after birth, *eLife*, eLife 8:e42135, 2019.
  - 675 41. K. H. Huang, M. B. Ahrens, T. W. Dunn & F. Engert, Spinal projection neurons control turning  
676 behaviors in zebrafish, *Current Biology*, 23(16):1566-73, 2013.