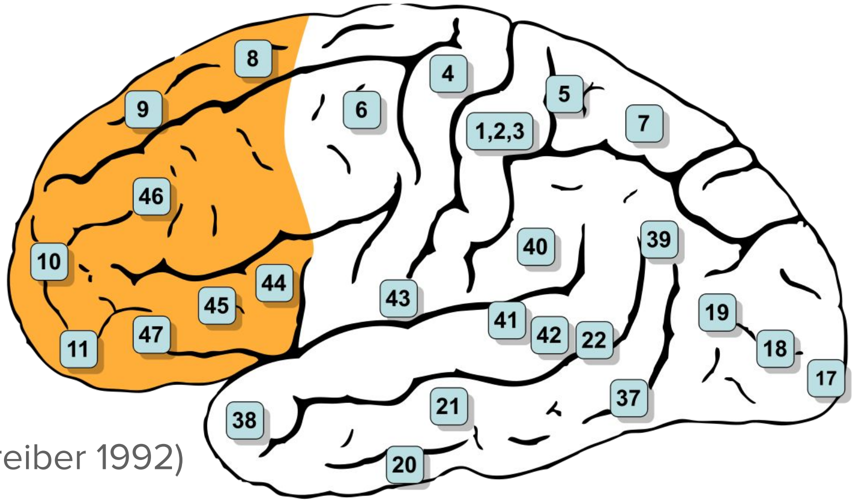
Context-dependent computation by recurrent dynamics in prefrontal cortex

Mante, Sussillo, Shenoy, and Newsome (2013)

Neural systems for decision making and executive function

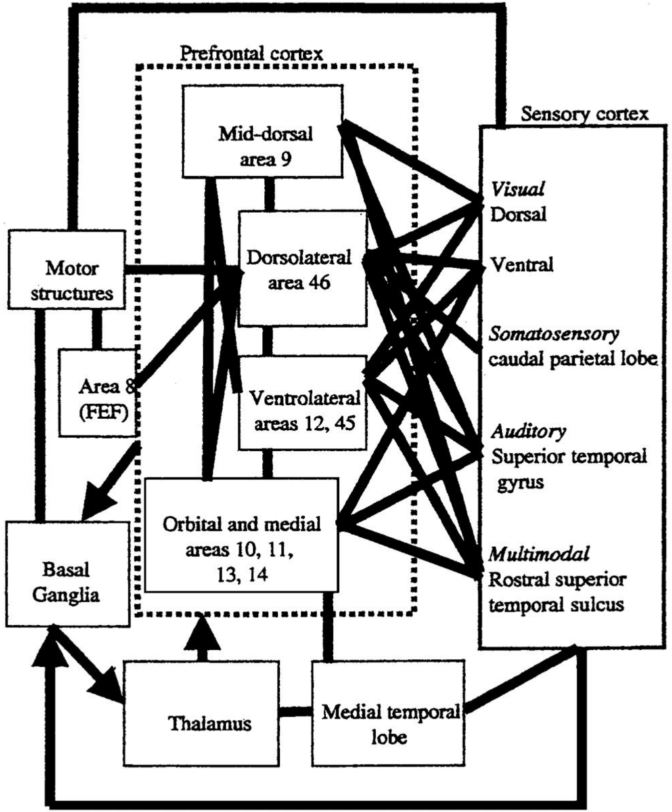
- Lesion studies have implicated **prefrontal cortex** (**PFC**) as a necessary component of the brain system underlying decision making in humans and non-human primates
- Characteristic prefrontal patient errors in the **Multiple Errands Test** (Shallice & Burgess, 1991):
  - Inefficiencies
  - Rule breaks
  - Interpretation failures
  - Task failures
- Characteristic error patterns in many tasks
  - Wisconsin Card Sorting (Milner, 1963)
  - Stroop task (Perrett 1974, Cohen & Servan-Schreiber 1992)
  - Etc.



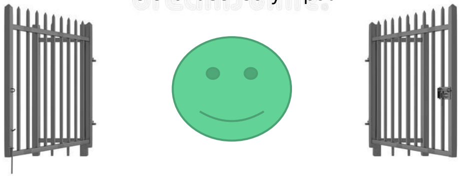
Neural systems for decision making and executive function

- Empirical neuroscience and computational modeling studies have confirmed and elaborated the role of PFC in decision making and executive function

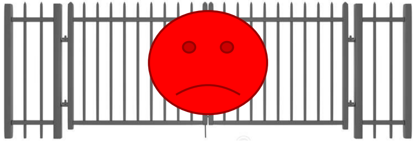
- An integrative theory of PFC (Cohen & Miller, 2001)
  - highlights widespread extrinsic and intrinsic connectivity
  - Representation of context and task goals in PFC, actively maintained through intrinsic connectivity
  - Extrinsic connectivity allows PFC to bias other cortical regions towards achieving task goals
    - PFC-mediated **gating** of irrelevant sensory inputs?



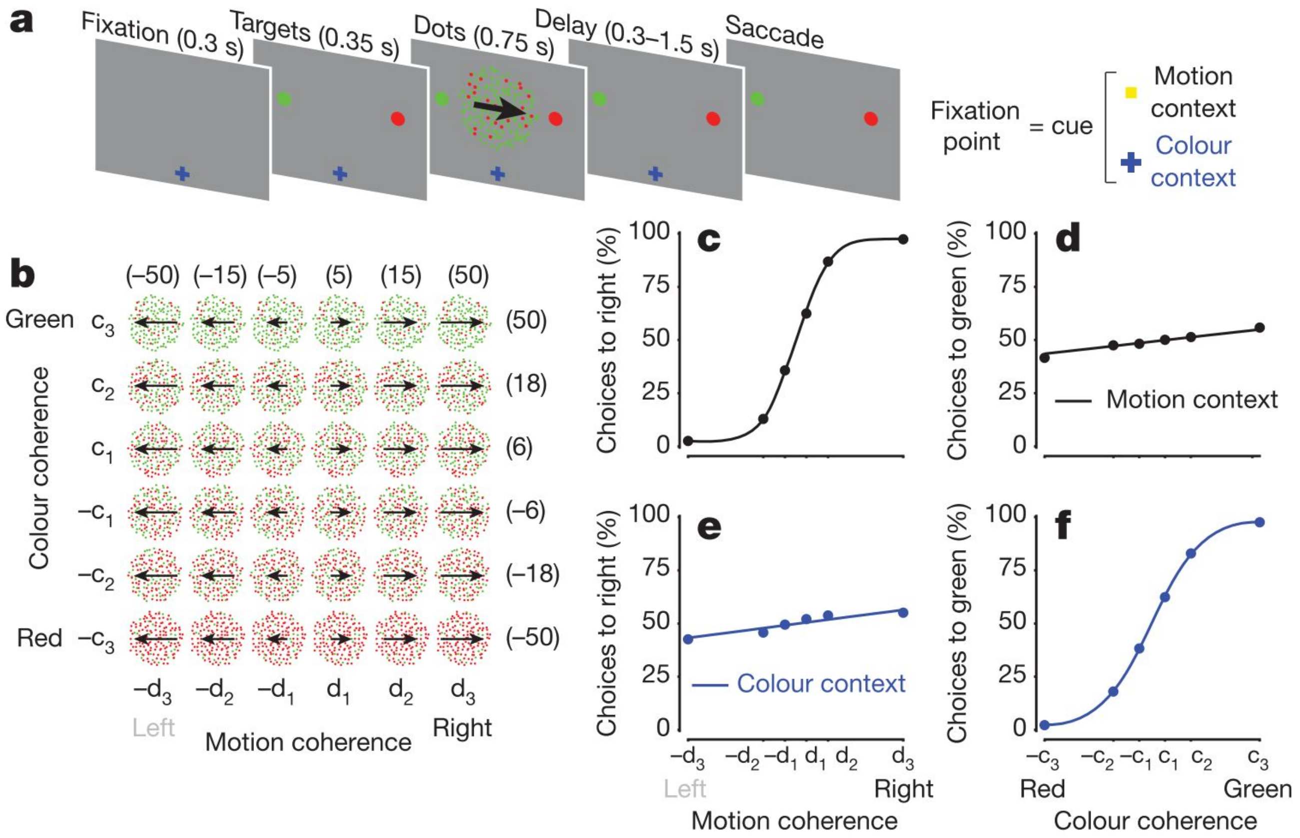
relevant sensory input



irrelevant sensory input

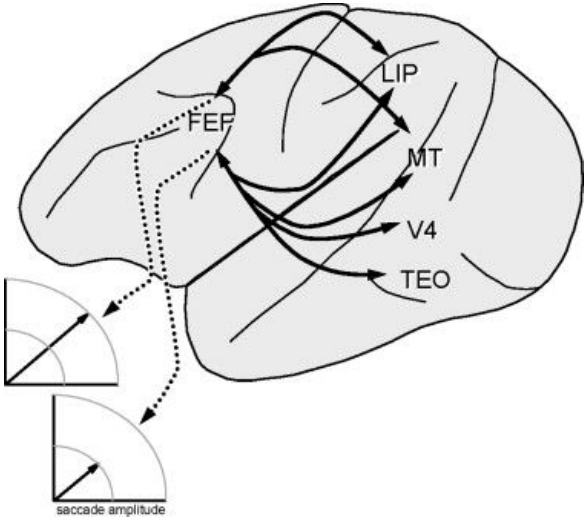


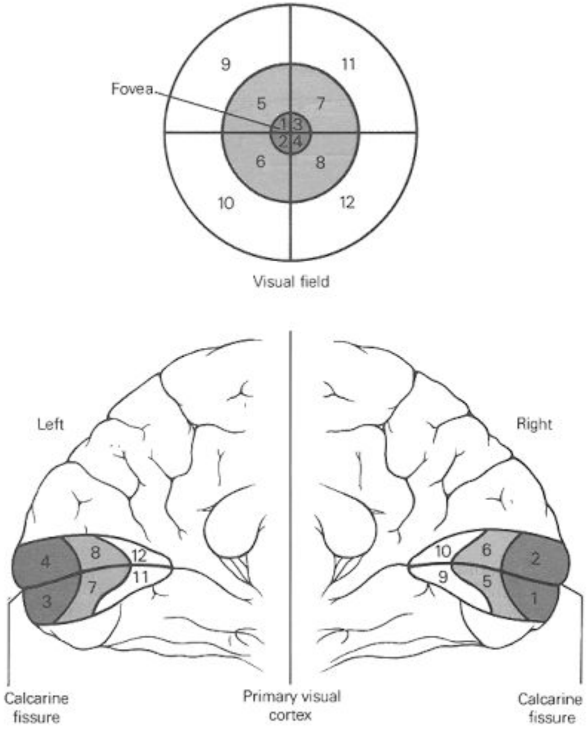
Behavioral task and psychophysical performance



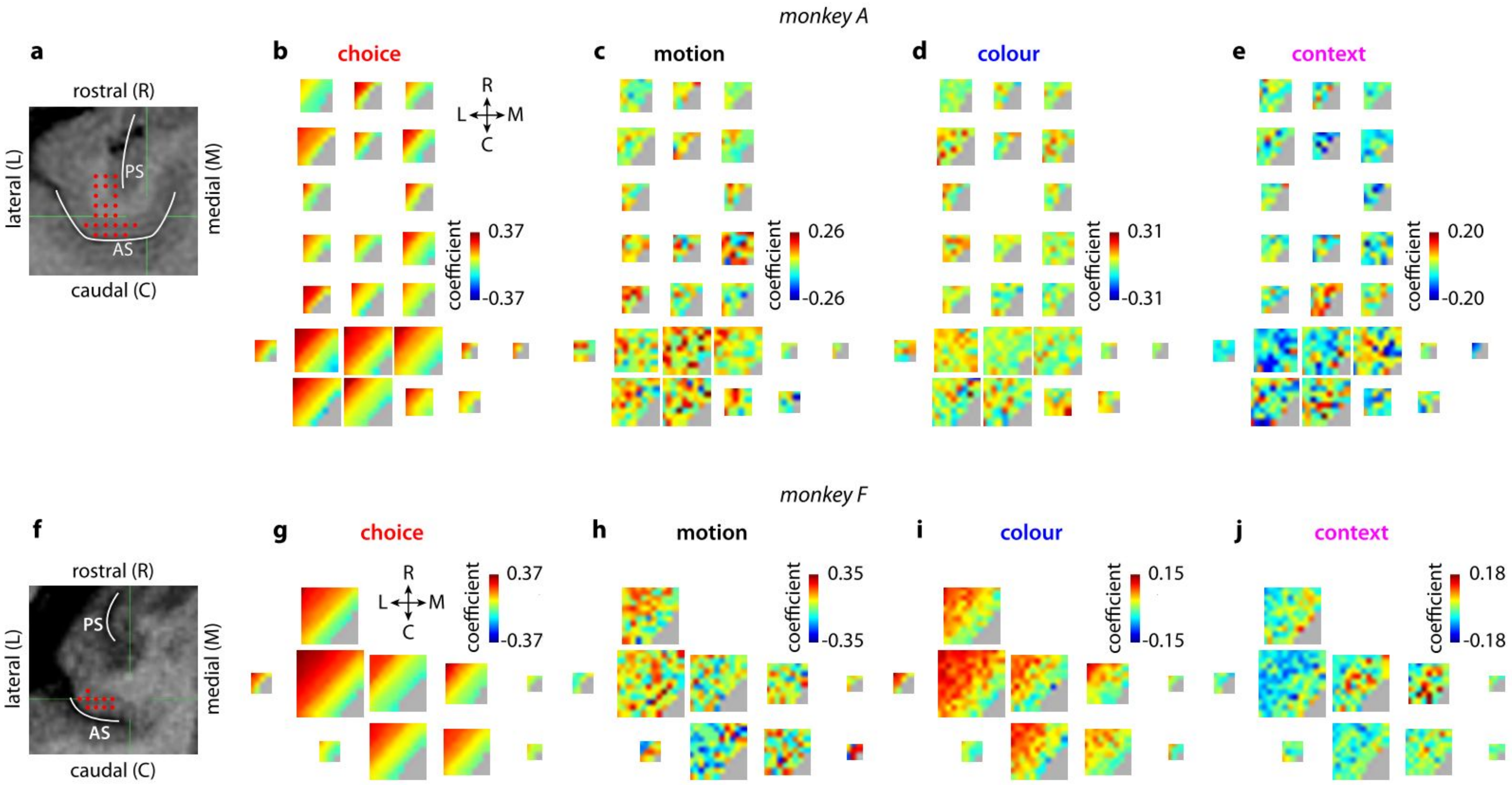
Frontal Eye Field (FEF)

- Involved in planning of **saccadic eye movements**
- Electrical stimulation elicits saccades
- Topographically organized in **retinotopic** coordinates of saccade targets





Recording from neural populations in FEF



Analyzing population-level responses

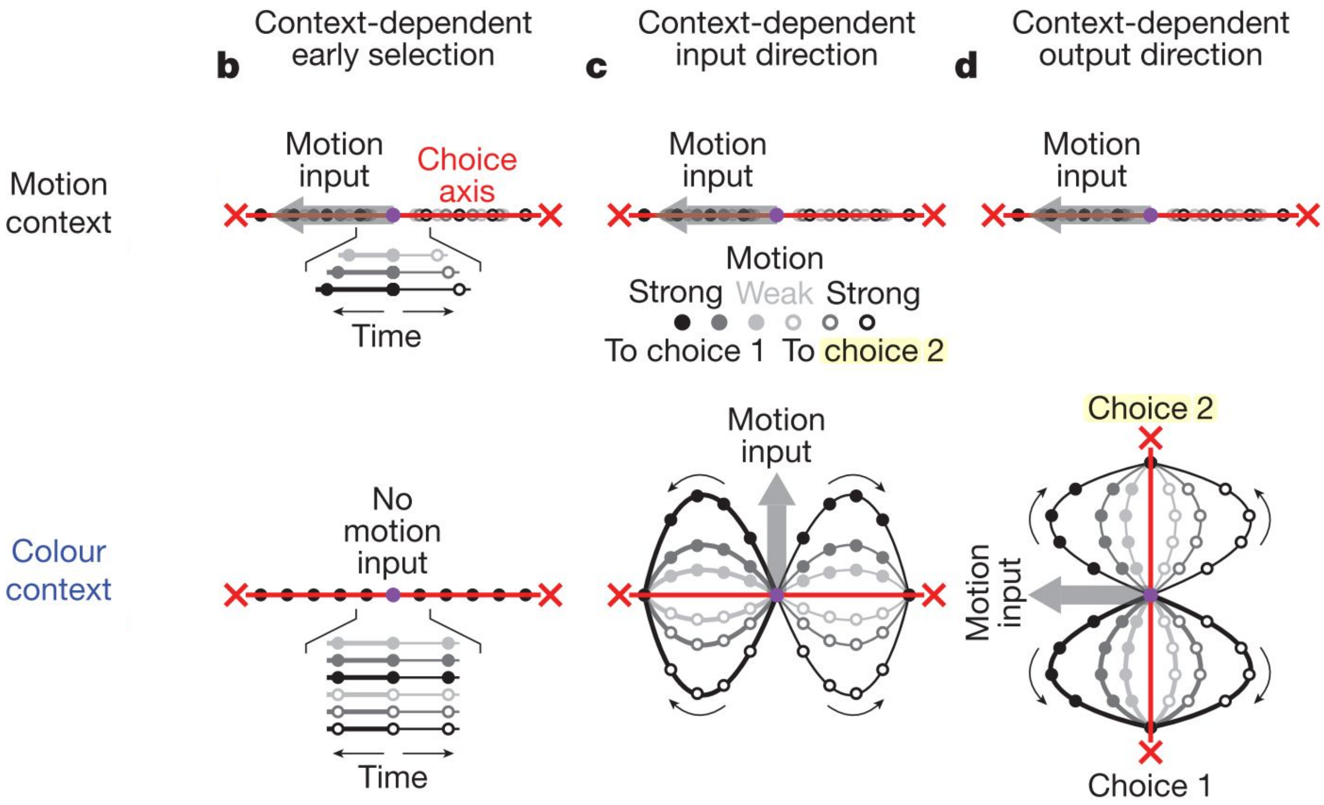
Stitch together condition average response over all neurons, using data collected from separate neurons in separate sessions

Perform dimensionality reduction (**PCA**) with samples corresponding to a condition at a given point in time

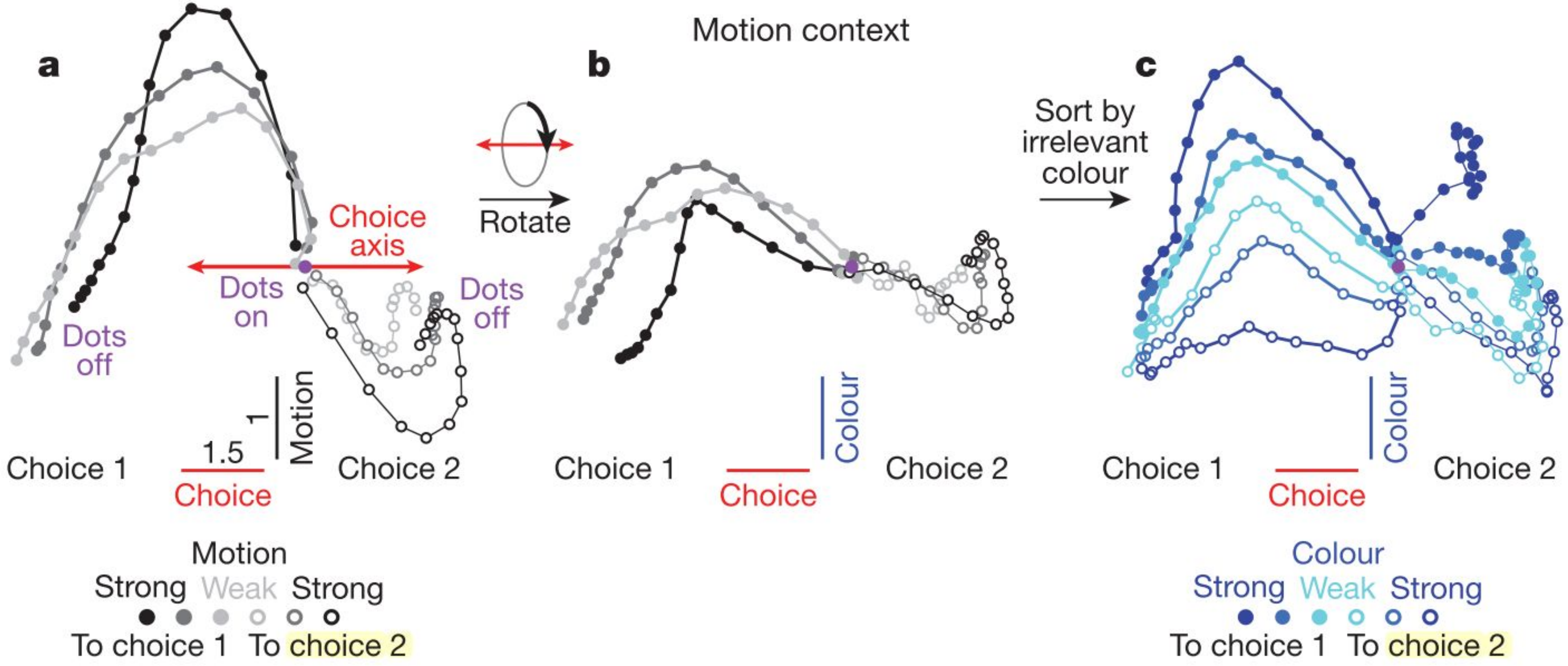
Targeted dimensionality reduction is then performed by regressing the PCs onto specific task components - color, motion, choice, and context

This is done while **orthogonalizing** the task components, such that each component explains unique variance - choice dimension soaks up variance first

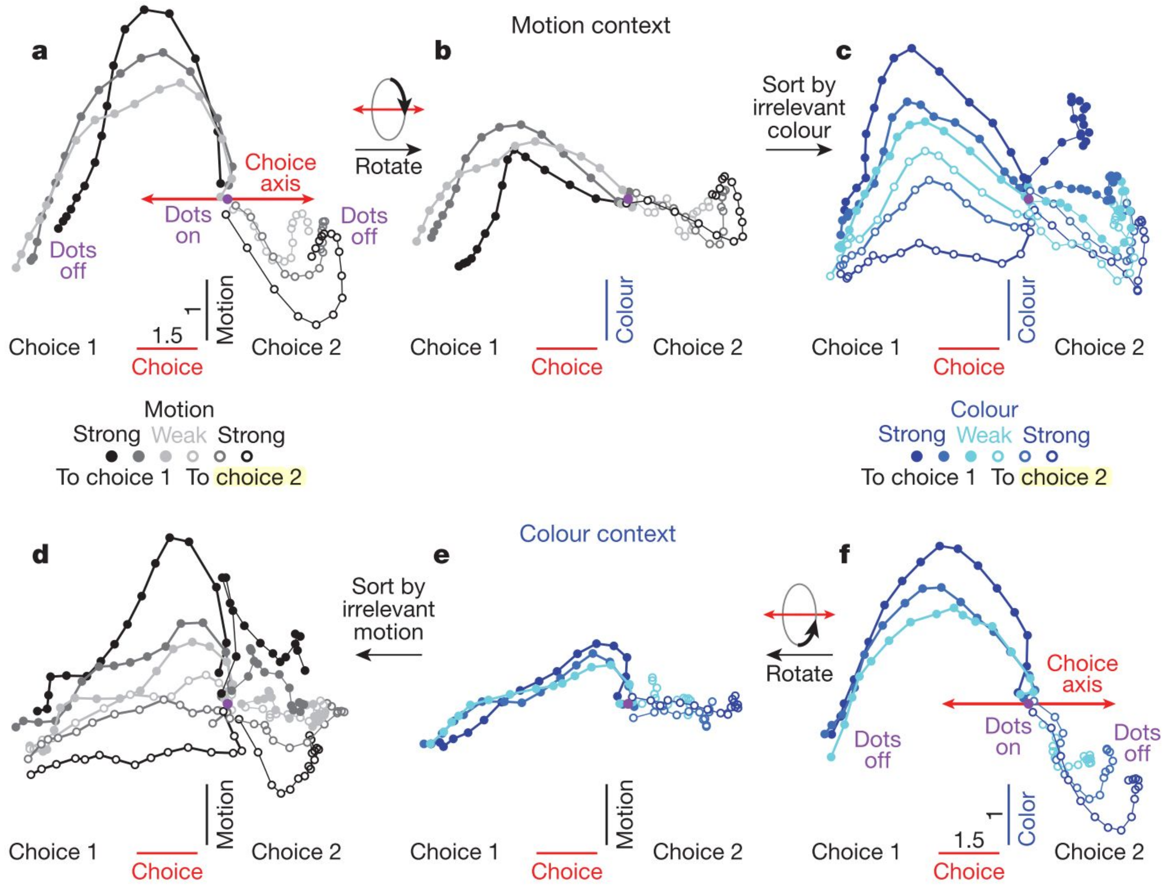
Models of selective integration



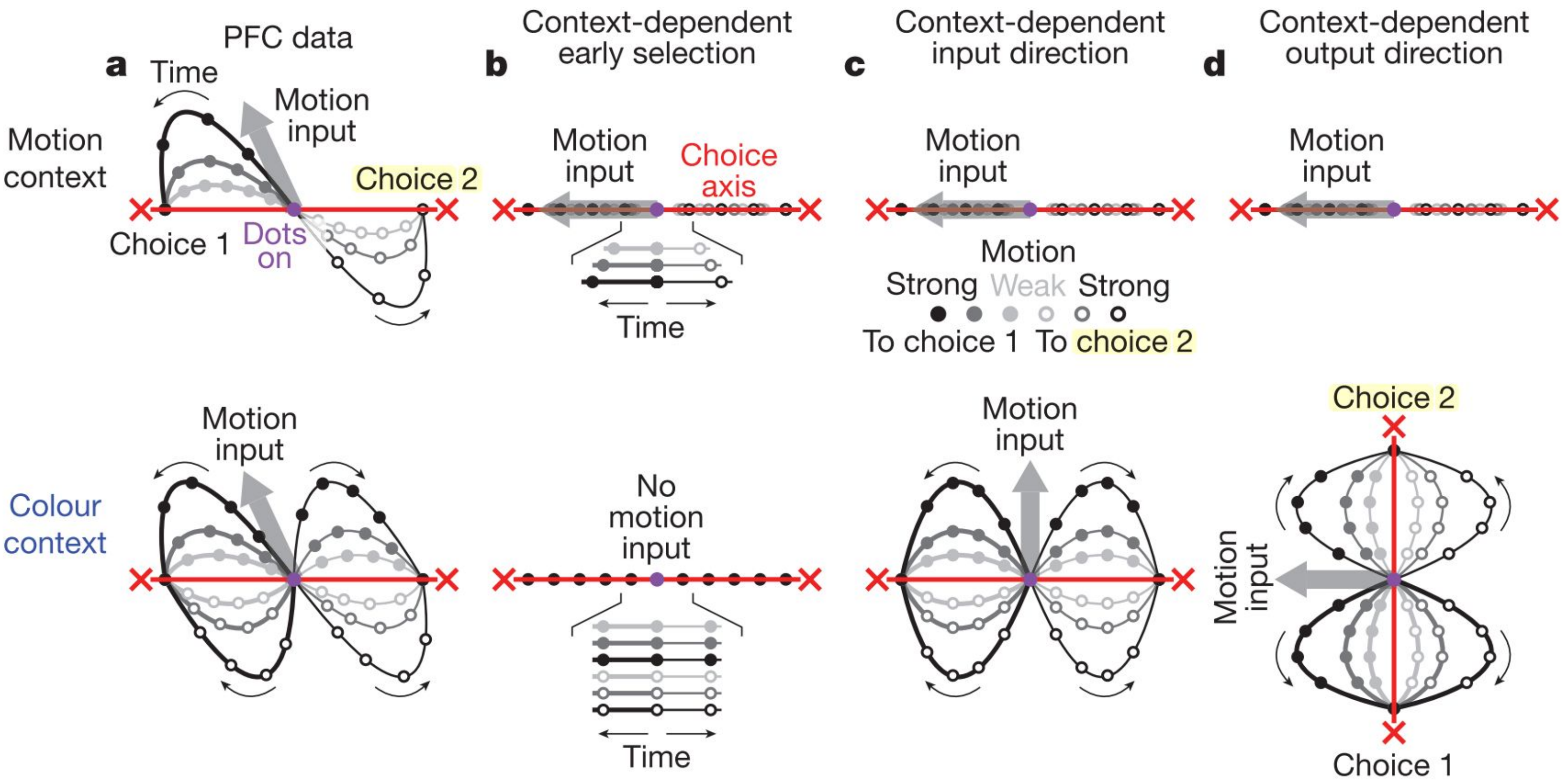
Dynamics of population responses in PFC



Dynamics of population responses in PFC

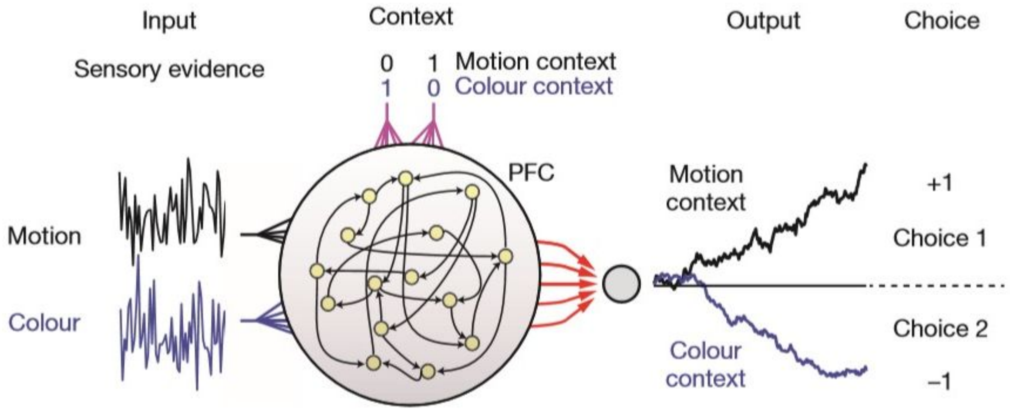


Models of selective integration are inconsistent with PFC



Recurrent models for selection & integration

- Identify mechanism via simulation
- Use a recurrent neural network (RNN)
  - Random weights
  - Trained via backpropagation through time



RNN

u_c - color
u_m - motion
u_{cc} - context color
u_{cm} - context motion

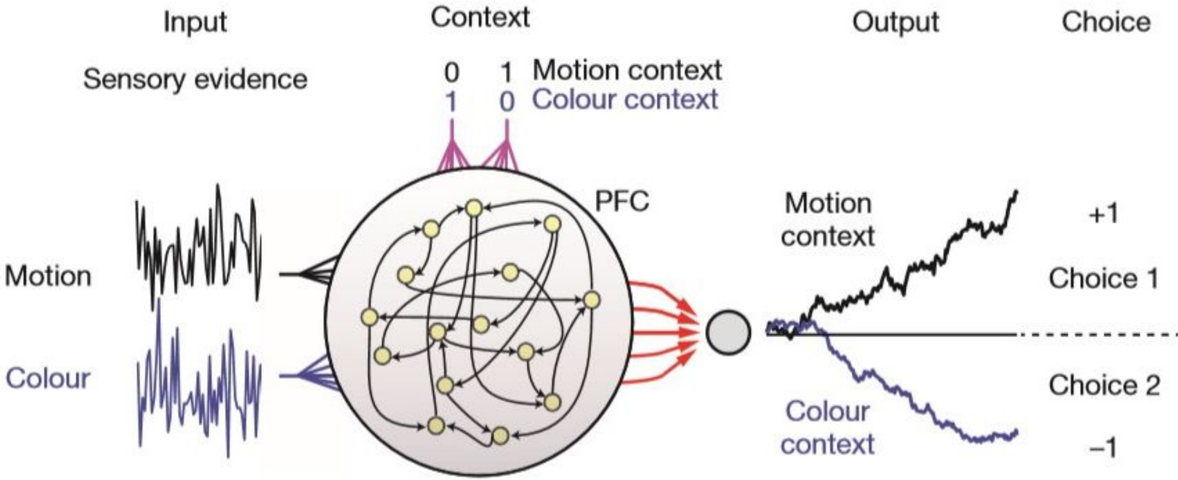
$\Delta x =$

First order update equation: $$\tau \dot{\boldsymbol{x}} = -\boldsymbol{x} + \boldsymbol{J}\boldsymbol{r} + \boldsymbol{b}^c u_c + \boldsymbol{b}^m u_m + \boldsymbol{b}^{cc} u_{cc} + \boldsymbol{b}^{cm} u_{cm} + \boldsymbol{c}^x + \rho_x$$

x - activation
$\dot{x}$ - rate of change of activation w.r.t time
J - weight matrix initialized randomly
r - *tanh(x)*
u - input
b - input weights
c - DC current
ρ - noise

$$u_m(t) = d_m + \rho_m(t)$$

$$u_c(t) = d_c + \rho_c(t).$$

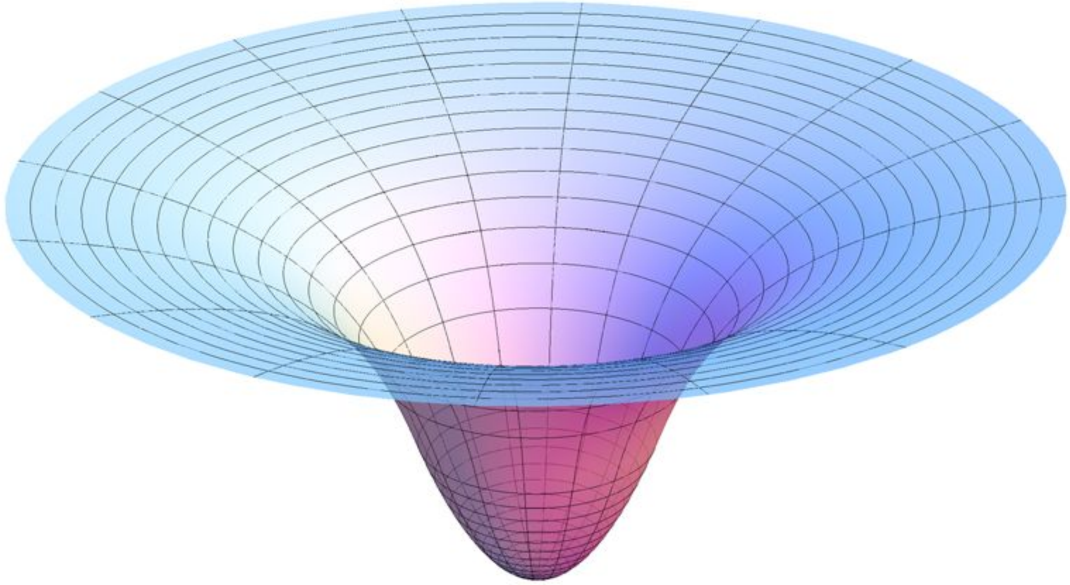


$$z = \boldsymbol{w}^T \boldsymbol{r} + c^z.$$

L_2 loss = Mean Squared Error

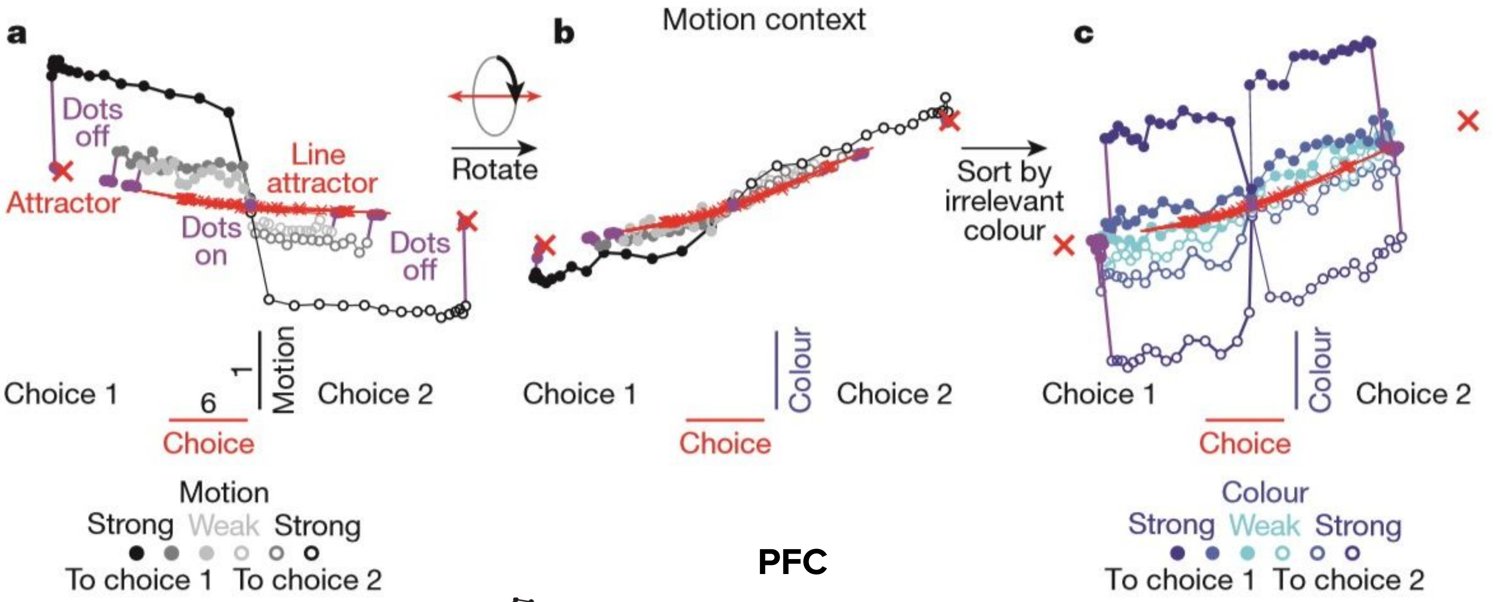
Attactors in state space

- Like a gravity well
- Stable states a network can go into

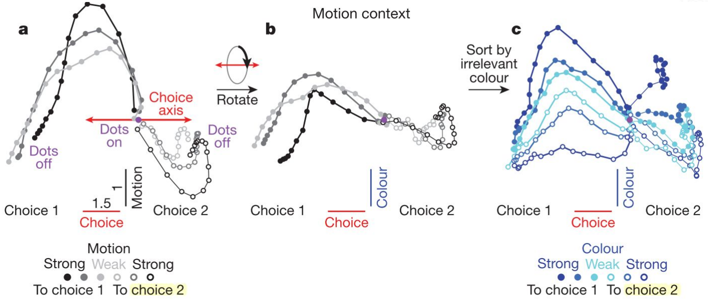


Fixed point analysis

RNN

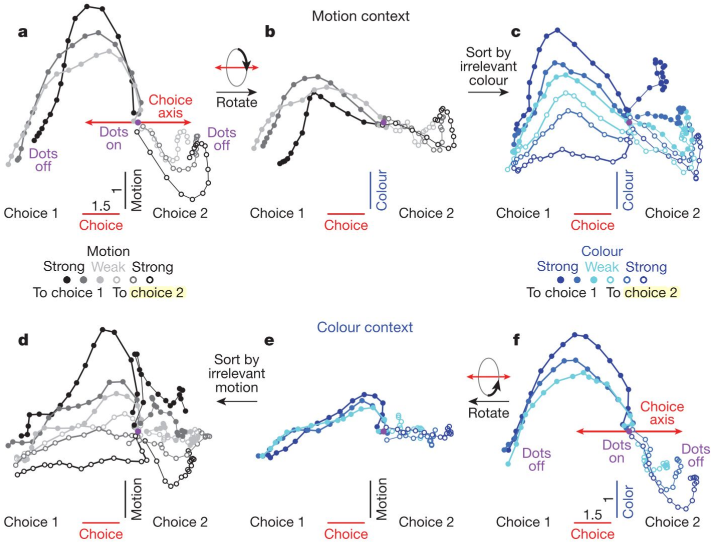


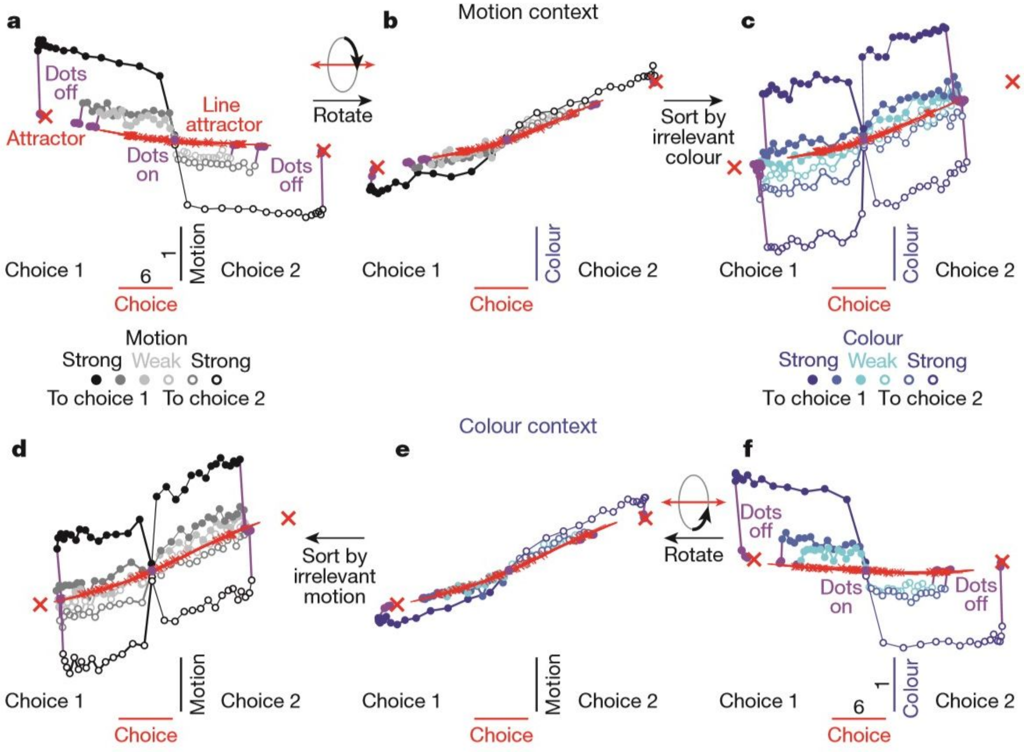
PFC



Fixed point analysis

- State space analysis
- Finds fixed points by optimizing the state space such that the step size is minimized

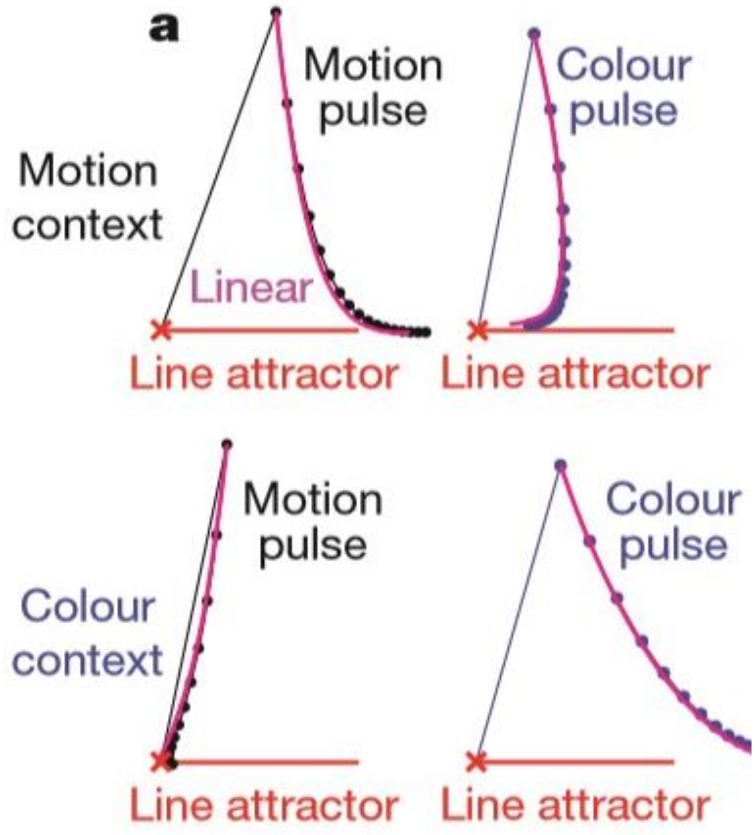




Analysis and proposed mechanisms

- Select corresponding context
- Initialize network on a fixed point
- Apply 1ms pulsed input
- Observe relaxation after stimulus stops

- Surprisingly, for the exact same weights, the network state depends strongly on the context



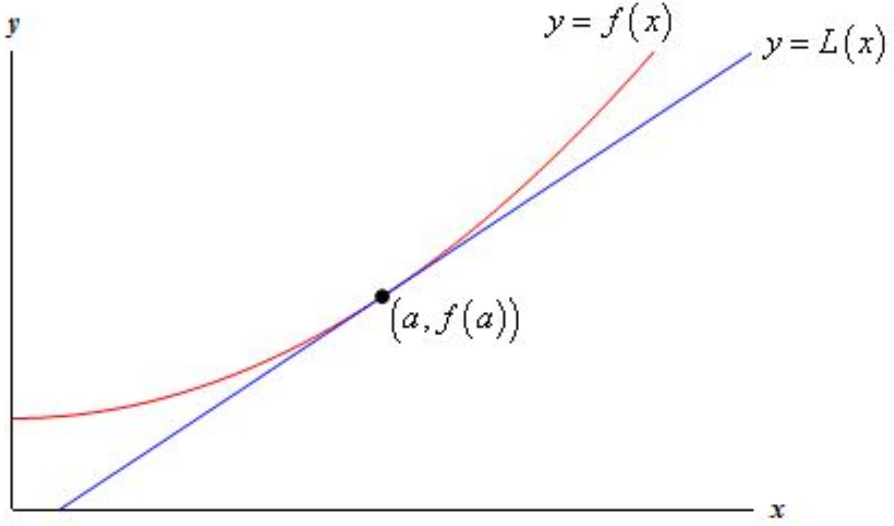
Local linearization

- Use a first order approximation of the local state space
- Allows for nice analysis using standard linear algebra tools

$$\dot{(\mathbf{x}^* + \delta\mathbf{x})} = \mathbf{F}'(\mathbf{x}^*)\delta\mathbf{x}$$

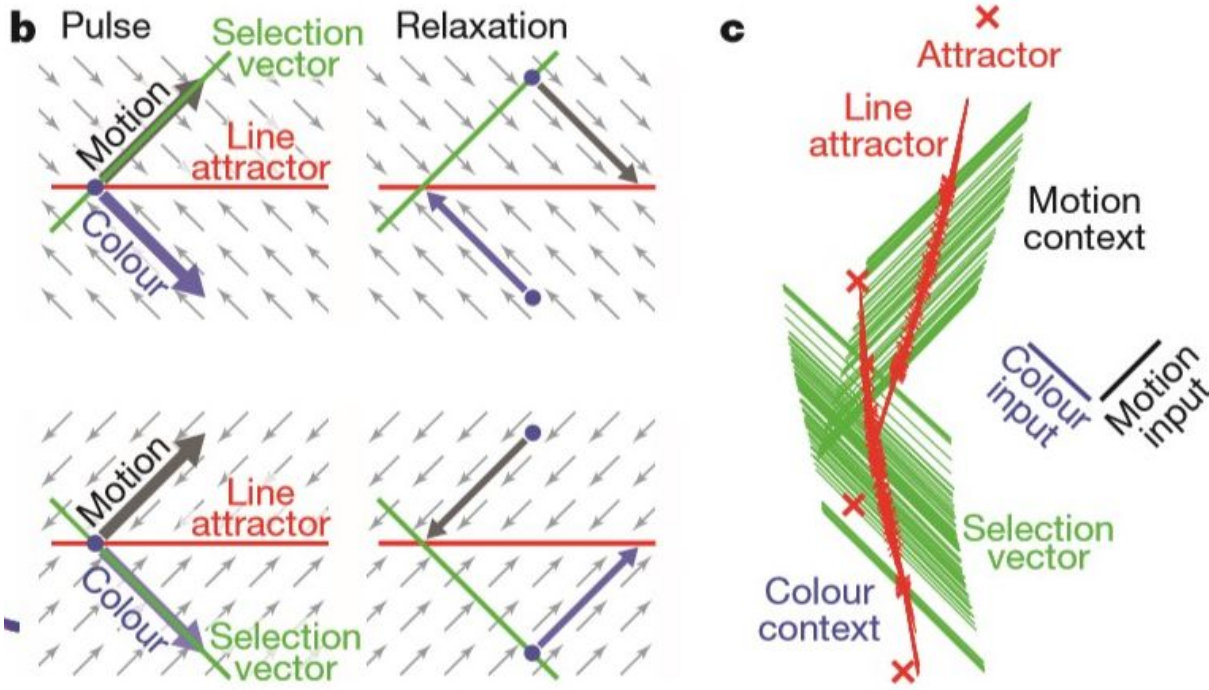
$$\dot{\delta\mathbf{x}} = \mathbf{F}'(\mathbf{x}^*)\delta\mathbf{x}$$

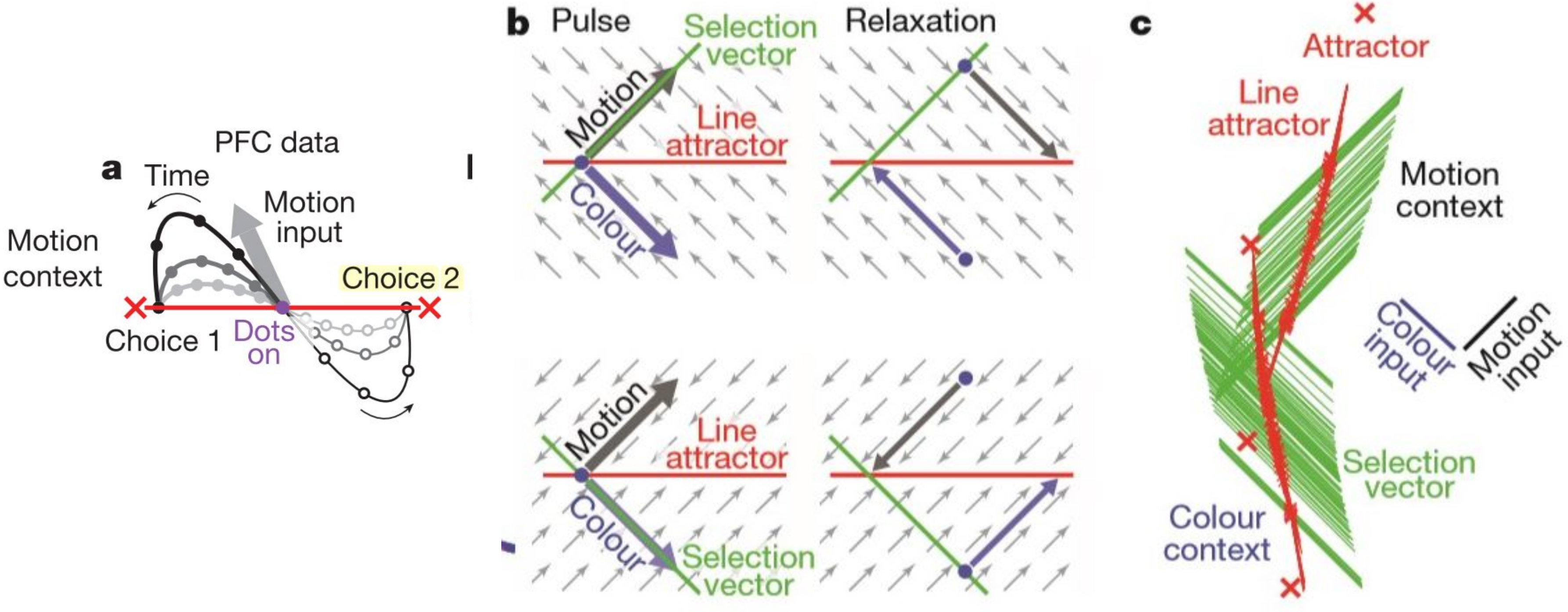
$$\dot{\mathbf{y}} = \mathbf{M}\mathbf{y}.$$



Evidence integration conditioned on context

- Eigenvector decomposition on the local linear approximation of the state space
- Allows the isolation of "selection vectors"
- Controls direction of relaxation
- Context determines selection vector
- Orthogonal to the irrelevant input




a
PFC data
Time
Motion input
Motion context
Choice 2
Choice 1
Dots on
b
Pulse
Selection vector
Motion
Line attractor
Colour
Relaxation
Motion
Line attractor
Colour
Selection vector
c
Attractor
Line attractor
Motion context
Colour input
Motion input
Selection vector
Colour context

Conclusion & Discussion

- Neural population responses in FEF exhibit **context-sensitive behavior**
- Cannot be explained by early selection (sensory gating) or context-sensitive input/output direction models
- A trained RNN can exhibit context-sensitive behavior and population responses emulated observed data from PFC
- The behavior of the high-dimensional RNN can be understood in a **low dimensional** framework using linear dynamical systems analyses
  - Context-sensitive **selection vector** determining responses to sensory inputs
  - **Line attractor** aligned with the choice dimension, yielding stable **persistent activity** representing the correct choice following stimulus extermination