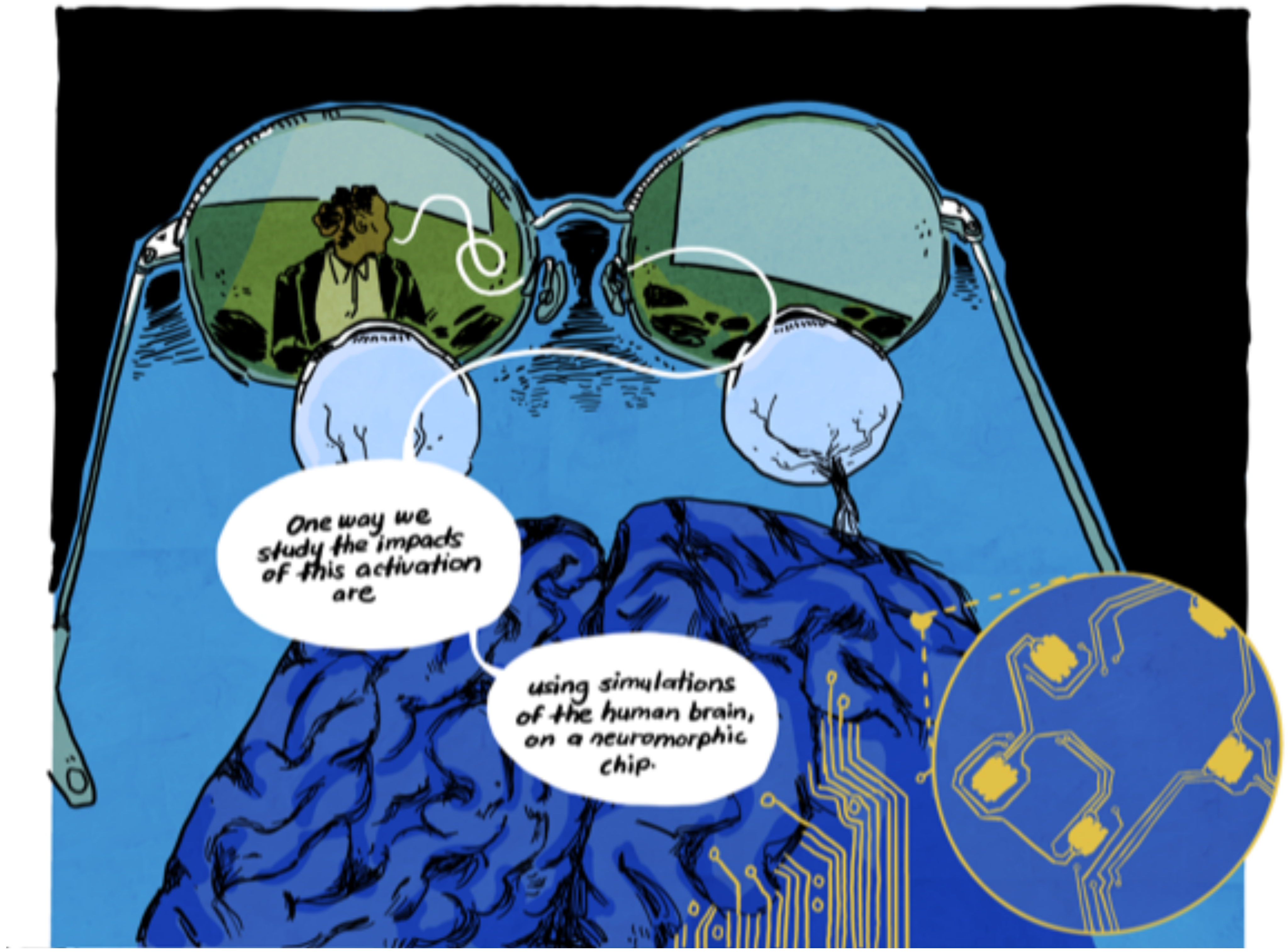
Dynamical systems for behavioral and neural data

Luca Mazzucato

Depts of Biology, Mathematics & Physics
Institute of Neuroscience
University of Oregon

CSHA Computational and Cognitive Neuroscience Summer School



Artwork by Audra McNamee

Dynamical systems for behavioral and neural data

- TODAY: hidden Markov models for **behavior**:
  - Slides
    - Spontaneous movements - HMM with Autoregressive observations
    - Naturalistic foraging - HMM with Linear Model emissions
    - HMM best practices
  - Live coding session
    - Mystery dataset
    - Fit Linear Model, Mixture of Linear Model
    - Sample and fit data w/ LM-HMM
    - Model selection for # of hidden states

Behavior is a complex dynamical system

Behavior is a complex dynamical system

Behavior is a complex dynamical system

words / **actions** ~100 ms

Whisking, sniffing, moving a paw, …

Behavior is a complex dynamical system

words / **actions** ~100 ms

Whisking, sniffing, moving a paw, …

sentences / **sequences** ~1 s

Walking, rolling the walnut, hiding

Behavior is a complex dynamical system

words / **actions** ~100 ms

Whisking, sniffing, moving a paw, …

sentences / **sequences** ~1 s

Walking, rolling the walnut, hiding

stories / **plans** ~1 min

Must get walnut!

Behavior is a complex dynamical system

words / **actions** ~100 ms

Whisking, sniffing, moving a paw, …

sentences / **sequences** ~1 s

Walking, rolling the walnut, hiding

stories / **plans** ~1 min

Must get walnut!

Behavior ⟵ Neural activity

Behavior is a complex dynamical system

words / **actions** ~100 ms

Whisking, sniffing, moving a paw, …

sentences / **sequences** ~1 s

Walking, rolling the walnut, hiding

stories / **plans** ~1 min

Must get walnut!

Behavior ⟷ Neural activity

Behavior is a complex dynamical system

words / **actions** ~100 ms

Whisking, sniffing, moving a paw, …

sentences / **sequences** ~1 s

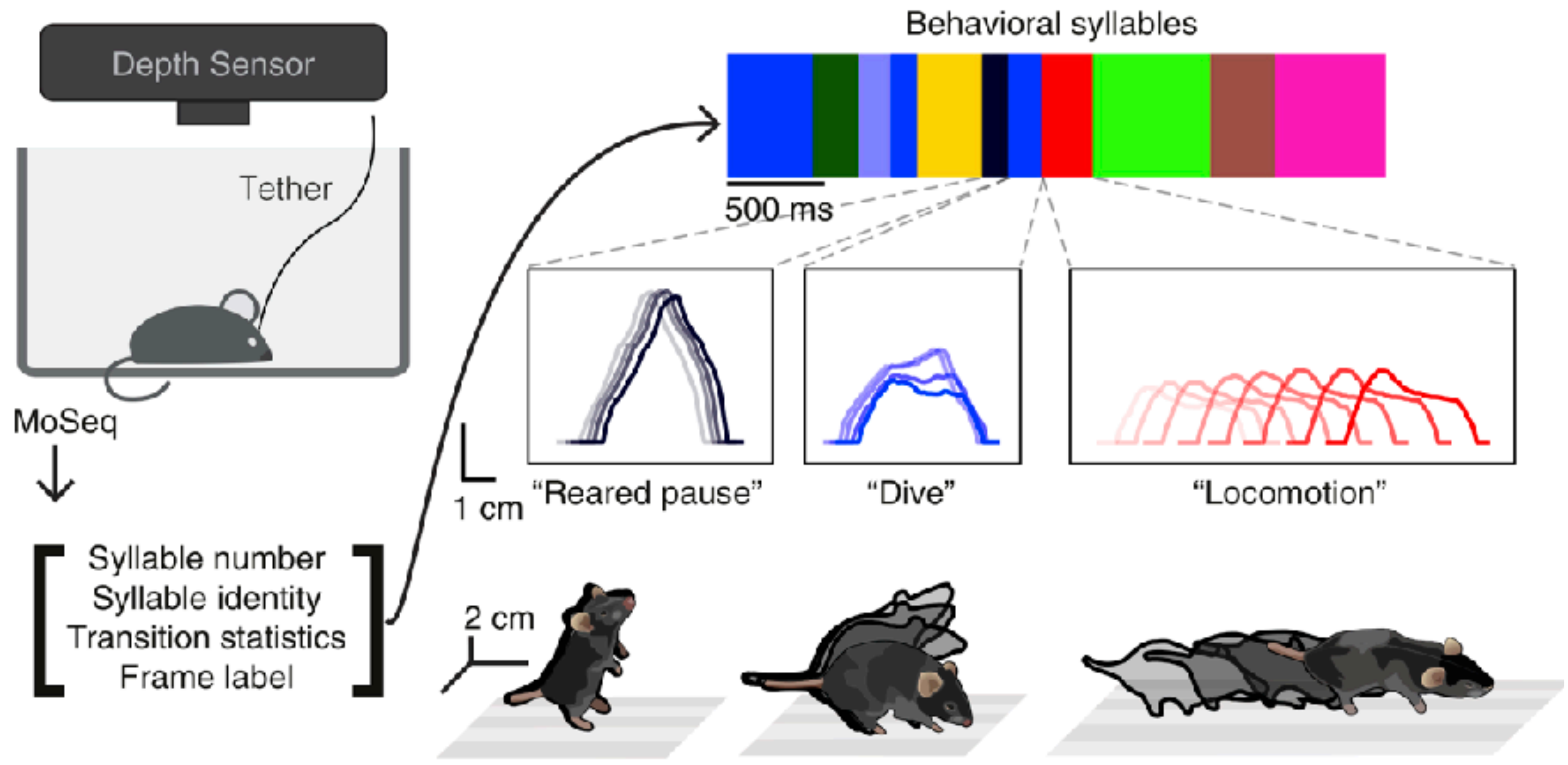
Walking, rolling the walnut, hiding

stories / **plans** ~1 min

Must get walnut!

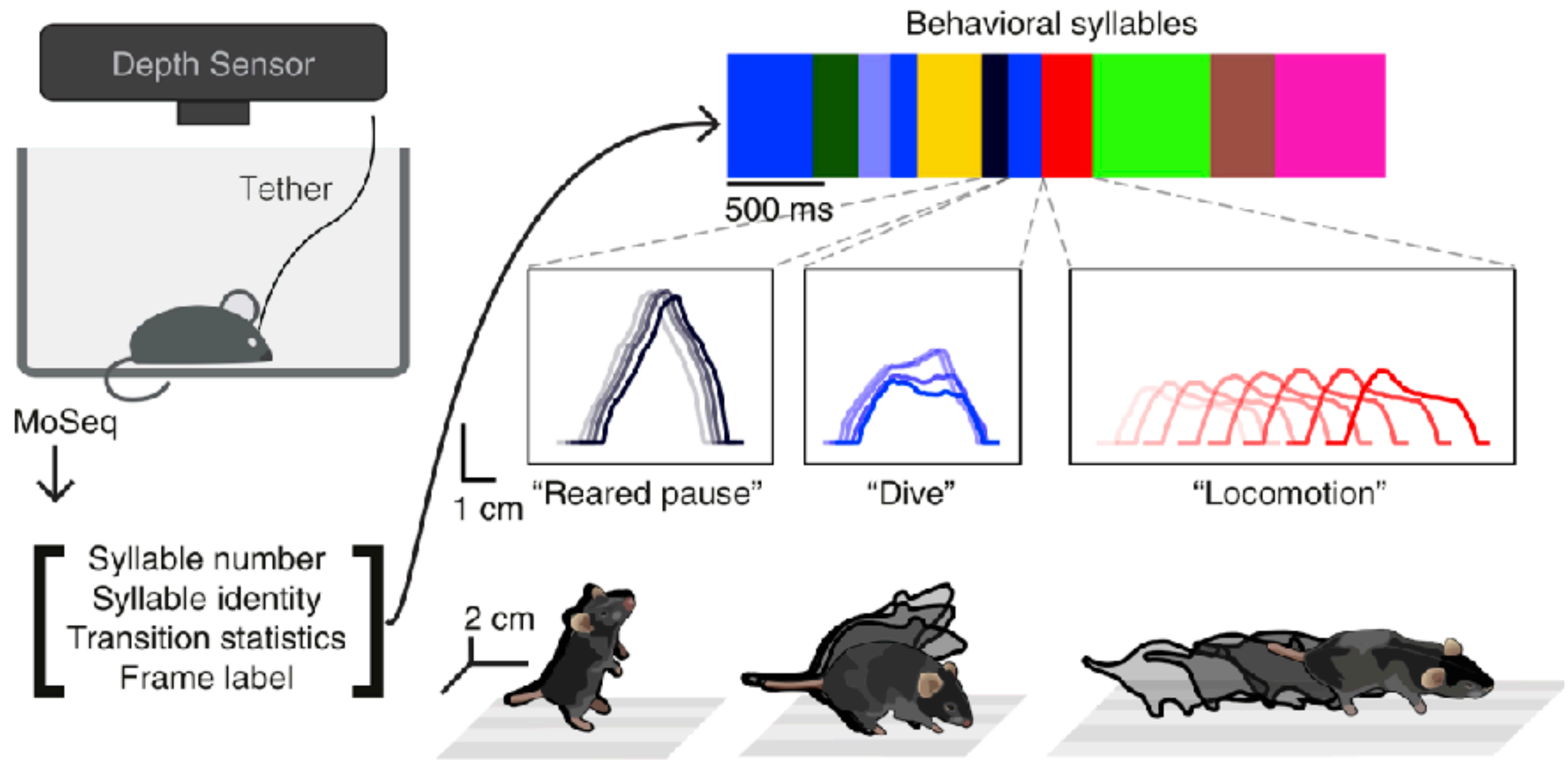
Behavior ⟷ Neural activity

Complex temporal structure in behavior



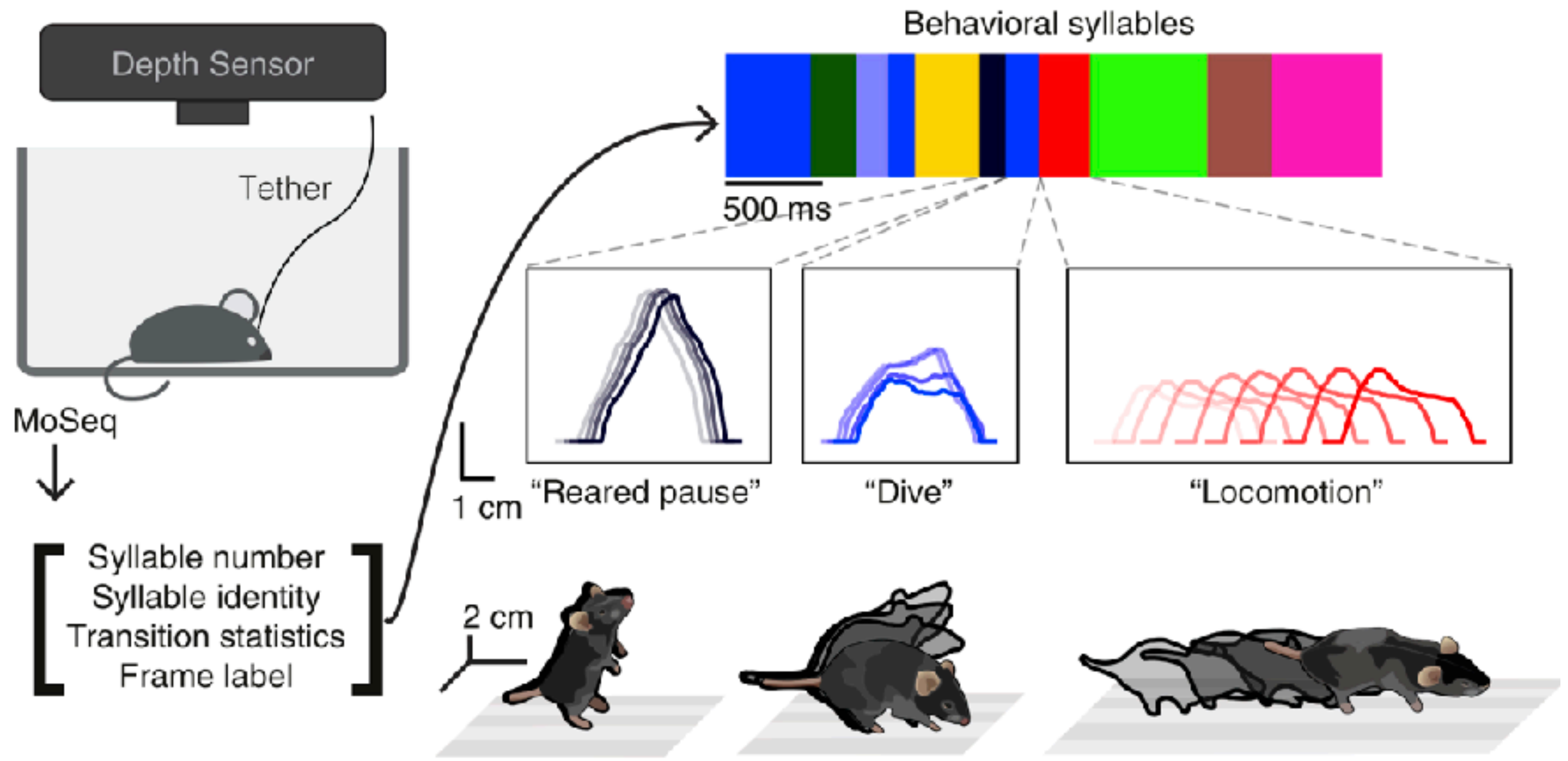
[Wiltschko et al, 2015]
[Markovits et al, 2018]
[Findley et al, 2021]

Complex temporal structure in behavior



[Wiltschko et al, 2015]
[Markovits et al, 2018]
[Findley et al, 2021]

Complex temporal structure in behavior

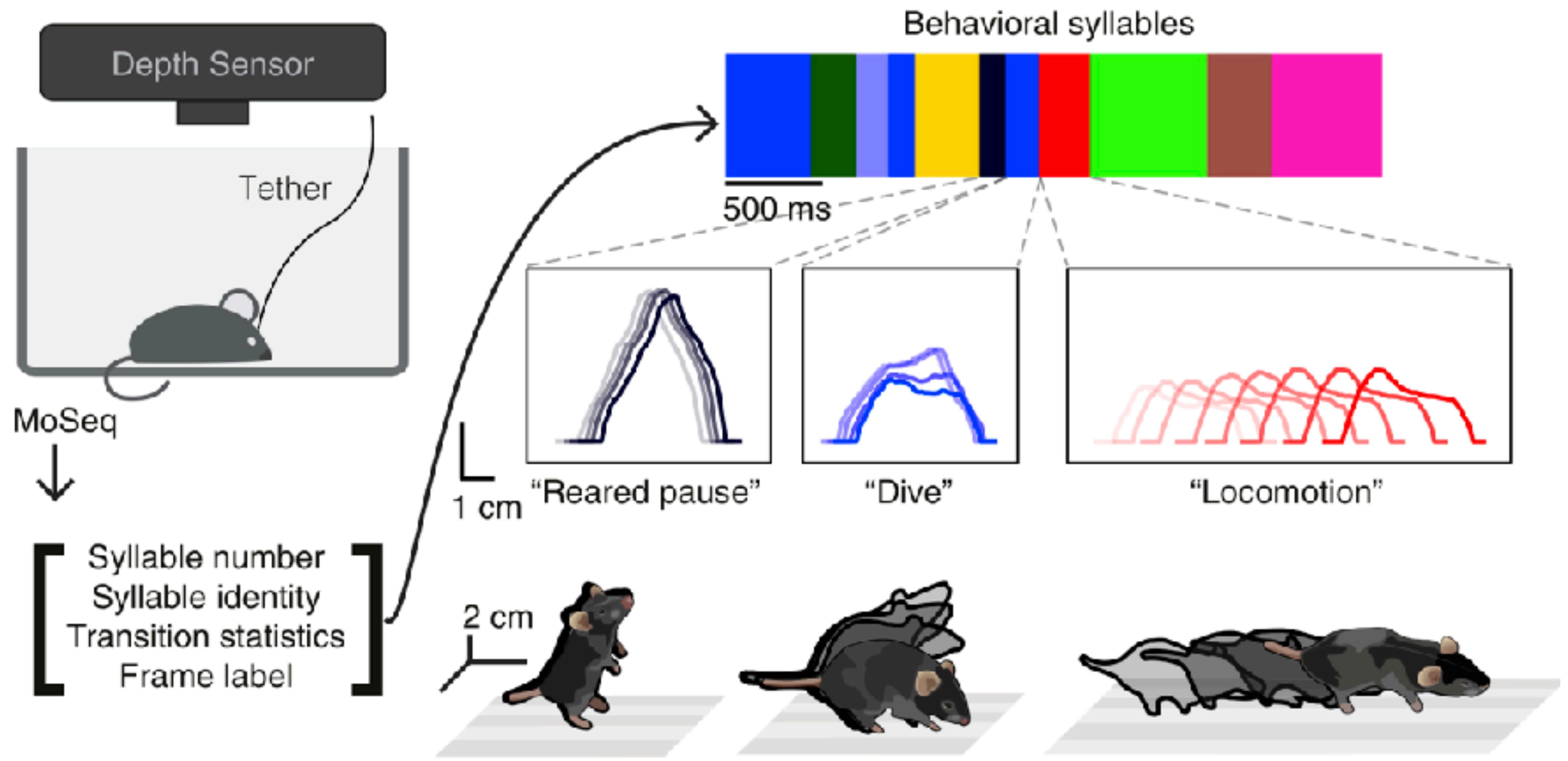


Variability sources:

1 - many syllables / actions

[Wiltschko et al, 2015]
[Markovits et al, 2018]
[Findley et al, 2021]

Complex temporal structure in behavior



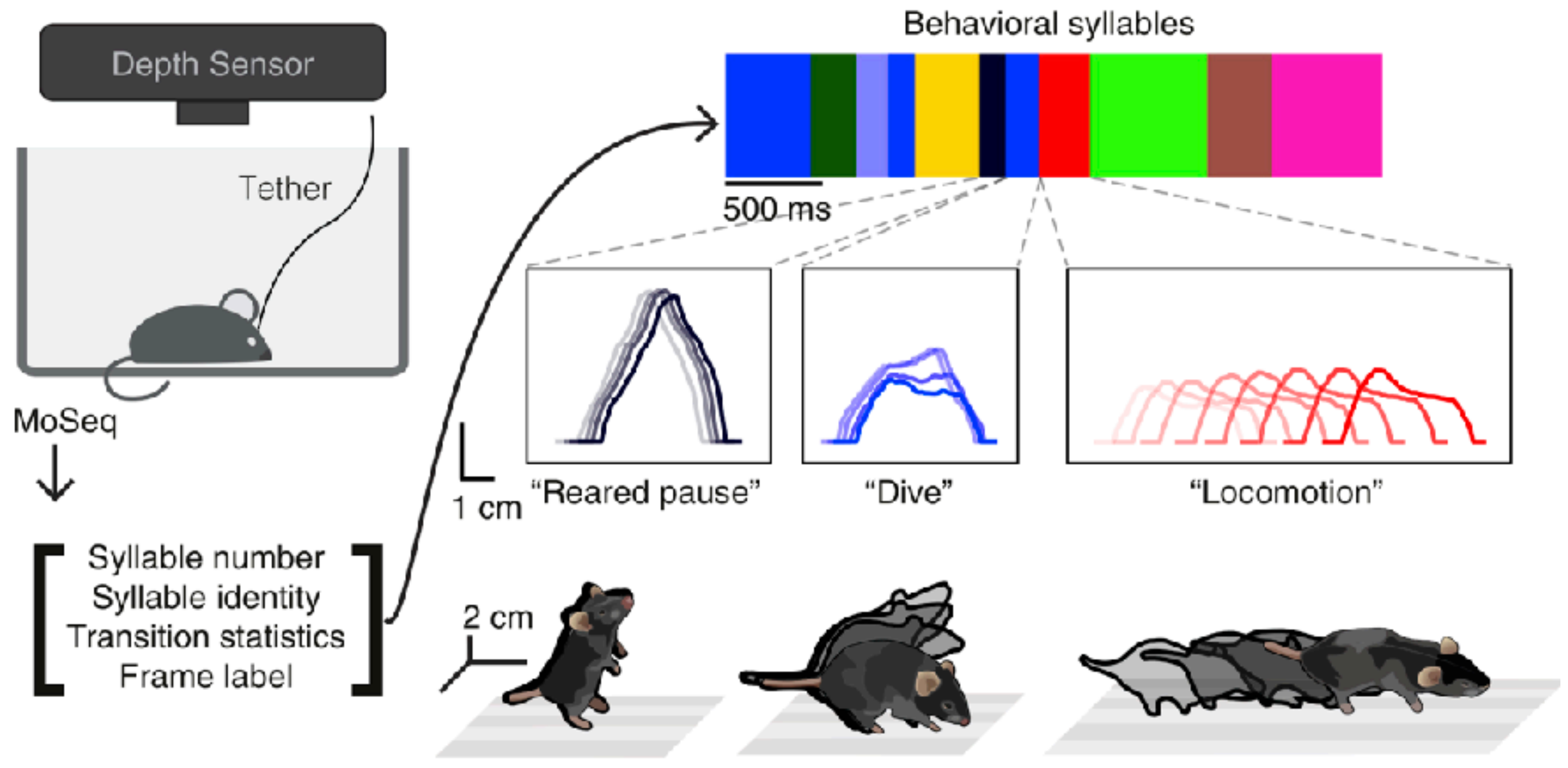
Variability sources:

1 - many syllables / actions

2 - many words / sequences

[Wiltschko et al, 2015]
[Markovits et al, 2018]
[Findley et al, 2021]

Complex temporal structure in behavior



Variability sources:

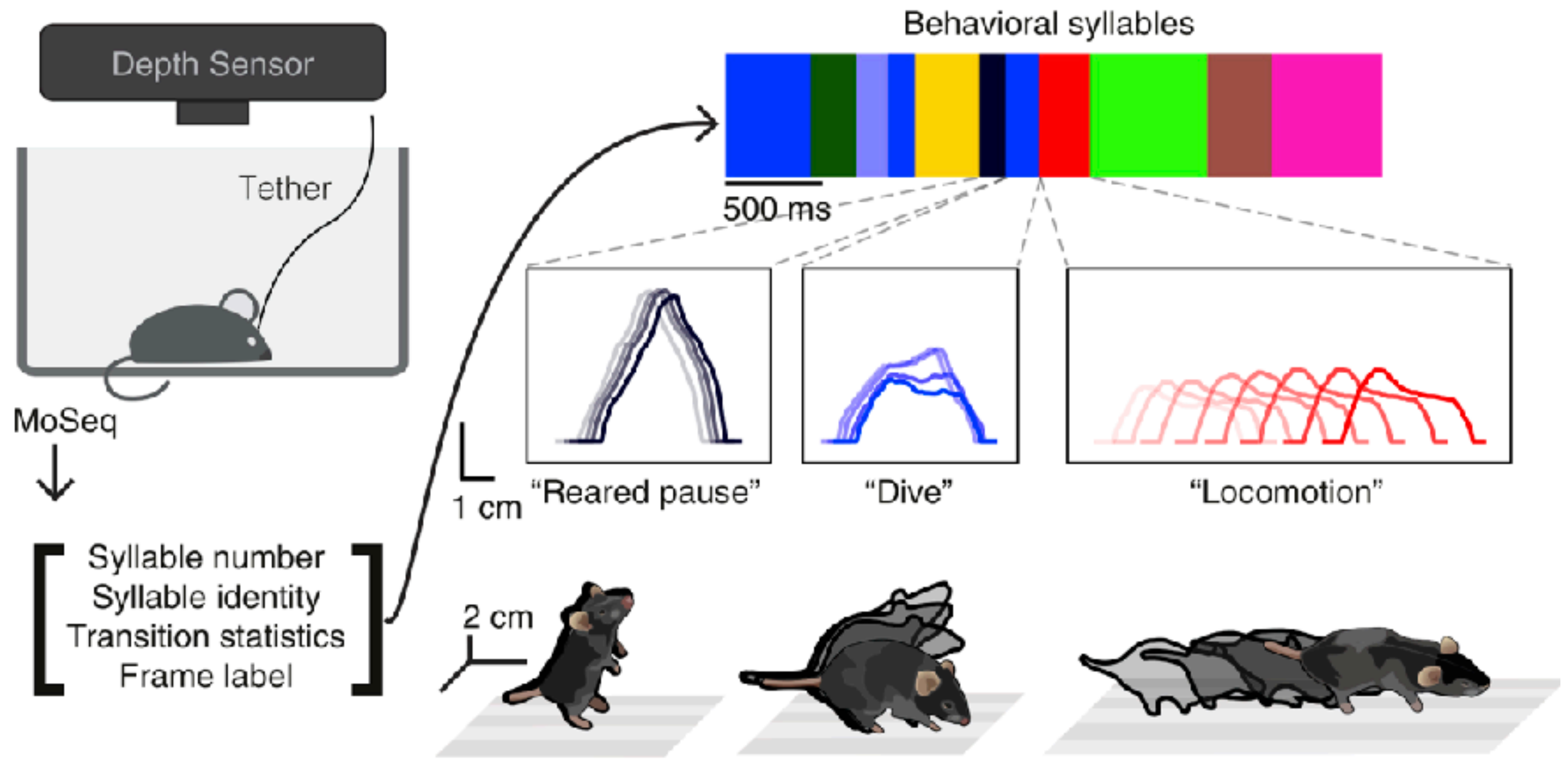
1 - many syllables / actions

2 - many words / sequences

3 - transition times (action duration)

[Wiltschko et al, 2015]
[Markovits et al, 2018]
[Findley et al, 2021]

Complex temporal structure in behavior



Variability sources:

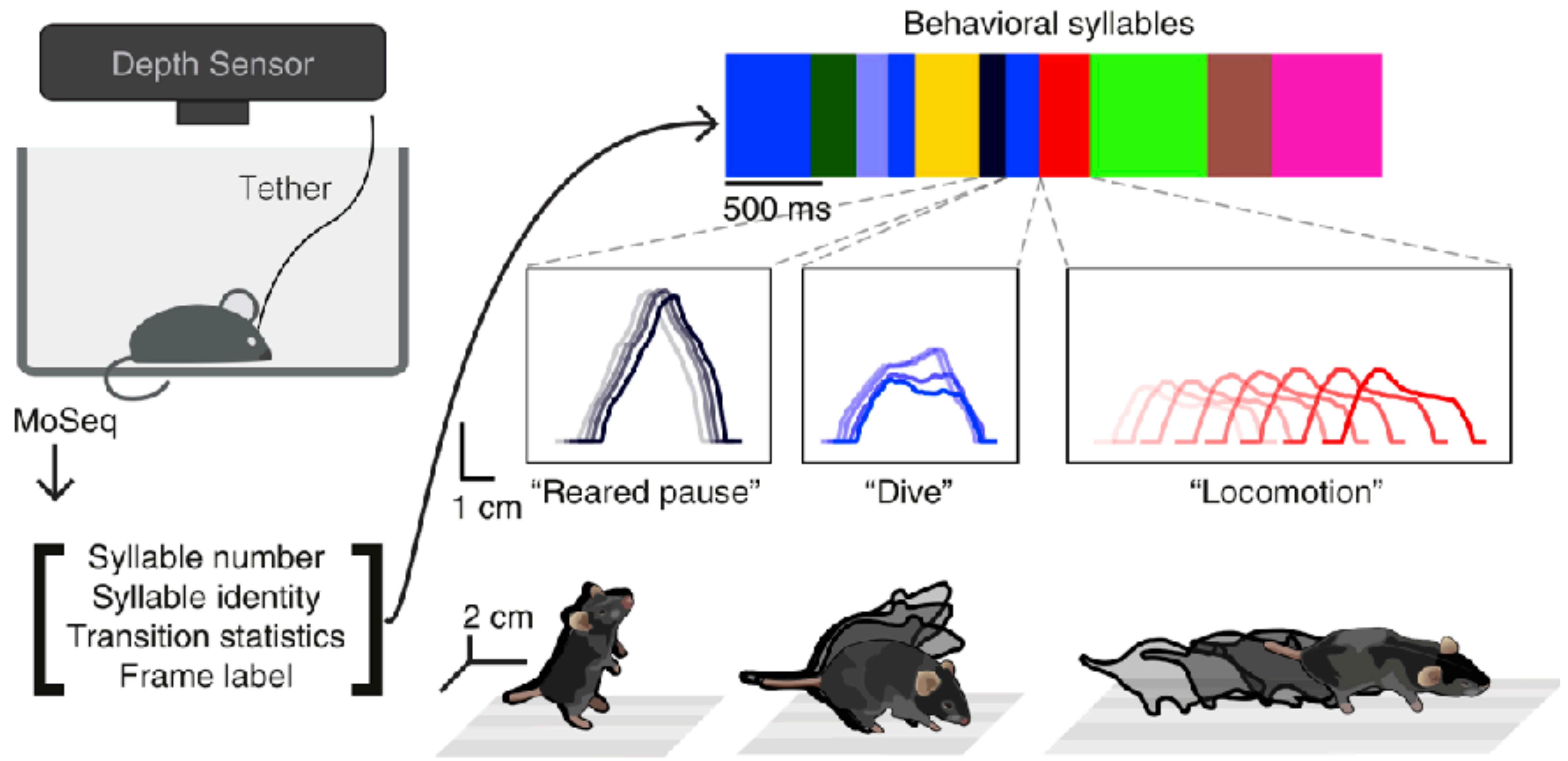
1 - many syllables / actions

2 - many words / sequences

3 - transition times (action duration)

[Wiltschko et al, 2015]
[Markovits et al, 2018]
[Findley et al, 2021]

Complex temporal structure in behavior



Variability sources:

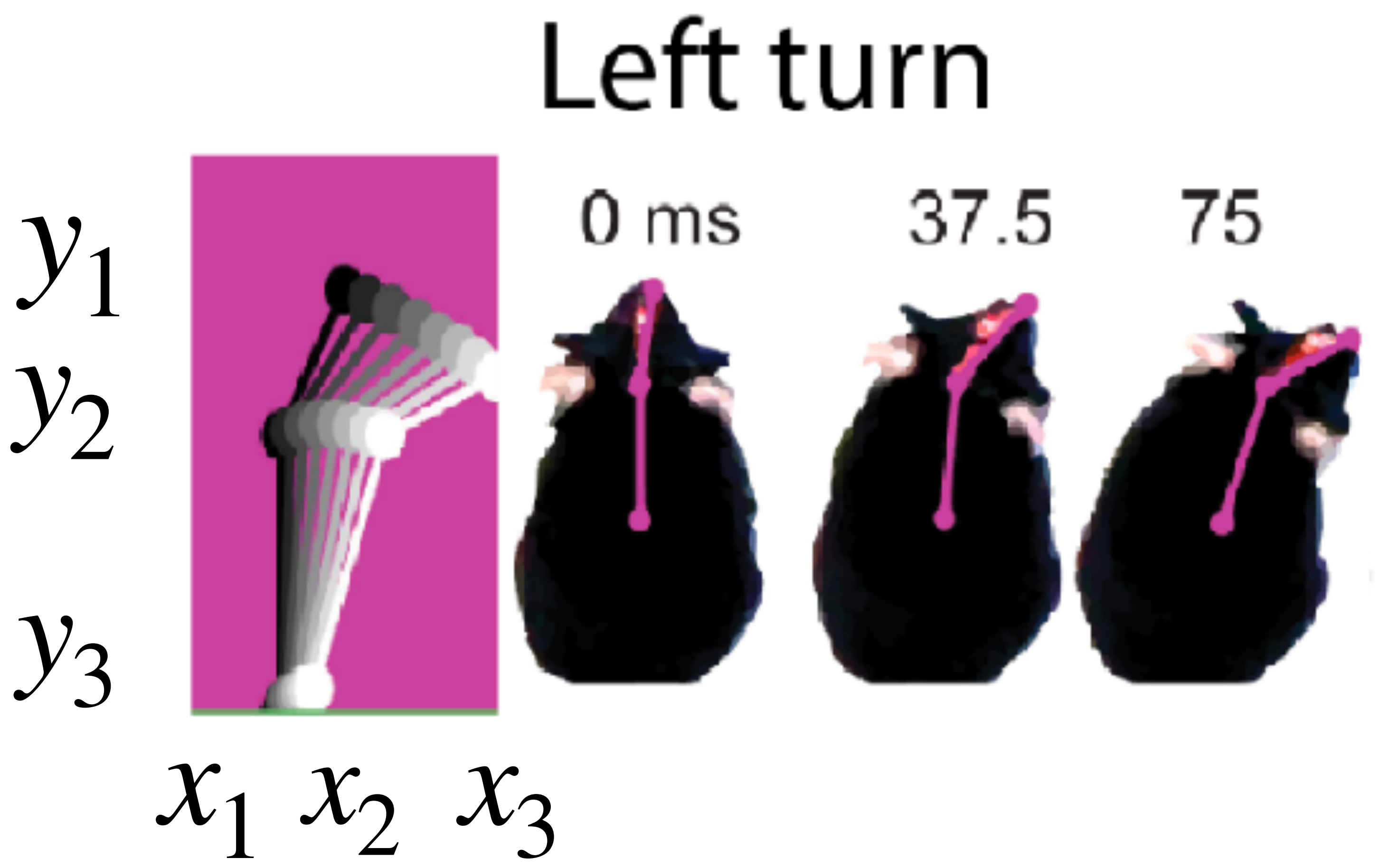
1 - many syllables / actions

2 - many words / sequences

3 - transition times (action duration)

[Wiltschko et al, 2015]
[Markovits et al, 2018]
[Findley et al, 2021]

Autoregressive dynamics



$$\vec{r} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ y_3 \end{pmatrix}$$

Autoregressive dynamics

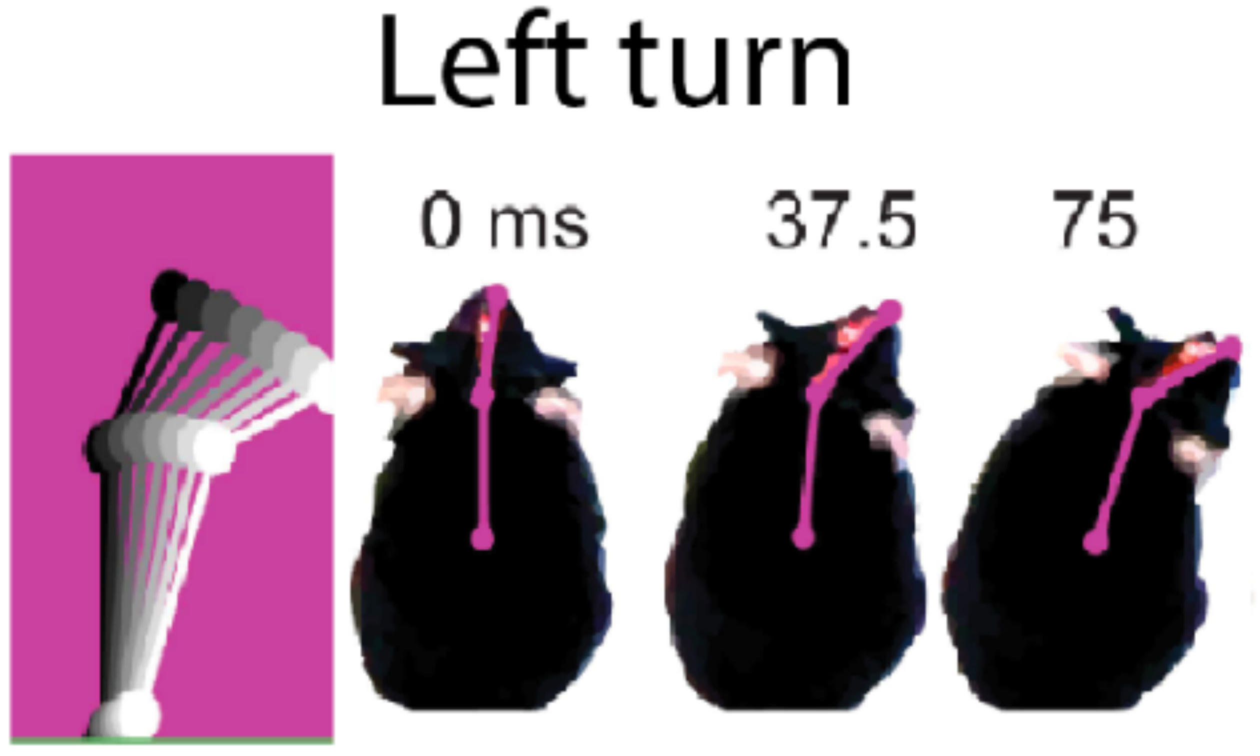
$$\vec{r}_{t+1} = A\vec{r}_t + \vec{b} + \vec{\epsilon}_t$$

A matrix = rotation

b vector = translation

[Wiltschko et al, 2015]
[Markovits et al, 2018]
[Findley et al, 2021]

Autoregressive dynamics



y_1 y_2 y_3

x_1 x_2 x_3

$$\vec{r} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ y_3 \end{pmatrix}$$

Autoregressive dynamics

$$\vec{r}_{t+1} = A\vec{r}_t + \vec{b} + \vec{\epsilon}_t$$

noise

A matrix = rotation

b vector = translation

$$p(\vec{r}_{t+1}) = N(A\vec{r}_t + \vec{b}, \Sigma)$$

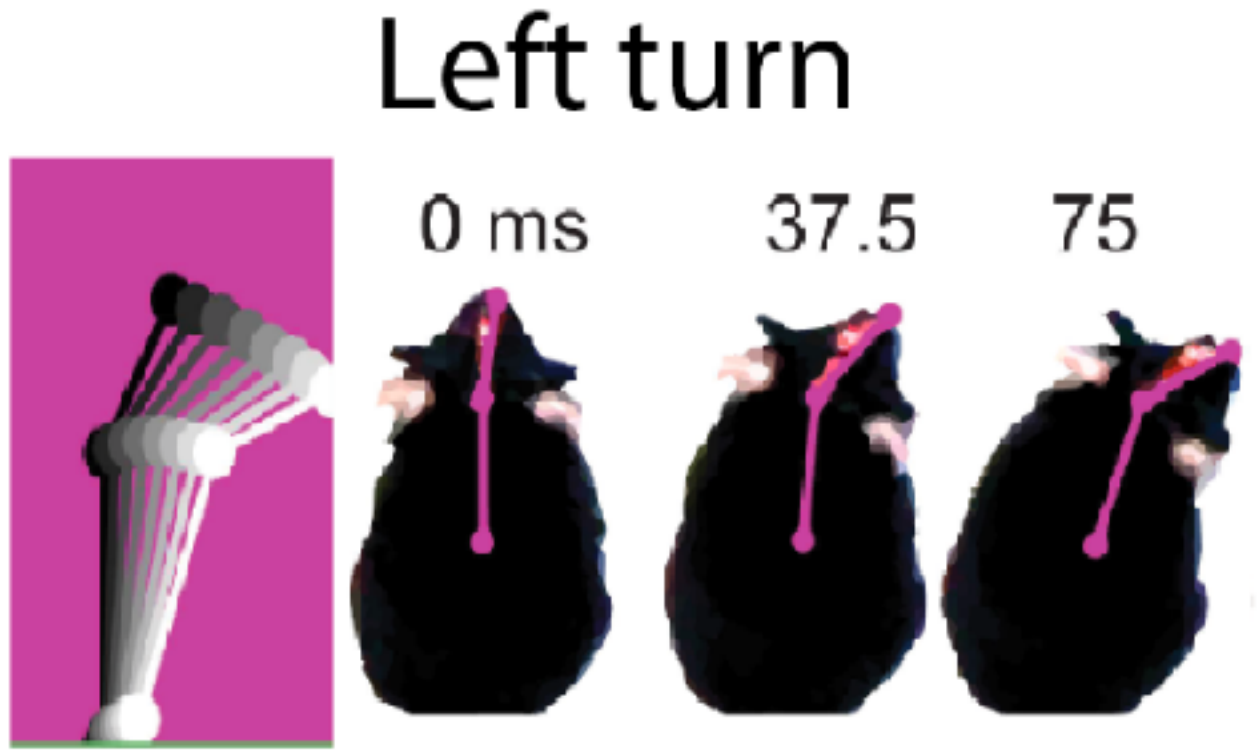
[Wiltschko et al, 2015]
[Markovits et al, 2018]
[Findley et al, 2021]

Autoregressive dynamics

y_1

y_2

y_3



$x_1 \; x_2 \; x_3$

$\Downarrow$

$$\vec{r} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ y_3 \end{pmatrix}$$

Autoregressive dynamics

noise

$$\vec{r}_{t+1} = A\vec{r}_t + \vec{b} + \vec{\epsilon}_t$$

A matrix = rotation

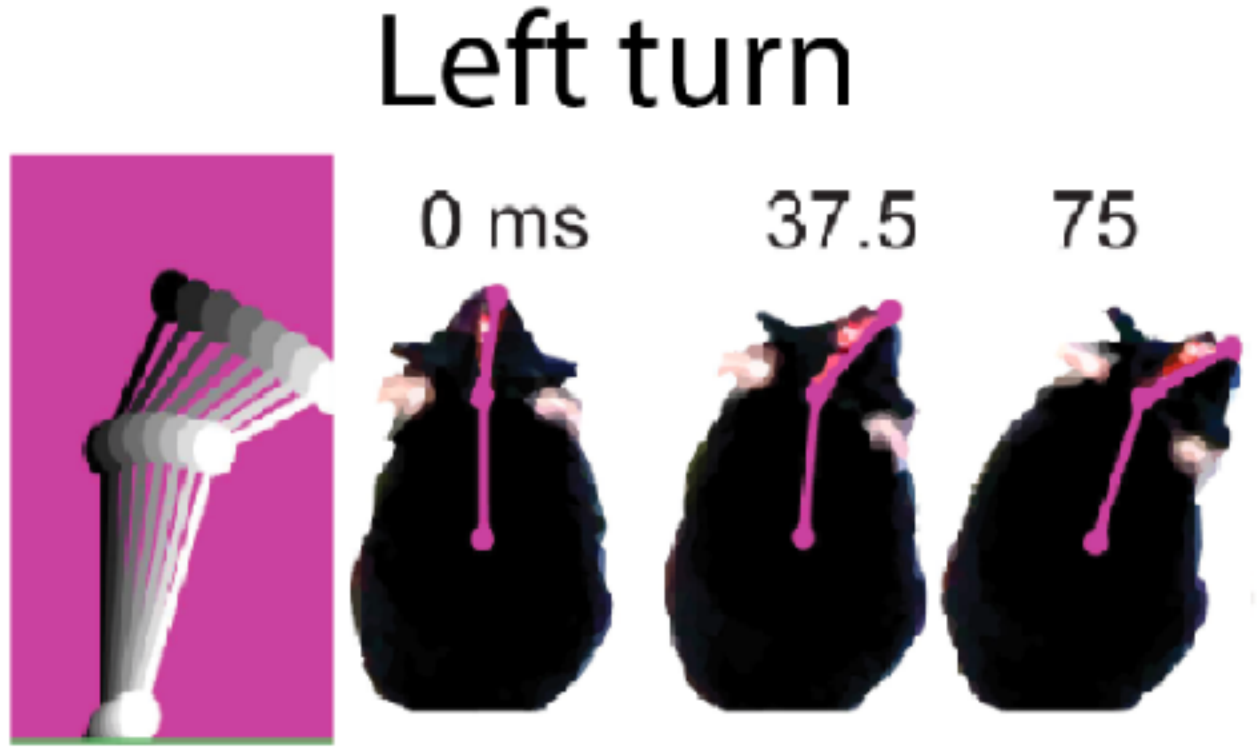
b vector = translation

$$p(\vec{r}_{t+1}) = N(A\vec{r}_t + \vec{b}, \Sigma)$$

$$\to r_{t-1} \to r_t \to r_{t+1} \to$$

[Wiltschko et al, 2015]
[Markovits et al, 2018]
[Findley et al, 2021]

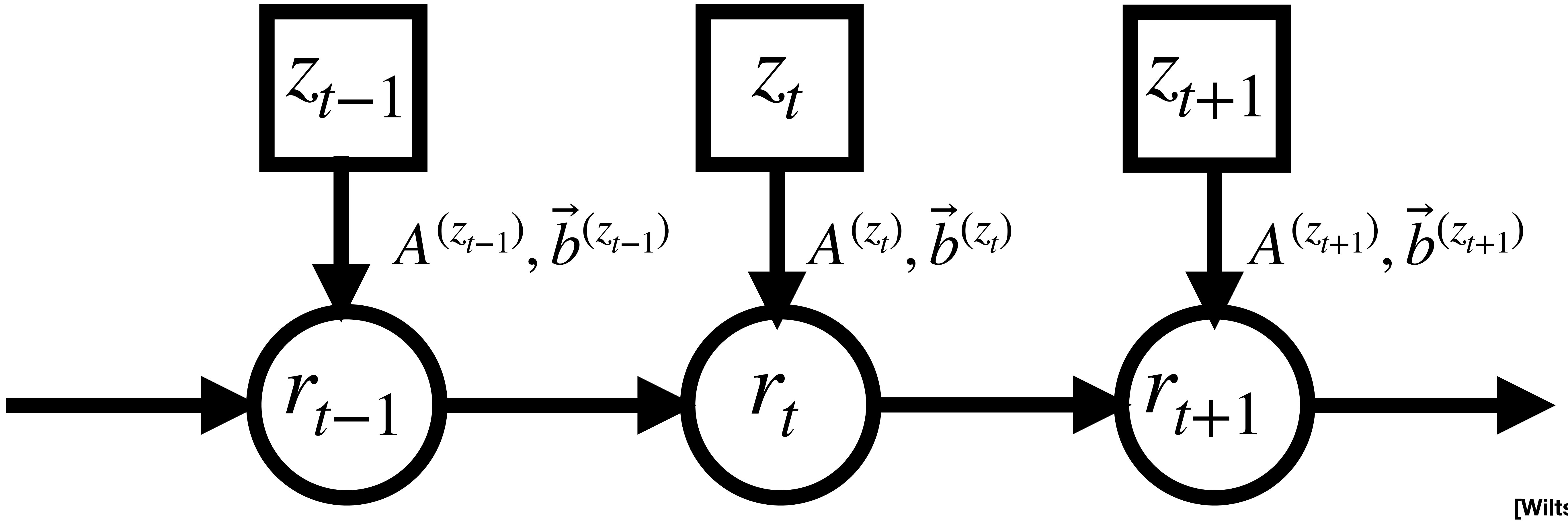
Autoregressive dynamics



Autoregressive dynamics

$$\vec{r}_{t+1} = A\vec{r}_t + \vec{b} + \vec{\epsilon}_t$$

$z_t \in 1,\ldots,K$ Hidden states (motifs)



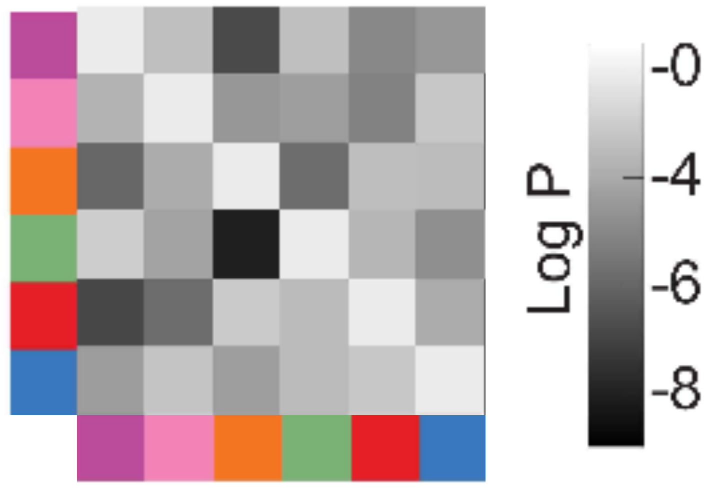
[Wiltschko et al, 2015]
[Markovits et al, 2018]
[Findley et al, 2021]

AR Hidden Markov model

Transition probabilities

$$p(z_t = k \,|\, z_{t-1} = l)$$

z_{t-1}

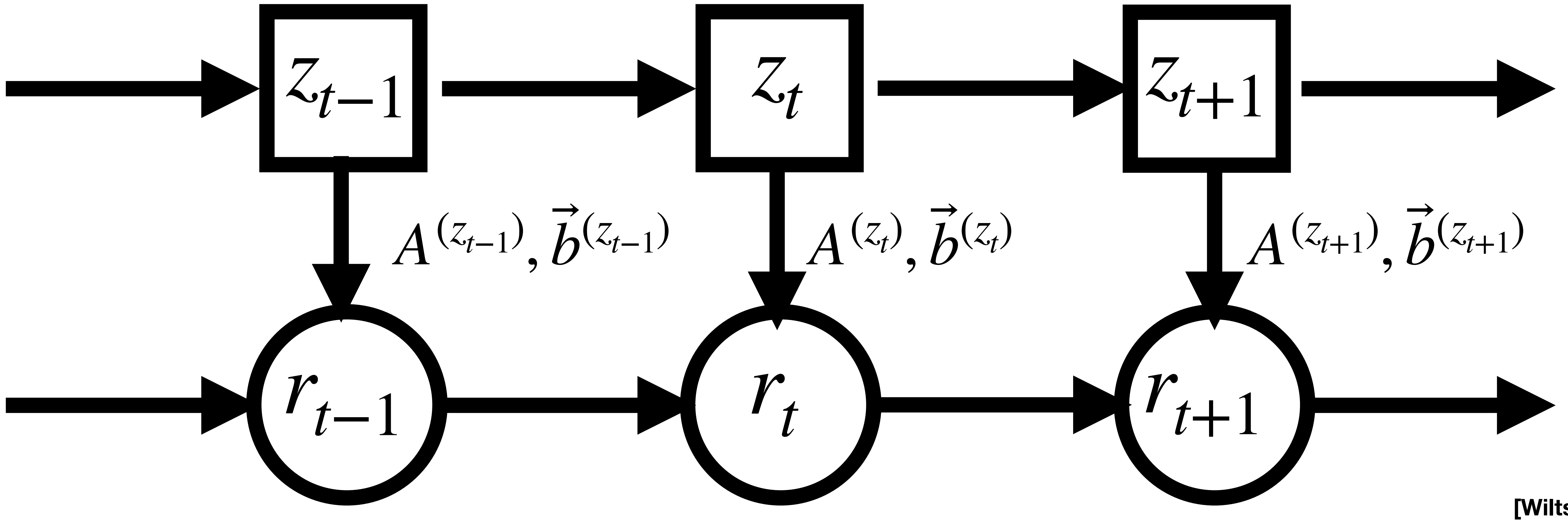


z_t

Autoregressive dynamics

$$\vec{r}_{t+1} = A\vec{r}_t + \vec{b} + \vec{\epsilon}_t$$

$z_t \in 1,\ldots,K$ Hidden states (motifs)

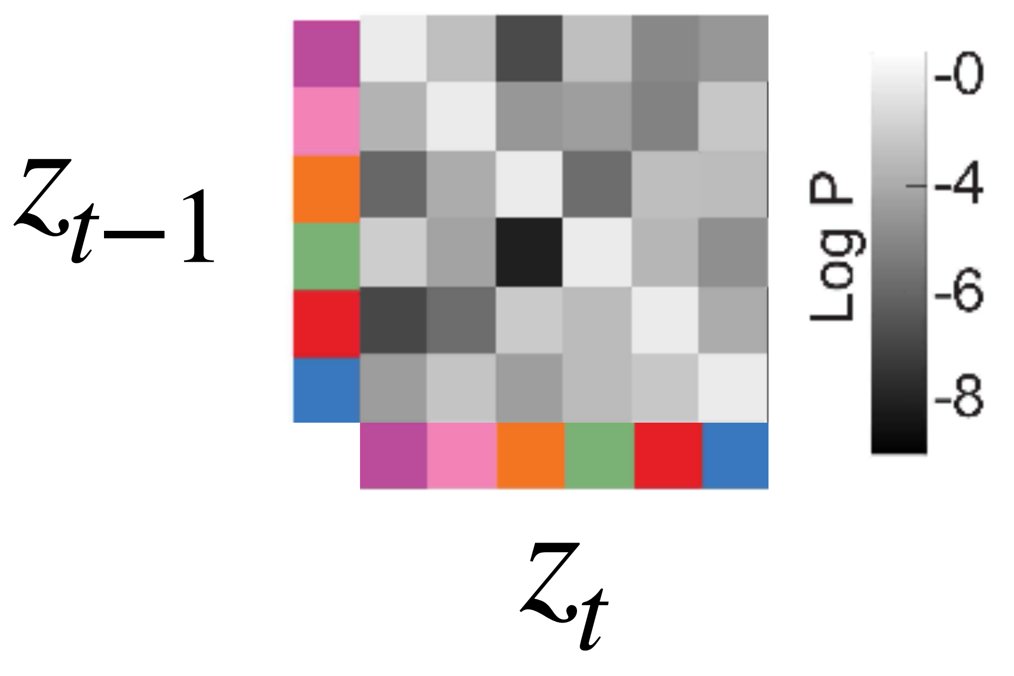


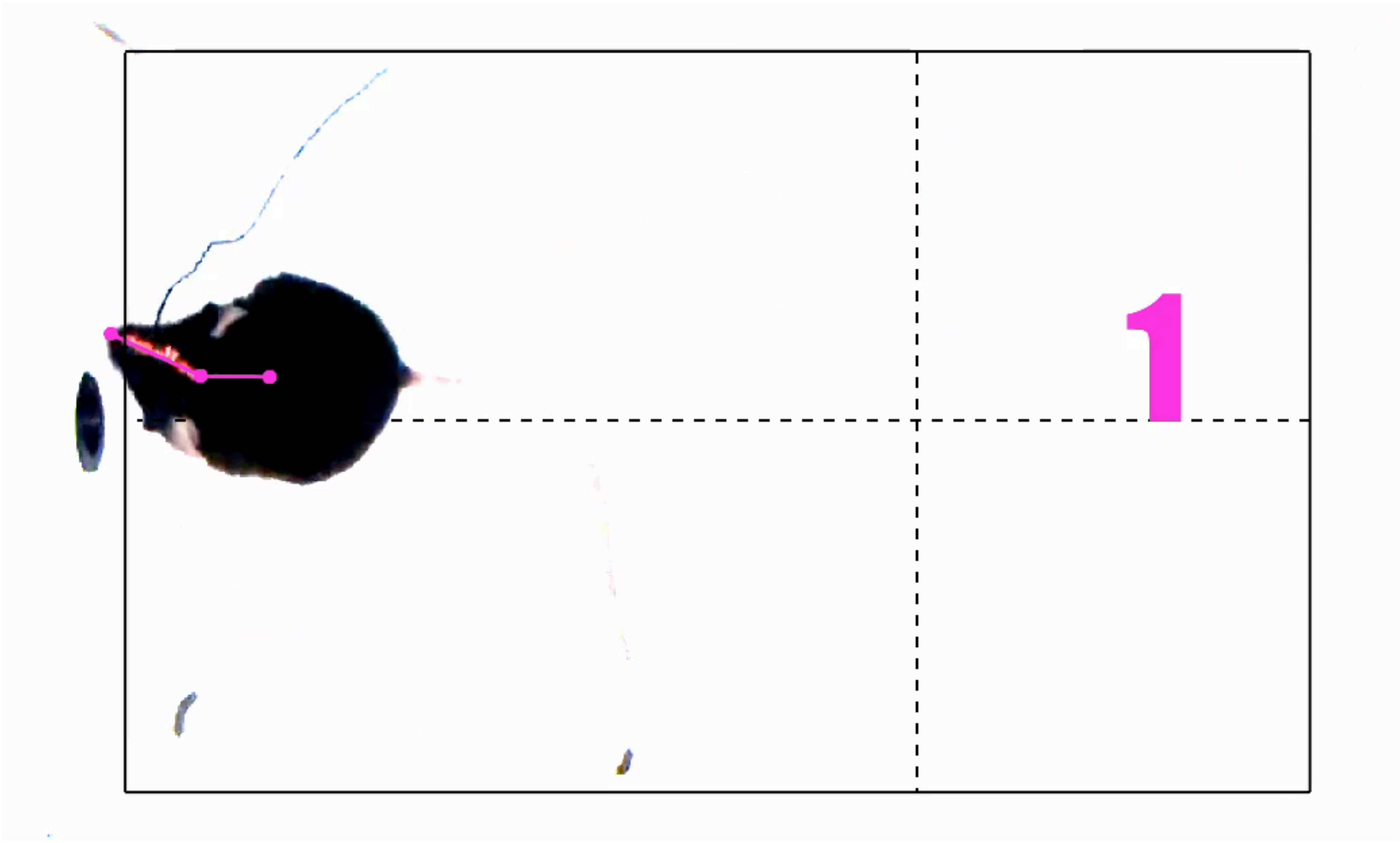
[Wiltschko et al, 2015]
[Markovits et al, 2018]
[Findley et al, 2021]

AR Hidden Markov model

Transition probabilities

$$p(z_t = k \,|\, z_{t-1} = l)$$



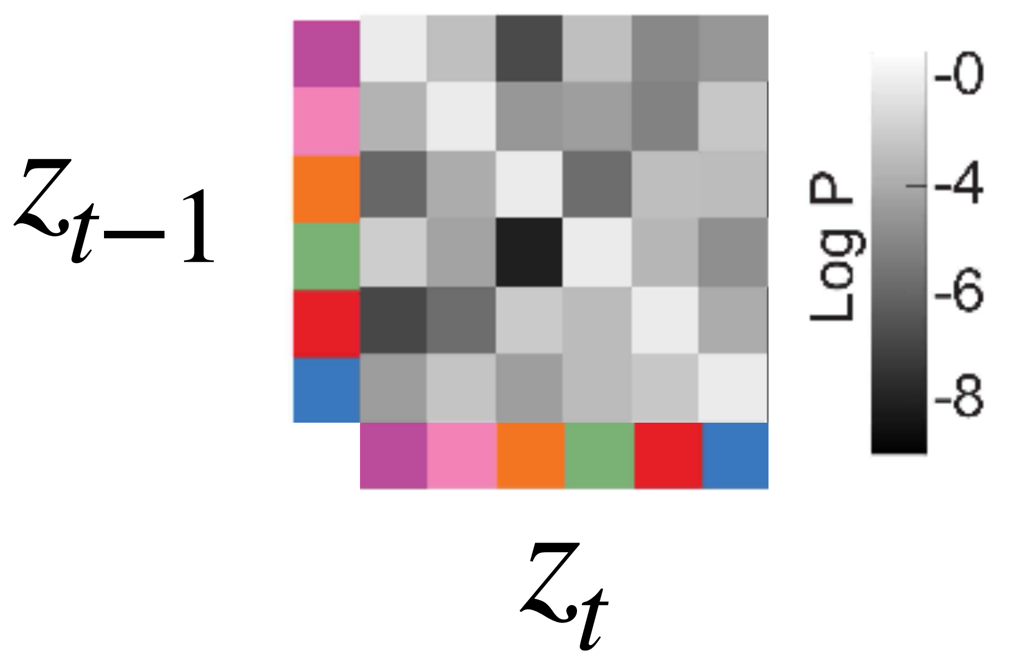


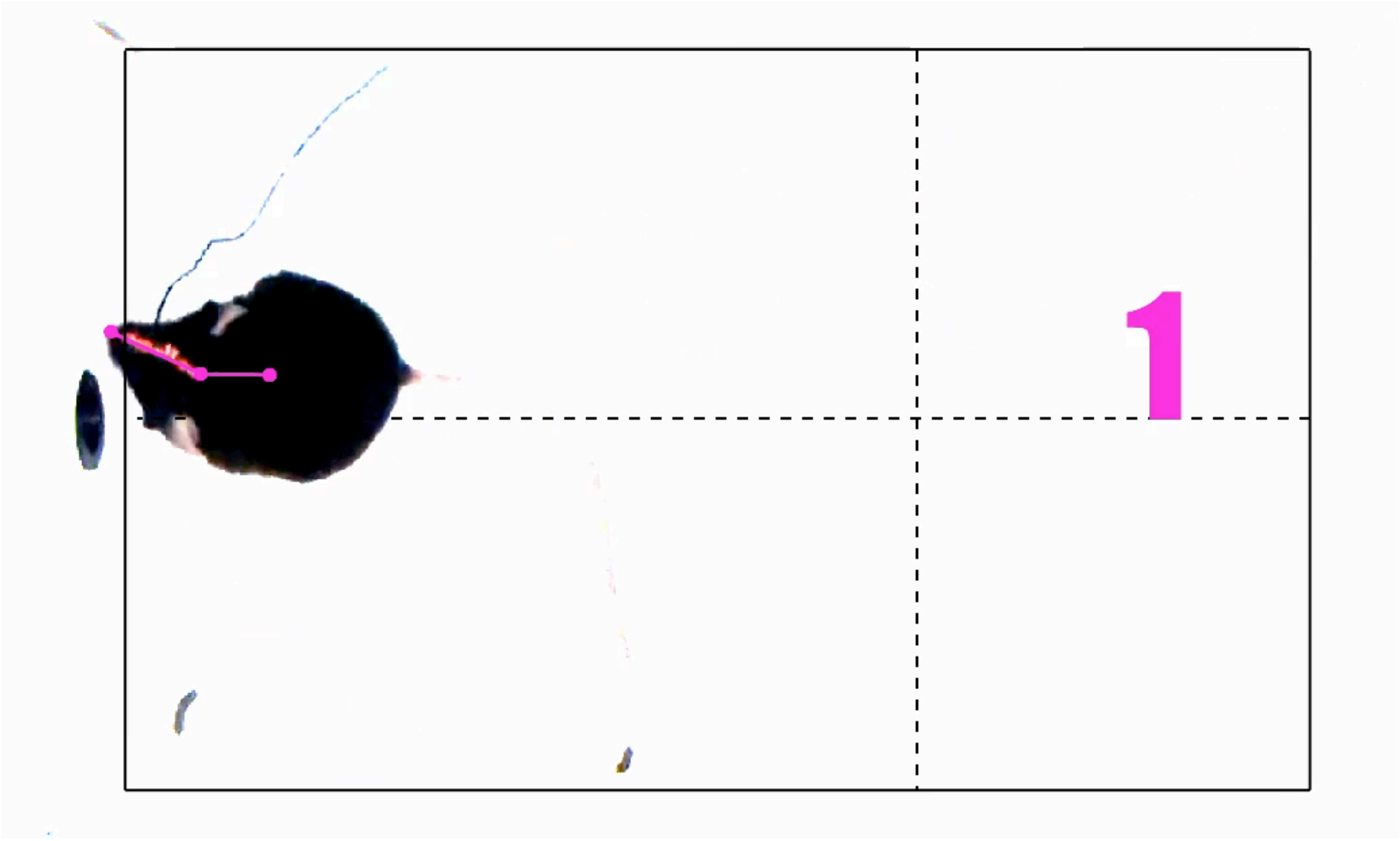
[Wiltschko et al, 2015]
[Markovits et al, 2018]
[Findley et al, 2021]

AR Hidden Markov model

Transition probabilities

$$p(z_t = k \mid z_{t-1} = l)$$





[Wiltschko et al, 2015]
[Markovits et al, 2018]
[Findley et al, 2021]

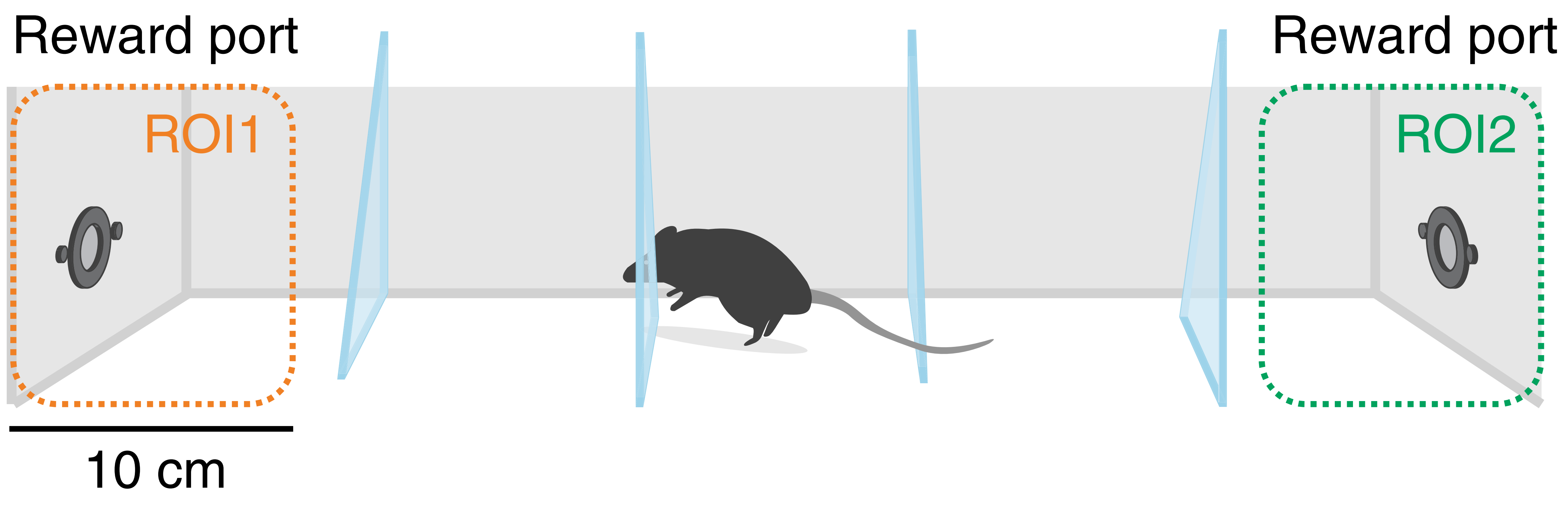
References

HMM: Theory

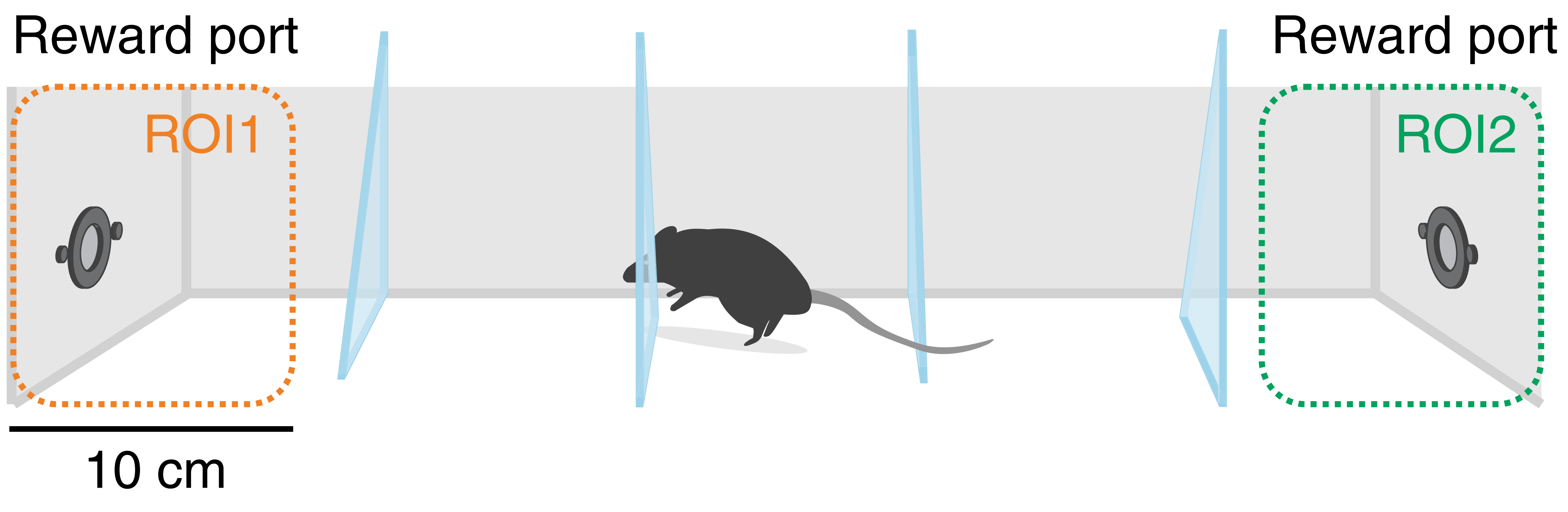
Applications to freely moving behavior

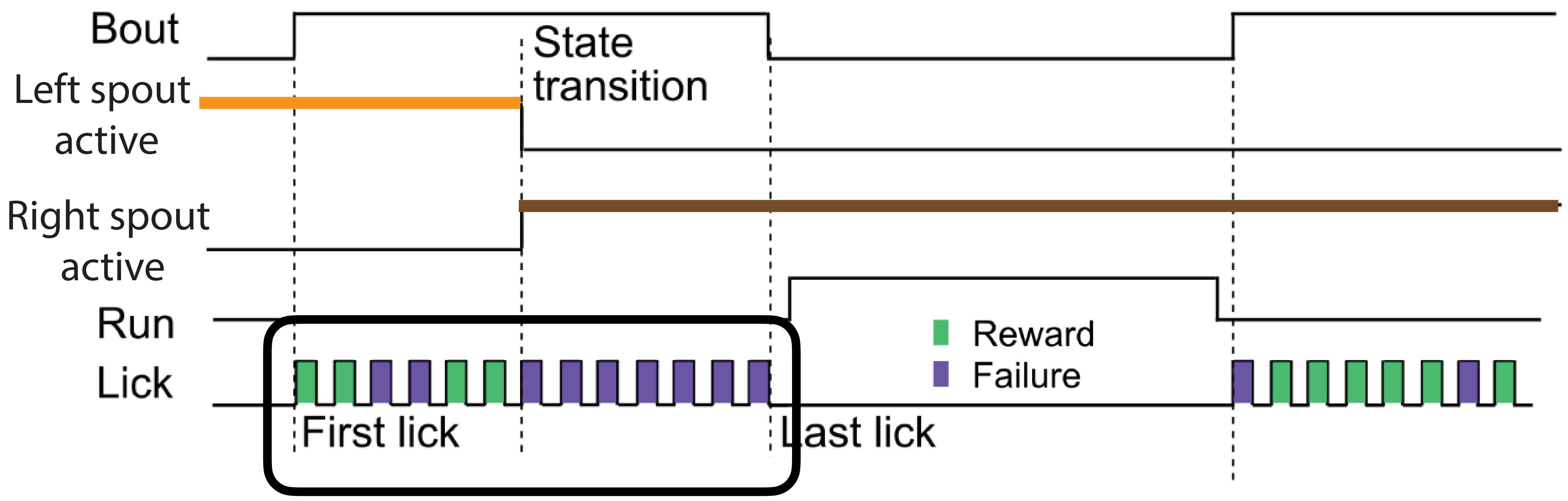
- AR-HMM:
  - Moseq *[Wiltschko et al, 2015, Markovits et al, 2018][Findley et al, 2021]*
- Switching Linear Dynamical Systems (SLDS)
  - Keypoint Moseq *[Weinreb et al, 2023]*

Naturalistic foraging



Naturalistic foraging



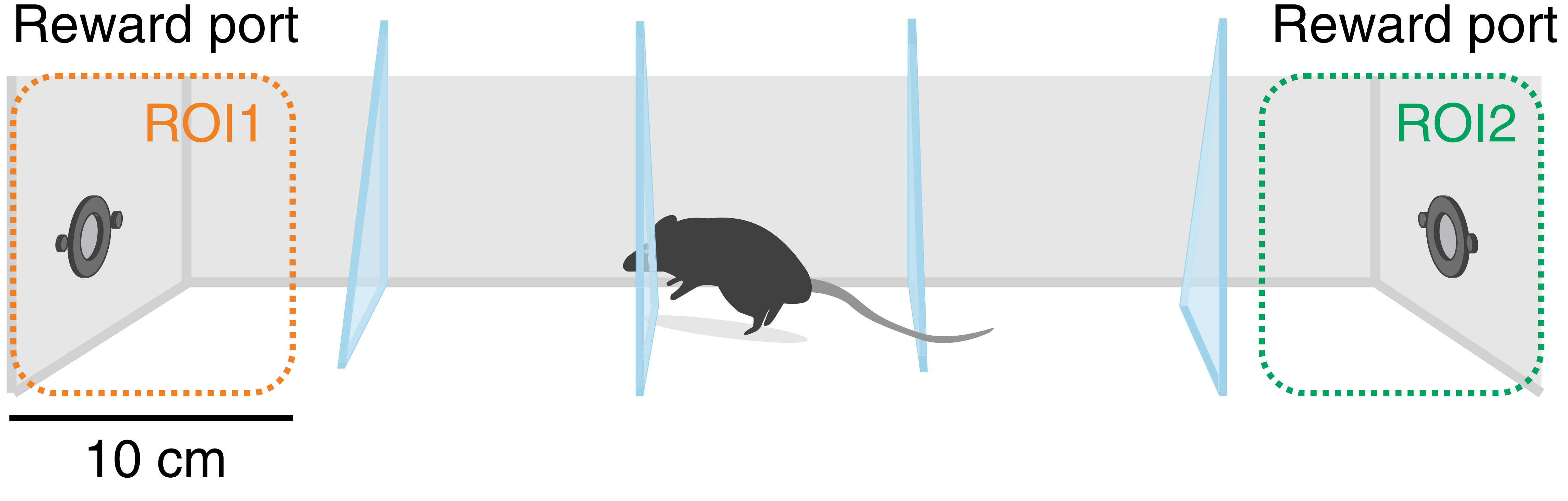


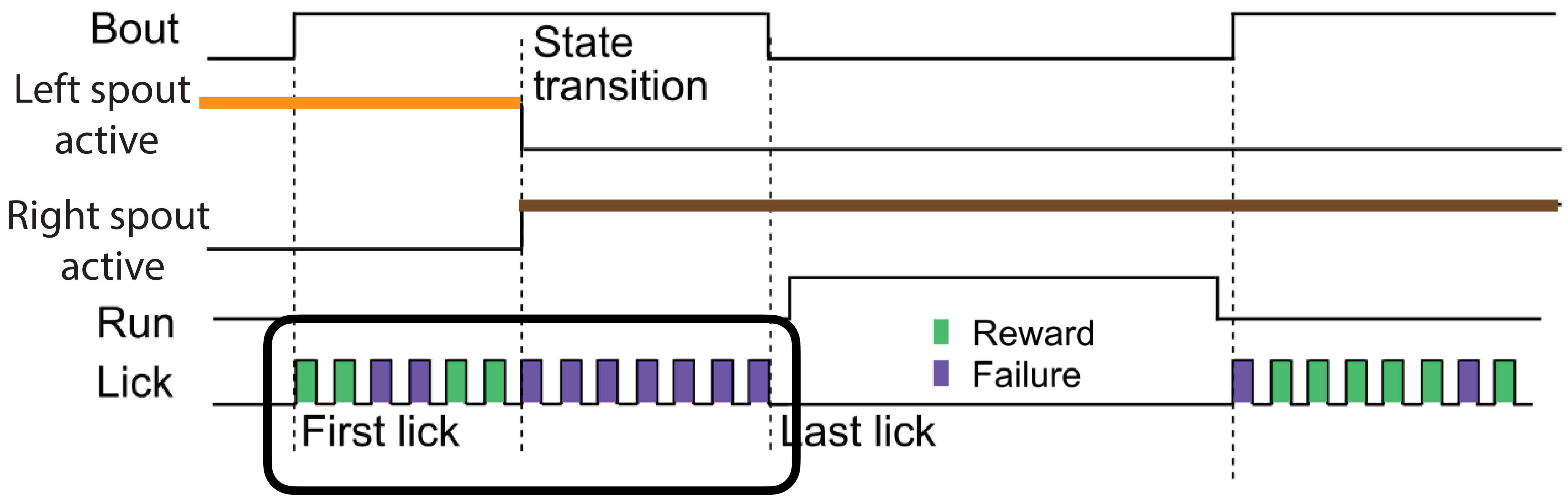
Rewarded lick —> port is active —> stay

Failures. —> prob release?

port inactive?

Naturalistic foraging

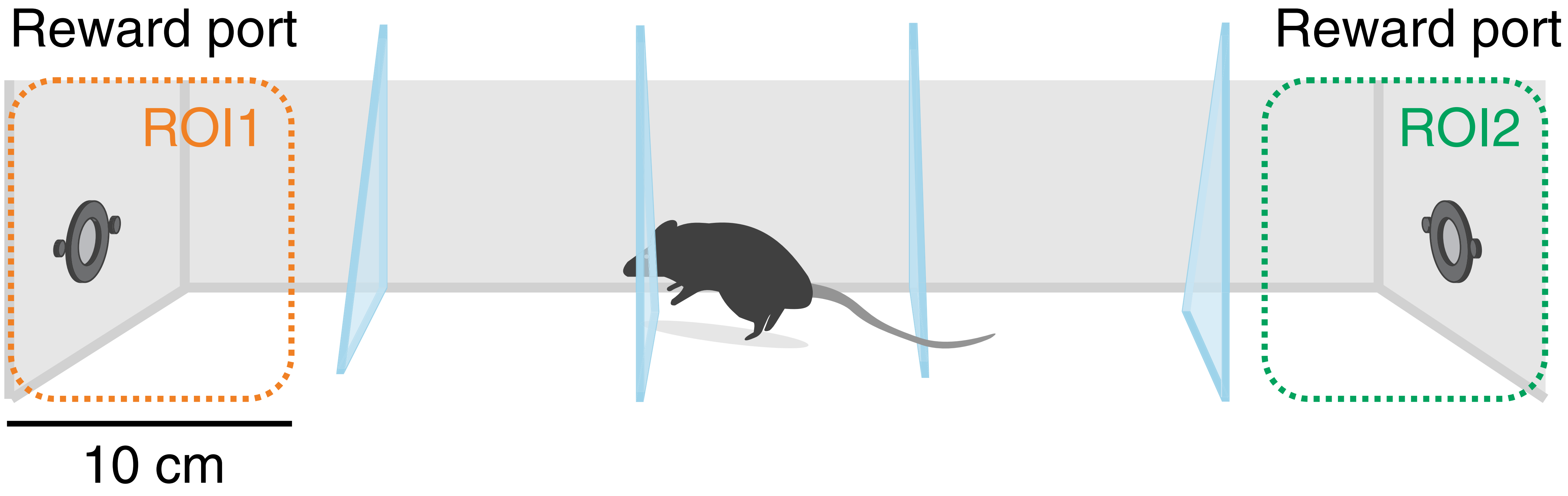


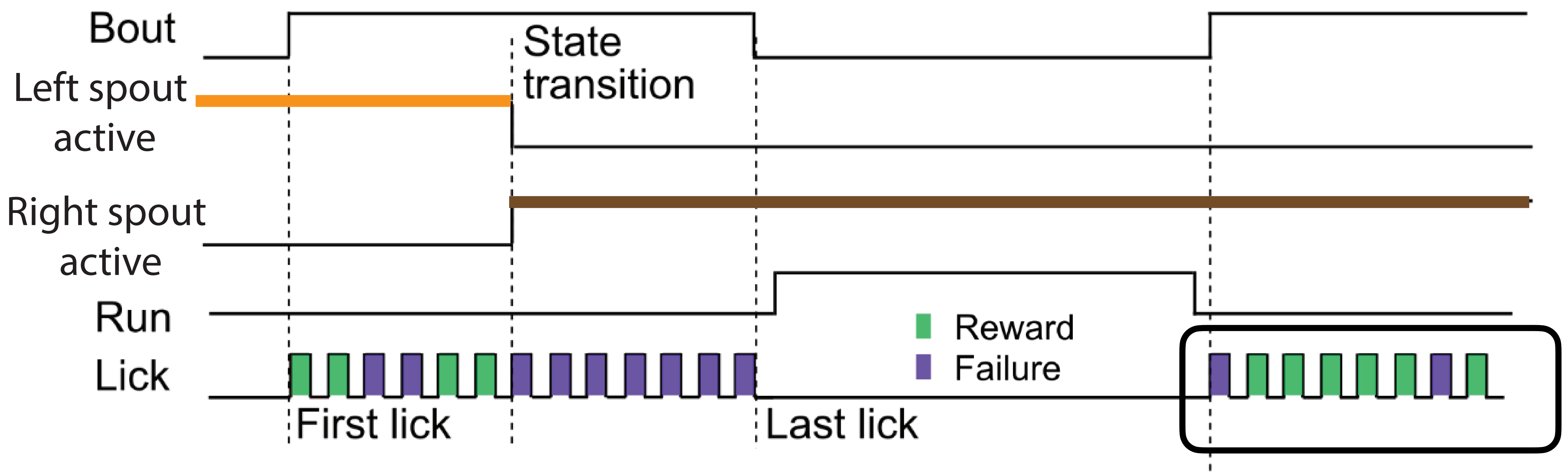


Rewarded lick —> port is active —> stay
Failures. —> prob release?
port inactive?

Integrate Rew & Fail, but how?

Naturalistic foraging



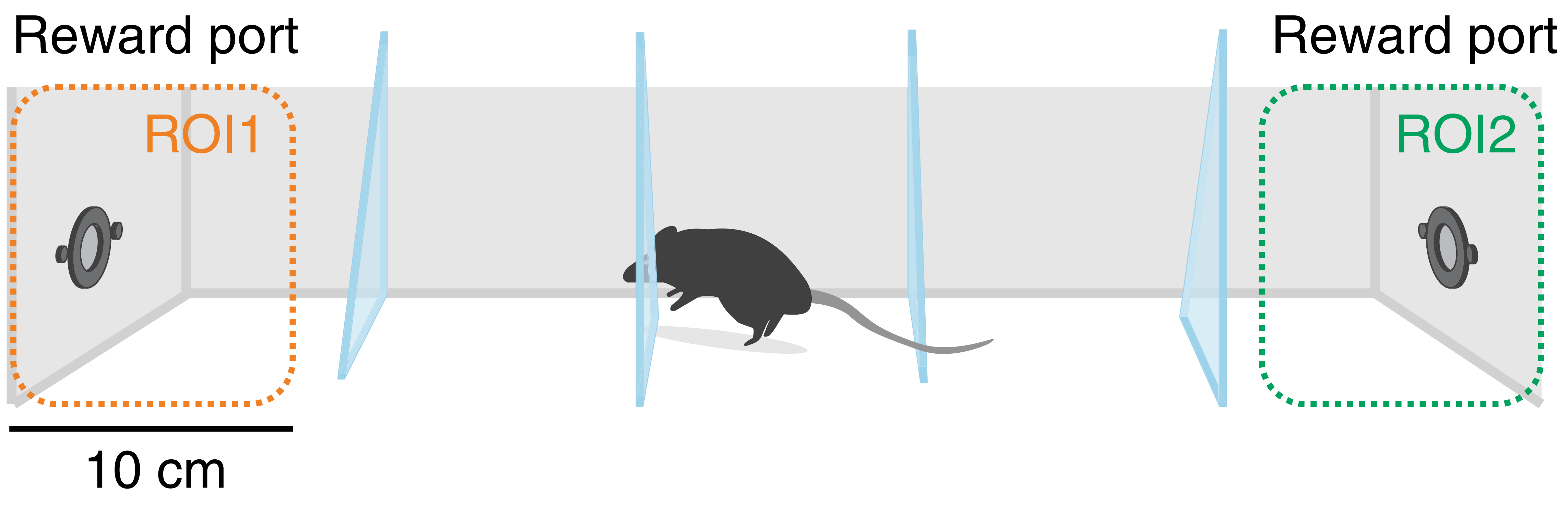


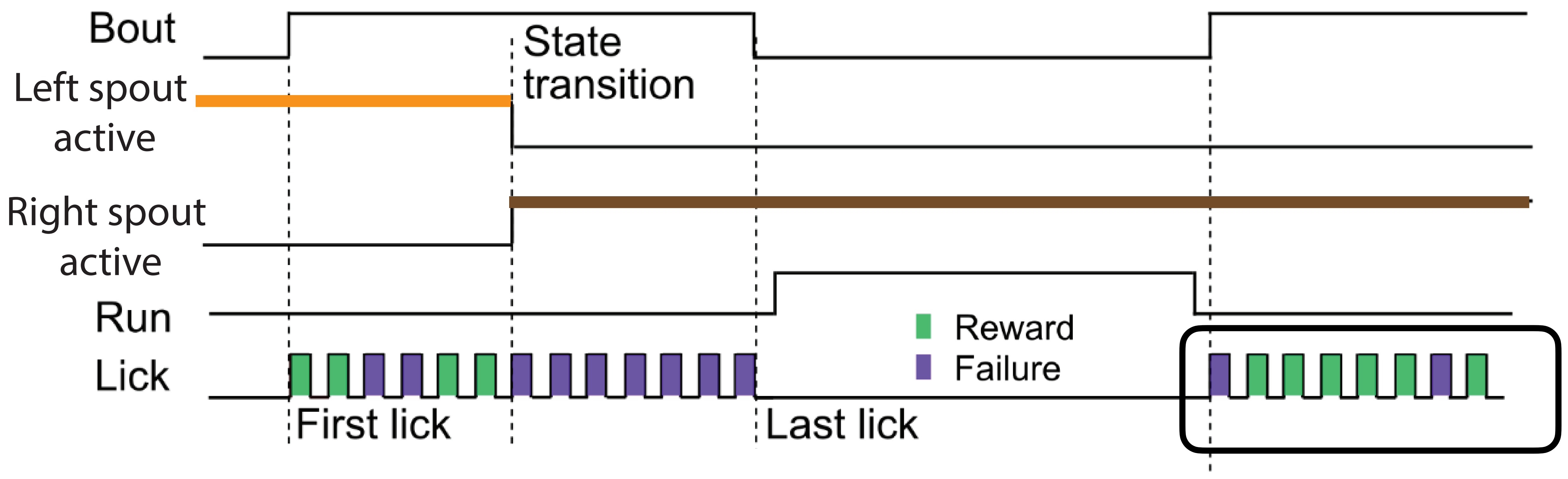
Rewarded lick —> port is active —> stay

Failures. —> prob release?
port inactive?

Integrate Rew & Fail, but how?

Naturalistic foraging





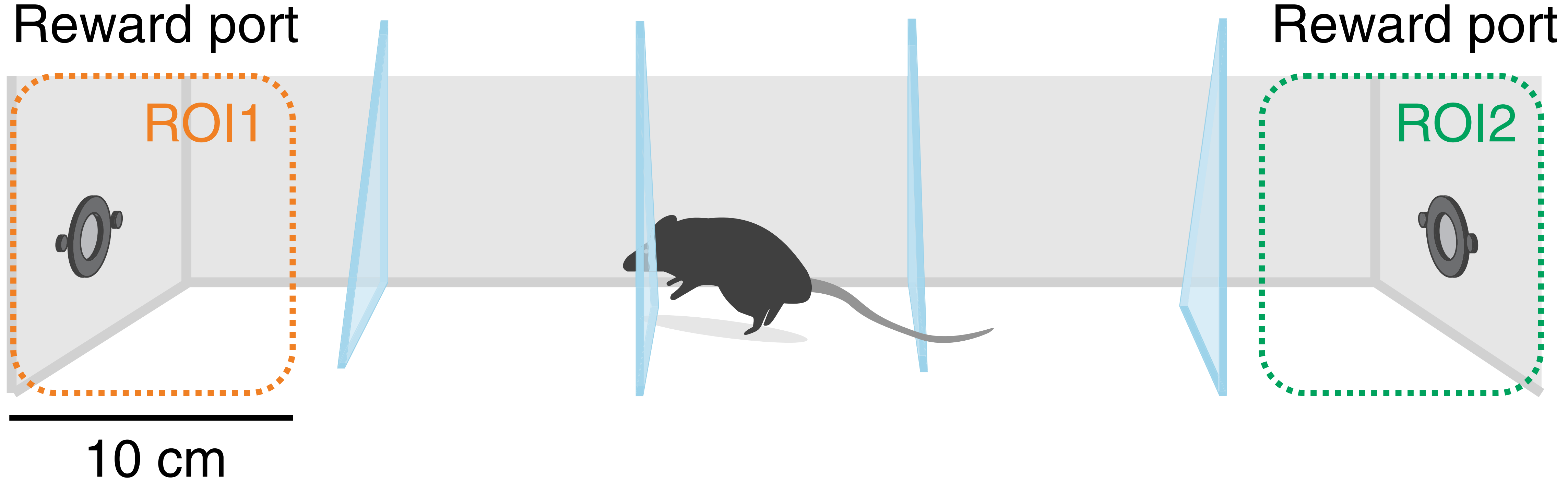
Rewarded lick —> port is active —> stay

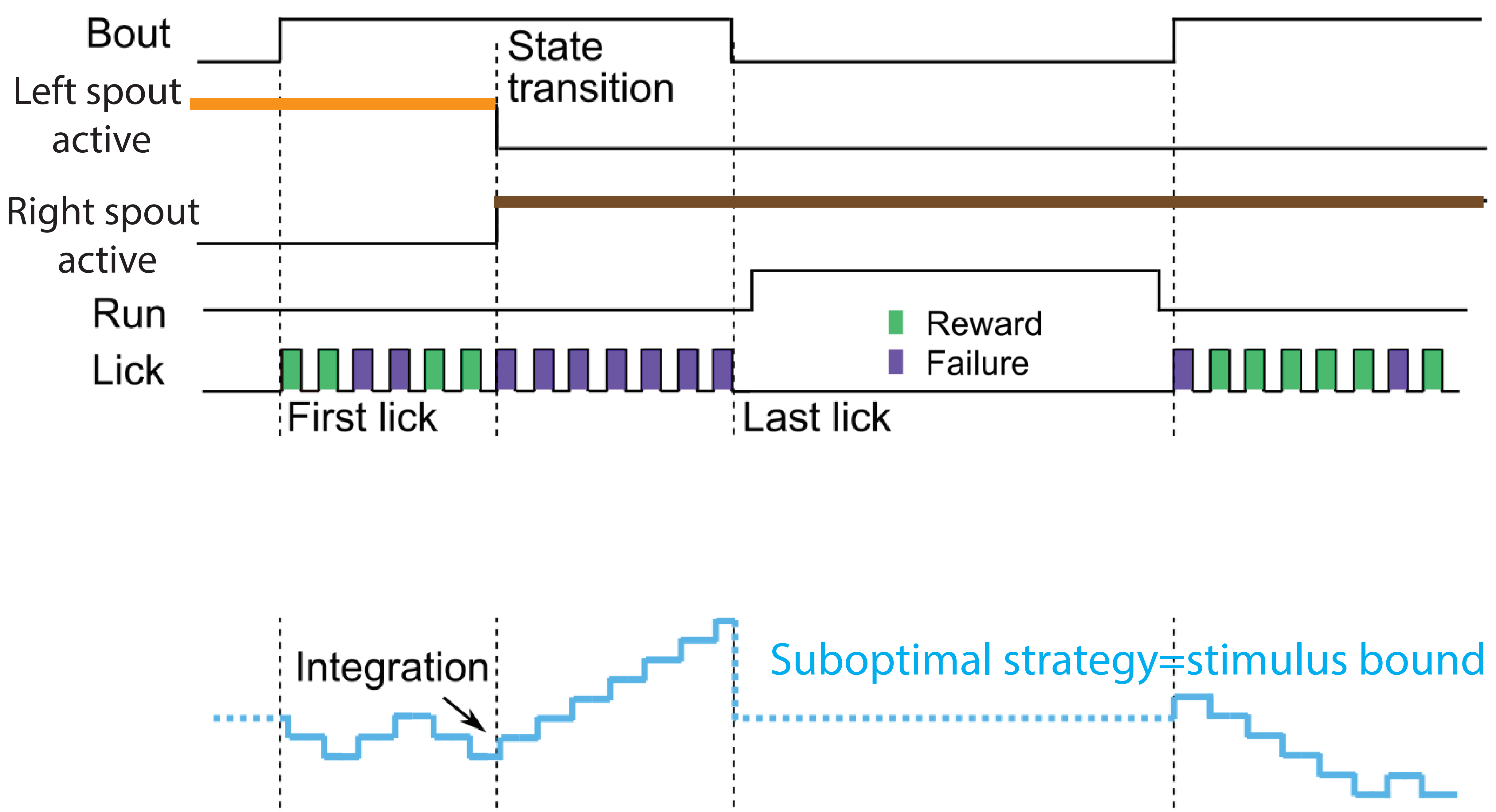
Failures. —> prob release?
port inactive?

Integrate Rew & Fail, but how?

…Alternative strategies for evidence accumulation

Naturalistic foraging



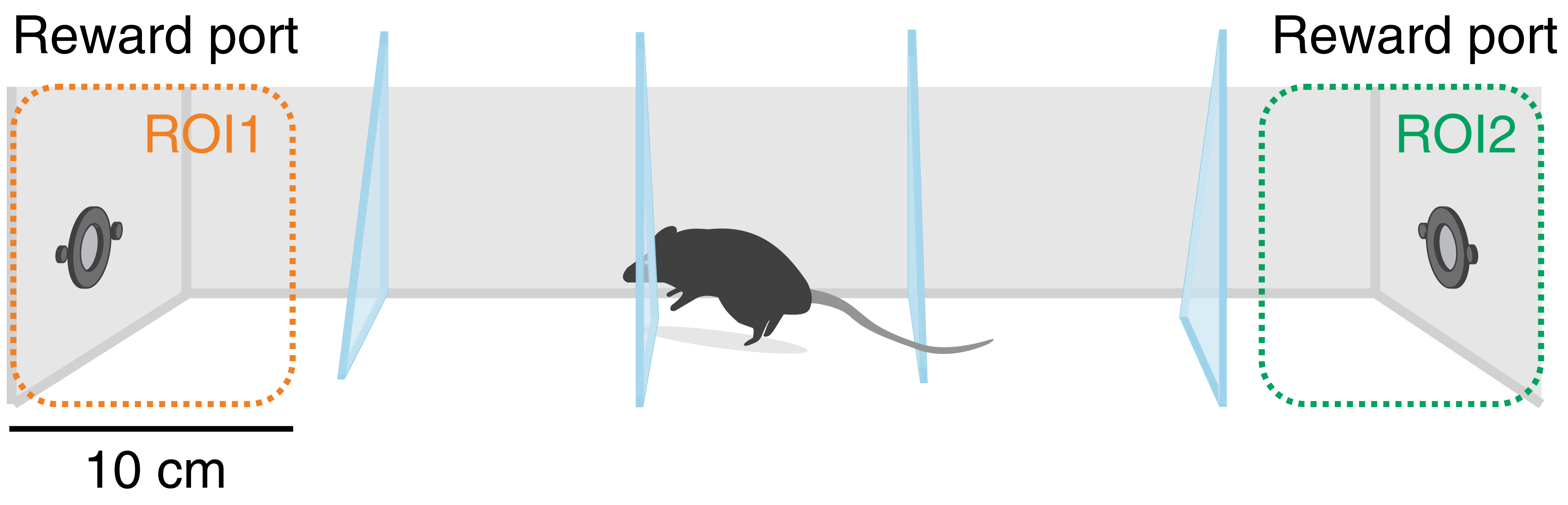


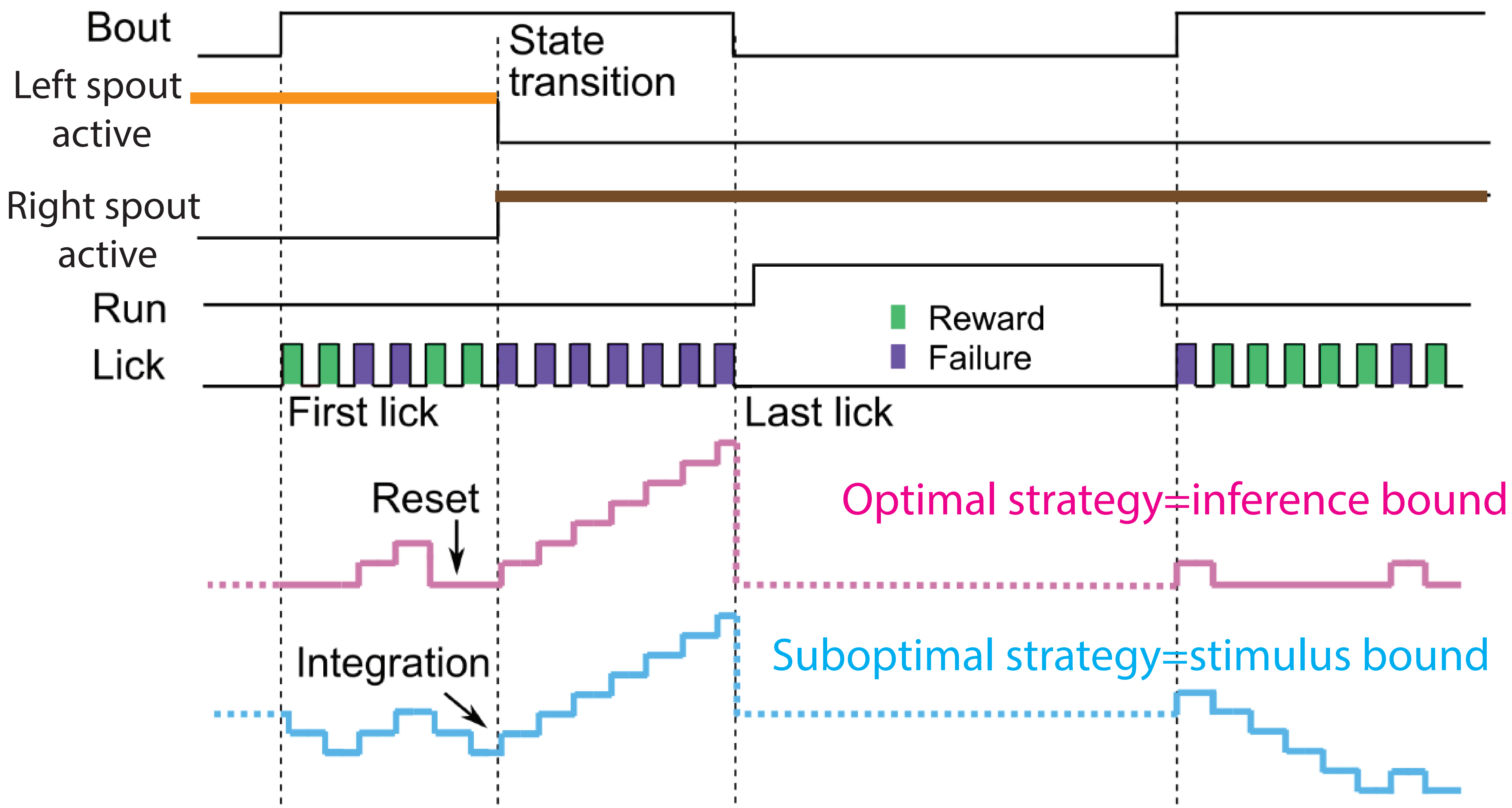
Rewarded lick —> port is active —> stay
Failures. —> prob release?
port inactive?

Integrate Rew & Fail, but how?

…Alternative strategies for evidence accumulation

Naturalistic foraging





Rewarded lick —> port is active —> stay

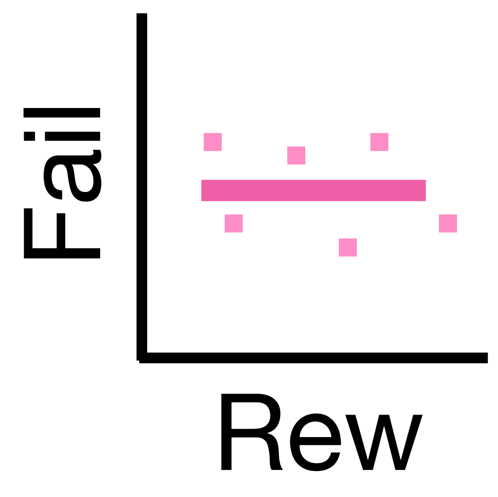
Failures. —> prob release?

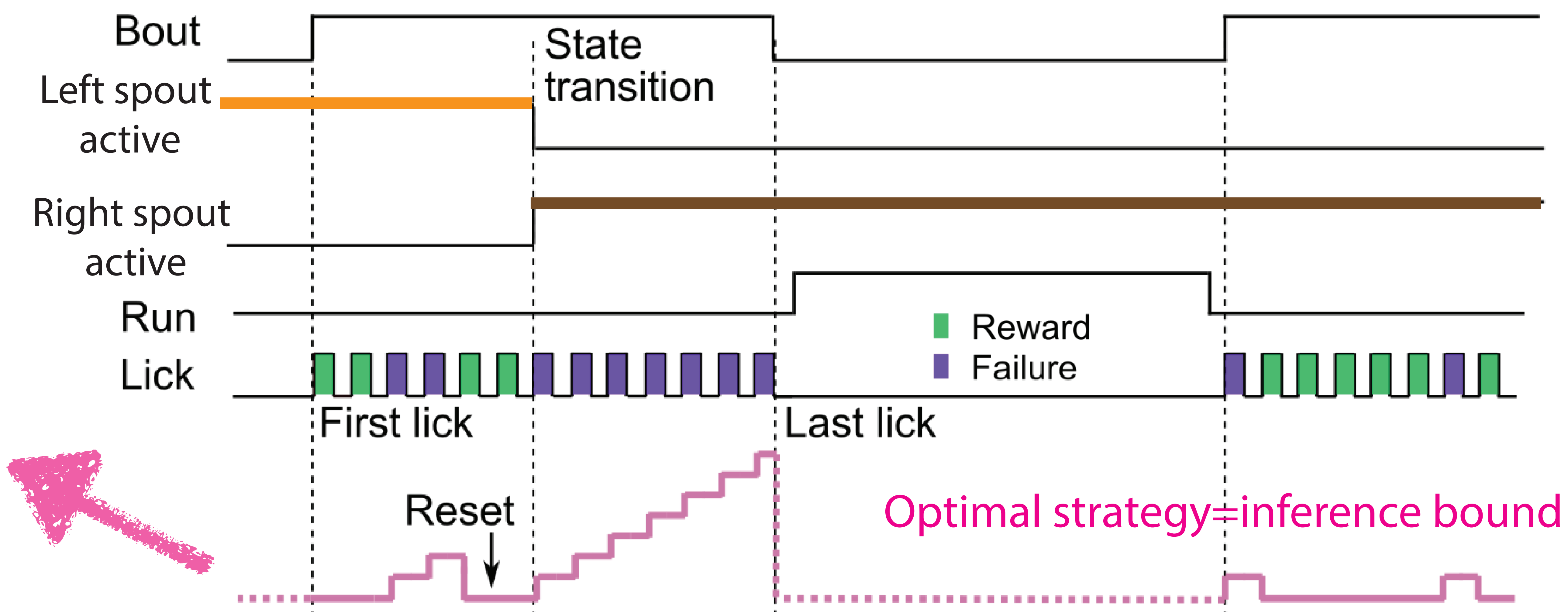
port inactive?

Integrate Rew & Fail, but how?

…Alternative strategies for evidence accumulation

Naturalistic foraging





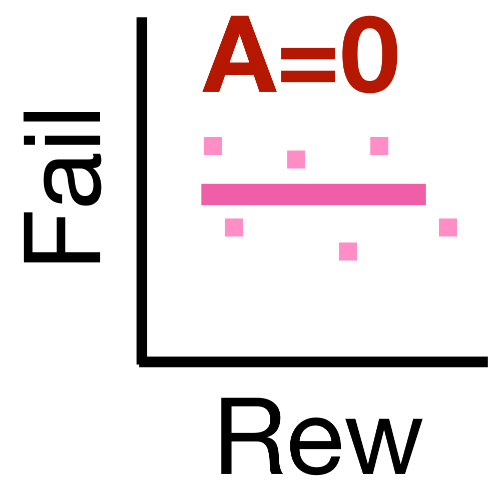
Naturalistic foraging

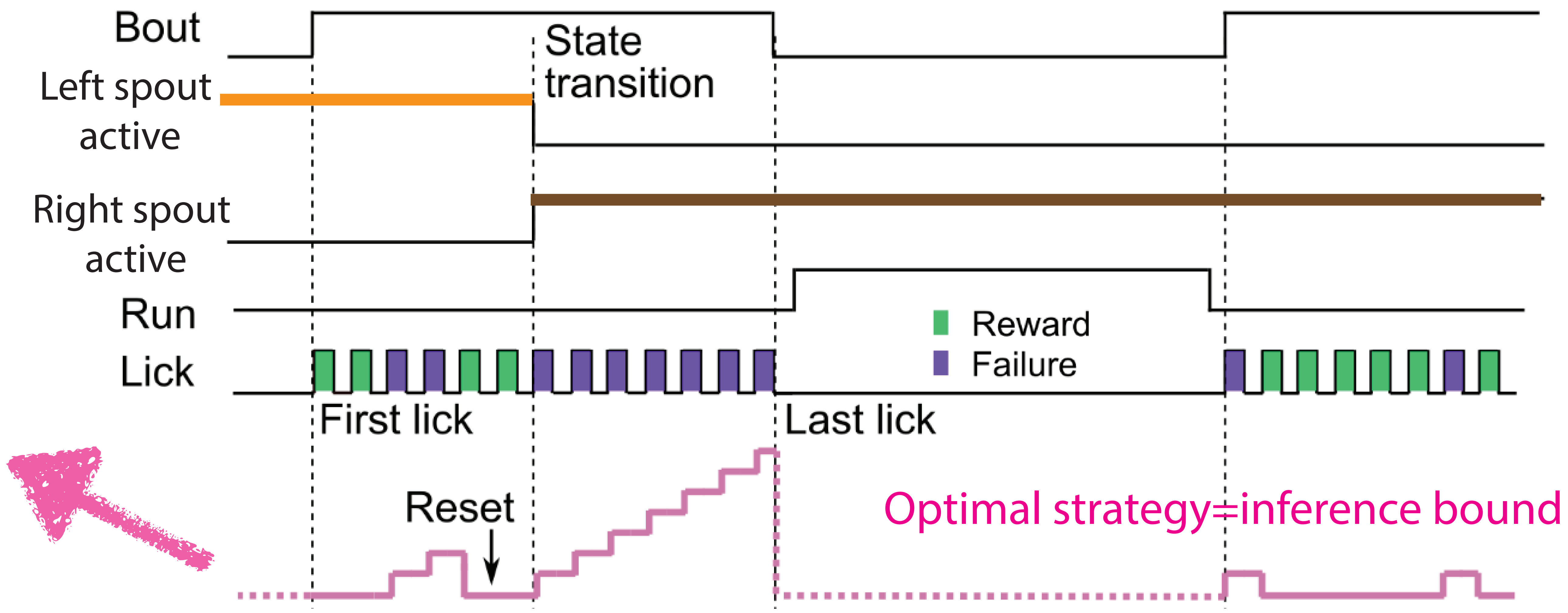
Strategies are **Linear Models**:

Fail = **A** x Rew + b

A = how much they weigh Rew

b = how many Fail they tolerate





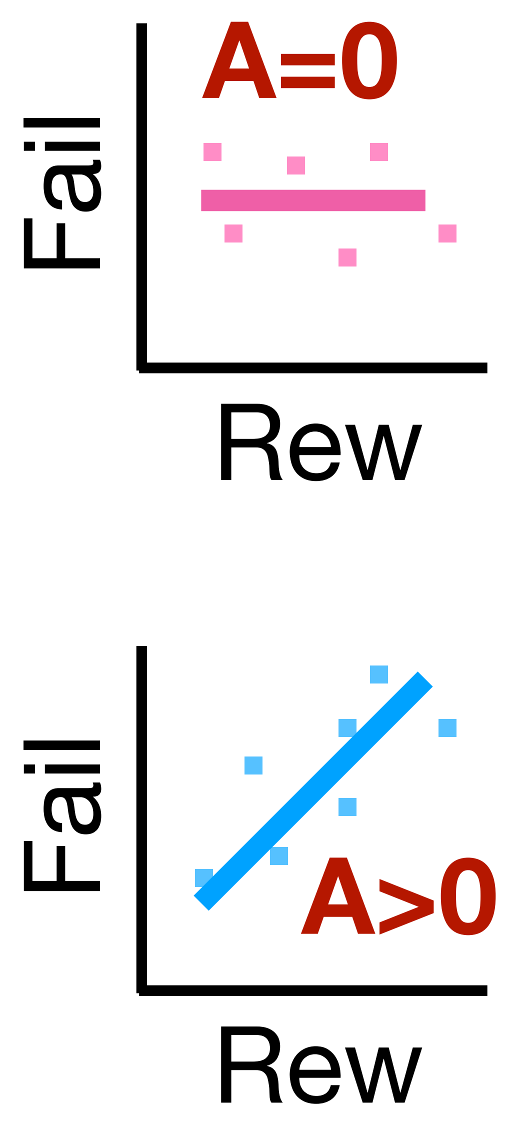
Naturalistic foraging

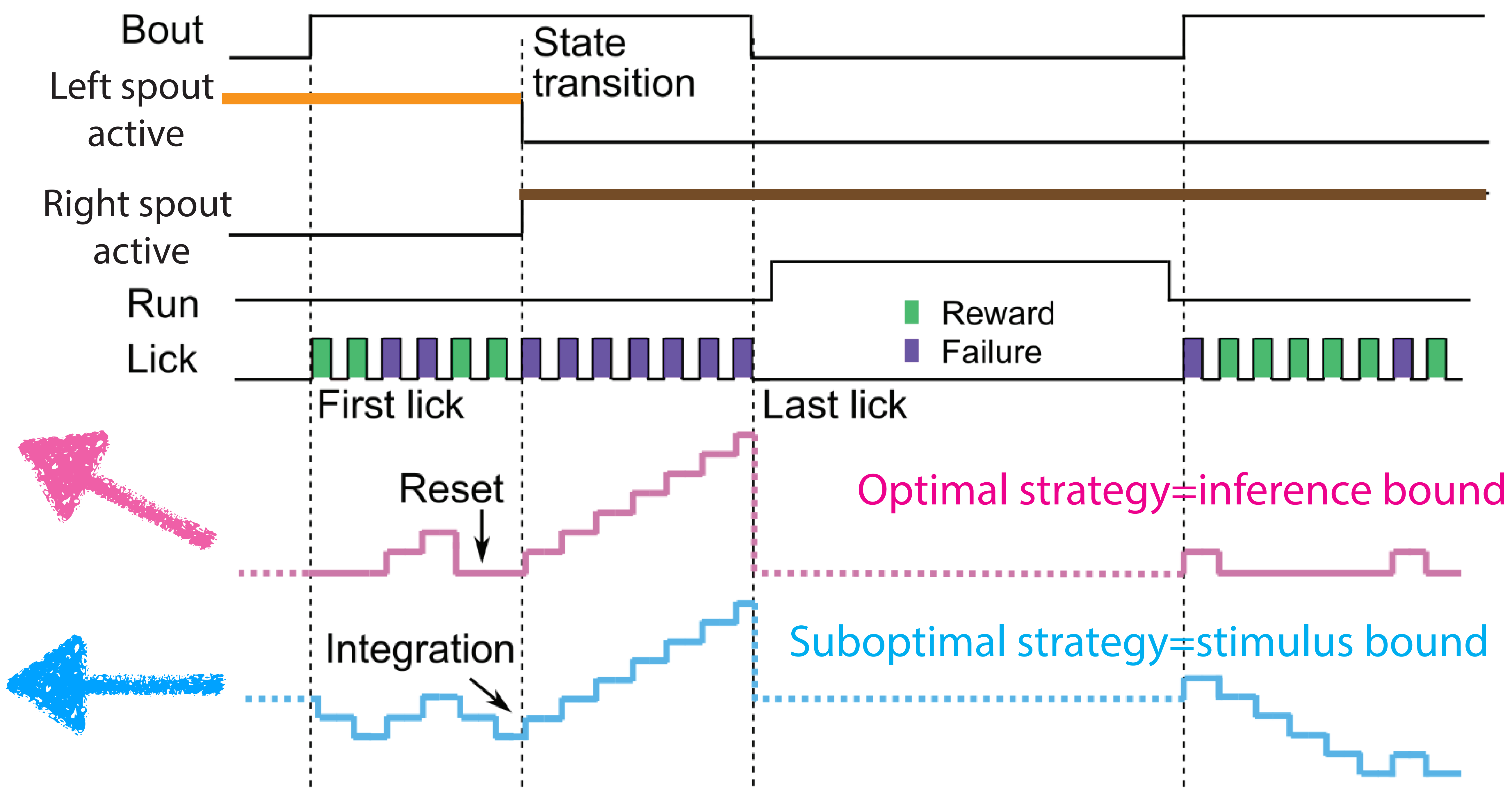
Strategies are **Linear Models:**

Fail = **A** x Rew + b

A = how much they weigh Rew

b = how many Fail they tolerate





Naturalistic foraging

Strategies are **Linear Models**:

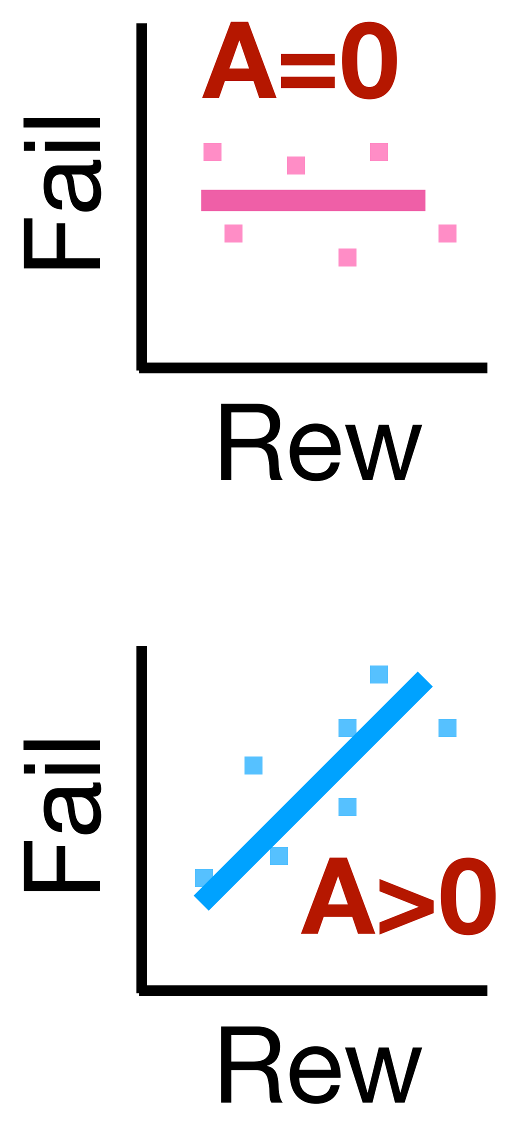
Fail = **A** x Rew + b

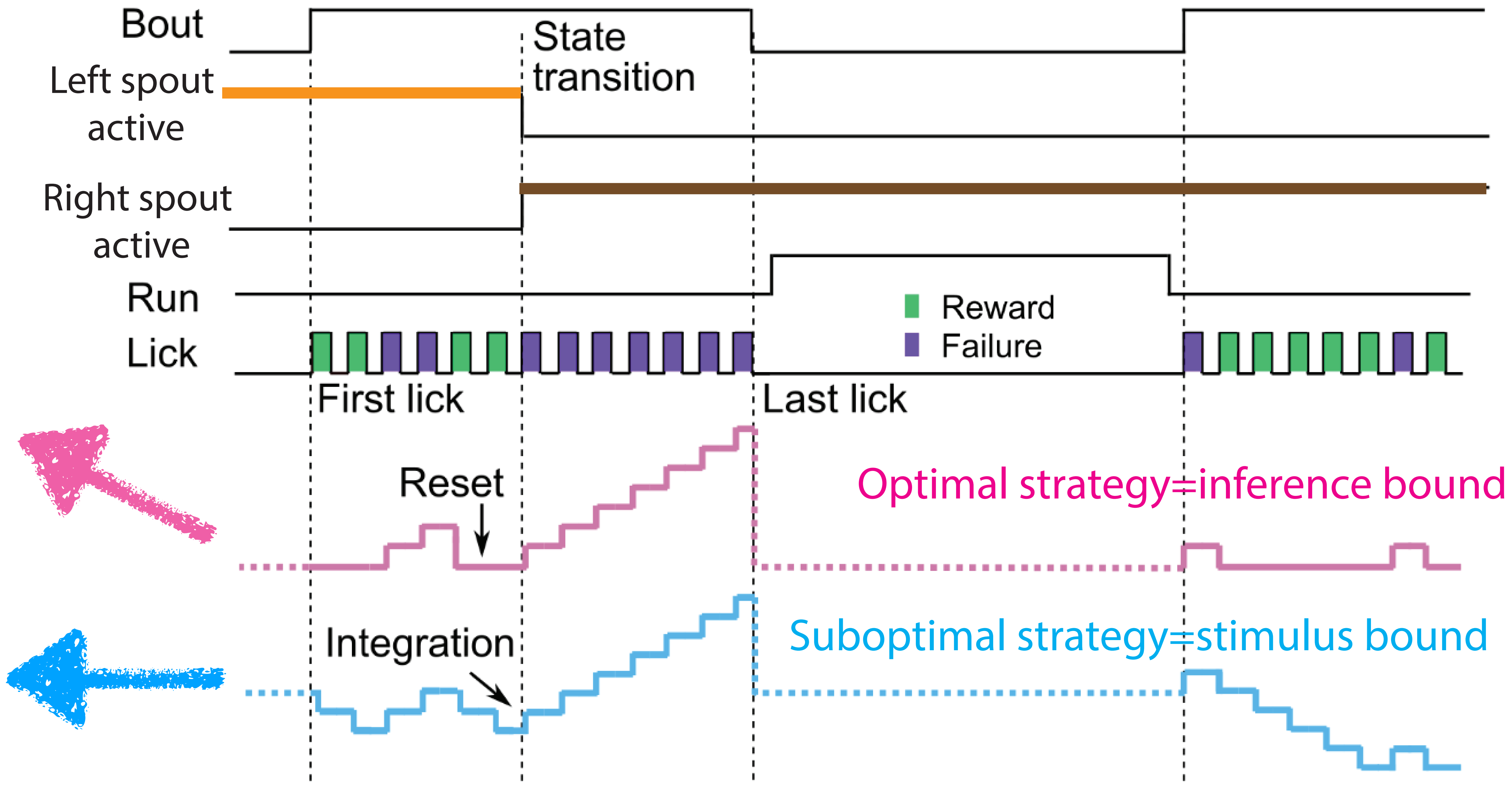
A = how much they weigh Rew

b = how many Fail they tolerate

Switching between strategies

Switching between linear models





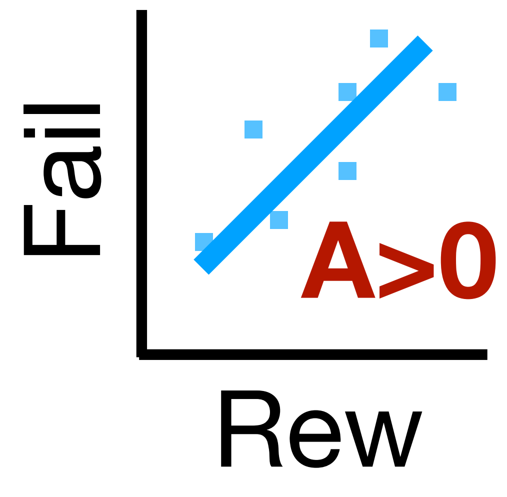
Naturalistic foraging

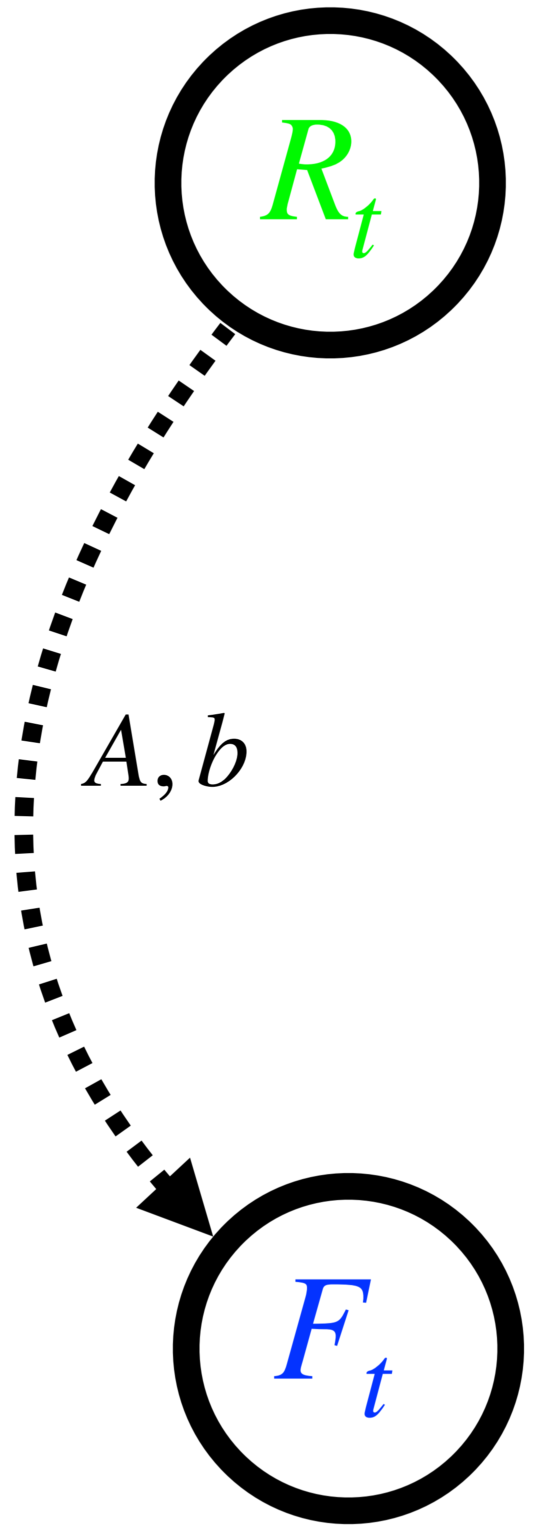
Strategies are **Linear Models**:

Fail = **A** x Rew + b

A = how much they weigh Rew

b = how many Fail they tolerate





$$p(F_t) = N(AR_t + b, \sigma)$$

Naturalistic foraging

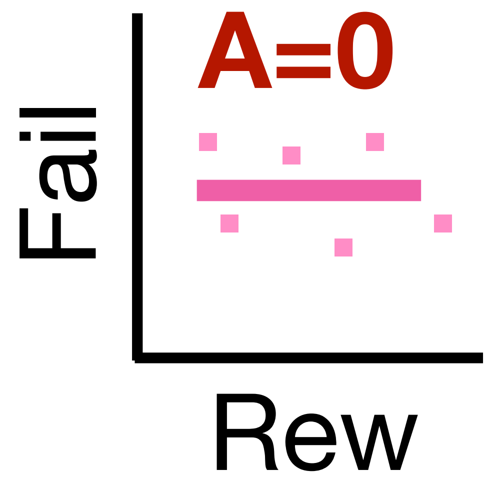
Strategies are **Linear Models**:

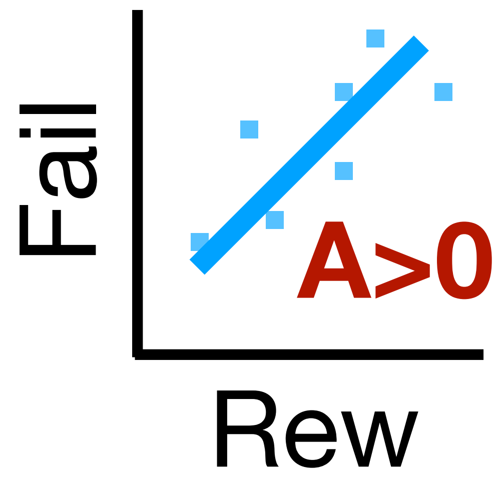
Fail = **A** x Rew + b

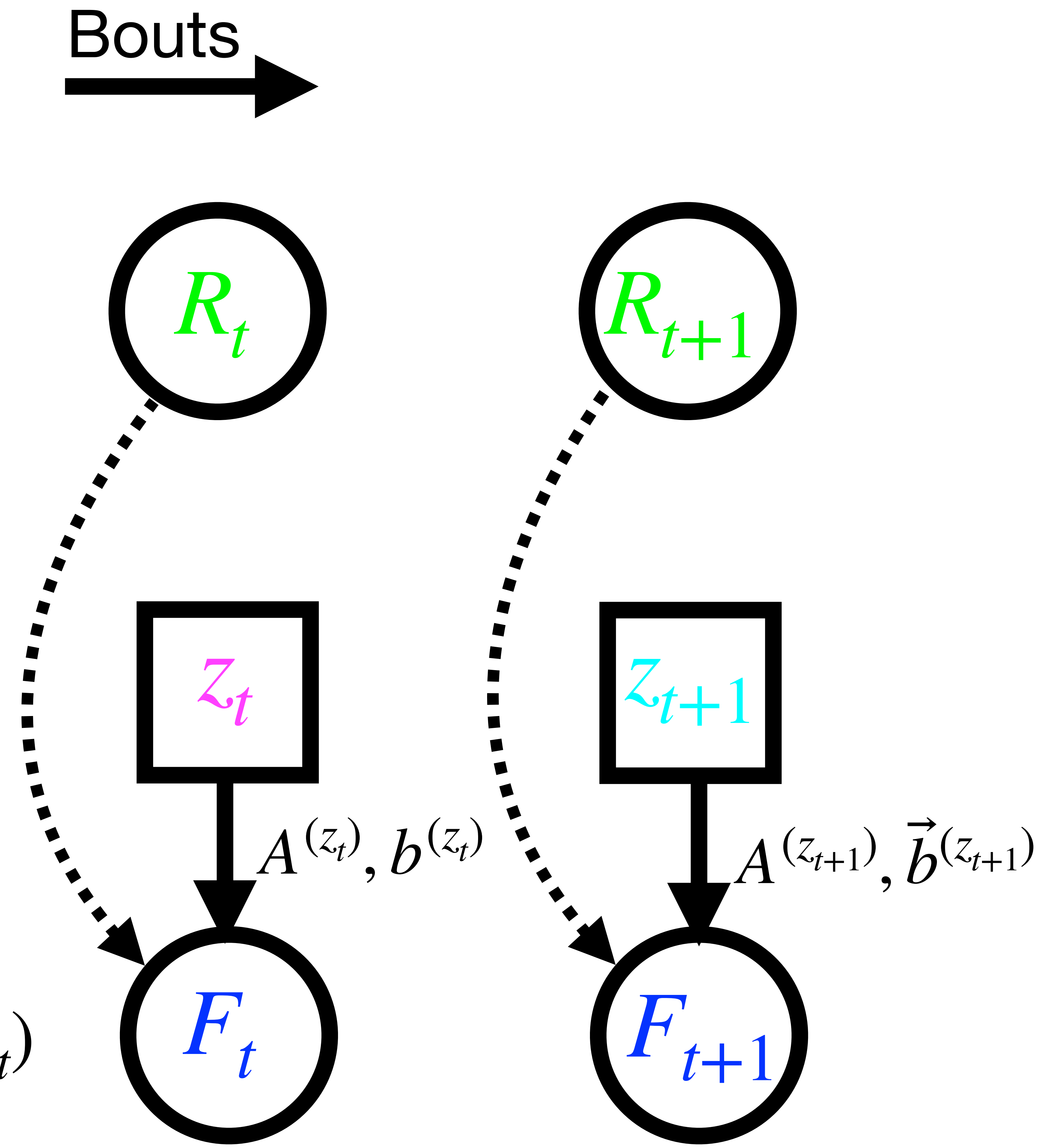
A = how much they weigh Rew

b = how many Fail they tolerate

Linear Model parameters A,b depend on hidden state z_t







$$p(F_t) = N(A^{(z_t)} R_t + b^{(z_t)}, \sigma^{(z_t)} \mid z_t)$$

Naturalistic foraging

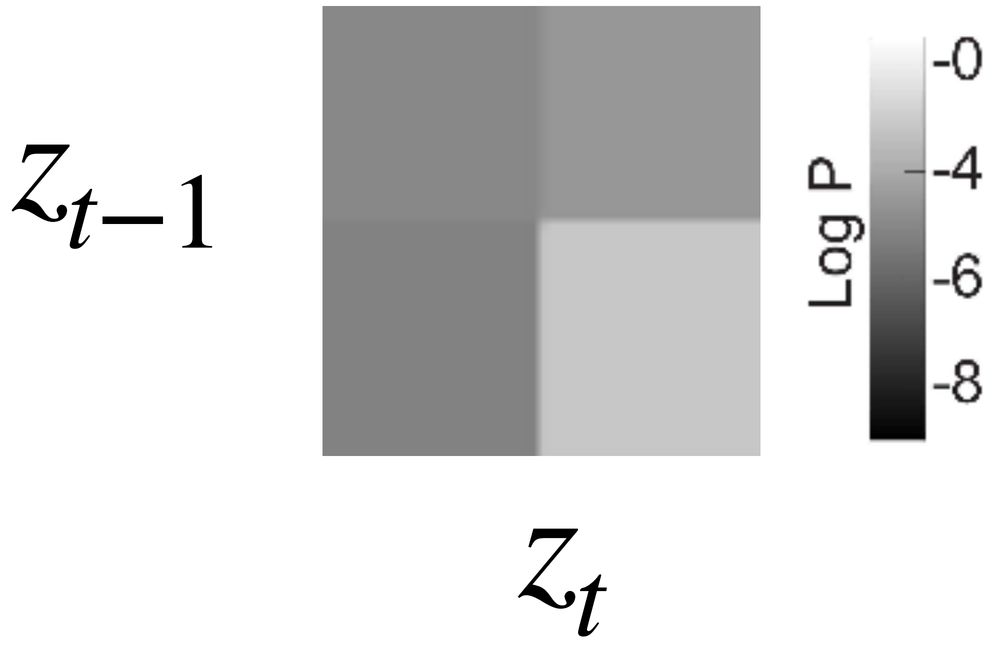
Strategies are **Linear Models**:

Fail = **A** x Rew + b

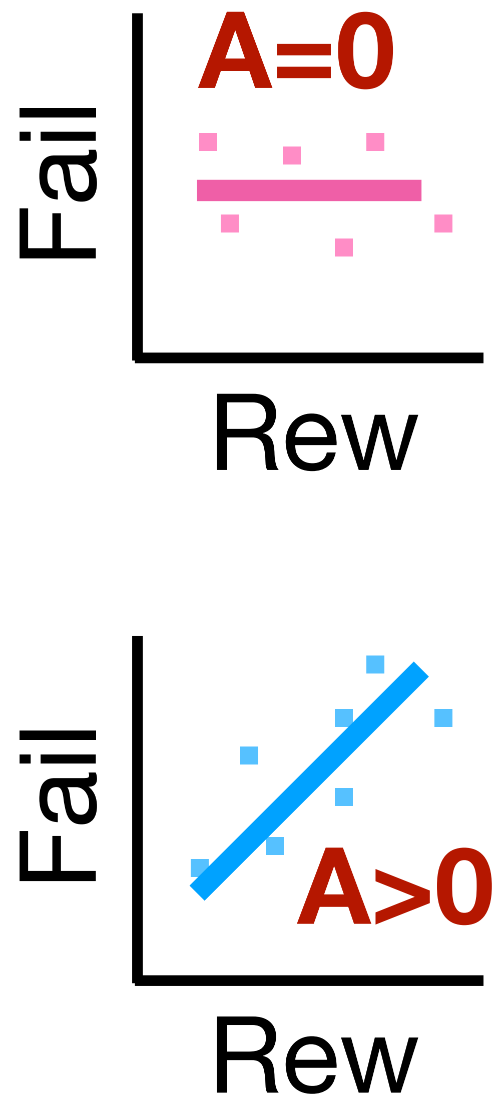
A = how much they weigh Rew

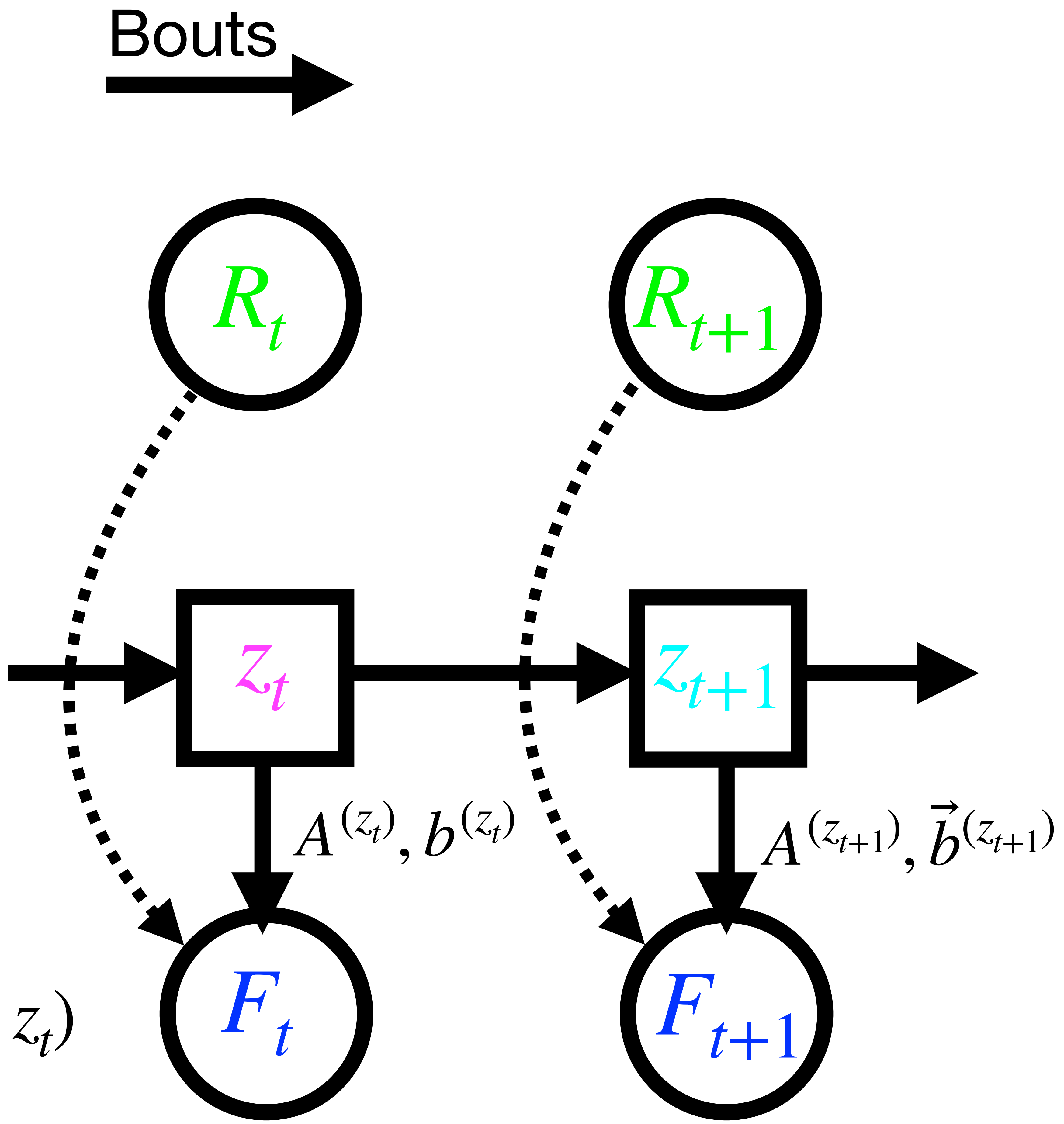
b = how many Fail they tolerate

$$p(z_t = k \,|\, z_{t-1} = l)$$



Hidden Markov model with Linear Model emissions





$$p(F_t) = N(A^{(z_t)}R_t + b^{(z_t)}, \sigma^{(z_t)} \,|\, z_t)$$

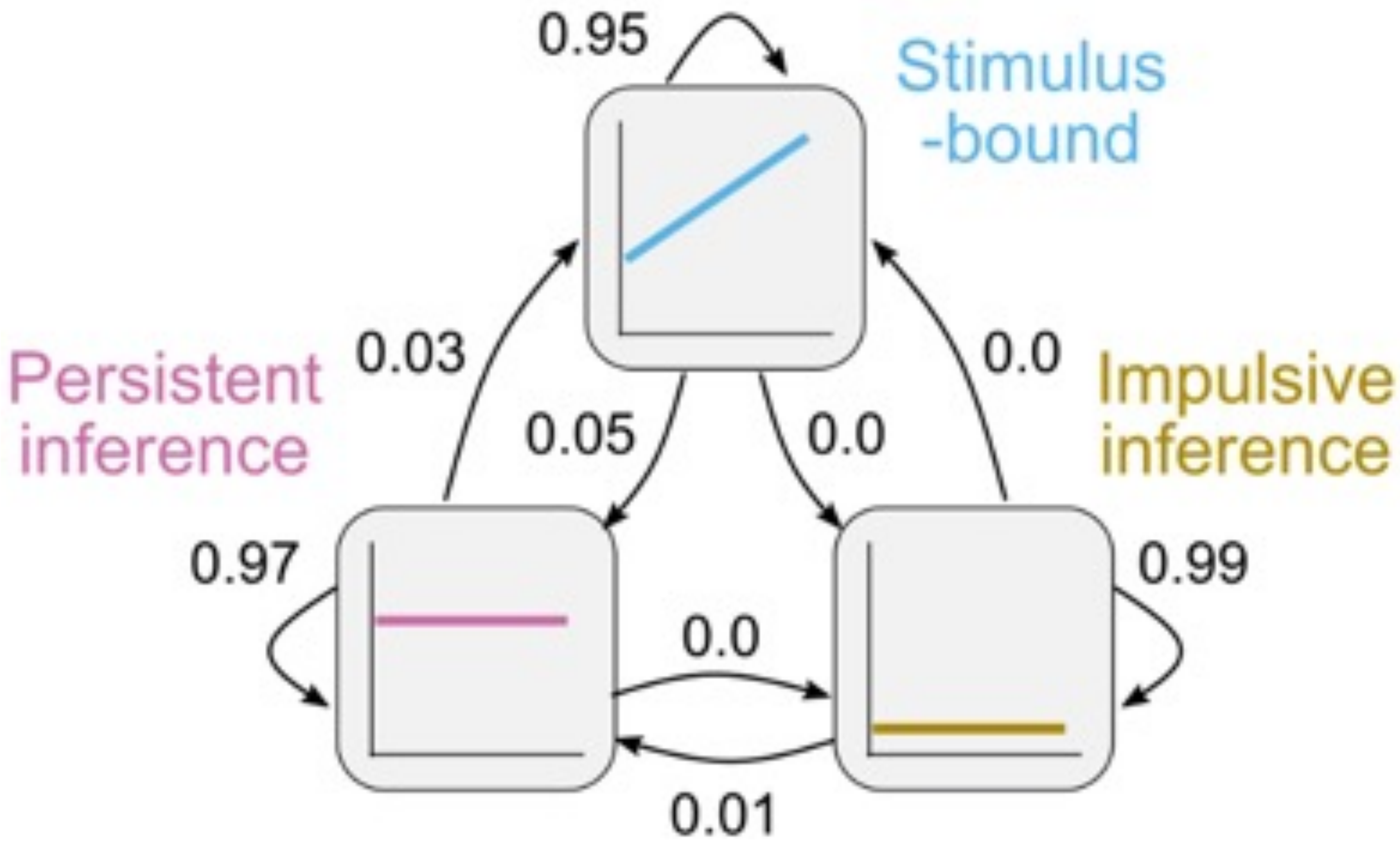
Naturalistic foraging

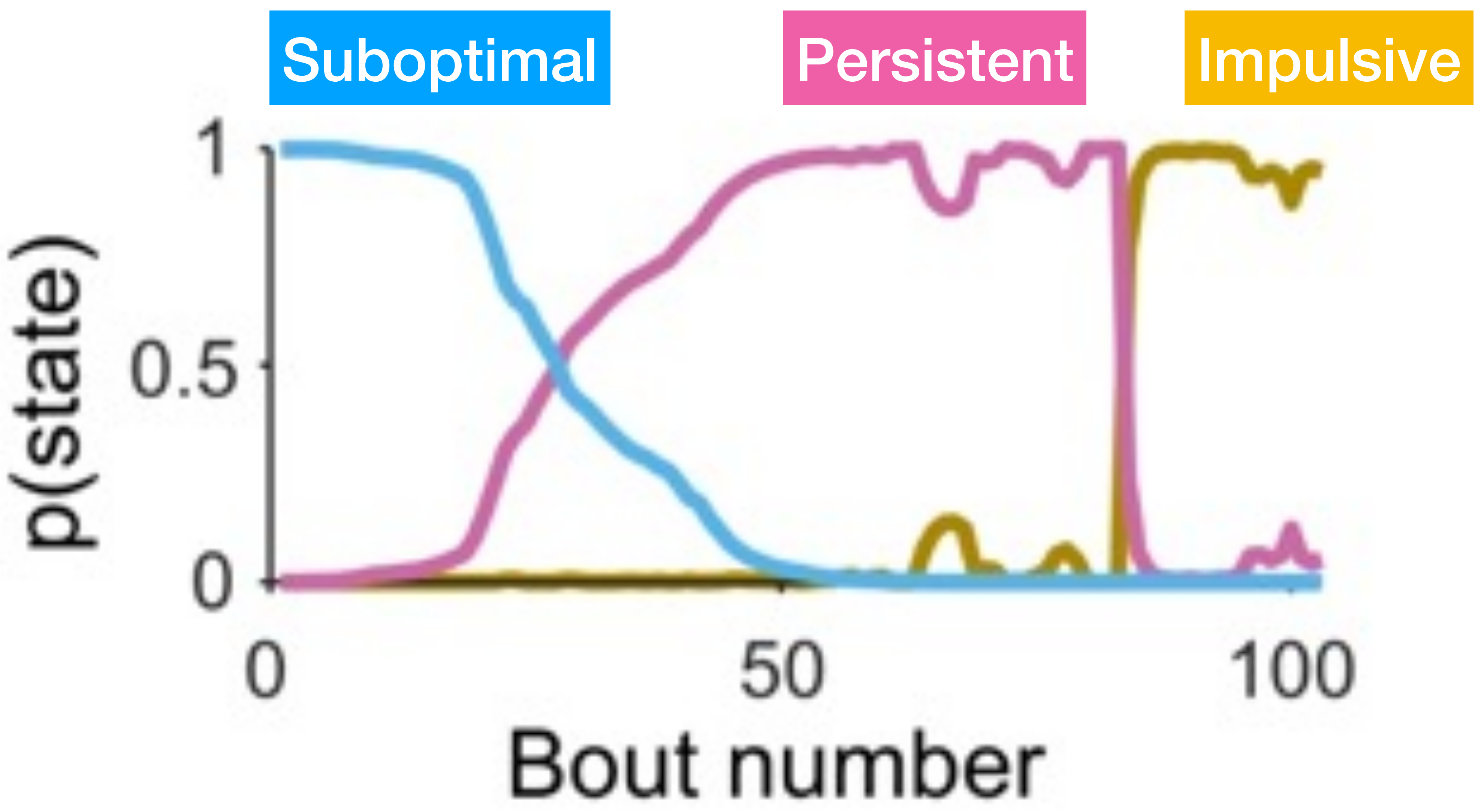
Strategies are **Linear Models**:

Fail = **A** x Rew + b

A = how much they weigh Rew

b = how many Fail they tolerate





[Cazettes et al 2023]

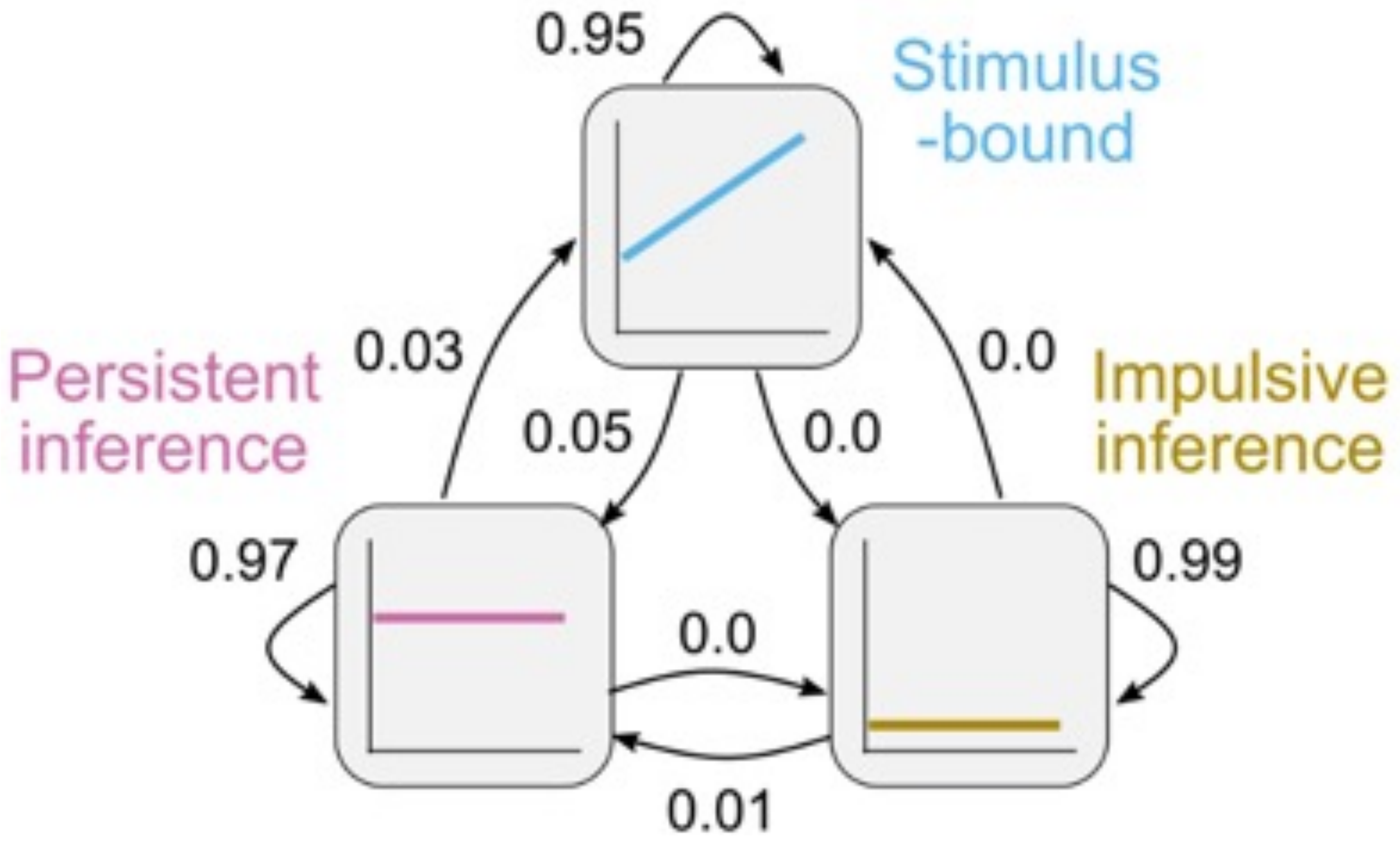
Naturalistic foraging

Strategies are **Linear Models**:

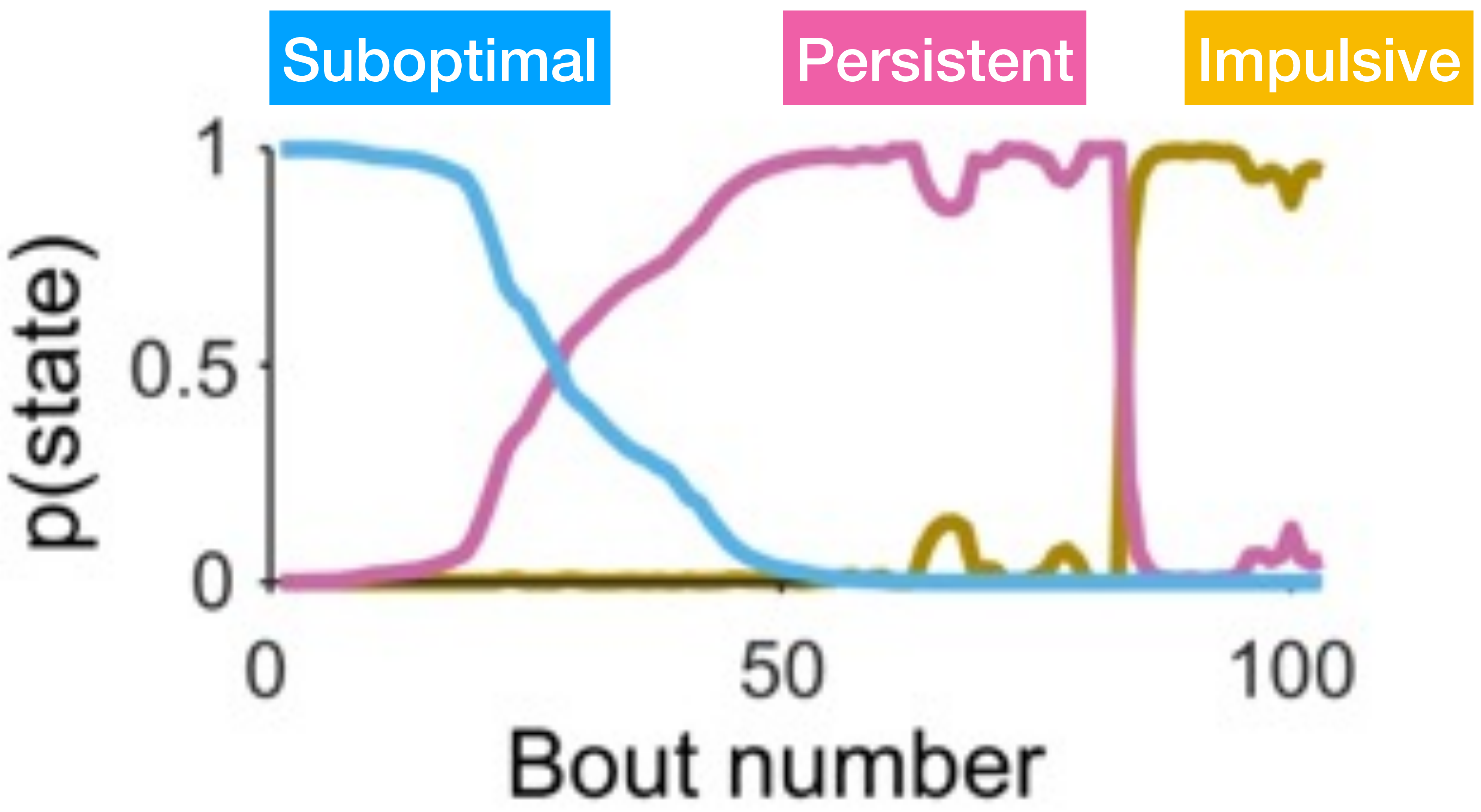
Fail = **A** x Rew + b

A = how much they weigh Rew

b = how many Fail they tolerate



Two optimal strategies, one suboptimal strategy.

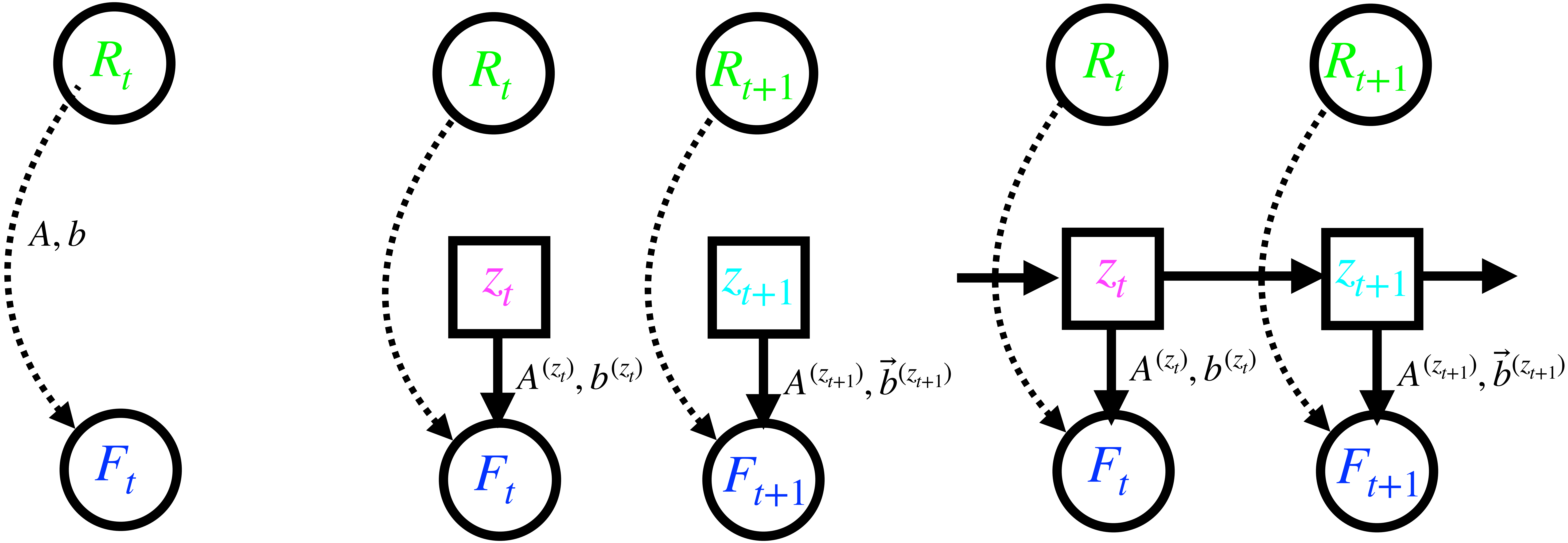


[Cazettes et al 2023]

Coding tutorial

- Load mystery dataset
- Fit models
- Select # hidden states

Fit Linear Model | **Fit mixture of LMs** | **Fit LM-HMM**



HMM with Linear Model emissions

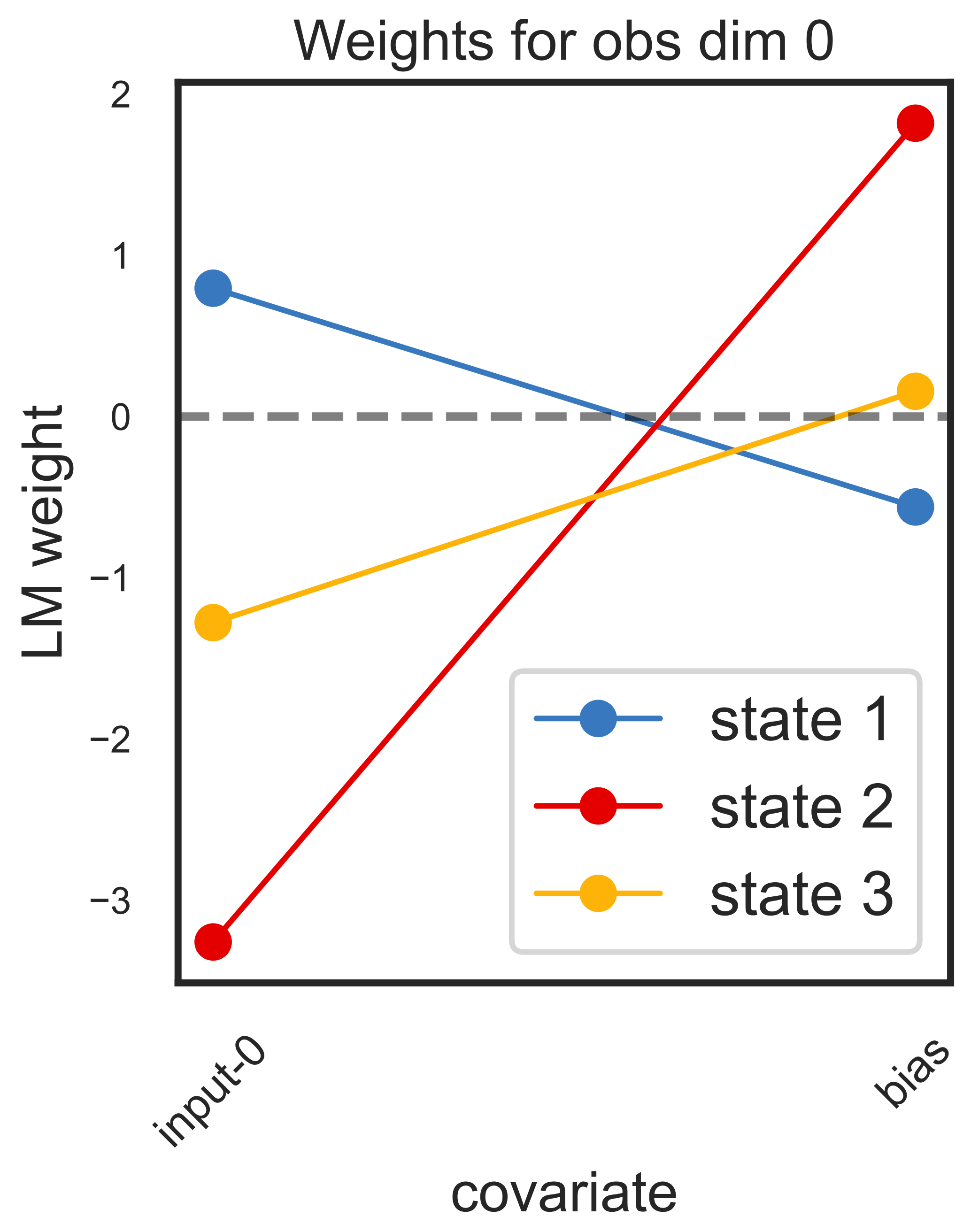
Observations n-dim: $\overrightarrow{F}_t$

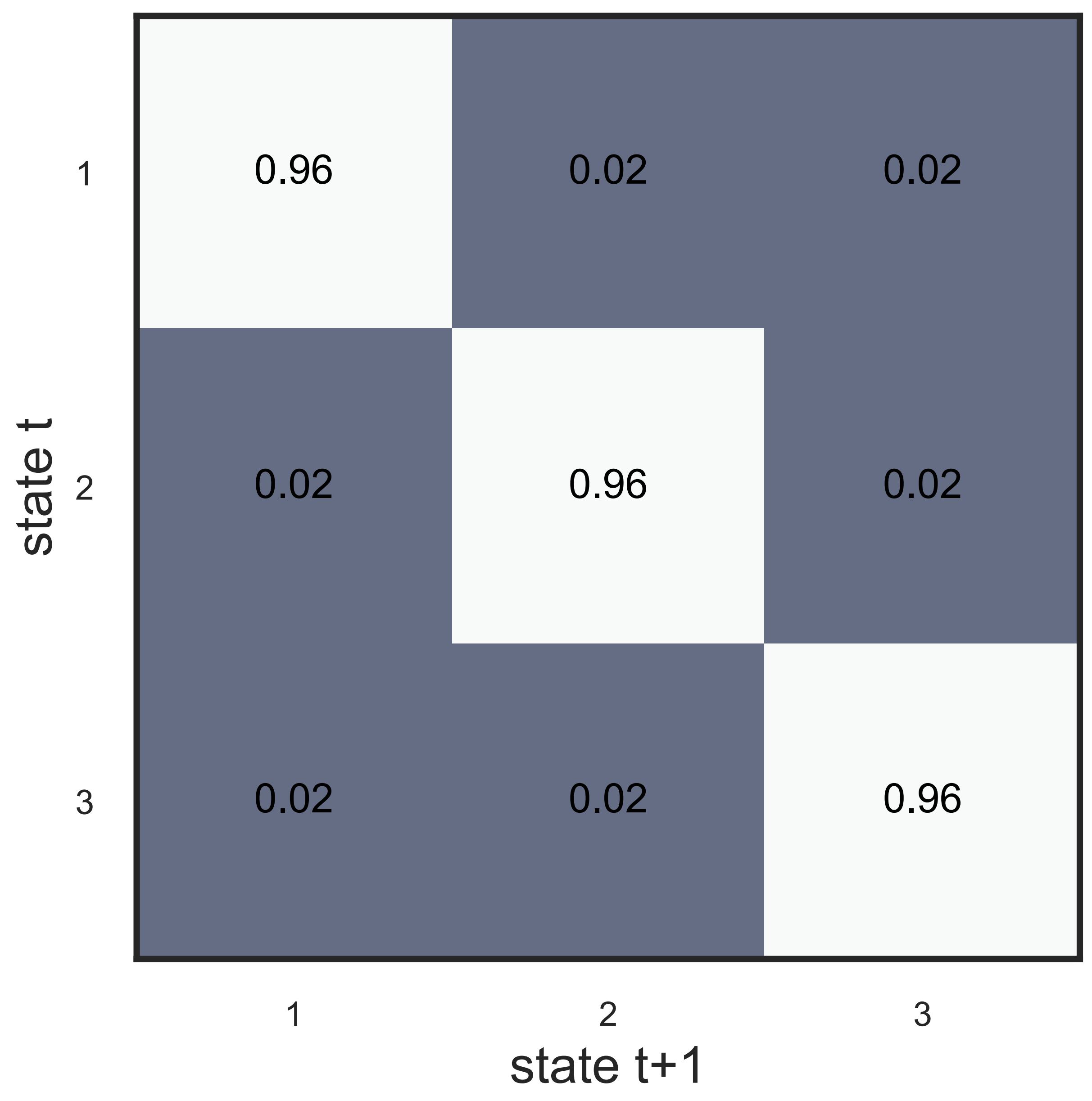
Inputs m-dim: $\overrightarrow{R}_t$

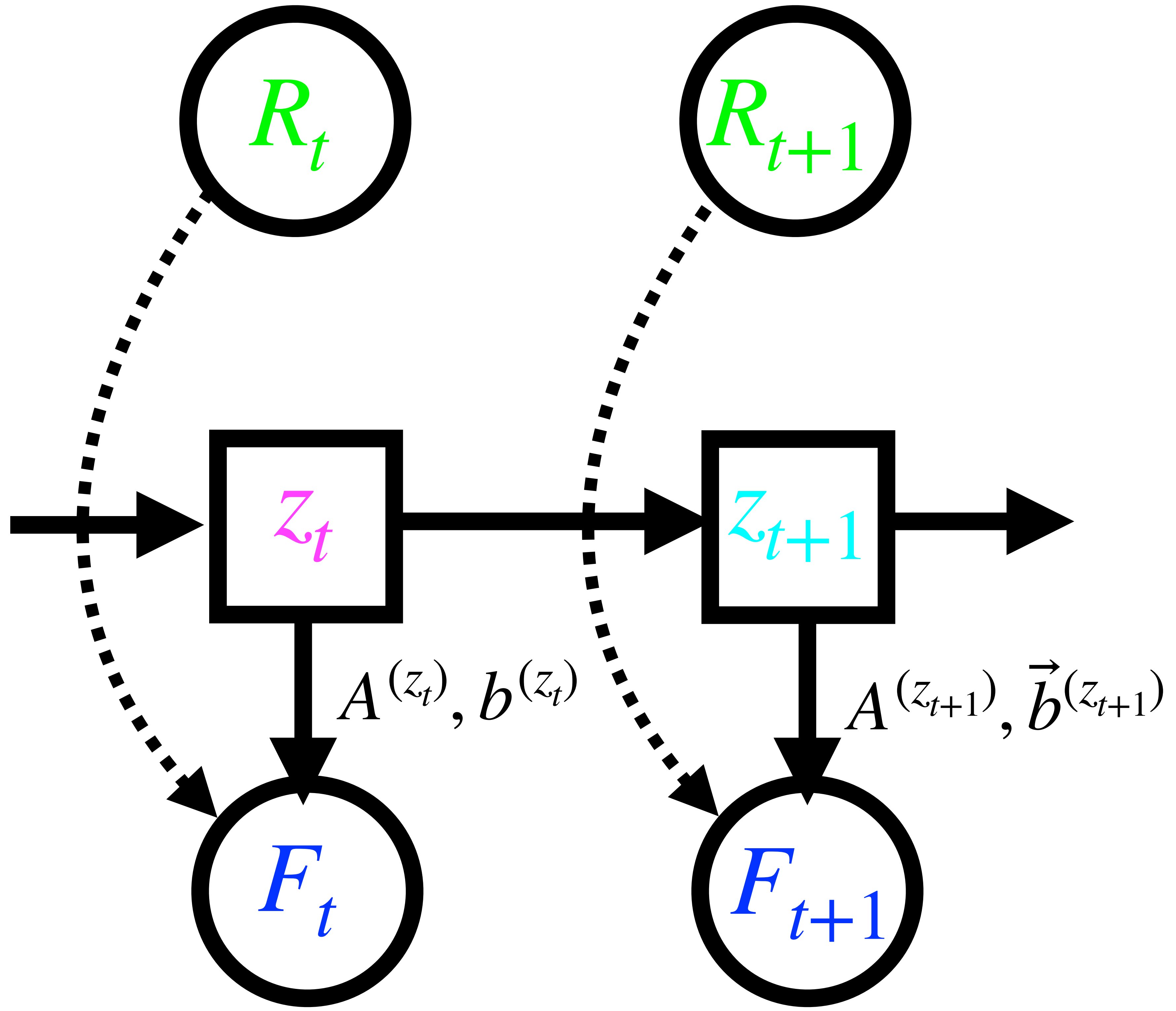
Model: $\overrightarrow{F}_t = A^{(z_t)}\ \overrightarrow{R}_t + b^{(z_t)} + \vec{\epsilon}_t$

Parameters with K hidden states
observation weights $A^{(k)}$, biases $b^{(k)}$,
noise cov $\Sigma^{(k)}$

Transition probabilities $T_{kl} = p(z_t = l \,|\, z_{t-1} = k)$
Initial probabilities: $\pi_k = p(z_0 = k)$







HMM with Linear Model emissions

Observations n-dim: $\overrightarrow{F}_t$

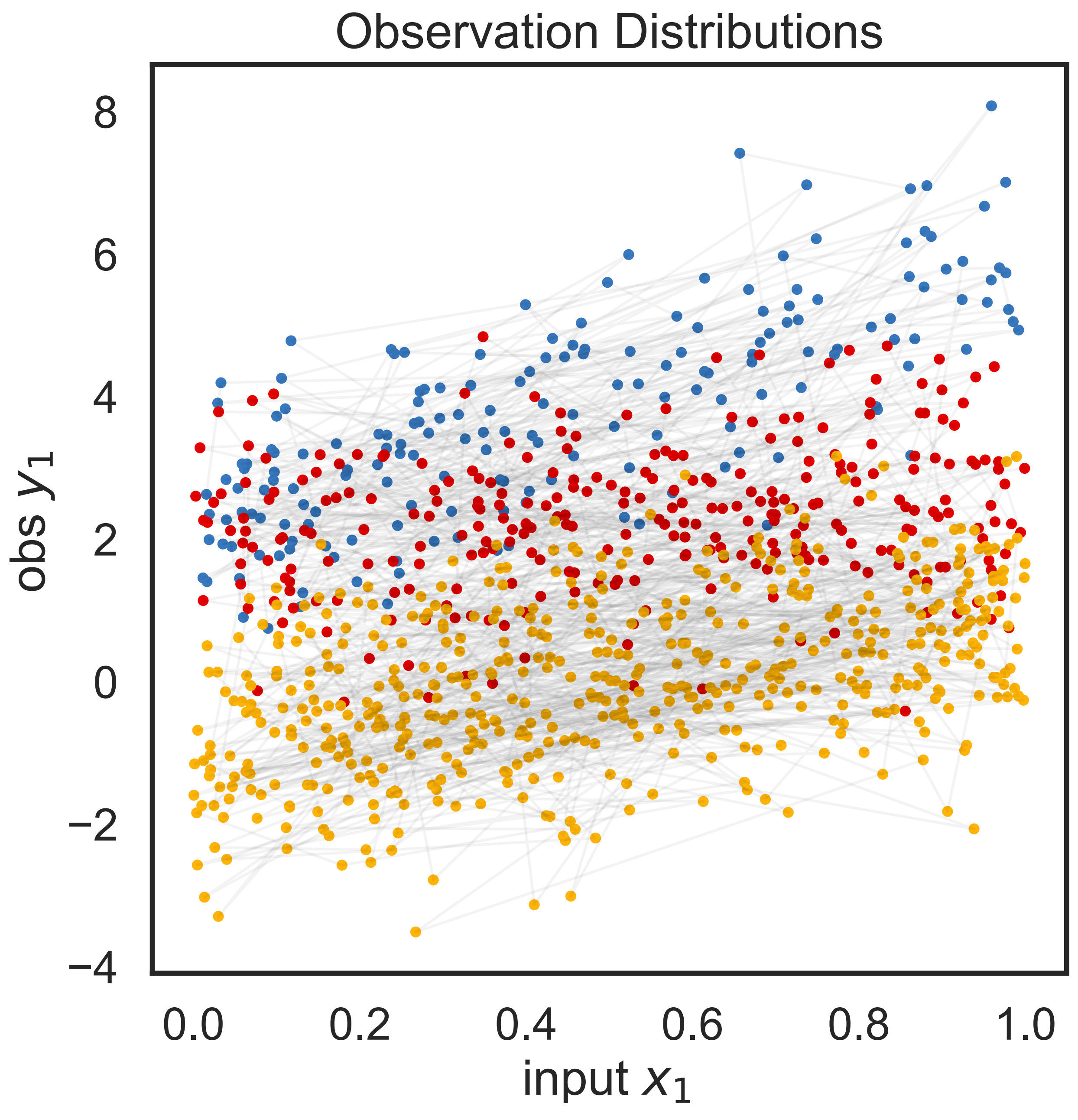
Inputs m-dim: $\overrightarrow{R}_t$

Model: $\overrightarrow{F}_t = A^{(z_t)} \overrightarrow{R}_t + b^{(z_t)} + \vec{\epsilon}_t$

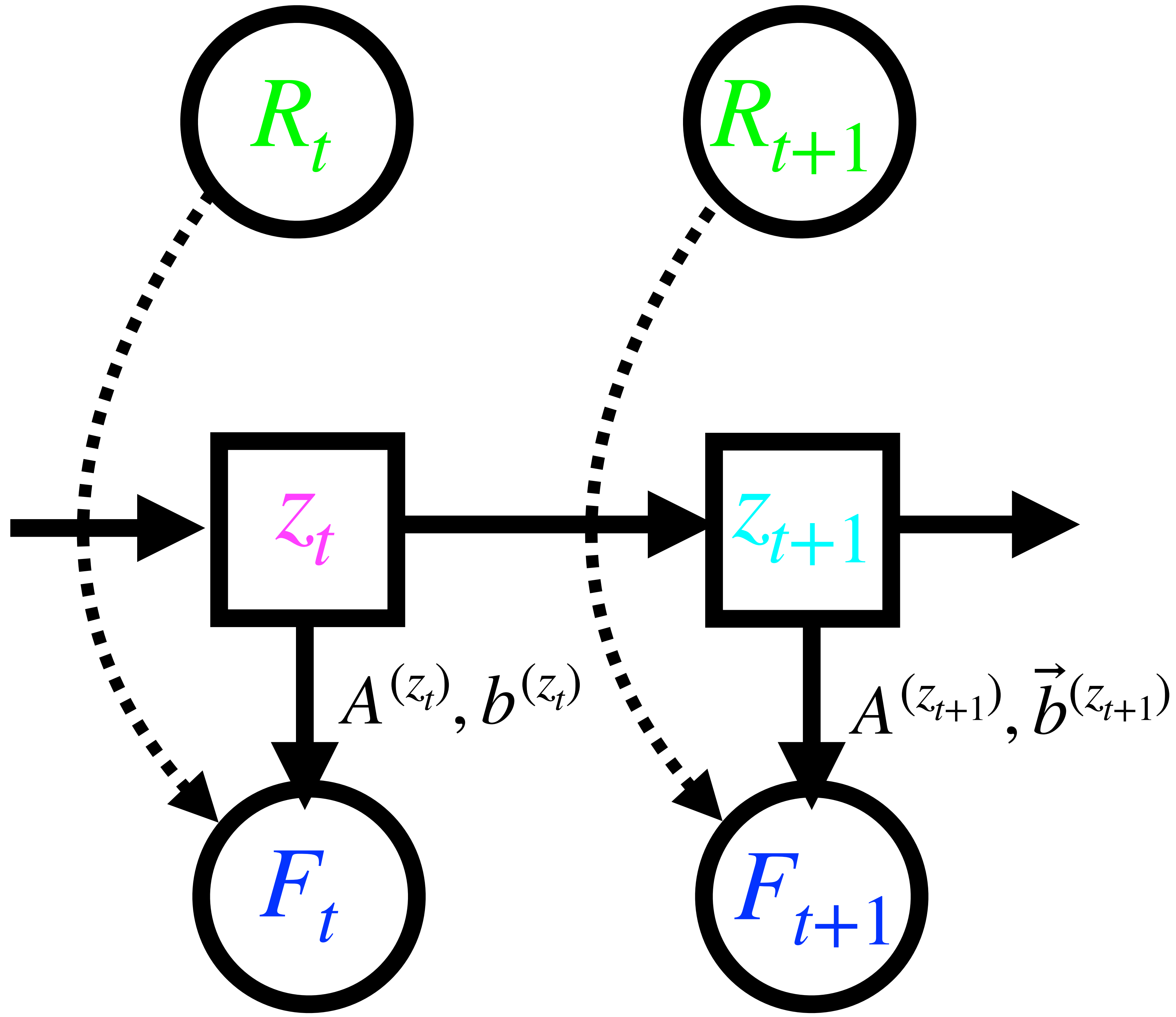
Parameters with K hidden states
observation weights $A^{(k)}$, biases $b^{(k)}$,
noise cov $\Sigma^{(k)}$

Transition probabilities $T_{kl} = p(z_t = l \,|\, z_{t-1} = k)$
Initial probabilities: $\pi_k = p(z_0 = k)$

$\overrightarrow{F}_t$



$\overrightarrow{R}_t$



HMM with Linear Model emissions

Observations n-dim: $\vec{F}_t$

Inputs m-dim: $\vec{R}_t$

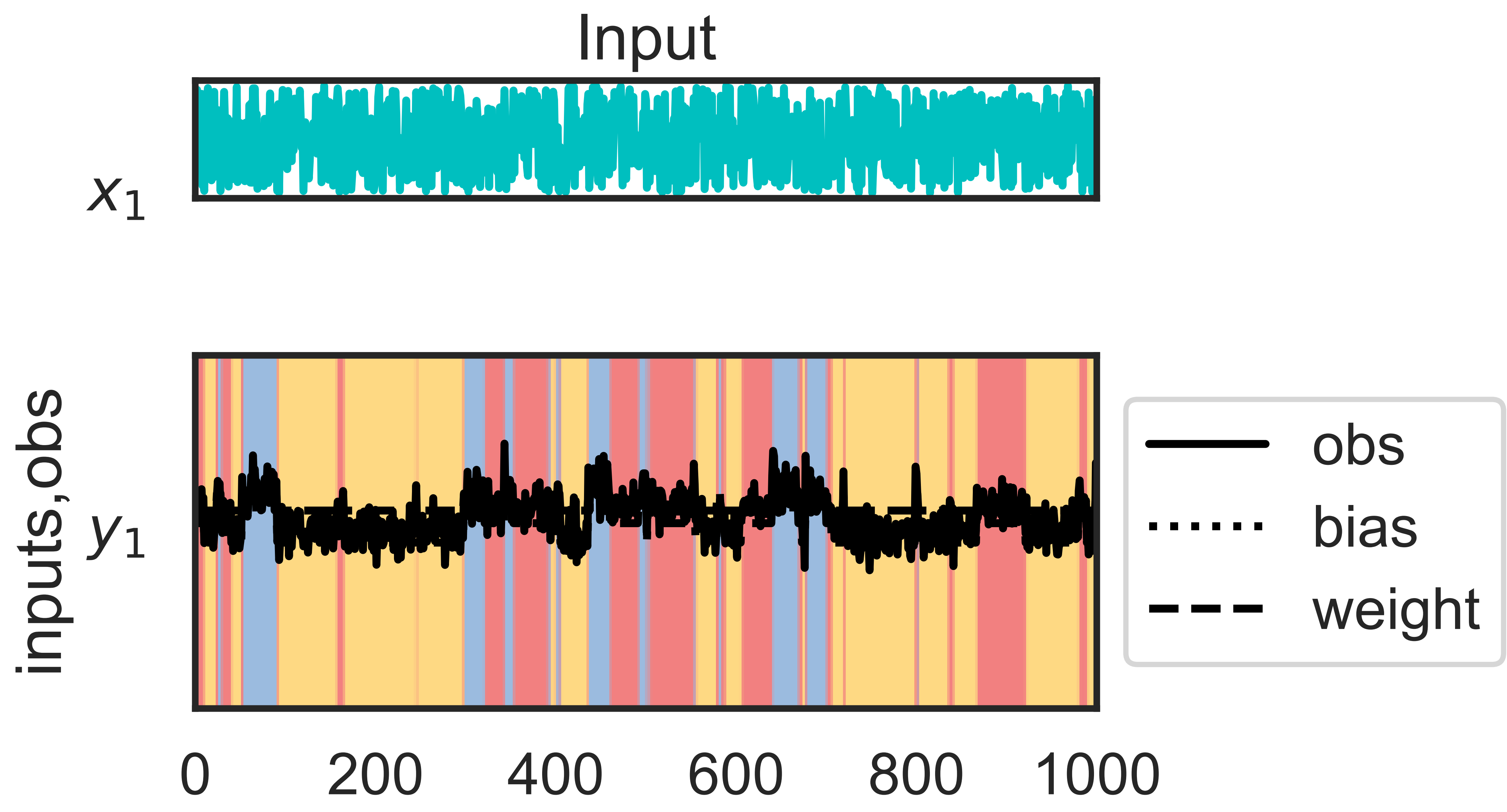
Model: $\vec{F}_t = A^{(z_t)}\vec{R}_t + b^{(z_t)} + \vec{\epsilon}_t$

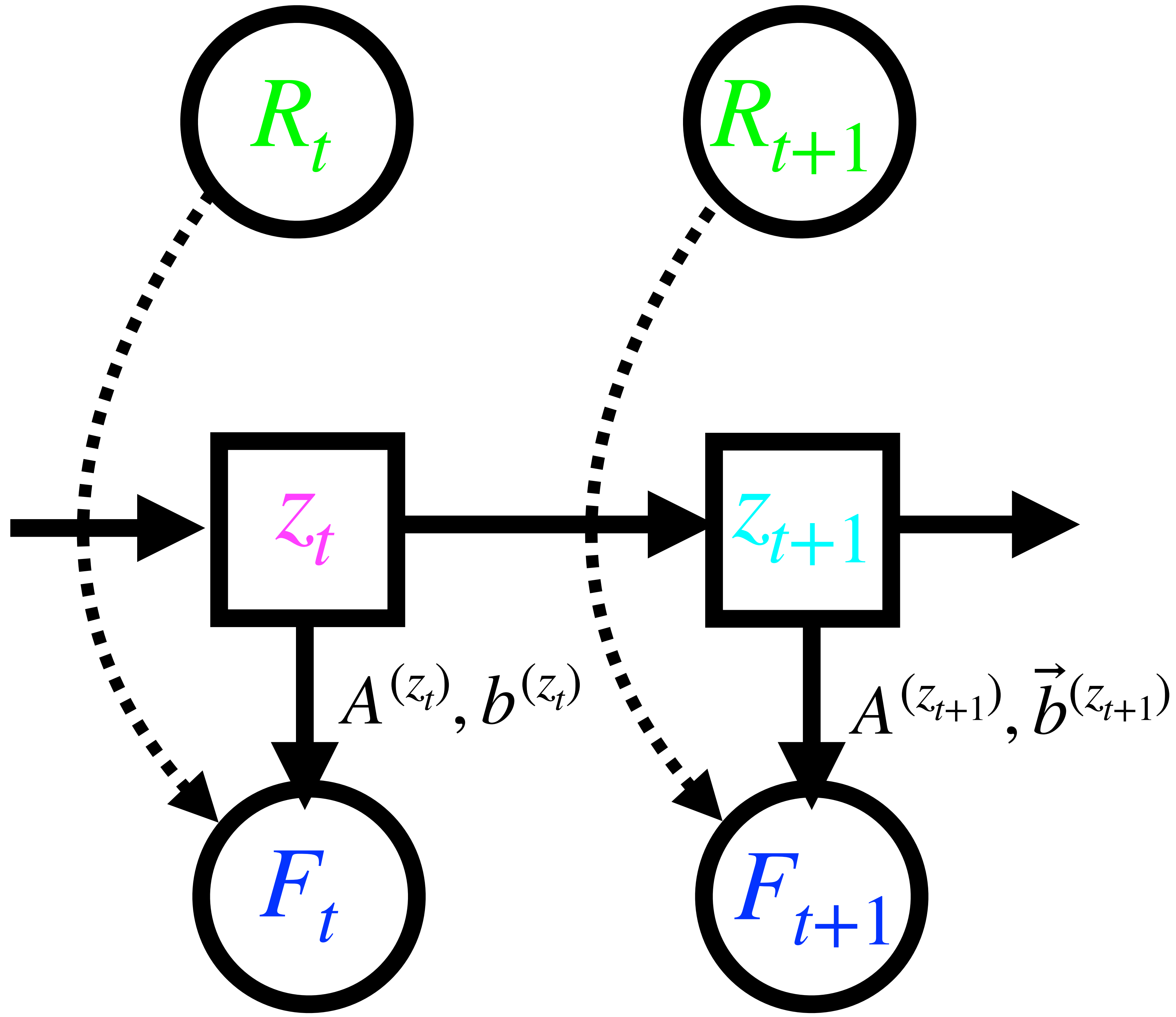
Parameters with K hidden states
observation weights $A^{(k)}$, biases $b^{(k)}$,
noise cov $\Sigma^{(k)}$

Transition probabilities $T_{kl} = p(z_t = l \mid z_{t-1} = k)$
Initial probabilities: $\pi_k = p(z_0 = k)$

$\vec{R}_t$

$\vec{F}_t$





HMM with Linear Model emissions

Observations n-dim: $\overrightarrow{F}_t$

Inputs m-dim: $\overrightarrow{R}_t$

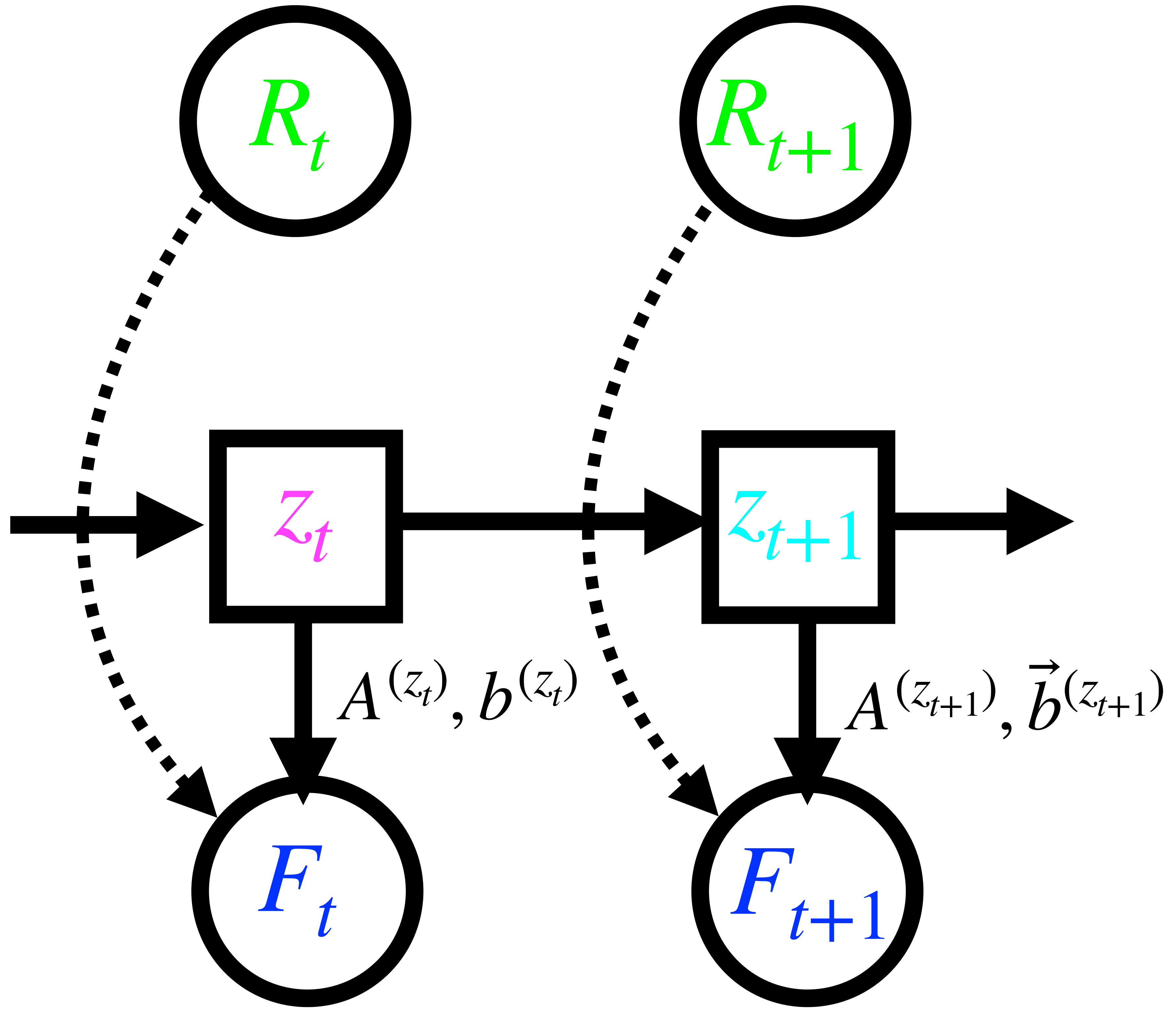
Model: $\overrightarrow{F}_t = A^{(z_t)}\,\overrightarrow{R}_t + b^{(z_t)} + \vec{\epsilon}_t$

Parameters with K hidden states
observation weights $A^{(k)}$, biases $b^{(k)}$,
noise cov $\Sigma^{(k)}$

Transition probabilities $T_{kl} = p(z_t = l \,|\, z_{t-1} = k)$
Initial probabilities: $\pi_k = p(z_0 = k)$

Train model parameters
using Expectation Maximization

Alternate E and M steps
E step: estimate probs with fixed params
M step: estimate params with fixed probs



HMM with Linear Model emissions

Observations n-dim: $\vec{F}_t$

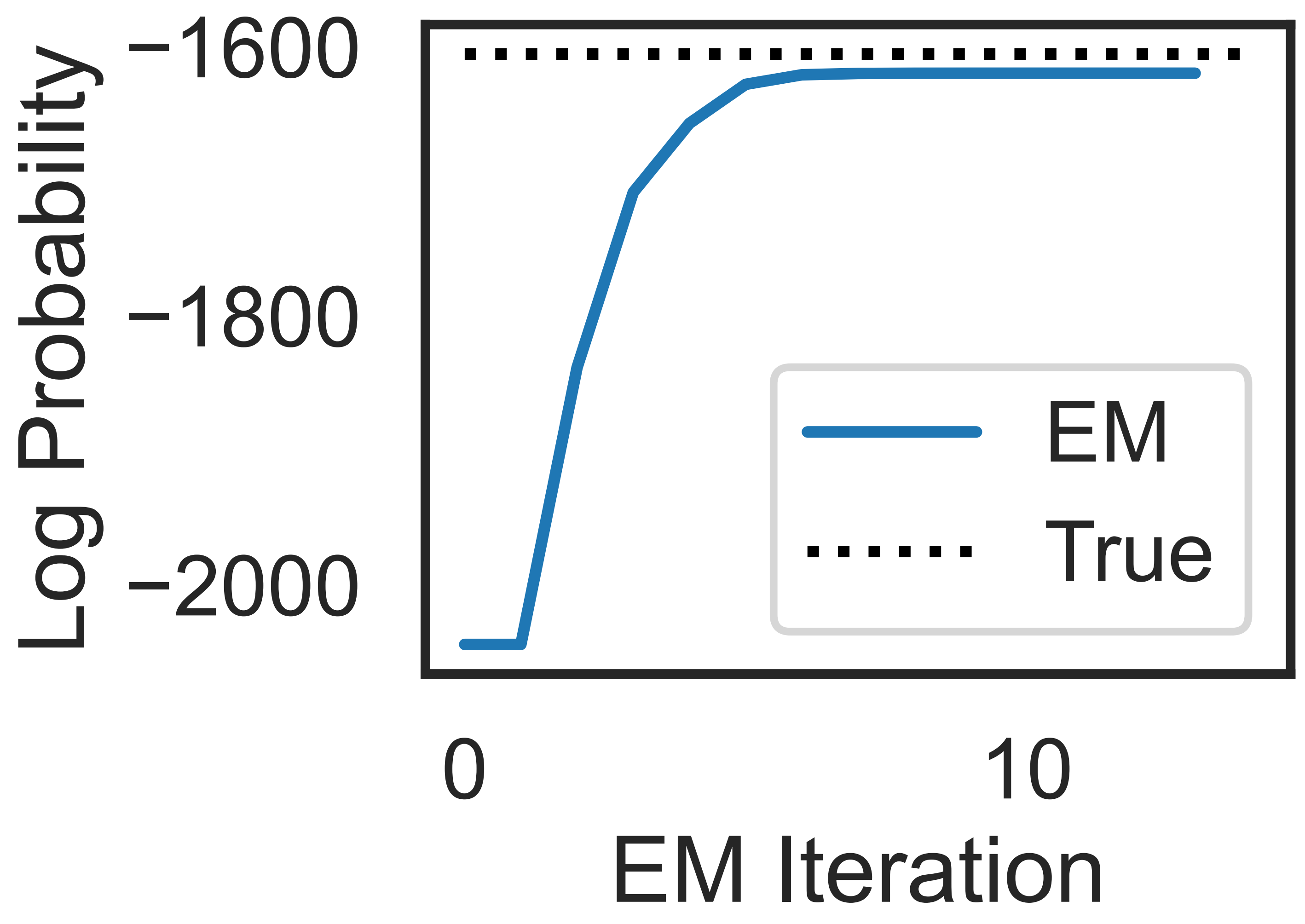
Inputs m-dim: $\vec{R}_t$

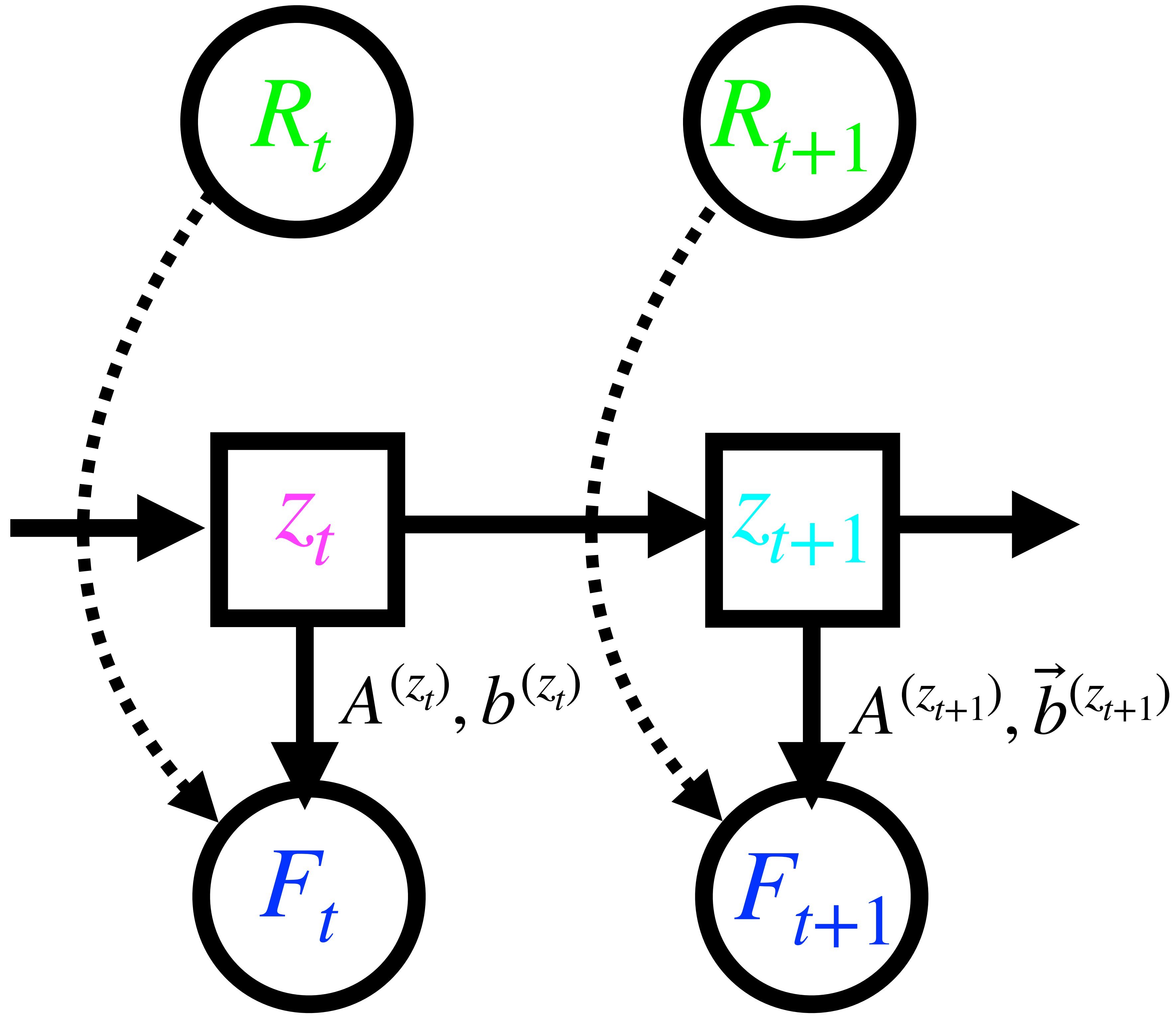
Model: $\vec{F}_t = A^{(z_t)}\, \vec{R}_t + b^{(z_t)} + \vec{\epsilon}_t$

Parameters with K hidden states
observation weights $A^{(k)}$, biases $b^{(k)}$,
noise cov $\Sigma^{(k)}$

Transition probabilities $T_{kl} = p(z_t = l \,|\, z_{t-1} = k)$
Initial probabilities: $\pi_k = p(z_0 = k)$

Train model parameters
using Expectation Maximization





HMM with Linear Model emissions

Observations n-dim: $\overrightarrow{F}_t$

Inputs m-dim: $\overrightarrow{R}_t$

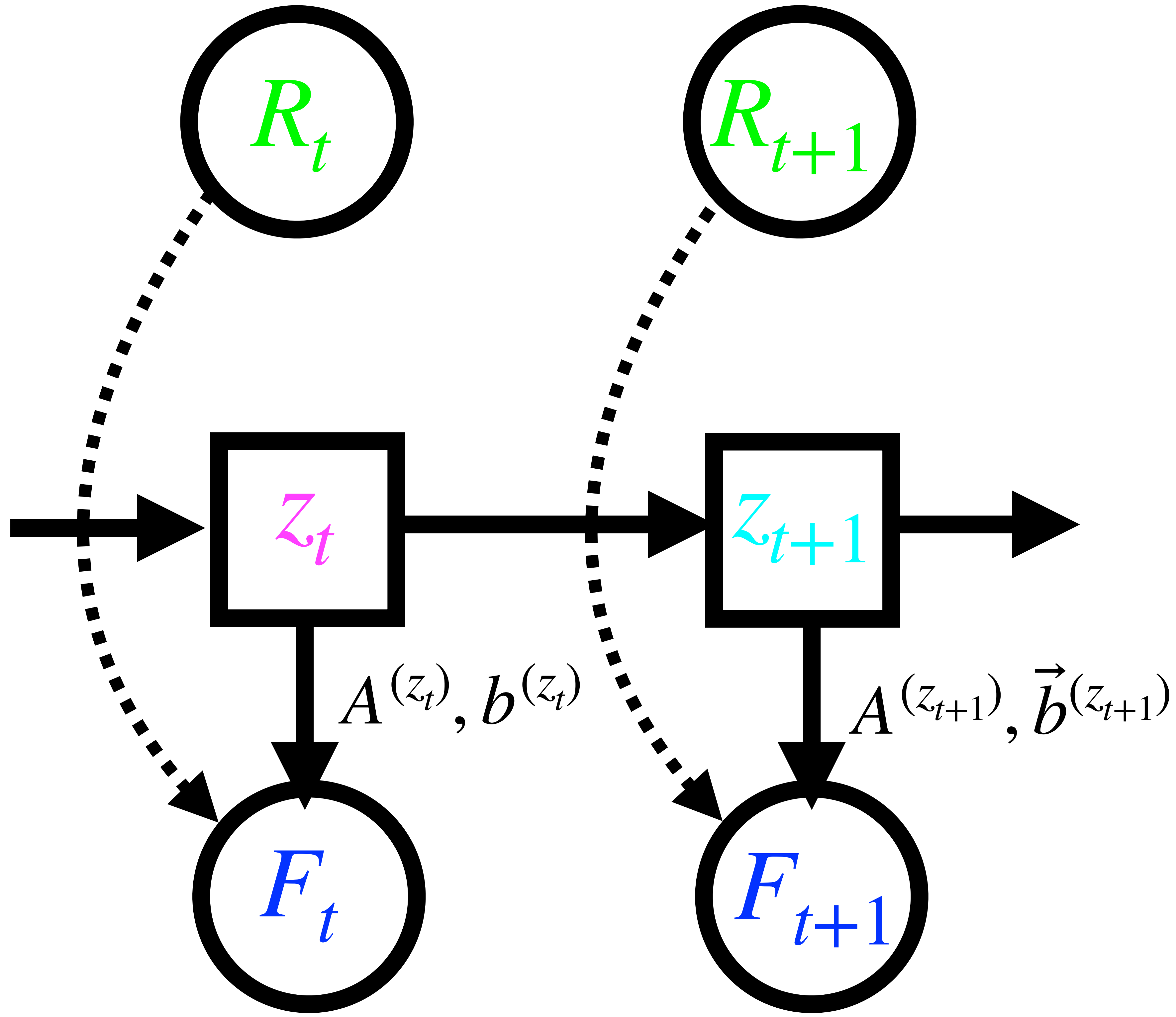
Model: $\overrightarrow{F}_t = A^{(z_t)} \overrightarrow{R}_t + b^{(z_t)} + \vec{\epsilon}_t$

Parameters with K hidden states
observation weights $A^{(k)}$, biases $b^{(k)}$,
noise cov $\Sigma^{(k)}$

Transition probabilities $T_{kl} = p(z_t = l \mid z_{t-1} = k)$
Initial probabilities: $\pi_k = p(z_0 = k)$

Train model parameters
using Expectation Maximization

EM is not convex
—> train many models with diverse
params initialization, pick best



HMM with Linear Model emissions

Observations n-dim: $\vec{F}_t$

Inputs m-dim: $\vec{R}_t$

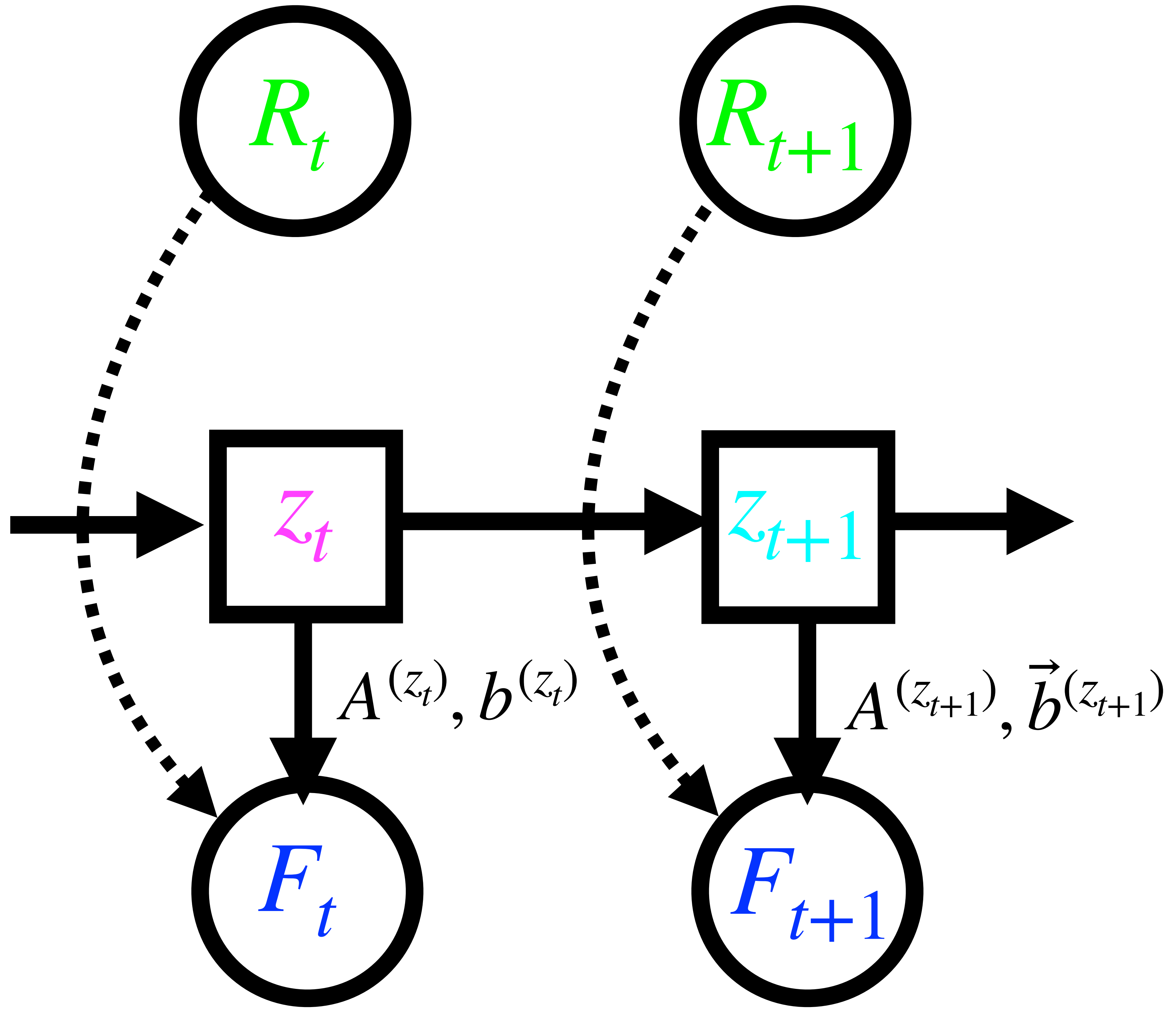
Model: $\vec{F}_t = A^{(z_t)}\,\vec{R}_t + b^{(z_t)} + \vec{\epsilon}_t$

Parameters with K hidden states
observation weights $A^{(k)}$, biases $b^{(k)}$,
noise cov $\Sigma^{(k)}$

Transition probabilities $T_{kl} = p(z_t = l \,|\, z_{t-1} = k)$
Initial probabilities: $\pi_k = p(z_0 = k)$

Train model parameters
using Expectation Maximization

EM is not convex
—> train many models with diverse
params initialization, pick best



Model selection: Pick # states

Observations n-dim: $\overrightarrow{F}_t$

Inputs m-dim: $\overrightarrow{R}_t$

Model: $\overrightarrow{F}_t = A^{(z_t)} \overrightarrow{R}_t + b^{(z_t)} + \vec{\epsilon}_t$

Parameters with K hidden states
observation weights $A^{(k)}$, biases $b^{(k)}$,
noise cov $\Sigma^{(k)}$

Transition probabilities $T_{kl} = p(z_t = l \,|\, z_{t-1} = k)$
Initial probabilities: $\pi_k = p(z_0 = k)$

How do we choose the
number of states?

Model selection: Pick # states

Observations n-dim: $\overrightarrow{F}_t$

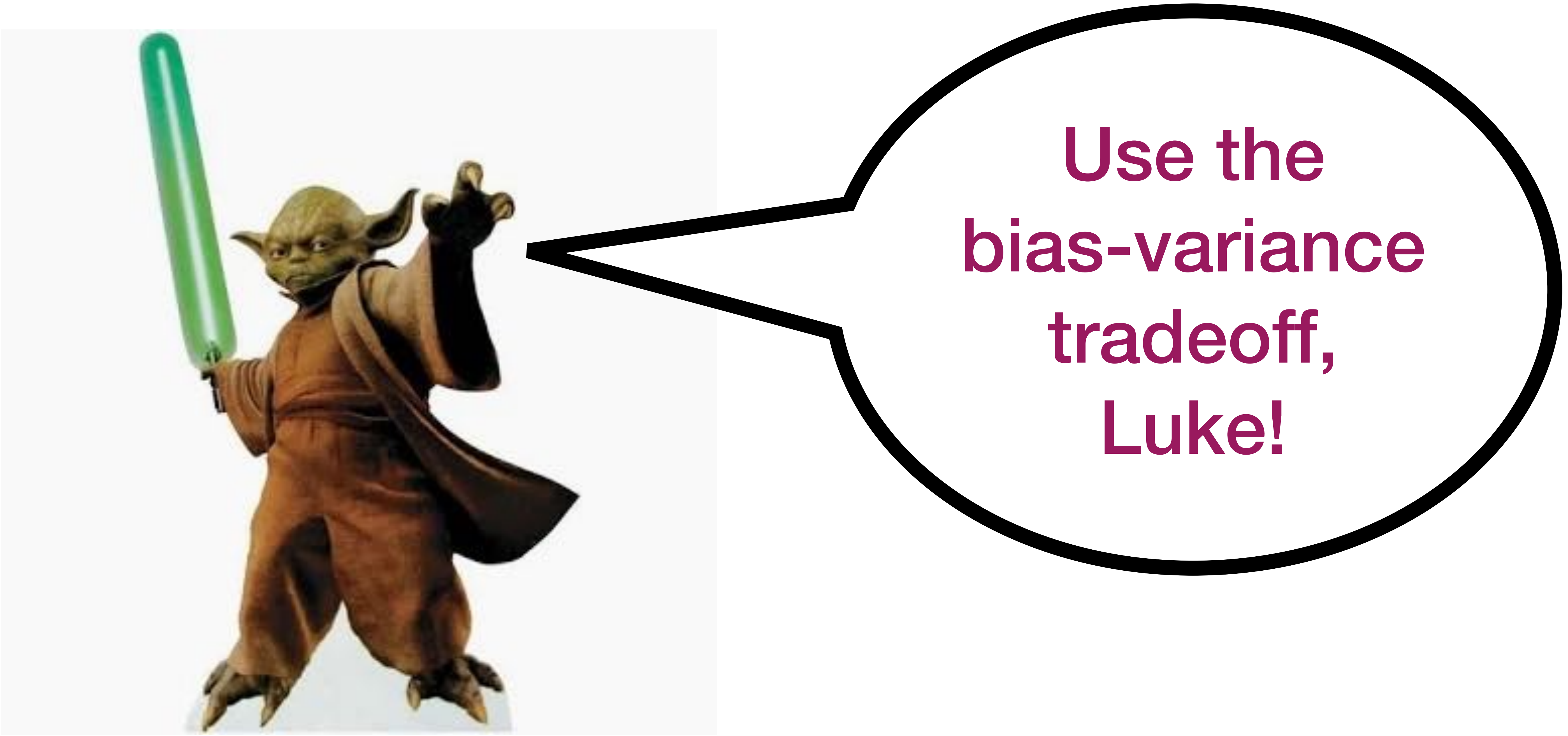
Inputs m-dim: $\overrightarrow{R}_t$

Model: $\overrightarrow{F}_t = A^{(z_t)} \overrightarrow{R}_t + b^{(z_t)} + \vec{\epsilon}_t$

Parameters with K hidden states
observation weights $A^{(k)}$, biases $b^{(k)}$,
noise cov $\Sigma^{(k)}$

Transition probabilities $T_{kl} = p(z_t = l \mid z_{t-1} = k)$
Initial probabilities: $\pi_k = p(z_0 = k)$

How do we choose the number of states?



Model selection: Pick # states

Observations n-dim: $\overrightarrow{F}_t$

Inputs m-dim: $\overrightarrow{R}_t$

Model: $\overrightarrow{F}_t = A^{(z_t)} \overrightarrow{R}_t + b^{(z_t)} + \vec{\epsilon}_t$

Parameters with K hidden states
observation weights $A^{(k)}$, biases $b^{(k)}$,
noise cov $\Sigma^{(k)}$

Transition probabilities $T_{kl} = p(z_t = l \,|\, z_{t-1} = k)$
Initial probabilities: $\pi_k = p(z_0 = k)$

Too few states —> underfit

Too many states —> overfit

Optimal # states —> best fit

Use the bias-variance tradeoff, Luke!

Model selection: xval

Observations n-dim: $\vec{F}_t$

Inputs m-dim: $\vec{R}_t$

Model: $\vec{F}_t = A^{(z_t)}\,\vec{R}_t + b^{(z_t)} + \vec{\epsilon}_t$

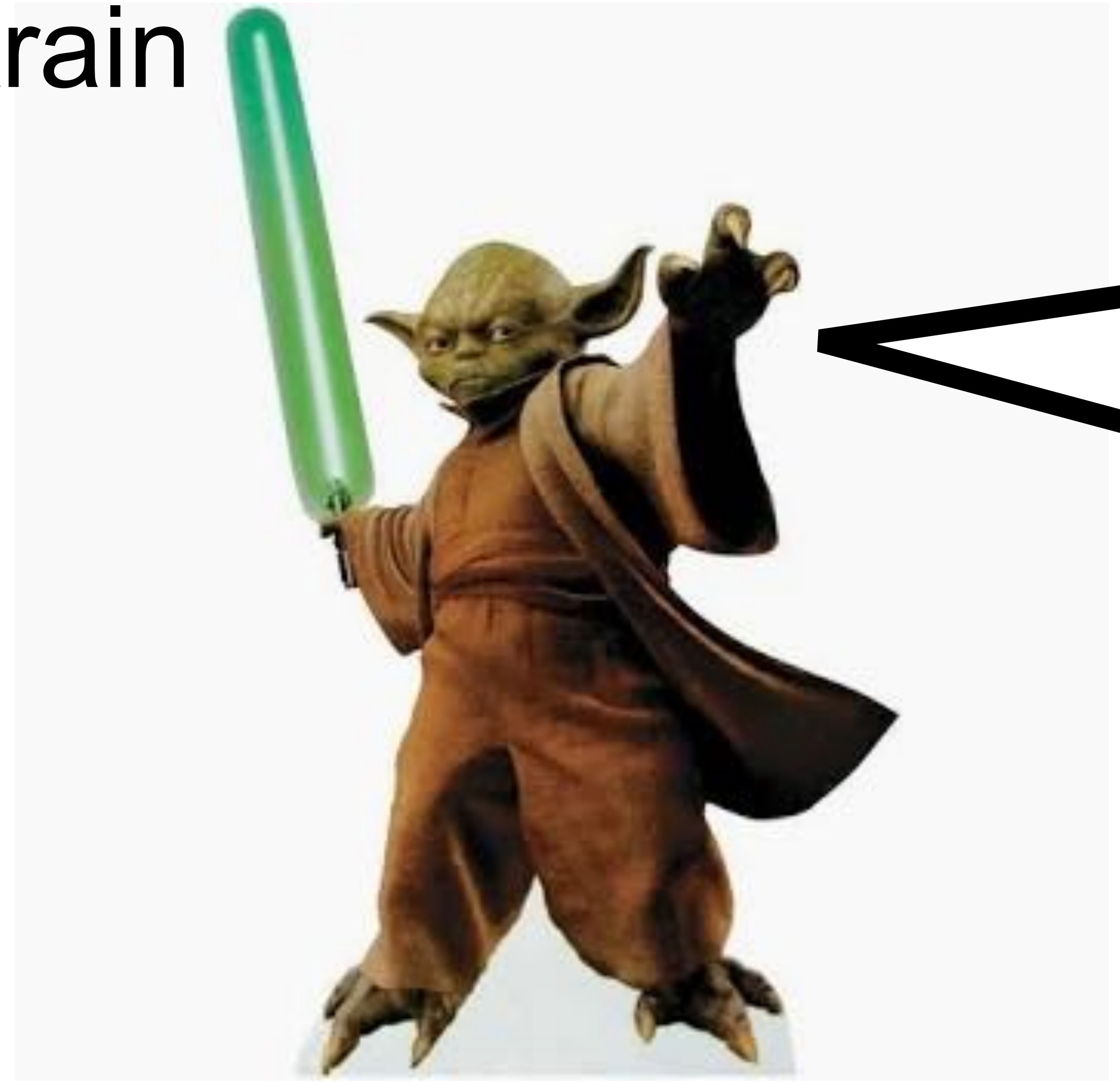
Parameters with K hidden states
observation weights $A^{(k)}$, biases $b^{(k)}$,
noise cov $\Sigma^{(k)}$

Transition probabilities $T_{kl} = p(z_t = l \,|\, z_{t-1} = k)$
Initial probabilities: $\pi_k = p(z_0 = k)$

1. Split data into training set S_train and test set S_test

2. Fit HMM with K states to S_train, estimate likelihood of test set LL(S_test|S_train)

3. Increase K and find best, or at least elbow



Use the bias-variance tradeoff, Luke!

Model selection: xval

Observations n-dim: $\overrightarrow{F}_t$

Inputs m-dim: $\overrightarrow{R}_t$

Model: $\overrightarrow{F}_t = A^{(z_t)} \overrightarrow{R}_t + b^{(z_t)} + \vec{\epsilon}_t$

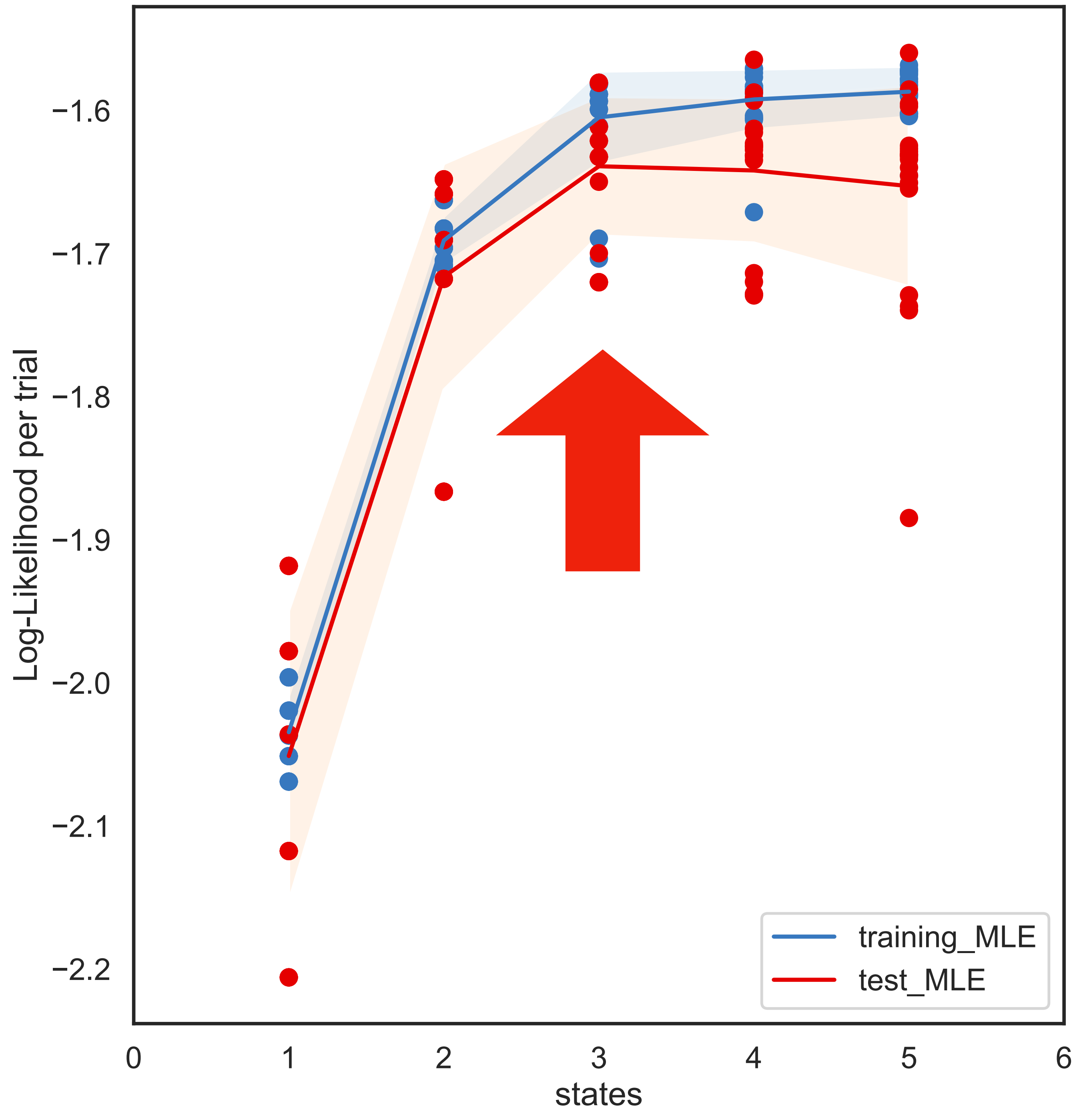
Parameters with K hidden states
observation weights $A^{(k)}$, biases $b^{(k)}$,
noise cov $\Sigma^{(k)}$

Transition probabilities $T_{kl} = p(z_t = l \mid z_{t-1} = k)$
Initial probabilities: $\pi_k = p(z_0 = k)$

1. Split data into training set S_train and test set S_test

2. Fit HMM with K states to S_train, estimate likelihood of test set LL(S_test|S_train)

3. Increase K and find best, or at least elbow



Model selection: BIC/AIC

Observations n-dim: $\overrightarrow{F}_t$

Inputs m-dim: $\overrightarrow{R}_t$

Model: $\overrightarrow{F}_t = A^{(z_t)} \overrightarrow{R}_t + b^{(z_t)} + \vec{\epsilon}_t$

Parameters with K hidden states observation weights $A^{(k)}$, biases $b^{(k)}$, noise cov $\Sigma^{(k)}$

Transition probabilities $T_{kl} = p(z_t = l \mid z_{t-1} = k)$

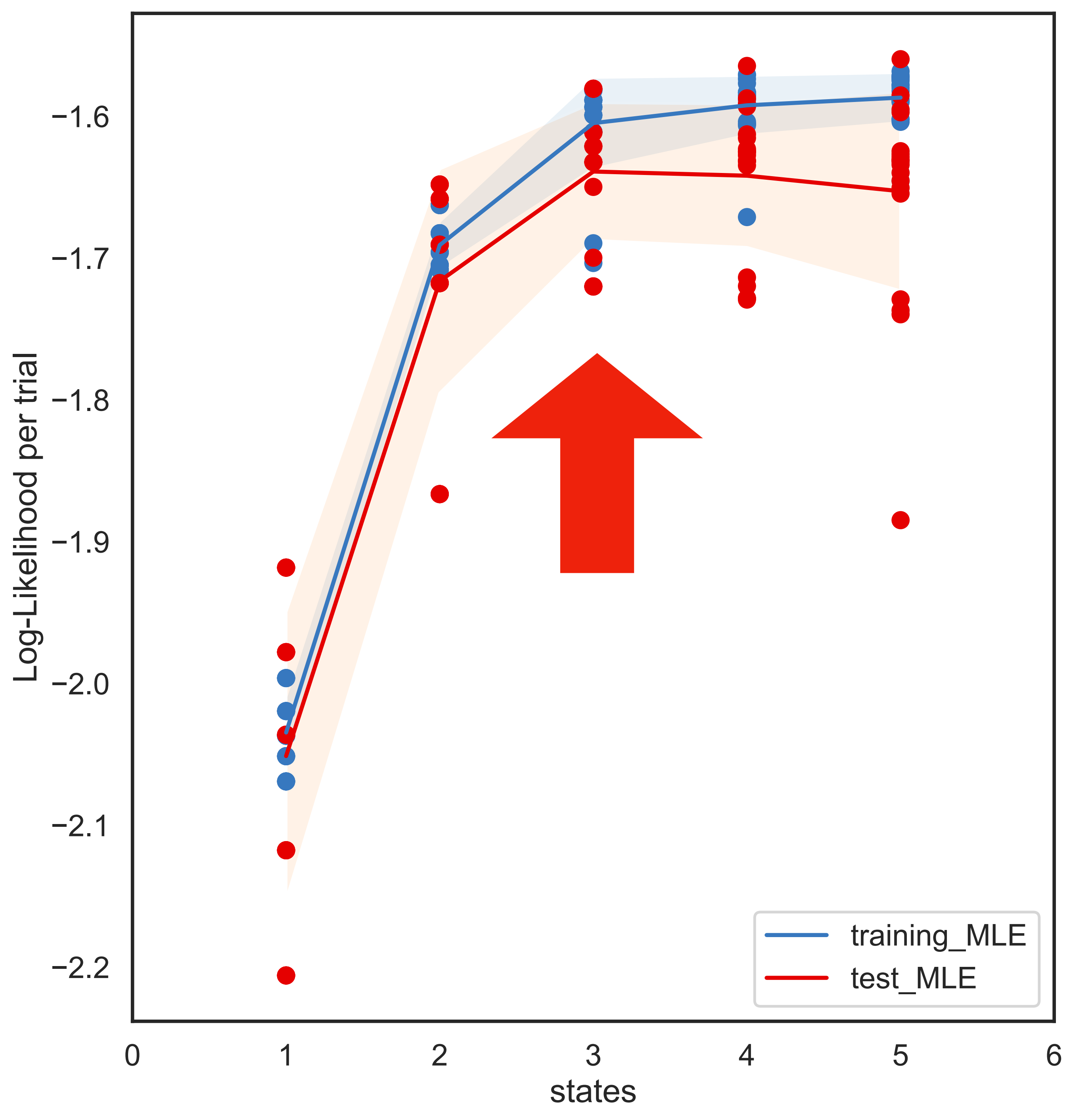
Initial probabilities: $\pi_k = p(z_0 = k)$

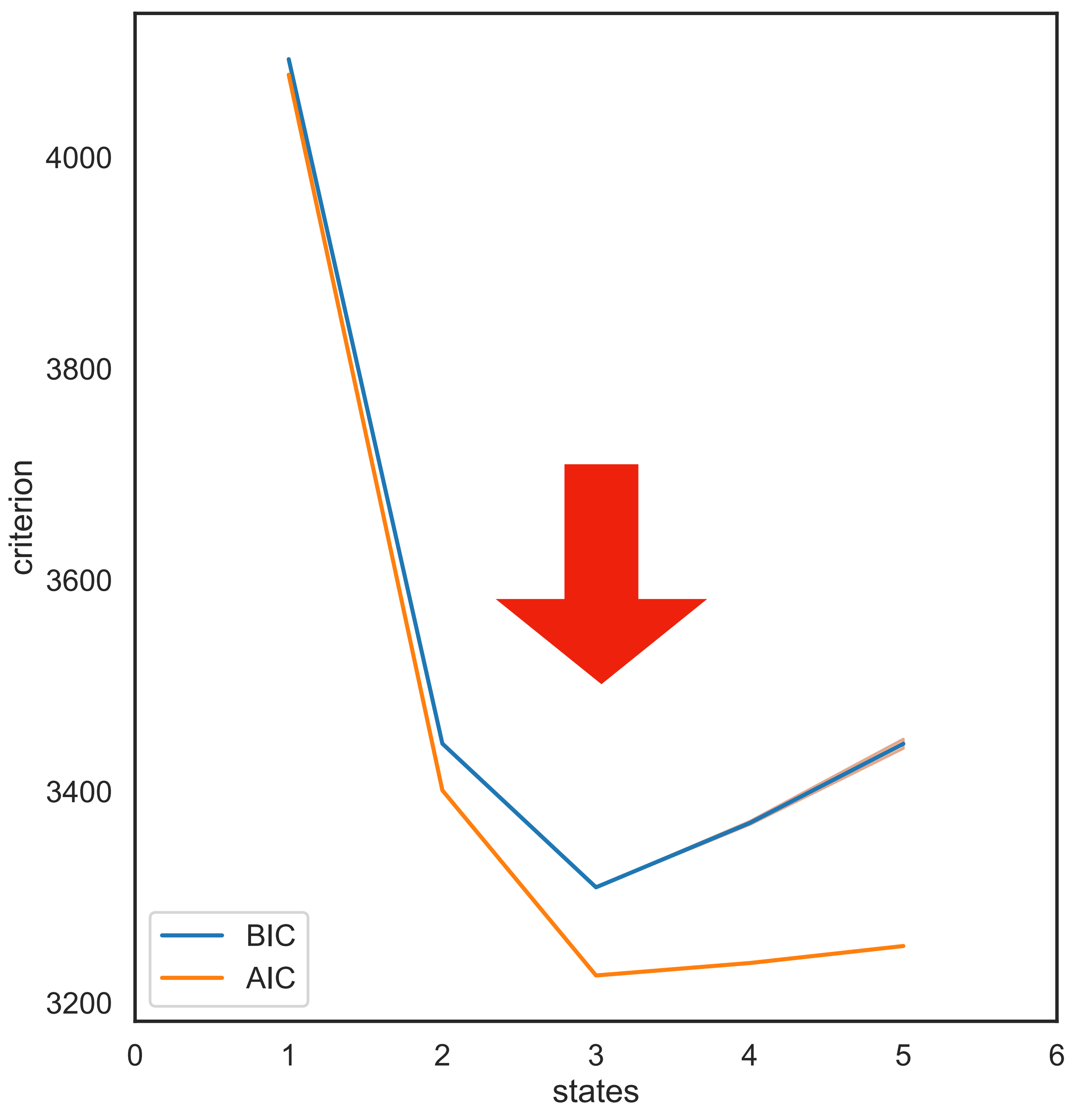
Penalize fit by # of parameters P:

BIC = - 2*LL + P*log(T)

AIC = - 2(LL + 2*P

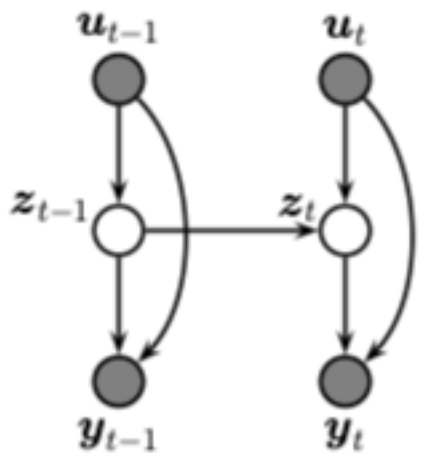
P = K(K-1) + K-1 + K*n*(m+1) + K*n





References

- State space model packages
  - https://github.com/lindermanlab/ssm
    - https://github.com/mazzulab/ssm
  - https://github.com/probml/dynamax



- Freely moving behavior
  - Wiltschko, Alexander B., et al. "Mapping sub-second structure in mouse behavior." (2015).
  - Findley, Teresa M., et al. "Sniff-synchronized, gradient-guided olfactory search by freely moving mice." (2021).

- Decision Making
  - Ashwood, Zoe C., et al. "Mice alternate between discrete strategies during perceptual decision-making." (2022).
  - Cazettes, Fanny, et al. "A reservoir of foraging decision variables in the mouse brain." (2023).