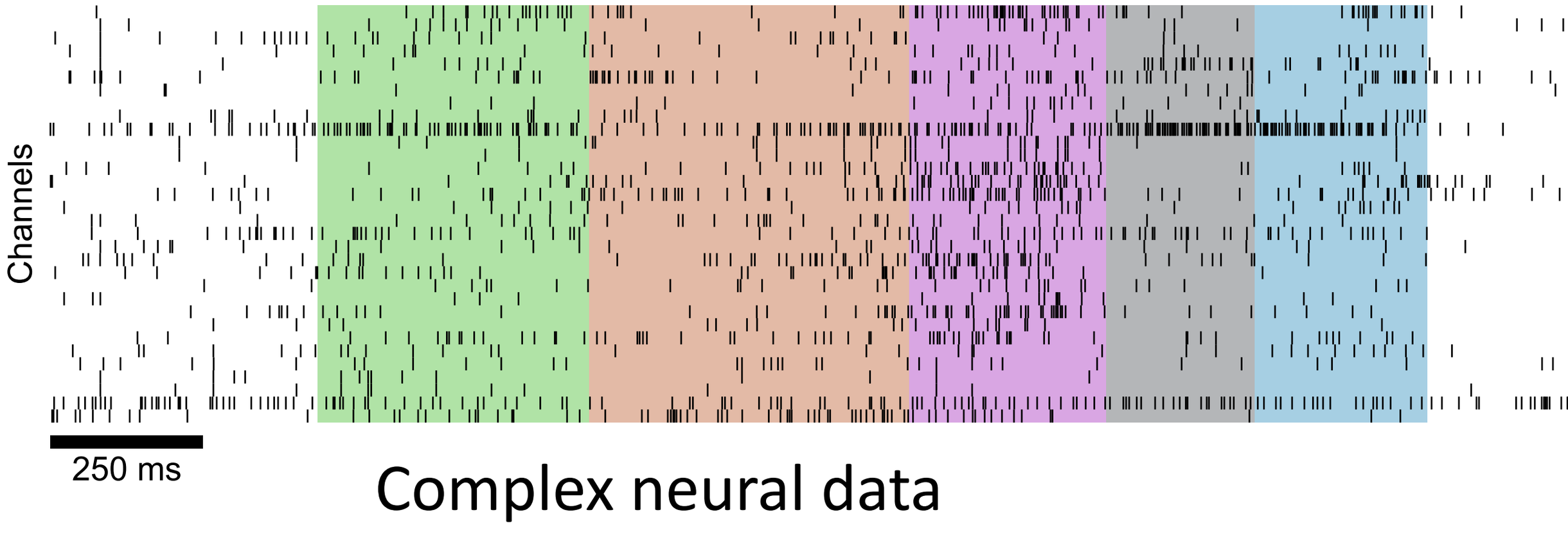
Computation through dynamics

Using recurrent neural networks to unveil mechanism in neural circuits

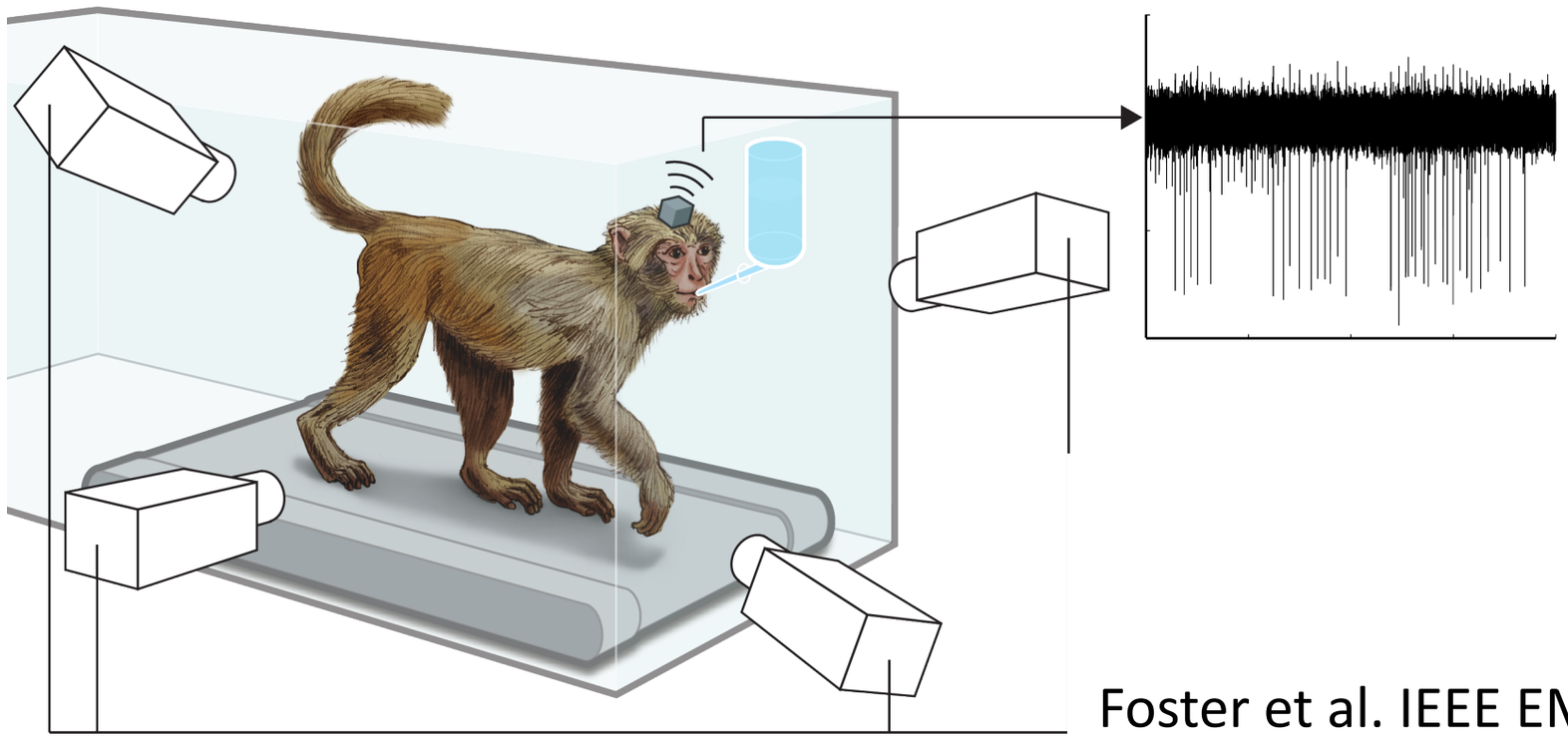
David Sussillo
with Valerio Mante and Bill Newsome

Table of contents

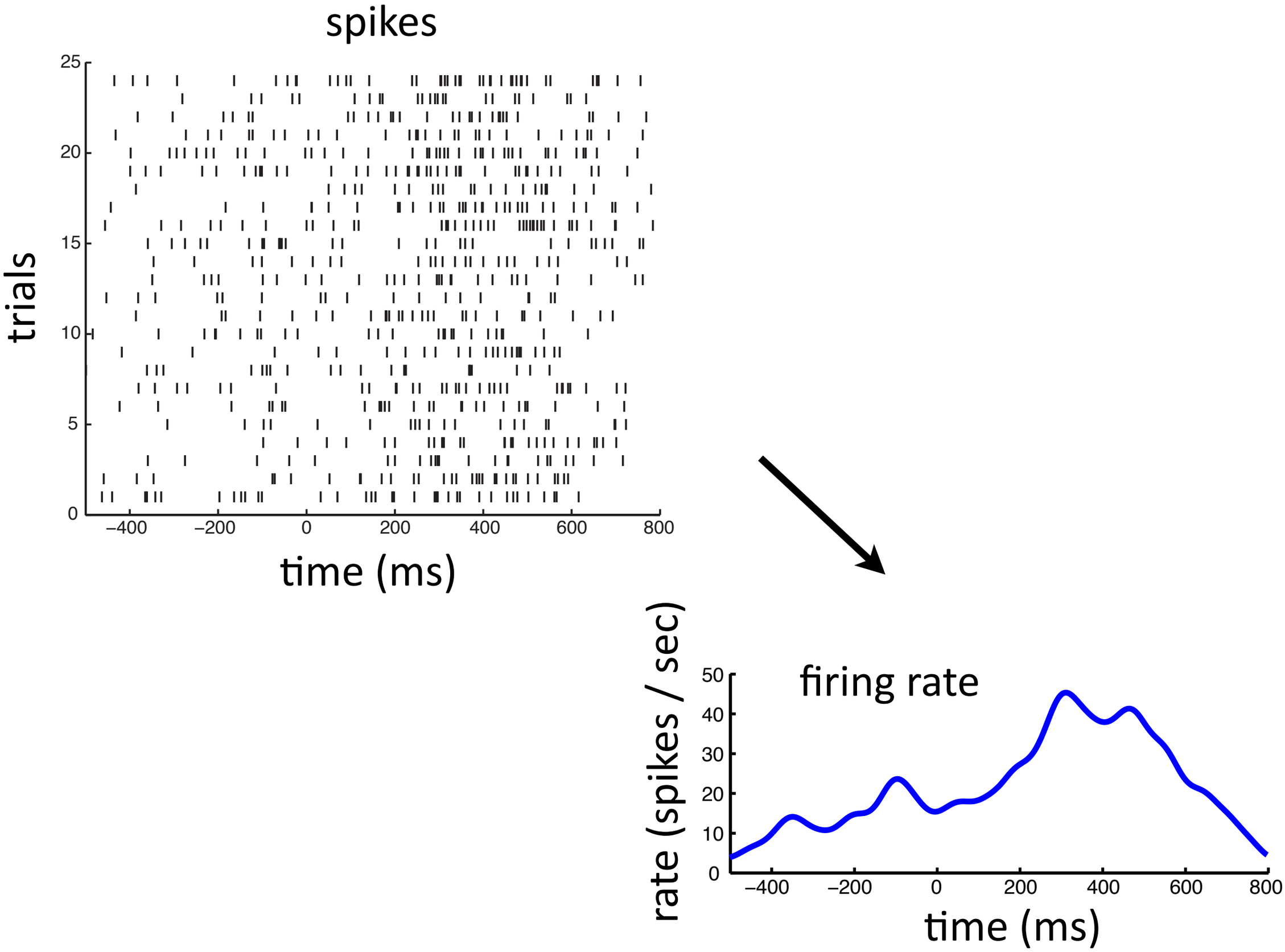


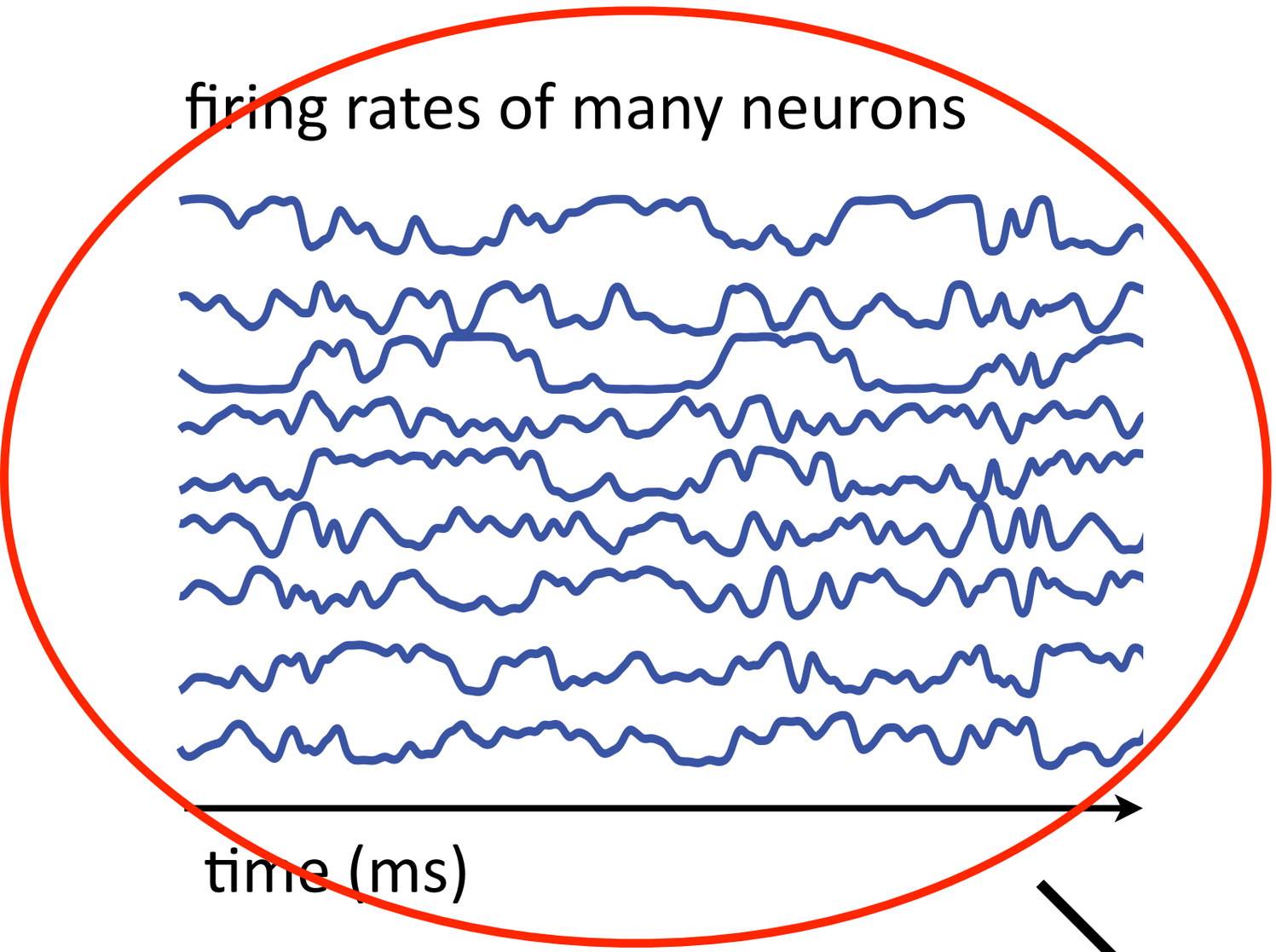
Complex neural data

Complex behavior



Foster et al. IEEE EMBS 2012

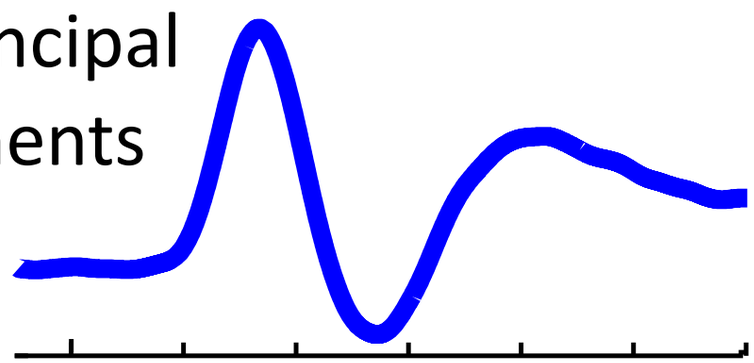

spikes
trials
25
20
15
10
5
0
−400
−200
0
200
400
600
800
time (ms)
firing rate
rate (spikes / sec)
50
40
30
20
10
0
−400
−200
0
200
400
600
800
time (ms)



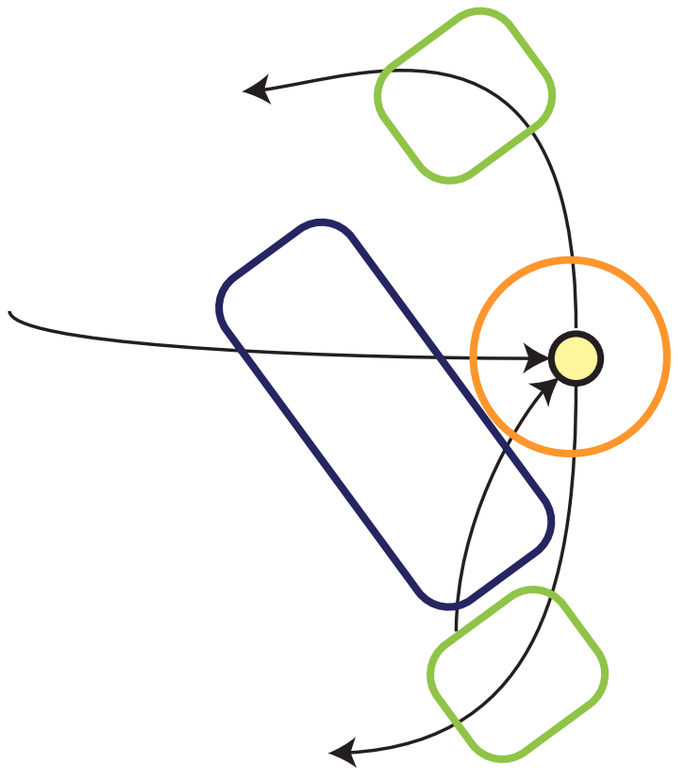
What are the biophysical correlates of these variables?

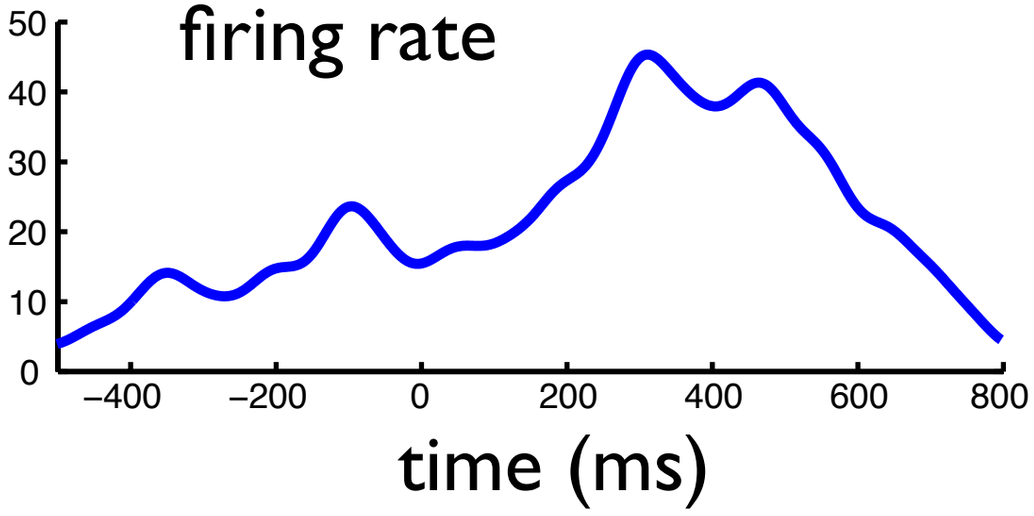
I work at the level of rates because we can make networks do interesting computations!

a few principal components



Recurrent Neural Networks (RNNs)

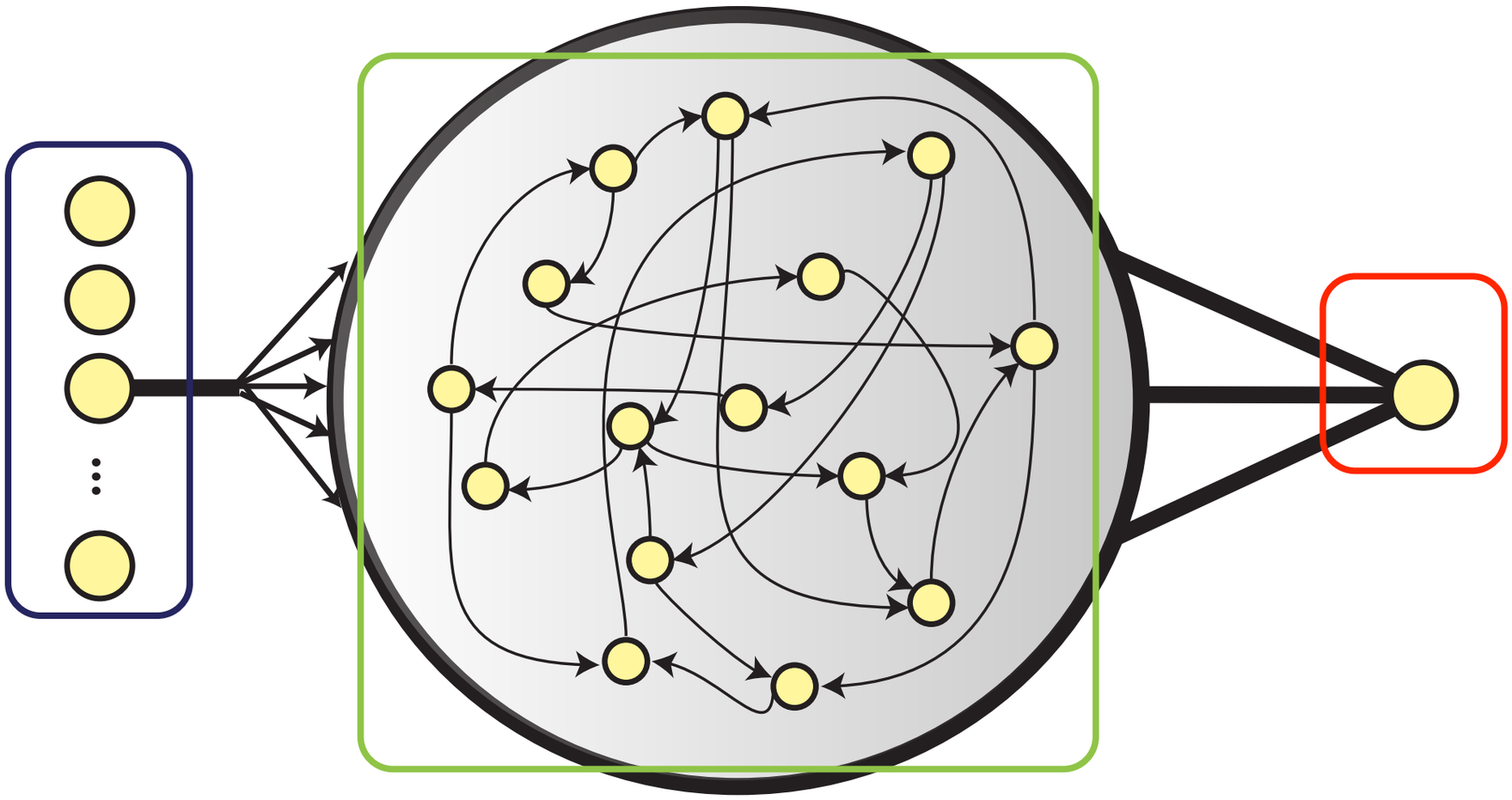


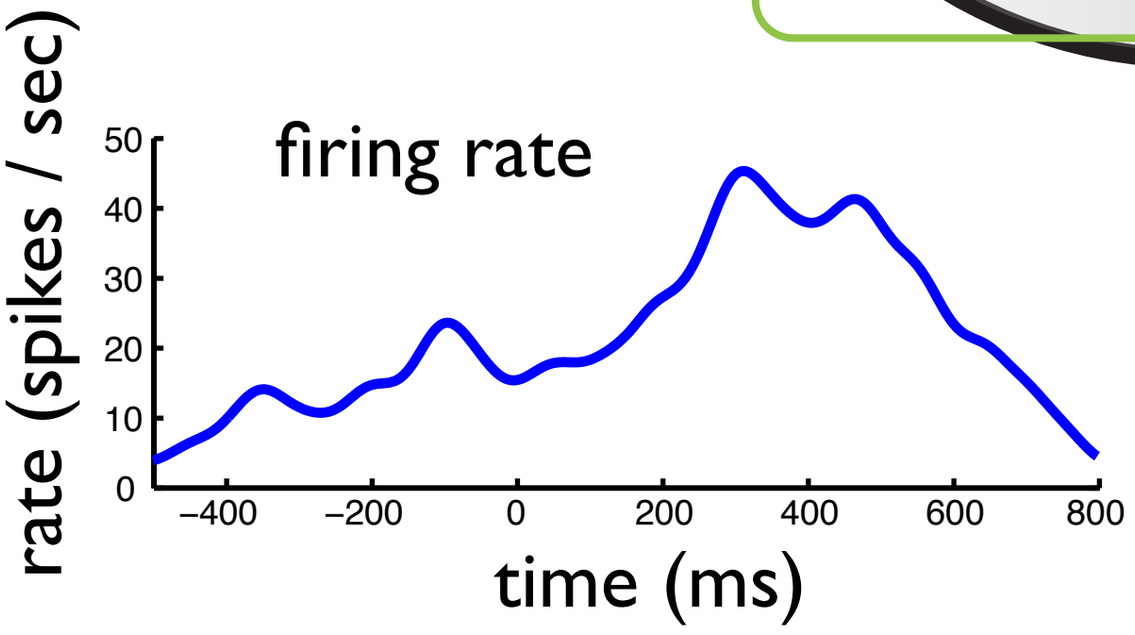


$$\tau \dot{x} = -x + b_i u_i + b_j u_j$$

$$r = [\tanh(x)]^+$$

Recurrent Neural Networks (RNNs)





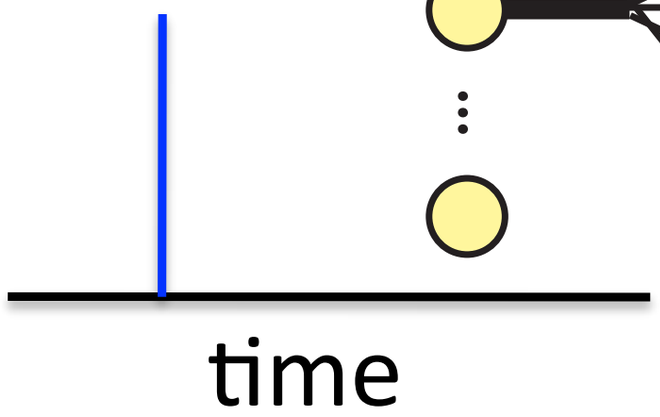
$$\tau \dot{\mathbf{x}} = -\mathbf{x} + \mathbf{J}\mathbf{r} + \mathbf{B}\mathbf{u}$$

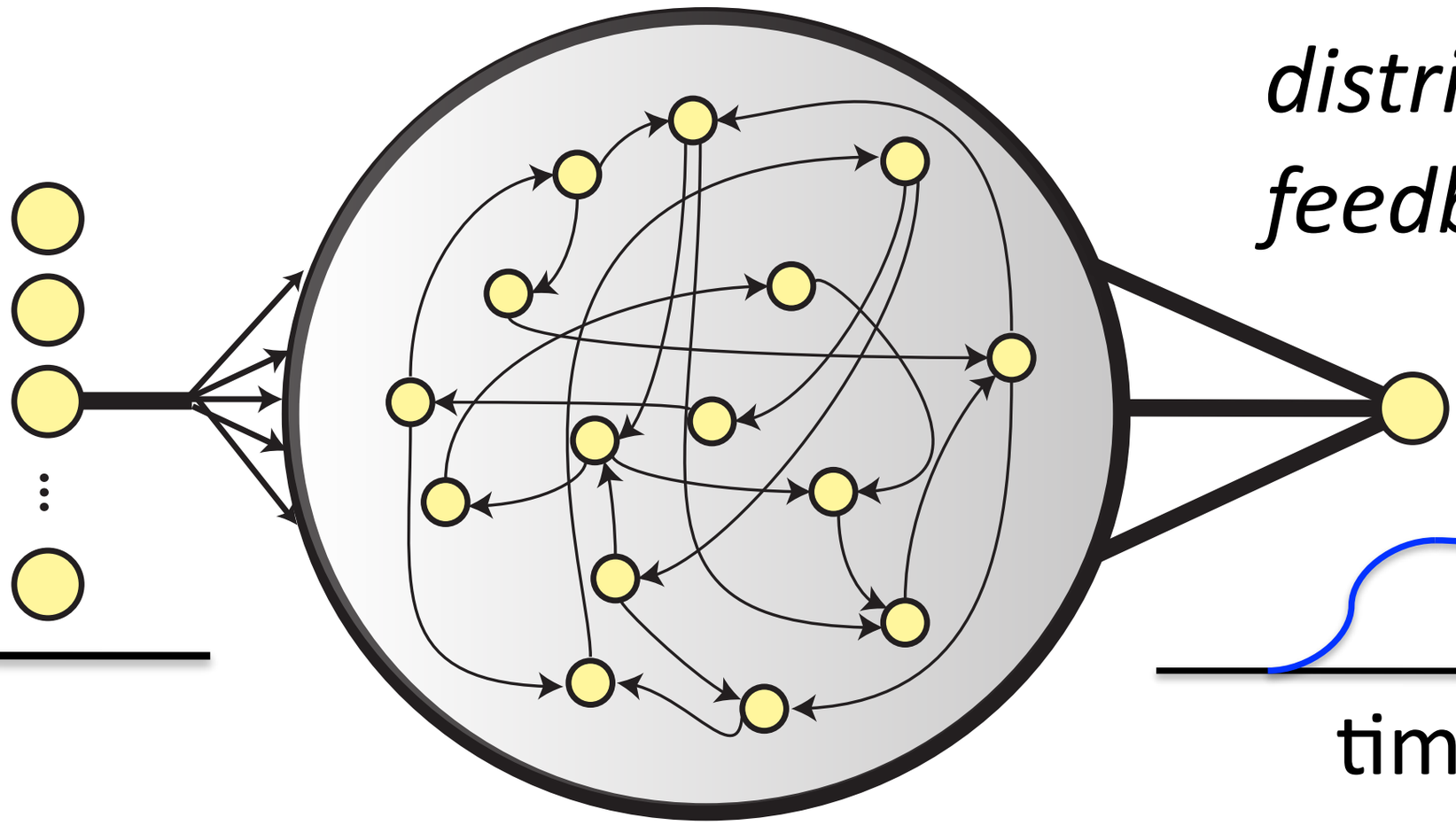
$$\mathbf{r} = [\tanh(\mathbf{x})]^+$$

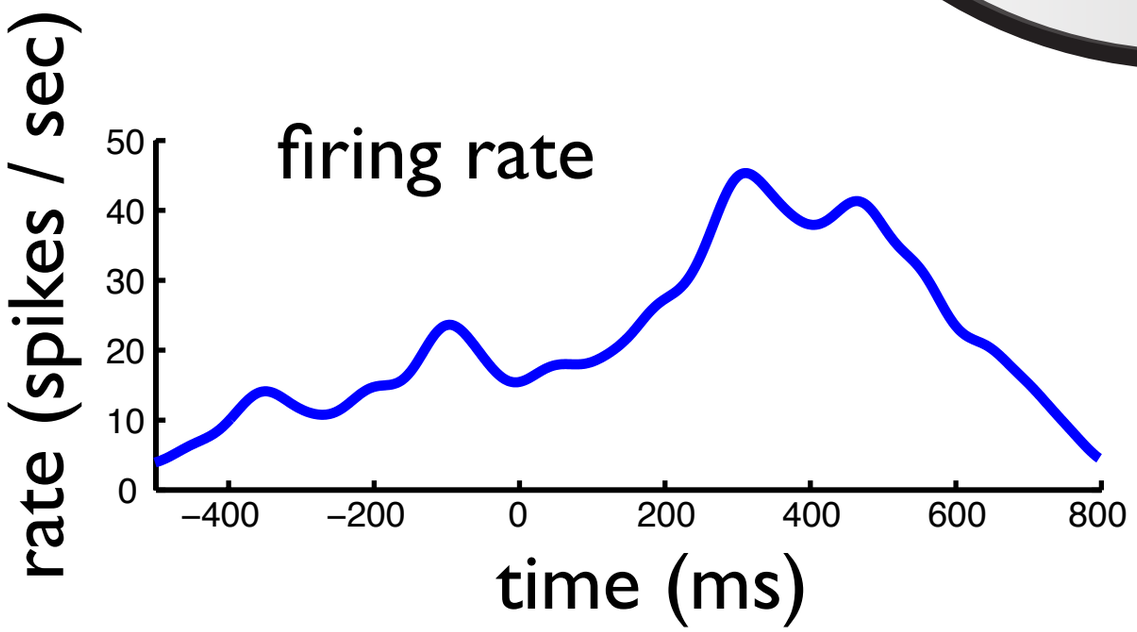
$$\mathbf{z} = \mathbf{W}\mathbf{r}$$

Recurrent Neural Networks (RNNs)

nonlinear
distributed
feedback





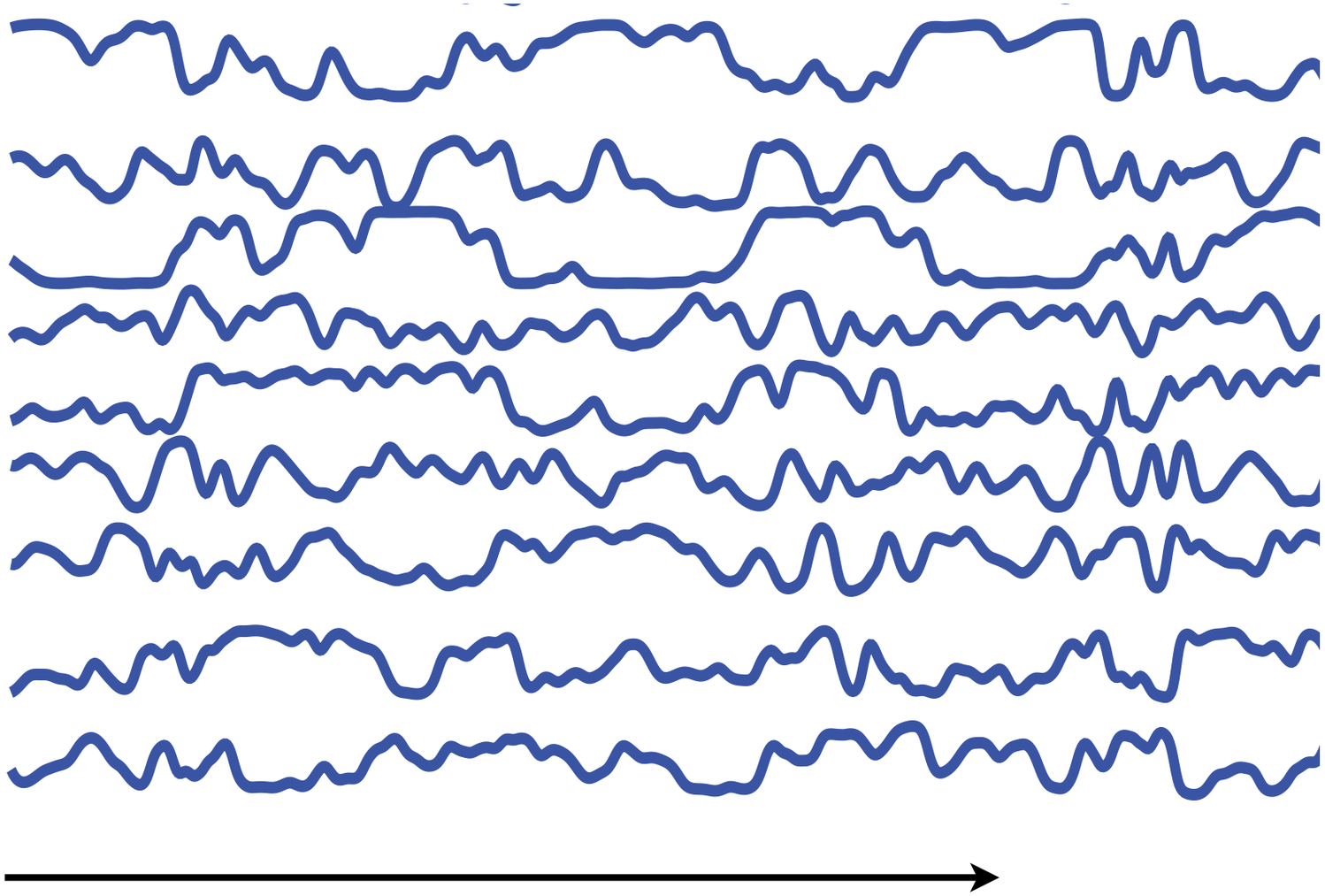


$$\tau\dot{\mathbf{x}} = -\mathbf{x} + \mathbf{J}\mathbf{r} + \mathbf{B}\mathbf{u}$$

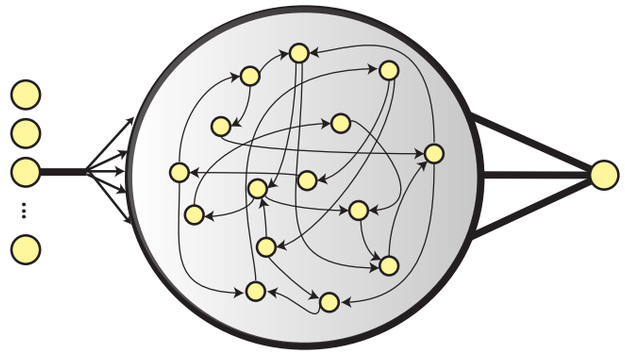
$$\mathbf{r} = [\tanh(\mathbf{x})]^+$$

$$\mathbf{z} = \mathbf{W}\mathbf{r}$$

Dynamics in RNNs (Spontaneous Activity)



time (ms)



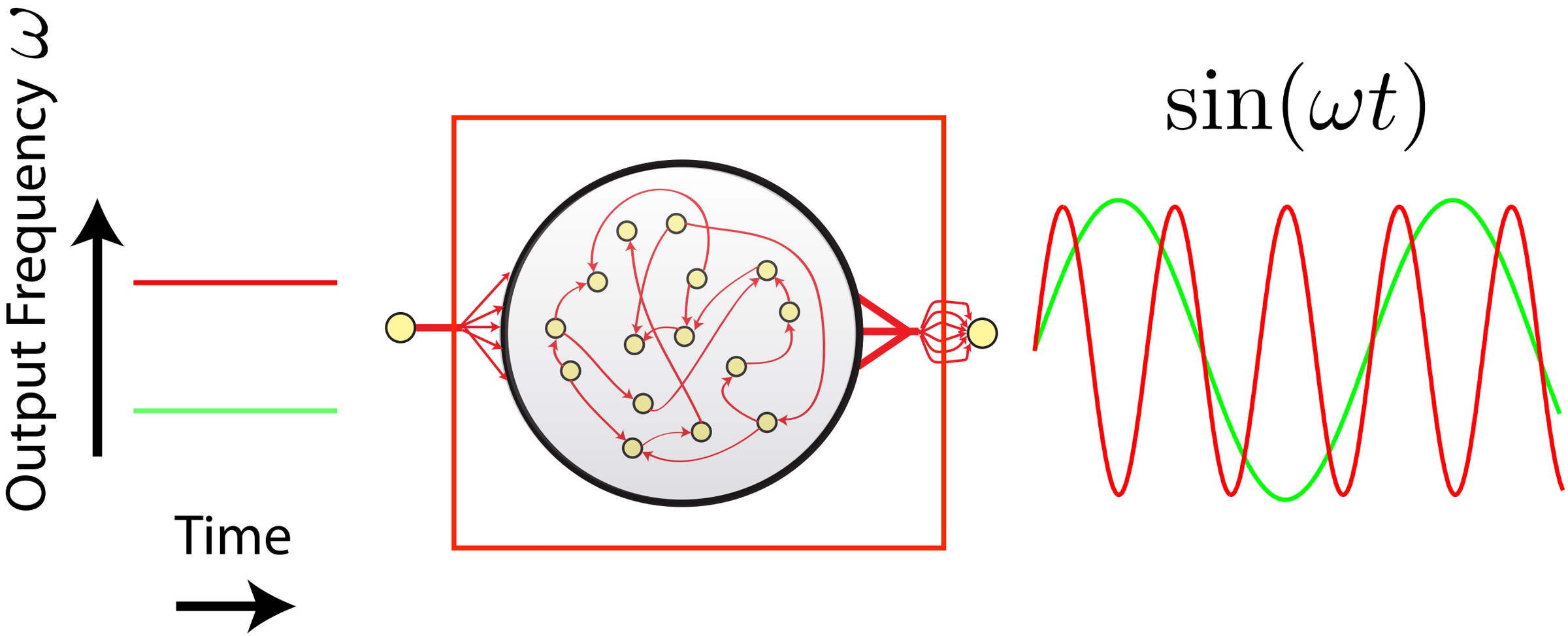
Sompolinsky et al., PRL 1988
Rajan et al., PRE 2010

Tools to understand how RNNs work

with Omri Barak

Sussillo* & Barak*,
Neural Computation 2013

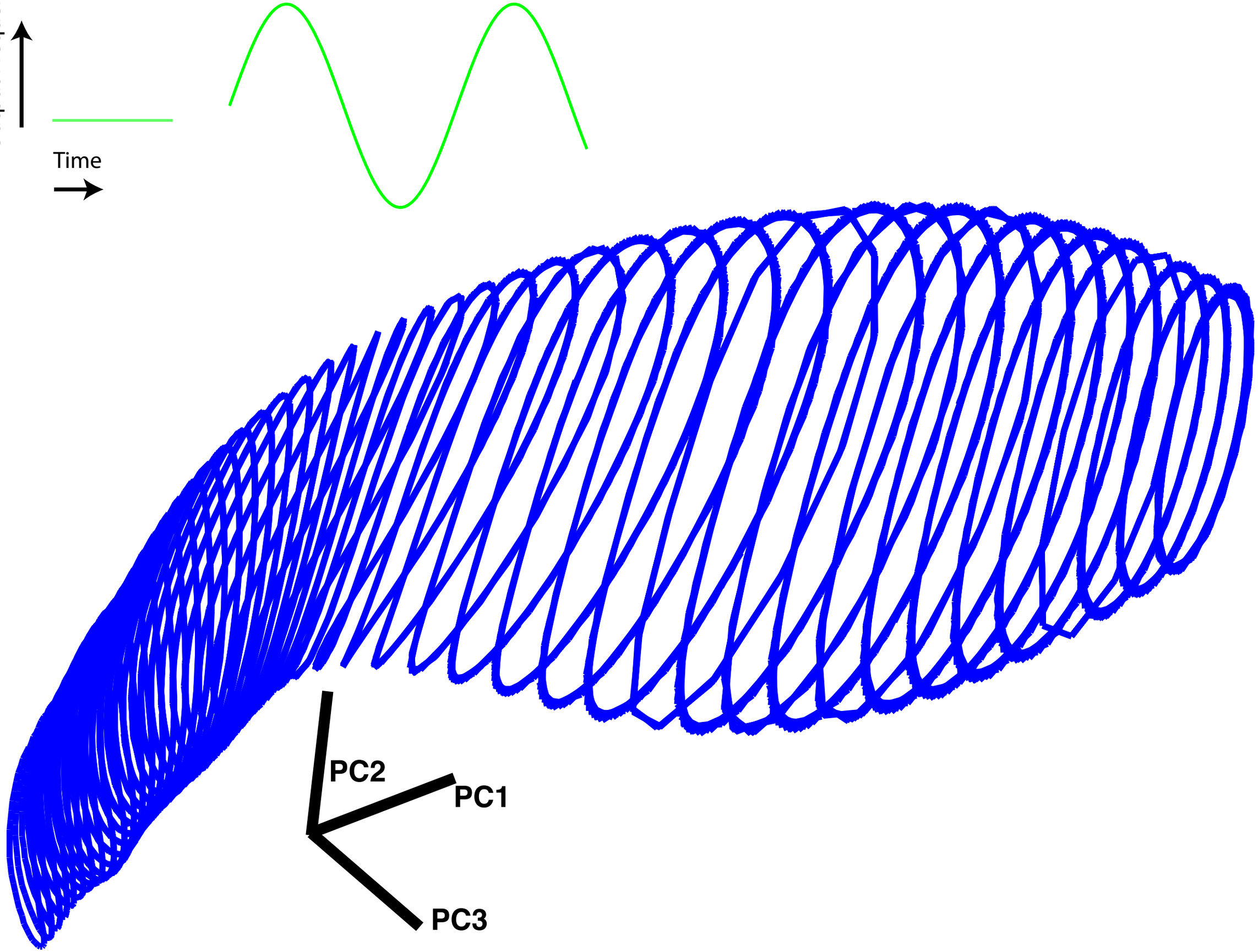
How does a sine-wave generator work?



Sussillo* & Barak*,
Neural Computation 2013
Martens & Sutskever, ICML 2011

Output Frequency

Time



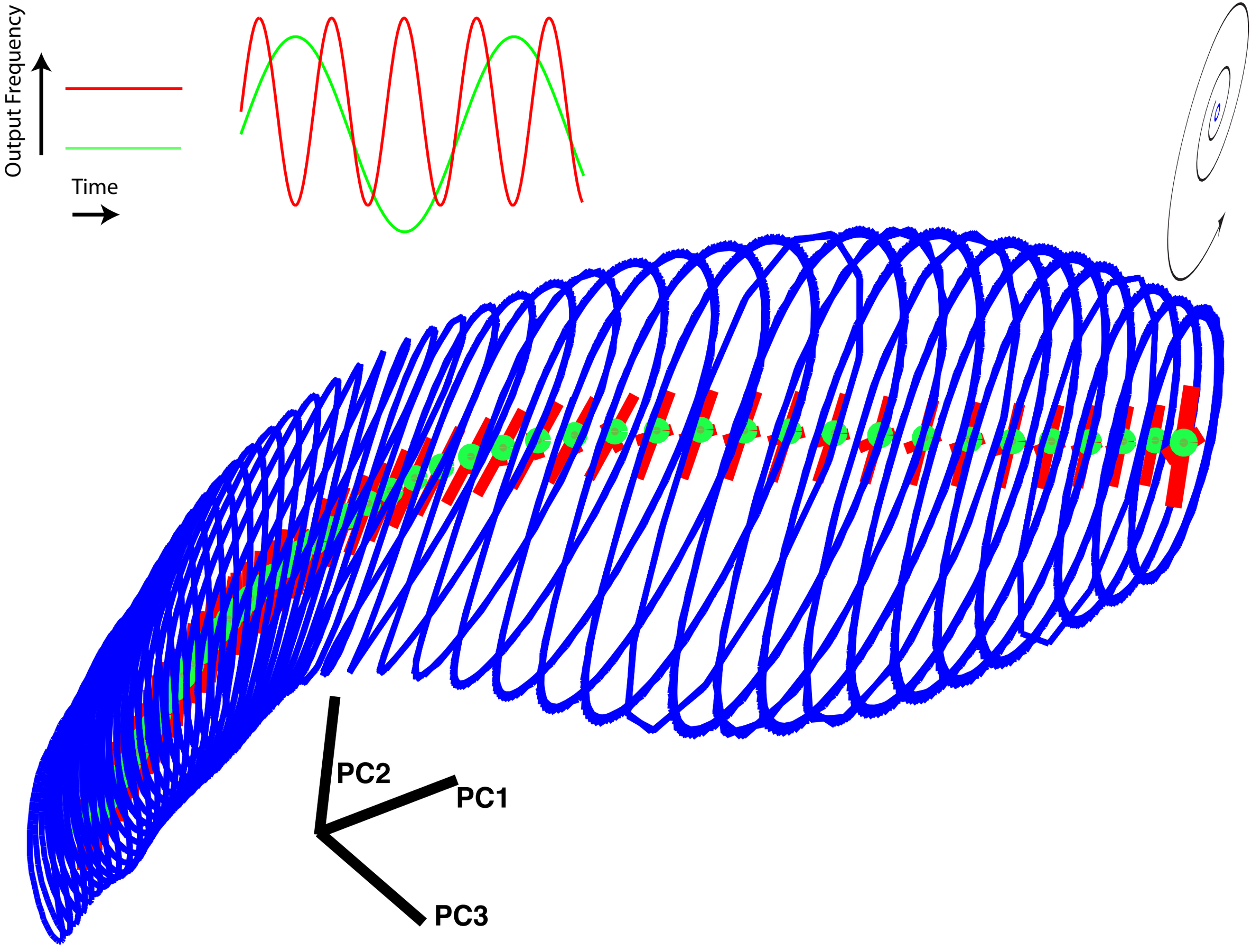
What is a fixed point?

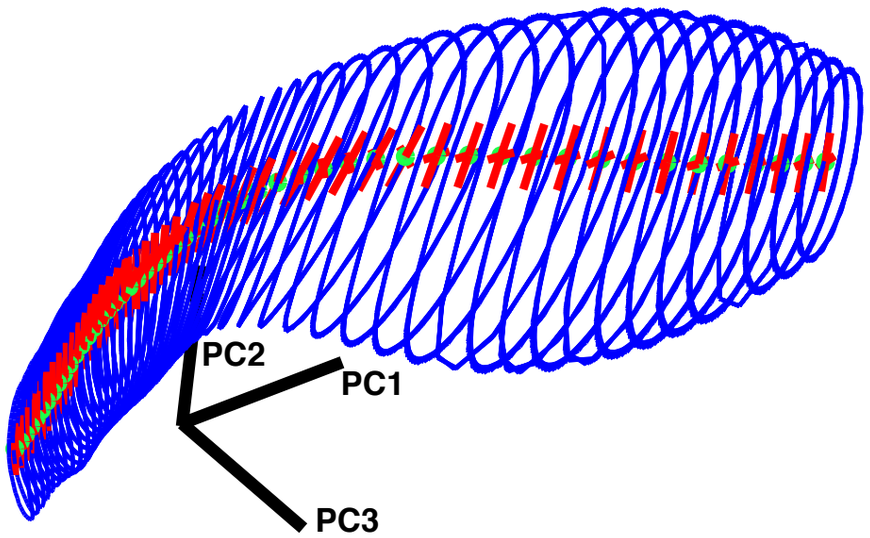
$\dot{\mathbf{x}} = \mathrm{F}(\mathbf{x})$ Any nonlinear dynamical system (e.g. neural circuit)

$\dot{\mathbf{x}} = \mathbf{0}$ Zero “motion”

Why are they important? $\dot{\mathbf{y}} = \mathbf{M}\mathbf{y}$

Sussillo* & Barak*,
Neural Computation 2013

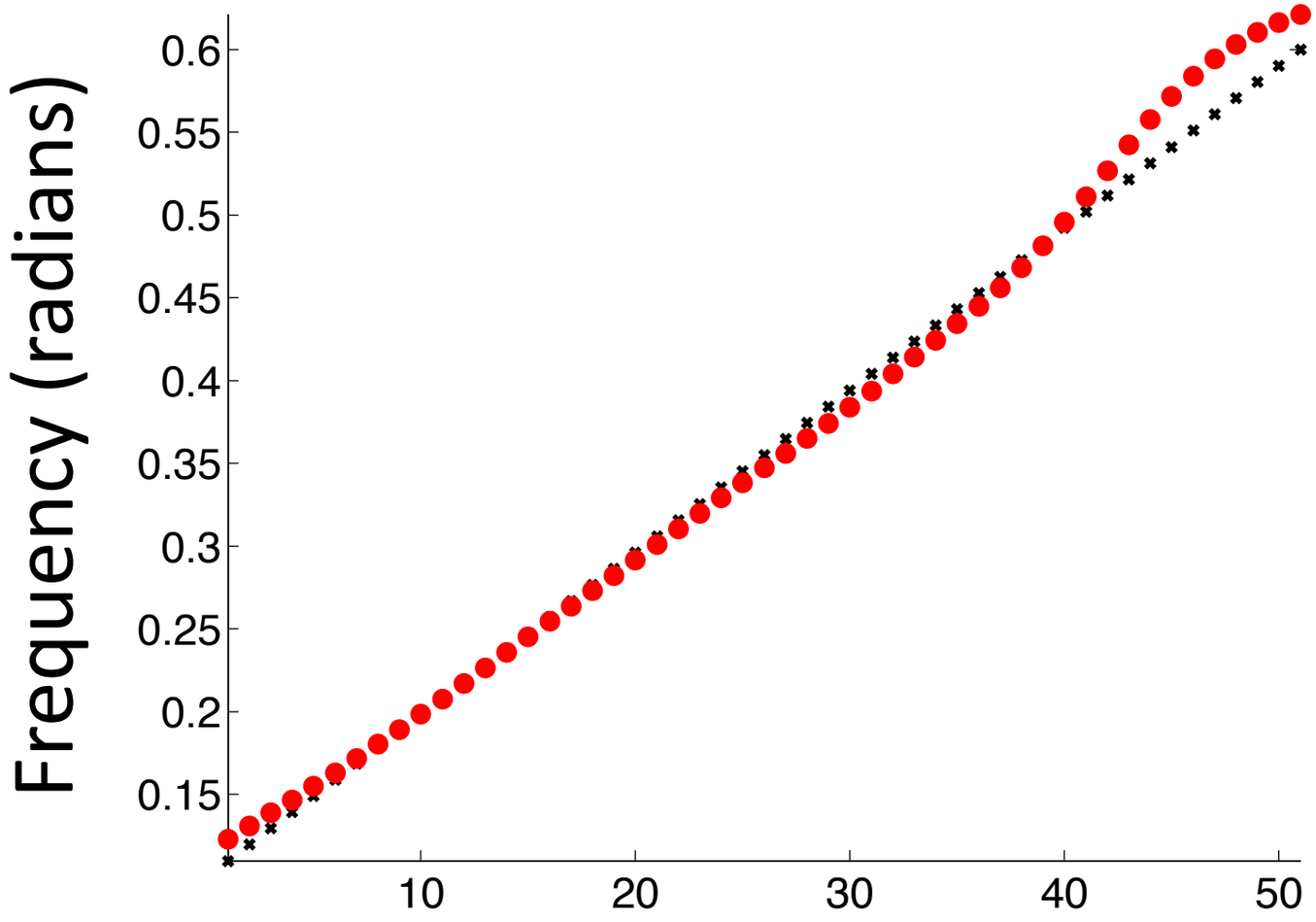

Output Frequency
Time
PC2
PC1
PC3



The linear system is a very good approximation!

Input frequency + + +

Linear system frequency • • •



Sussillo* & Barak*,
Neural Computation 2013

Conclusions from technical part

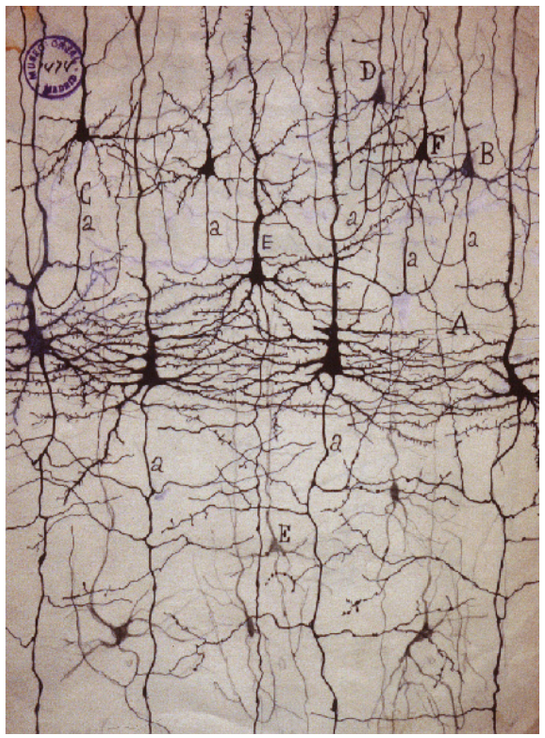
- Recurrent neural networks are a natural model class for modeling cortical phenomenon: dynamical, nonlinear, distributed.
- Recent advances have enabled the training of RNNs.
- In “simple” cases, one can understand how an RNN implements its computation in the language of dynamical systems (e.g. fixed points, saddle points, oscillations).
- One simple description of an RNN is as a bunch of linear systems tiling the state space.

Contextual decision making (data)

with Valerio Mante and Bill Newsome

Mante*, Sussillo*, Shenoy & Newsome

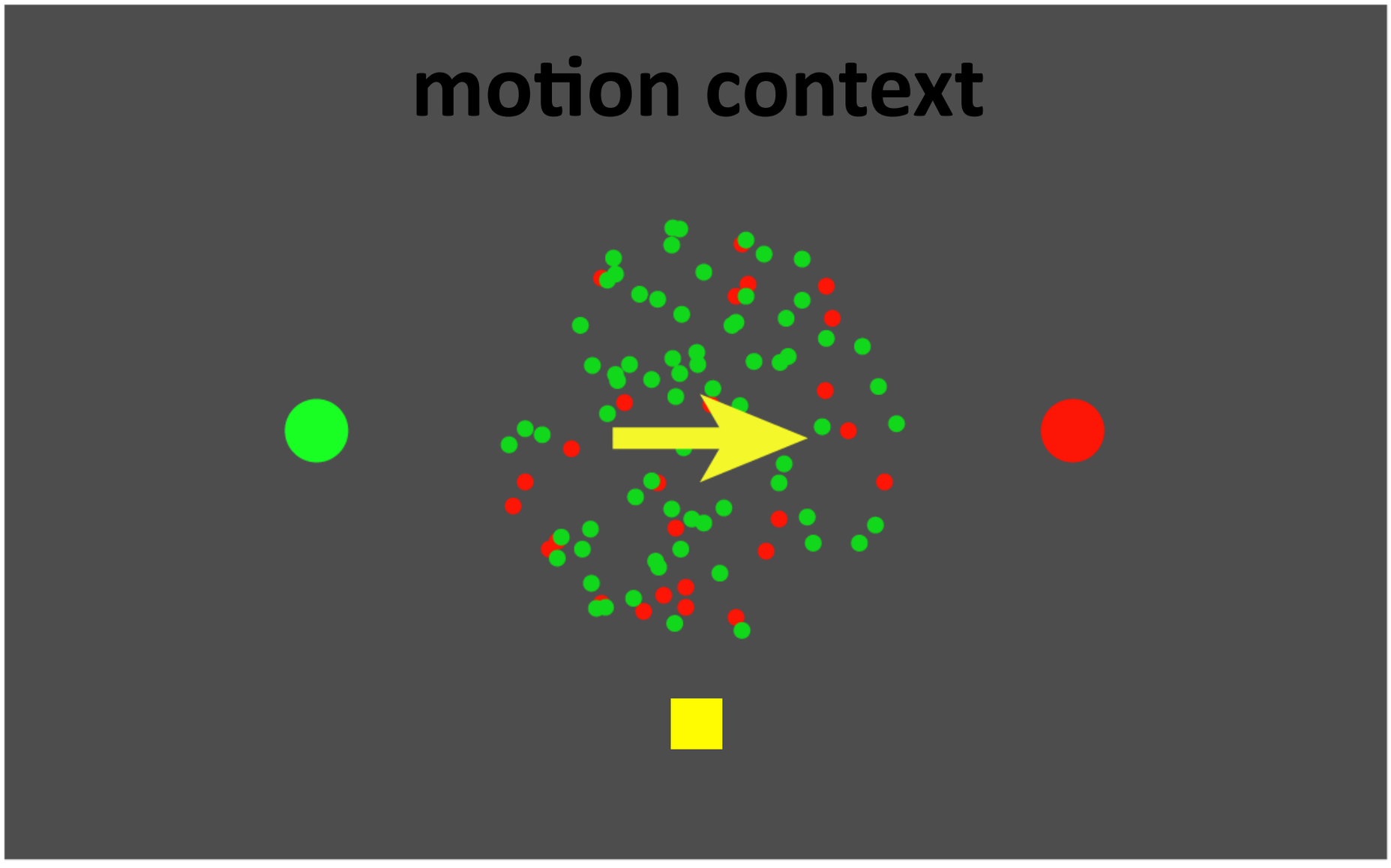
Computations in cortical circuits are flexible



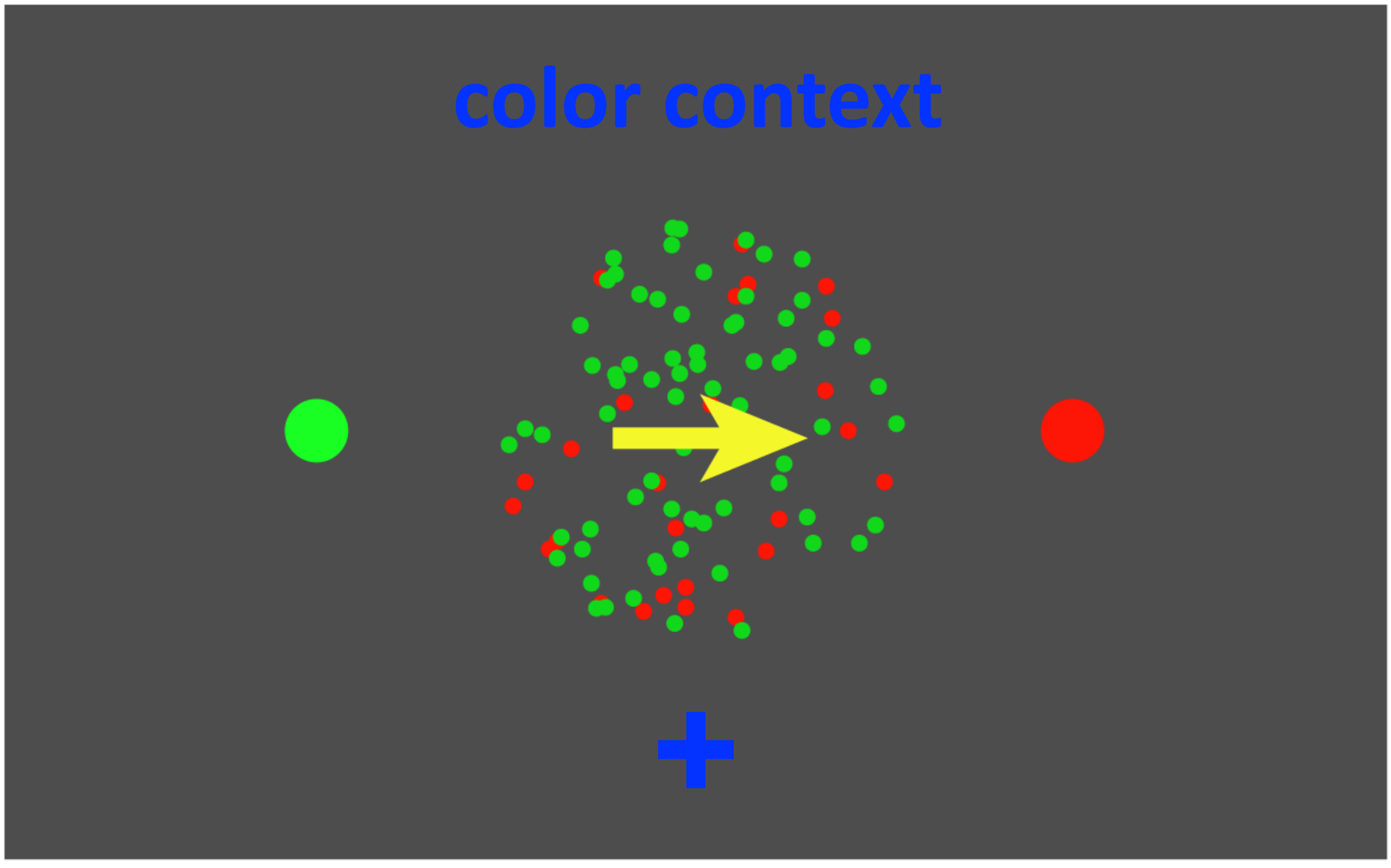
Prefrontal cortex
contributes to
flexibility of decisions

Attend relevant stimuli
Ignore irrelevant stimuli
Suppress inappropriate responses
Represent context

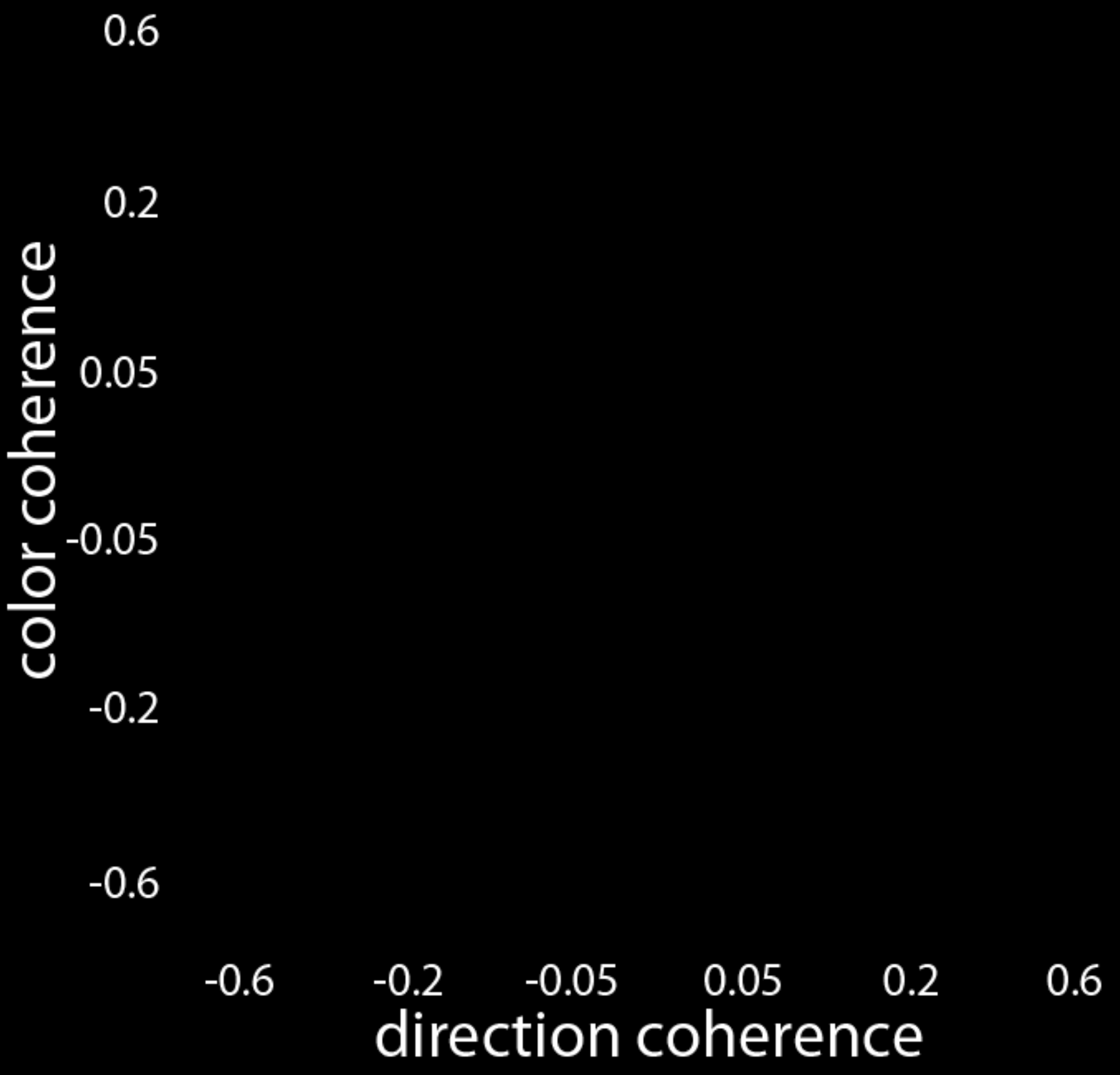
Context-dependent gating in monkeys



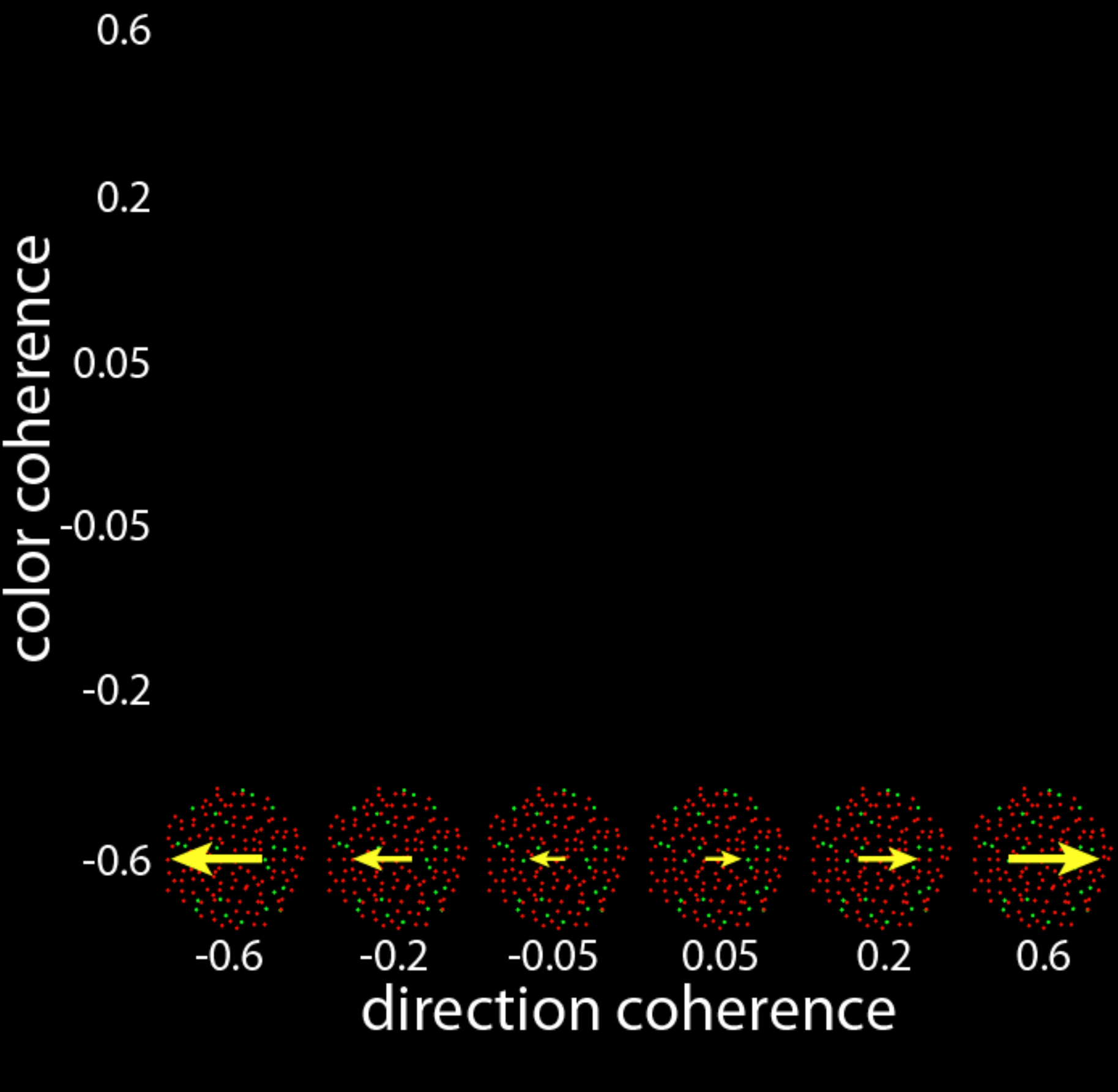
Context-dependent gating in monkeys



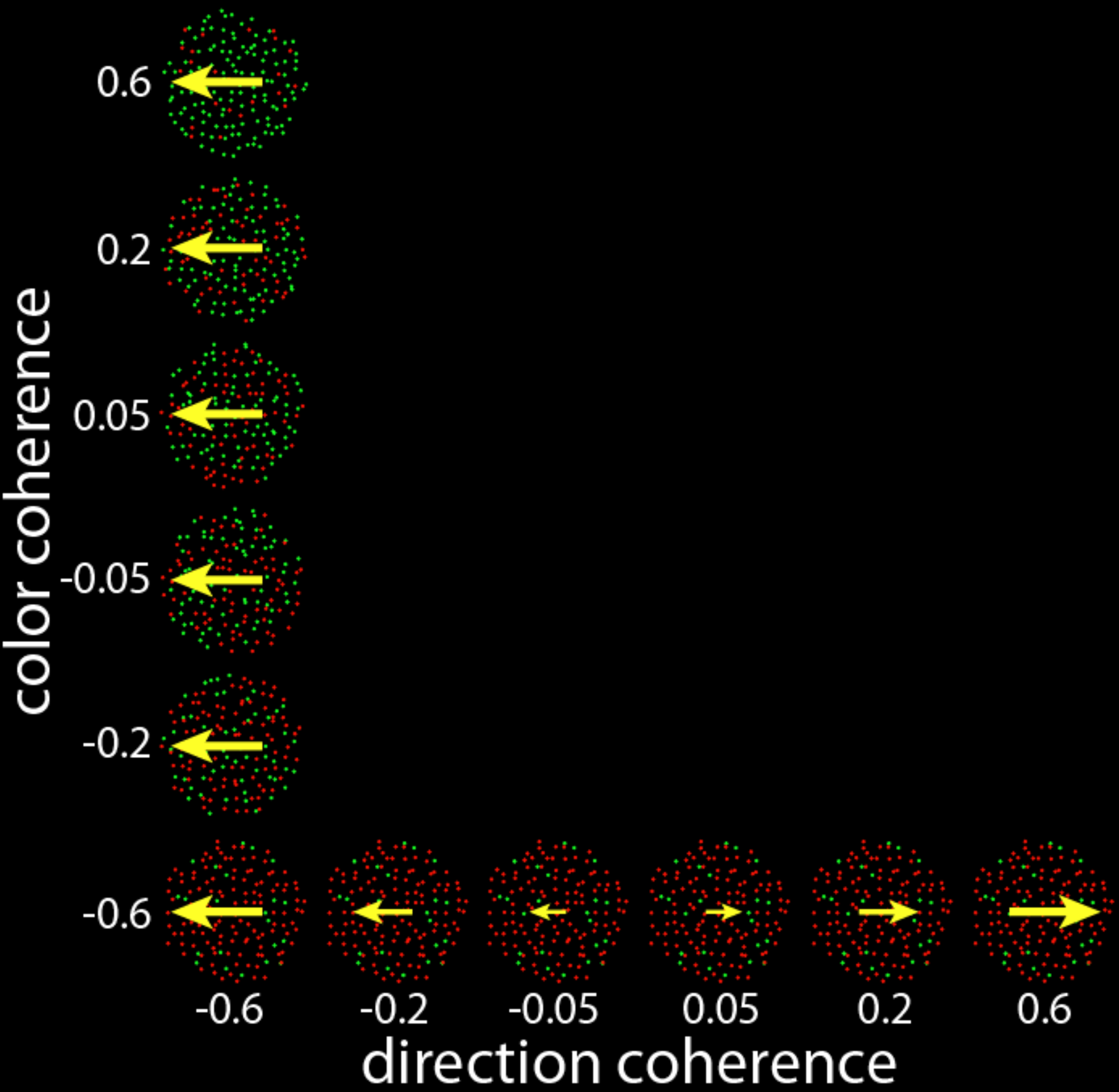
Stimuli


0.6
0.2
0.05
-0.05
-0.2
-0.6
color coherence
-0.6
-0.2
-0.05
0.05
0.2
0.6
direction coherence

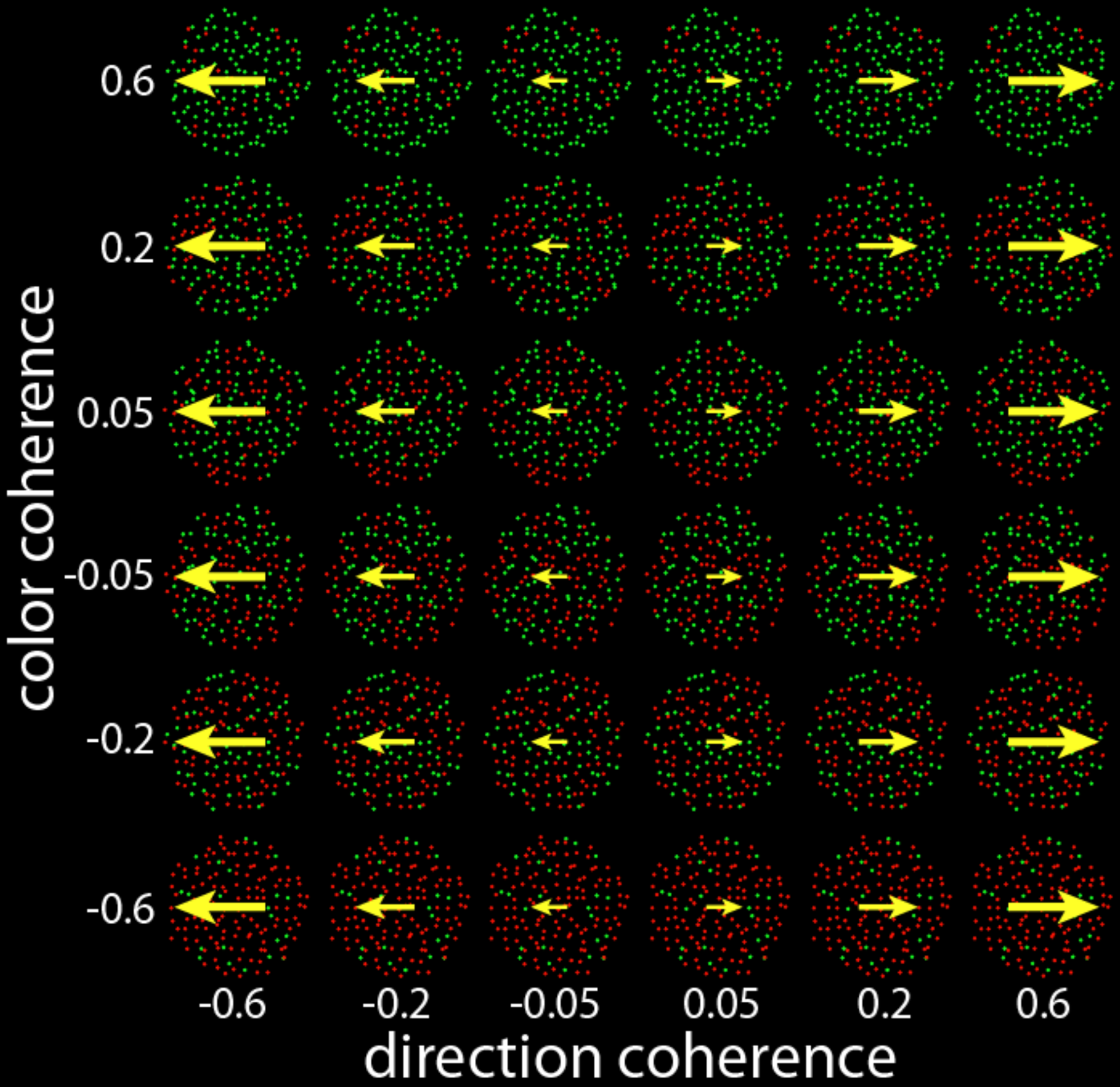
Stimuli


0.6
0.2
0.05
-0.05
-0.2
-0.6
color coherence
-0.6
-0.2
-0.05
0.05
0.2
0.6
direction coherence

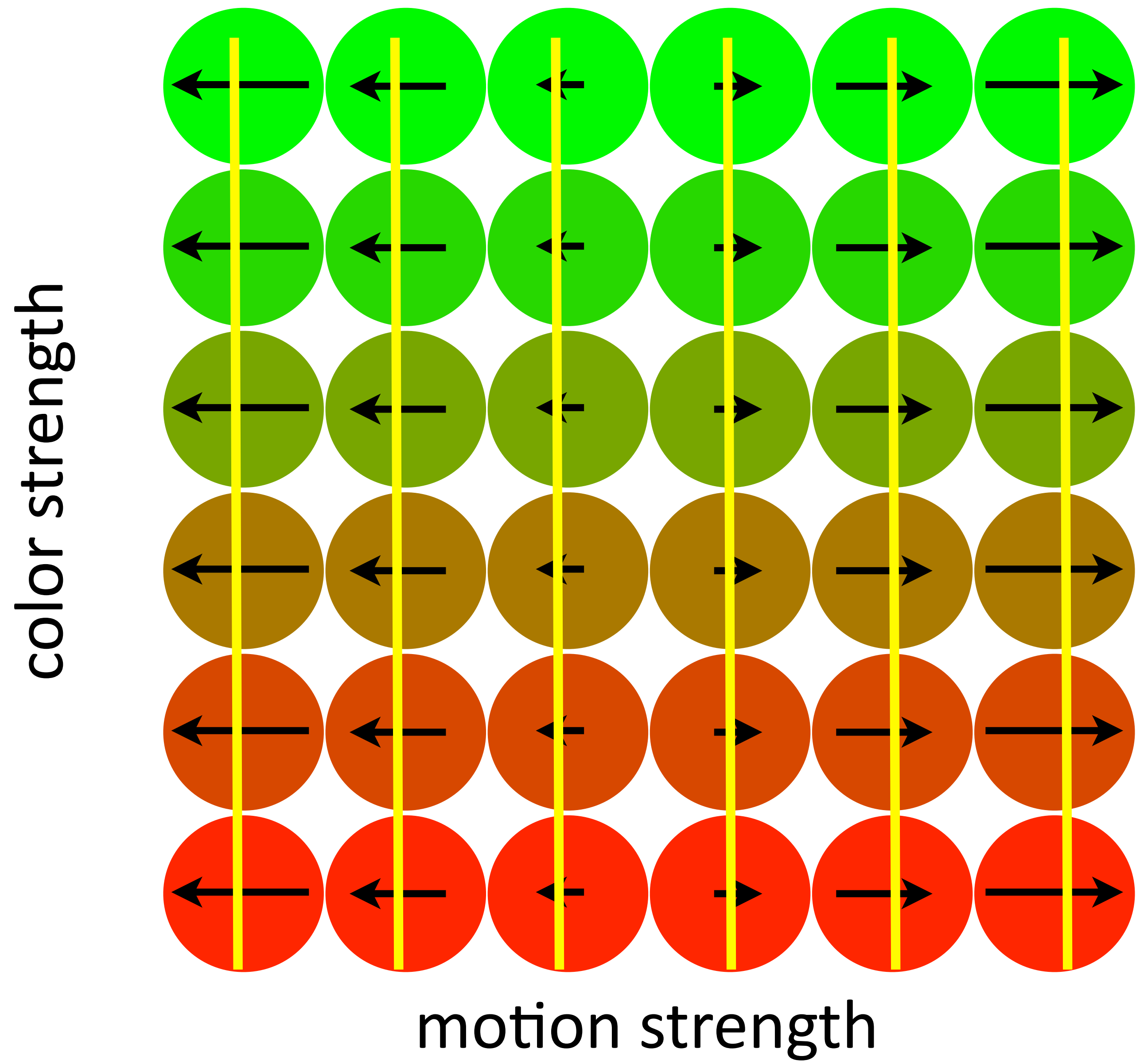
Stimuli


0.6
0.2
0.05
-0.05
-0.2
-0.6
color coherence
-0.6
-0.2
-0.05
0.05
0.2
0.6
direction coherence

Stimuli

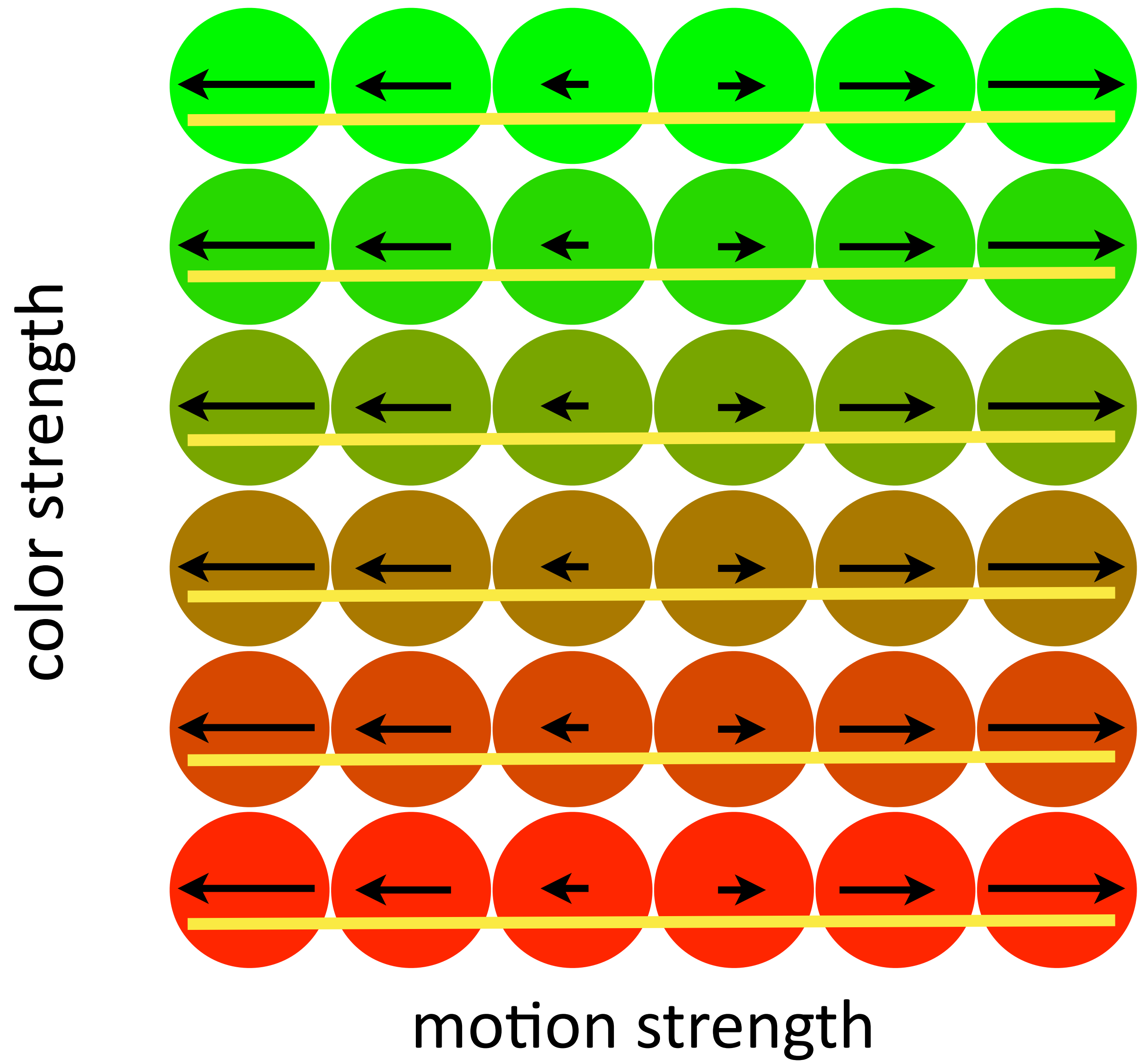

0.6
0.2
0.05
-0.05
-0.2
-0.6
color coherence
-0.6
-0.2
-0.05
0.05
0.2
0.6
direction coherence

Averaging over color shows effects of motion



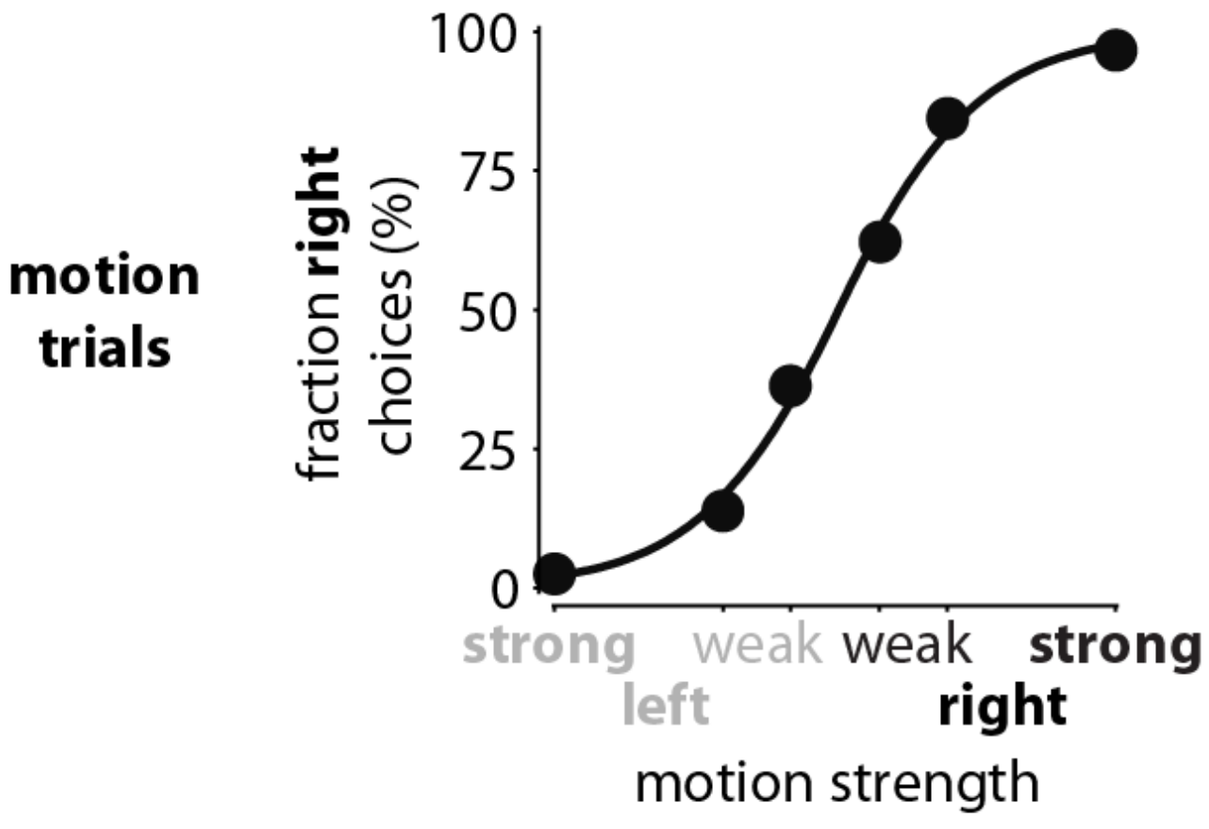
“Average over”

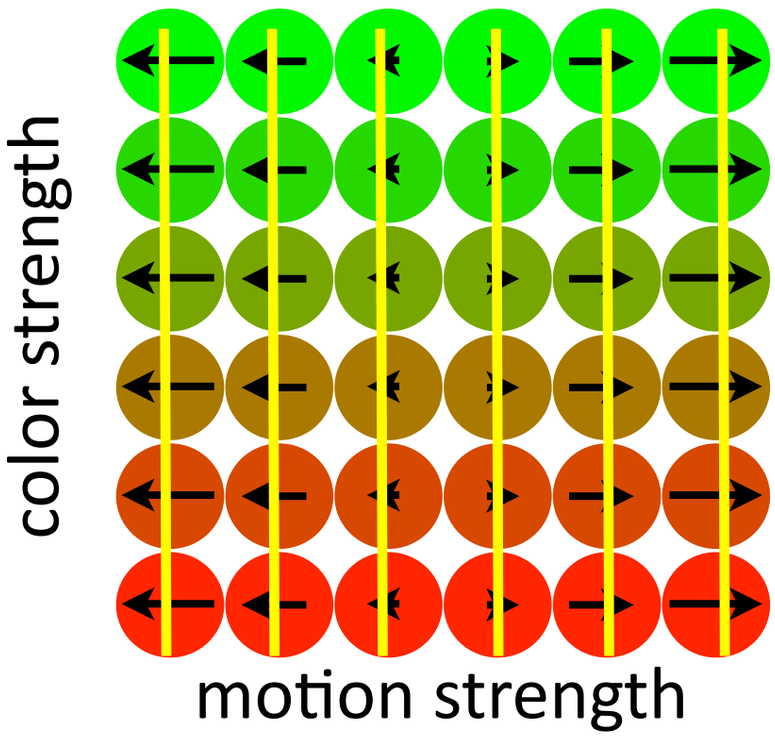
Averaging over motion shows effects of color

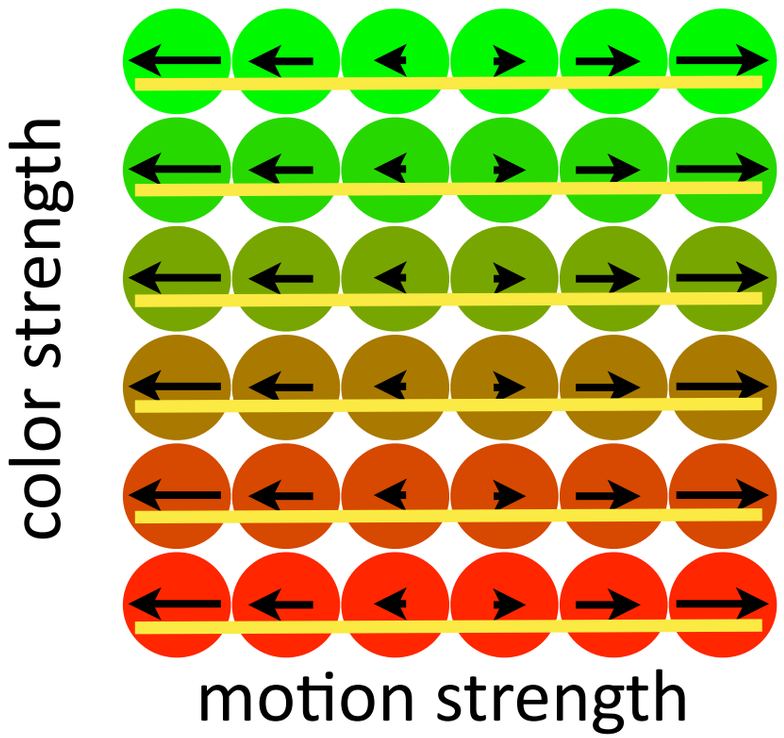


“Average over”

Behavior

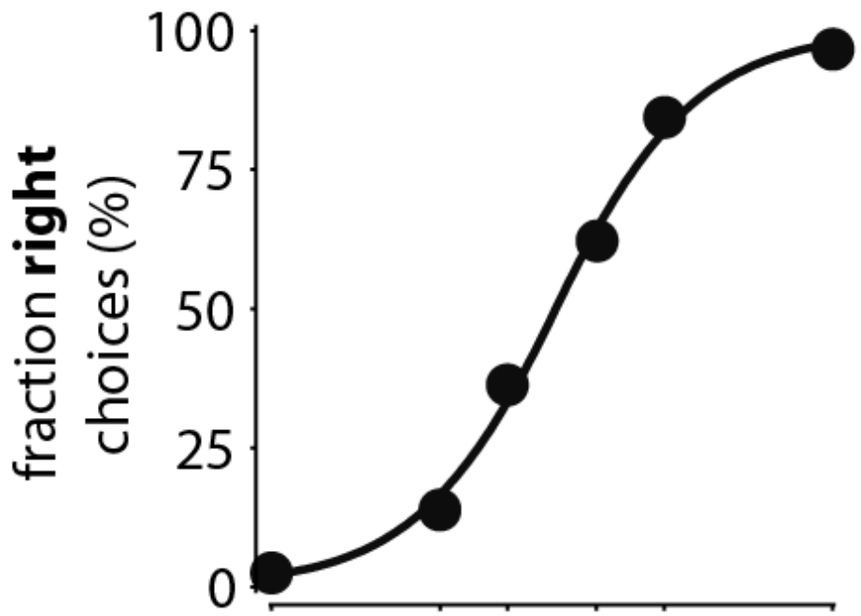

motion
trials
fraction right
choices (%)
100
75
50
25
0
strong
left
weak
weak
strong
right
motion strength

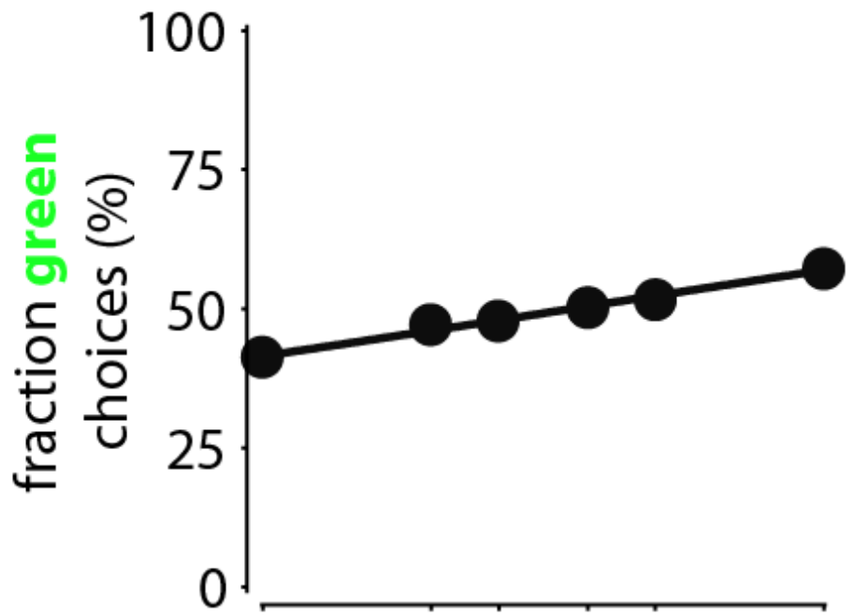

color strength
motion strength


color strength
motion strength

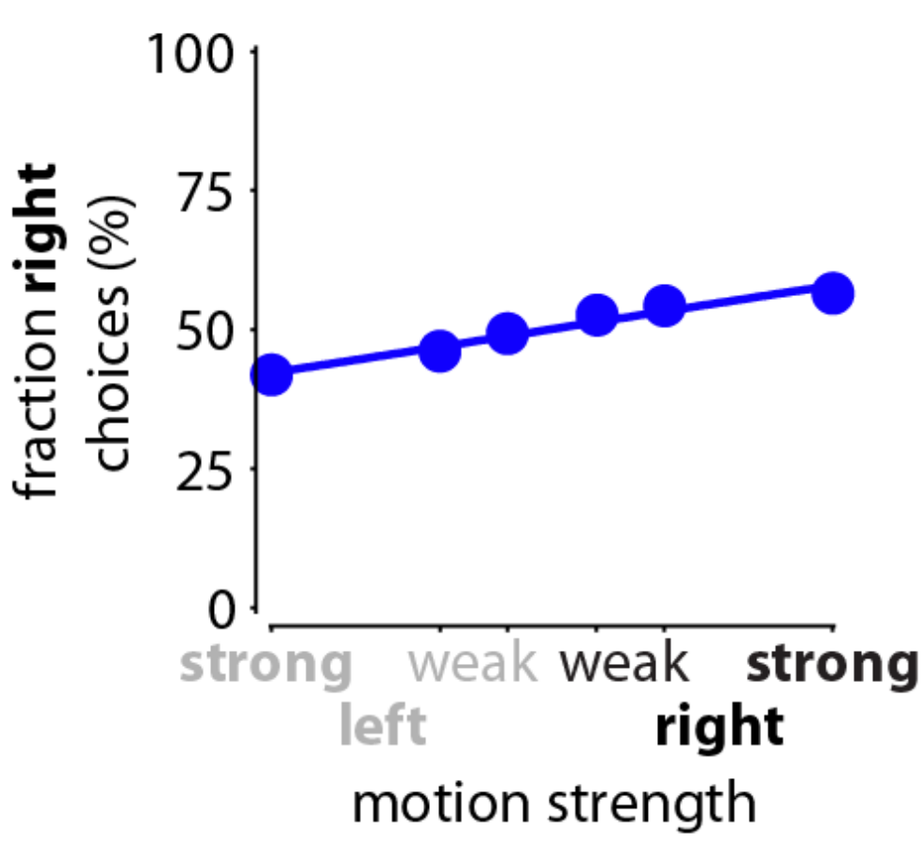
Behavior

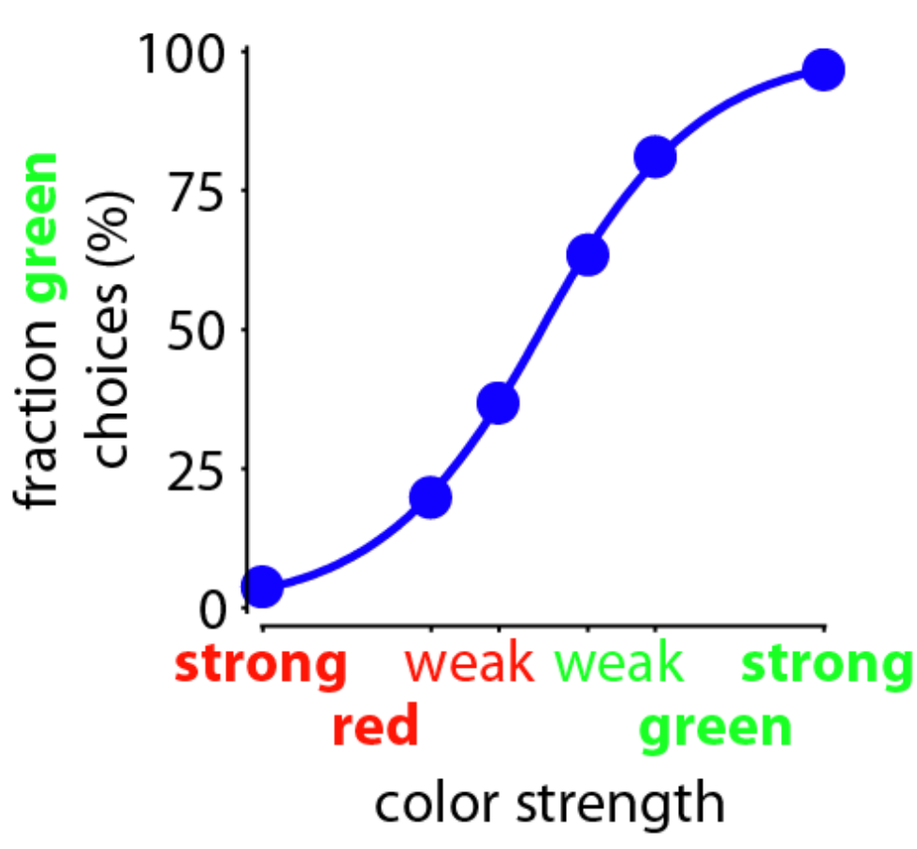
motion trials



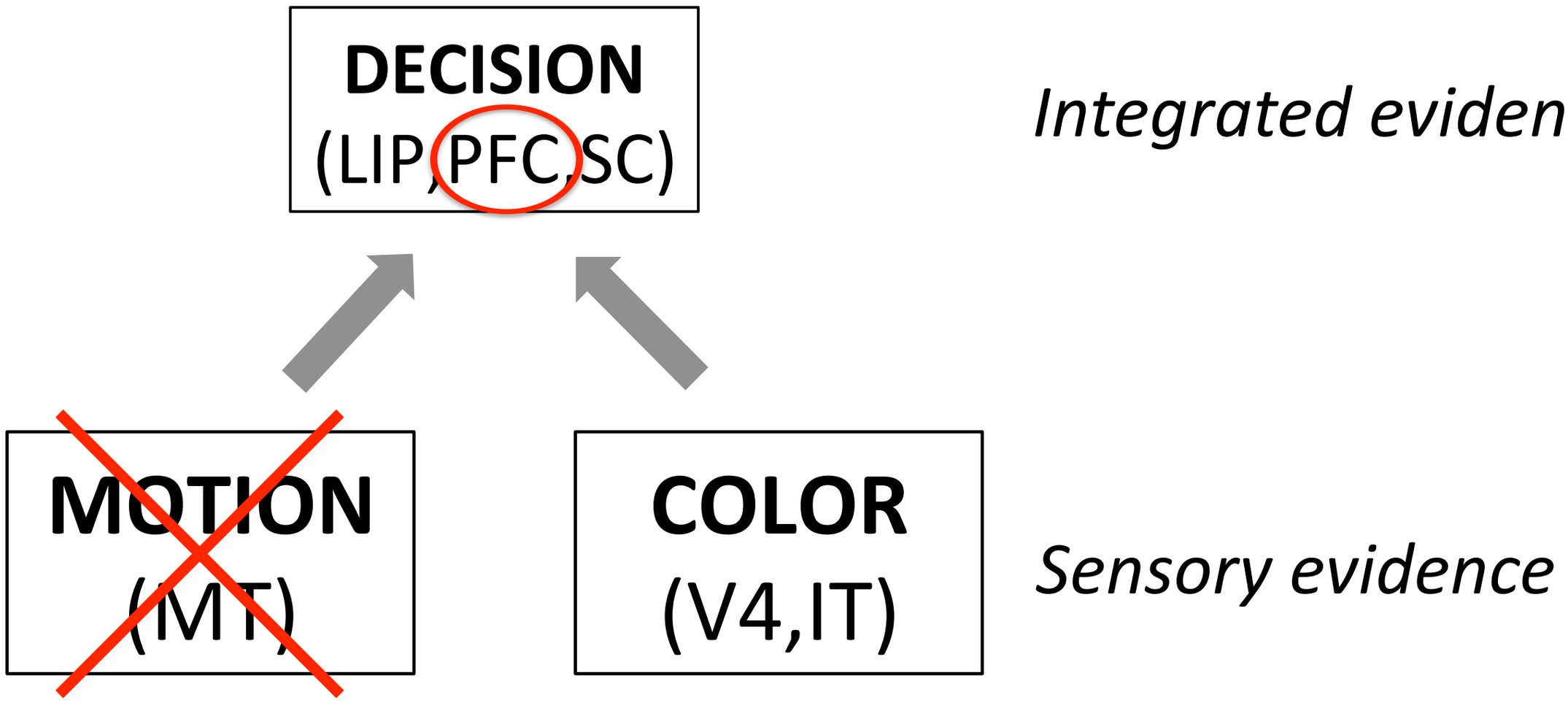


color trials



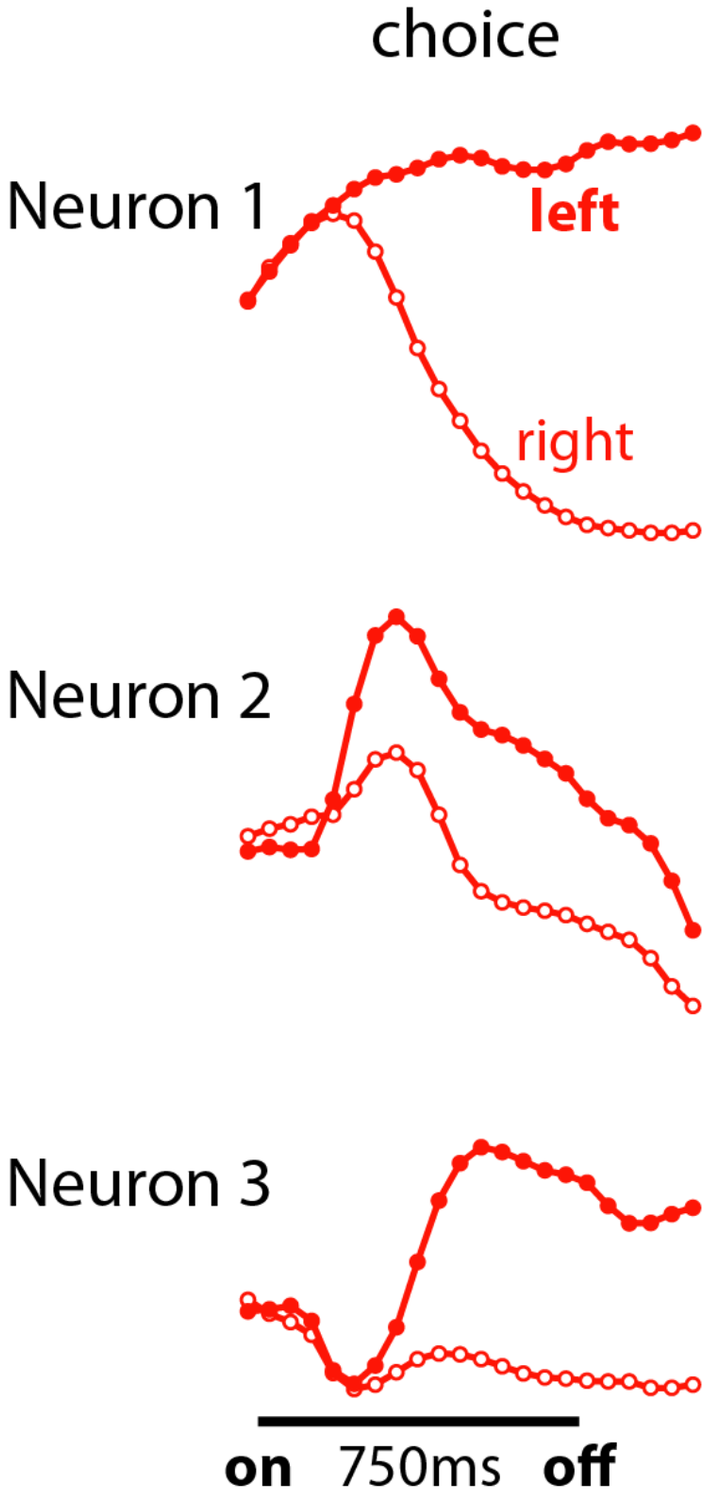


Where are sensory inputs selected?

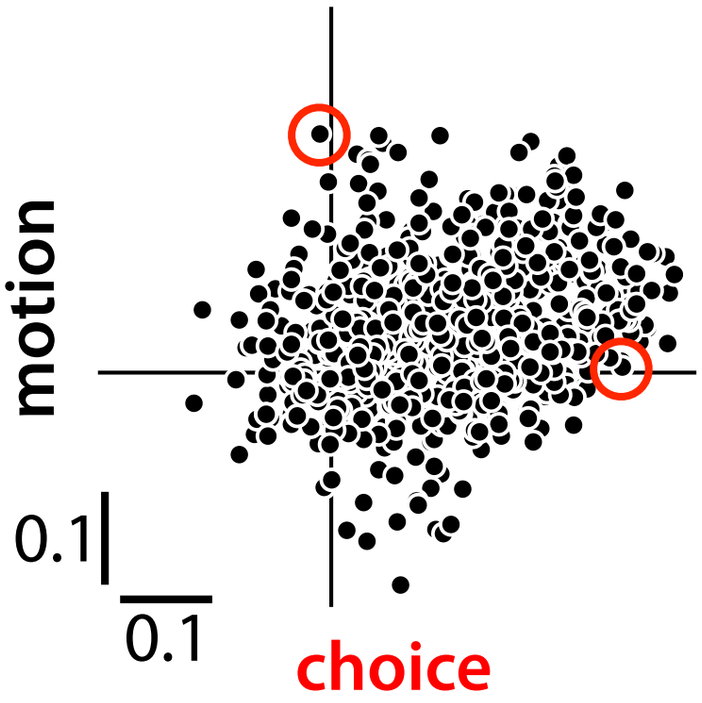


One could easily frame this work in the context of routing information in the brain.

Mixed signals in FEF neurons

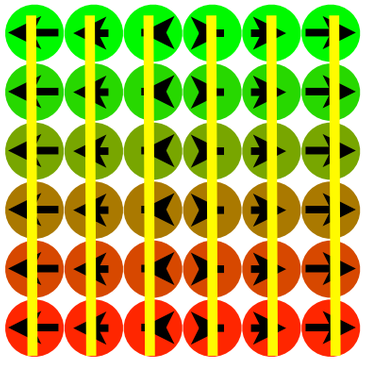


Mixed signals in FEF neurons



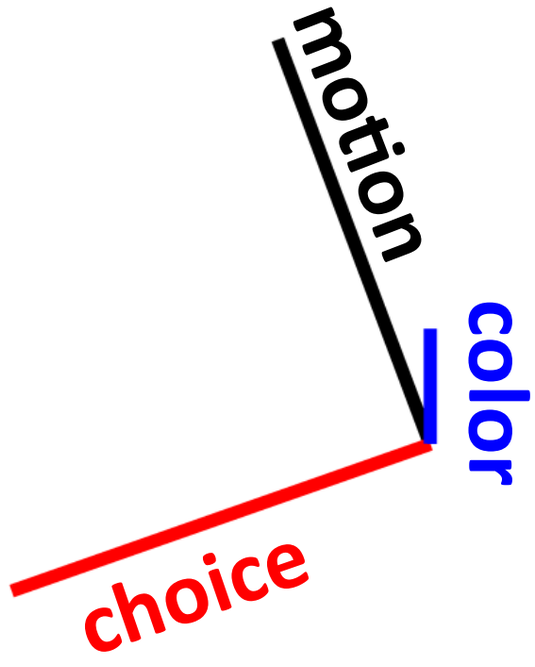
Verbal aside on how to make sense of this data via a state-space.

PFC population response during motion context

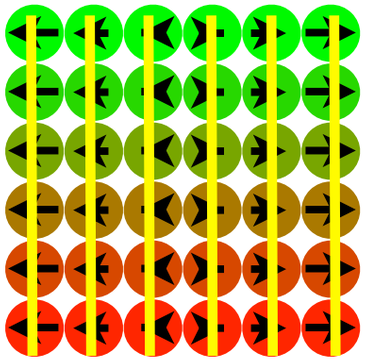


“dots on” to “dots off” (750ms)

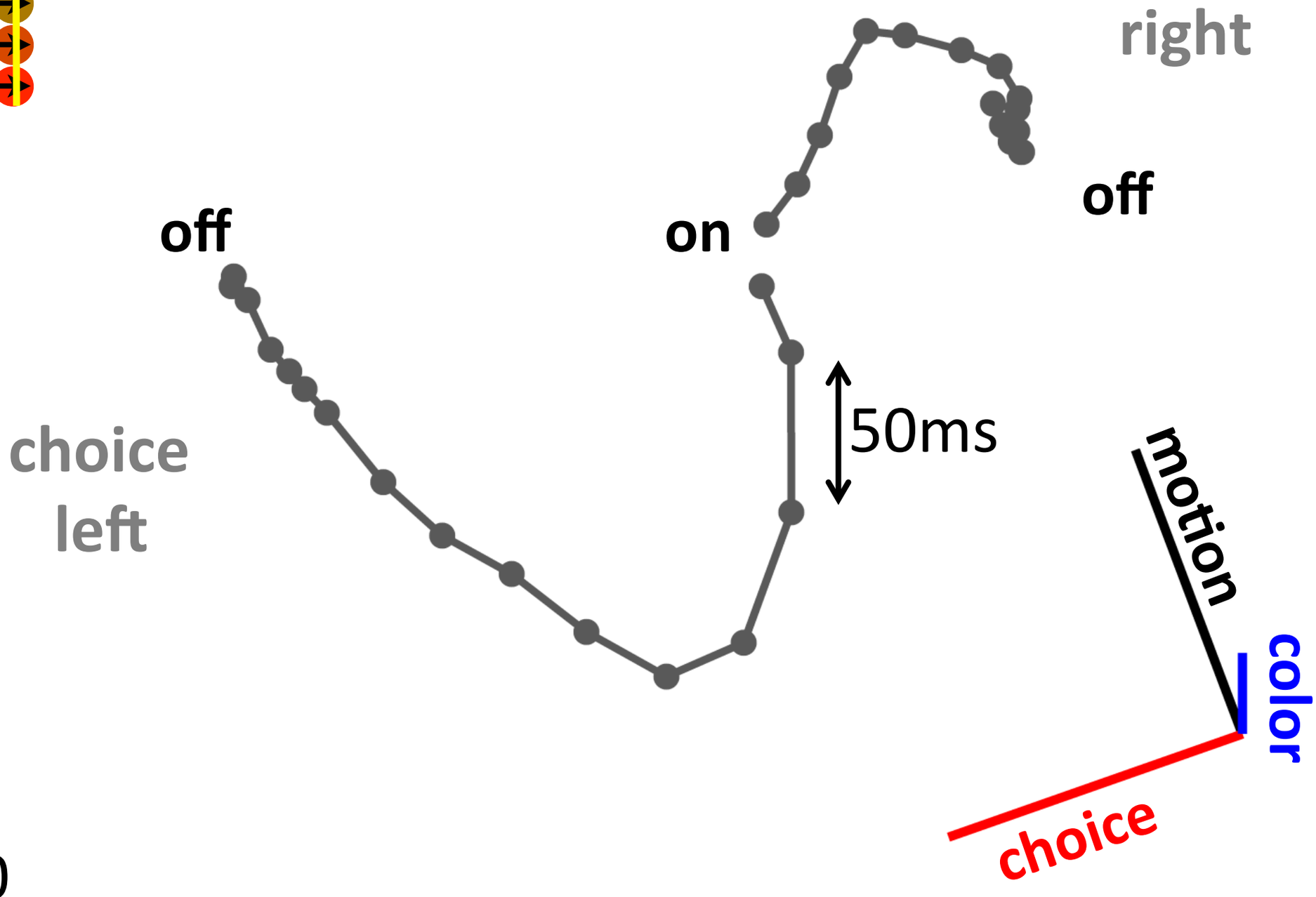
Correct trials only!



PFC population response during motion context

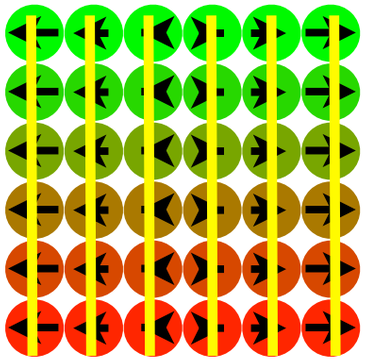


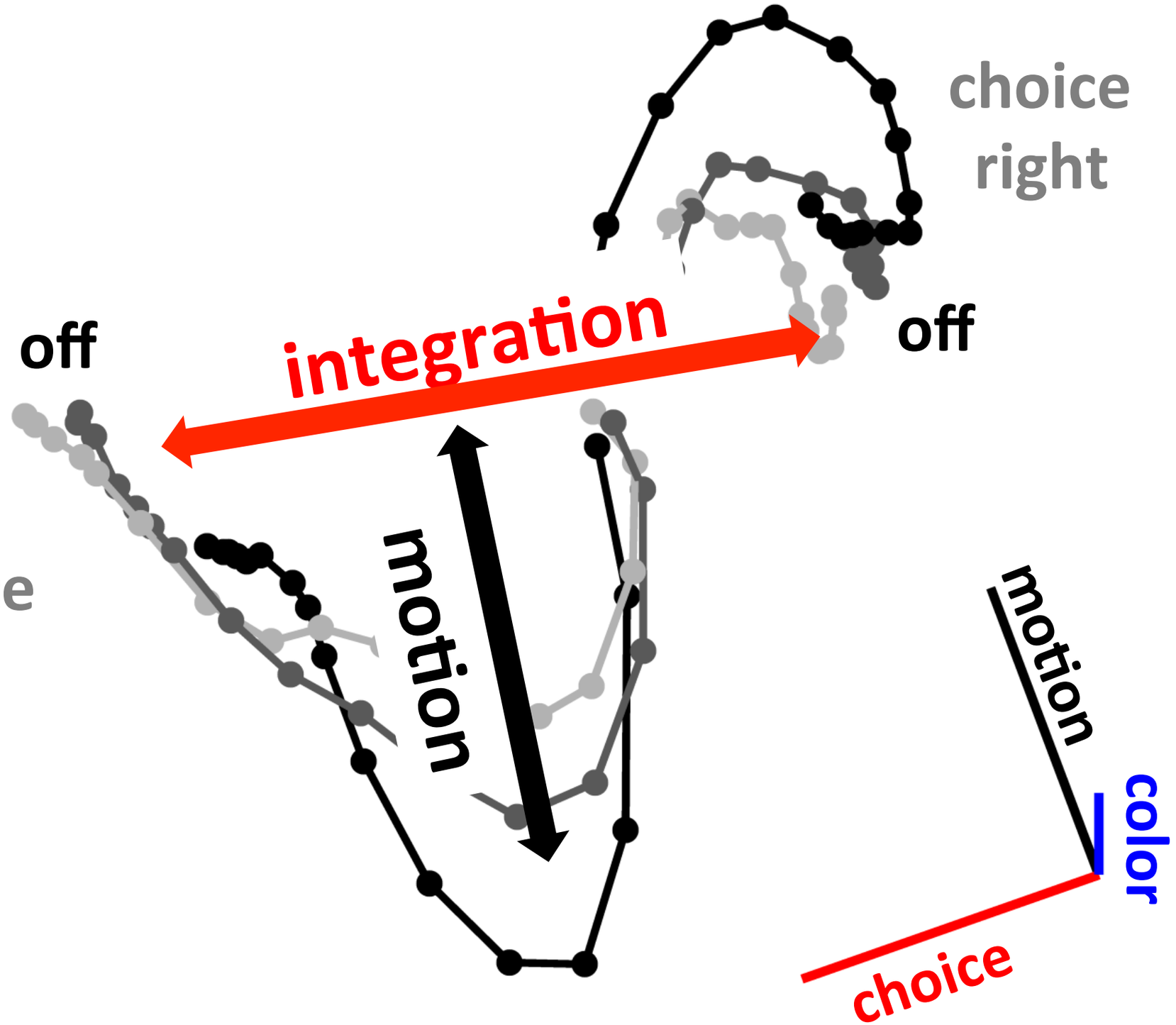
"dots on" to "dots off" (750ms)



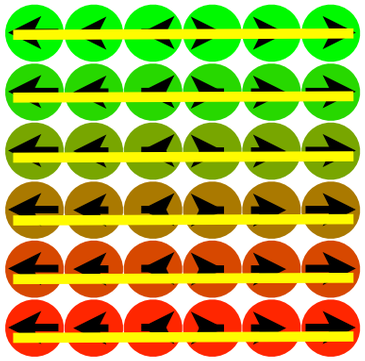
N > 250

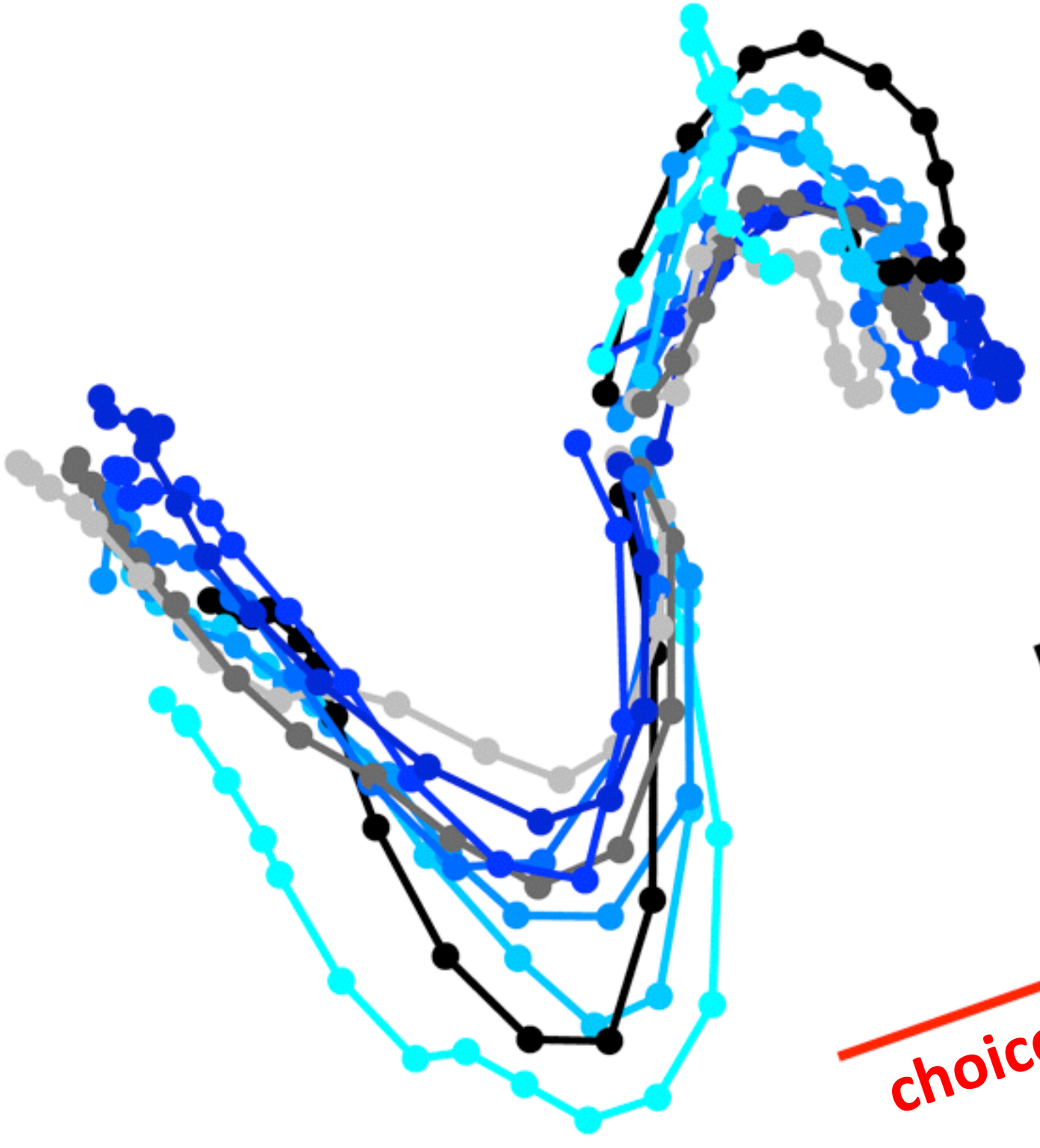
PFC population response during motion context



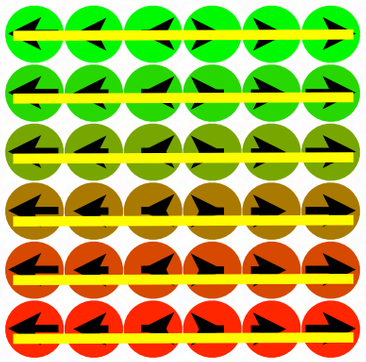


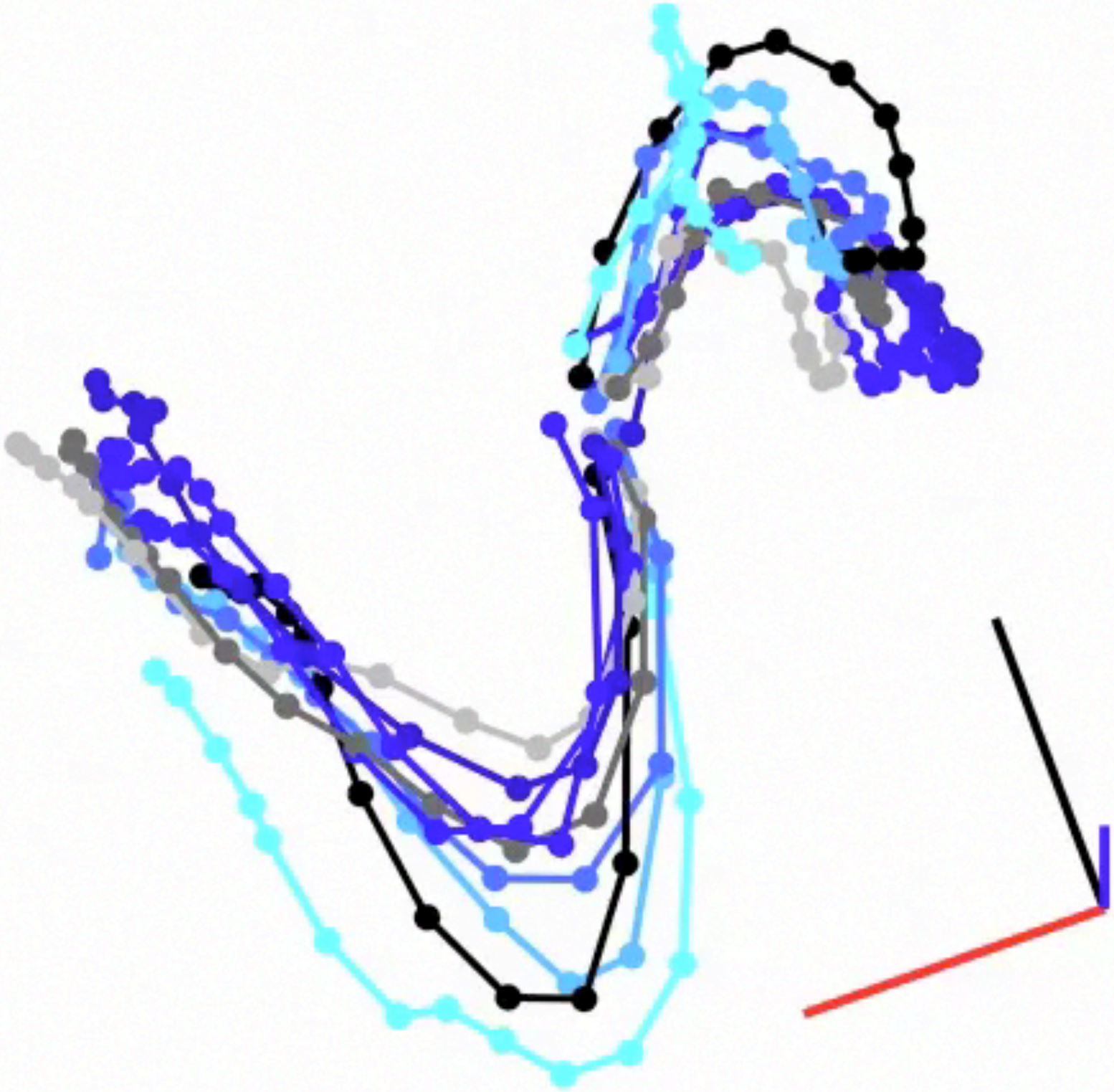
PFC population response during motion context



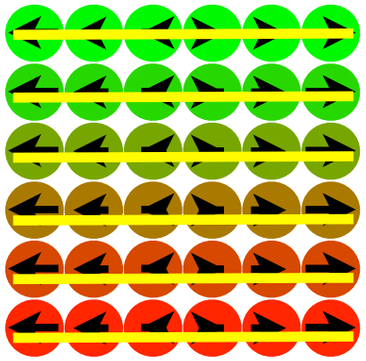


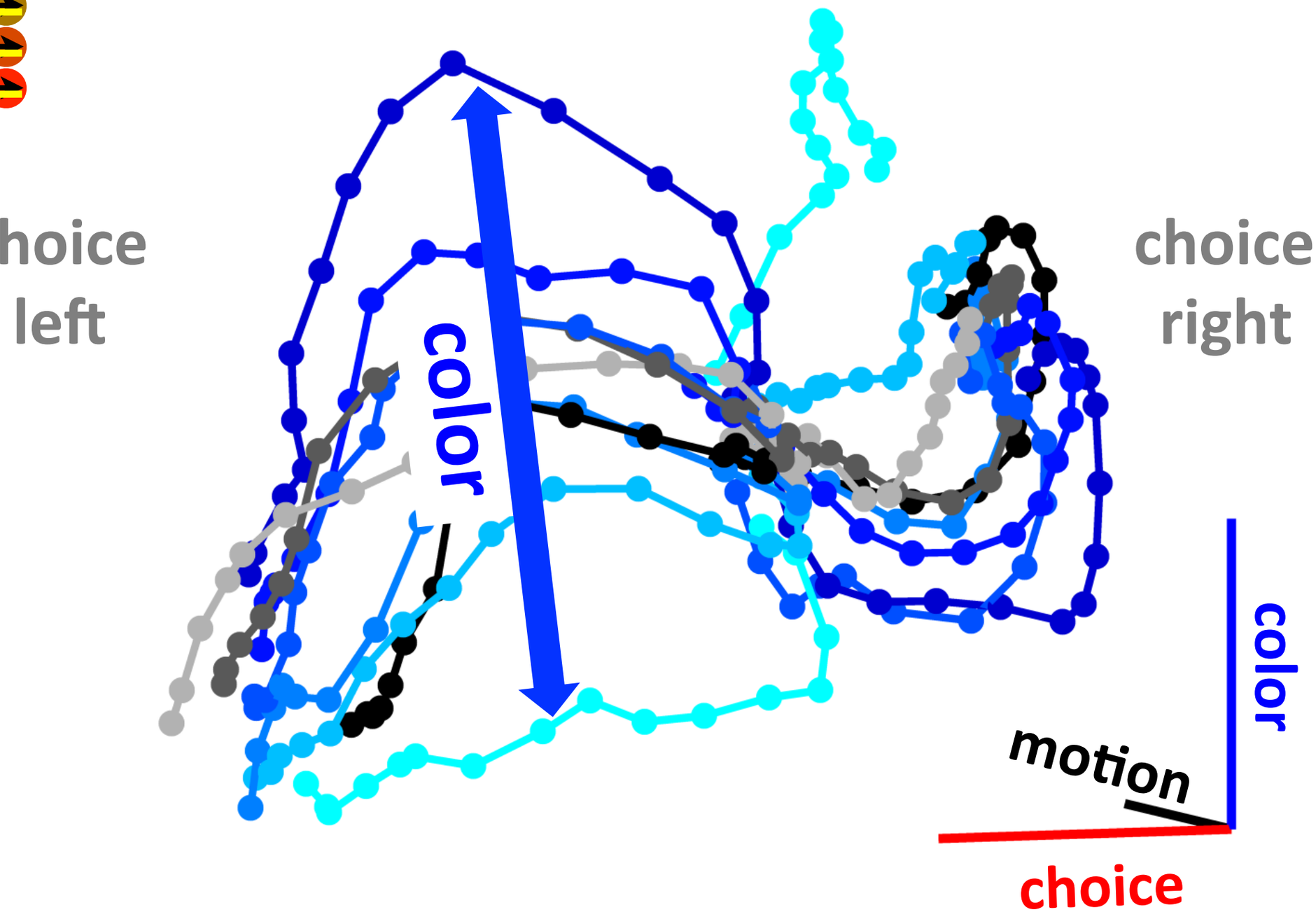
PFC population response during motion context



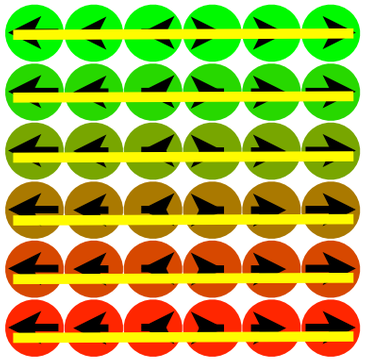


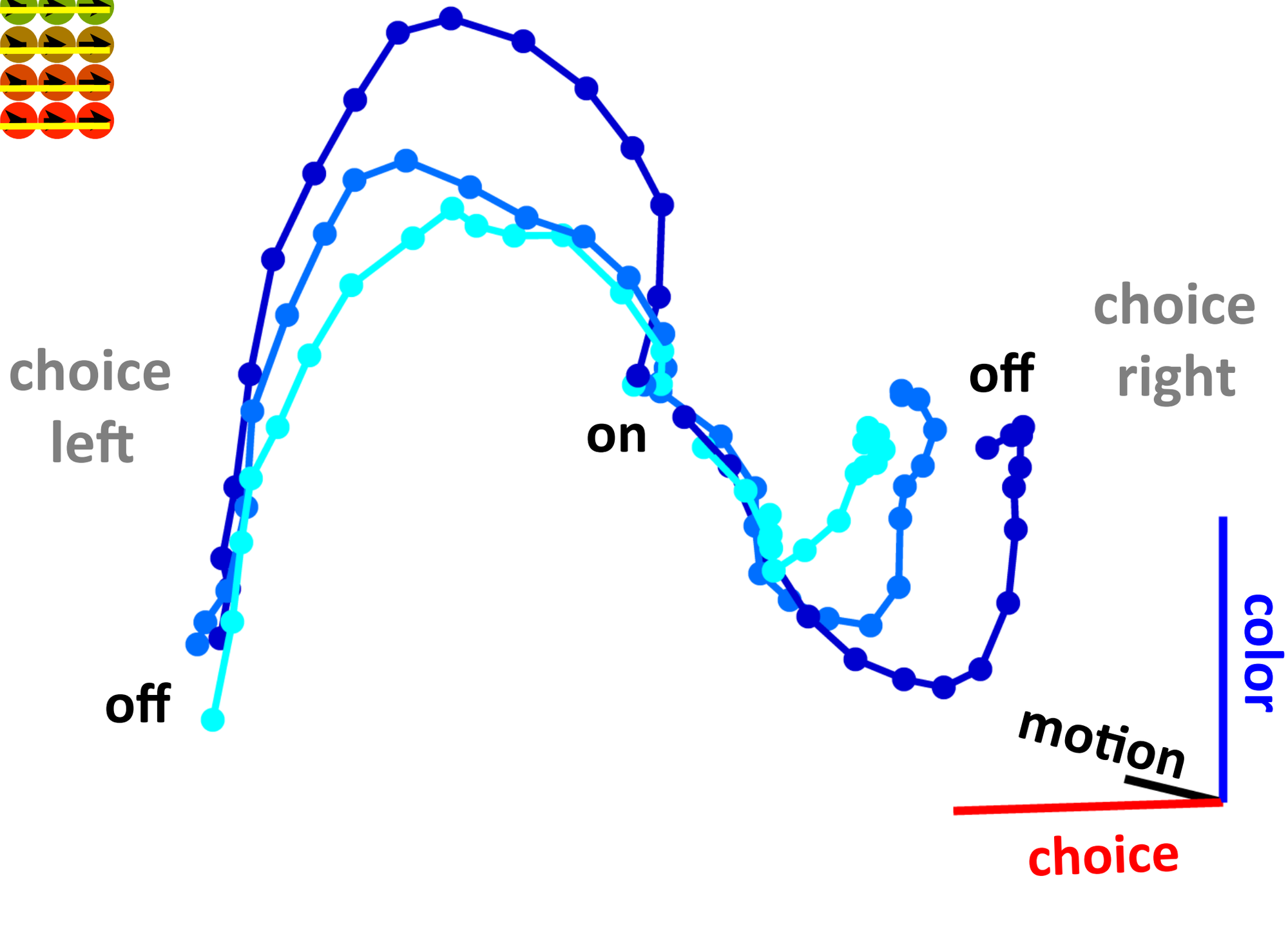
PFC population response during motion context



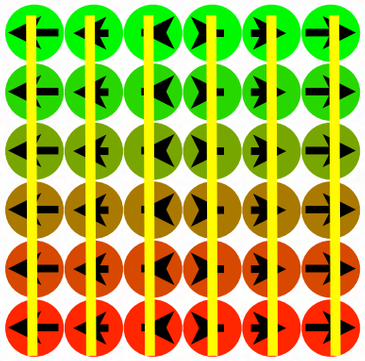


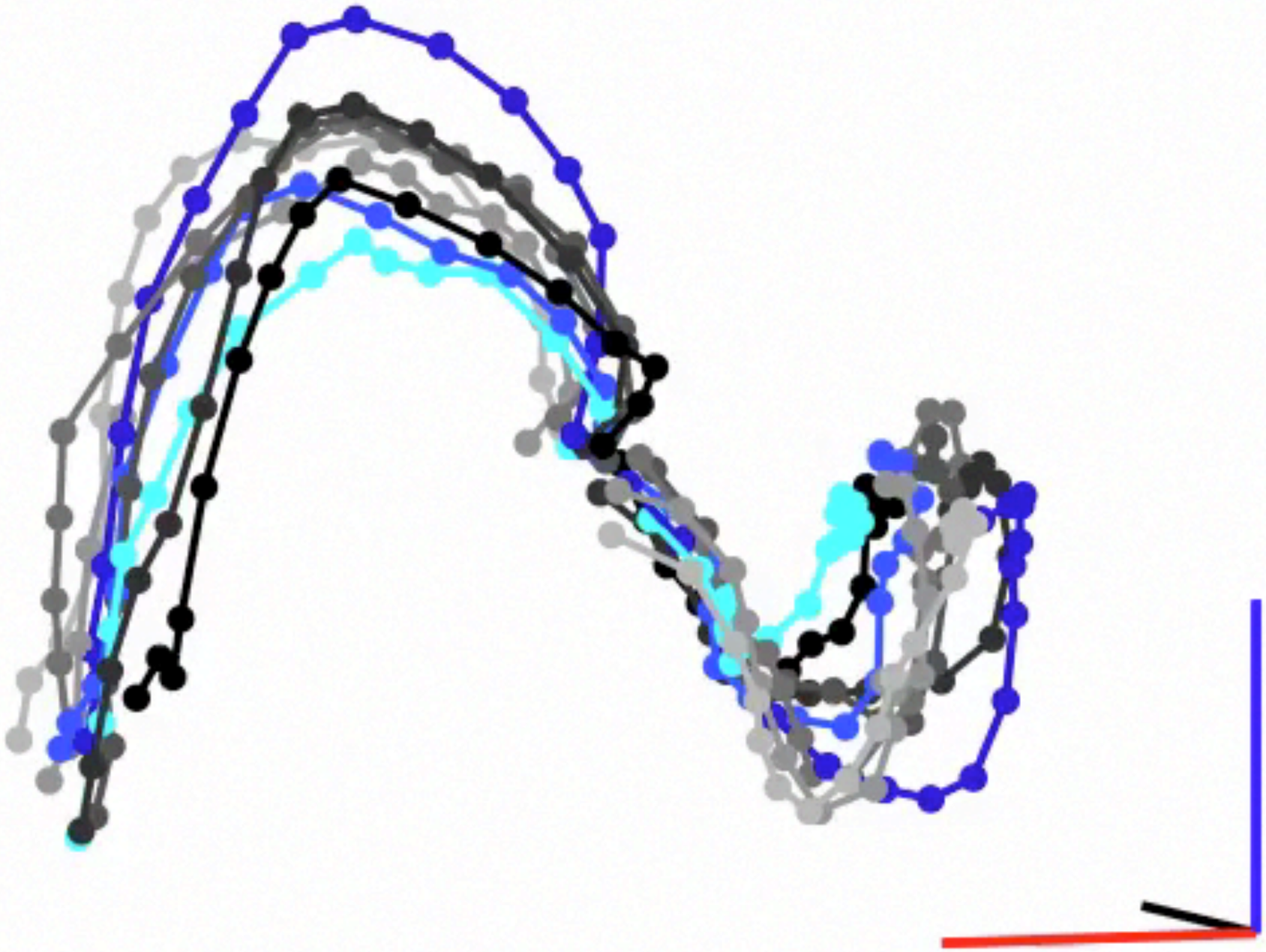
PFC population response during color context



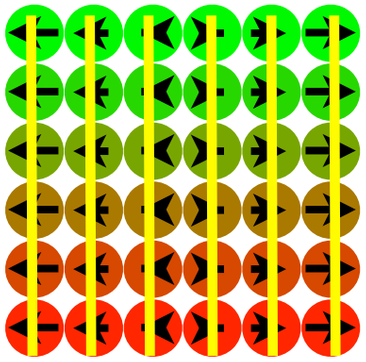


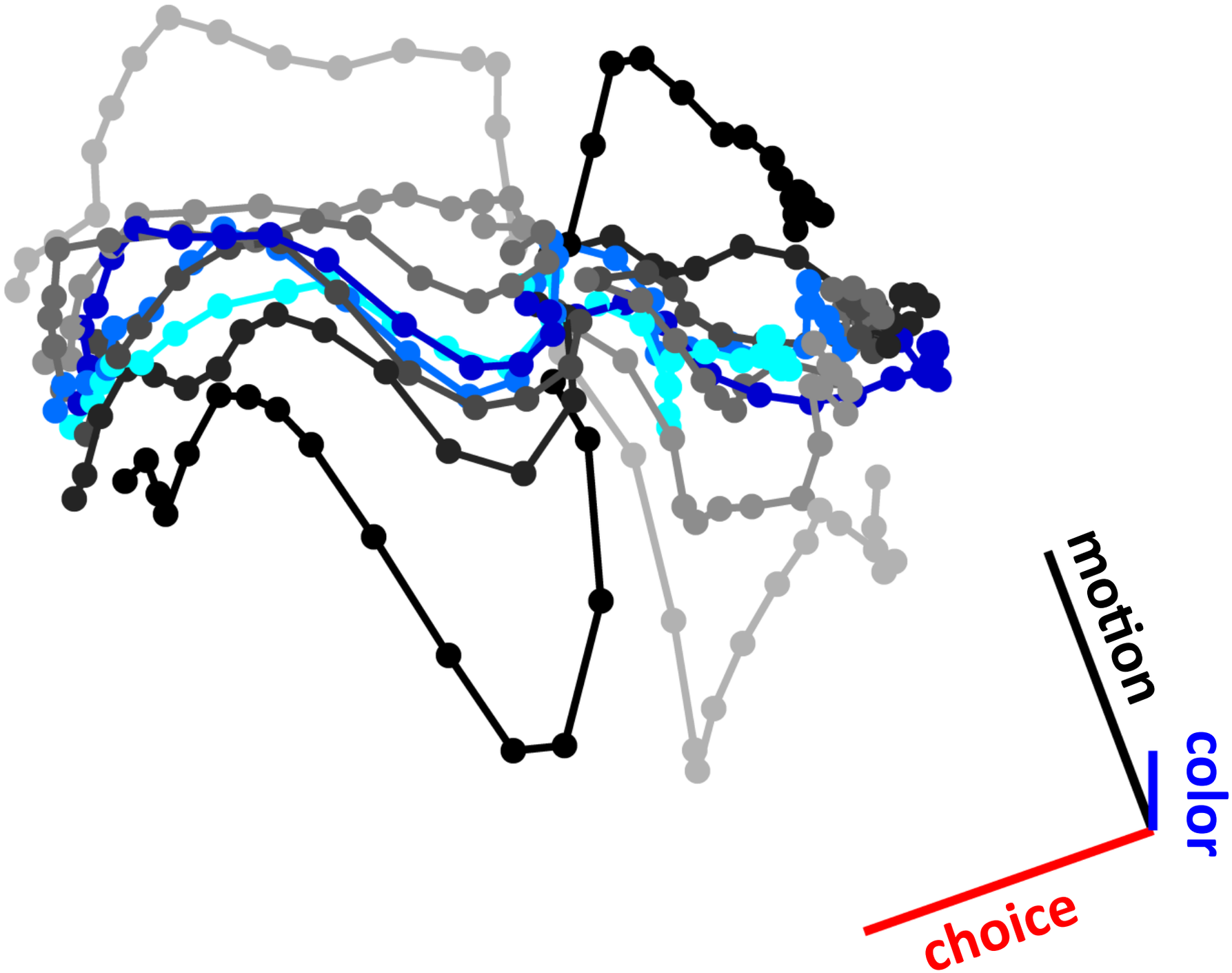
PFC population response during color context



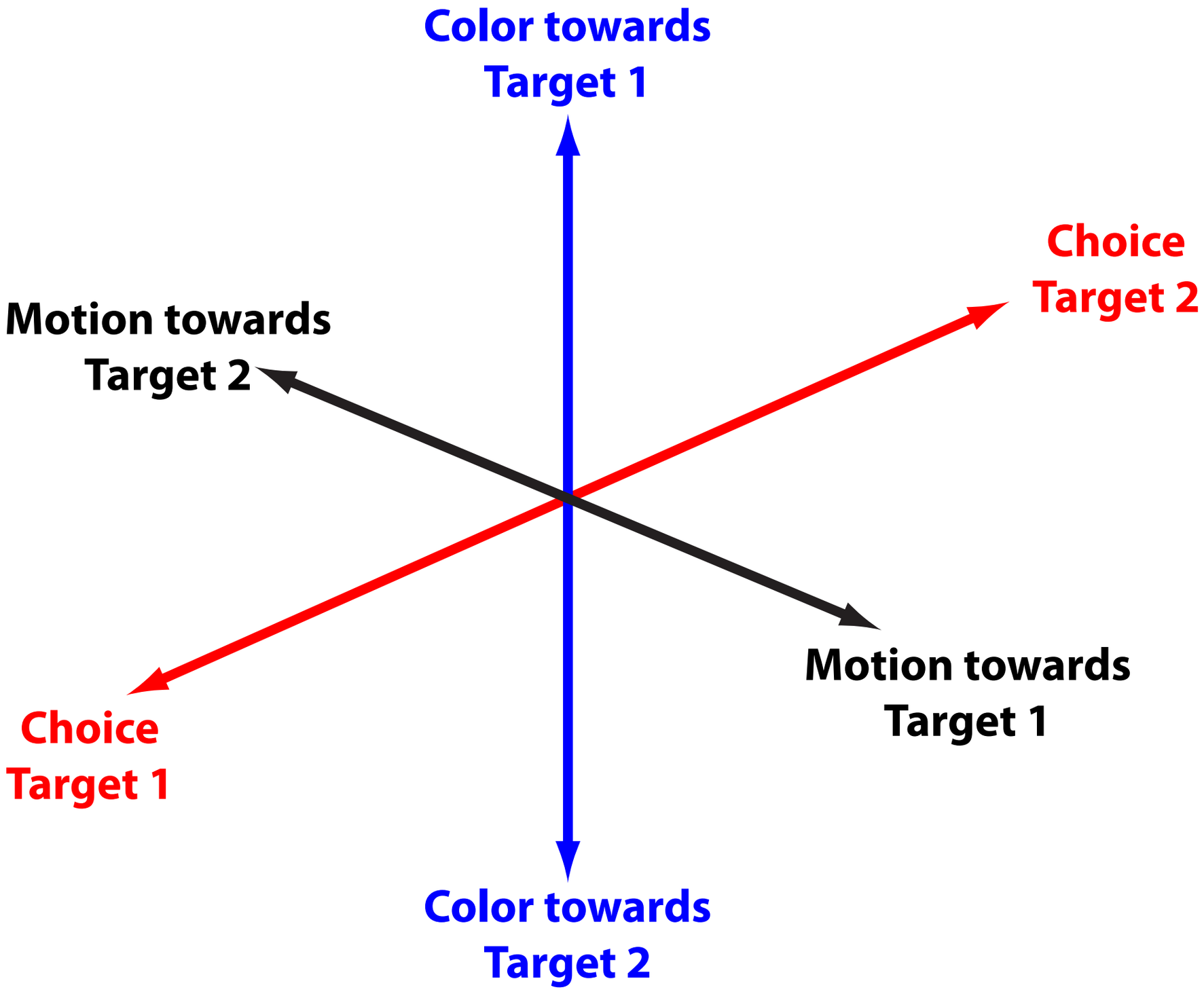


PFC population response during color context

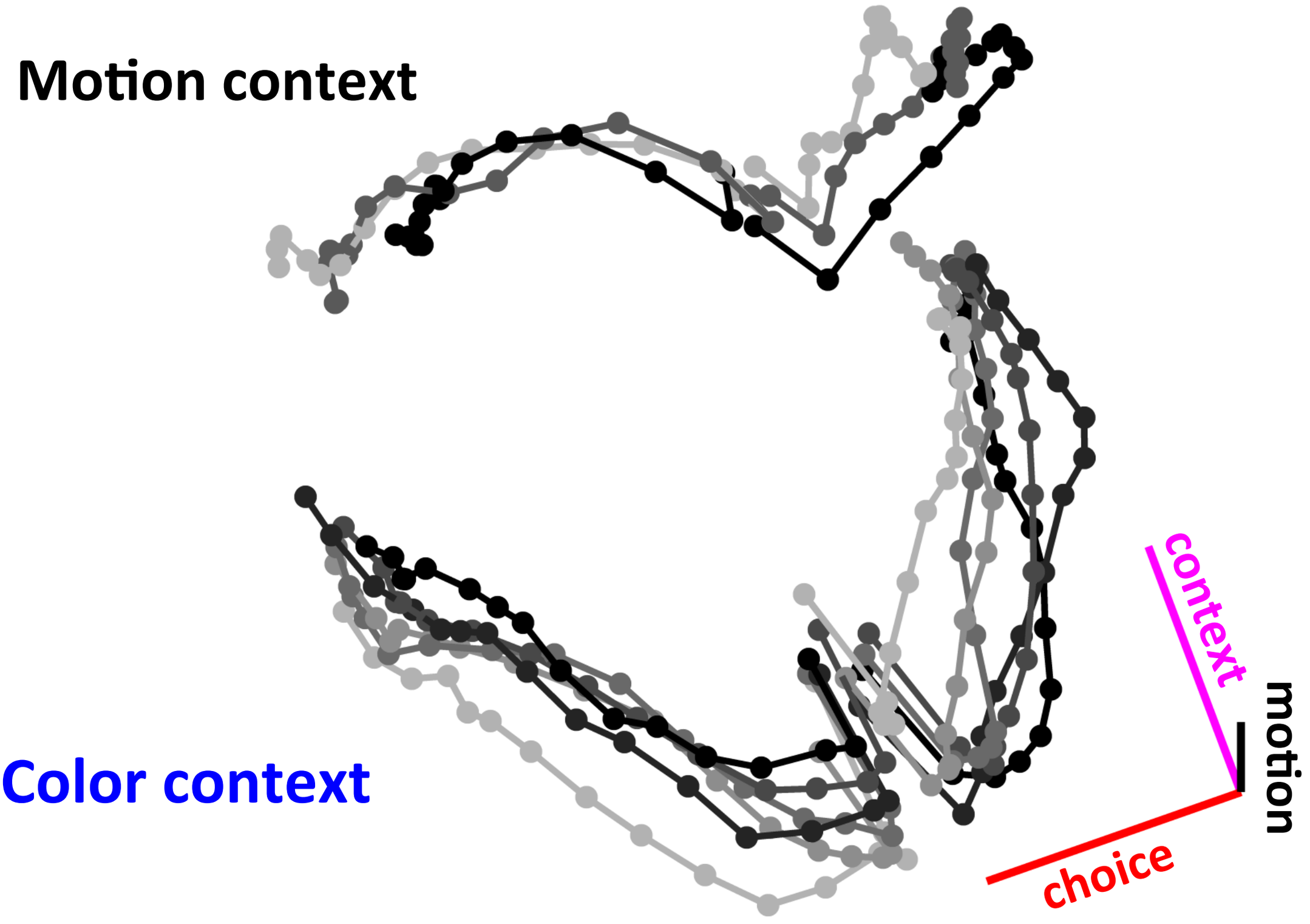




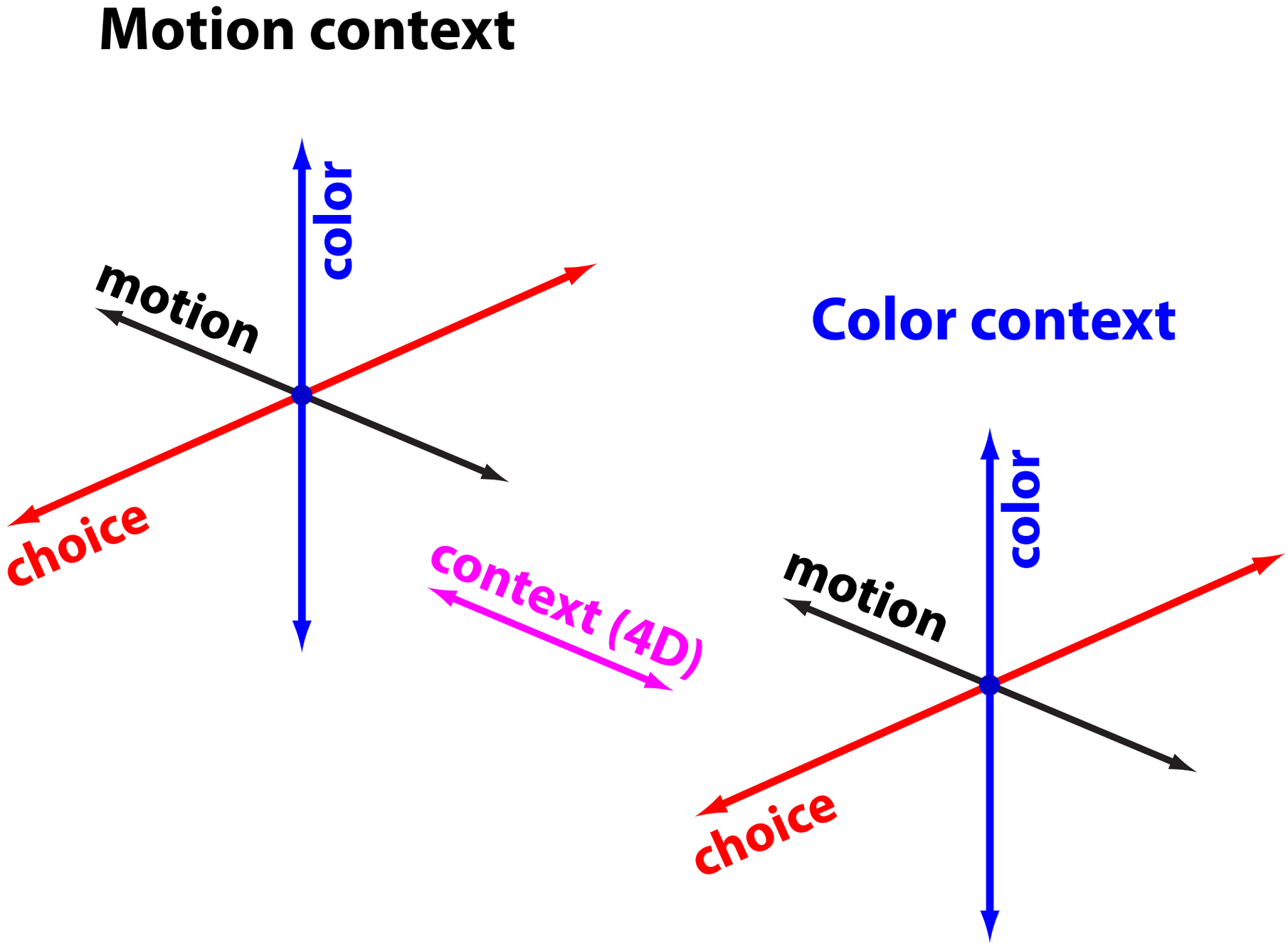
Choice and input signals in PFC



Representation of context in PFC

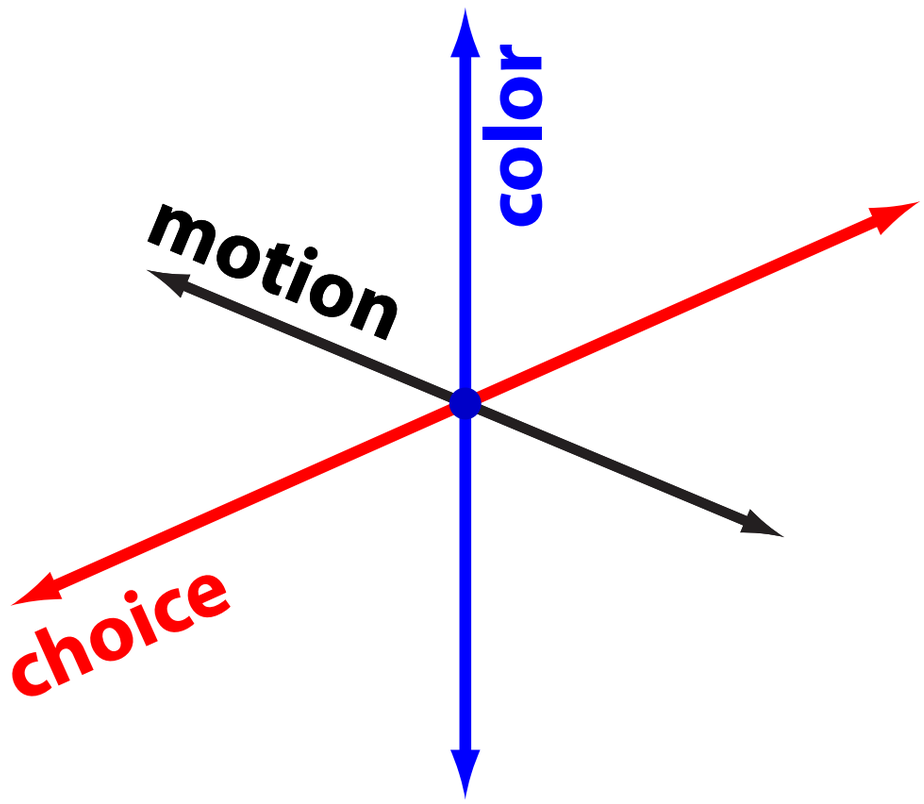


The structure of task related signals in PFC

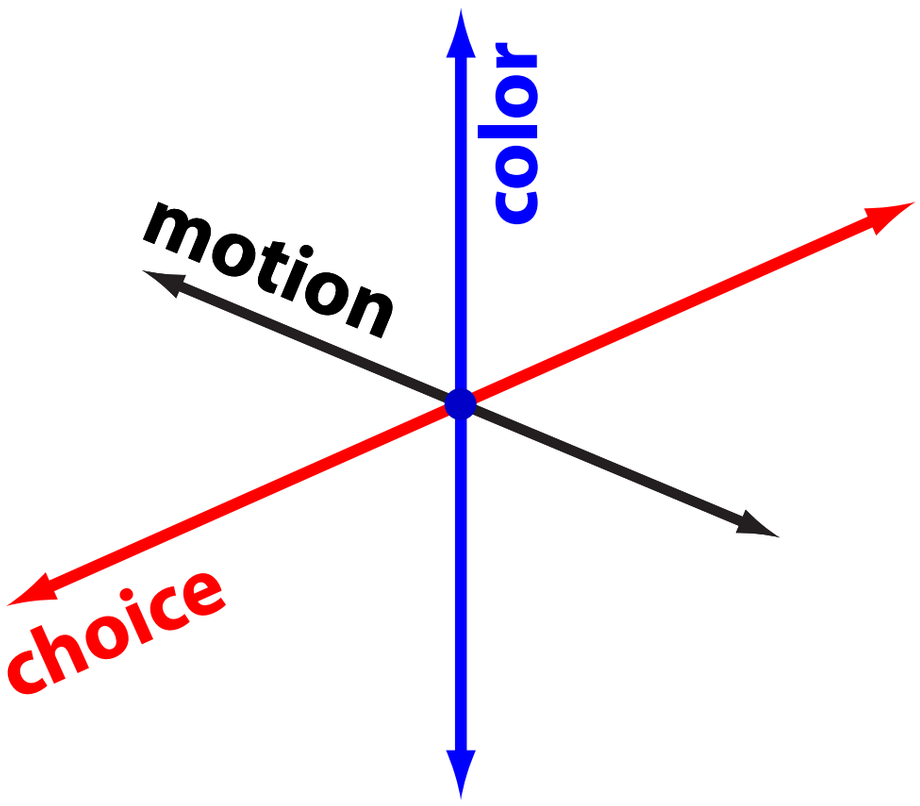


How does selective integration occur?

Motion context

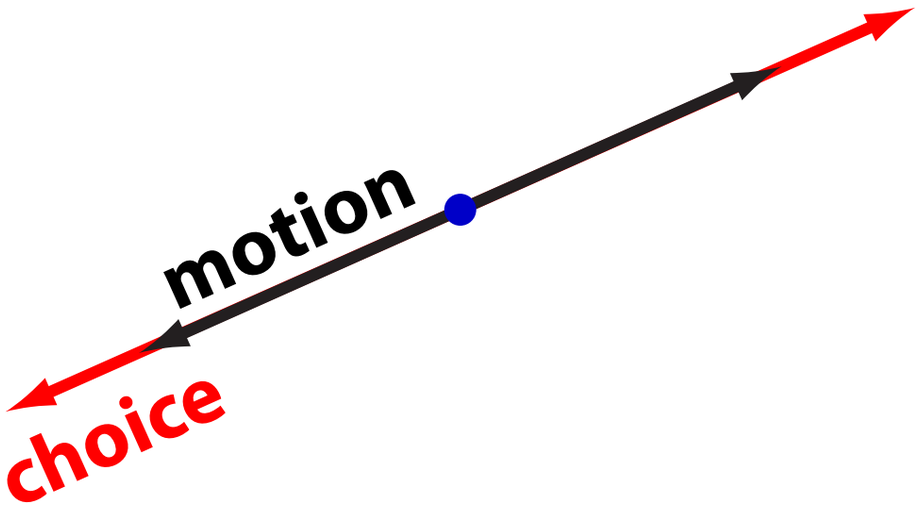


Color context



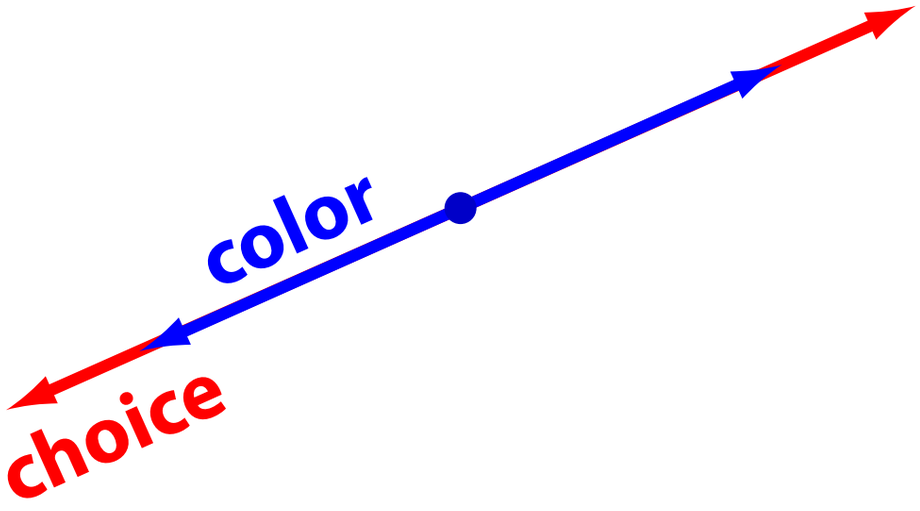
How does selective integration occur?

Motion context



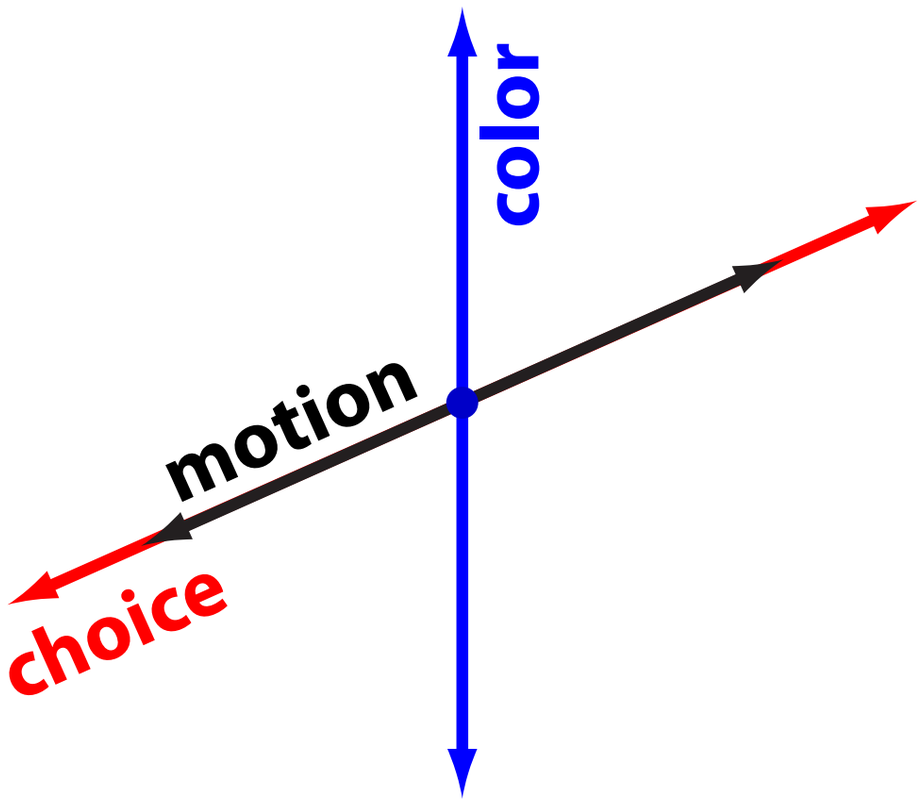
Context-dependent
gating *(“attention”)*

Color context



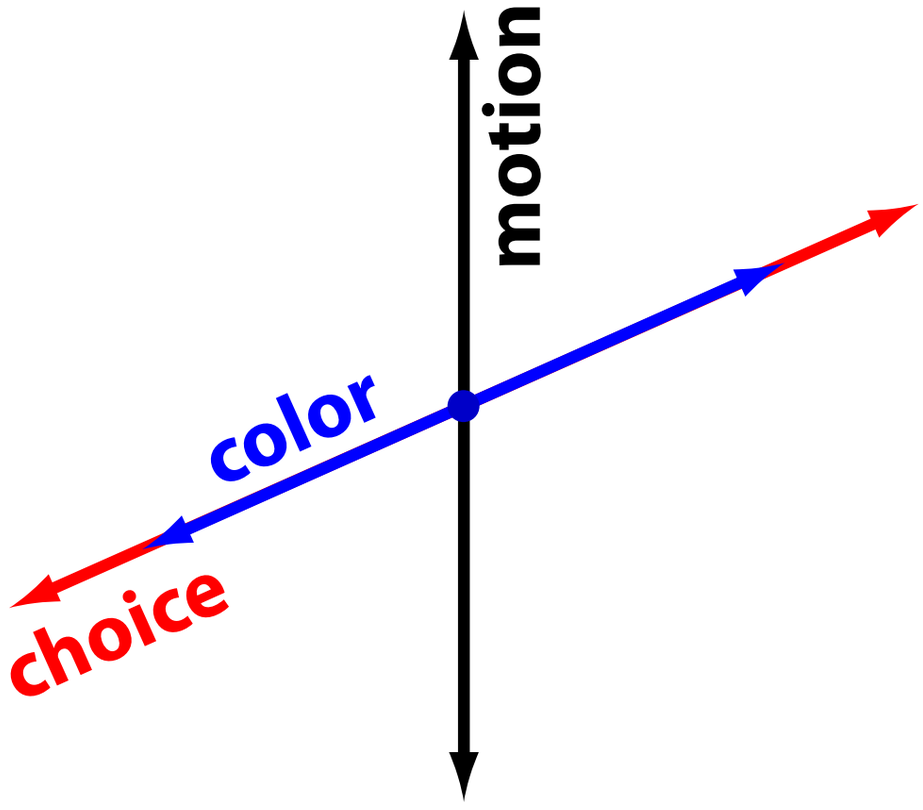
How does selective integration occur?

Motion context



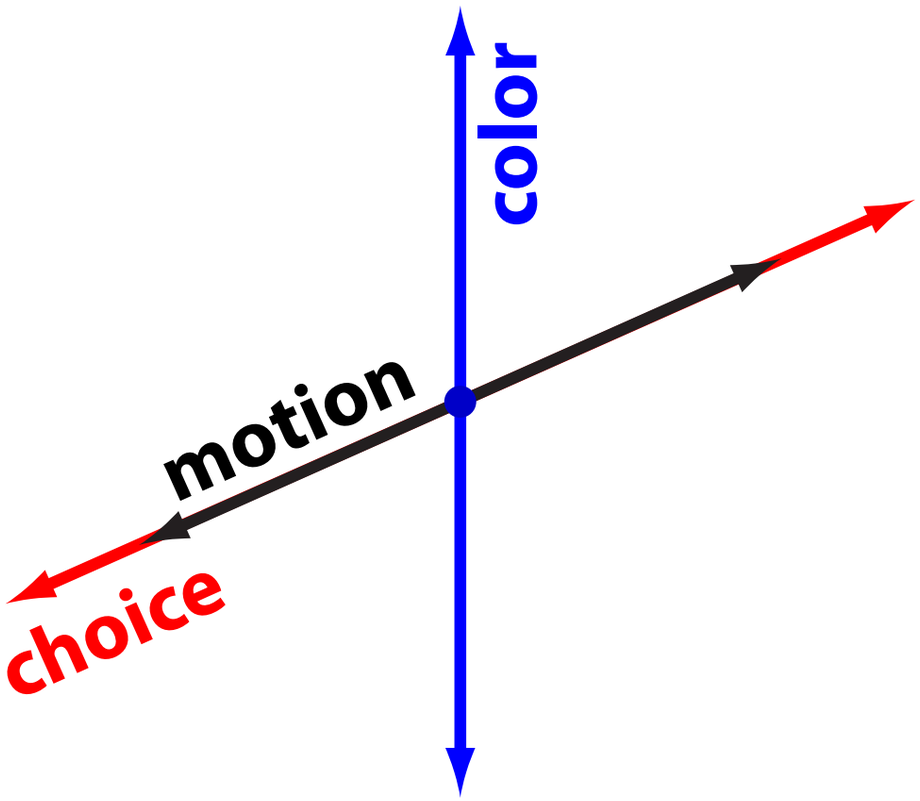
Context-dependent ***input*** *direction*

Color context



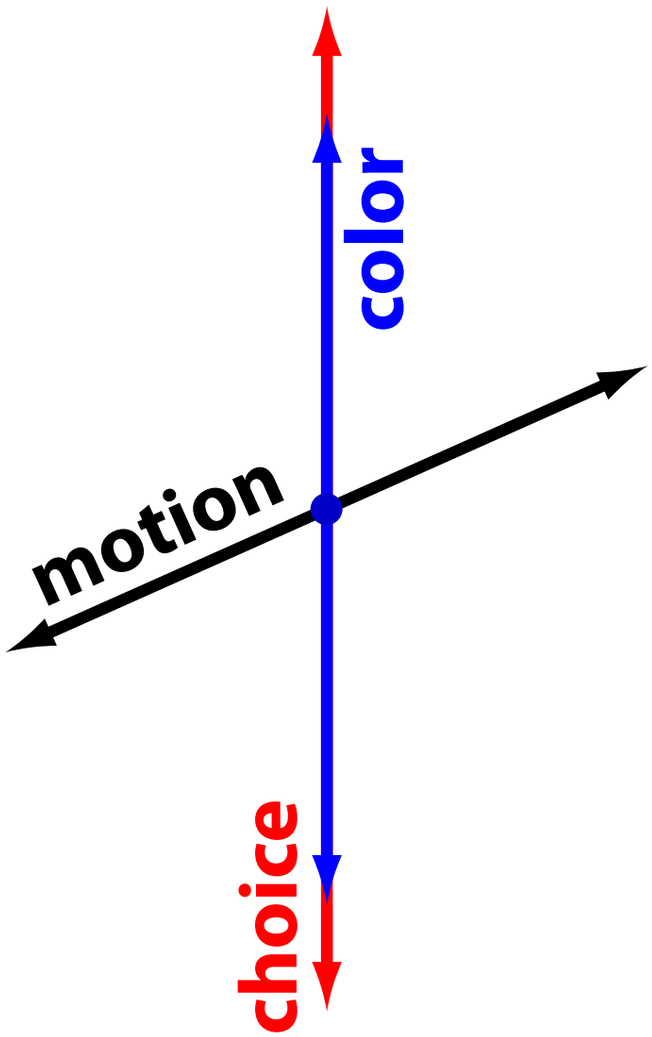
How does selective integration occur?

Motion context



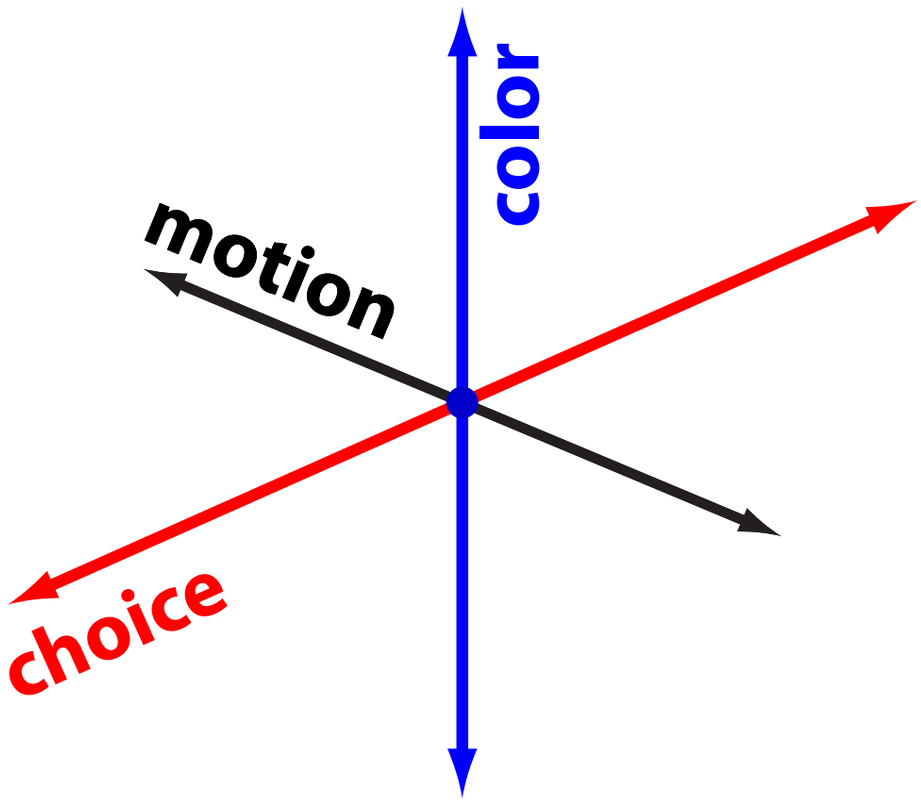
Context-dependent
choice *direction*

Color context

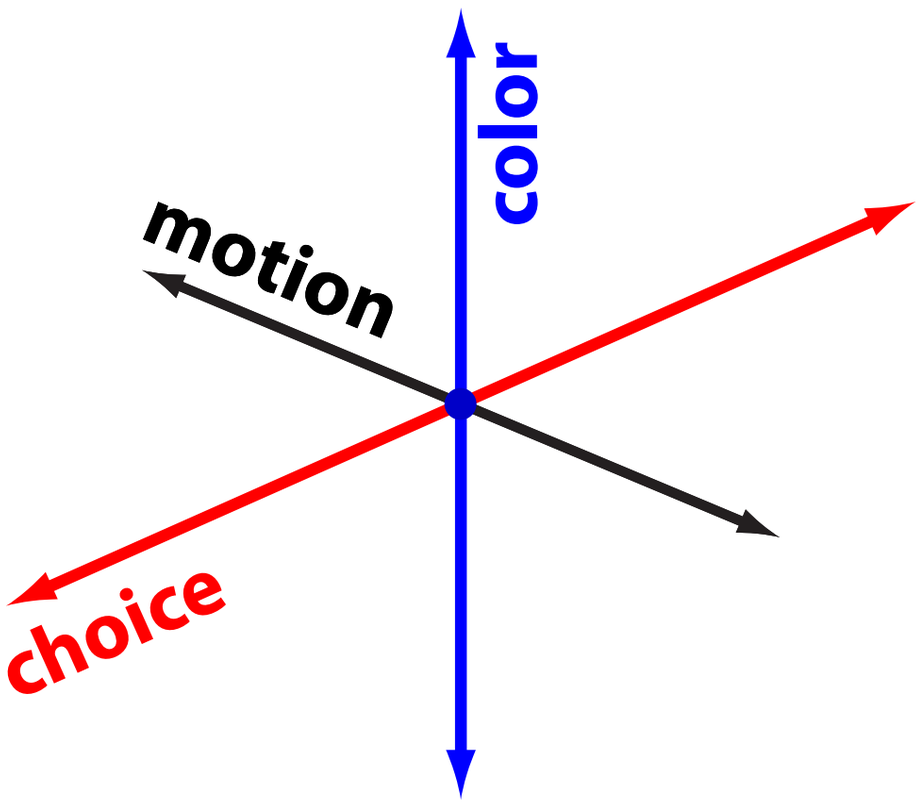


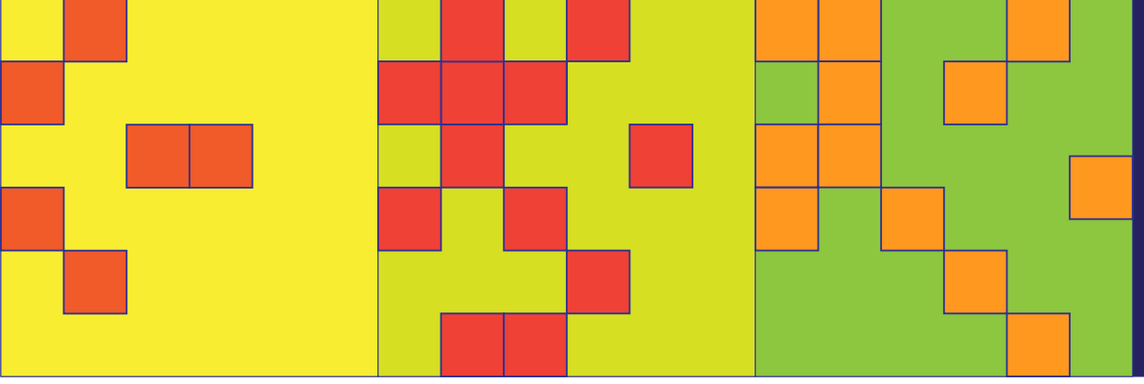
How does selective integration occur?

Motion context



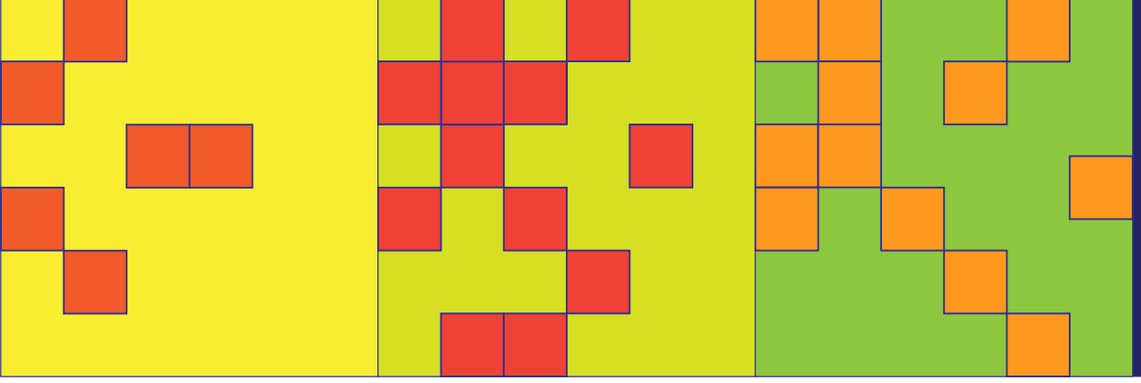
Color context





Conclusions from data

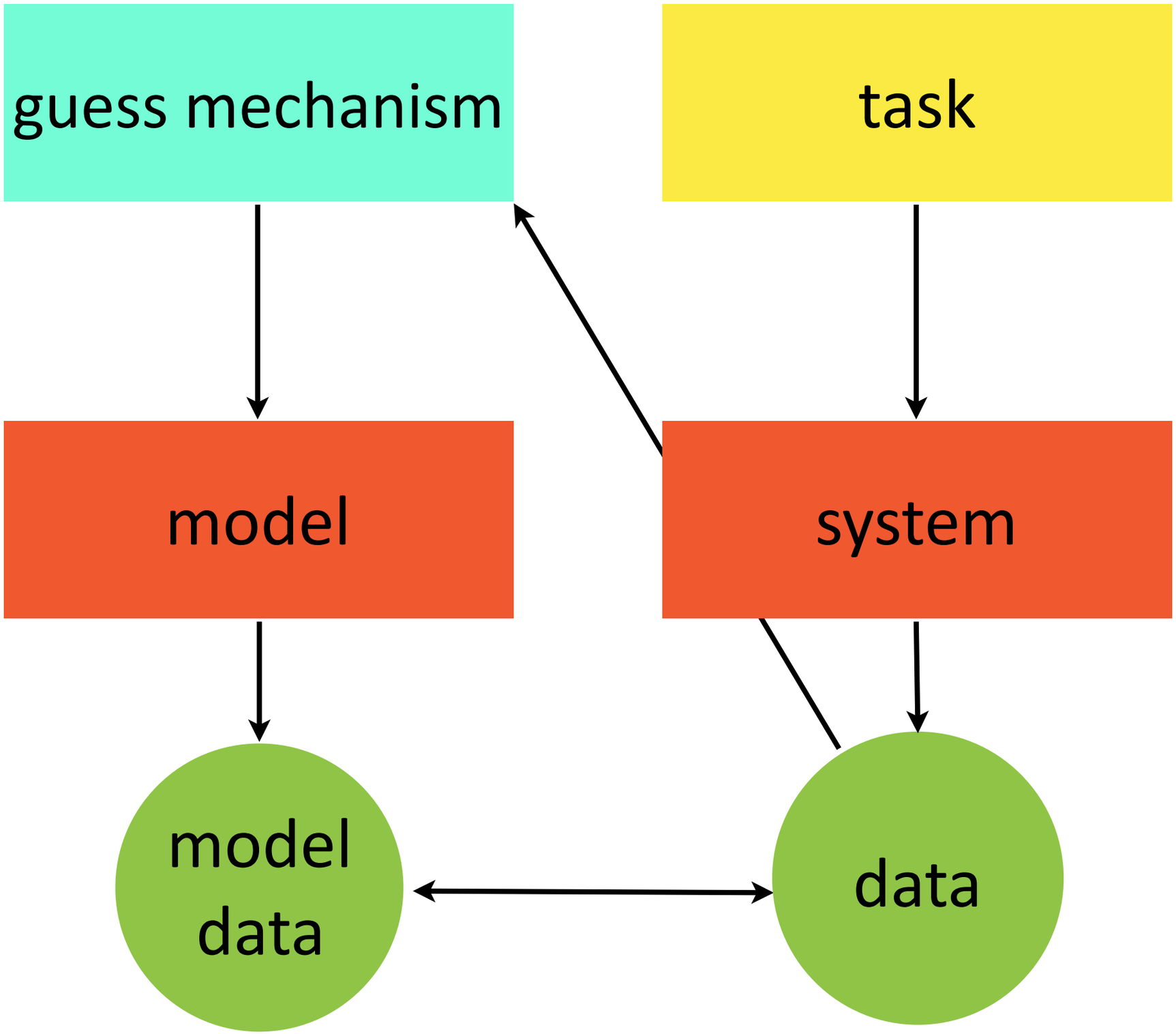
- Task-relevant variables are mixed in the responses of single neurons, but separable and systematically represented in the population.
- Irrelevant inputs are not filtered-out. Selection of relevant inputs occurs late, possibly within PFC.
- Sensory inputs elicit population responses that differ from those corresponding to a choice.
- The directions of choice and of the inputs are largely independent of context (only shift in state space)



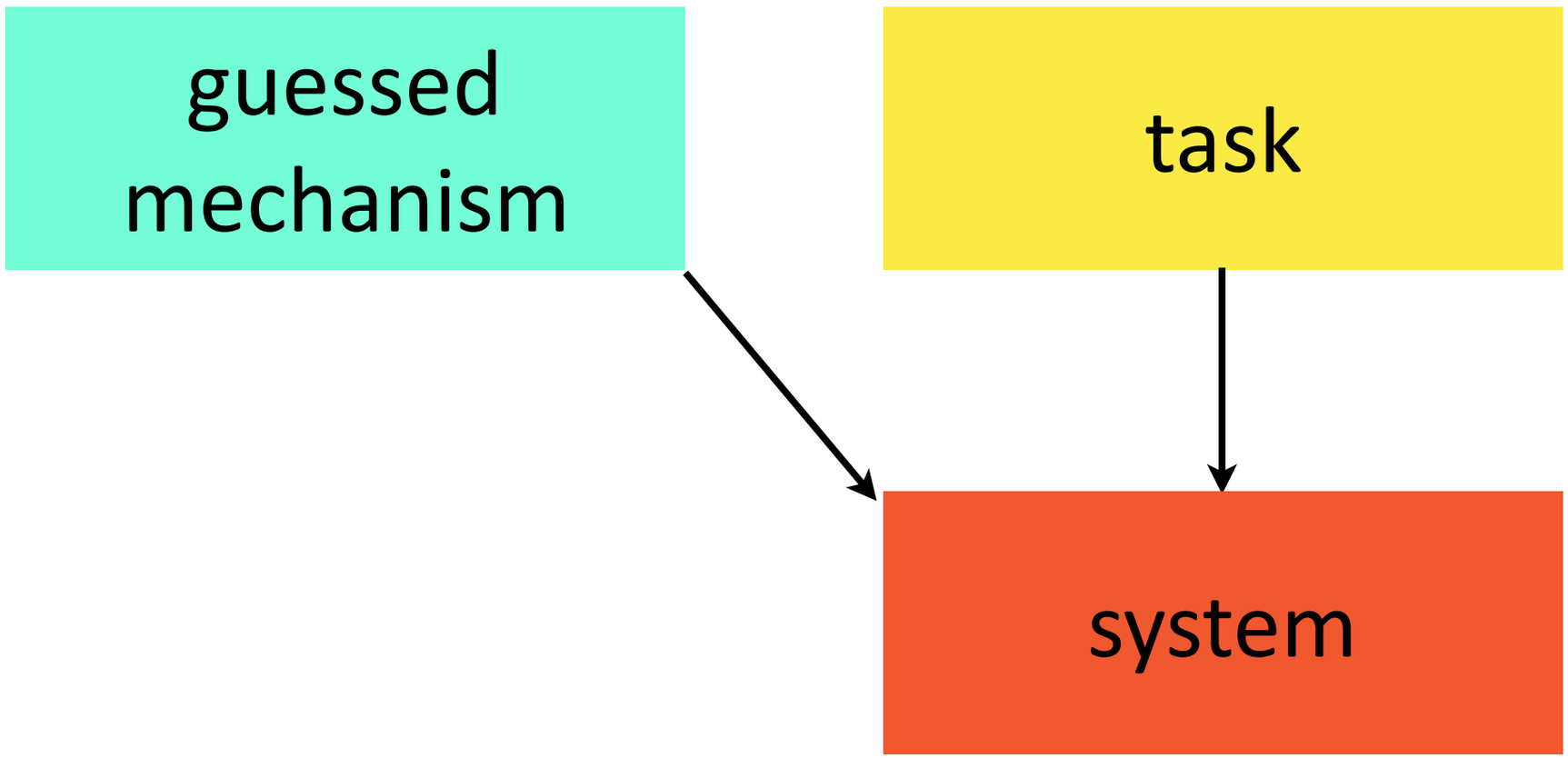
Contextual decision making (model)

How could selective integration occur?

Traditional Modeling Framework

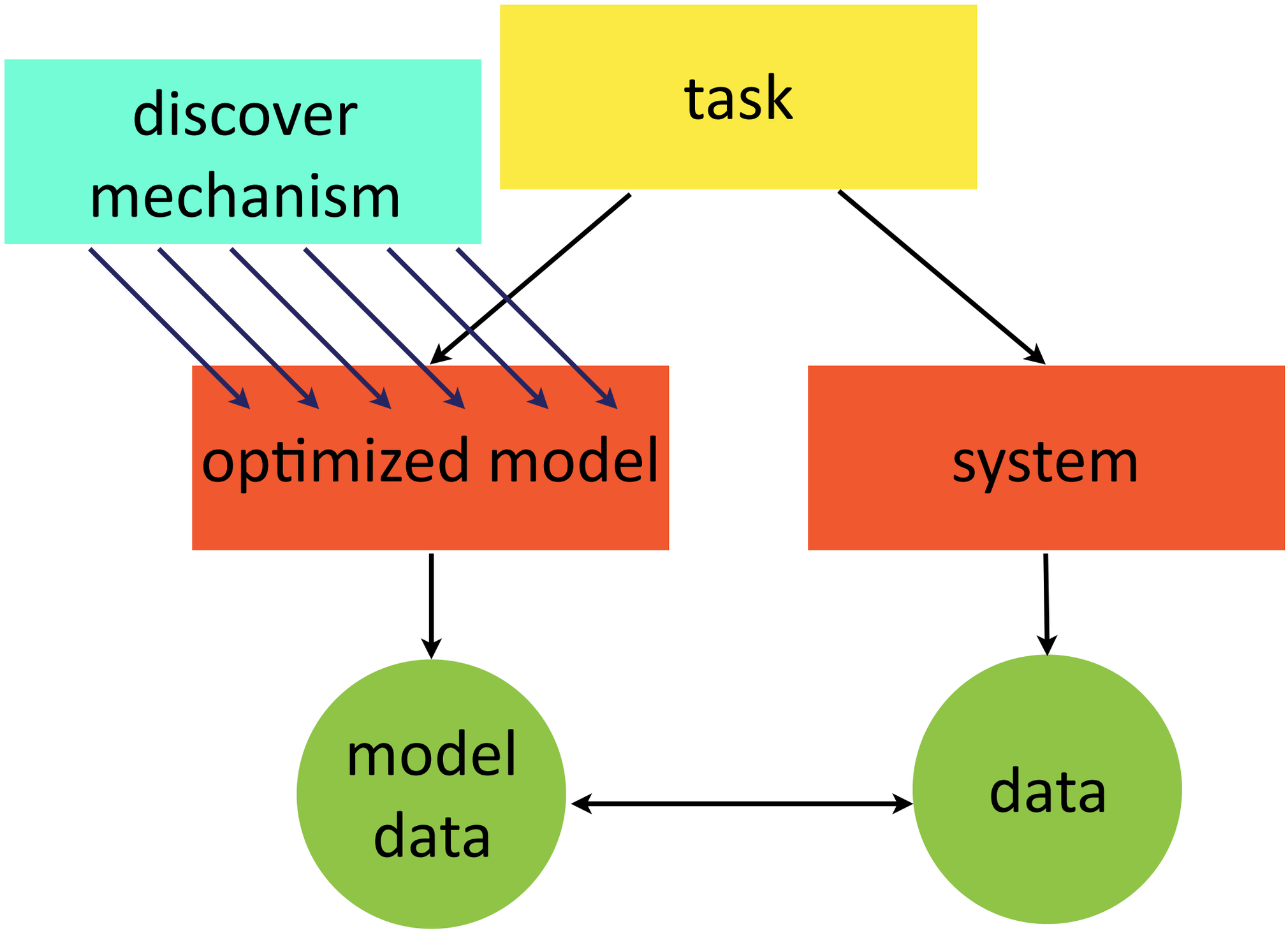


Traditional Modeling Framework

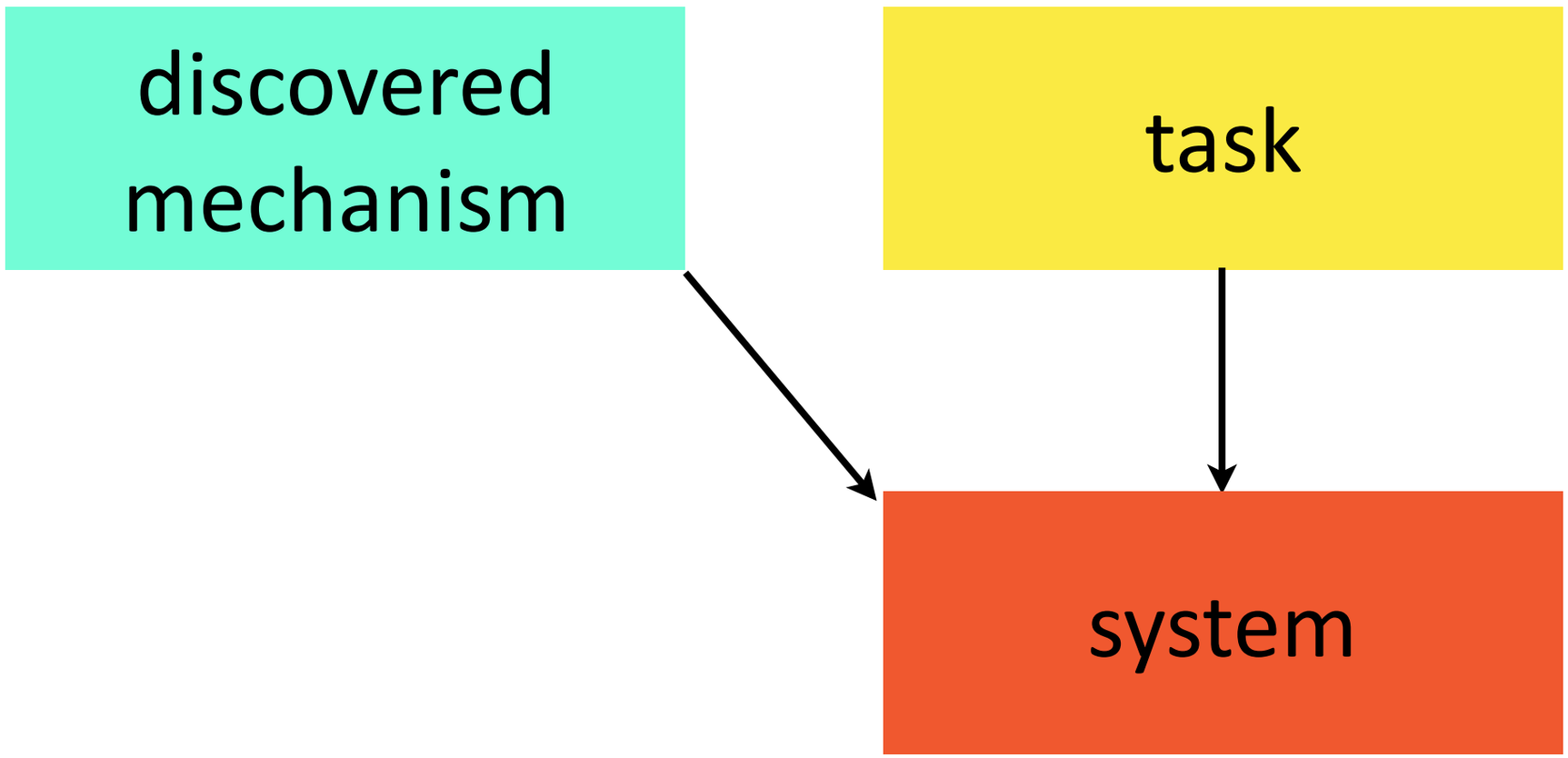


But what should the solutions look like?
Are we too clever?
Not clever enough?

Optimized Modeling Framework



Optimized Modeling Framework

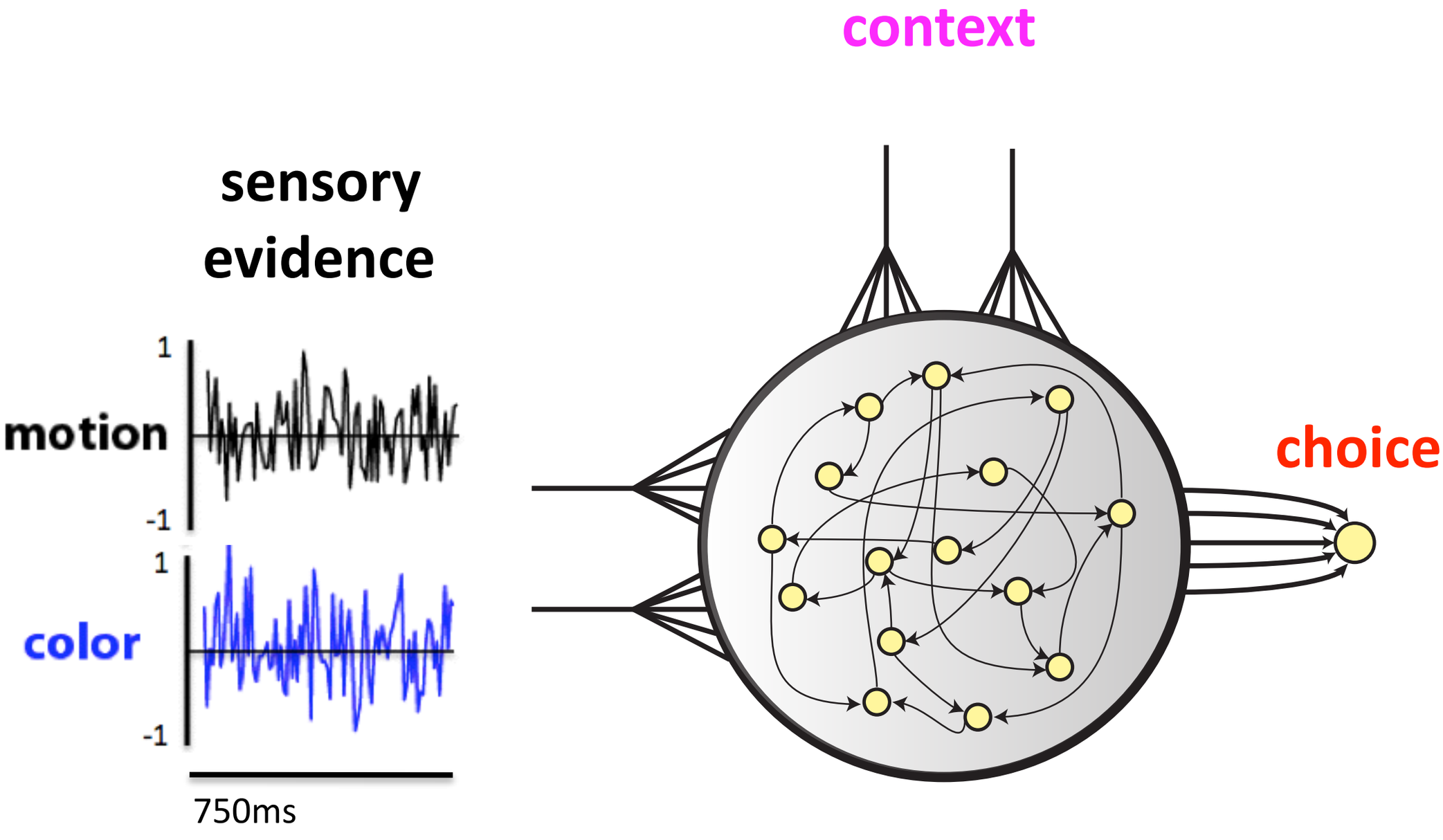


This is a concrete and detailed hypothesis generating mechanism.

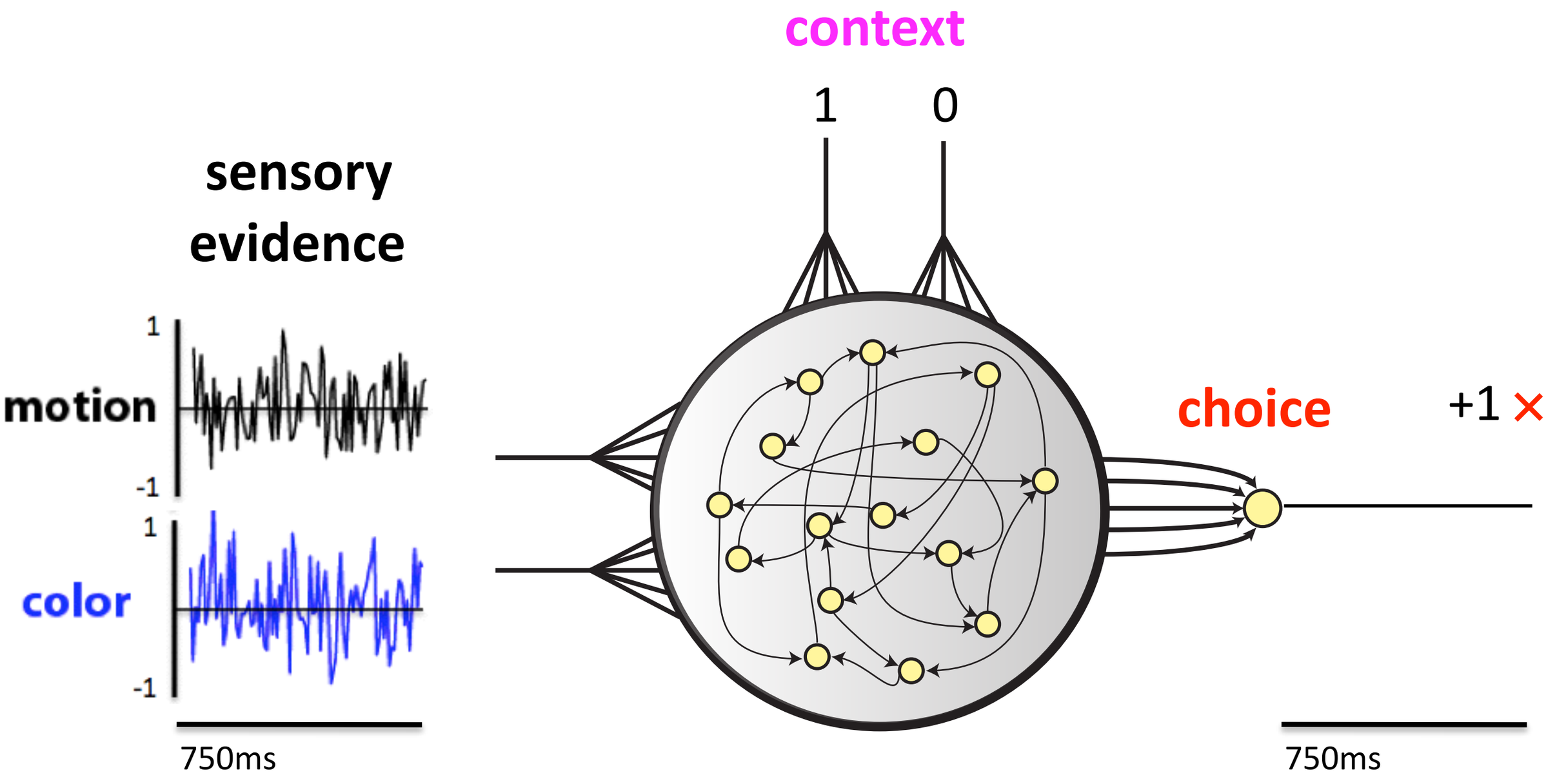
Zipser & Andersen, 1988

Fetz, 1993

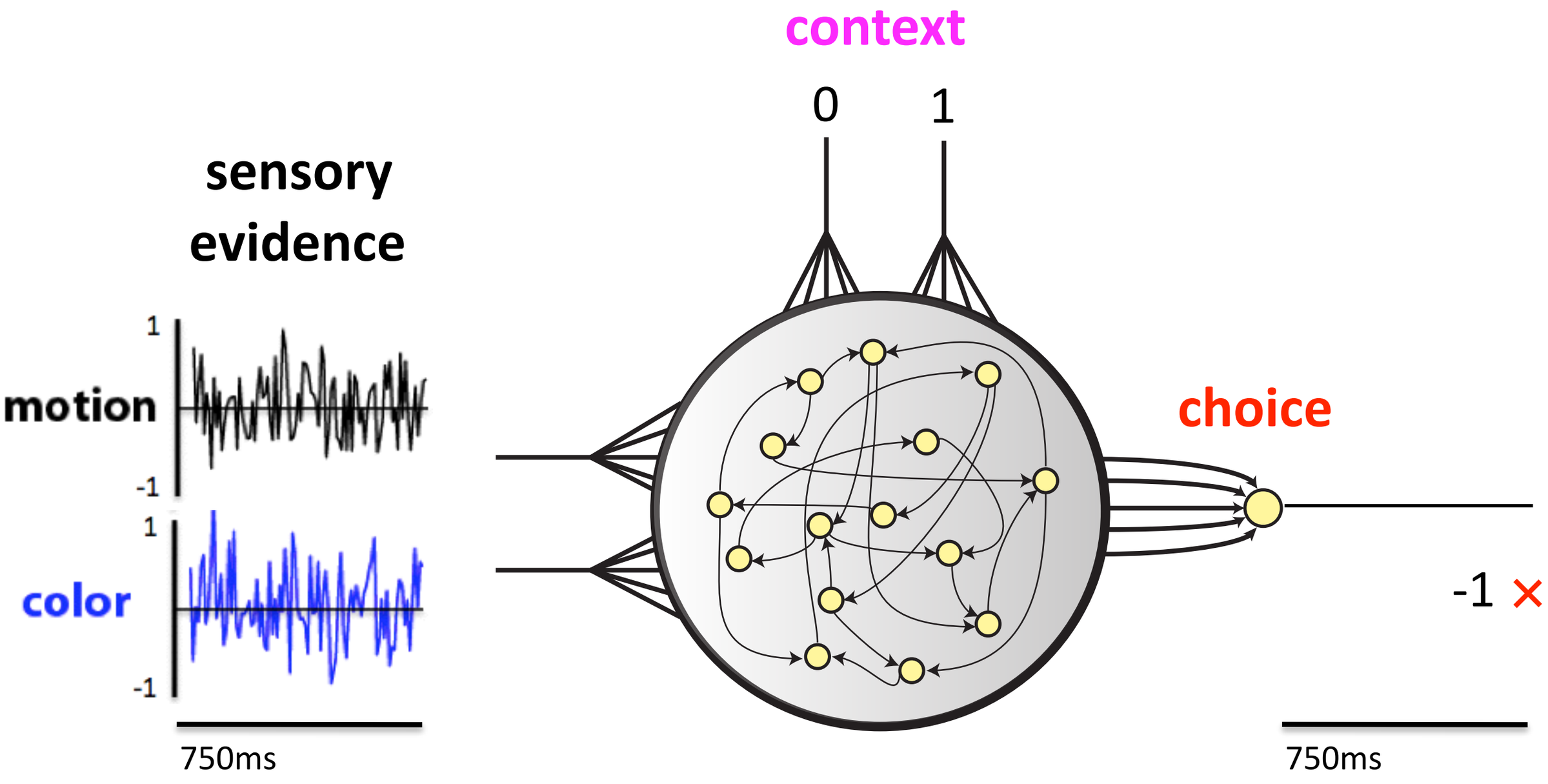
A neural-network model of selective integration



A neural-network model of selective integration

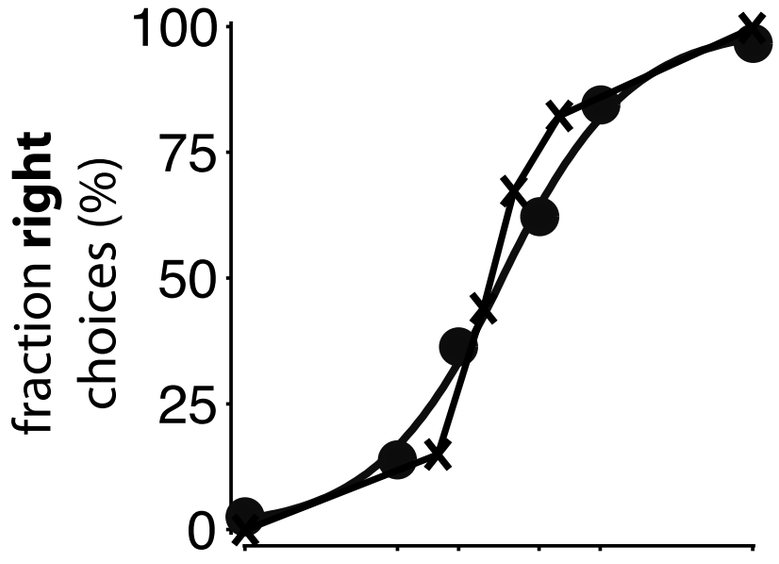


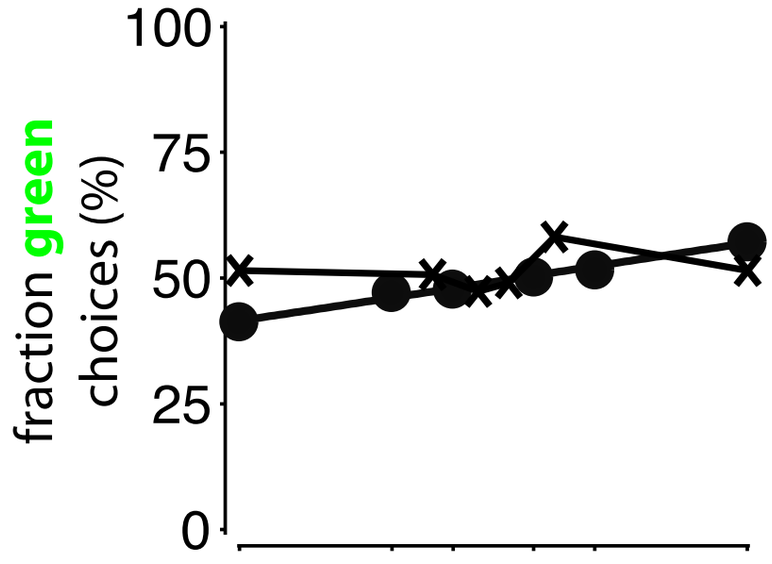
A neural-network model of selective integration



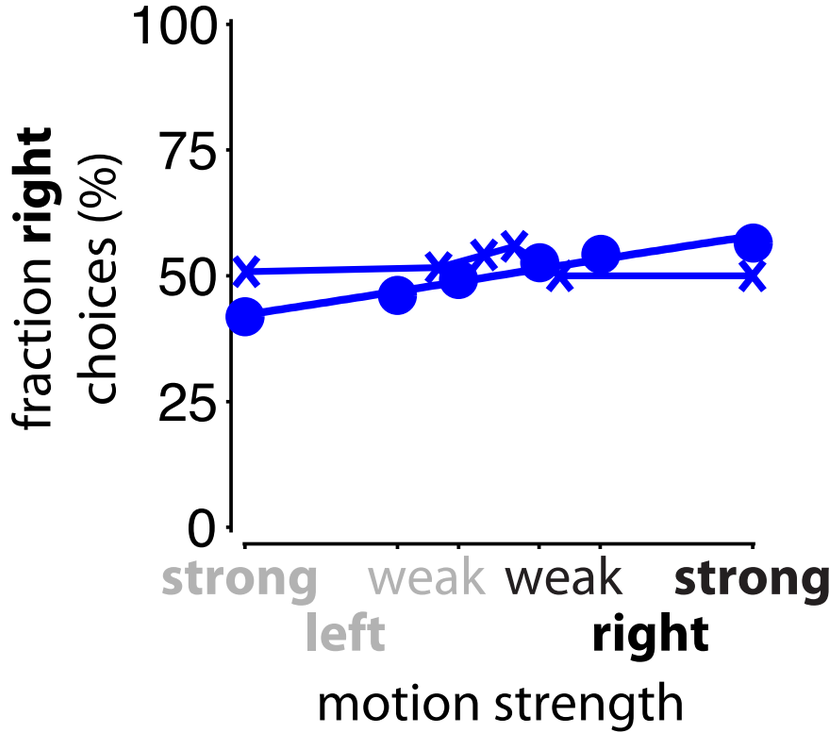
Model “Behavior”

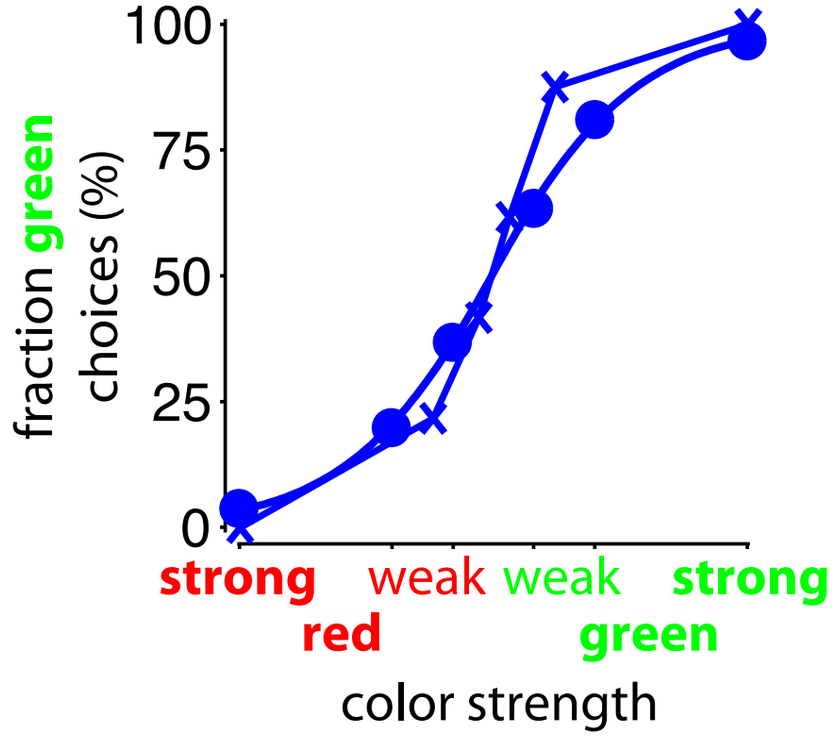
motion trials



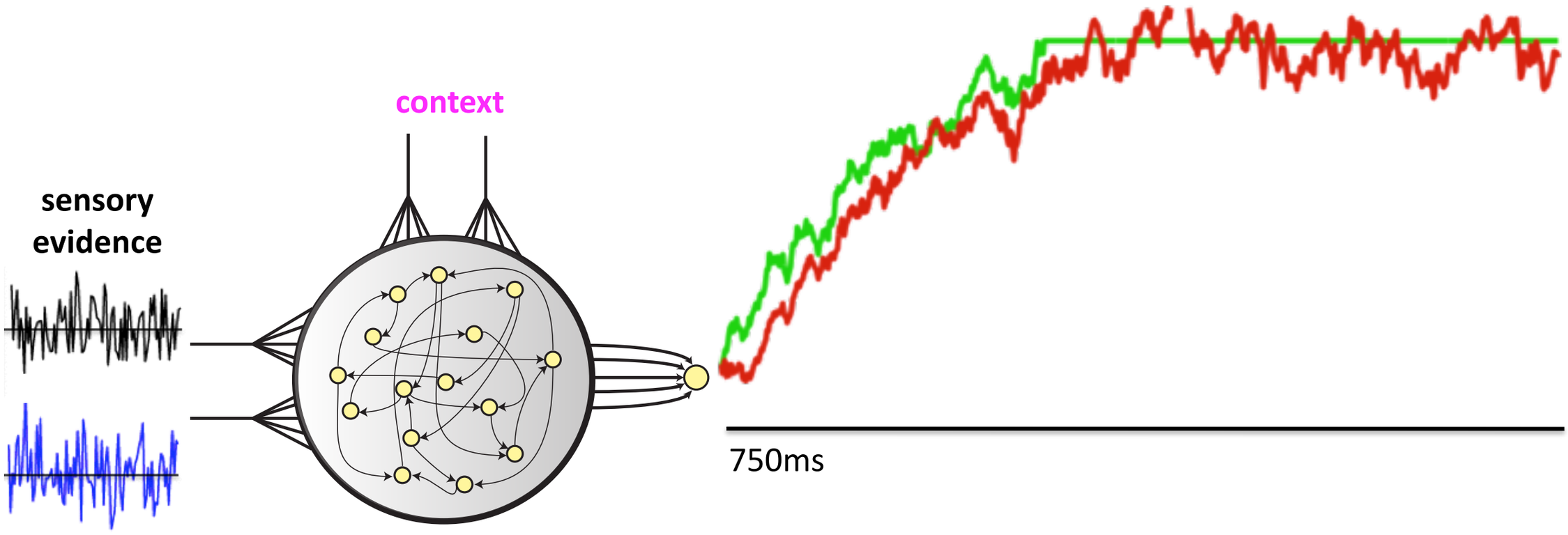


color trials





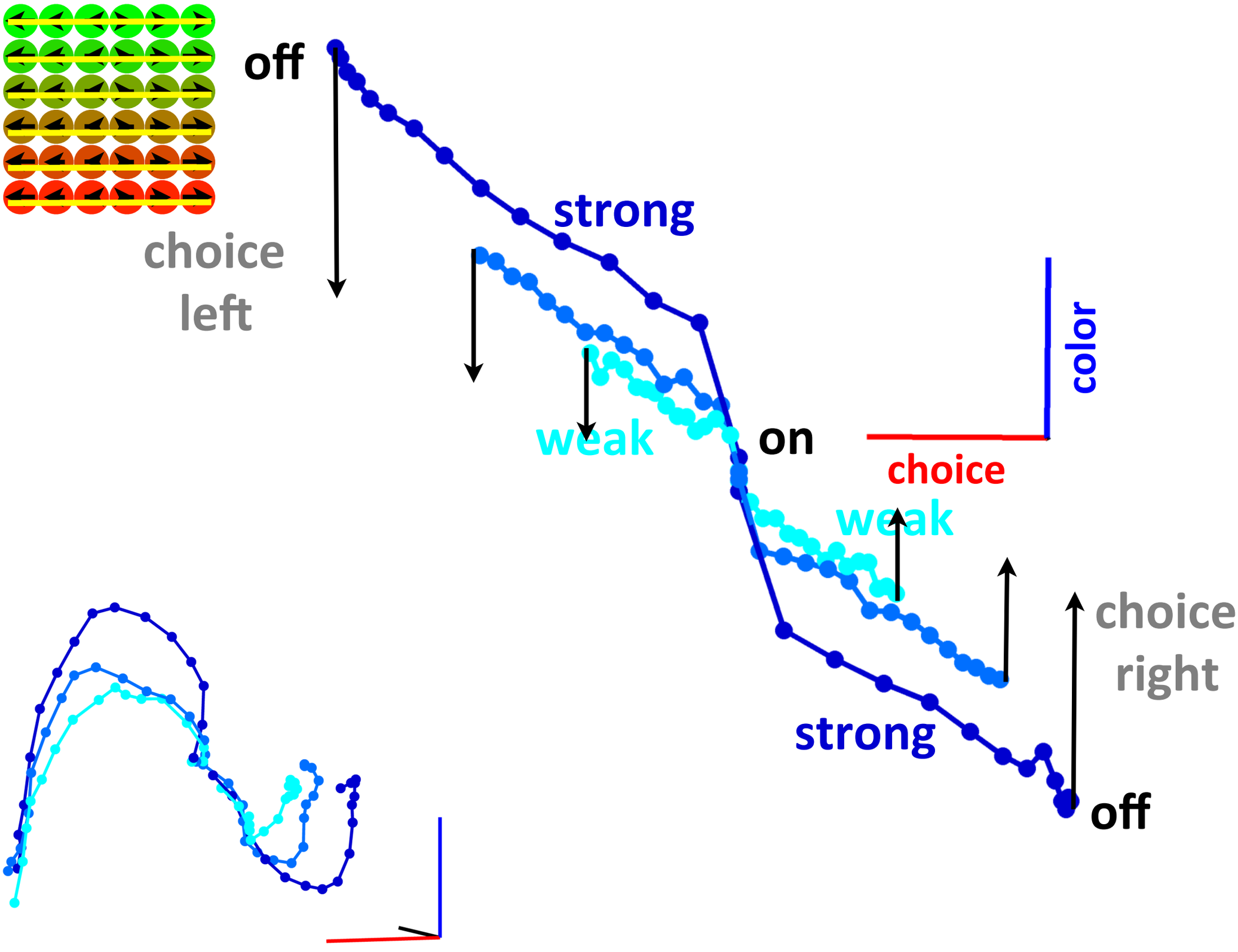
The trained network creates a bounded integrator



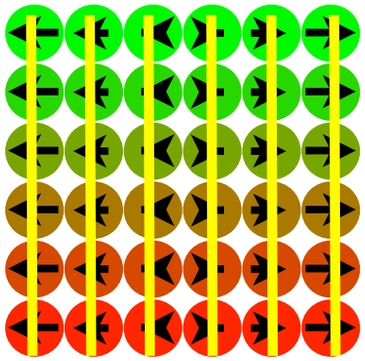
Network Output

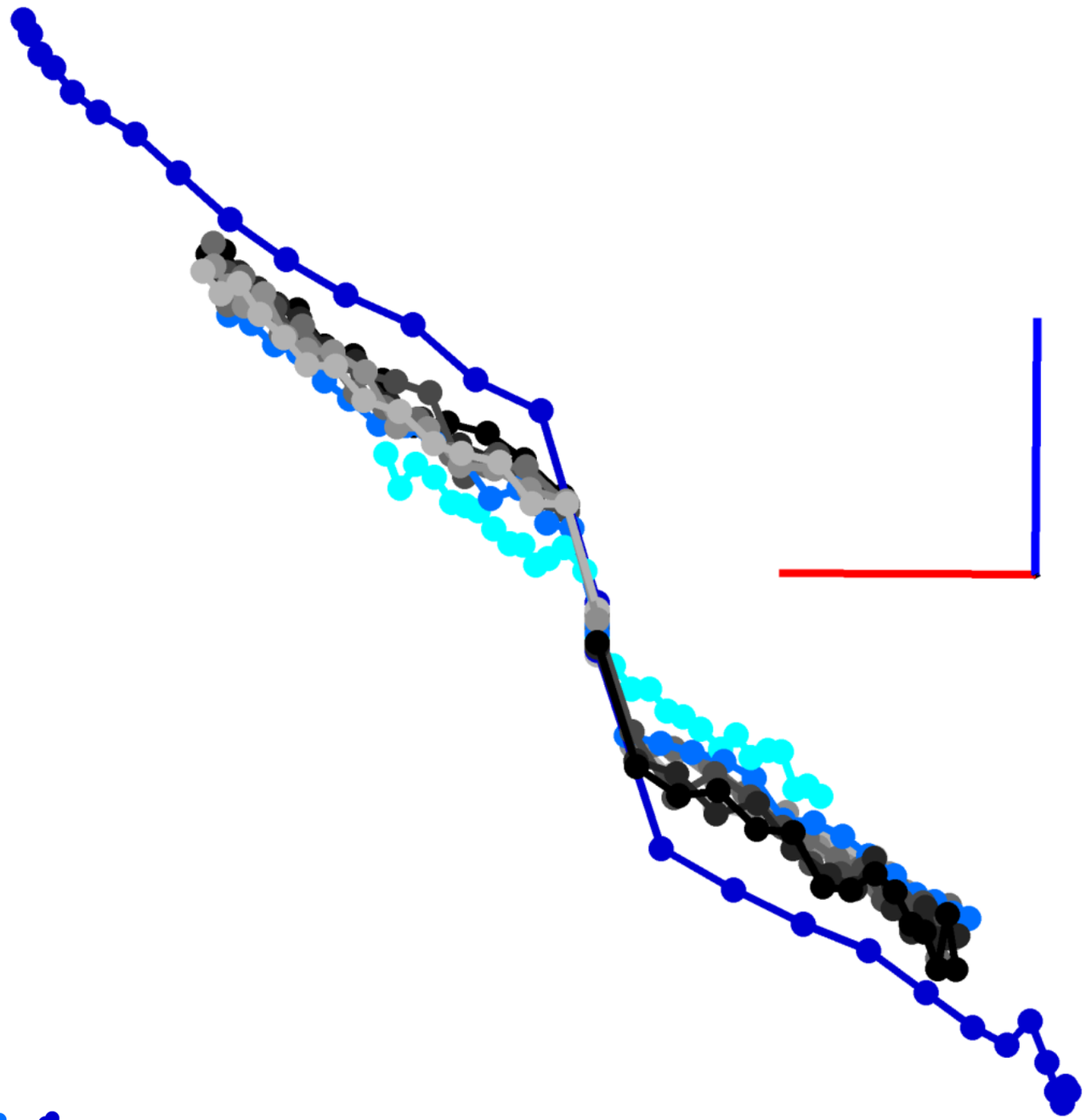
Bounded Integrator

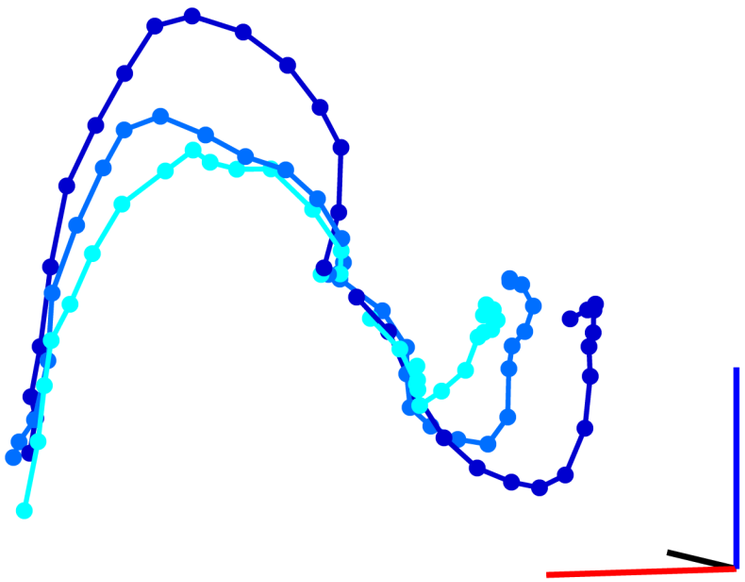
Model trajectories during color trials

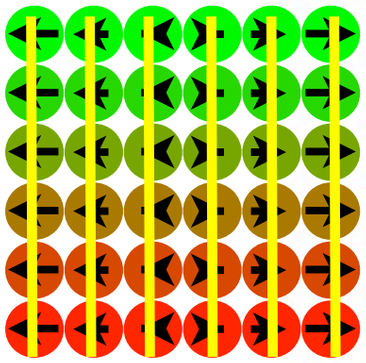


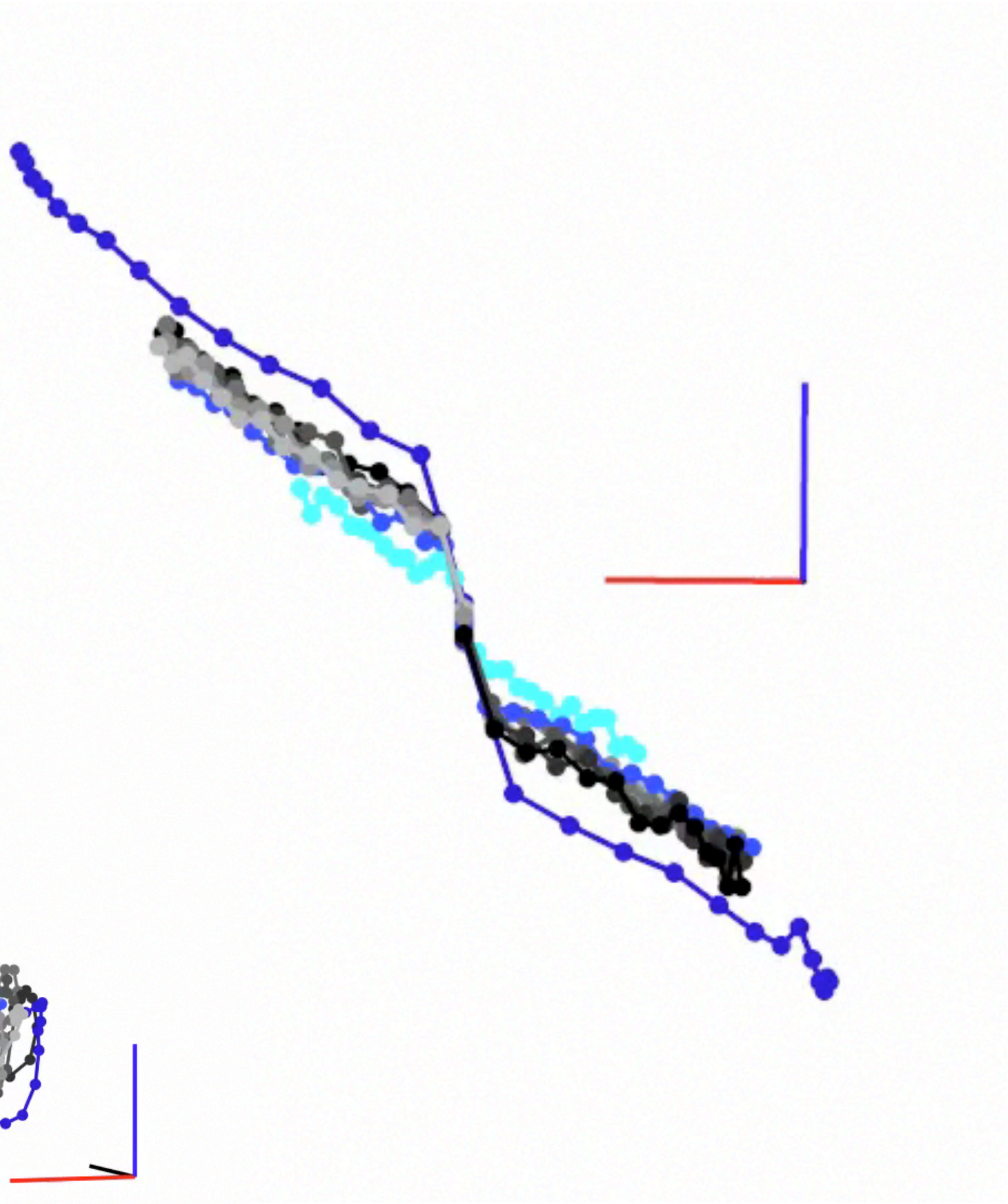
Model trajectories during color trials

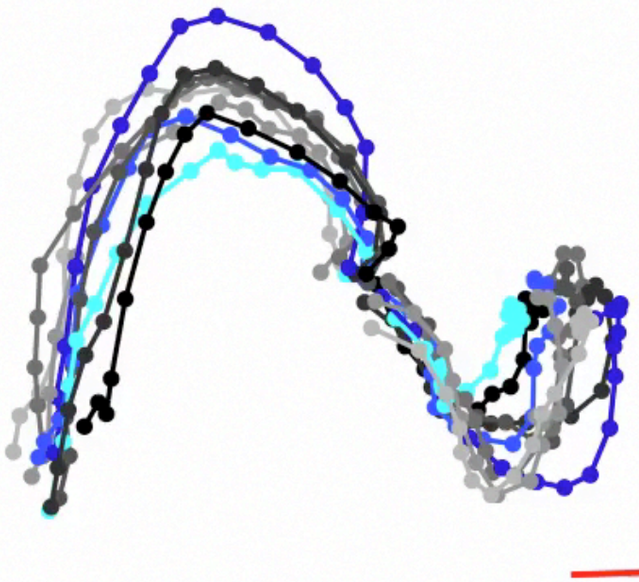




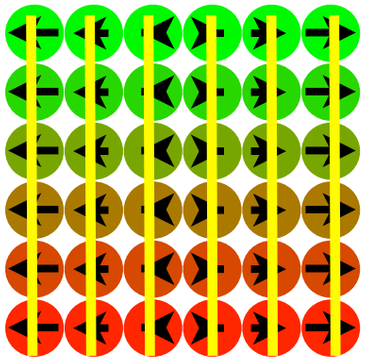




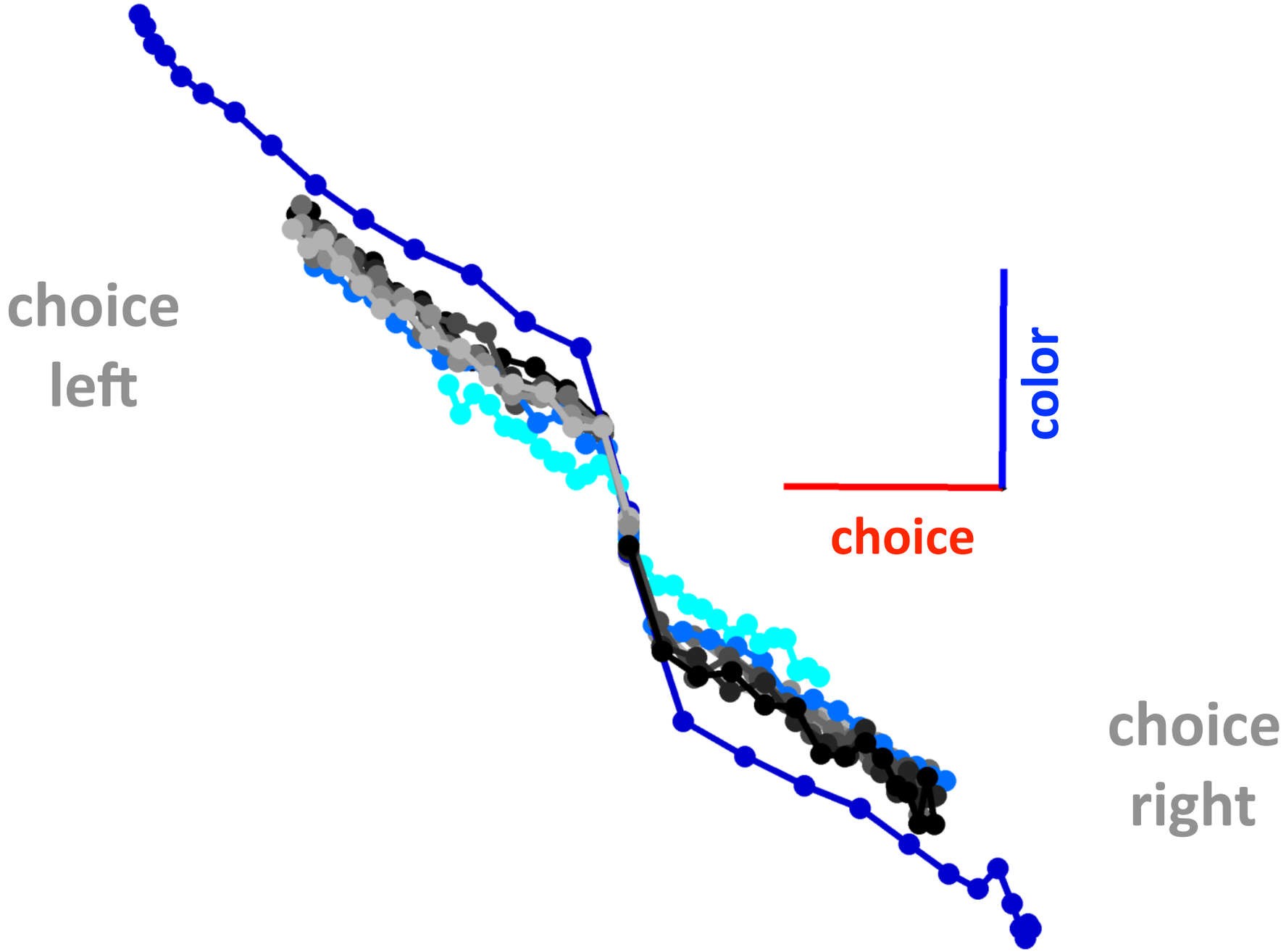




Model trajectories during color trials



How does integration happen?



What is a fixed point?

$$\dot{\mathbf{x}} = \mathrm{F}(\mathbf{x})$$

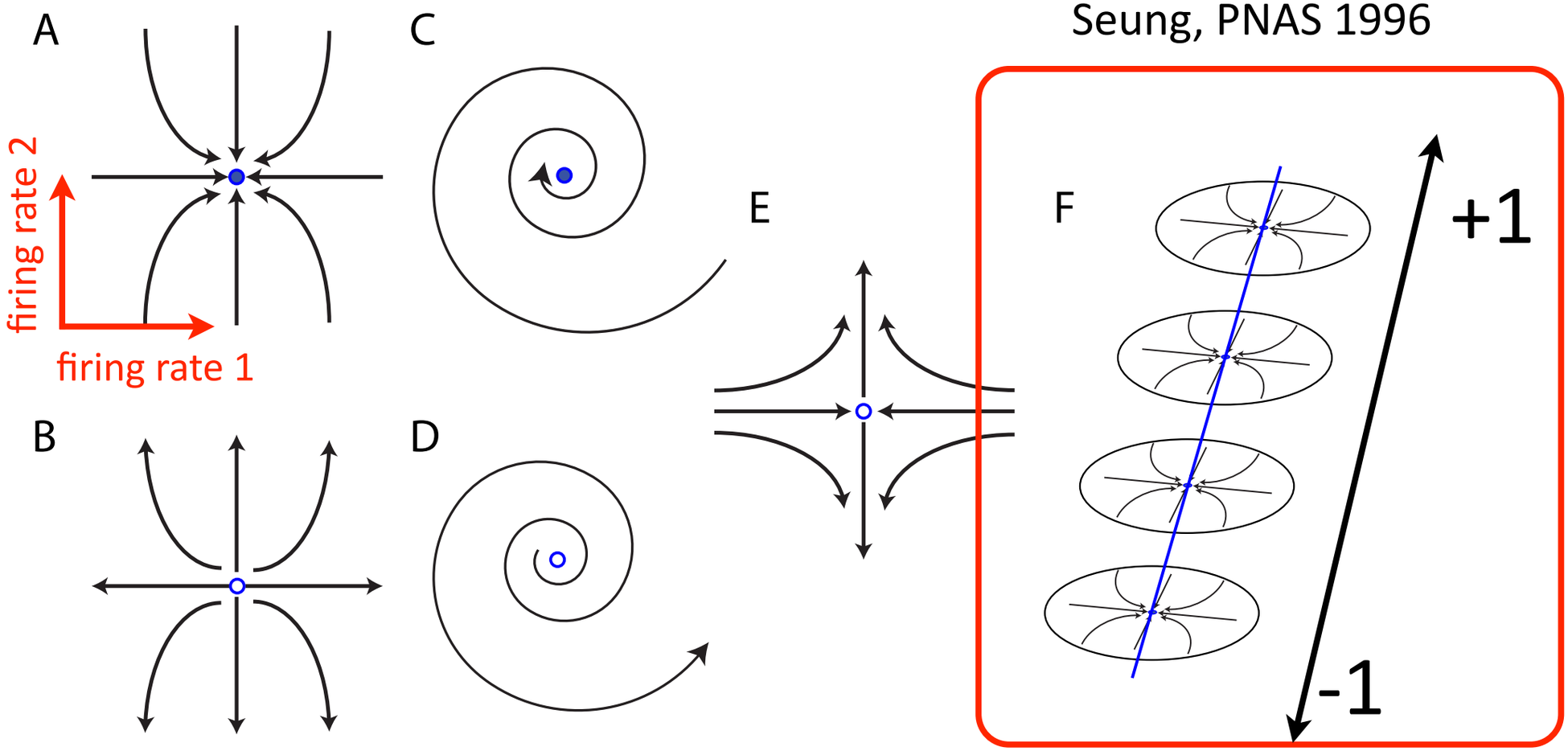
Any nonlinear dynamical system
(e.g. neural circuit)

$$\dot{\mathbf{x}} = \mathbf{0}$$

Zero “motion”

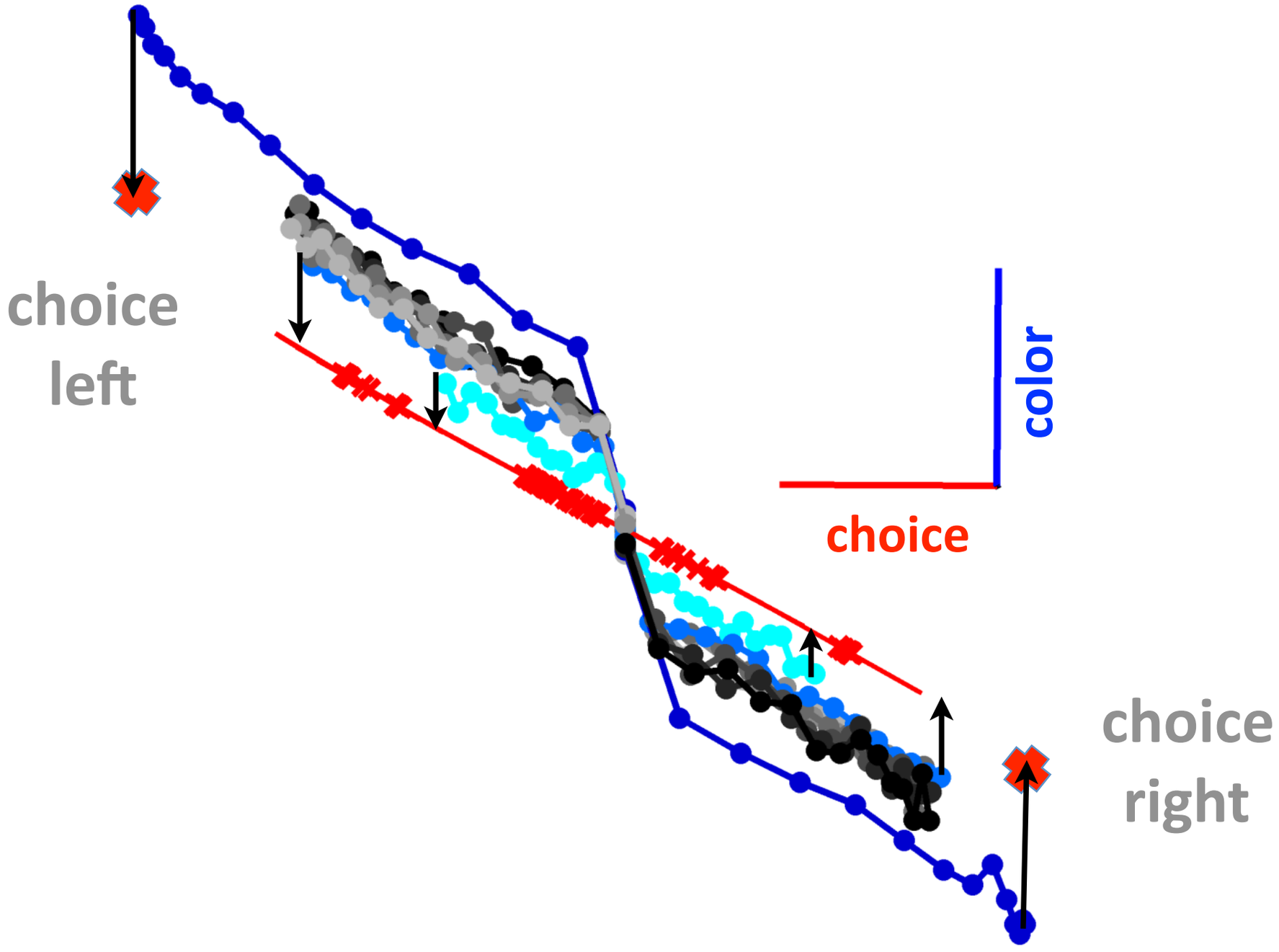
Why are they important?

$$\dot{\mathbf{y}} = \mathbf{M}\mathbf{y}$$

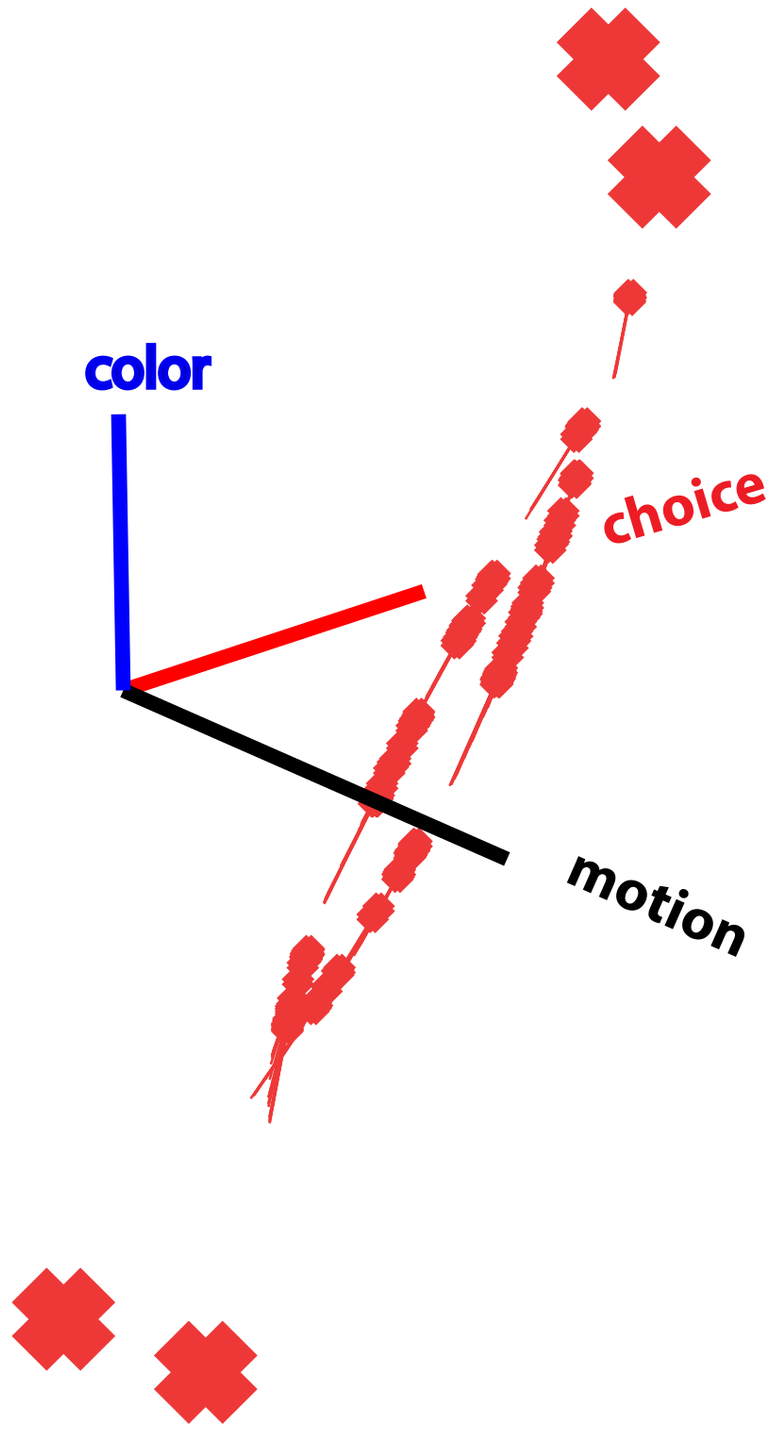


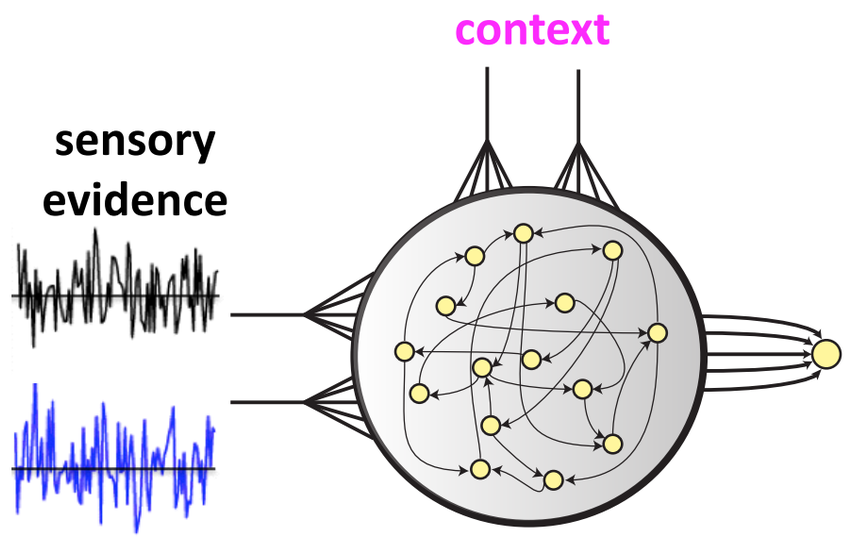
Seung, PNAS 1996

Fixed points make a line attractor



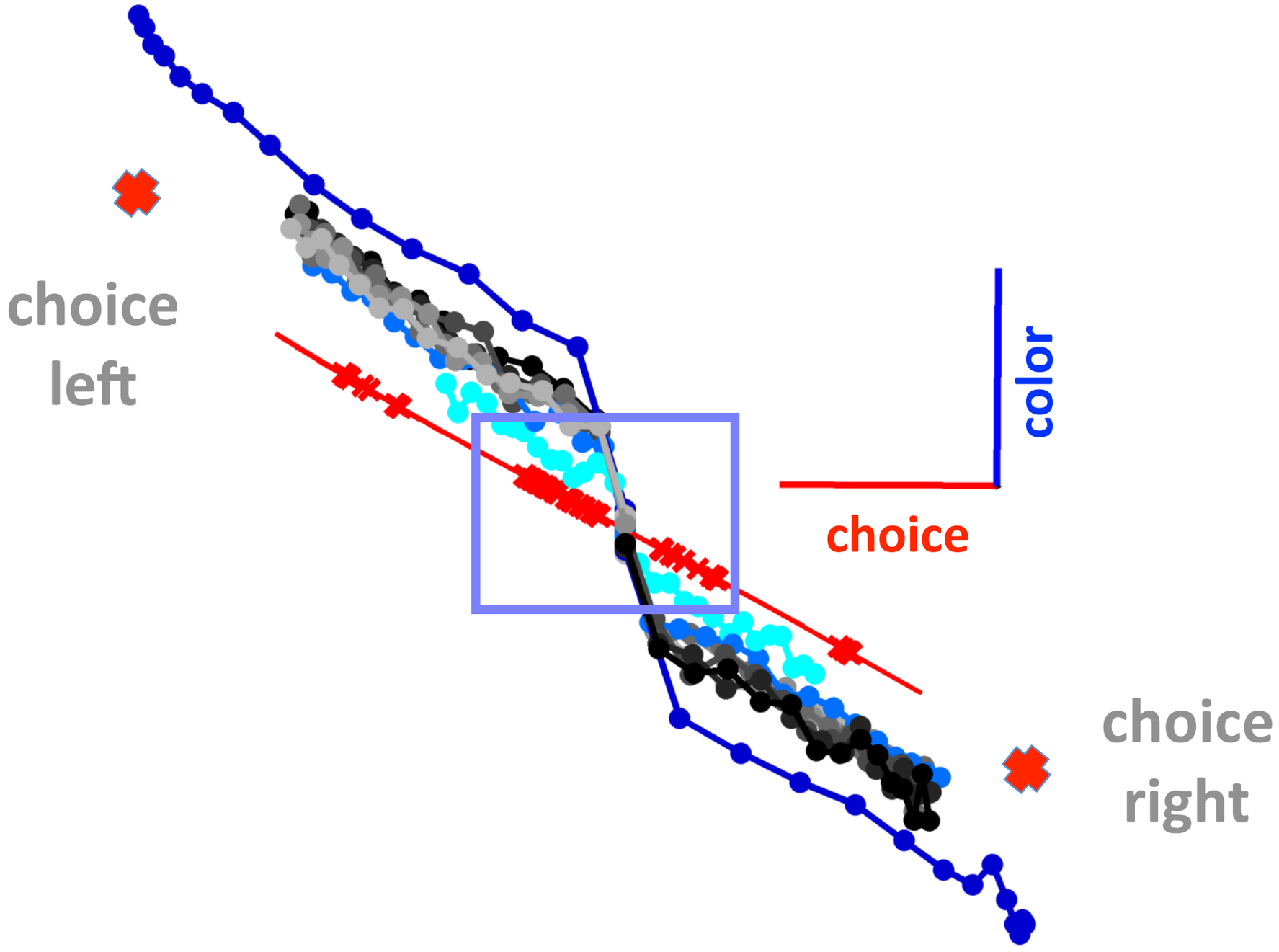
Two line attractors for two contexts





The line attractors are context dependent and never exist at the same time.

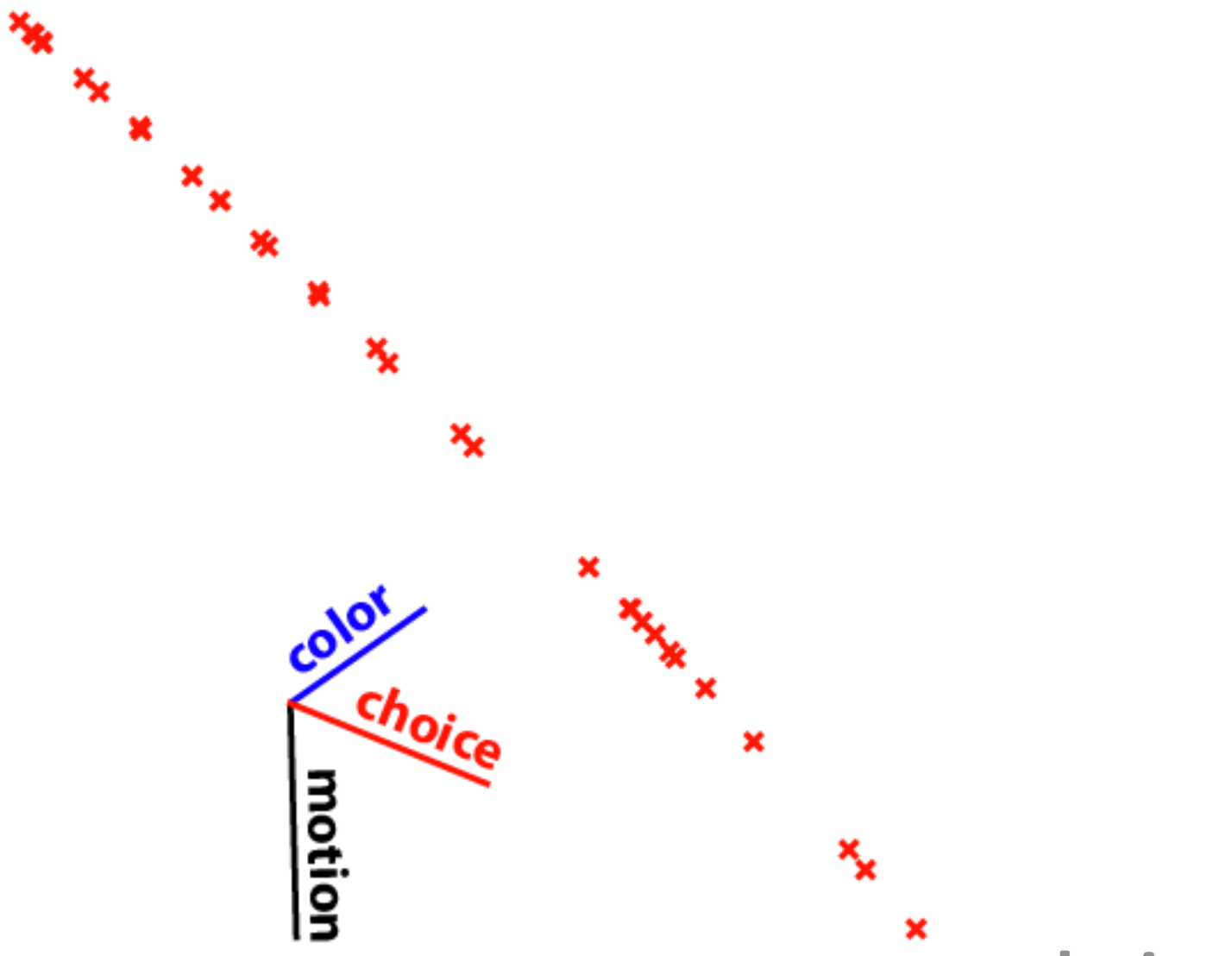
Fixed points make a line attractor



A simulated perturbation experiment

choice
left

color context

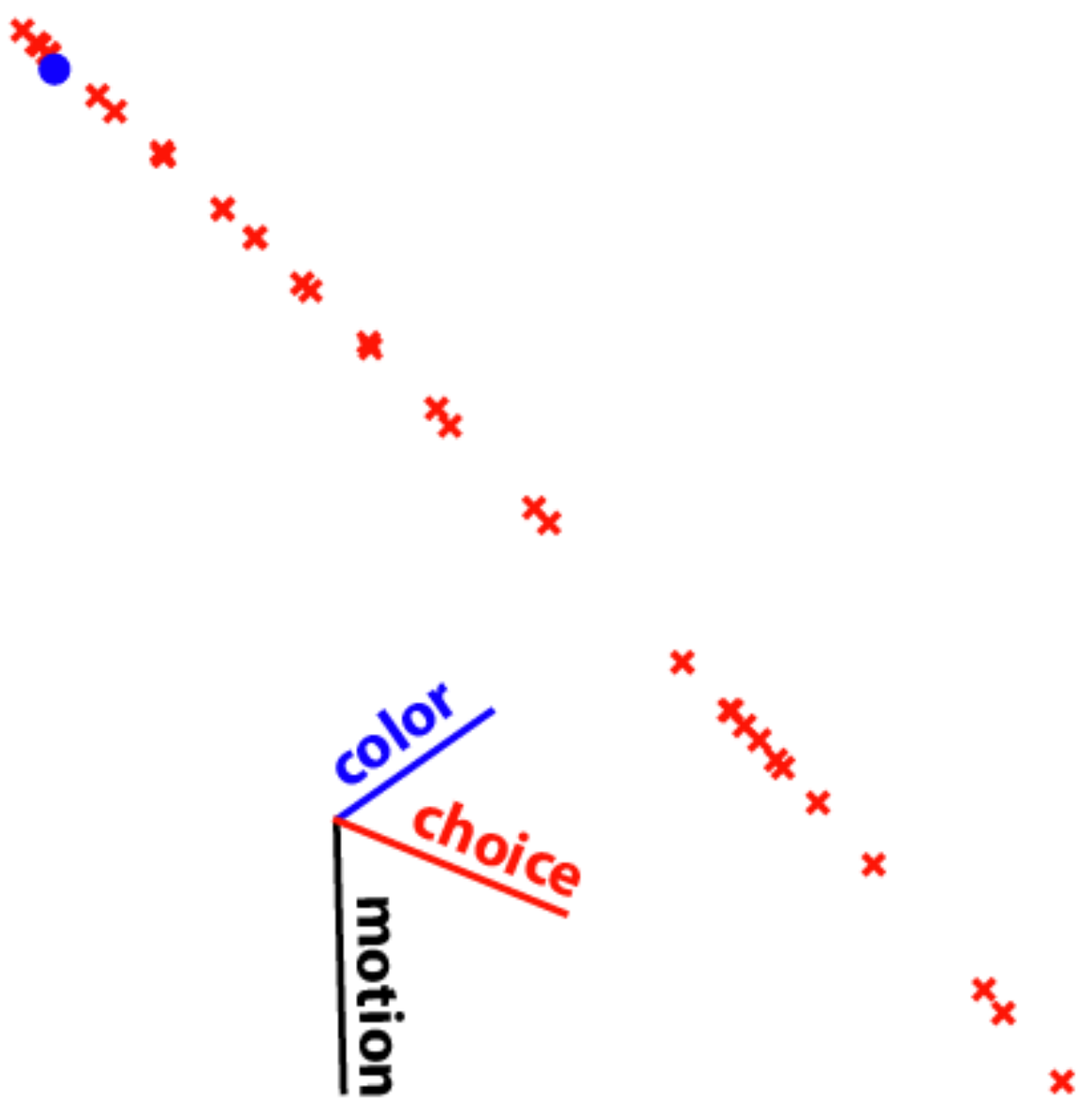


choice
right

A simulated perturbation experiment

choice
left

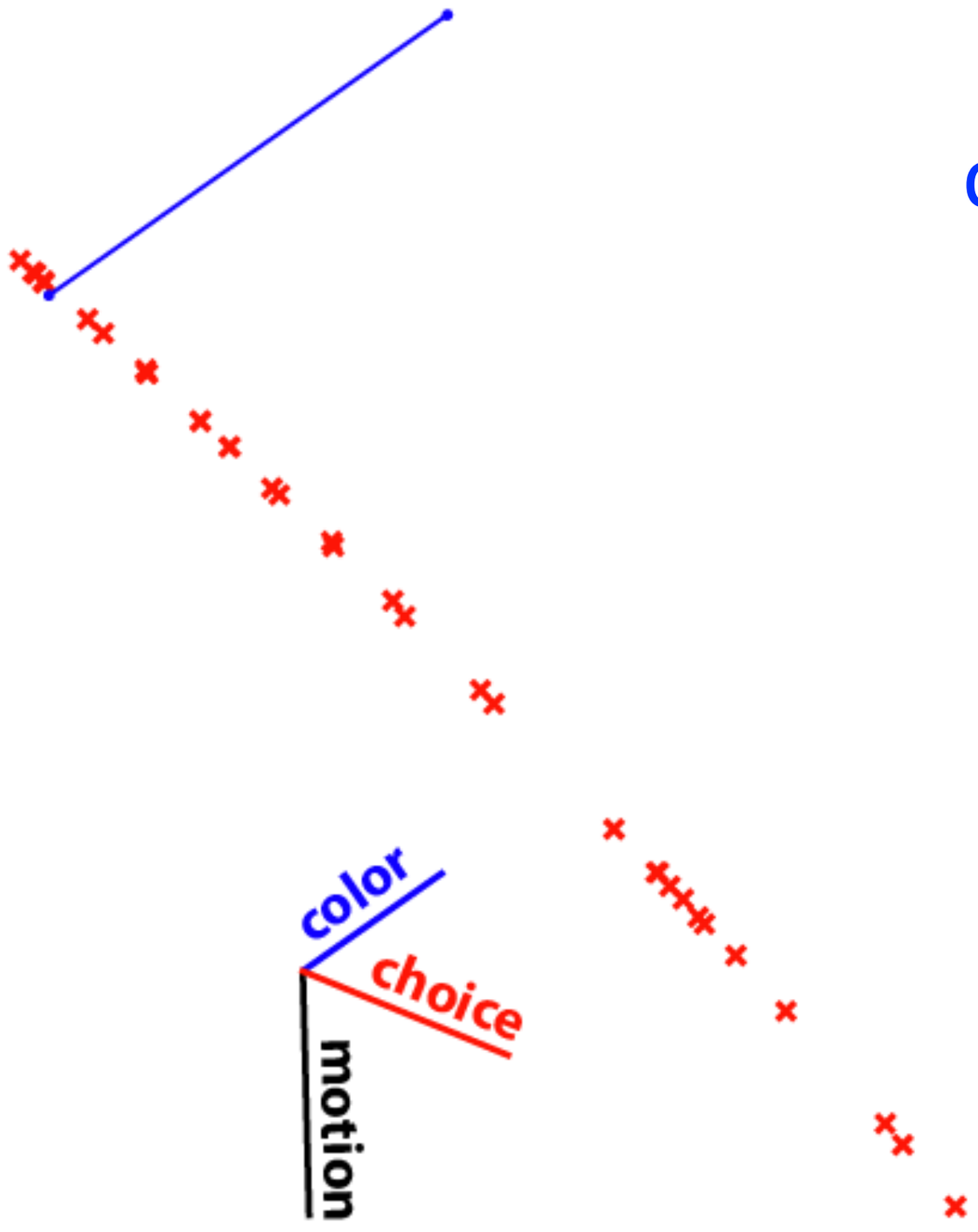
color context



choice
right

A simulated perturbation experiment

choice
left

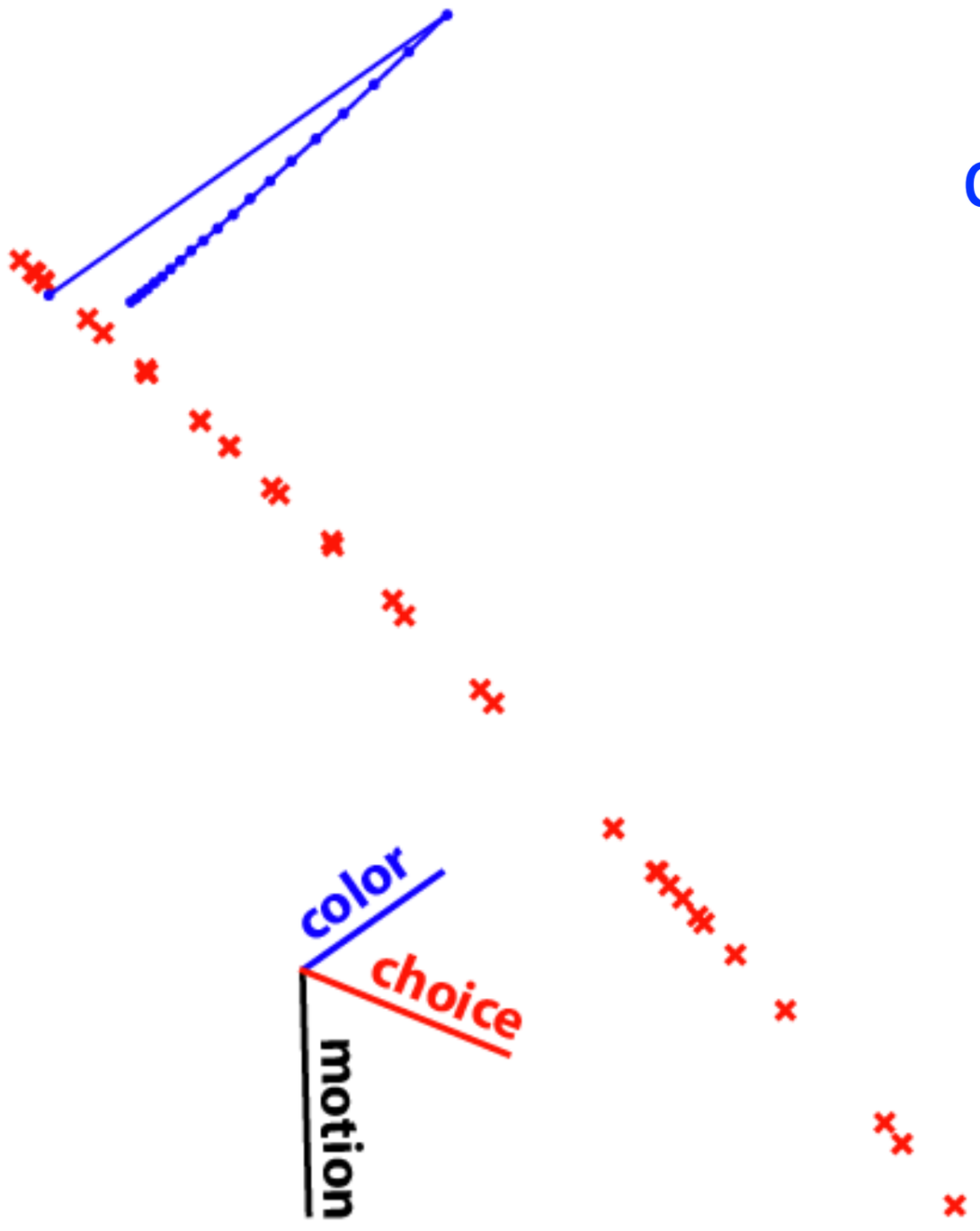


color context

choice
right

A simulated perturbation experiment

choice
left

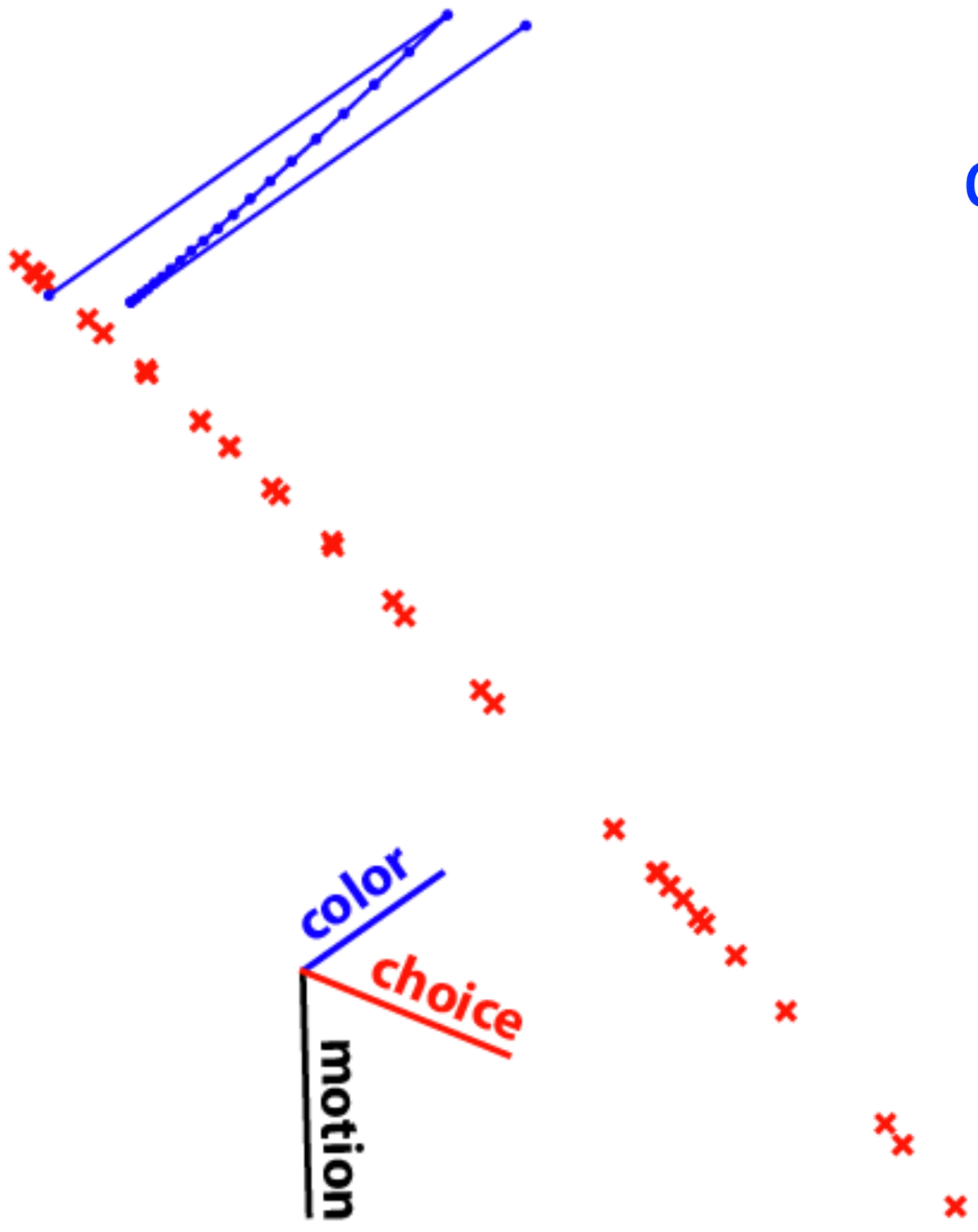


color context

choice
right

A simulated perturbation experiment

choice
left

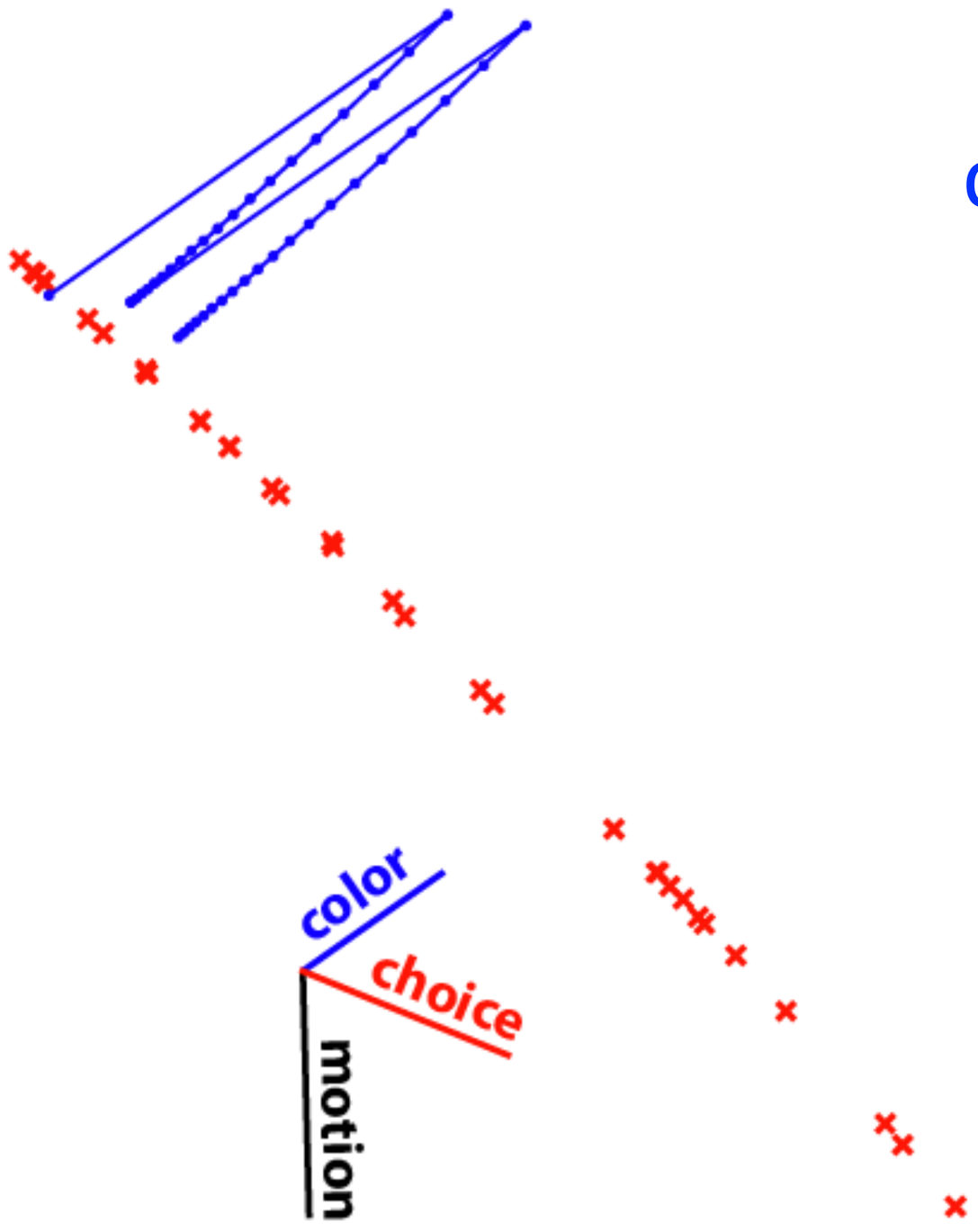


color context

choice
right

A simulated perturbation experiment

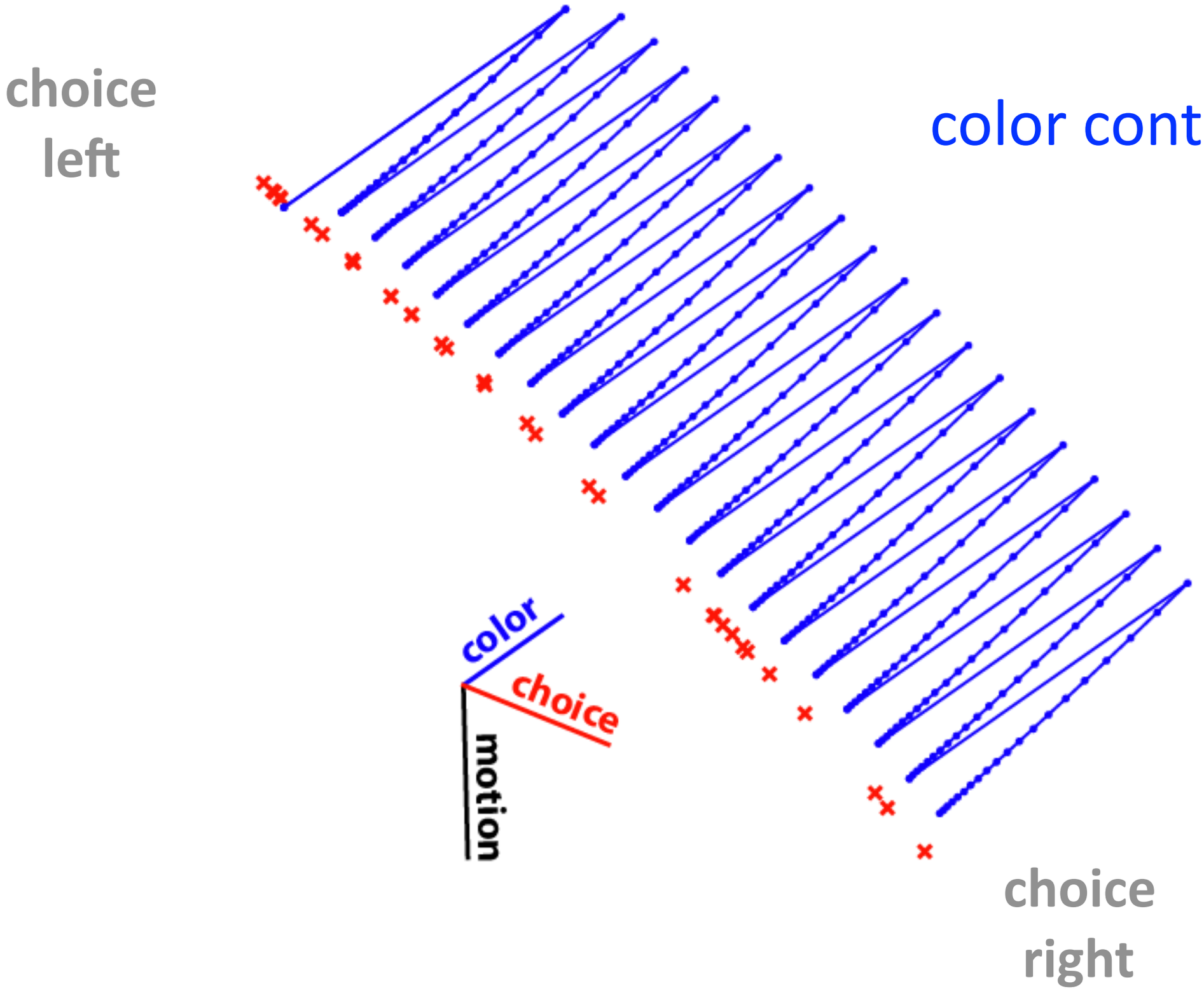
choice
left



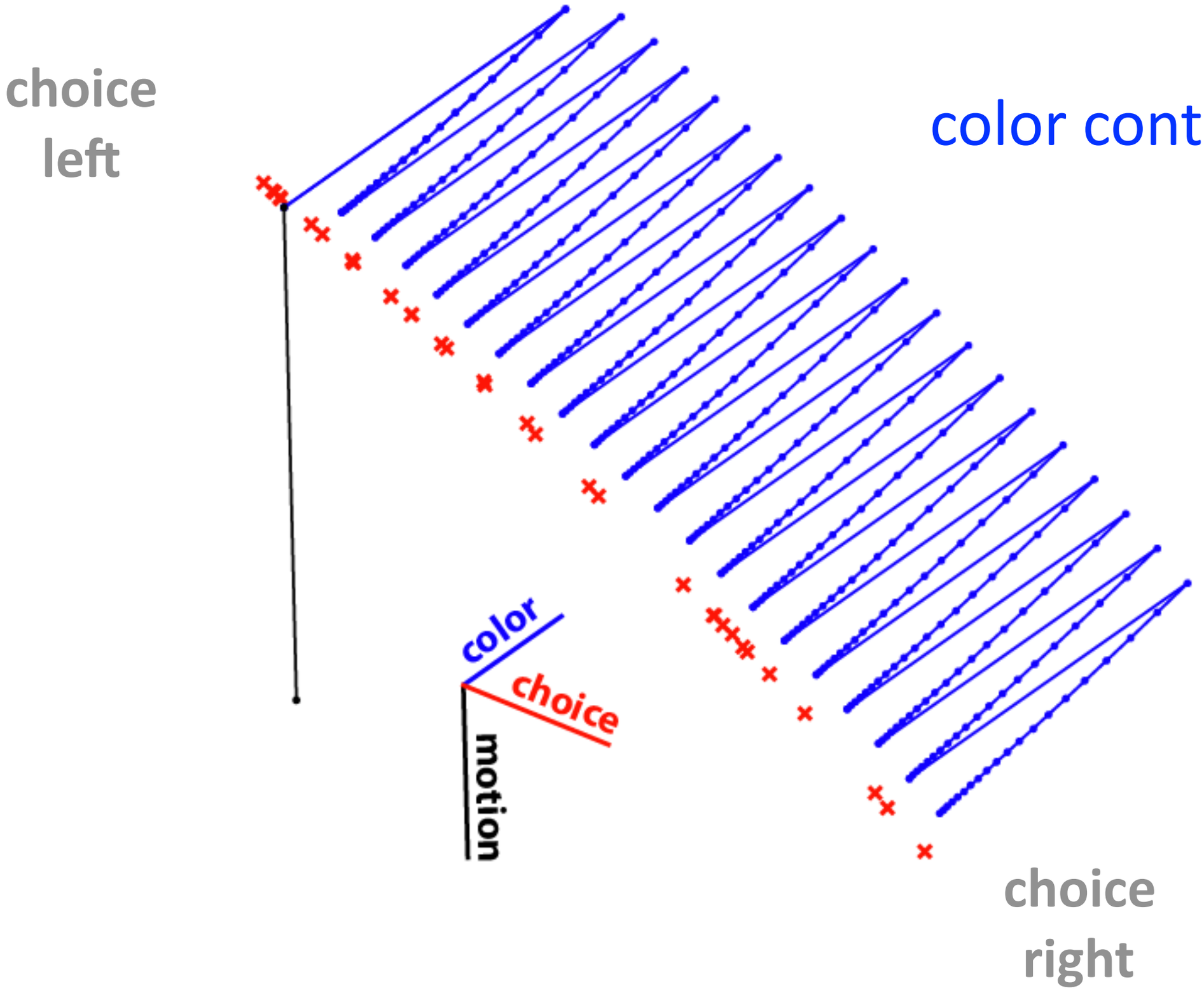
color context

choice
right

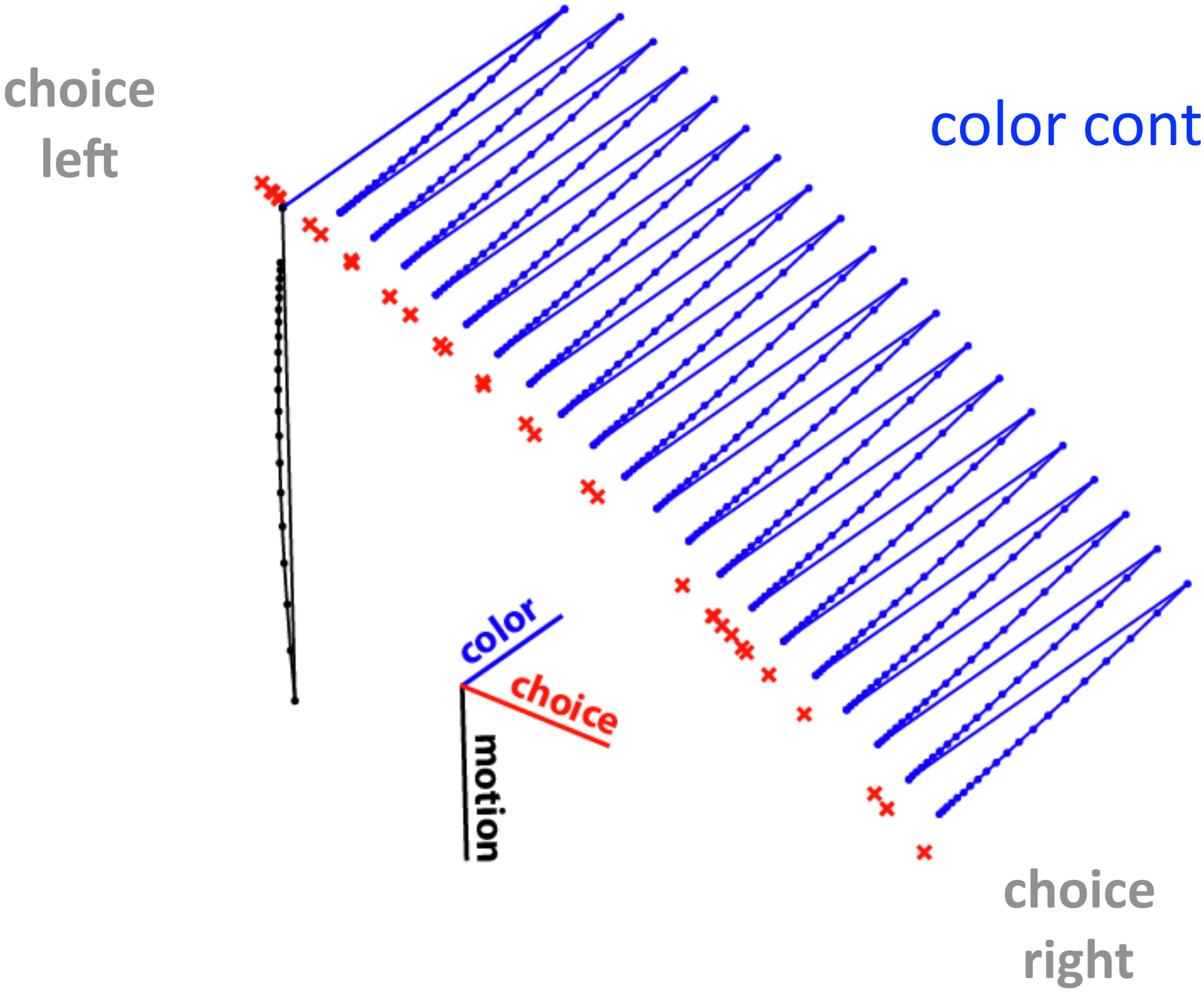
A simulated perturbation experiment



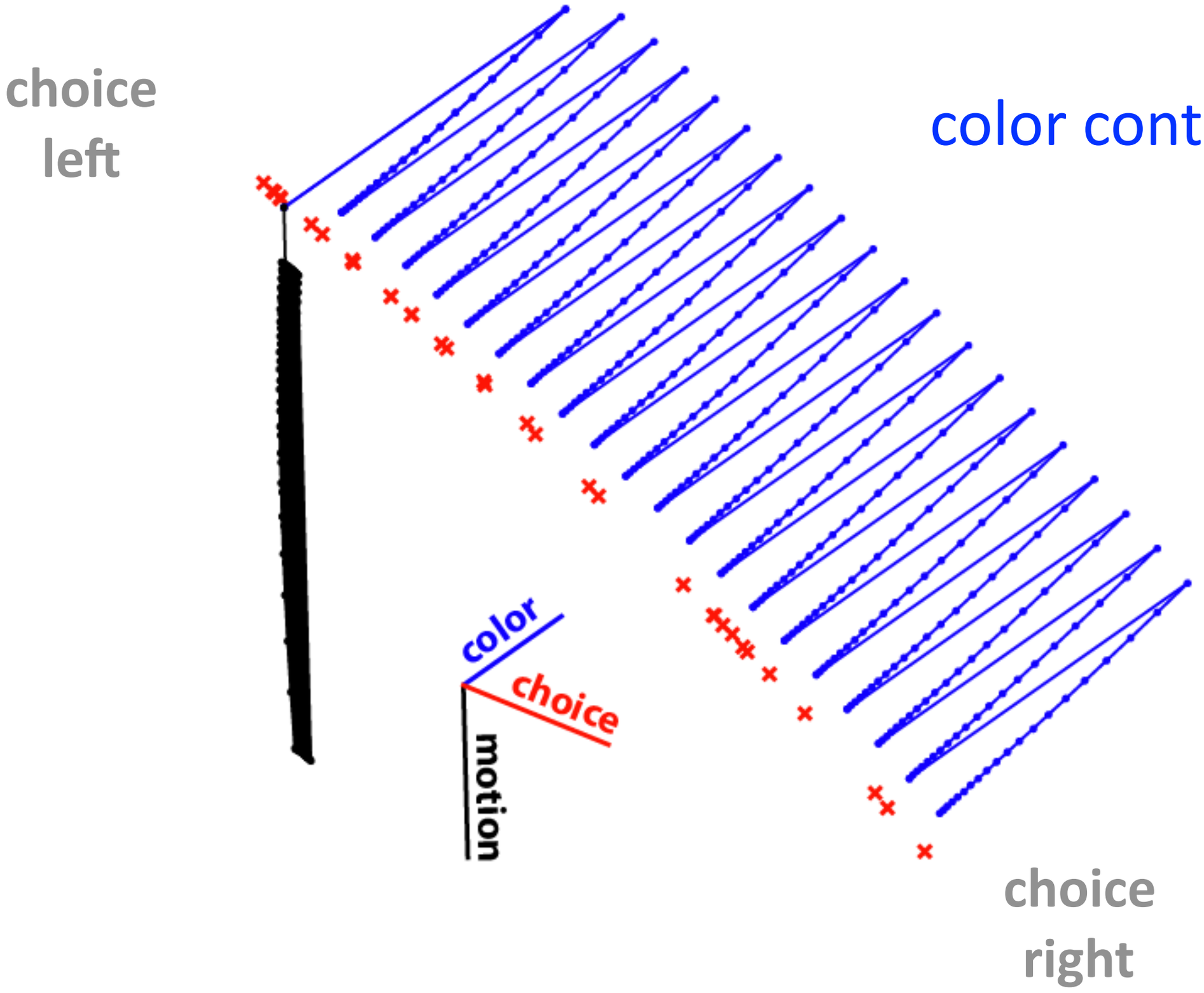
A simulated perturbation experiment



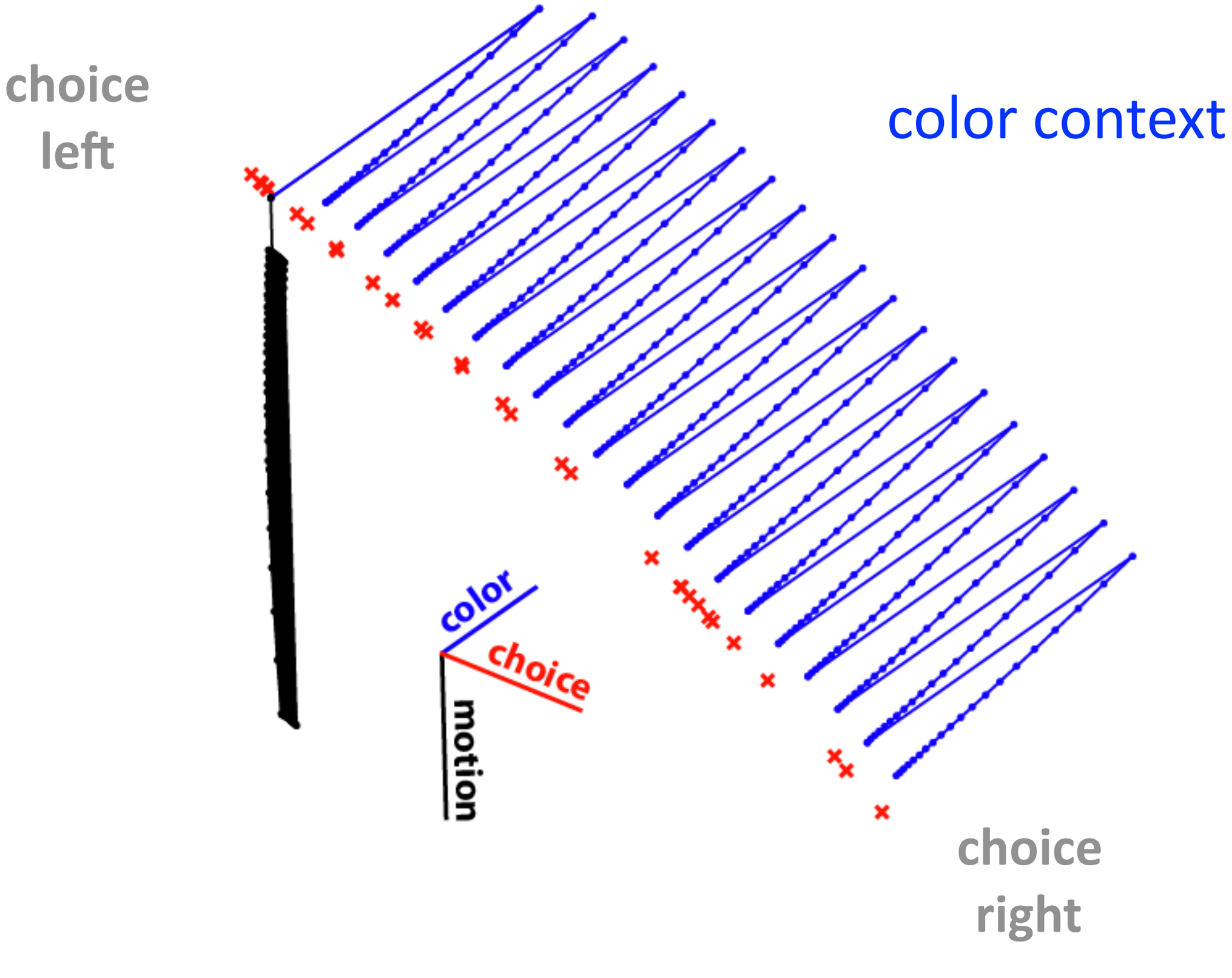
A simulated perturbation experiment



A simulated perturbation experiment

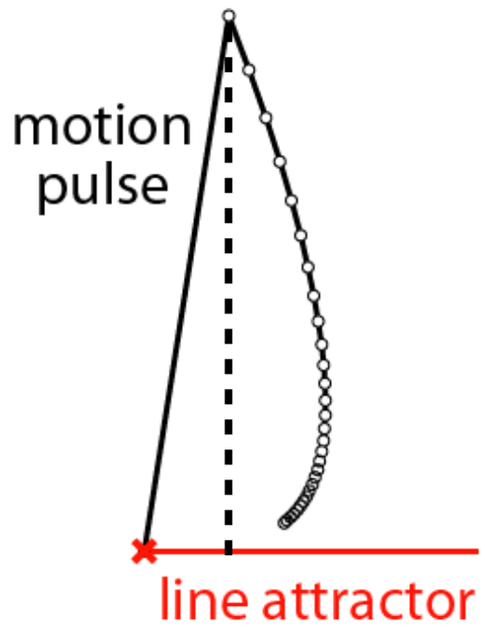


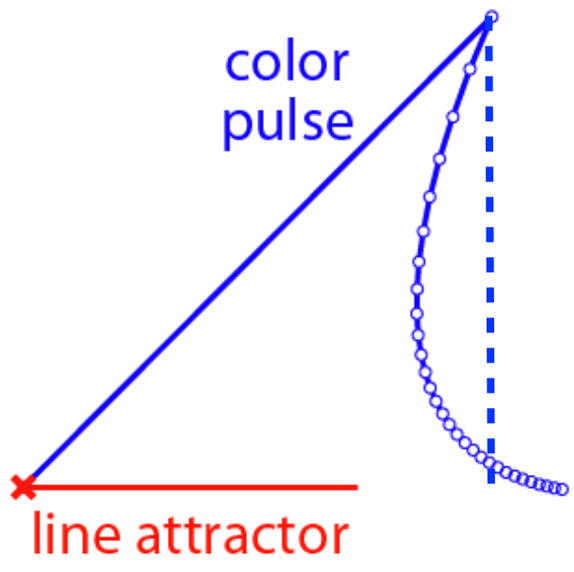
So what causes this difference between integration of color and ignoring motion?



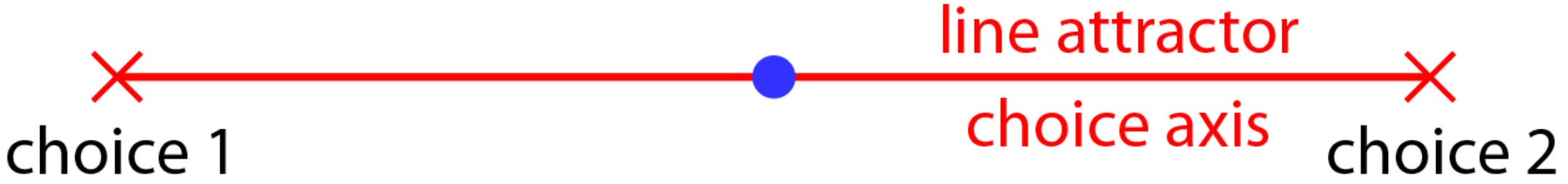
Projections onto the line attractor

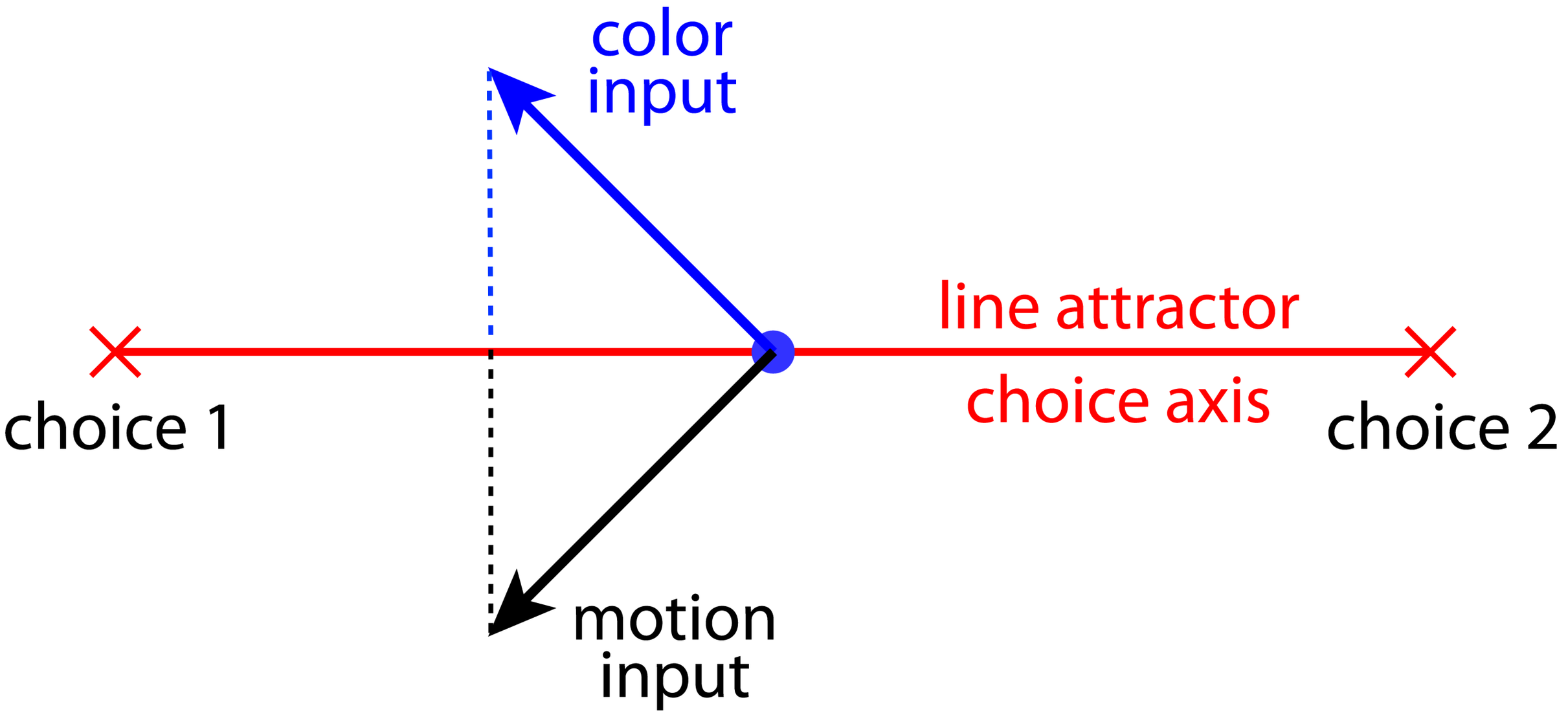
color trials



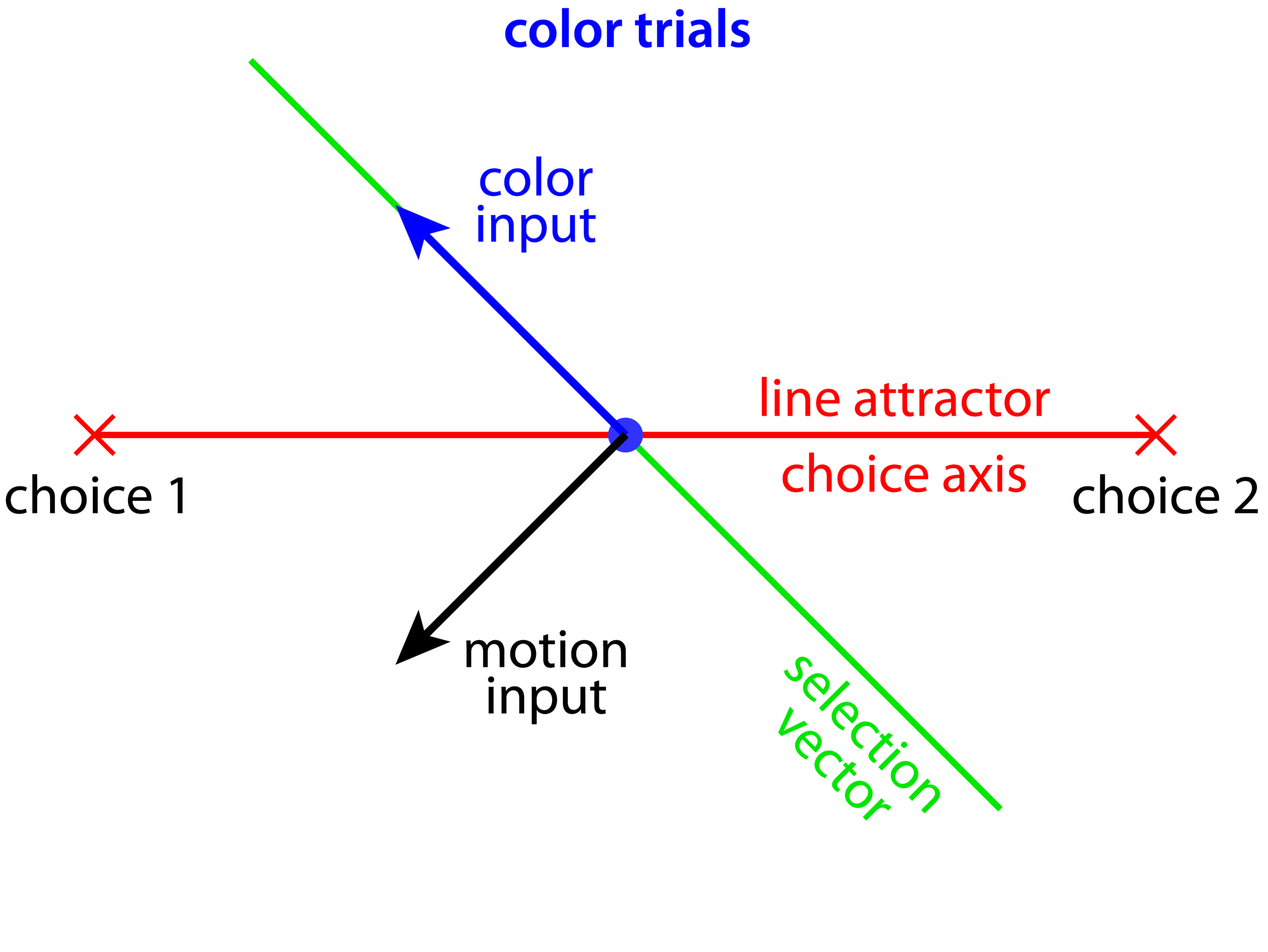


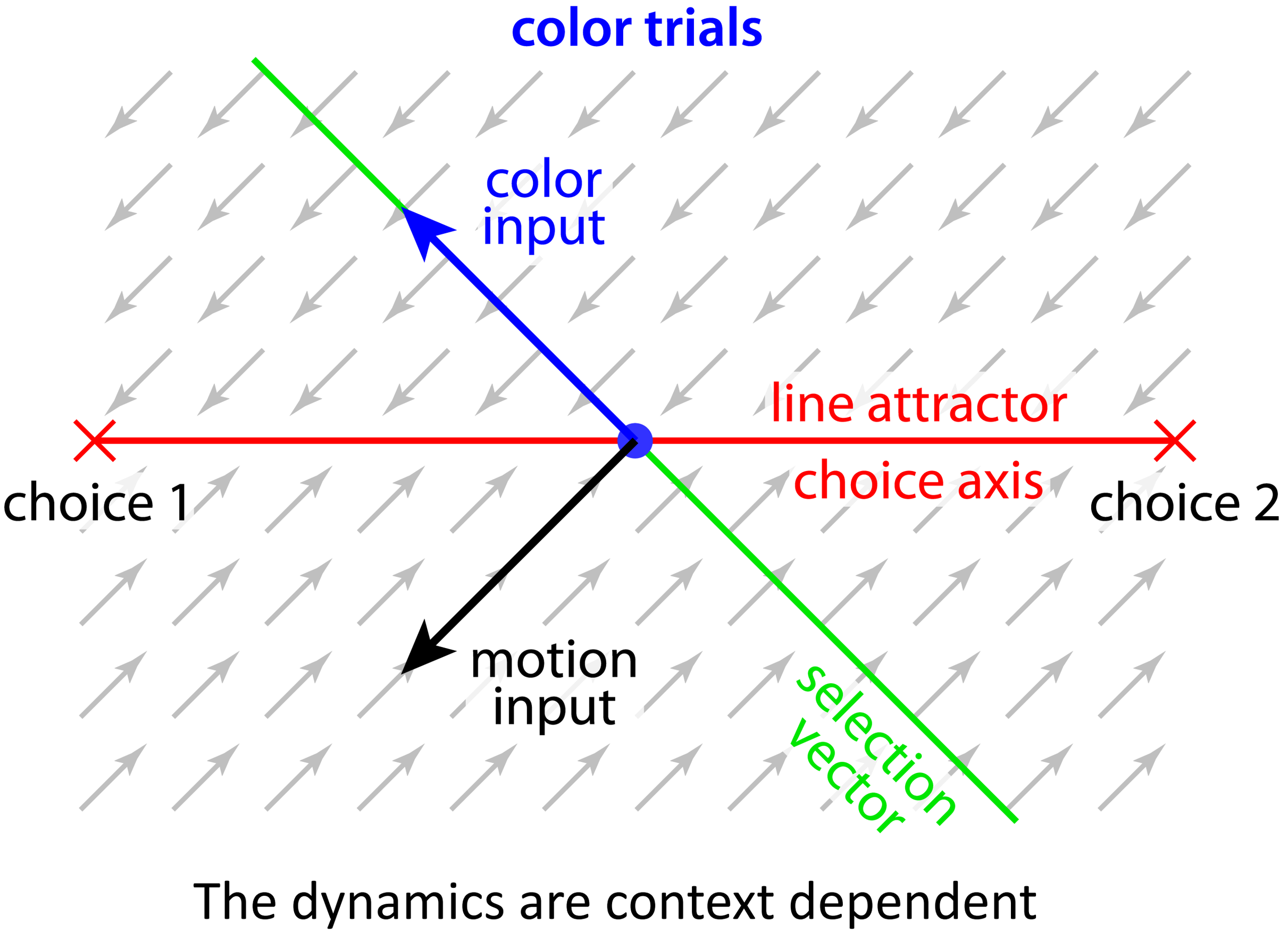
color trials



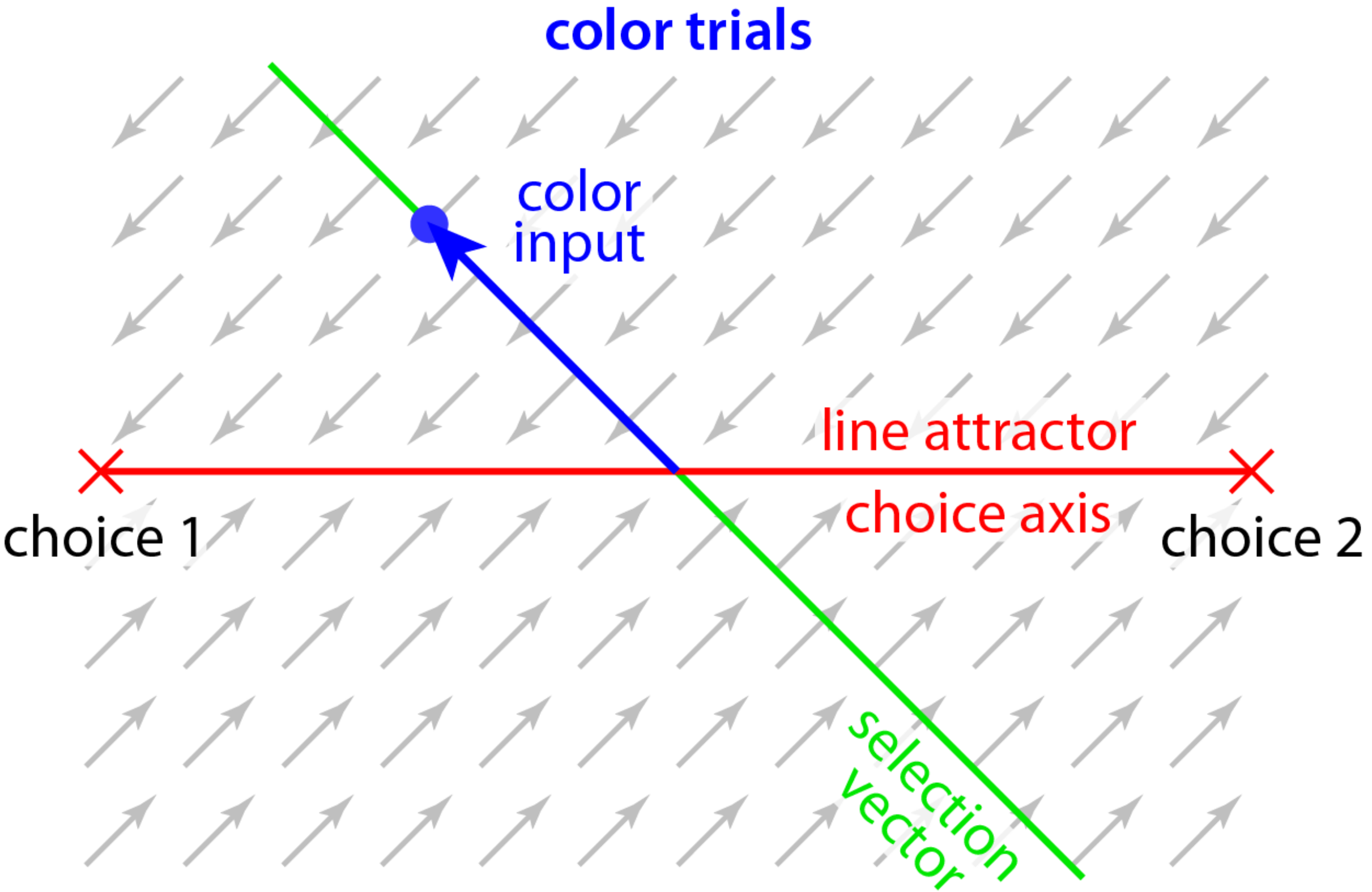


These vectors we've talked about are context **independent**

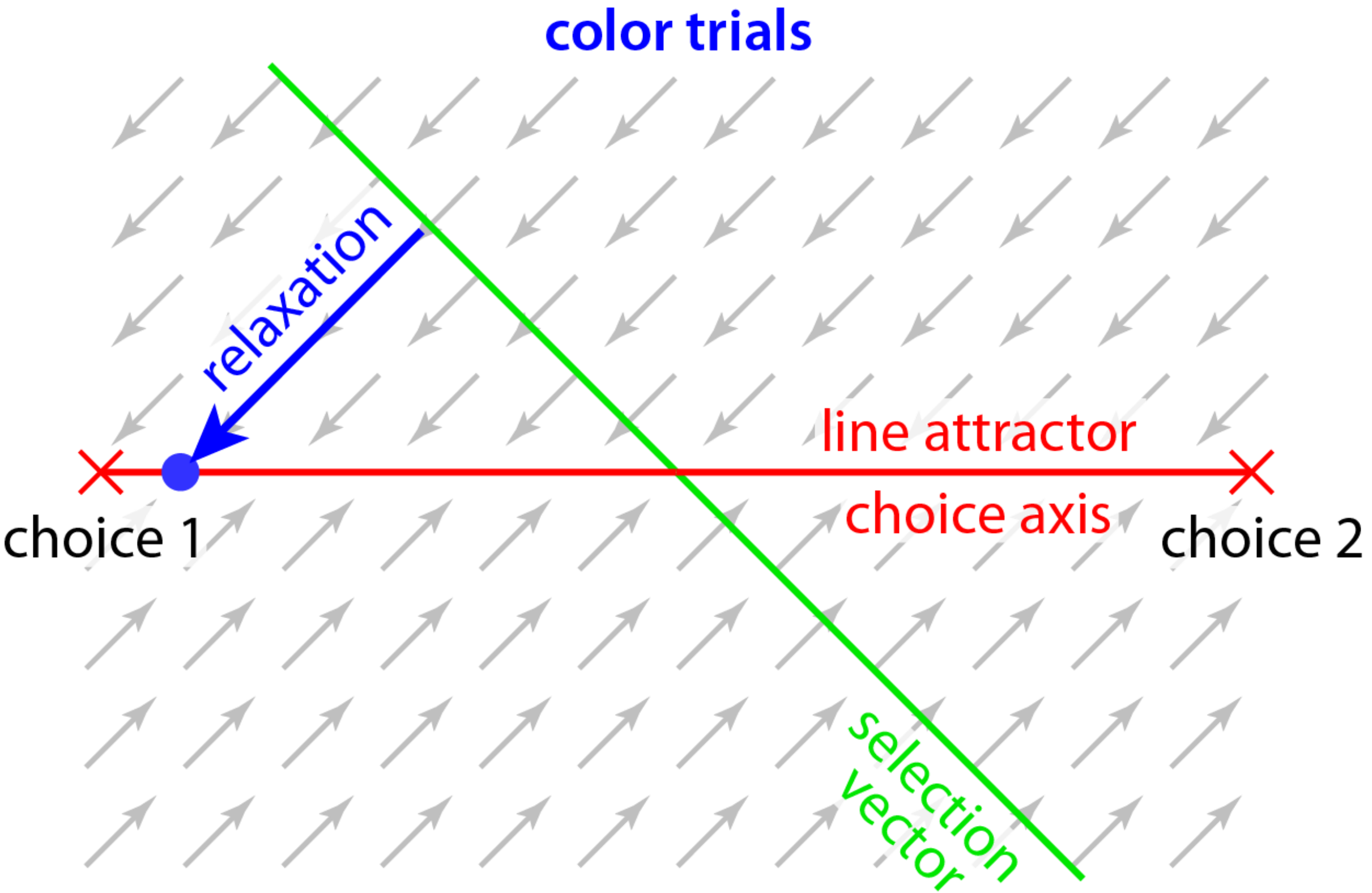

color trials
color
input
line attractor
choice axis
choice 1
choice 2
motion
input
selection
vector



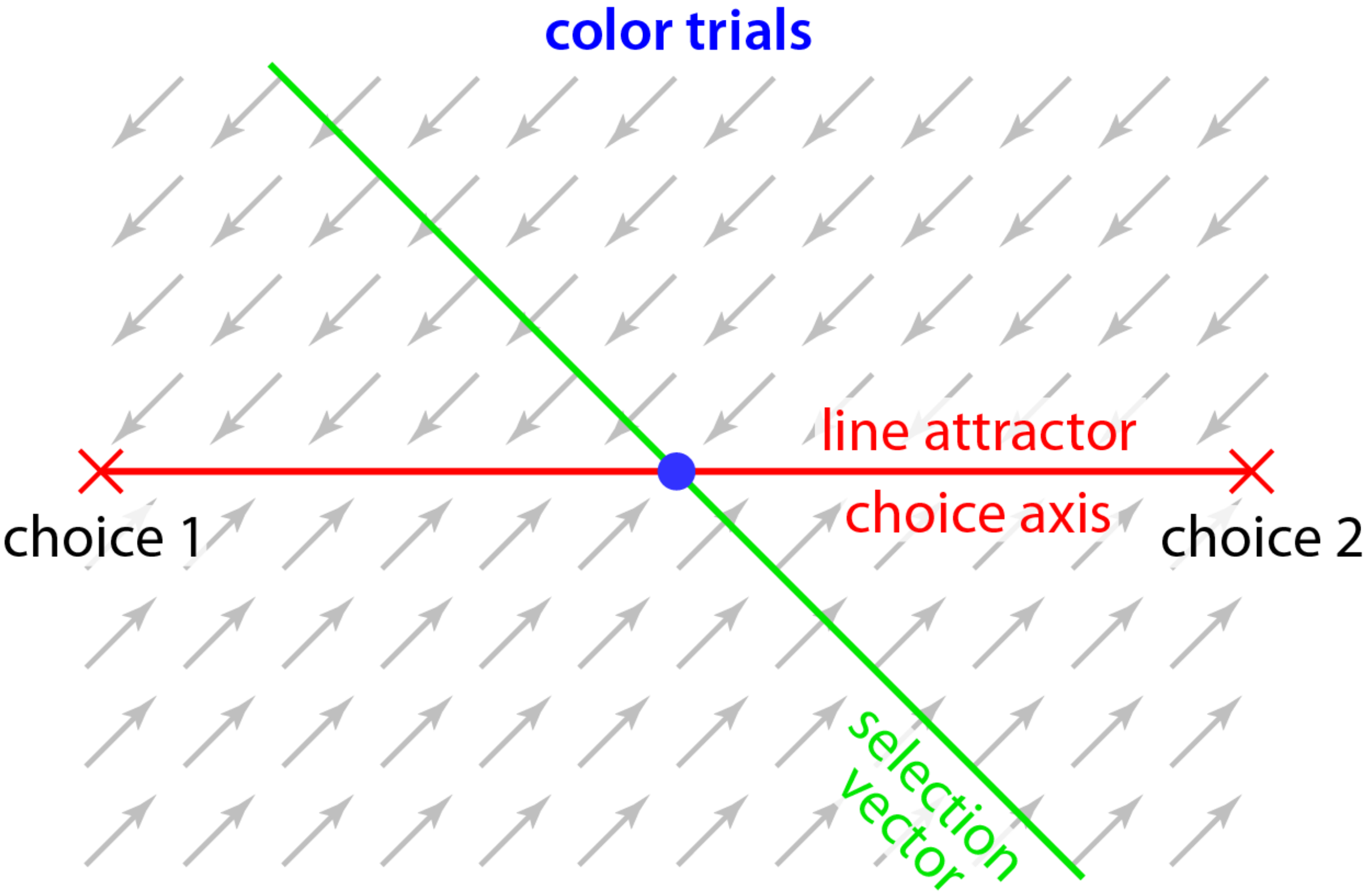
The dynamics are context dependent



The dynamics are context dependent



The dynamics are context dependent



The dynamics are context dependent

color trials

line attractor

choice axis

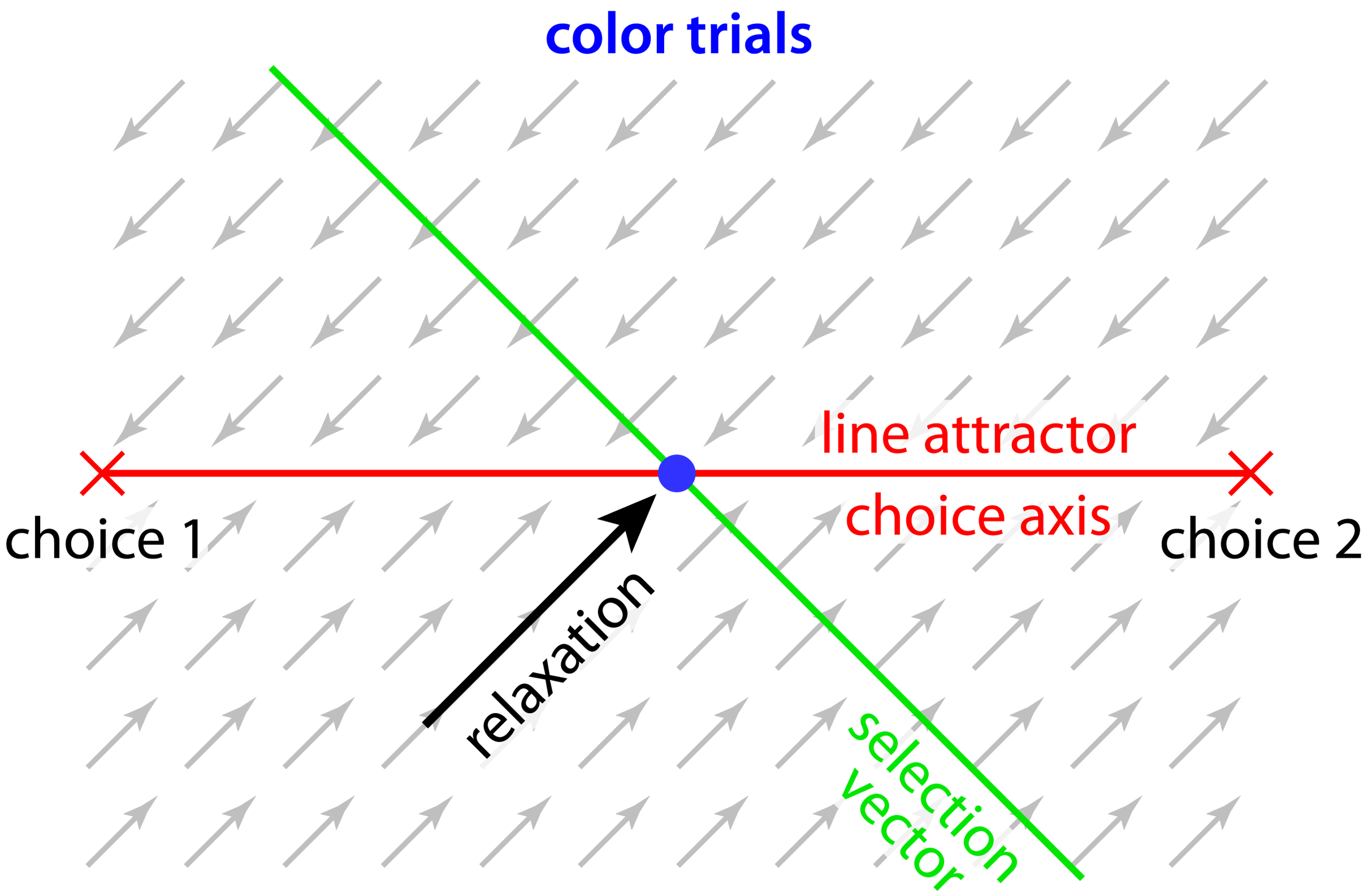
choice 1

choice 2

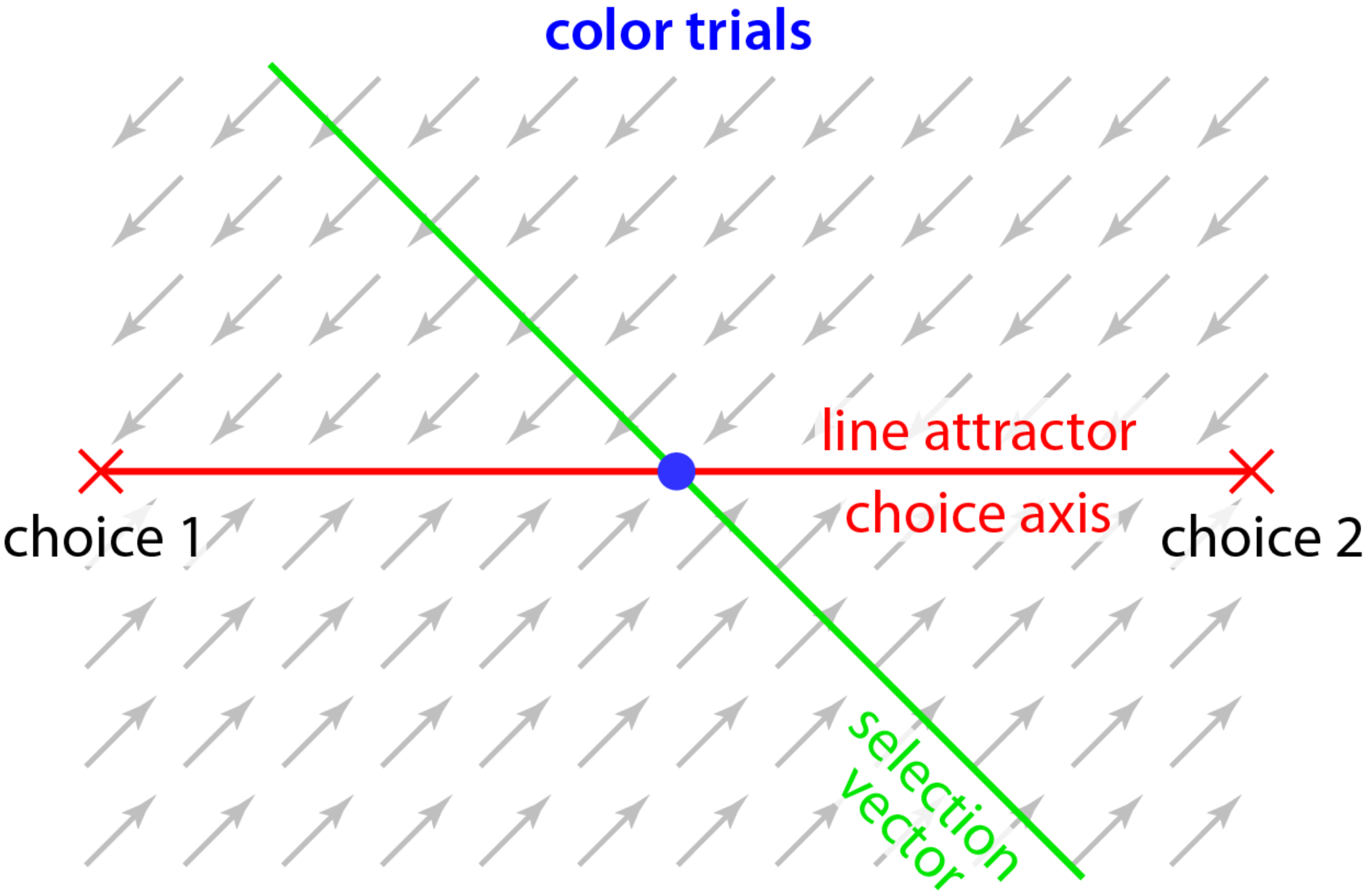
motion input

selection vector

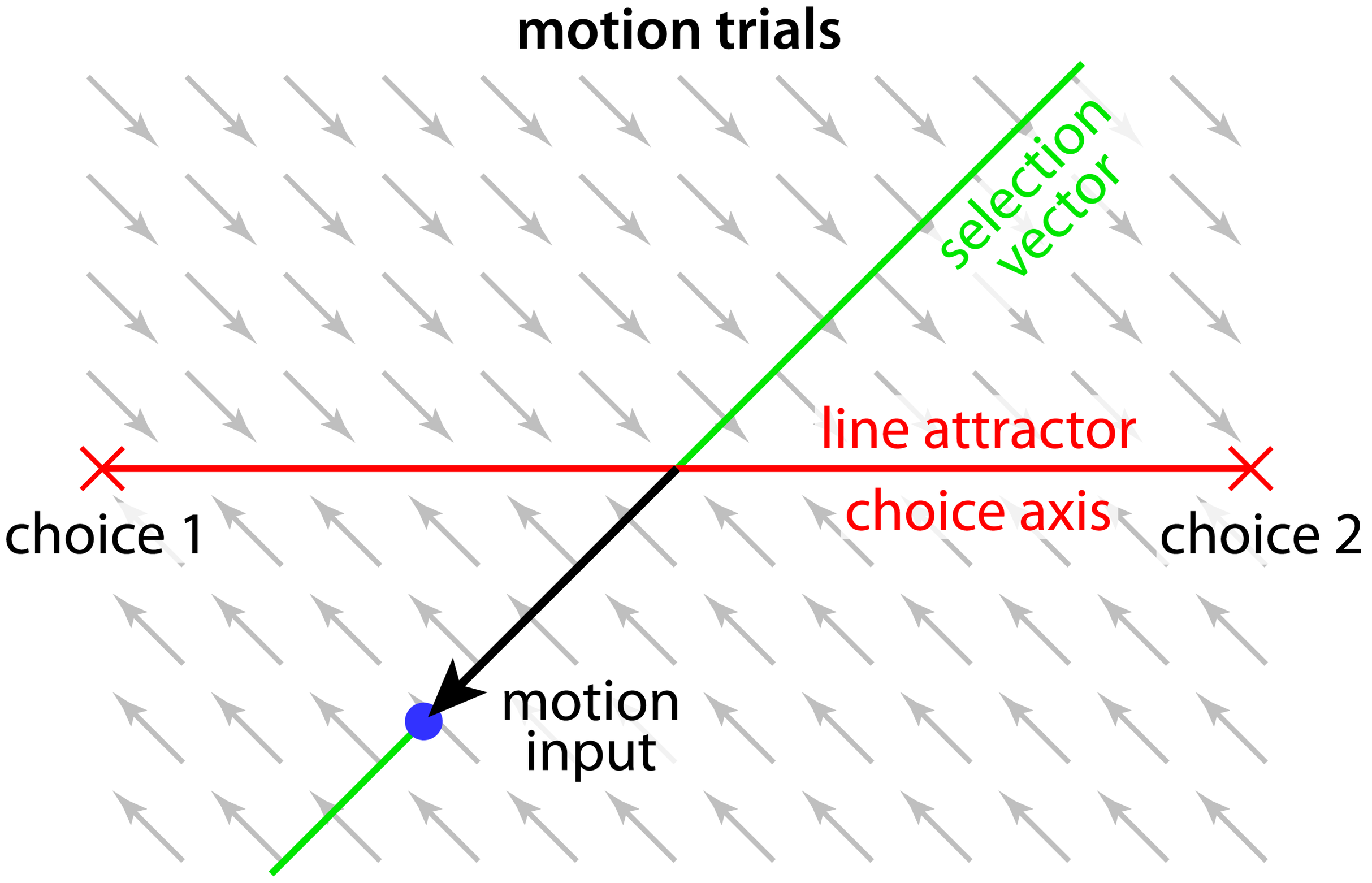
The dynamics are context dependent



The dynamics are context dependent

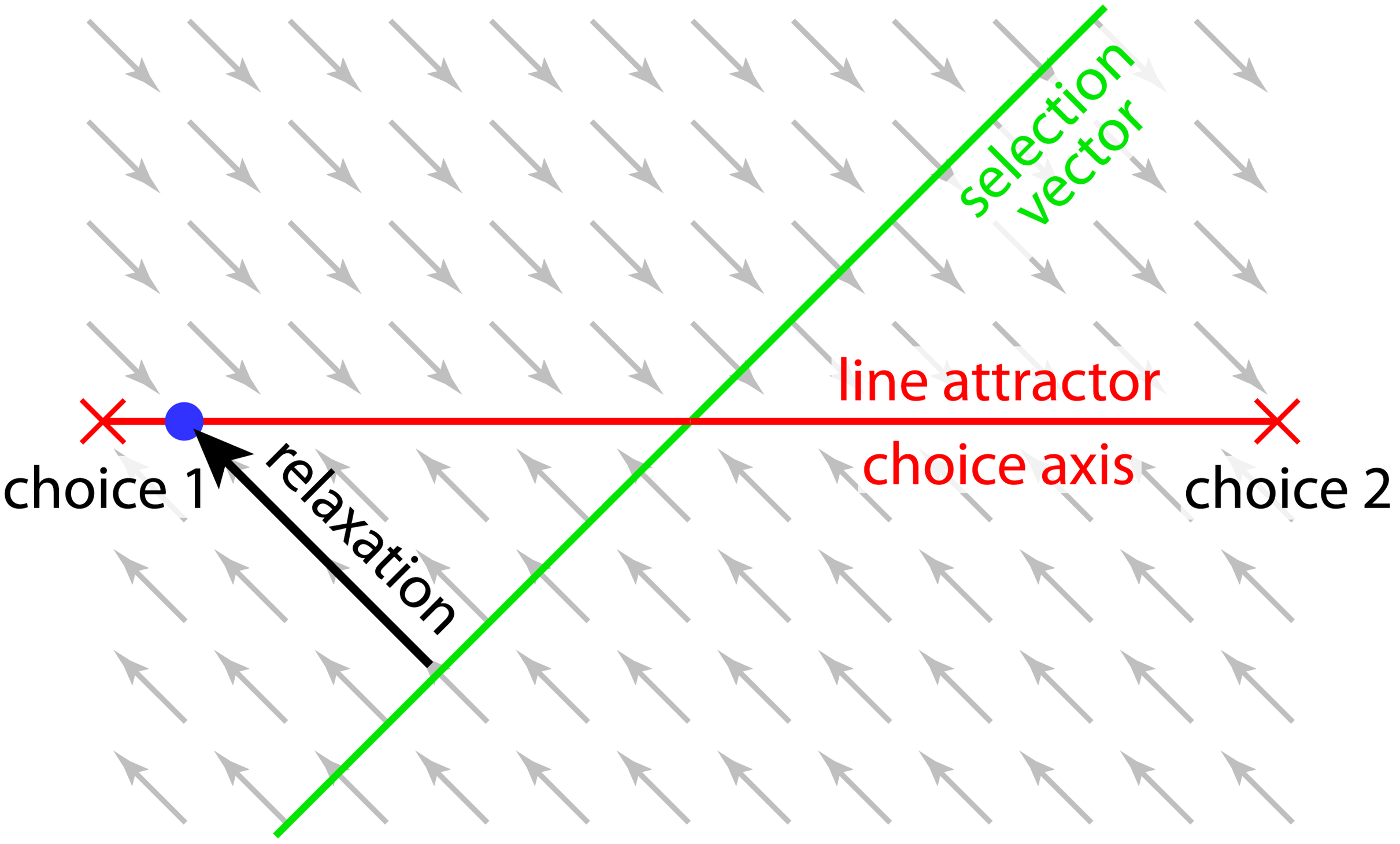


The dynamics are context dependent

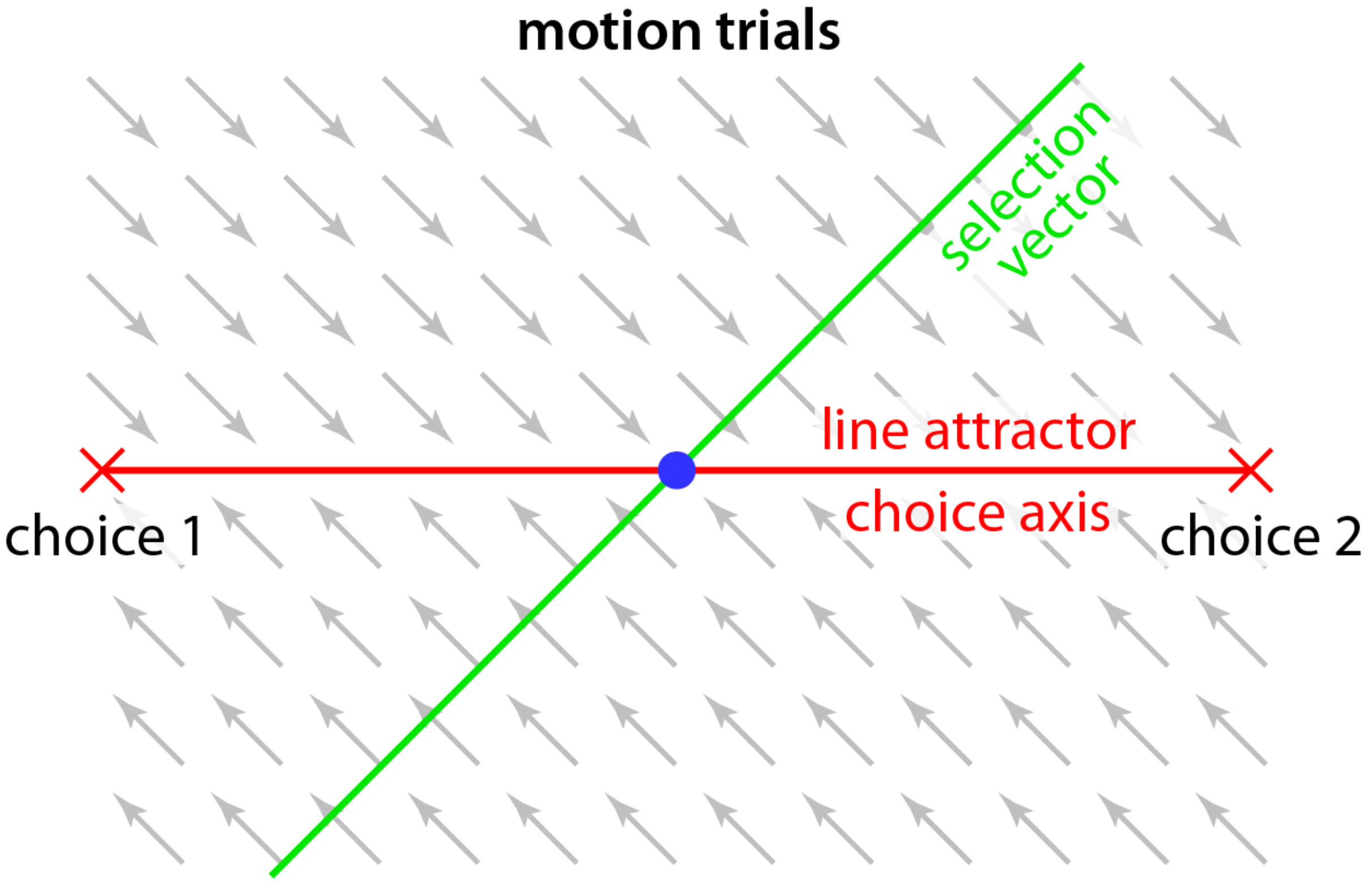


The dynamics are context dependent

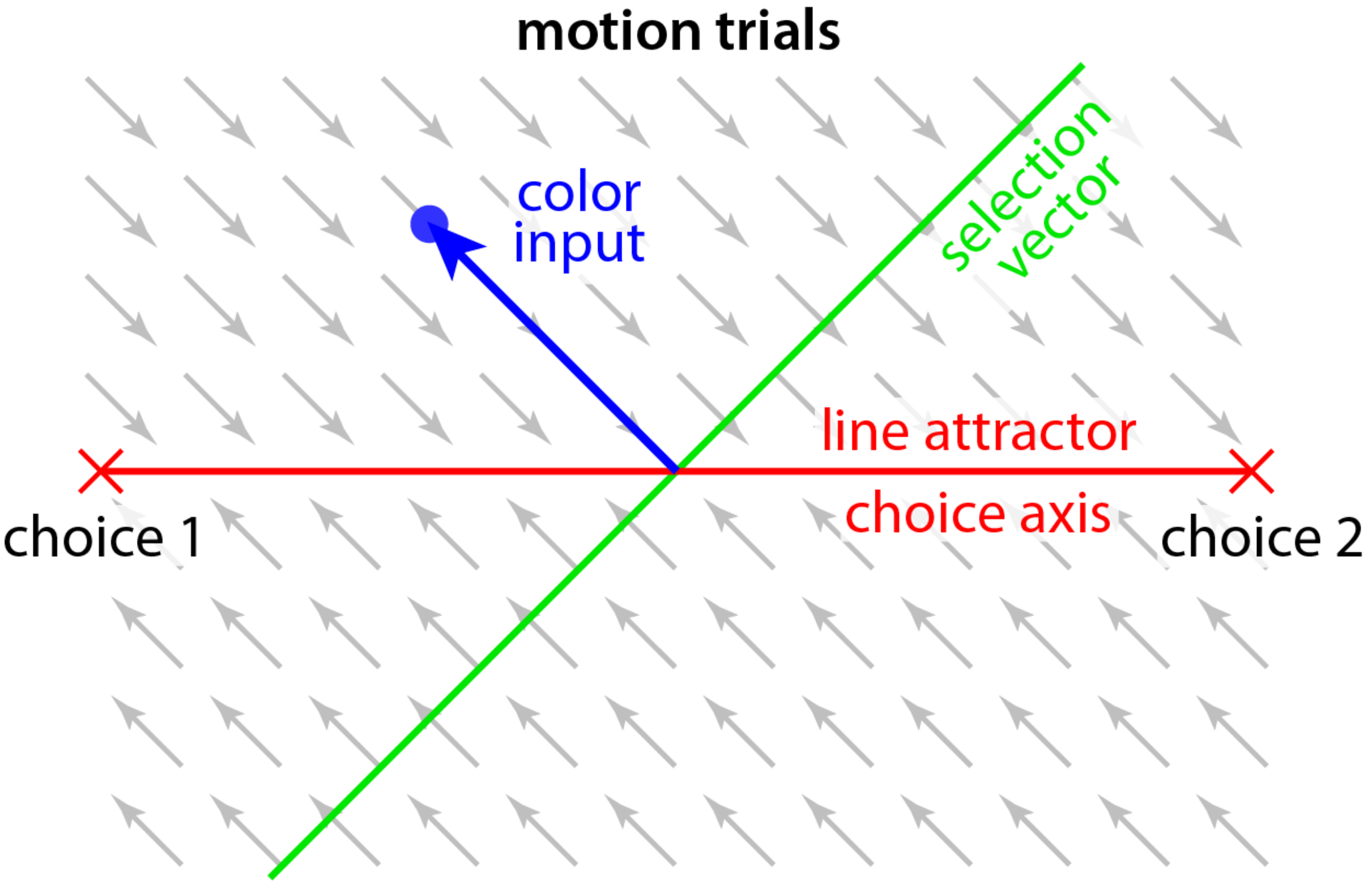
motion trials



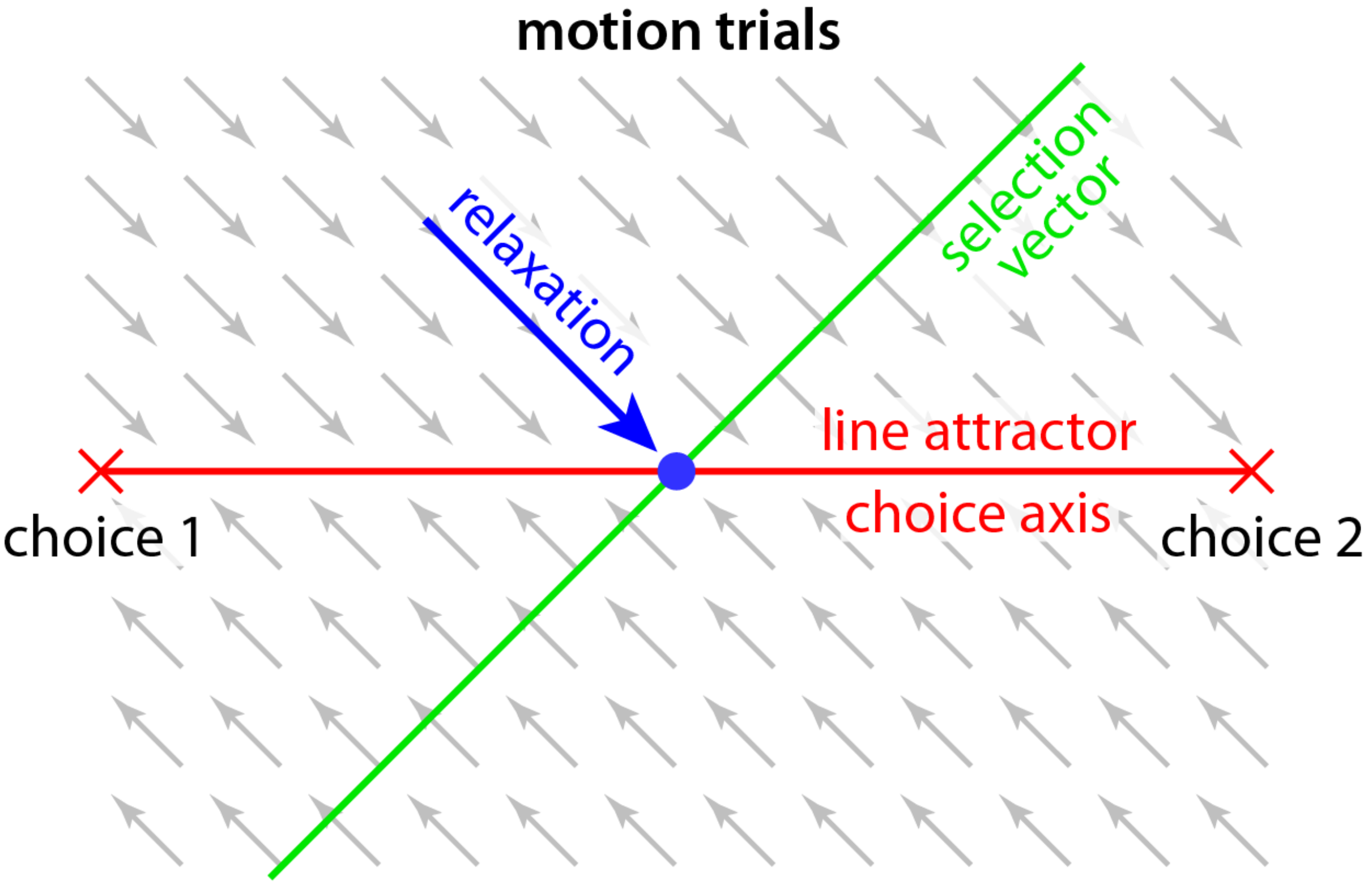
The dynamics are context dependent



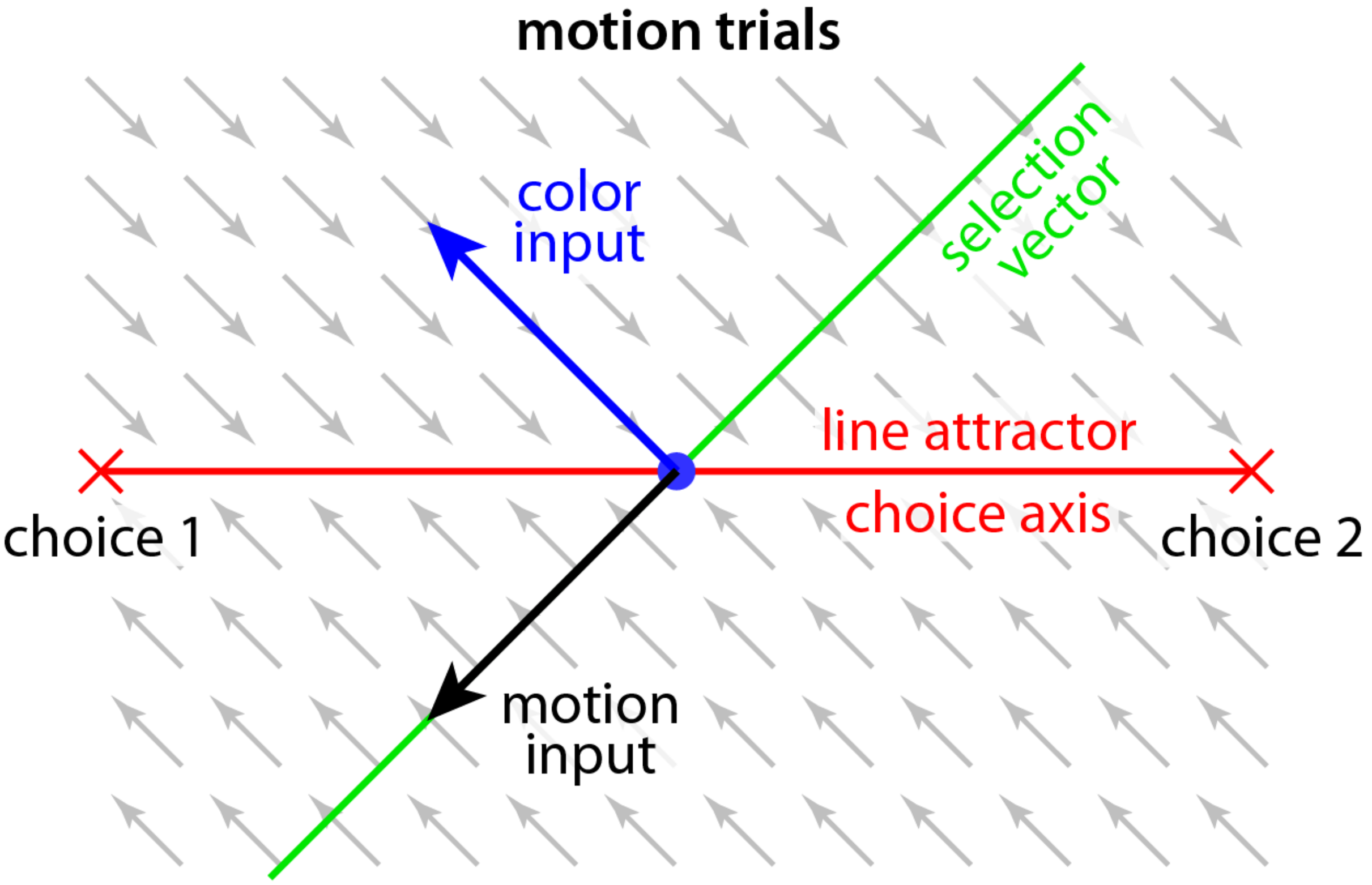
The dynamics are context dependent



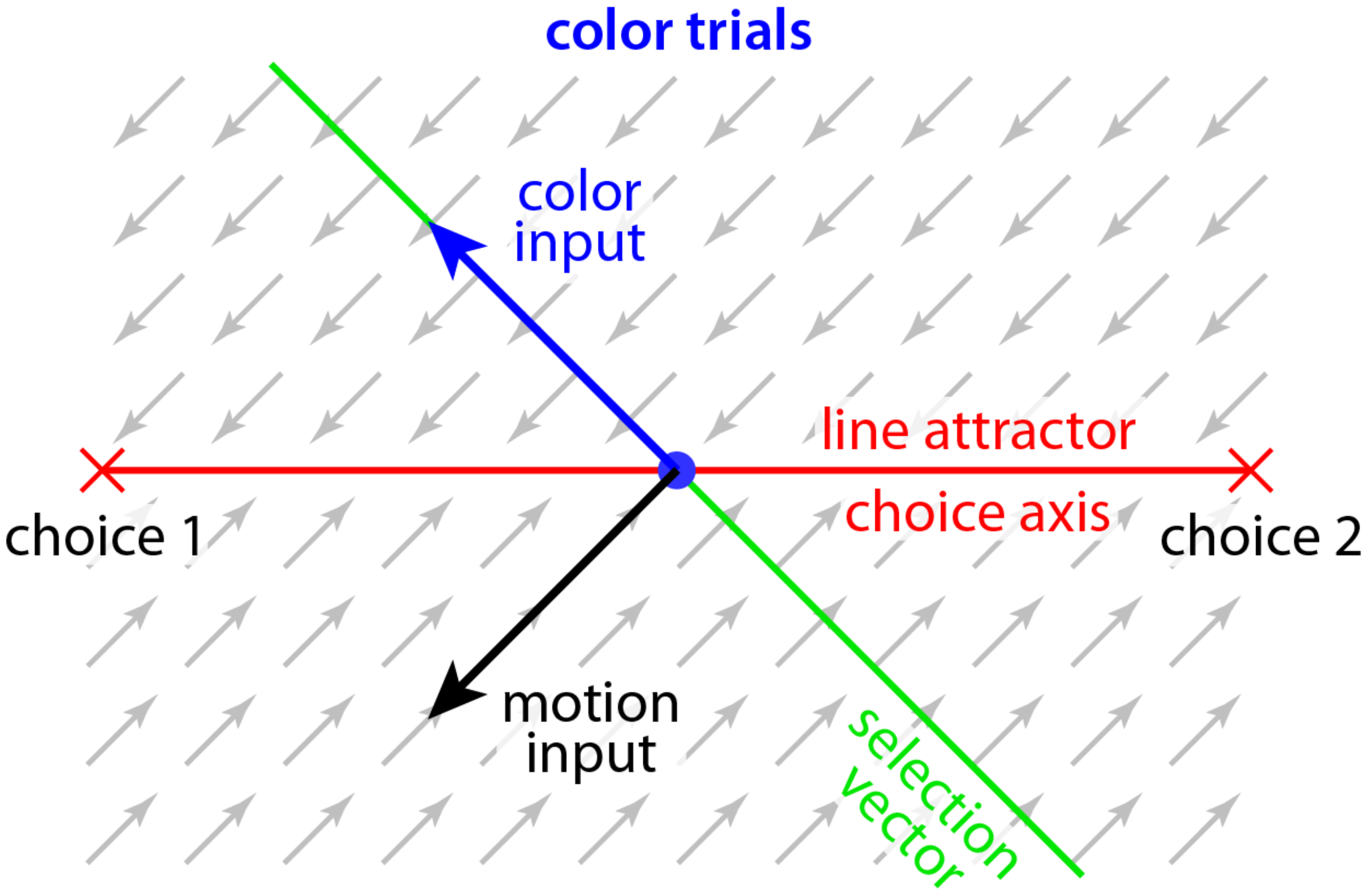
The dynamics are context dependent



The dynamics are context dependent

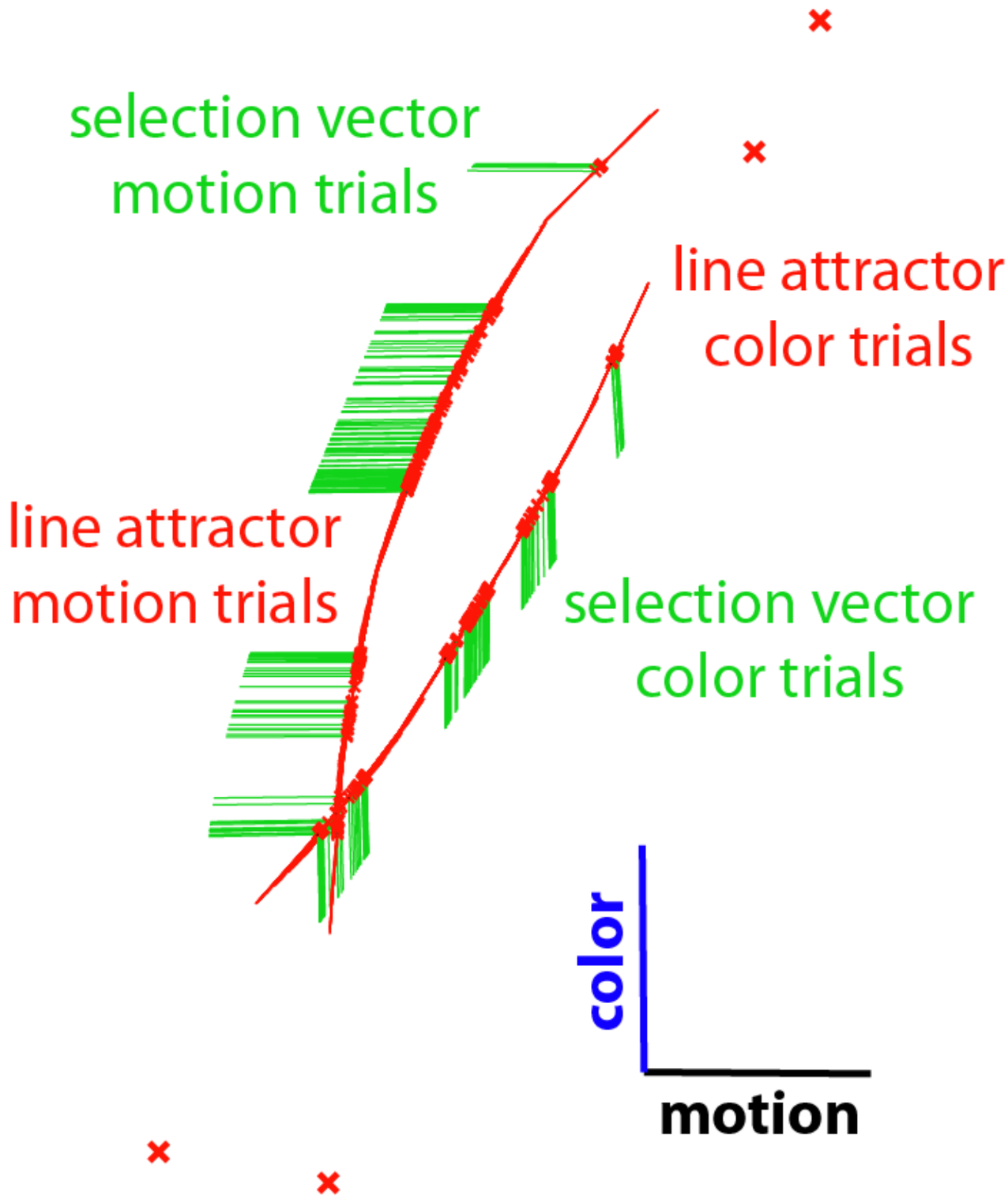


The dynamics are context dependent



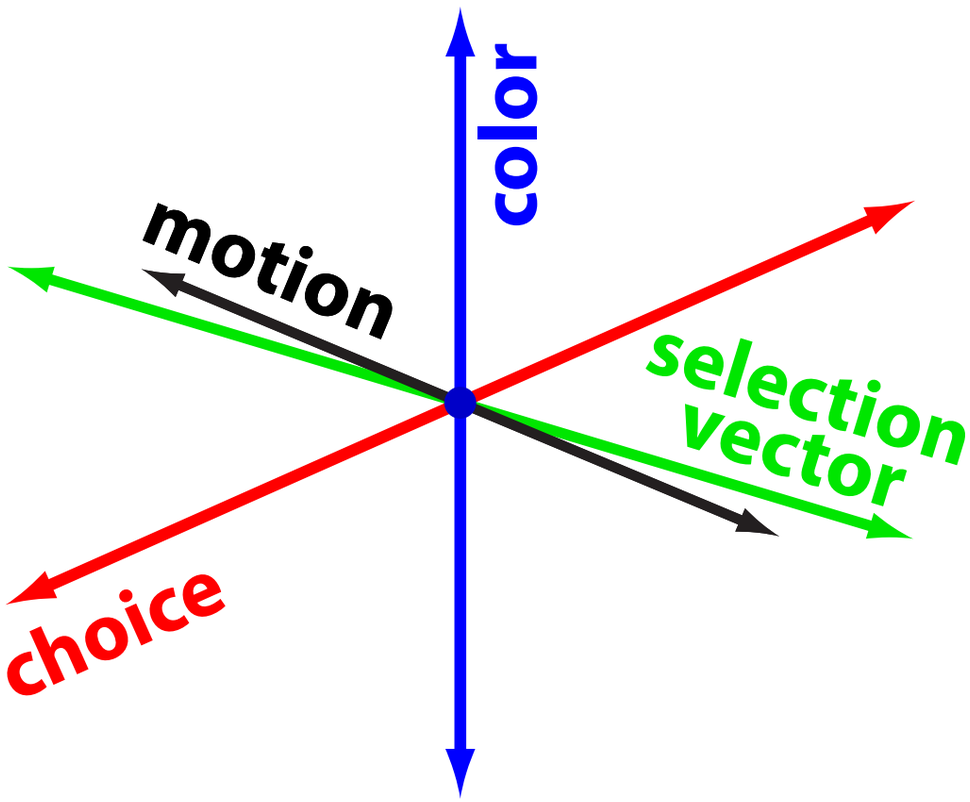
The dynamics are context dependent

Flexible selection and integration



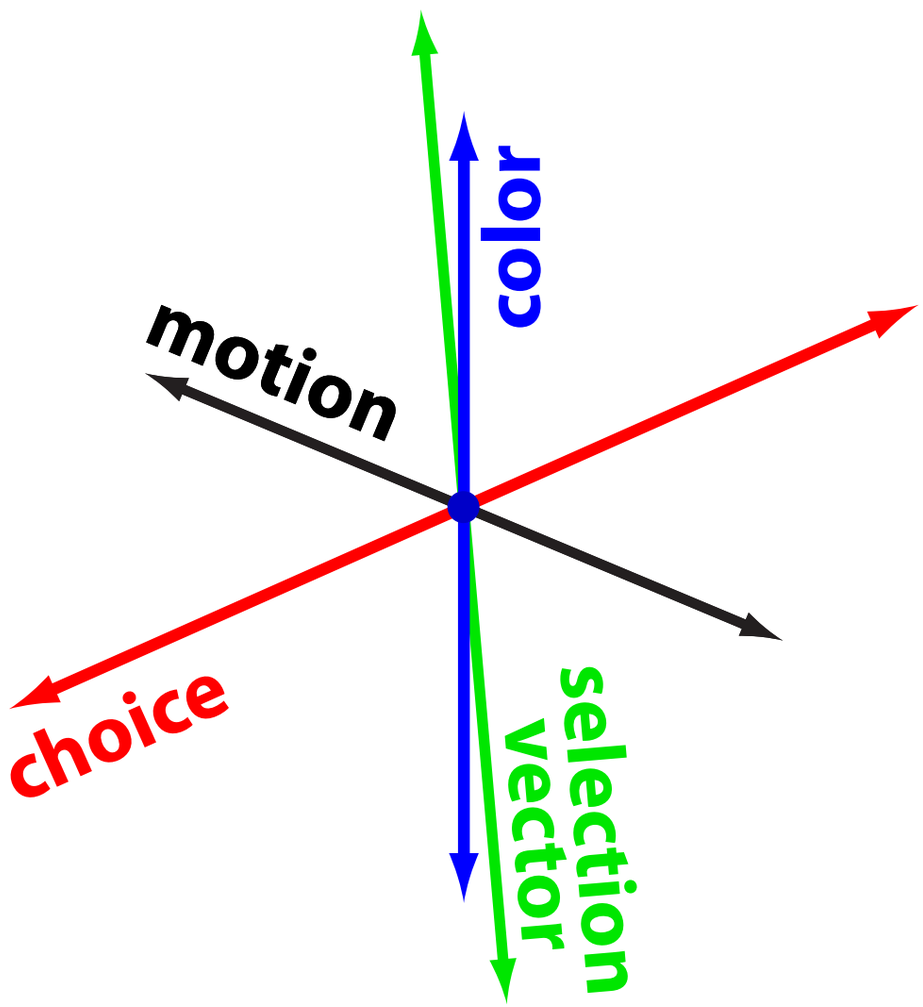
How does selective integration occur?

Motion context

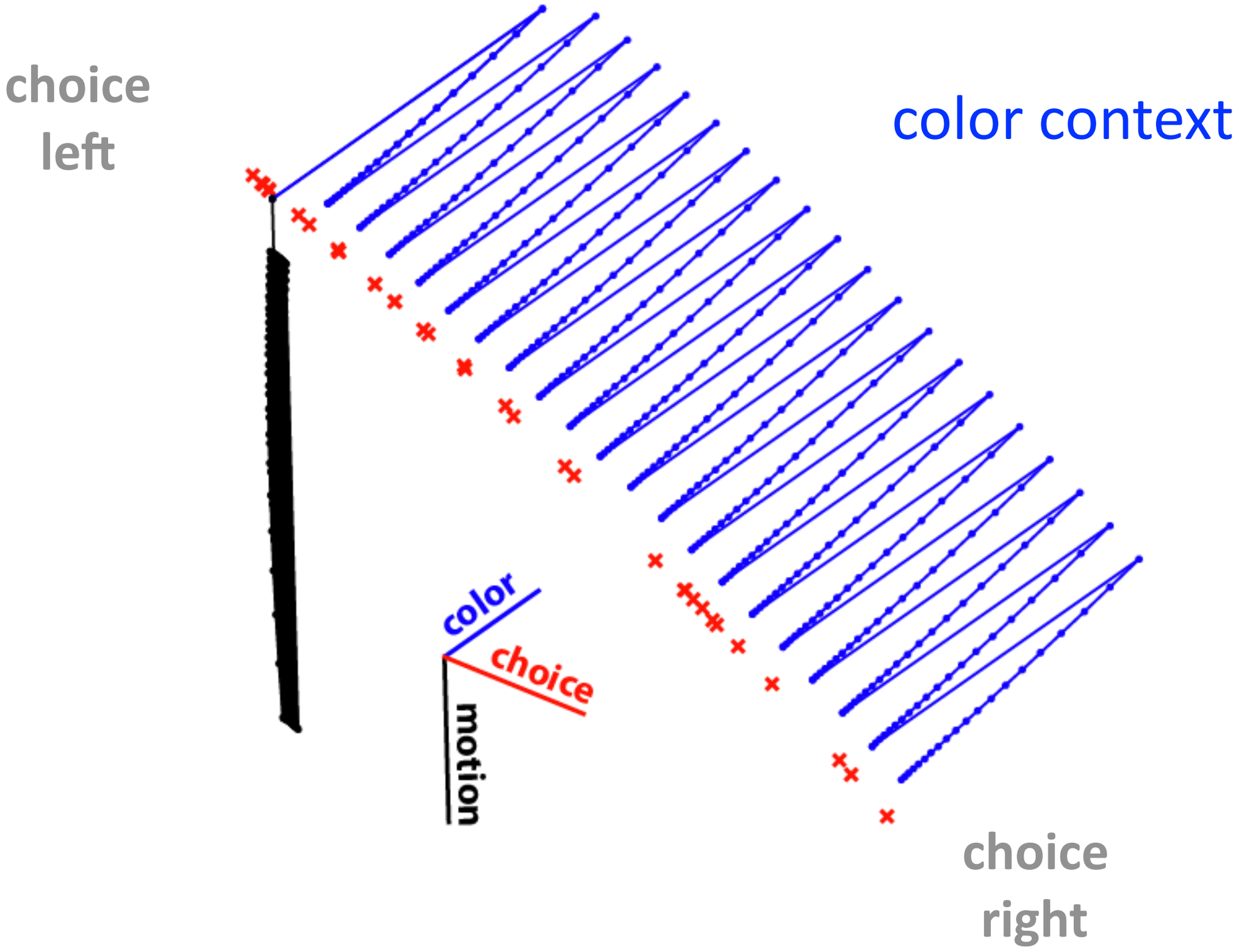


Context-dependent
selection vector

Color context



A prediction of the model



Conclusions from model

- We trained an abstract model to make a contextual decision based on two noisy input streams.
- The model made a contextual integrator with bounds.
- Like the data, the model represents the relevant and irrelevant inputs in separable dimensions.
- Two context dependent line attractors are responsible for the integration.
- Network dynamics generated through feedback, not input gating, are responsible for context dependent integration.
- The network is flexibly reconfigured by the context input, which is seen as two different line attractors in state space.

Final conclusions

Gating of sensory signals:

- Does not require modulation of sensory responses.
- Is not about suppressing the irrelevant input, but about selecting the relevant input in state space.
- Is one aspect of a dynamical process occurring in the same cortical circuit as integration of evidence.
- Everything is happening at the population level.
- Our works suggests a *possible* mechanism, which is *not* exclusive of others.

Computation through dynamics: mixed, separable representations are contextually and dynamically linked to generate the desired output.

Acknowledgments

Krishna Shenoy
Mark Churchland
Matt Kaufman
Cindy Chestek
Dan O'Shea
Cora Ames
Werapong Goo
Justin Foster
Paul Nuyujukian
Jonathan Kao
Joline Fan
Eric Trautmann
Sergey Stavisky
Chand Chandrasekaran

Bill Newsome
Valerio Mante
Roozbeh Kiani
Vince McGinty
Daniel Kimmel
Leo Segrue
Diogo Peixioto
Nazli Emadi

Larry Abbott
Mark Churchland
Omri Barak
Valerio Mante

Funding

This research was supported by an NIH Director's Pioneer award, Howard Hughes Medical Institute, and grants from NIH-CRCNS, DARPA REPAIR, NSF, the Helen Hay Whitney Foundation, the Burroughs Wellcome Fund, and the Christopher and Dana Reeve Foundation and a NSF GRFP.