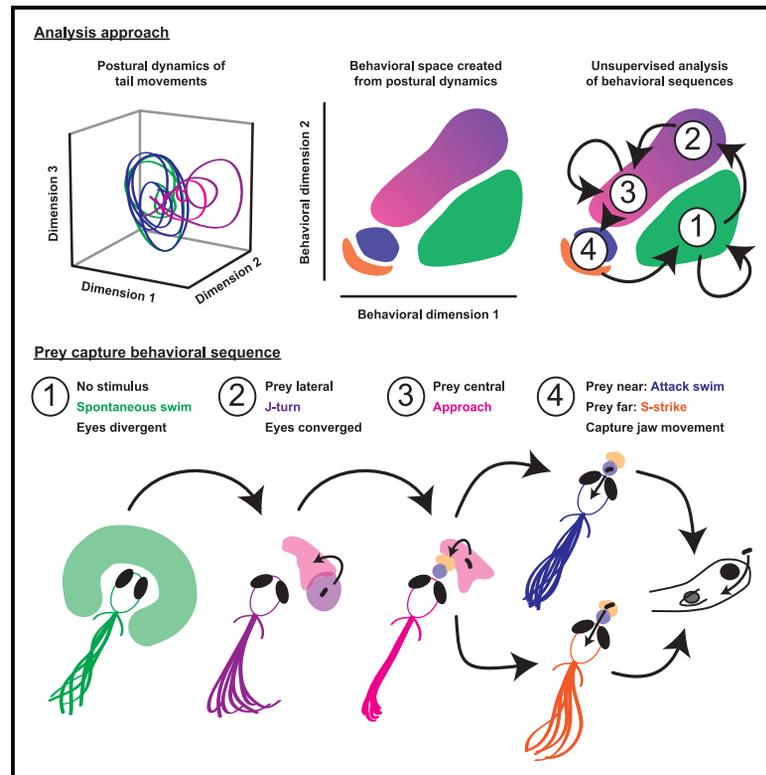


Deconstructing Hunting Behavior Reveals a Tightly Coupled Stimulus-Response Loop

Graphical Abstract



Authors

Duncan S. Mearns,
Joseph C. Donovan,
António M. Fernandes,
Julia L. Semmelhack, Herwig Baier

Correspondence

hbaier@neuro.mpg.de

In Brief

Behavioral sequences are shaped by an animal's interaction with its environment. Mearns et al. find that, during hunting, zebrafish larvae adaptively interact with their prey through continuous integration of visual information. When prey reach a precise location in the binocular visual field, larvae perform a stereotyped strike maneuver.

Highlights

- Zebrafish hunting consists of stereotyped transitions in a behavioral continuum
- Chaining successive hunting bouts depends on short-term high-acuity visual cues
- Larvae capture prey from below with stereotyped jaw and tail movements
- Precise positioning of prey in the strike zone depends on binocular visual cues



Deconstructing Hunting Behavior Reveals a Tightly Coupled Stimulus-Response Loop

Duncan S. Mearns,^{1,2} Joseph C. Donovan,¹ António M. Fernandes,¹ Julia L. Semmelhack,^{1,3} and Herwig Baier^{1,4,*}

¹Max Planck Institute of Neurobiology, Department of Genes – Circuits – Behavior, Am Klopferspitz 18, 82152 Martinsried, Germany

²Graduate School of Systemic Neurosciences, 82152 Martinsried, Germany

³Hong Kong University of Science and Technology, Division of Life Science, Clearwater Bay, Hong Kong, China

⁴Lead Contact

*Correspondence: hbaier@neuro.mpg.de

<https://doi.org/10.1016/j.cub.2019.11.022>

SUMMARY

Animal behavior often forms sequences, built from simple stereotyped actions and shaped by environmental cues. A comprehensive characterization of the interplay between an animal's movements and its environment is necessary to understand the sensorimotor transformations performed by the brain. Here, we use unsupervised methods to study behavioral sequences in zebrafish larvae. We generate a map of swim bouts, revealing that fish modulate their tail movements along a continuum. During prey capture, larvae produce stereotyped sequences using a subset of bouts from a broader behavioral repertoire. These sequences exhibit low-order transition dynamics and immediately respond to changes in visual cues. Chaining of prey capture bouts is disrupted in visually impaired (*lakritz* and *blumenkohl*) mutants, and removing the prey stimulus during ongoing behavior in closed-loop virtual reality causes larvae to immediately abort the hunting sequence. These results suggest that the continuous integration of sensory information is necessary to structure the behavior. This stimulus-response loop serves to bring prey into the anterior dorsal visual field of the larvae. Fish then release a capture strike maneuver comprising a stereotyped jaw movement and tail movements fine-tuned to the distance of the prey. Fish with only one intact eye fail to correctly position the prey in the strike zone, but are able to produce the strike itself. Our analysis shows that short-term integration of binocular visual cues shapes the behavioral dynamics of hunting, thus uncovering the temporal organization of a goal-directed behavior in a vertebrate.

INTRODUCTION

Quantitative descriptions of behavior are essential if we are to fully understand the brain [1]. Such descriptions have provided a framework for interrogating the genetic and neural basis of behavior in worms, flies, and mice [2–5]. It is believed that

complex, flexible behavior arises as a result of animals chaining together simpler, more stereotyped movements [6–8]. These movements have been termed motor primitives [9], synergies [10], movemes [11], or behavioral modules [5, 7, 12–14]. However, whether such modules truly constitute stereotyped, invariant movements or whether they merely reflect extremes in a behavioral continuum remains unclear [12, 14–17].

Animals chain their movements into sequences to achieve their desired goal. Such sequences may be stereotyped, with the behaviors chaining together in a similar way with each occurrence [18]. Alternatively, sequences may be flexible, with the ordering of modules different each time they appear. In such cases, sequences may be hierarchically organized, whereby switching between modules is stochastic but influenced by longer-term behavioral states or sensory stimuli [5, 19–21]. Classically, fixed-action patterns have described behaviors in which a single “releasing” stimulus is sufficient to drive an entire sequence of movements [8]. Alternatively, in a stimulus-response loop each movement generates a new sensory stimulus for the animal that, itself, serves as a trigger for the next movement in the chain [22].

Capturing prey is an innate and essential behavior for the survival of many animals. Zebrafish larvae hunt protists that float in the water column [23–25]. Both real and virtual prey presented to restrained animals can produce isolated orienting swim bouts and eye convergence, hallmarks of prey capture in zebrafish larvae [26–28]. Such movements could form a stimulus-response loop, whereby movements of the tail and eyes bring prey to the near-anterior visual field of the animals [16, 29]. However, it is not clear whether this would be implemented by gradual changes in the kinematics of bouts over the course of a hunting sequence [16, 23], or as a result of discrete switches between more stereotyped motor patterns [14]. One possibility is that different stages of the behavior have a different organization. For example, animals might dynamically modulate their movements to adjust to the position of the prey during pursuit, but resort to more stereotyped motor patterns when consuming prey [22]. Studies of prey capture have predominantly focused only on tail, jaw, or fin movements, and it is not known how these movements are coordinated over the behavioral sequence [16, 23, 30, 31]. Moreover, it has been speculated that binocular overlap in the visual field produced during eye convergence is necessary to bring prey into a “strike zone” that triggers the final capture bout [16, 26, 32]. The necessity of binocular vision for hunting, however, has not been tested.



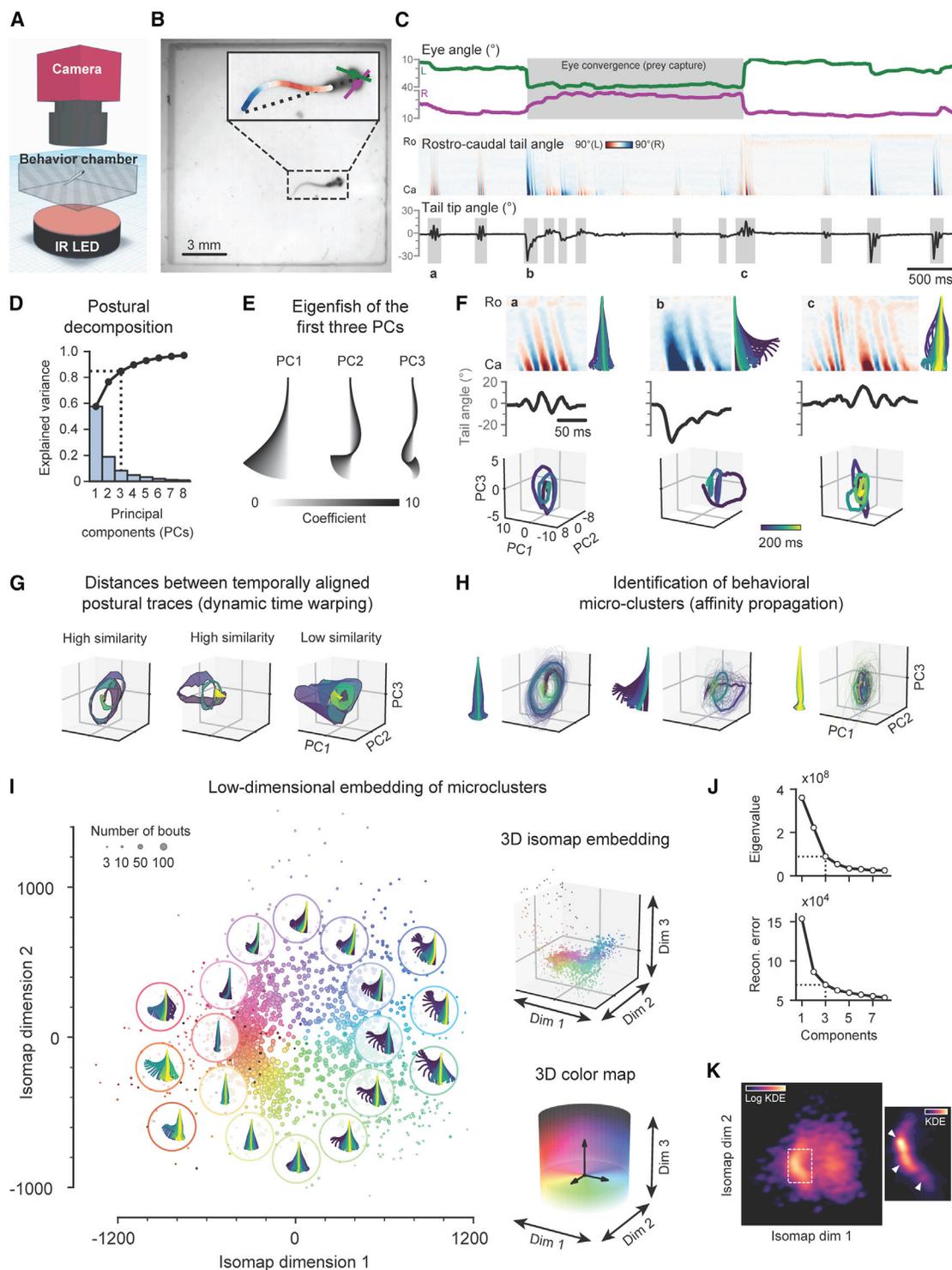


Figure 1. Zebrafish Larva Bouts Lie on a Behavioral Continuum

(A) Schematic of the setup used to record behavioral data.

(B) Example high-speed video frame. Inset overlaid with tail and eye tracking.

(C) Eye and tail kinematics extracted from 6 s of recording. Ro, rostral; Ca, caudal.

(D) Principal component analysis (PCA) of tail shapes. Explained variance (bars) and cumulative explained variance (points) of the first eight components. We retained three components (dotted line).

(E) “Eigenfish” of the first three PCs.

(legend continued on next page)

Here, we present a new computational framework for decomposing an animal's movements into kinematic and transition components. We apply our pipeline to the bouts of zebrafish larvae and find a continuum of behaviors. We use unsupervised analysis of behavioral transitions to reveal stereotyped chaining during prey capture through a subset of these bouts. These sequences emerge from a stimulus-response loop in which the fish's movements generate new stimuli that trigger subsequent bouts in the chain. Prey capture sequences are disrupted when we interfere with visual processing using genetic mutants or remove visual cues after the behavior is initiated. Further investigating the capture strike, we show discreteness in tail and jaw movements during this behavior. In addition, our approach reveals changes in the prey capture dynamics of monocularly blinded fish, impeding their ability to bring prey into the final "strike zone."

RESULTS

Zebrafish Swim Bouts Lie on a Low-Dimensional Behavioral Continuum

We sought a representation of zebrafish behavior that would reveal whether different bout types are truly stereotyped and discrete, or whether the kinematics are more graded. We recorded and tracked individual zebrafish larvae (7–8 days post-fertilization [dpf]; $n = 45$) hunting live prey (paramecia) in a custom-built behavioral arena (Figures 1A–1C; Video S1; STAR Methods). We analyzed 57,644 individual swim bouts with a pipeline involving four main steps. First, we reduced the dimensionality of our tracked data (Figures 1D–1F). Second, we computed the distance between each pair of bouts (Figure 1G). Third, we combined similar bouts into exemplars by micro-clustering (Figure 1H). Lastly, we performed non-linear embedding using the distances between these exemplars to generate the final behavioral space (Figures 1I–1K and S1).

Mechanical and neural constraints impose limits on the possible posture configurations for an animal; thus, movements are generally believed to lie on relatively low-dimensional manifolds [7, 12, 17, 33, 34]. As our tracking provides overly redundant measurements along the tail, we performed principal component analysis (PCA) on the sequence of all tail postures. Three PCs explained 85% of the variance in tail shape (Figure 1D). These PCs define postural modes and can be represented by a set of tail shapes known as "eigenfish" [17, 34, 35] (Figure 1E). As posture is dynamic over time, these shapes trace a trajectory in three-dimensional coordinate space (Figure 1F; Video S2). Retaining more than three postural modes did not

alter the final structure of the behavioral space (Figure S1A). Thus, the tail kinematics of zebrafish larva bouts can be accurately represented in this lower-dimensional space.

To measure the distance between different bouts of temporally varying postures, we used dynamic time warping (DTW), which handles temporal offsets and small variations better than Euclidean distance [36, 37] (Figure 1G). From these distances, we generated micro-clusters (each represented by a single exemplar) using affinity propagation [38] (Figure 1H). Micro-clustering serves numerous functions that aid with embedding and subsequent analysis. It robustly handles the non-metric distances obtained from DTW [38], removes rare or spurious events, and prevents common trajectories from dominating the embedding. Furthermore, we can conveniently embed data collected in different experiments by aligning to exemplars. Altering the number of exemplars did not strongly influence the structure of the final behavioral space (Figure S1B).

Next, we performed isomap embedding [39] using the DTW distances between exemplars (Figure 1I), which preserves global behavior structure preferentially to local structure. We could not discern any additional structure if we embedded all bouts, suggesting that we were not losing information about the structure of behavior by only embedding exemplars (Figure S1E). We found that three dimensions were sufficient to capture most of the variation in bout kinematics, as indicated by an elbow in both the eigenvalues and reconstruction error of the embedding (Figure 1J). Inspecting bouts that are represented in different regions of the behavioral space (Figure 1I), we found capture strikes (far left) [14, 40], forward scoots (lower left) [24, 41], routine turns (right) [24, 41], and J-turns characteristic of prey capture behavior (top) [25]. Our results suggest that turn angle, swimming speed, and tail laterality are the dominant kinematic features that define global bout structure (Figure S1F). Overall, we observed a high degree of continuity across bouts, consistent with earlier descriptions of zebrafish behavior [16, 23, 36].

To confirm that the apparent continuity in the behavior was not an artifact of the isomap embedding, we also performed t-SNE (Figure S1C) and UMAP (Figure S1D). While UMAP also revealed a more continuous behavioral space, t-SNE suggested a small degree of local clustering may be lost in our approach, mostly within the forward scoots (Figure S1C). Closer inspection of our isomap space, accounting for local structure preserved in the micro-clusters, recapitulated these local densities (Figure 1K). Thus, our results suggest a structure to zebrafish behavior that is largely continuous, with some local densities representing more stereotyped types of forward swim.

(F) Three example bouts, labeled a–c in (C). For each bout, we show curvature along rostral-caudal axis of the tail over time, trajectory through the first three PCs, and sequence of tail shapes reconstructed from PCs.

(G) Bout pairs aligned using dynamic time warping (DTW). High similarity, small sum of Euclidean distances between aligned points, represented by ribbon stretched between trajectories. Low similarity, large sum of distances.

(H) Bout micro-clusters identified with affinity propagation (thin lines). Thick line, exemplar. Left subpanels: tail movement represented by the exemplar.

(I) Isomap embedding of DTW distances between exemplars. Left: projection onto the first two dimensions. Right: projection onto first three dimensions. Points colored according to position within a hue-lightness cylinder centered on origin.

(J) Eigenvalues (top) and reconstruction errors (bottom) of the isomap embedding. We retained three components (dotted line).

(K) Kernel density estimation of bout density over first two isomap dimensions. Right: forward scoot region of the behavioral space (white dotted outline). Arrowheads: local densities.

See also Figure S1 and Videos S1 and S2.

Behavioral Sequences Are Built from a Small Number of Simple Chaining Rules

Next, we investigated the temporal organization of bout sequences for structure and stereotypy (Figure 2A). To this end, we decomposed the one-step transition frequency matrix between micro-clusters into a set of “transition modes,” each of which captures some feature of temporal behavioral dynamics (Figures 2B and S2B). Since the matrix is sparse, we first smoothed transitions using bout distances in behavioral space, as bouts with very similar kinematics likely share similar transitions (Figure S2A). We were particularly interested in three types of transition mode that can be obtained from this decomposition: the “common” mode, symmetric modes (S-modes), and anti-symmetric modes (A-modes). The common mode reflects the most commonly visited micro-clusters across all transitions. S-modes recover groups of bouts that contribute to a “sticky” behavioral state [42]. A-modes highlight dynamic transitions between groups of bouts. Each transition mode is associated with a singular value, which describes how much that mode contributes toward the overall transition structure.

The majority of transitions in our data were explained by a combination of the common mode, one S-mode, and one A-mode (Figure 2C, elbow in the singular values). To test whether the dynamics captured by these modes were consistent across animals, we tested our ability to predict the transitions performed by half of the animals using transition modes obtained from the other half (Figure 2D; STAR Methods). We found that the first S-mode (S1) improved our behavior prediction by 31% over a null model consisting only of the common mode (31.4 ± 4.5 , mean \pm SD), and the first A-mode (A1) improved our behavior prediction by 2% (2.0 ± 0.54 , mean \pm SD).

Next, we investigated which transitions were represented by each significant transition mode (Figures 2E–2G). We found that S1 separated approach swims from larger spontaneous swims and turns (Figure 2F). A1 appeared to represent transitions through J-turns, approach swims, and capture strikes, bouts classically associated with prey capture (Figure 2G). To confirm this, we used eye convergence as an independent measure of hunting behavior (Figure 2H) [26]. We identified bouts that occurred when the eyes were and were not converged (“prey capture” and “spontaneous,” respectively) and found that these correlated with S1 (Spearman’s $r = 0.71$) (Figures 2F and 2I). Furthermore, by studying changes in eye convergence, we could identify bouts that occurred at the beginning, middle, and end of a hunting sequence. This demonstrated that A1 accurately recovered behavioral sequences during prey capture ($r = 0.45$, 0.41 , 0.41 ; A1 with early, mid, late prey capture) (Figures 2G and 2J). We then recorded fish swimming in the absence of prey and mapped their bouts into our behavioral space by alignment to the exemplars. Performing SVD of this spontaneous behavior revealed a dominant common mode and no significant A-modes (Figure S3). Thus, zebrafish larvae have a unique repertoire of bouts reserved for hunting and produce stereotyped and dynamic sequences using these bouts.

Seven Clusters in the Behavioral Space Correspond to Broad Bout Types

We next assigned labels to different parts of the behavioral space. This approach has three benefits. First, it allows us to

relate different regions of our space to bout types identified through other expert- and machine-annotated descriptions of behavior [14, 41] (Figures 3B and 3C). Second, larger clusters provide more robust states for Markov chain analysis than micro-clusters (Figures 3D–3G). Third, by integrating over many bouts, we can generate maps of stimuli that are associated with different regions of the behavioral space (Figure 3H).

To generate such labels, we generated a new behavioral space, accounting for both kinematic similarity and similarity in chaining structure (Figure 3A; STAR Methods). Kinematic similarity between exemplars was determined using DTW distances (Figure 1G). Chaining similarity was determined by finding exemplars that contributed similar weights to the transition modes (Figures 2F and 2G). We multiplied these numbers and proceeded with isomap embedding to generate the new space. Hierarchical clustering separated bouts into seven types (Figure S4A), corresponding to J-turns, approach swims, “slow 1” swims, capture strikes, “slow 2” swims, high-angle turns (HATs), and routine turns described in previous studies [14, 25, 41] (Figure 3B; Video S3). J-turns, approach swims, and capture strikes occupy the prey capture region of the behavioral space and predominantly occurred when the eyes were converged (Figures 3C and S4B). “Slow 2” swims, HATs, and routine turns were performed by fish during spontaneous swimming. “Slow 1” swims contributed to both prey capture and spontaneous behavior. Thus, taking transition dynamics into account, we could relate regions of our behavioral space to known bout types of zebrafish larvae and demonstrate the differential contributions of these swims to spontaneous and prey capture behavior.

Prey Capture Sequences Are Maintained through Tight Stimulus-Response Loops

Next, we wanted to test whether transition dynamics were shaped by short- or longer-term memory processes. We constructed a family of models using our behavioral clusters as states in a Markov chain and quantified how longer memories affected prediction of the next bout in a sequence (Figures 3D and 3E; STAR Methods). The default model contained no memory, and for all bout clusters (excluding capture strikes) prediction of the following bout could be improved by including knowledge of the previous bout label (J-turn, 43%; approach, 47%; slow 1, 8%; slow 2, 71%; HAT, 13%; routine turn, 118%). Our prediction of bouts following capture strikes was generally poor, possibly due to more variability within this cluster (Figure 3B). Further increasing the memory of previous bouts in a sequence did not improve prediction for bouts following J-turns, approach swims, or “slow 2” swims, though it did slightly improve performance for “slow 1” swims, HATs, and routine turns (7%, 5%, and 26%, respectively) (Figures 1E and S4C). This indicates that the preceding bout is a strong predictor of transition dynamics, which is compatible with “memoryless” behavior driven by a stimulus-response loop.

To investigate the most common transitions underlying predictable behavioral sequences, we visualized transitions as an ethogram (Figure 3F). The most likely transitions were from J-turns and approaches to another approach (>40%). Moreover, we found elevated transition probabilities between “slow 2” swims, HATs, and routine turns (11%–42%), and high probabilities of transitioning to “slow 1” swims from most other behaviors

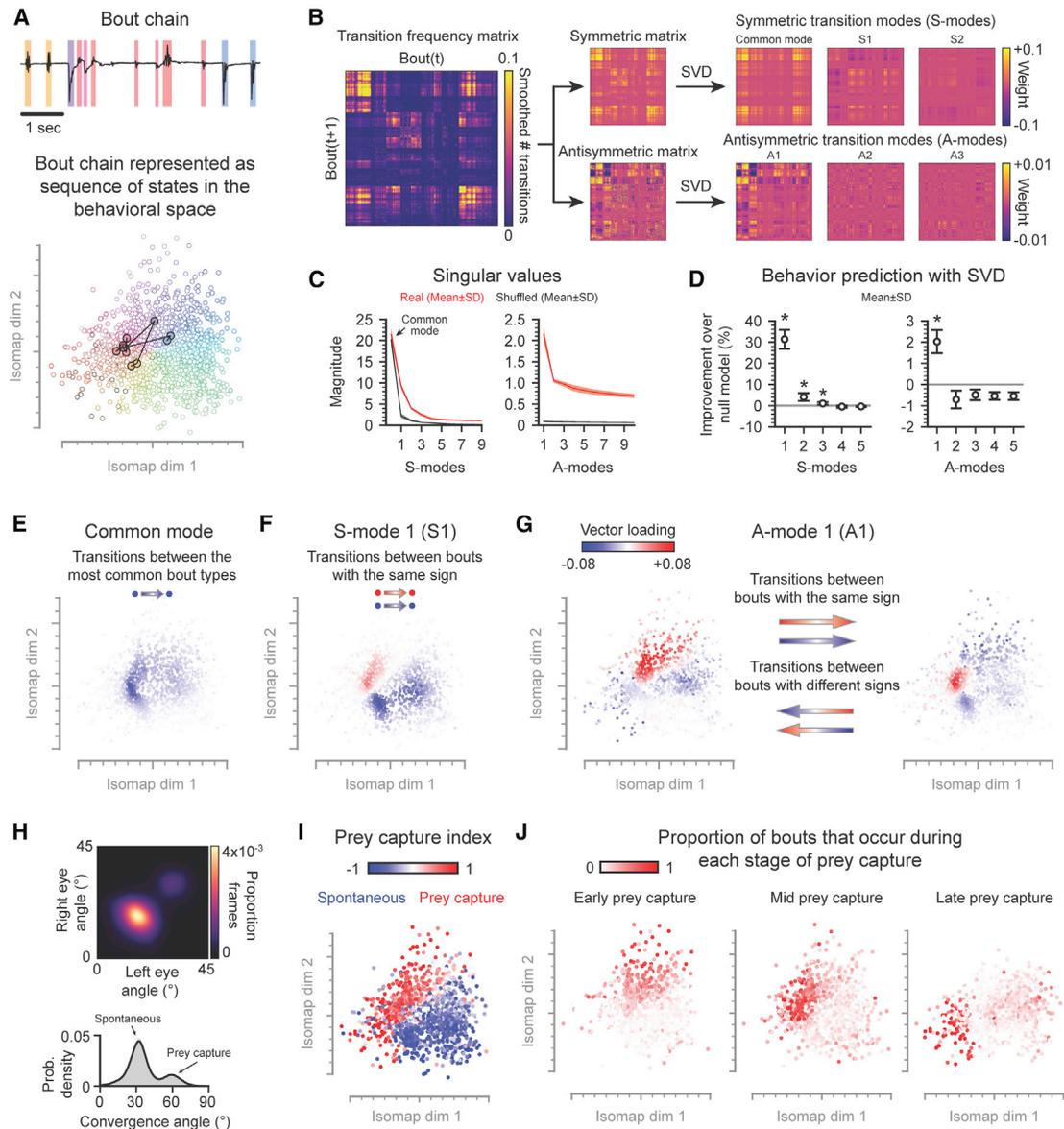
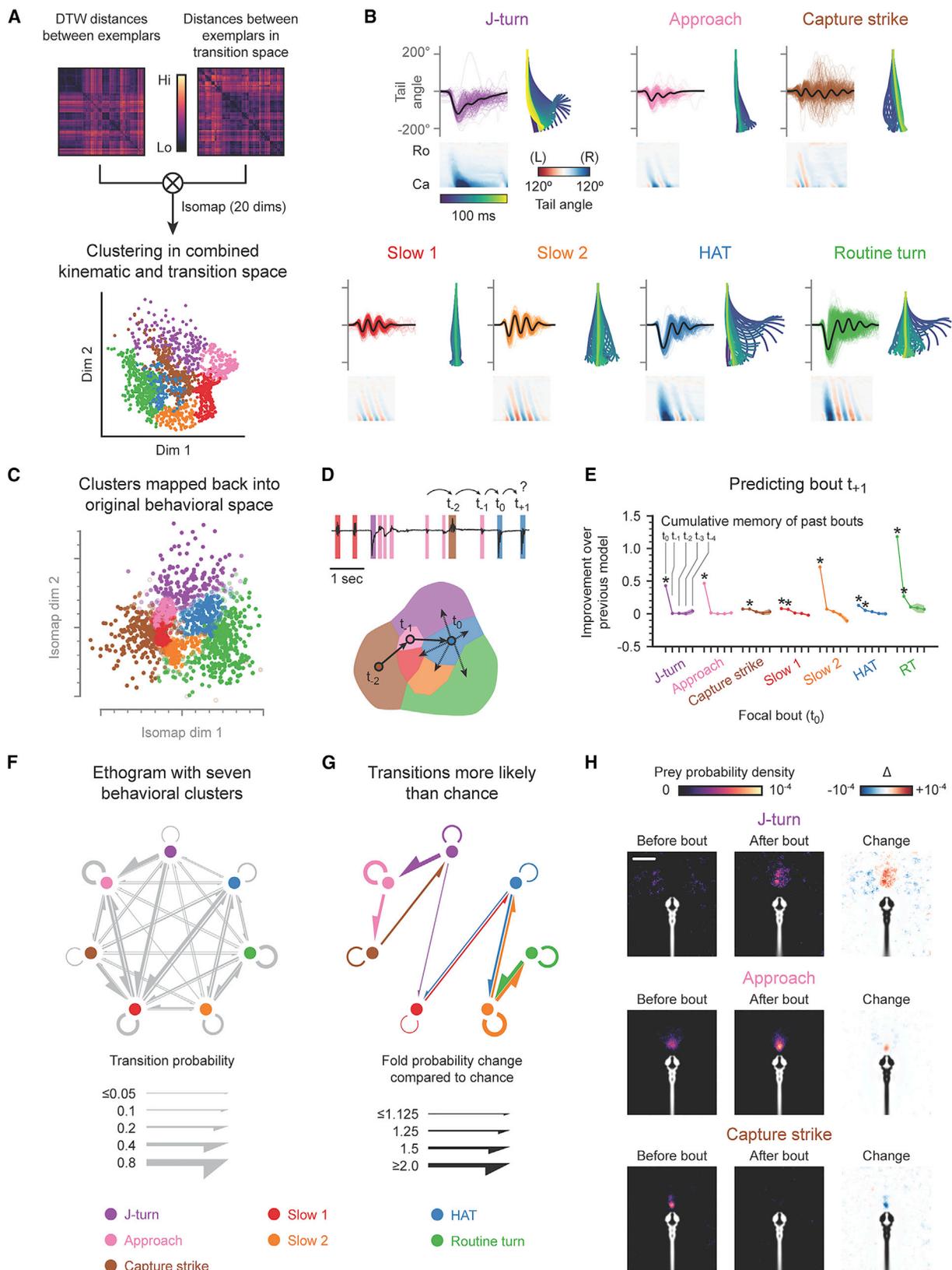


Figure 2. Singular-Value Decomposition Reveals Behavioral Dynamics

- (A) Top: tail tip angle trace from Figure 1C. Bouts color-coded according to position in the behavioral space (bottom).
 (B) Decomposition of the transition frequency matrix (smoothed, obtained from all bouts) into transition modes by singular-value decomposition (SVD).
 (C) Singular values of the symmetric (left) and antisymmetric (right) components of the transition frequency matrix. Mean \pm SD over 10,000 partitions of the data (shaded area). Observed transitions (red), shuffled transitions (gray).
 (D) Improvement of behavioral prediction each transition mode offers over common mode. Mean \pm SD over 10,000 partitions. Left: S-modes. Right: A-modes. * $p < 0.001$, permutation test.
 (E–G) Contribution of each micro-cluster to selected transition modes.
 (E) Common mode. Color intensity reflects how often each position in space was visited.
 (F) S1, represented by a single vector. Transitions occur between bouts with the same sign (color).
 (G) A1, represented by a pair of vectors. Bouts with a given sign on the left transition to bouts with the same sign on the right. Bouts with a given sign on the right transition to bouts with the opposing sign on the left.
 (H) Top: 2D histogram of eye angles across all frames. Bottom: bimodal distribution of eye convergence angles across all frames. Local minimum in the eye convergence distribution provides a threshold for defining prey capture behavior.
 (I) Prey capture index of each micro-cluster. Index defined as (# prey capture bouts – # spontaneous bouts)/(# total bouts).
 (J) Proportion of bouts in each micro-cluster that occur during early (eyes converge during bout), middle (eyes are converged), or late (eyes de-converge during bout) prey capture.

See also Figures S2 and S3.



(legend on next page)

(15%–39%). To investigate which of these transitions were significant, we compared the observed probabilities to randomly shuffled bout sequences (Figure 3G). This corroborated the results of our SVD analysis, revealing unidirectional chaining through J-turns, approaches, and capture strikes that were >50% more likely than chance, as well as enriched transitions between spontaneous clusters (17%–127% increase).

We next calculated the probability density of a paramecium being within a certain area of the visual field before and after each type of bout (Figures 3H and S4D; STAR Methods). We found that J-turns were initiated when prey lie in the lateral visual field (~45° from midline) and served to center prey in the anterior visual field (~1° from midline) (Figure 3H, top). This new stimulus correlated with the onset of approach swims, which moved prey toward a stereotyped position in the near-anterior visual field (Figure 3H, middle). “Slow 1” swims occurred if prey were further away (Figure S4D). Prey in the near-anterior visual field (~0.4 mm from the fish) were associated with capture strikes (Figure 3H, bottom). Thus, the successive transformation of the visual scene as a result of the fish’s own motion could account for the stereotyped sequence through bouts we observed during prey capture. We do not find such stereotyped stimuli associated with spontaneous swims (Figure S4D), so bidirectional transitions between different types of forward swim and turn during this behavior are likely mediated by internal rather than external factors [43, 44].

Genetic Disruptions of Vision Do Not Change Bout Kinematics, but Alter Behavioral Dynamics

To explore how sensory cues orchestrate prey capture sequences, we tested fish larvae with perturbations of their visual system. *Lakritz* mutants (*lak^{th241}*) [45, 46] are blind due to a null mutation in *atonal homolog 7 (atoh7)*, a transcription factor necessary for RGC development (Figure 4A, middle). SVD revealed that key transition modes associated with prey capture were absent in *lak* mutants (absolute cosine similarity: 0.59, common; 0.13, S1; 0.06, A1) compared to wild-type sibling controls (similarity: 0.96, common; 0.88, S1; 0.72, A1) (Figures 4B and 4C, middle). We also performed pharmacogenetic ablation of RGCs after fish had already had prey capture experience [47] (Figure S5A). Ablated fish lacked normal prey capture

dynamics, which were present in sibling controls (similarity: 0.58, common; 0.13, S1; 0.12, A1) (Figures S5B–S5D). These results demonstrate that the stereotyped behavioral sequences that occur during prey capture are innate and, under normal light conditions, depend on vision.

We next investigated the effect of a more subtle mutant phenotype on the structure of prey capture behavior. *Blumenkohl* mutants (*blu^{tc257}*) [46] carry a mutation in *vglut2a*, a gene encoding one of the vesicular glutamate transporters expressed by RGCs. These mutants grow larger RGC axonal arbors in the tectum and show decreased visual acuity (Figure 4A, bottom) [48]. Mutants performed more spontaneous bouts than heterozygous sibling controls (similarity between common modes: 0.64) (Figure 4B, bottom). Controls exhibited normal prey capture dynamics (similarity compared to wild-types: 0.95, common; 0.93, S1; 0.87, A1), while mutants lacked the prey capture-associated transition modes (similarity compared to control: 0.64, common; 0.08, S1; 0.10, A1) (Figures 4C and S6A–S6C). Studying stimuli associated with different bout types in these animals revealed that mutants performed J-turns when prey were closer, consistent with reduced visual acuity and accounting for the observed reduction in their hunting activity (Figures 4D, S6D, and S6E).

Virtual Reality Experiment Reveals that Hunting Requires a Sustained Prey Stimulus

Fish larvae may make moment-to-moment decisions during hunting. Alternatively, a single releasing stimulus may be sufficient to drive the behavior to completion. We developed a free-swimming virtual prey capture assay that allowed us to control the visual cues presented to an animal (Figure 4E; STAR Methods). Projecting small white dots moving with paramecium-like kinetics on the surface of the water reliably triggered prey capture in 40% of animals tested (17/41, >5% time spent with eyes converged when stimulus present). We presented animals with “persisting” trials during which larvae were allowed to hunt virtual prey objects *ad libitum*, and “vanishing” trials during which prey disappeared as soon as eye convergence was detected online (Figure 4F). Hunting sequences were significantly shorter during vanishing trials, and this was significant across all responsive fish (mean 1.38 versus 0.82 s) (Figure 4G). Hunting sequences also consisted of fewer bouts during vanishing trials,

Figure 3. Prey Capture Dynamics Arise through a Stimulus-Response Loop

(A–C) Clustering bouts.

(A) Generating a combined kinematic-transition space. Seven behavioral clusters are defined using hierarchical clustering in 20 dimensions of this space (first two shown).

(B) Example bouts from seven behavioral clusters. Subpanels: individual tail angle traces in color with the average in black (top left); tail kinematics of a representative bout (bottom); tail reconstruction of the representative bout (right). HAT, high-angle turn.

(C) Isomap from Figure 1I showing location of behavioral clusters.

(D and E) Predicting the next bout in a behavioral sequence.

(D) Top: bout sequence from Figure 1C color-coded according to cluster. Bottom: schematic partitioning of the behavioral space into bout types, showing part of the above bout sequence.

(E) Incremental improvement over previous model prediction when successively adding memory of previous bouts in a sequence. Mean improvement \pm SEM (* $p < 0.01$, Student’s *t* test, Bonferroni correction).

(F) Quantitative ethogram of zebrafish swimming behavior during prey capture. Colored circles represent behavioral clusters; gray arrows indicate probability of transitioning between clusters.

(G) Transition probabilities between clusters significantly higher than chance (* $p < 0.05$, permutation test, Holm-Bonferroni correction). Arrows show fold change in probability compared to shuffled data.

(H) Transformation of prey probability density in the visual scene during J-turns, approaches, and capture strikes in fish-centered coordinates. Images are thresholded using the 85th percentile. Average fish outline in white (left, center) or black (right). Scale bar, 1 mm.

See also Figure S4 and Video S3.

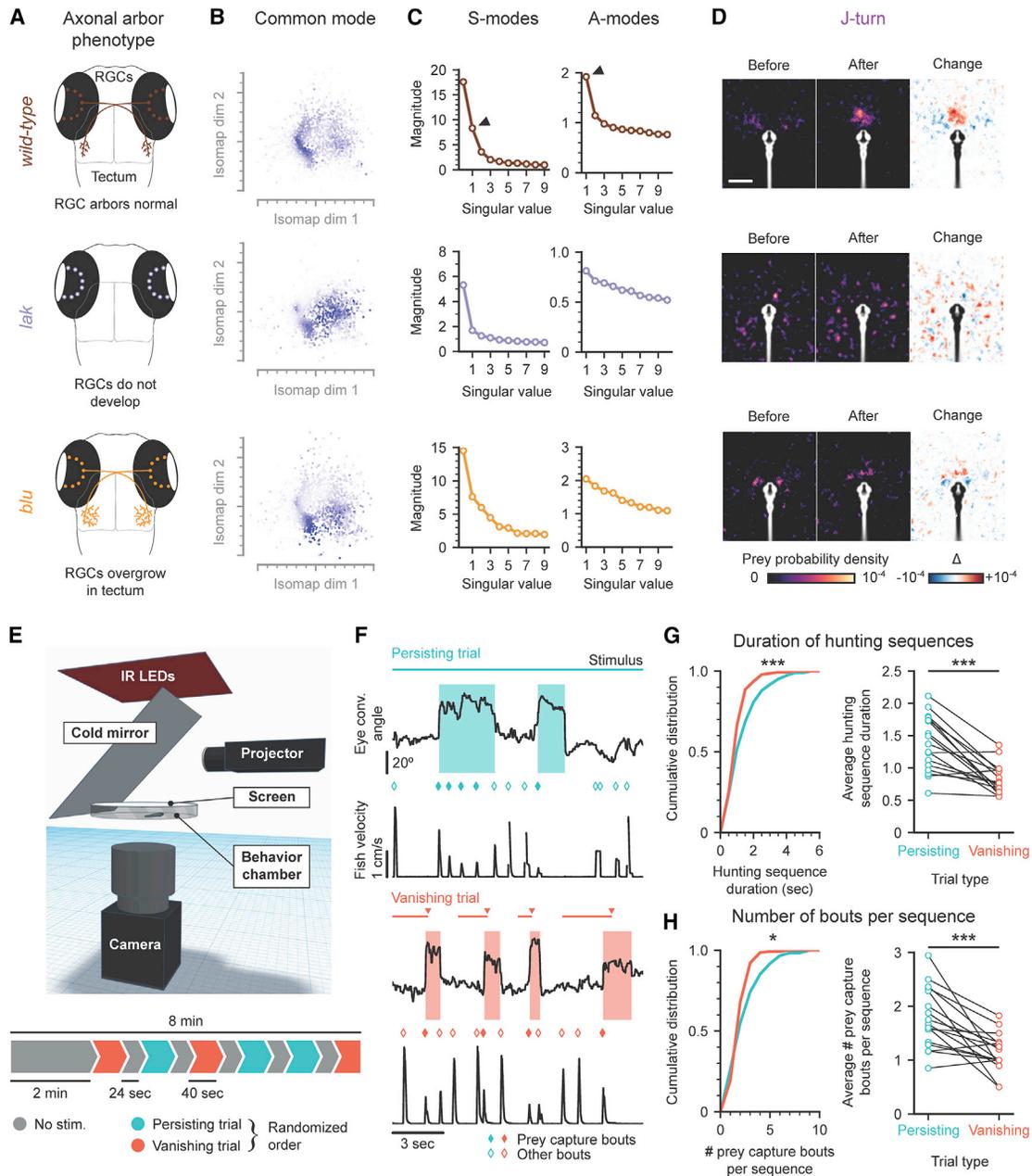


Figure 4. Chronic and Acute Disruption in Virtual Reality of Visual Cues Impairs Prey Capture

(A–D) Behavioral dynamics in mutants.

(A) Schematic anatomical phenotype of *lakritz* (*lak*) and *blumenkohl* (*blu*) mutants.

(B) Common mode obtained from the SVD of transition matrices after mapping bouts from wild-type sibling control (top), *lak* (middle), and *blu* (bottom) animals into the behavioral space from Figure 1.

(C) Singular values associated with S- and A-modes. Black arrowheads: prey capture-associated modes present in controls but absent or disrupted in mutants.

(D) Stimulus maps associated with J-turns.

(E–H) Acute disruption of visual cues during prey capture.

(E) Setup and experimental design for a virtual prey capture assay. Animals are presented with six 40 s stimulus trials, interspersed with intervals with no stimuli.

Persisting trials: virtual prey are always present. Vanishing trials: virtual prey disappear as soon as eye convergence is detected.

(F) Example period from persisting (top) and vanishing (bottom) trials. Horizontal bars show when stimulus was present. Arrowheads indicate when eye convergence was detected online and the stimulus removed. Diamonds: automatic bout detection. Shaded regions: post hoc determination of eye convergence.

(G) Left: cumulative distribution of hunt durations (** $p < 0.001$, Kolmogorov-Smirnov test). Right: average prey capture sequence duration during persisting and vanishing trials (** $p < 0.001$, Wilcoxon signed-rank test).

(H) Left: cumulative distribution of bout chain lengths (* $p < 0.05$, Kolmogorov-Smirnov test). Right: average number of bouts per hunting sequence (** $p < 0.001$, Wilcoxon signed-rank test).

See also Figures S5 and S6.

consistent across animals (median 1.7 versus 1.0 bouts per sequence) (Figure 4H).

Together, these results demonstrate that animals decide on the next bout in a sequence based on the available visual cues at a given moment. We conclude that prey capture is an example of a stimulus-response loop that serves to bring the prey to the near-anterior visual field of zebrafish, terminating with a capture strike.

Distance to the Prey Determines the Choice of Strike Type

We next analyzed the maneuvers that precede the capture of prey. Previously, strikes have been classified as ram or suction [16, 23], low or high velocity [40], and long or short duration [14]. We similarly found a bimodal distribution in capture strike duration, with one peak around 100 ms and a second peak around 200 ms (Figure 5A). Across all strike durations, however, we noticed that fish consumed prey after a stereotyped time (~50 ms), and that long-duration strikes resulted from a second, variable bout occurring immediately after the capture event. Re-embedding the first 50 ms of capture strike dynamics to produce a behavioral sub-space using our PCA-DTW-isomap pipeline revealed two clusters in the capture strike sub-space displaying markedly different postural dynamics (Figures 5B–5D; STAR Methods). We termed these maneuvers the attack swim and the S-strike (Figure 5E; Video S4). Almost all S-strikes were followed by a post-capture bout (“long duration”) whereas attack swims contributed to both long- and short-duration strikes (Figure 5A). This variation in strike dynamics suggests that the behavior does not represent a single stereotyped movement, but rather two possible capture strategies employed in different contexts.

To test whether different strikes might be selected in response to different stimuli, we investigated the prey probability density distribution in the visual field of animals prior to the two bout types. Both were triggered when prey was in the near-anterior visual field (Figure 5F); however, we noticed strong and characteristic fin abduction that only occurred prior to the onset of S-strikes [31]. This suggested that the animals were making preparatory movements before S-strikes, so we studied the evolution of the prey stimulus over time for hunting sequences ending in the two strike types (Video S5). We found prey position in the anterior visual field started to diverge approximately 250 ms prior to swim onset (Figures 5G and 5H). S-strikes occurred with a higher probability when prey was centered in the visual field 0.6 mm away. For attack swims, prey were less likely to be centered, and the behavior typically occurred once the prey was within 0.4 mm of the fish. These results indicate that larvae are sensitive to the distance of a prey item in the center of their visual field. Most commonly, larvae will perform an attack swim once prey reach a “strike zone.” If, however, the prey becomes centered in the visual field at a greater distance, larvae may resort to using their fins to keep the prey centered until it reaches the strike zone and then release an S-strike.

Larvae Perform Stereotyped Jaw Movements to Capture Prey in the Dorsal Visual Field

Animals exploit the three-dimensional water column during naturalistic behavior [49], and we wondered whether this was

also true during prey capture. Recording hunting behavior from the side revealed that larvae made full use of the vertical dimension and performed prominent cranial and jaw movements during this behavior (Video S6). To better quantify these movements, we modified our recording setup so that we could simultaneously track tail, jaw, and cranial movements with two views in a single camera (Figures 6A–6C; Video S7; STAR Methods). We found that the majority of jaw movements performed by larvae were initiated immediately after a swim bout (Figure 6D). We then applied our PCA-DTW-isomap pipeline to embed jaw movements in a new behavioral space, which revealed two distinct clusters (Figures 6E and 6F). The larger cluster corresponds to a relatively slow, low-amplitude depression of the jaw with little or no movement of the cranium and could relate to early buccal pumping in the larva or chemosensation (Figure 6G, left). The rarer jaw movement was highly stereotyped, comprising a rapid, large-amplitude depression of the jaw concurrent with cranial elevation (Figure 6G, right; Video S6). This movement was exclusively associated with capture events and was preceded by an attack swim or S-strike, or occurred in isolation as a “suction” capture [16, 30] (Figure 6H). Thus, different capture strategies in zebrafish larvae emerge by combining different stereotyped tail kinematics with stereotyped jaw kinematics.

Hunting episodes were associated with both changes in pitch and moving up and down in the water column (Figure 6C). Larvae had a mean preferred orientation of $\sim 7^\circ$ and rotated to $\sim 12^\circ$ prior to the onset of a capture (Figure 6I). Capture events occurred when prey were in the dorsal visual field (Figure 6J), which would correspond to a ventral position on the retina.

Larvae Use Binocular Vision to Position Prey in the Strike Zone

It has been proposed that larvae converge their eyes during prey capture in order to create an area of binocular overlap [16, 26]. To test the role of binocular vision in prey capture, we removed the lens from either one or both eyes (lensectomy or “delensing”) of 7-dpf larvae and tested their behavior the following day (Figure 7A; STAR Methods; sham = 16, unilateral = 18, bilateral = 16). Unilaterally delensed animals were able to converge their eyes to the same extent as controls, indicating that the procedure did not prevent eye movements; however, bilaterally delensed animals did not perform eye convergence, suggesting that lensectomy effectively blinded animals to prey items (Figure 7B). Unilaterally delensed animals spent less time engaged in prey capture than controls (median proportion of time: 0.25, sham, versus 0.15, unilateral) (Figure 7C) and detected prey only about half as often (median detection rate: 10 times per minute, sham, versus 6 times per minute, unilateral) (Figure 7D). Unilaterally delensed animals only initiated J-turns toward prey located on the same side as their intact eye (Figure 7E). Thus, binocular vision is not required to initiate hunting behavior.

We wanted to know how subsequent prey capture dynamics were affected in one-lensed animals. We used SVD to identify transition modes in control and unilaterally delensed groups and found a subtle but significant difference in A1 encoding prey capture sequence dynamics (similarity: 0.73)

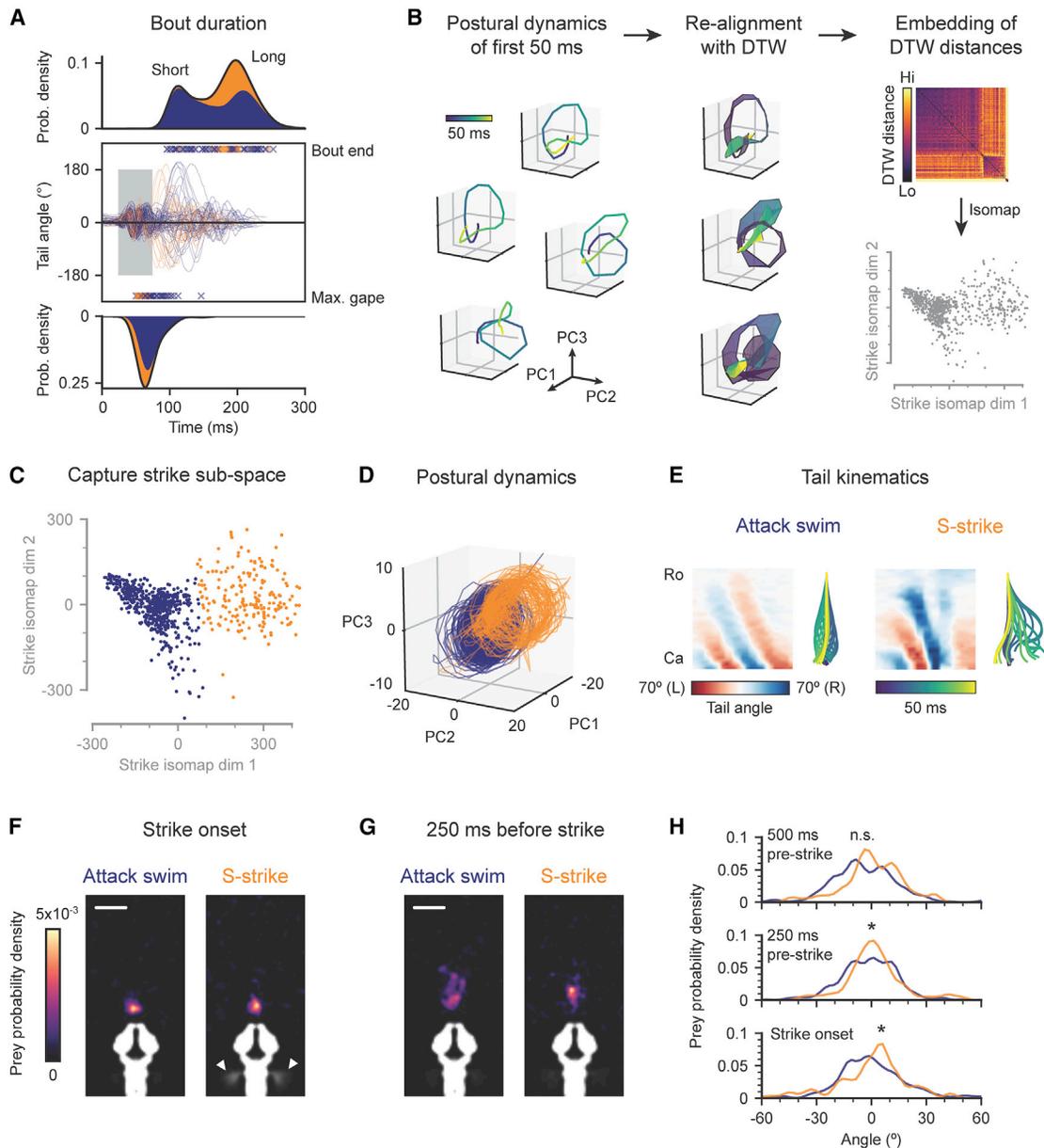


Figure 5. Prey Pursuit Concludes with Either of Two Distinct Capture Maneuvers

(A) Capture strikes consist of a capture phase and a variable post-capture phase. Top: stacked KDE of capture strike durations. Attack swims (blue) and S-strikes (orange). Middle: tail tip angle over time for capture strikes. Gray window: initial 50 ms capture phase. Bottom: KDE over jaw opening times.

(B) Pipeline for generating capture strike sub-space. Isomap embedding of DTW distances between initial capture phases of all strikes.

(C) K-means clustering (two clusters) in the capture strike sub-space. Blue, attack swims; orange, S-strikes.

(D) Trajectories through PCs for attack swims and S-strikes.

(E) Representative examples of an attack swim and an S-strike. Tail kinematics (left) and reconstructed bout (right).

(F) Prey probability density prior to attack swims (left) and S-strikes (right). White arrowheads: fin abduction.

(G) Prey probability density 250 ms prior to the onset of attack swims (left) and S-strikes (right).

Scale bar in (F) and (G), 500 μ m.

(H) Prey probability density as a function of azimuthal angle at different time points in hunting sequences resulting in attack swims (blue) or S-strikes (orange) (*p < 0.05, permutation test using energy statistics to compare distributions; n.s., not significant).

See also Videos S4 and S5.

(Figure 7F). Specifically, transitions to capture strikes from approach swims were under-represented in the unilaterally delensed group (dotted outline). Unilaterally delensed animals

ended their hunting sequences with a capture strike only about half as often as controls (Figure 7G). We wanted to test whether this decrease in strike rate affected both attack

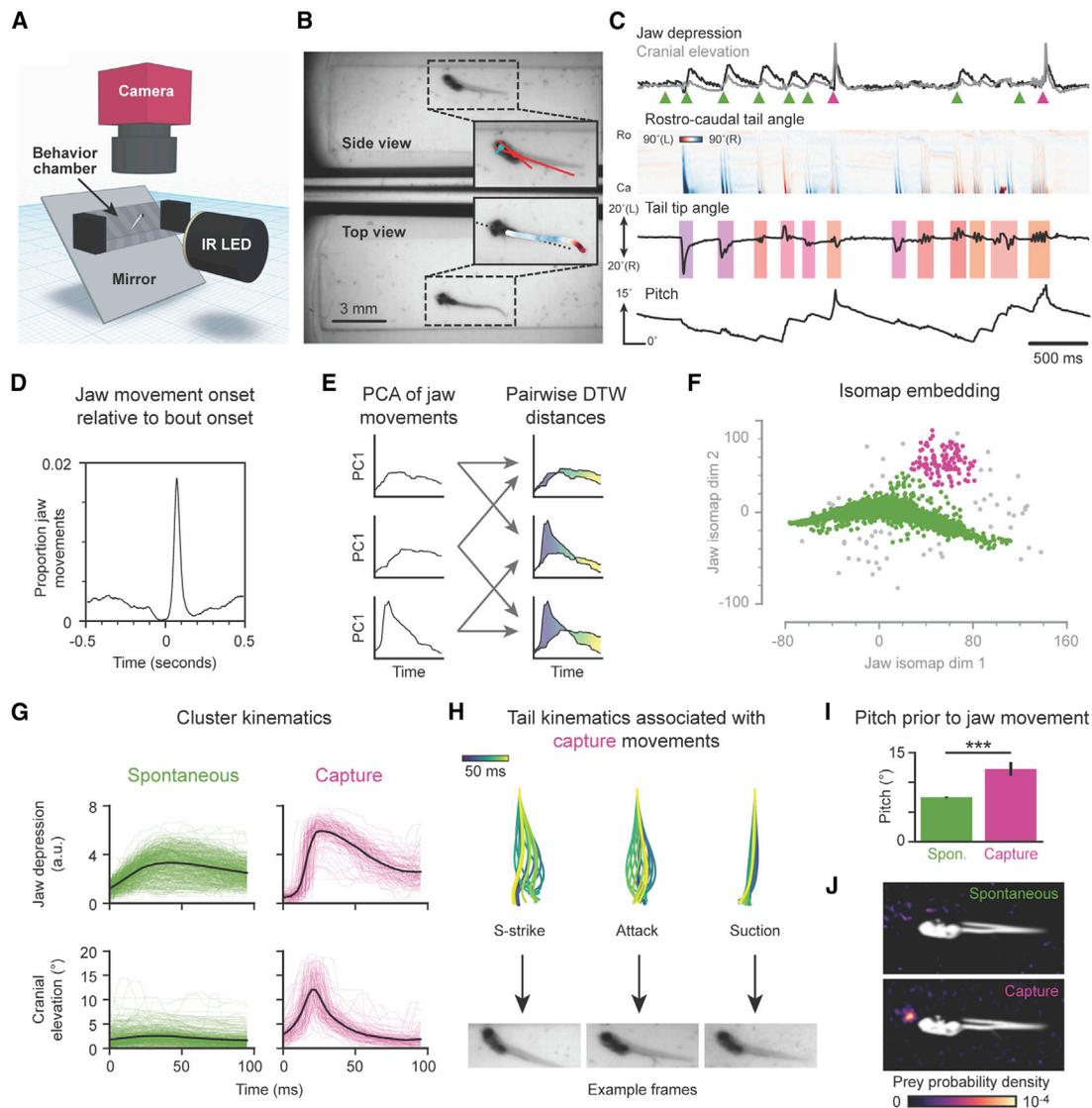


Figure 6. Larvae Capture Prey from Below with Stereotyped Jaw Kinematics

(A) Schematic of the setup used to record behavior simultaneously from above and from the side.

(B) Example frame; insets are overlaid with tail and jaw tracking.

(C) Jaw and tail kinematics from 4 s of behavioral recording. Top: depression of the jaw (black) and elevation of the cranium (gray). Arrowheads: jaw movement onsets. Spontaneous movements (green); capture strikes (magenta). Middle: tail tracking. Bouts color-coded according to nearest exemplar in behavioral space. Bottom: pitch of the fish.

(D) Cross-correlation between bout onsets and jaw movement onsets.

(E) Generating a jaw movement behavioral space. Jaw movements projected onto first PC followed by pairwise alignment and distance calculation with DTW. (F) Two-dimensional isomap embedding of DTW distances between jaw movements. Clustered using HDBSCAN into spontaneous jaw movements (green), capture jaw movements (magenta), and noise (gray).

(G) Jaw depression (top) and cranial elevation (bottom) for spontaneous (left) and capture (right) jaw movements. Colored traces: individual movements. Black lines: average.

(H) Example tail movements (top) preceding the shown capture jaw movement (bottom).

(I) Pitch of fish prior to bouts containing spontaneous and capture jaw movements (***two-tailed $p < 0.001$, unpaired Student's t test). Mean \pm SEM.

(J) Prey probability density prior to bouts containing spontaneous (top) and capture (bottom) jaw movements. White: average outline of fish. Anterior is left. See also Videos S6 and S7.

swims and S-strikes, so we again generated a capture strike sub-space (Figure 7H; cf. Figure 5B). We found that one-lensed animals performed only about a third as many attack swims (median number per animal: 14, sham; 4.5, unilateral)

and almost no S-strikes (median number per animal: 3.5, sham; 0, unilateral) (Figure 7I).

The reduction in capture strikes in unilaterally blinded animals suggested that larvae may be using binocular cues to

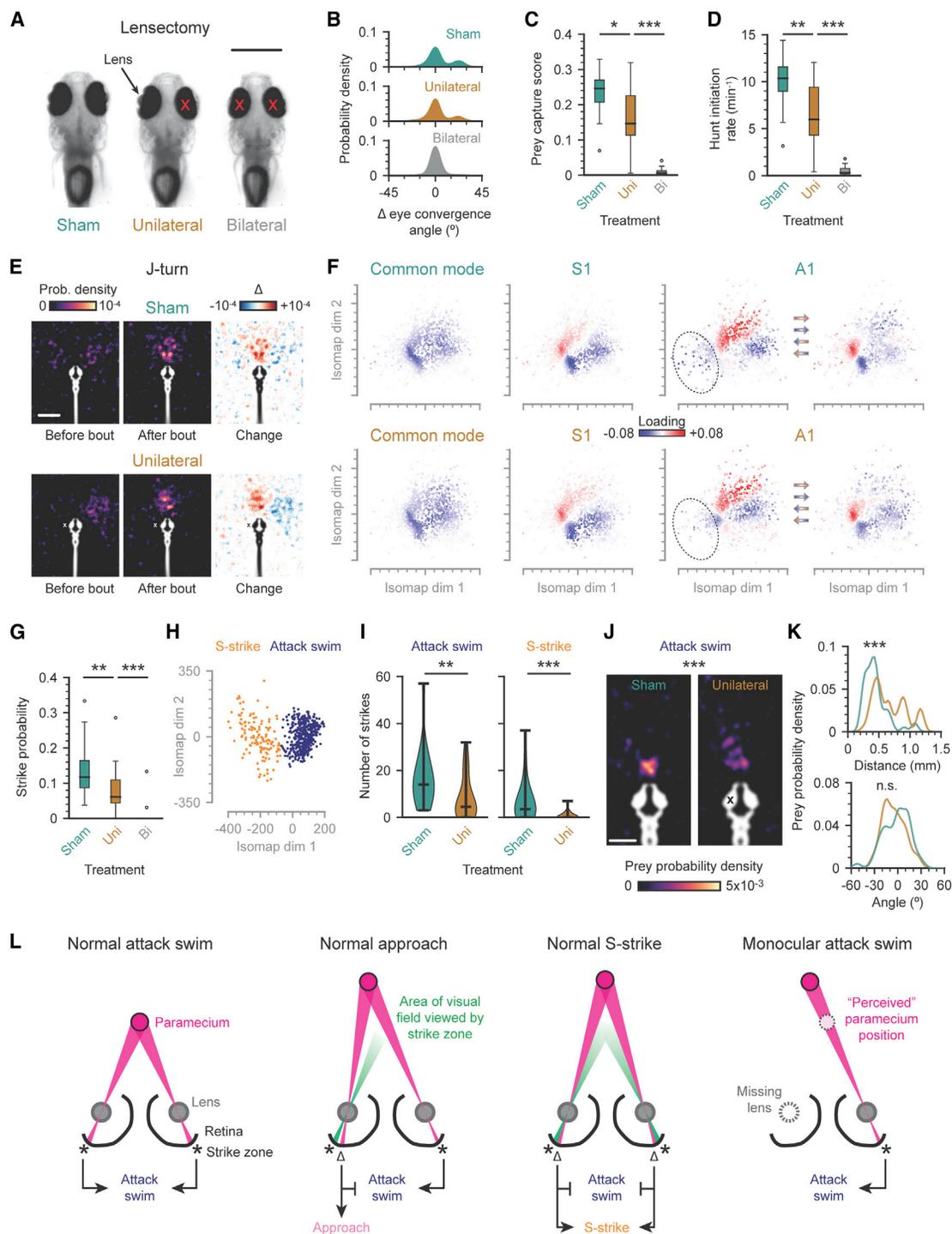


Figure 7. Larvae Use Binocular Cues to Position Prey on the Strike Zone

(A) Lensectomy. Lens dissected from one (unilateral) or both (bilateral) eyes (red crosses). Scale bar, 500 μm .

(B) Eye convergence distributions in sham, unilaterally, and bilaterally delensed larvae. Aligned to the resting eye convergence angle (0°).

(C) Proportion of time spent engaged in prey capture (prey capture score) for each treatment group.

(D) Number of times animals converged their eyes each minute (hunt initiation rate) for each treatment group.

(E) Stimulus maps for sham and unilaterally delensed animals during J-turns. Animals that had their right lens removed are mirrored (white cross). Scale bar, 1 mm.

(F) Transition modes of sham (top) and unilaterally delensed (bottom) animals. Differences between common modes not significant ($p > 0.05$). Differences between S1 and A1: $p < 0.05$, permutation test on absolute cosine similarity. Dotted outline: approximate location of capture strikes in the space.

(G) Probability that hunting sequences terminate with a capture strike for animals in each treatment group.

(legend continued on next page)

judge distance to the prey. Studying the probability distribution in the visual field revealed that prey was less likely to fall in the strike zone of unilaterally delensed animals (Figure 7J; Video S8). These animals initiated attack swims when prey were further away (0.5 mm versus 0.4 mm) and skewed $\sim 10^{\circ}$ – 15° in the direction of their blinded eye (Figures 7J and 7K). Thus, it appears that larvae will typically perform the capture strike when prey fall on the temporal-ventral retina of both eyes (Figure 7L). If a fish brings the prey into the temporal-ventral retina of one eye, but is missing input from the other eye, it may trigger an attack swim prematurely. Furthermore, animals may use binocular cues to determine that prey is centered but too far from the strike zone, triggering an S-strike maneuver.

DISCUSSION

Our analysis reveals that zebrafish larvae capture prey through a stimulus-response loop relying on immediate binocular cues. We identify stereotyped bout sequences that emerge from low-order transition rules. Chronic “blurring” of the visual scene in genetic mutants impairs prey detection and removal of visual cues after initiation of behavior causes animals to abort the hunting sequence early. Bout sequences position the prey in the center of the proximal dorsal visual field, triggering a capture strike. Depending on the distance to the prey, animals will perform one of two possible strike maneuvers. Disrupting binocular vision impedes an animal’s ability to position the prey centrally in the visual field and causes premature release of a capture strike.

Different embedding approaches can highlight different structural aspects when visualizing behavior in lower dimensions. Previously, t-SNE has revealed local structure in the behavior of flies and fish [12, 14]. We found that isomap embedding performed well at separating distinct and rarer bout types such as J-turns and capture strikes (Figures 1I and 3C), but was less effective than t-SNE at separating different kinds of forward scoot (Figures 1I and S1C). Some of this local structure was, however, retained in an intermediate clustering step (Figures 1H and 1K). We also revealed distinct types of capture maneuvers and jaw movements using isomap embedding, and thus this approach may be complementary to t-SNE and UMAP when investigating the structure of behavior.

Previous studies of the temporal structure of behavior have employed Markov models [5, 19, 21]. We identified common transitions shared across different regions of our behavioral space,

and this aided the delineation of cluster boundaries subsequently used for Markov chain analysis (Figures 3A–3C). We did not find strong evidence of higher-order dynamics in our analysis of sequences built from these clusters (Figures 3D and 3E), suggesting that transition modes provide an accurate account of zebrafish larva prey capture behavior (Figures 3F and 3G). Recent work in this area has also found the preceding bout strongly influences the next, although longer-term dynamics, including satiety state, can have some subtle effects [50]. Future work may seek to generalize the analysis of transition modes to also account for behavioral changes over longer timescales.

We demonstrated that short-term visual cues on the order of 100s of milliseconds are necessary to maintain hunting behavior (Figures 4E–4H). Our new closed-loop virtual reality assay extends the possibilities for studying the visual cues necessary for the formation and maintenance of hunting behavior under more naturalistic free-swimming conditions, as previous analyses have been restricted to embedded preparations [26, 28, 29]. Short integration windows for deciding the next behavior have been observed in thermal navigation of larvae [51] and social affiliation of juvenile zebrafish [52]. Thus, stimulus-response loops driving behavioral chaining might not be specific to prey capture, but provide a more general mechanism underlying goal-directed behavior in zebrafish.

It has been proposed that a specialized UV-sensitive zone in the ventral retina could facilitate targeting prey from below [53–55]. We demonstrate that larvae do indeed orient themselves beneath the prey over the course of a hunting sequence (Figures 6I and 6J). Dedicated neural circuits in the retina and pretectum could mediate the initial formation of the stimulus-response loop [28, 55–58], subsequently maintained by the optic tectum [32, 59, 60] and the nucleus isthmi [61]. Future work may seek to identify and characterize the neural circuitry underlying prey capture strikes (Figure 5). When the eyes are converged, prey in the temporal-ventral retina is represented symmetrically in the anterior regions of both tecta. We posit the anterior tectum contains specialized circuitry for implementing the appropriate capture maneuver. The S-strike and attack swim may be driven by separate command-like neuronal populations [59], or alternatively by different activity patterns within a common population. These swims are performed in conjunction with a stereotyped jaw movement to capture prey (Figures 6F–6H; Videos S6 and S7). Producing invariant actions in response to stereotyped “releasing” stimuli has long been considered an efficient way to ensure reproducible outcomes in innate behaviors [8]. We speculate that by linking different releasing stimuli to stereotyped swims sharing a common jaw

(H) Capture strike sub-space, generated as outlined in Figure 5B. Capture strikes clustered into two types using K-means clustering. Attack swims (blue). S-strikes (orange).

(I) Number of strikes performed by animals split by strike type (attack swim versus S-strike) and treatment group (sham versus unilateral).

(J) Prey probability density at the onset of attack swims in sham (left) and unilaterally delensed (right) animals. Cross indicates eye with lens removed (right animals mirrored). Scale bar, 500 μ m.

(K) Prey probability densities prior to attack swims at various distances (top) and azimuthal angles (bottom) in the visual field for sham and unilaterally delensed animals. Positive angles signify azimuthal position toward the sighted eye.

(L) Model for the binocular control of capture strikes. Far left: a paramecium falling on the strike zone triggers an attack swim. Center left: a paramecium falling outside the strike zone triggers an approach. Center right: a paramecium centered in the visual field but outside the strike zone triggers an S-strike. Far right: a paramecium falling outside the strike zone in a larva with its left lens removed (gray dotted outline) triggers an inappropriate attack swim. Δ signifies offset between current prey position (magenta) and strike zone (asterisk/green cone).

(C), (D), (G), and (I): * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Mann-Whitney U test. (J) and (K): *** $p < 0.001$; n.s., not significant ($p = 0.15$), permutation test on energy distances between distributions. See also Video S8.

movement, the nervous system of the zebrafish larva has evolved an efficient means to produce reliable, yet flexible, behavior.

Since eye convergence was first identified as a hallmark of prey capture in zebrafish larvae, there has been speculation about a potential role of binocular vision in this behavior [16, 26]. We demonstrate that, in addition to reducing the rate of prey detection, loss of binocular vision reduces the probability to strike once a prey capture sequence has begun (Figures 7A–7G). One possible explanation is that the stimulus-response loop that brings prey to the near-anterior visual field is monocularly driven, with visual cues from each eye independently and stochastically releasing the next bout in the sequence. In this case, loss of vision in one eye would reduce the probability of a bout being generated, giving prey time to move out of the field of view of the intact eye. Alternatively, animals may integrate cues from both eyes to bring the prey optimally into the strike zone of both retinas. While we show that binocular cues are not required to release the capture strike behavior (Figures 7G–7I), they do appear to serve a role in the appropriate positioning of prey in the strike zone (Figures 7J–7L). These results suggest that binocular vision could be used as a mechanism to determine distance to prey in zebrafish larvae.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **LEAD CONTACT AND MATERIALS AVAILABILITY**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Wild-type zebrafish larvae
 - Mutant zebrafish larvae
 - Genetic ablation of retinal ganglion cells (RGCs)
- **METHOD DETAILS**
 - Free-swimming behavioral assay with real prey
 - Tail and eye tracking
 - Generating a behavioral space
 - Mapping kinematic features and eye convergence into the behavioral space
 - Mapping bouts from other experiments into the behavioral space
 - Singular-value decomposition of behavioral transitions
 - Investigating transition modes with toy data
 - Eye convergence analysis
 - Clustering in a combined kinematic-transition space
 - Modeling transitions between clusters
 - Paramecium tracking
 - Generating stimulus maps
 - Free-swimming virtual prey capture assay
 - Capture strike analysis
 - Free-swimming behavioral assay in three dimensions
 - Jaw tracking
 - Generating a behavioral space of jaw movements
 - Paramecium tracking from the side
 - Lensectomy
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Identification of significant transition modes
 - Comparison of transition modes across experimental datasets

- Simplex projection to test higher order behavioral dynamics
- Identification of significant transitions in an ethogram
- Statistical analysis of virtual prey experiments
- Quantification of prey distance during capture strikes
- Quantification of lensectomy experiments
- **DATA AND CODE AVAILABILITY**

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cub.2019.11.022>.

ACKNOWLEDGMENTS

The authors would like to thank all members of the Baier lab and colleagues at the Max Planck Institute of Neurobiology for discussions and helpful comments. In particular, we thank T. Helmbrecht, J. Larsch, S. Reinert, and D. Guggiana Nilo for constructive feedback on the manuscript. We thank M. Dal Maschio and N. Dolensek for especially helpful discussions. We thank V. Štih and L. Petrucco for help setting up and designing the virtual prey capture assay. We also thank K. Slanchev for providing *lakritz* and *blumenkohl* mutants. Schematics in Figures 4A and 5SA were adapted with permission from unpublished work by Y. Kölsch. Funding was provided by the Max Planck Society.

AUTHOR CONTRIBUTIONS

Conceptualization, D.S.M., J.L.S., and H.B.; Methodology, D.S.M., J.C.D., A.M.F., and H.B.; Investigation, D.S.M.; Software, D.S.M. and J.C.D.; Formal Analysis, D.S.M.; Visualization, D.S.M.; Writing – Original Draft, D.S.M.; Writing – Review & Editing, J.C.D., A.M.F., J.L.S., and H.B.; Supervision, J.L.S. and H.B.; Funding Acquisition, H.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 19, 2019

Revised: September 30, 2019

Accepted: November 6, 2019

Published: December 19, 2019

REFERENCES

1. Krakauer, J.W., Ghazanfar, A.A., Gomez-Marin, A., MacIver, M.A., and Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. *Neuron* 93, 480–490.
2. Cande, J., Namiki, S., Qiu, J., Korff, W., Card, G.M., Shaevitz, J.W., Stern, D.L., and Berman, G.J. (2018). Optogenetic dissection of descending behavioral control in *Drosophila*. *eLife* 7, e34275.
3. Kato, S., Kaplan, H.S., Schrödel, T., Skora, S., Lindsay, T.H., Yemini, E., Lockery, S., and Zimmer, M. (2015). Global brain dynamics embed the motor command sequence of *Caenorhabditis elegans*. *Cell* 163, 656–669.
4. Vogelstein, J.T., Park, Y., Ohyama, T., Kerr, R.A., Truman, J.W., Priebe, C.E., and Zlatic, M. (2014). Discovery of brainwide neural-behavioral maps via multiscale unsupervised structure learning. *Science* 344, 386–392.
5. Wiltschko, A.B., Johnson, M.J., Iurilli, G., Peterson, R.E., Katon, J.M., Pashkovski, S.L., Abraira, V.E., Adams, R.P., and Datta, S.R. (2015). Mapping sub-second structure in mouse behavior. *Neuron* 88, 1121–1135.
6. Anderson, D.J., and Perona, P. (2014). Toward a science of computational ethology. *Neuron* 84, 18–31.

7. Egnor, S.E.R., and Branson, K. (2016). Computational analysis of behavior. *Annu. Rev. Neurosci.* **39**, 217–236.
8. Tinbergen, N. (1951). *The Study of Instinct* (Oxford University Press).
9. Flash, T., and Hochner, B. (2005). Motor primitives in vertebrates and invertebrates. *Curr. Opin. Neurobiol.* **15**, 660–666.
10. Bizzi, E., and Cheung, V.C. (2013). The neural origin of muscle synergies. *Front. Comput. Neurosci.* **7**, 51.
11. Del Vecchio, D., Murray, R.M., and Perona, P. (2003). Decomposition of human motion into dynamics-based primitives with application to drawing tasks. *Automatica* **39**, 2085–2098.
12. Berman, G.J., Choi, D.M., Bialek, W., and Shaevitz, J.W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. *J. R. Soc. Interface* **11**, 20140672.
13. Brown, A.E.X., Yemini, E.I., Grundy, L.J., Jucikas, T., and Schafer, W.R. (2013). A dictionary of behavioral motifs reveals clusters of genes affecting *Caenorhabditis elegans* locomotion. *Proc. Natl. Acad. Sci. USA* **110**, 791–796.
14. Marques, J.C., Lackner, S., Félix, R., and Orger, M.B. (2018). Structure of the zebrafish locomotor repertoire revealed with unsupervised behavioral clustering. *Curr. Biol.* **28**, 181–195.e5.
15. Katsov, A.Y., Freifeld, L., Horowitz, M., Kuehn, S., and Clandinin, T.R. (2017). Dynamic structure of locomotor behavior in walking fruit flies. *eLife* **6**, e26410.
16. Patterson, B.W., Abraham, A.O., MacIver, M.A., and McLean, D.L. (2013). Visually guided gradation of prey capture movements in larval zebrafish. *J. Exp. Biol.* **216**, 3071–3083.
17. Szigeti, B., Deogade, A., and Webb, B. (2015). Searching for motifs in the behaviour of larval *Drosophila melanogaster* and *Caenorhabditis elegans* reveals continuity between behavioural states. *J. R. Soc. Interface* **12**, 20150899.
18. Long, M.A., Jin, D.Z., and Fee, M.S. (2010). Support for a synaptic chain model of neuronal sequence generation. *Nature* **468**, 394–399.
19. Berman, G.J., Bialek, W., and Shaevitz, J.W. (2016). Predictability and hierarchy in *Drosophila* behavior. *Proc. Natl. Acad. Sci. USA* **113**, 11943–11948.
20. Seeds, A.M., Ravbar, P., Chung, P., Hampel, S., Jr., Midgley, F.M., Jr., Mensh, B.D., and Simpson, J.H. (2014). A suppression hierarchy among competing motor programs drives sequential grooming in *Drosophila*. *eLife* **3**, e02951.
21. Tao, L., Ozarkar, S., Beck, J.M., and Bhandawat, V. (2019). Statistical structure of locomotion and its modulation by odors. *eLife* **8**, e41235.
22. Ewert, J.-P. (1987). Neuroethology of releasing mechanisms: prey-catching in toads. *Behav. Brain Sci.* **10**, 337–368.
23. Boria, M.A., Palecek, B., Budick, S., and O'Malley, D.M. (2002). Prey capture by larval zebrafish: evidence for fine axial motor control. *Brain Behav. Evol.* **60**, 207–229.
24. Budick, S.A., and O'Malley, D.M. (2000). Locomotor repertoire of the larval zebrafish: swimming, turning and prey capture. *J. Exp. Biol.* **203**, 2565–2579.
25. McElligott, M.B., and O'malley, D.M. (2005). Prey tracking by larval zebrafish: axial kinematics and visual control. *Brain Behav. Evol.* **66**, 177–196.
26. Bianco, I.H., Kampff, A.R., and Engert, F. (2011). Prey capture behavior evoked by simple visual stimuli in larval zebrafish. *Front. Syst. Neurosci.* **5**, 101.
27. Muto, A., Ohkura, M., Abe, G., Nakai, J., and Kawakami, K. (2013). Real-time visualization of neuronal activity during perception. *Curr. Biol.* **23**, 307–311.
28. Semmelhack, J.L., Donovan, J.C., Thiele, T.R., Kuehn, E., Laurell, E., and Baier, H. (2014). A dedicated visual pathway for prey detection in larval zebrafish. *eLife* **3**, e04878.
29. Trivedi, C.A., and Bollmann, J.H. (2013). Visually driven chaining of elementary swim patterns into a goal-directed motor sequence: a virtual reality study of zebrafish prey capture. *Front. Neural Circuits* **7**, 86.
30. Hernández, L.P., Barresi, M.J.F., and Devoto, S.H. (2002). Functional morphology and developmental biology of zebrafish: reciprocal illumination from an unlikely couple. *Integr. Comp. Biol.* **42**, 222–231.
31. McClenahan, P., Troup, M., and Scott, E.K. (2012). Fin-tail coordination during escape and predatory behavior in larval zebrafish. *PLoS ONE* **7**, e32295.
32. Gahtan, E., Tanger, P., and Baier, H. (2005). Visual prey capture in larval zebrafish is controlled by identified reticulospinal neurons downstream of the tectum. *J. Neurosci.* **25**, 9294–9303.
33. Brown, A.E.X., and de Bivort, B. (2018). Ethology as a physical science. *Nat. Phys.* **14**, 653–657.
34. Stephens, G.J., Johnson-Kerner, B., Bialek, W., and Ryu, W.S. (2008). Dimensionality and dynamics in the behavior of *C. elegans*. *PLoS Comput. Biol.* **4**, e1000028.
35. Girdhar, K., Gruebele, M., and Chemla, Y.R. (2015). The behavioral space of zebrafish locomotion and its neural network analog. *PLoS ONE* **10**, e0128668.
36. Jouary, A., and Sumbre, G. (2016). Automatic classification of behavior in zebrafish larvae. *bioRxiv*. <https://doi.org/10.1101/052324>.
37. Sakoe, H., and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.* **26**, 43–49.
38. Frey, B.J., and Dueck, D. (2007). Clustering by passing messages between data points. *Science* **315**, 972–976.
39. Tenenbaum, J.B., de Silva, V., and Langford, J.C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* **290**, 2319–2323.
40. Westphal, R.E., and O'Malley, D.M. (2013). Fusion of locomotor maneuvers, and improving sensory capabilities, give rise to the flexible homing strikes of juvenile zebrafish. *Front. Neural Circuits* **7**, 108.
41. Mirat, O., Sternberg, J.R., Severi, K.E., and Wyart, C. (2013). ZebraZoom: an automated program for high-throughput behavioral analysis and categorization. *Front. Neural Circuits* **7**, 107.
42. Fritzsche, D., Virnik, E., Szyld, D.B., and Mehrmann, V. (2008). An SVD approach to identifying metastable states of Markov chains. *Electron. Trans. Numer. Anal.* **29**, 46–69.
43. Dunn, T.W., Mu, Y., Narayan, S., Randlett, O., Naumann, E.A., Yang, C.-T., Schier, A.F., Freeman, J., Engert, F., and Ahrens, M.B. (2016). Brain-wide mapping of neural activity controlling zebrafish exploratory locomotion. *eLife* **5**, e12741.
44. Wolf, S., Dubreuil, A.M., Bertoni, T., Böhm, U.L., Bormuth, V., Candelier, R., Karpenko, S., Hildebrand, D.G.C., Bianco, I.H., Monasson, R., and Debrégeas, G. (2017). Sensorimotor computation underlying phototaxis in zebrafish. *Nat. Commun.* **8**, 651.
45. Kay, J.N., Finger-Baier, K.C., Roeser, T., Staub, W., and Baier, H. (2001). Retinal ganglion cell genesis requires *lakritz*, a Zebrafish *atonal* Homolog. *Neuron* **30**, 725–736.
46. Neuhauss, S.C.F., Biehler, O., Seeliger, M.W., Das, T., Kohler, K., Harris, W.A., and Baier, H. (1999). Genetic disorders of vision revealed by a behavioral screen of 400 essential loci in zebrafish. *J. Neurosci.* **19**, 8603–8615.
47. Fernandes, A.M., Larsch, J., Donovan, J.C., Helmbrecht, T.O., Mearns, D., Kölsch, Y., Maschio, M.D., and Baier, H. (2019). Neuronal circuitry for stimulus selection in the visual system. *bioRxiv*. <https://doi.org/10.1101/598383>.
48. Smear, M.C., Tao, H.W., Staub, W., Orger, M.B., Gosse, N.J., Liu, Y., Takahashi, K., Poo, M.M., and Baier, H. (2007). Vesicular glutamate transport at a central synapse limits the acuity of visual perception in zebrafish. *Neuron* **53**, 65–77.
49. Horstick, E.J., Bayley, Y., Sinclair, J.L., and Burgess, H.A. (2017). Search strategy is regulated by somatostatin signaling and deep brain photoreceptors in zebrafish. *BMC Biol.* **15**, 4.
50. Johnson, R.E., Linderman, S., Panier, T., Wee, C.L., Song, E., Herrera, K.J., Miller, A., and Engert, F. (2019). Probabilistic models of larval

- zebrafish behavior reveal structure on many scales. *Curr. Biol.* <https://doi.org/10.1016/j.cub.2019.11.026>.
51. Haesemeyer, M., Robson, D.N., Li, J.M., Schier, A.F., and Engert, F. (2015). The structure and timescales of heat perception in larval zebrafish. *Cell Syst.* *1*, 338–348.
 52. Larsch, J., and Baier, H. (2018). Biological motion as an innate perceptual mechanism driving social affiliation. *Curr. Biol.* *28*, 3523–3532.e4.
 53. Schmitt, E.A., and Dowling, J.E. (1999). Early retinal development in the zebrafish, *Danio rerio*: light and electron microscopic analyses. *J. Comp. Neurol.* *404*, 515–536.
 54. Zimmermann, M.J.Y., Nevala, N.E., Yoshimatsu, T., Osorio, D., Nilsson, D.-E., Berens, P., and Baden, T. (2018). Zebrafish differentially process color across visual space to match natural scenes. *Curr. Biol.* *28*, 2018–2032.e5.
 55. Yoshimatsu, T., Schröder, C., Berens, P., and Baden, T. (2019). Cellular and molecular mechanisms of photoreceptor tuning for prey capture in larval zebrafish. *bioRxiv*. <https://doi.org/10.1101/744615>.
 56. Antinucci, P., Folgueira, M., and Bianco, I.H. (2019). Pretectal neurons control hunting behaviour. *eLife* *8*, e48114.
 57. Muto, A., Lal, P., Ailani, D., Abe, G., Itoh, M., and Kawakami, K. (2017). Activation of the hypothalamic feeding centre upon visual prey detection. *Nat. Commun.* *8*, 15029.
 58. Wee, C.L., Song, E.Y., Johnson, R.E., Ailani, D., Randlett, O., Kim, J.Y., Nikitchenko, M., Bahl, A., Yang, C.T., Ahrens, M.B., et al. (2019). A bidirectional network for appetite control in larval zebrafish. *eLife* *8*, e43775.
 59. Helmbrecht, T.O., Dal Maschio, M., Donovan, J.C., Koutsouli, S., and Baier, H. (2018). Topography of a visuomotor transformation. *Neuron* *100*, 1429–1445.e4.
 60. Bianco, I.H., and Engert, F. (2015). Visuomotor transformations underlying hunting behavior in zebrafish. *Curr. Biol.* *25*, 831–846.
 61. Henriques, P.M., Rahman, N., Jackson, S.E., and Bianco, I.H. (2019). Nucleus isthmi is required to sustain target pursuit during visually guided prey-catching. *Curr. Biol.* *29*, 1771–1786.e5.
 62. Štih, V., Petrucco, L., Kist, A.M., and Portugues, R. (2019). Stytra: an open-source, integrated system for stimulation, tracking and closed-loop behavioral experiments. *PLoS Comput. Biol.* *15*, e1006699.
 63. van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* *9*, 2579–2605.
 64. McInnes, L., Healy, J., and Melville, J. (2018). UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv*, arXiv:1802.03426. <https://arxiv.org/abs/1802.03426>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Metronidazole	Sigma-Aldrich	M3761
Tricaine (MS-222)	Sigma-Aldrich	A5040
Deposited Data		
Processed behavioral data	This paper	https://doi.org/10.17632/mw2mmpdz3g.1
Experimental Models: Organisms/Strains		
Tg(ath5:QF2)mpn405	[47]	N/A
Tg(QUAS:epNTR-tagRFP)mpn165	[47]	N/A
<i>lakritz</i>	[46]	<i>th241</i>
<i>blumenkohl</i>	[46]	<i>tz257</i>
Software and Algorithms		
Custom tracking and behavior analysis code	This paper	https://bitbucket.org/mpinbaierlab/mearns_et_al_2019
StreamPix 5	Norpix	https://www.norpix.com/
Python 2.7	Anaconda2	https://www.anaconda.com/
Python 3.4	Anaconda3	https://www.anaconda.com/
Stytra	[62]	http://www.portugueslab.com/stytra/
OpenCV	OpenCV	https://opencv.org/
VirtualDub	VirtualDub	http://www.virtualdub.org/
Tinkercad	Tinkercad	https://www.tinkercad.com/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for reagents should be directed to and will be fulfilled by the Lead Contact, Herwig Baier (hbaier@neuro.mpg.de). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All animal procedures conformed to the institutional guidelines set by the Max Planck Society and were approved under licenses from the regional government of Upper Bavaria (Regierung von Oberbayern).

Wild-type zebrafish larvae

For all experiments, unless noted otherwise, we obtained TLN (nacre) embryos from an outcross of TLN homozygous to TL/TLN heterozygous adults. Until 3 days post fertilization (dpf) embryos were raised in Danieau's solution (17 mM NaCl, 2 mM KCl, 0.12 mM MgSO₂, 1.8 mM Ca(NO₃)₂, 1.5 mM HEPES) at a density of 60 embryos per 50 mL at 28°C with a 14 h-10 h light-dark cycle. Thereafter, embryos were transferred to new dishes containing fish system water and raised at a density of 30 larvae per 50 mL until behavioral testing at 7 dpf or 8 dpf. At 5 dpf and 6 dpf, a few drops of dense paramecia culture (*Paramecium multimicronucleatum*, Carolina Biological Supply Company, Burlington, NC) were added to each dish and larvae were allowed to feed *ad libitum*.

Mutant zebrafish larvae

For experiments with mutants, we used *lakritz* (*lak^{th241}*) and *blumenkohl* (*blu^{tz257}*) mutants [46] in a TL background. *Lak* mutants were obtained from a heterozygous in-cross. Homozygous mutants could be clearly identified by their dark color compared to sibling controls (mixture of heterozygotes and wild types) in a visual background adaptation (VBA) assay. *Blu* mutants were obtained by outcrossing heterozygous females to homozygous males. Similar to *lak*, mutants could be identified unambiguously with a VBA assay. Larvae were raised as described above, except they were not fed at 5 and 6 dpf, and thus their naive prey capture ability was assayed at 7 dpf. This was to minimize potential confounding effects of experience-dependent improvement in prey capture efficacy between groups.

Genetic ablation of retinal ganglion cells (RGCs)

Chemogenetic ablation of RGCs was performed using the Q-system driving the expression of an enhanced version of nitroreductase (epNTR). Double transgenic larvae (*Tg(ath5:QF2)mpn405*; *Tg(QUAS:epNTR-tagRFP)mpn165*) [47] were fed at 5dpf and 6dpf with

paramecia and then incubated with 5mM metronidazole (MTZ, Sigma Aldrich) in fish water containing 0.2% DMSO, for 24 h. The MTZ solution was washed out and larvae were allowed to recover overnight before behavioral experiments were performed at 8dpf. Controls were siblings only expressing *QUAS:epNTR-tagRFP* and treated similarly with MTZ.

METHOD DETAILS

Free-swimming behavioral assay with real prey

Experiments with real prey were conducted using a custom-built behavioral setup. Behavior arenas were produced by flooding a 35 mm Petri dish with 2% agarose (Biozym, Germany), with an acrylic square (15 × 15 mm, 5 mm deep) placed in the center. Once the agarose had set, the acrylic square was removed producing a hollow chamber with transparent walls. Single larvae were introduced to the chamber along with a drop of culture containing approximately 50-100 paramecia. The chamber was filled to the top with fish system water and a glass coverslip was placed over the chamber to flatten the meniscus. This provided a clean, transparent chamber where behavior could be observed and tracked.

Behavior experiments were performed in a climate-controlled box kept at $28 \pm 1^\circ\text{C}$ between 3 and 12 h after lights on. Each larva was recorded for 20 min using a high speed camera (PhotonFocus, MV1-D1312-160-CL, Switzerland), fitted with an objective (Sigma 50 mm f/2.8 ex DG Macro, Japan), connected to a frame grabber (Teledyne DALSA X64-CL Express, Ontario, Canada). The camera was positioned over the behavior arena, which was lit from below with a custom-built near infrared LED array (LEDs from OSRAM Licht AG). Behavior was filmed at 500 frames per second (fps) with a frame size of 500 × 500 pixels covering an area slightly larger than the arena (Figure 1B), providing a final resolution of approximately 0.03 mm/pixel. The aperture of the camera objective was adjusted such that the fish was in focus throughout the entire depth of the arena. Recording was performed using StreamPix 5 software (NorPix, Quebec, Canada) and individual trials were initiated through a custom written Python script. Each 20 min session was split into 20x 1 min recording trials, with < 1 s between the end of one trial and the beginning of the next, to keep video files to a manageable size. If frames were dropped during a trial, the recording was stopped to prevent problems in subsequent analyses. Videos were compressed offline in VirtualDub with Xvid compression before tracking was performed.

Tail and eye tracking

Tracking was performed using custom-written Python scripts. Each frame was tracked independently. Each frame was divided by a background image, calculated as the median of every 100th frame over all trials from a given animal. The frames were then thresholded and contours extracted using OpenCV. The largest contour in the image was taken as the outline of the fish and all other pixels were discarded. Then, the histogram of pixel values of the fish was normalized and a second threshold was applied to find the three largest contours within the fish, corresponding to the two eyes and swim bladder. The eyes were identified automatically as the two contours with the nearest centroids and left and right identities were assigned using the sign of the vector product between lines connecting the swim bladder to these two points. The heading of the fish was defined by a vector starting in the center of the swim bladder and passing through the midpoint between the eye centroids. The angle of each eye was calculated from the image moments of their contours and was defined as:

$$\frac{1}{2} * \arctan\left(\frac{2 * u_{11}}{u_{20} - u_{02}}\right),$$

where u_{ij} is the corresponding central moment. The eye angles in an egocentric reference were calculated as the difference between the heading angle and absolute orientation of the eyes, and eye convergence defined as the difference between the eye angles. A 100 ms median filter was applied to smooth the traces obtained from each eye while preserving edges. The two thresholds used for tracking were set manually for each fish. In frames where the eye contours could not be detected through thresholding, we instead applied a watershed algorithm to obtain contours and then proceeded as above.

Due to the dark pigmentation of *lak* and *blu* mutants, there was insufficient contrast to segment the eyes from the surrounding skin using either thresholding or watershed analysis. For this reason, eye tracking could not be performed in these animals. To calculate the heading in this case, we used the second threshold to segment the head and body of the fish from the tail, for which we identified the minimum enclosing triangle using OpenCV. The heading was then defined as a vector passing through the apex and centroid of this triangle, and the position of the swim bladder was estimated as lying midway between these two points.

To track the tail of the fish, we skeletonized the contour obtained after applying the first threshold described above. We started the tracking from the point on this skeleton nearest to the swim bladder. We used a custom-written algorithm to identify the longest path through the skeletonized image that started at this point, ended at the tip of a branch, and began in the opposite direction of the heading vector. We then linearly interpolated 51 equally spaced points along this path to obtain the final tail points.

The tail tip angle was defined as the angle between the midline of the fish (provided by the heading vector) and a vector between the center of the swim bladder and the last point of the tail. This angle is used to help visualize the sinusoidal oscillation of the tail, but was not used as the basis of any analysis in the paper.

We vectorized the tracked tail points for kinematic analysis in a similar manner to what has been previously described [34, 35]. Briefly, we calculated the angle between the midline (defined by the heading vector) and a vector drawn between each adjacent pair of tail points, providing a 50 dimensional representation of the tail in each frame. A three frame median filter was applied to the heading angle and tail kinematics to remove single frame noise.

The mean tail tip curvature was computed as the mean of the last ten points of the tail angle vector, and was used for bout segmentation. Bouts were detected by applying a threshold to the smoothed absolute value of the first derivative of this mean tail tip curvature. Uncharacteristically long bouts detected with this method were further split by finding inflection points in the smoothed absolute value of the mean tail tip curvature convolved with a cosine kernel.

Generating a behavioral space

To generate our behavioral space, we excluded any bouts during which the tail of the fish hit the wall of the behavior chamber. This was to ensure that only the fish's self-generated motion – and not motion artifacts introduced from distortion of the tail by the wall – was considered when mapping the behavioral space. Consequently, not all the bouts we observed could be mapped into the space.

To describe bouts in terms of their postural dynamics, we performed principal component analysis (PCA) on the tail kinematics across all bout frames. Data were normalized before applying PCA by subtracting the mean tail shape and dividing by the standard deviation.

The next step in generating the behavioral space involved computing the distance between every pair of bouts with dynamic time warping (DTW) [37]. DTW finds an alignment between two time series that minimizes a cost function, which is the sum of the Euclidean distances between each pair of aligned points. In our analysis, we only allowed trajectories to be warped within a 10 ms time window. For bouts of different lengths, we padded the end of the shorter bout with zeros until it was the same length as the longer bout. We performed each alignment twice, reversing the sign of all the values for one of the trajectories the second time, and considered the distance between two bouts to be: $\min(DTW(t_1, t_2), DTW(t_1, -t_2))$, thus effectively ignoring the left/right polarity of the bouts.

To generate the behavioral space, we performed a round of affinity propagation [38] prior to embedding, using the negative DTW distance between a given pair of bouts as a measure of their similarity. We varied the affinity parameter in this step, testing values ranging from 400 to 4000 (number of clusters ranging from 1634 to 179), and found it did not have a strong impact on the structure of the final behavioral space (Figure S1B). Therefore, we used the median similarity between bouts as the preference for the clustering (~400). Doing so provided 2,802 clusters, of which we excluded any clusters containing fewer than three bouts, thus ensuring that only repeatedly observable motor patterns were used for generating the behavioral space. As a final quality check, we manually inspected every cluster exemplar and removed incorrectly identified bouts, which usually was the result of tracking artifacts from a paramecium crossing the tail of the fish. The final number of clusters that we kept for embedded was 1,744.

Since affinity propagation identifies an exemplar to represent each cluster, we produced our final behavioral space by performing isomap embedding [39] of these exemplars. For the isomap embedding, we constructed a nearest-neighbors graph of the exemplars using their DTW distances, and calculated the minimum distance between each pair of points in this graph. The isomap components correspond to the eigenvectors of this graph distance matrix.

In addition to isomap, we also performed t-SNE [63] and UMAP [64] of the data. As with isomap, we performed the embedding using the precomputed DTW distances between the 1,744 exemplars. For t-SNE, we systematically varied the perplexity (10, 20, 50) and learning rate (10, 100, 1000) of the embedding to see whether this significantly changed the visualization of the behavioral space (Figure S1C). Similarly with UMAP, we systematically varied the nearest neighbors parameter trying values between 5 and 50 (Figure S1D).

Isomap, t-SNE and UMAP embeddings in Figure S1E were obtained using the precomputed DTW distances between all bouts in the dataset. We used 20 nearest neighbors (or 20 perplexity, in the case of t-SNE) to generate these spaces.

Mapping kinematic features and eye convergence into the behavioral space

With our PCA-DTW-isomap approach, each point in the behavioral space represents a small cluster of bouts. For each bout, we calculated the mean speed, angle through which the fish turned, maximum angular velocity of the fish, and the time at which the maximum angular velocity occurred (turn onset). In Figure S1F, we show the median of each of these features over a cluster. Similarly, we could calculate the proportion of bouts in each cluster that occurred during spontaneous, early, mid, or late prey capture as defined below (see Eye convergence analysis). The prey capture index was defined as:

$$\frac{\# \text{prey capture bouts in cluster} - \# \text{spontaneous bouts in cluster}}{\# \text{bouts in cluster}}$$

Mapping bouts from other experiments into the behavioral space

To map bouts from new experiments into the behavioral space, we extracted tail kinematics and identified bouts as described above (see Tail and eye tracking). The postural dynamics of each new bout was projected onto the first three principal components obtained from the main dataset to bring it into the same space as bouts from that dataset. Then, each bout was mapped to one of the 1,744 exemplars identified in “Generating a behavioral space” using dynamic time warping (DTW), with the nearest exemplar having the smallest DTW distance to the bout. In this way, each new bout could be projected into the three dimensional behavioral space defined by the 1,744 exemplars.

Singular-value decomposition of behavioral transitions

To identify transition modes, we generated a transition frequency matrix, M , where M_{ij} contains the number of transitions from micro-cluster j to micro-cluster i , where each micro-cluster is a small cluster of bouts in the behavioral space identified with affinity propagation (see Generating a behavioral space). This matrix included all the transitions from all animals for a given experiment.

Since there are more than 3 million ($1,744^2$) possible transitions between motifs, and only 44,154 transitions in our largest dataset, the matrix M is necessarily sparse. This would hinder the identification of common dynamical motifs, and so we performed smoothing on matrix M by blurring similar transitions into each other. To achieve this, we took advantage of the fact that nearby points in our behavioral space encode bouts with similar postural dynamics. We computed a weighting matrix, W , where $W_{ij} \equiv e^{-a \cdot d(p_i, p_j)}$, $d(p_i, p_j)$ is the Euclidean distance between a pair of points in the three-dimensional behavioral space, and a is a smoothing factor (see Figure S2A).

We normalized matrix W so that the columns summed to one and then smoothed the transitions in matrix M with the transformation: $M_{smooth} = WMW^T$.

To distinguish between symmetric transitions (i.e., those that occur in both direction), and antisymmetric transitions (i.e., those in which transitions in one direction outweigh those in the other), we decomposed the smoothed matrix, M_{smooth} , into its symmetric and antisymmetric parts, where:

$$M_{smooth} = M_{symmetric} + M_{antisymmetric}$$

$$M_{symmetric} = 1/2 (M_{smooth} + M_{smooth}^T)$$

$$M_{antisymmetric} = 1/2 (M_{smooth} - M_{smooth}^T)$$

The symmetric and antisymmetric transition modes were found by taking the SVD of these two matrices respectively.

Every real or complex matrix, A , can be factorized using the singular-value decomposition (SVD) into three matrices such that:

$$A = USV^T$$

The columns of U and rows of V^T define two sets of orthonormal basis vectors and S is a diagonal matrix containing the singular values, ordered from largest to smallest. The SVD describes the transformation performed by matrix, A . Under this transformation, each row of the matrix, V^T , is mapped to the corresponding column of U and scaled by the associated singular value. Therefore, this decomposition provides an unbiased description of the most common transitions between micro-clusters.

A symmetric matrix, such as $M_{symmetric}$, geometrically defines a scaling transformation. Consequently, its singular-value decomposition is the same as its eigendecomposition: spaces U and V are the same and S contains the eigenvalues. As such, the n^{th} transition mode of $M_{symmetric}$ can be written:

$$\vec{v}_n \cdot \sigma_n \cdot \vec{v}_n^T$$

where \vec{v}_n is the singular vector with corresponding singular value, σ_n . To visualize symmetric transition modes (S-modes), we found the contribution of each micro-cluster to the singular vector defining that transition mode and show this in the behavioral space.

An antisymmetric matrix, such as $M_{antisymmetric}$, describes a set of orthogonal rotations. As such, spaces U and V are related by a 90° rotation and each transition mode can be written:

$$\begin{pmatrix} \vec{v}_1 & \vec{v}_2 \end{pmatrix} \begin{pmatrix} 0 & -\sigma_n \\ \sigma_n & 0 \end{pmatrix} \begin{pmatrix} \vec{v}_1 & \vec{v}_2 \end{pmatrix}^T$$

where \vec{v}_1 and \vec{v}_2 are orthonormal, and σ_n is the corresponding singular value. Positive values in \vec{v}_1 map to positive values in \vec{v}_2 , positive values in \vec{v}_2 map to negative values in \vec{v}_1 , negative values in \vec{v}_1 map to negative values in \vec{v}_2 and negative values in \vec{v}_2 map to positive values in \vec{v}_1 :

$$\begin{array}{ccc} v_1^+ & \rightarrow & v_2^+ \\ \uparrow & & \downarrow \\ v_2^- & \leftarrow & v_1^- \end{array}$$

These are the transitions we represent with arrows in figures displaying A-modes.

Investigating transition modes with toy data

To generate a toy behavioral space, we generated nine clusters with a 2D Gaussian distribution of points (SD = 0.06) and cluster centers drawn randomly from a 2D uniform distribution between 0 and 1. Each cluster contained a random number of points (drawn from a normal distribution; mean = 50, SD = 5). We selected a random seed that produced significant overlap between clusters in the 2D hypothetical behavioral space. Each point in this space was considered analogous to a micro-cluster in our isomap space.

We manually distributed the nine clusters between three hypothetical “states.” For “separated behavioral states,” each state consisted of three clusters and each of the nine clusters belonged to only a single state. For “overlapping behavioral states,” each state consisted of four clusters and three of the clusters were shared between two states.

To generate symmetric transition structure between the three states, we used a left stochastic matrix, P_S , with values:

$$P_S = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

To generate antisymmetric transition structure between the three states, we used a left stochastic matrix, P_A , with values:

$$P_A = \begin{pmatrix} 0.3 & 0.1 & 0.7 \\ 0.5 & 0.1 & 0.1 \\ 0.2 & 0.8 & 0.2 \end{pmatrix}$$

To generate transition frequency matrices, we performed Monte Carlo simulation with 2000 transitions (separated behavioral states) or 10,000 transitions (overlapping states) using one of the stochastic matrices above. For each transition, we randomly selected a “micro-cluster” from the next state in the sequence. After the simulation was complete, we added random transitions to the matrix for a final total of 2500 transition (separated states) or 12,500 transition (overlapping states).

We then proceeded to perform smoothing of the transition matrices obtained in our simulations, split matrices into symmetric and antisymmetric components, performed SVD on each of these matrices, and performed hierarchical clustering in the space defined by transition modes (first two S-modes for symmetric, first A-mode for antisymmetric). For separated behavioral states we generated three clusters and for overlapping behavioral states we generated six clusters (three regions of the behavioral space that contribute to a single state, three regions of the behavioral space that contribute to multiple states). We then computed a confusion matrix, C , where C_{ij} is the number of points assigned cluster label i by our hierarchical clustering and whose ground-truth label is j .

Eye convergence analysis

To identify periods of eye convergence, we calculated a kernel density estimation (Gaussian kernel, bandwidth = 2.0) of the eye convergence angles across all frames for a given fish. This distribution was bimodal (eyes converged or unconverged) and therefore we defined the eye convergence threshold as the antimode (least frequent value between the two modes). To identify spontaneous, early, mid, and late prey capture bouts, we calculated the mean eye convergence angle over the first and last 20 ms of a bout, and concluded the eyes were converged if this number was above the threshold. Bouts were classified as spontaneous if the eyes were unconverged at the beginning and end of a bout; early prey capture if the eyes were unconverged at the beginning and converged at the end of the bout; mid prey capture if the eyes were converged at the beginning and end of the bout; and late prey capture if the eyes were converged at the beginning and unconverged at the end of the bout.

Clustering in a combined kinematic-transition space

To define behavioral clusters, we combined information about bouts’ kinematics and transitions to generate a new combined kinematic-transition space. Kinematic similarity between exemplars was computed from DTW distances as described above. We constructed a transition space by combining the first two non-common symmetric transition modes, S1, and S2, and the pair of vectors defining the first antisymmetric transition mode, A1, and then calculating an orthogonal basis (using the Gram-Schmidt process).

We multiplied the kinematic DTW distances between exemplars by the distance between exemplars in transition space (generated by combining singular vectors as described above). Each exemplar was then represented by a feature vector, with each feature being the similarity to every other exemplar. We performed dimensionality reduction on this new feature space (using isomap), retaining 20 components. We then performed ward hierarchical clustering in this reduced space. We calculated the silhouette score associated with different thresholds for defining cluster boundaries, and identified a local maximum at seven. The bouts belonging to each cluster were inspected and confirmed to align well with previous annotations of zebrafish behavior.

In [Figure 3C](#), we colored points in the original behavioral space based on the cluster they were assigned in the combined kinematic-transition space. The transparency value in that graph was determined by the number of nearest neighbors that were assigned the same cluster label.

To produce average traces for the tail tip angle in [Figure 3B](#), we aligned all exemplars belonging to a given cluster using dynamic time warping and took the average of the aligned traces. The representative examples we show are real bouts with a similar tail angle trace to the average.

Modeling transitions between clusters

For this analysis, we first identified every uninterrupted chain containing at least two bouts in our data which could be assigned a behavioral cluster, i.e., only chains of bouts from within a single recording trial (see *Free-swimming behavioral assay with real prey*) and that could be embedded in the behavioral space (see [Generating a behavioral space](#)). We then tested the ability of a series of Markov models – ranging from zeroth to fifth order – to predict each subsequent bout. For this purpose, we modeled each cluster as a state in a Markov process (allowing transitions to the same state, since fish can perform the same type of bout twice in a row). Each of our models contained seven states (from our behavioral clusters), s_1, s_2, \dots, s_7 , and we denote the current state, t_0 , the next state t_{+1} , the previous state t_{-1} , etc.

A zeroth order Markov model does not know the current state and therefore guesses the next state based simply on the distribution of bouts across all states:

$$P(t_{+1} = s_i | t_0 = s_j) = P(s_i)$$

In a first order Markov model, the current state is known. To predict the next state, we considered all other times the current state was visited (t_n) and observed which bout occurred next in the sequence:

$$P(t_{+1} = s_i | t_0 = s_j) = P(t_{n+1} = s_i | t_n = s_j)$$

For the second-order Markov model, we took into account the last two states in a chain when predicting the next state:

$$P(t_{+1} = s_i | t_0 = s_j, t_{-1} = s_k) = P(t_{n+1} = s_i | t_n = s_j, t_{n-1} = s_k)$$

Similarly, for Markov models up to order, m , we predicted the next state:

$$P(t_{+1} = s_i | t_0 = s_j, t_{-1} = s_k, \dots, t_{1-m} = s_n) = P(t_{n+1} = s_i | t_n = s_j, t_{n-1} = s_k, \dots, t_{n+1-m} = s_n)$$

Paramecium tracking

To track paramecia within frames, we performed background division (see [Tail and eye tracking](#)) followed by Gaussian blurring (using a 3×3 pixel kernel) over the image. Next, we applied an adaptive threshold to the image and rotated and centered the fish in the frame using the midpoint between the eyes and the heading obtained from tracking. To identify paramecia, we applied a median filter over the image (5×5 pixels), then performed gray erosion (3×3 flat structuring element) and thresholding. We counted contours with an area > 3 pixels as paramecia, and used the centroids of these contours to determine prey position in the visual field.

Generating stimulus maps

To obtain maps of prey probability density, we performed paramecium tracking as described above for all bouts belonging to a given behavioral cluster (depending on the figure, using either the first and last frame as defined by bout segmentation, or defined time points). Normalized histograms of prey density were obtained after overlaying paramecium positions in all images and masking out the area of the image containing the fish (identified through thresholding of an average image). These histograms were then smoothed with a Gaussian kernel. To identify regions of the visual field with a higher paramecium density, we applied the same process to $\sim 90,000$ randomly chosen frames from all videos and subtracted this baseline density from the image. We then threshold the resulting density plot using the 85th percentile, setting all pixels below this value to zero, and overlaid the average image of the fish.

We obtained time series of paramecium density by aligning video sequences to the onset of the capture strike and proceeded to analyze frames as described above.

Free-swimming virtual prey capture assay

We designed a setup to present virtual prey stimuli to individual zebrafish larvae via a screen on the surface of the water. We adapted a previously described virtual reality setup (Stytra) which allows real-time tracking and presentation of visual stimuli [62]. Behavior arenas were produced by flooding a 55 mm Petri dish with 2% agarose (Biozym, Germany), with a circle mold (18×18 mm, 5 mm deep) placed in the center. Once the agarose had set, the mold was removed producing a hollow chamber with transparent walls. Individual animals are placed in the arena with fish water and a projection filter (3008 - Tough Frost, Rosco) was placed on top. Visual prey-like stimuli were projected onto this filter via a cold mirror (45° AOI, 101×127 mm, Edmund optics).

Animals were recorded at 300 fps with a Ximea MQ013MG-ON camera placed below the arena using a 25 mm lens (Edmund Optics Nr. 59-872). Diffuse infrared illumination for imaging was provided from the top using an IR LED array (RAYMAX 25, 120°). A white light source was also provided from the top. Stimuli were presented to the fish via the projector (DLP LightCrafter 4500, 912×1140 pixels) but blocked before the camera by an IR band-pass filter. Three lenses were combined together to reduce the size of the projected stimuli on the screen (one 150 mm lens, AC508-150-A and two 75 mm lenses, AC508-075-A-ML). Image acquisition, real-time processing and stimulus generation were performed on a Desktop PC running Stytra [62]. Briefly, the background was modeled using a mean image from several previous frames. The difference between the current frame and the thresholded background was computed. The eyes and swim bladder were identified as the largest contours in the thresholded image. The heading of the fish and the angle of each eye were computed using the vector starting in the center of the swim bladder and passing through the midpoint between the eye centroids (see [Tail and eye tracking](#)). Each experiment started with a two minute pre-trial period (no stimulus) in order for the fish adapt to the arena. This period was used to compute a distribution of spontaneous eye convergence angles. The 98th percentile of the data was computed from this distribution and subsequently used during virtual prey presentation as a threshold to detect eye convergence (prey capture). Virtual prey stimuli consisted of gray dots of a single pixel moving linearly with a speed matching real paramecia. For each experiment, 50 virtual prey items were presented on each trial. We presented animals with six trials, lasting 40 s each. In three of the trials virtual prey disappeared for 2 s each time eye convergence was detected online (“vanishing” trials). In the other three trials, prey remained visible the whole time (“persisting” trials). The order of persisting and vanishing trials was randomized for each fish. Five periods of no stimulus presentation lasting 24 s each were interspersed in between each trial. Analysis was performed with custom-written Python scripts.

Capture strike analysis

We defined capture strikes as bouts that belonged to a micro-cluster containing > 50% late prey capture bouts. To determine the moment of capture in [Figure 5A](#), we selected 100 random capture strikes and manually annotated the frames where the jaw was maximally extended.

For subsequent analysis, we only considered the 50 ms time window shown in [Figure 5A](#) (24-74 ms after the bout onset as determined by our bout segmentation algorithm) and proceeded with our general DTW-isomap embedding algorithm as described above (see [Generating a behavioral space](#)). To generate the capture strike subspace, we computed the DTW distance between each pair of strikes, only allowing warping within a 6 ms (3 frames) time window. We used the resulting pairwise distance matrix directly for isomap embedding, keeping the first two dimensions. Note we did not need to perform an intermediate affinity propagation micro-clustering step, due to the small size of this dataset. We then performed k-means clustering ($k = 2$) to classify strikes.

Free-swimming behavioral assay in three dimensions

To record behavior simultaneously from above and from the side, we designed a new chamber. A 3 mL transparent, unfrosted plastic cuvette was with flooded with 2% agarose. An acrylic rod (20 × 5 × 5 mm) was inserted into the liquid agarose, which was allowed to set, after which the rod was removed leaving behind a hollow chamber. As before, individual larvae were introduced into the chamber with a drop of paramecia culture topped up with fish system water. The opening was plugged with a small piece of acrylic cut to match the cross section of the chamber (5 × 5 mm). The cuvette was placed on its side on top of a glass coverslip suspended above a mirror angled at 45°. The high speed camera was positioned above this setup in such a way as to allow the fish in the chamber as seen from above as well as the reflected side view from the mirror to be visible within the field of view of the camera. The IR LED array was rotated by 90°, allowing the chamber to be illuminated from the side and from below (via the mirror) with a single light source. We reduced the aperture of the camera objective so that the entire arena was in focus in both views and offset the decrease in luminance by increasing the exposure time of each frame. Consequently, for this experiment we achieved a frame rate of 400 fps. As described above, data from each fish was split into 20x 1 min recording trials.

To record jaw movements during prey capture with higher spatial resolution in [Video S6](#), we used two cameras (PhotonFocus, MV1-D1312-160-CL, Switzerland) and two light sources and filmed a number of fish swimming in a custom-built transparent chamber. We waited for one of the fish to start hunting a paramecium in the field of view of both cameras and manually triggered the recording. Frame acquisition was synchronized using StreamPix 5 and a dual camera frame grabber.

Jaw tracking

As for the single view setup, each frame was tracked independently offline using custom-written Python scripts. Each frame was divided by a background image, calculated as the median of every 100th frame over a recording trial. The upper and lower halves of the frame were tracked separately. The lower half of the frame, containing the image of the fish as seen from above, was tracked as described above. Fish were only tracked from the side when their heading was within $\pm 45^\circ$ of the imaging plane to minimize artifacts arising as a result of foreshortening. Frames were thresholded and contours extracted using OpenCV. The largest contour in the image was taken as the outline of the fish and all other pixels were discarded. Then, the histogram of pixel values of the fish was normalized and a second threshold was applied to find a contour enclosing the head and body of the fish. The pitch and angle of the cranium were calculated using image moments of these two contours respectively, with cranial elevation defined as the difference between them.

To find the point of the base of the jaw, we first identified the centroid of the head-body contour and the vector defined by the cranium angle (i.e., orientation of this contour in the frame). We extended this vector from the centroid until it intersected the fish contour. Next, we found the midpoint between the centroid and this intersection point. From this midpoint, we extended a vector orthogonal to cranium angle vector until it intersected the fish contour at the base of the jaw. Jaw depression was defined as the Euclidean distance between the midpoint and this intersection point.

The cranial elevation angle and jaw depression were smoothed with an edge-preserving five-frame median filter. Then, we applied a high-pass filter by subtracting the baseline of these two kinematic features over a recording. To compute this baseline, we first calculated a 250 ms rolling minimum, and then computed the one-second rolling mean of this rolling minimum. This provided a relatively stable baseline for identifying jaw movements, despite changes in pitch and azimuth of the fish over a recording. To segment jaw movements, we identified periods when the baseline-adjusted jaw depression, smoothed with a 50 ms rolling average, was above a predetermined threshold and defined movement onset and offset as inflection points in this smoothed trace.

Generating a behavioral space of jaw movements

To generate the jaw movement behavioral space in [Figure 6F](#), we performed PCA on the jaw depression and cranial elevation traces across movement frames (see [Jaw tracking](#)). We calculated the DTW distance (warping bandwidth = 10 ms) between each pair of movements projected onto the first principal component ([Figure 6E](#)), and performed isomap embedding using the resulting distance matrix. To identify clusters, we used Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) (hdbscan library, Python).

Paramecium tracking from the side

We tracked paramecium and generated stimulus maps from the side in a similar manner as described above (see [Paramecium tracking](#) and [Generating stimulus maps](#)). To generate stimulus maps, we aligned frames using the centroid of the contour outlining the head and the pitch of the fish in the water. Baseline paramecium density was calculated from ~18,000 randomly selected frames.

Lensectomy

For lensectomy experiments, larvae were allowed to feed on paramecia *ad libitum* between 5 and 6 dpf. At 7 dpf, larvae were anaesthetized in 0.02% tricaine (MS-222, Sigma-Aldrich) and then embedded in 2% low melting point agarose (Invitrogen). Larvae were kept anaesthetized for the entire surgery. Some of the agarose was removed from around the eyes using a scalpel to facilitate access. Animals undergoing sham, unilateral, or bilateral lensectomy were all treated in the same way up to this point and mounted alongside each other to ensure as similar treatment conditions as possible. Lensectomies were performed with borosilicate glass needles pulled to a sharp point. A single horizontal incision was made in the transparent cuticle of the eye and the lens was removed, taking care not to damage other parts of the eye such as the retinal pigment epithelium. This was performed on either one or both eyes. Unilateral lensectomies included animals that had either the left or right lens removed. Animals were freed from the agarose and allowed to recover overnight in fish system water. Immediately prior to behavioral testing at 8 dpf, animals were inspected for swelling or other aberrations to the eye, and excluded from the experiment if this was seen. Animals were tested in the 2D prey capture assay as described above. After behavioral testing, each larva was re-embedded in agarose and a snapshot of its head obtained with a camera mounted to a stereomicroscope to confirm that the lens had not regenerated.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical tests were performed using the SciPy library in Python, or, if appropriate tests were not available in this library, we wrote new code to perform statistical tests. For plots, the type of error bar and p values are indicated in the figure legends.

Identification of significant transition modes

Significant transition modes in [Figure 2D](#) were identified by permutation testing and cross validation. Animals were split into training and test groups (50% animals in each group). Transition matrices were constructed for these two groups and SVD was used to identify transition modes of the training group. For each mode we constructed a transition model by adding the common transition mode to it. We compared these training models with the test group's transition matrix by computing the sum of squares error between them. This was repeated across 10,000 splits into training and test datasets. We tested the significance of each mode with a one-sided t test on the distribution of sum-of-square errors across all partitions of the data (significance level = 0.01), for modes that provided a better prediction than the common-mode model alone.

Comparison of transition modes across experimental datasets

To compare transition modes across datasets and treatment groups, we calculated the absolute value of the cosine similarity between the first four symmetric modes (S-modes) or first four antisymmetric mode pairs (A-modes). To test significance, we performed a permutation test. We permuted the labels between the two groups 1000 times (or if an exact test required less than 1000 comparisons, used that) and calculated the pairwise similarity of equivalent modes. This generated a null distribution of similarities against which we could calculate a p value for the observed similarity between unshuffled groups. For calculating the similarity between antisymmetric modes, we compared both left and right singular vectors to account for rotations and took the maximum. In the [Results](#) section, we report the similarity between like transition modes (i.e., always comparing S1 to S1 and A1 to A1 etc.). In [Figures S3C](#), [S5D](#), and [S6C](#) we show pairwise similarities between different transition modes.

Simplex projection to test higher order behavioral dynamics

To test whether higher-order behavioral dynamics were present in behavioral data ([Figure 3E](#)), we took each focal bout and inspected the preceding bouts in the chain. We identified all other occasions that this sequence of states occurred in the rest of the data and calculated which state was most likely to occur next. We calculated the probability of predicting the next bout correctly across all instances of a given behavioral cluster to obtain a distribution. We compared the probability distributions obtained by considering the past n-1 bouts in a chain with those obtained comparing the past n bouts in the chain using a one-tailed Student's t test and corrected for multiple comparisons with a Bonferroni correction (significance level = 0.01).

Identification of significant transitions in an ethogram

To identify which first-order transitions between behavioral clusters were significant ([Figure 3F](#)), we used a permutation test. We shuffled the order of bouts *within* each fish 1000 times and recomputed the first-order Markovian transition probability matrices. This gave a reference distribution of transition probabilities between each pair of modules from which we could calculate one-tailed p values. We considered significant transitions as those that had a p value < 0.05 after applying a Holm-Bonferroni correction ($7^2 = 49$ comparisons).

Statistical analysis of virtual prey experiments

We compared the cumulative distributions of bout duration and number of prey capture bouts per sequence across all trials with a Kolmogorov-Smirnov test. To compare averages for these two metrics, we calculated the mean per fish and then compared control and test trials with a Wilcoxon signed rank test.

Quantification of prey distance during capture strikes

We tracked paramecia in the front central visual field of the larvae and calculated the probability density as described above. We compared groups with a permutation test on energy distances. We shuffled labels between groups (attack swim versus S-strike or unilateral versus sham) 1,000,000 times to calculate a null distribution of energy distances. We calculated the one-tailed p value by comparing the observed energy distance to this null distribution. We calculated the energy statistic in two dimensions to compare 2D stimulus maps, or one dimension to compare distances and angles.

Quantification of lensectomy experiments

In [Figure 7](#), we tested time spent in prey capture (prey capture score), hunt initiation rate, strike probability, and number of capture strikes of each type between sham and unilateral, or unilateral and bilateral conditions with a Mann-Whitney U-test.

DATA AND CODE AVAILABILITY

The code generated during this study and sample data are available on Bitbucket (https://bitbucket.org/mpinbaierlab/mearns_et_al_2019). Processed data generated during this study are available at Mendeley data (<https://doi.org/10.17632/mw2mmpdz3g.1>). The full dataset generated during this study has not been deposited in a public repository due to its large size but is available from the corresponding author on request.