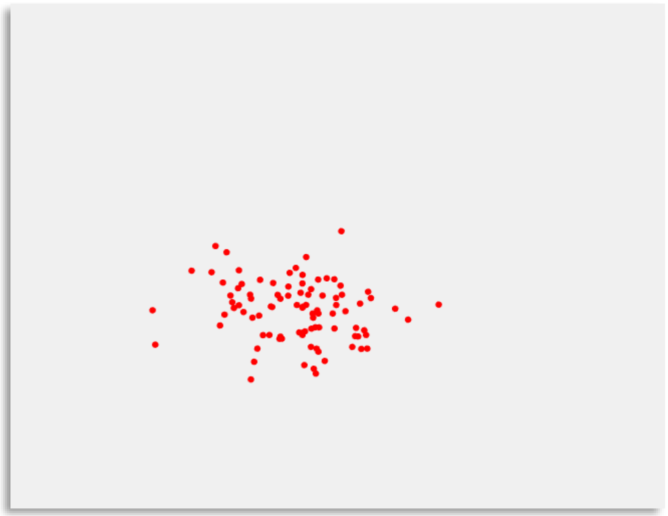
Active inference

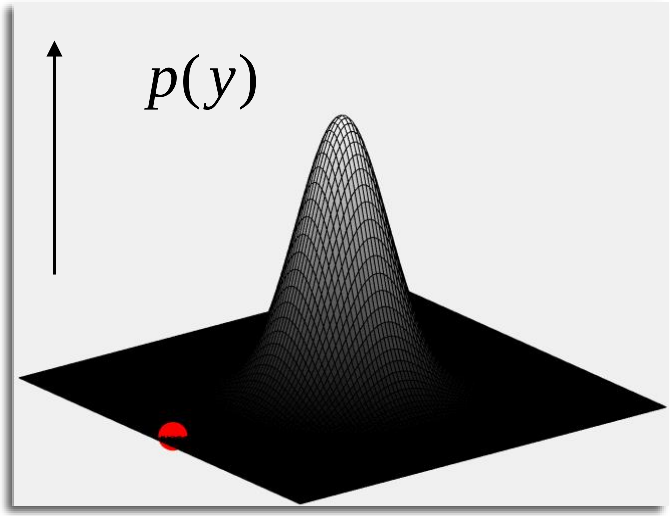
Computational psychiatry course

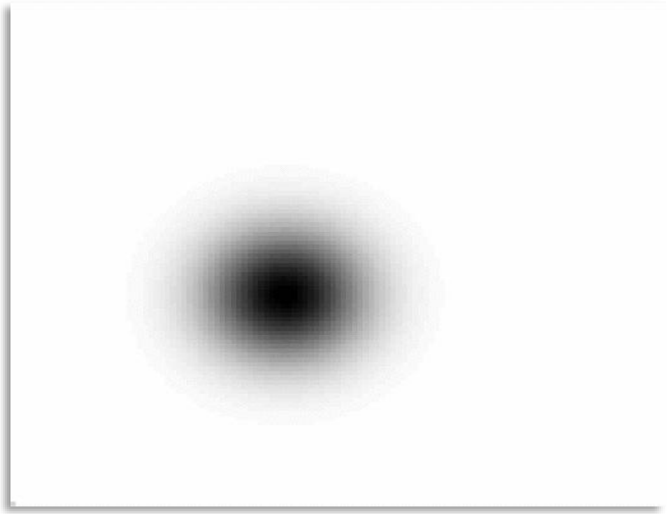
2022

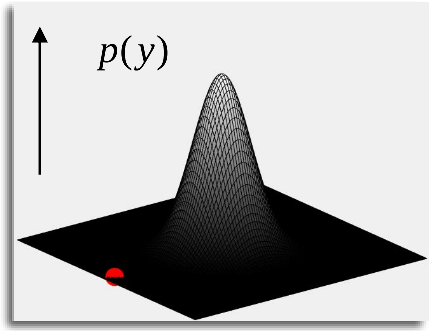
Active inference
Generative models
Exploitation
Exploration
Movement
Hierarchy

Active inference
Generative models
Exploitation
Exploration
Movement
Hierarchy

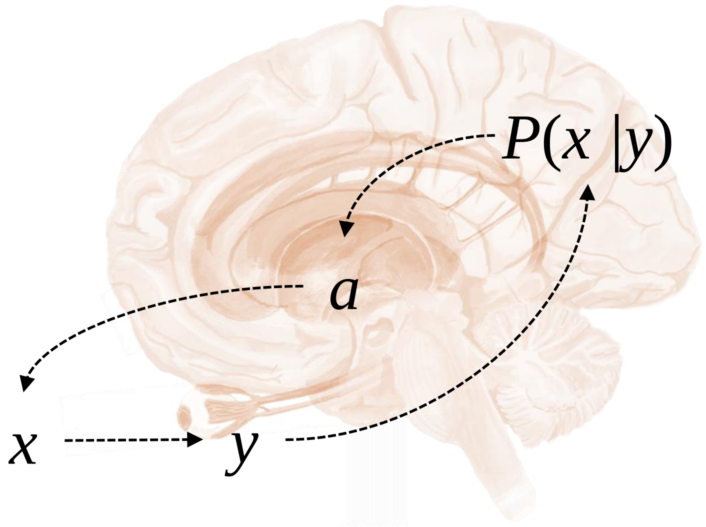



$p(y)$



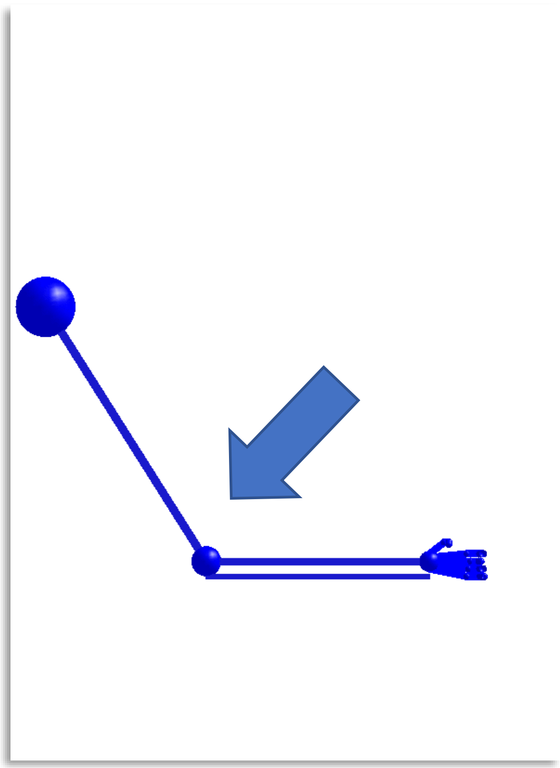


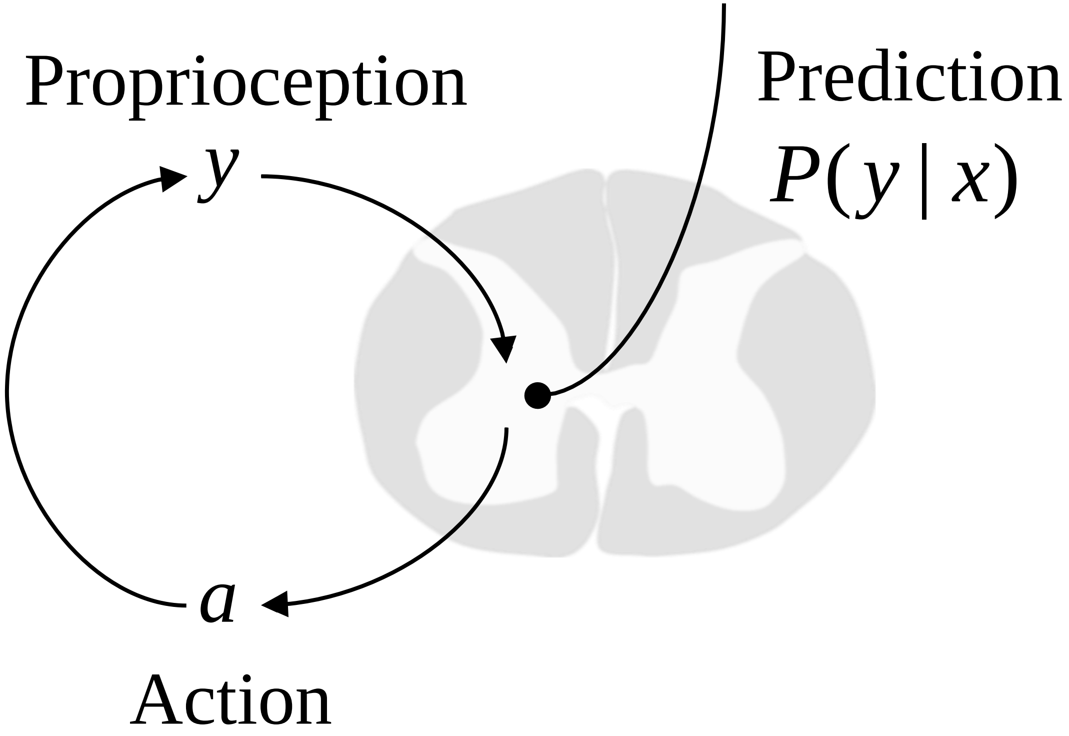
Self-evidencing – Active Inference



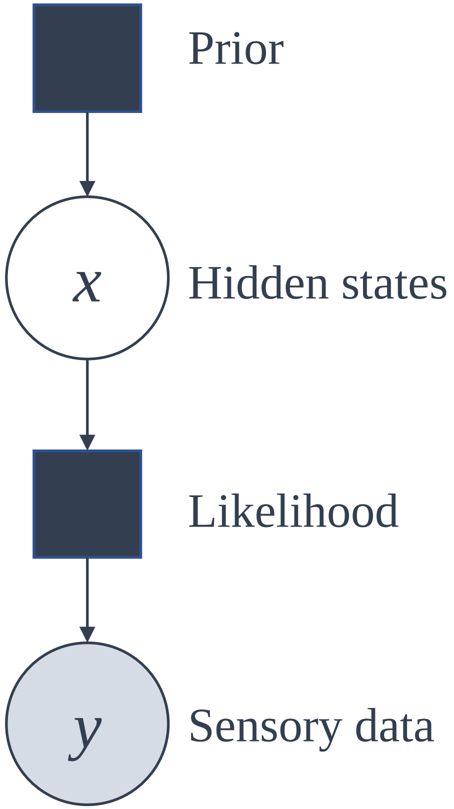
$$\underbrace{\underbrace{P(y\,|\,x)}_{\text{Likelihood}}\underbrace{P(x)}_{\text{Prior}}}_{\text{Generative model}} = \underbrace{\underbrace{P(x\,|\,y)}_{\text{Posterior}}\underbrace{P(y)}_{\text{Evidence}}}_{\text{Inverted model}}$$

Reflexes



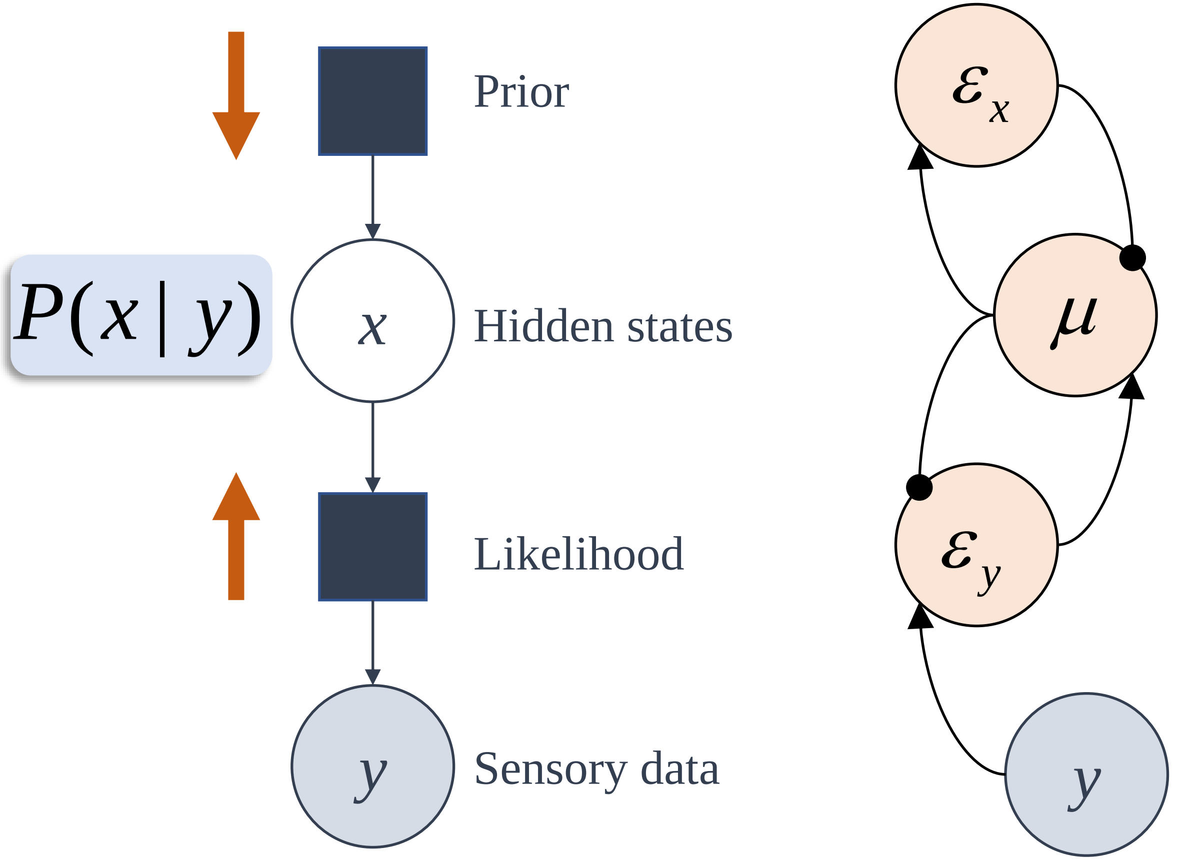


Models and messages



$$P(x \mid y)P(y) = P(y \mid x)P(x)$$

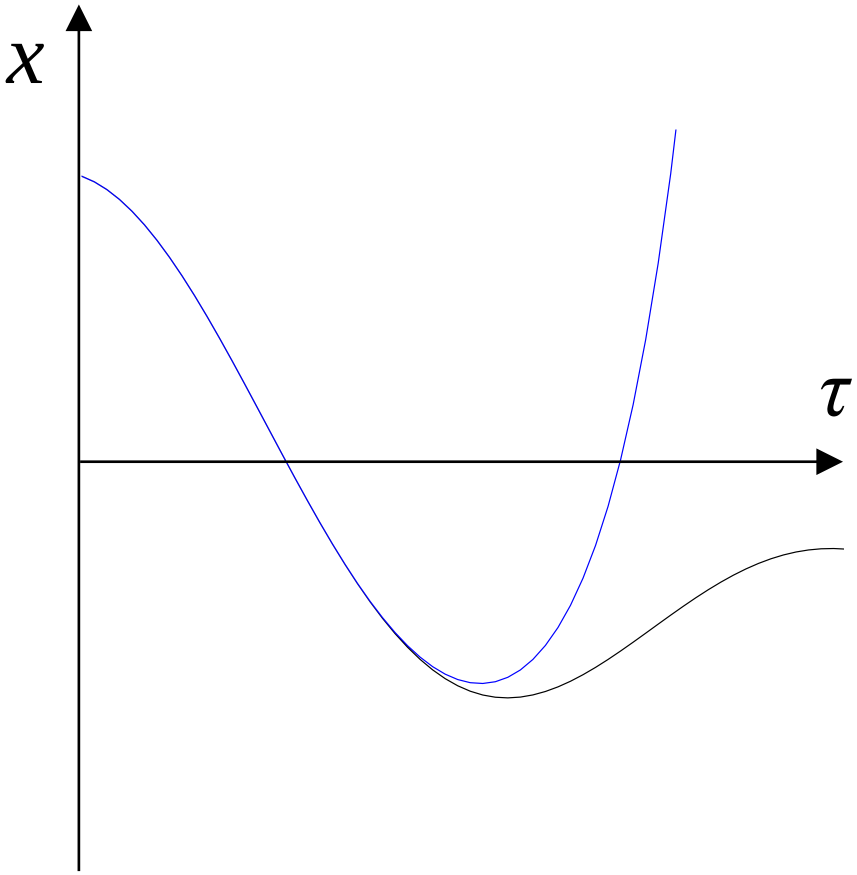
Models and messages

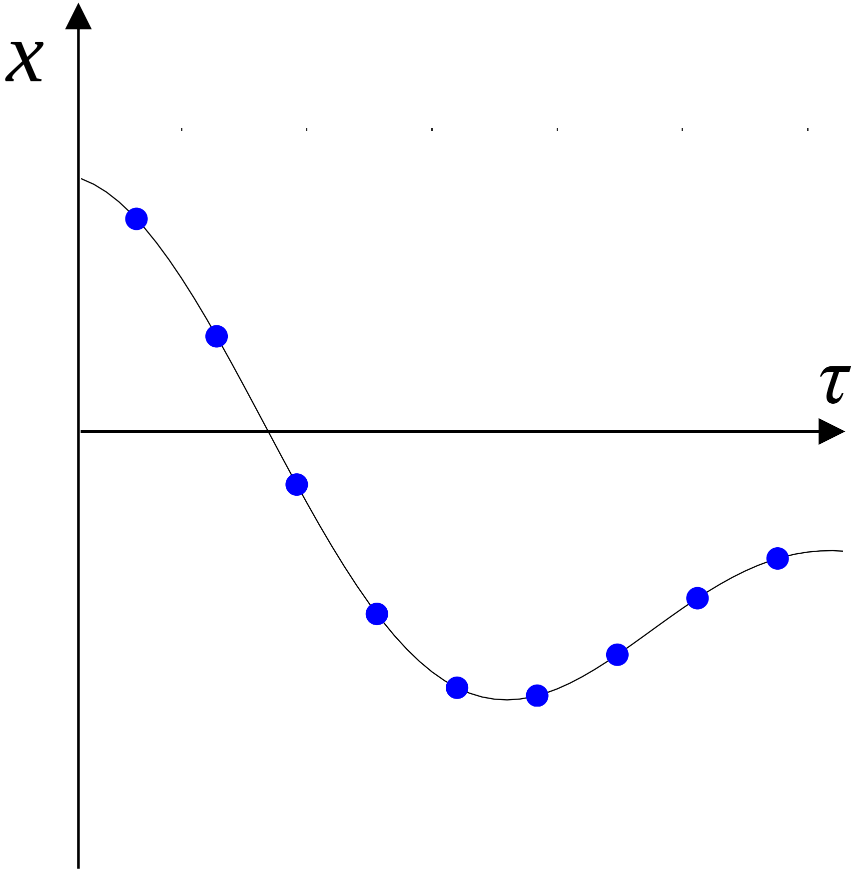


$$P(x \mid y) \propto P(y \mid x)P(x)$$

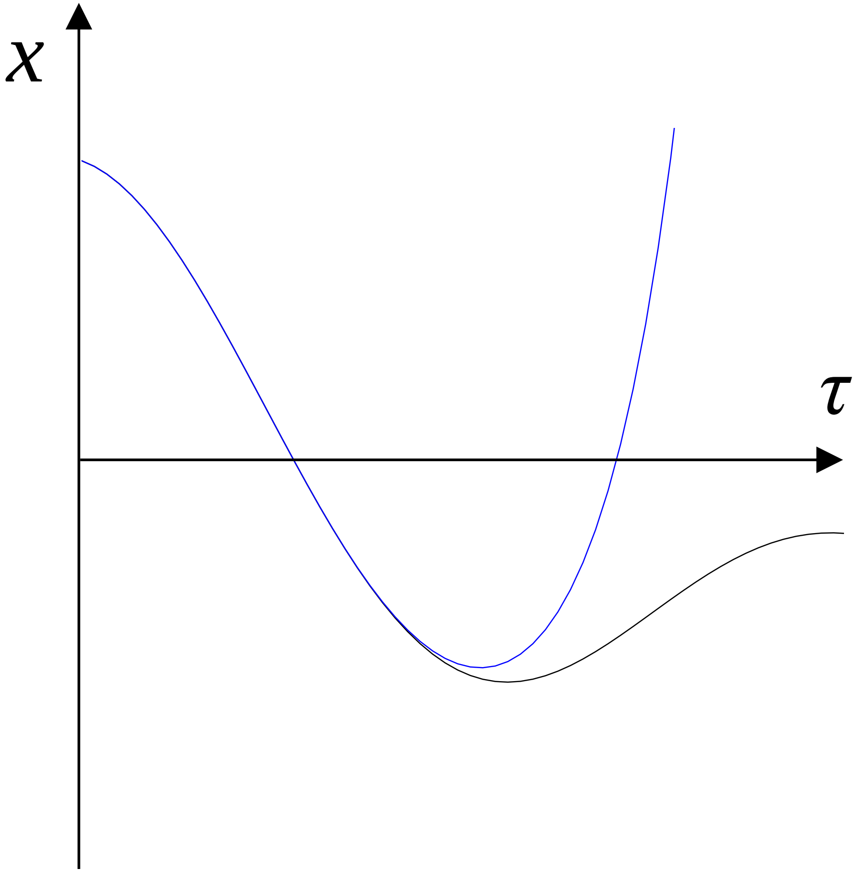
Active inference
Generative models
Exploitation
Exploration
Movement
Hierarchy

Representing dynamics in generative models

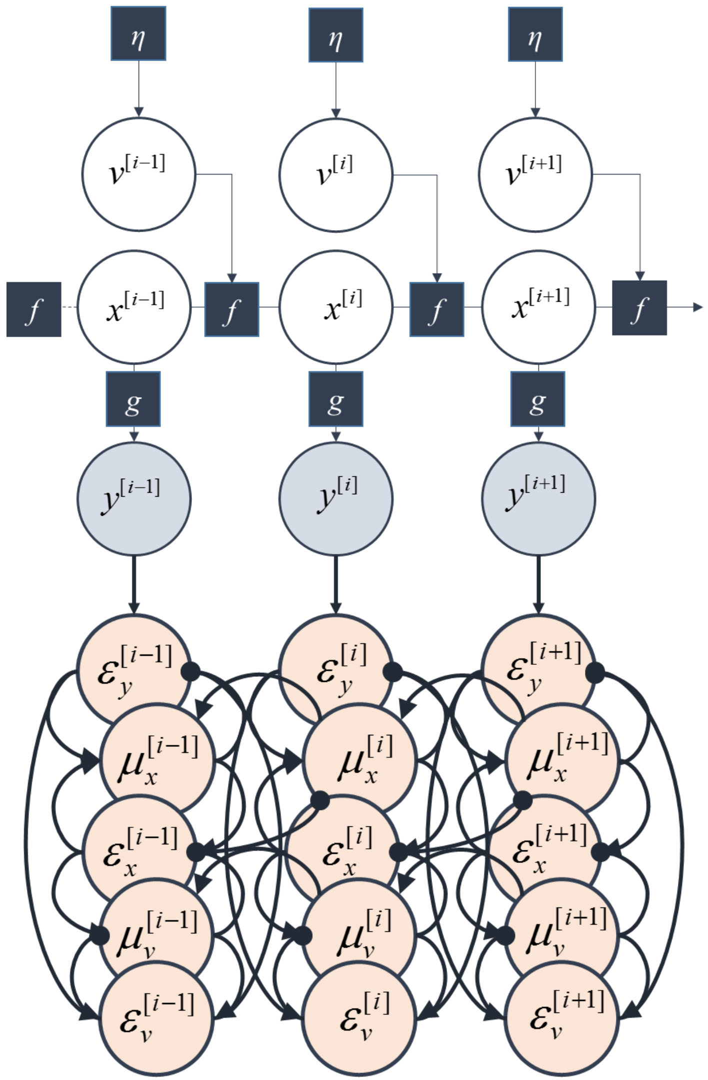




Continuous time models



$$x(\tau) \approx x_0 + \tau x_0' + \tfrac{1}{2}\tau^2 x_0'' + \tfrac{1}{6}\tau^3 x_0''' + \cdots$$



Generative model

$$p(\tilde{y},\tilde{x},\tilde{v}) = \prod_i p(v^{[i]})\,p(x^{[i+1]} \mid x^{[i]},v^{[i]})\,p(y^{[i]} \mid x^{[i]},v^{[i]})$$

$$p(y^{[i]} \mid x^{[i]},v^{[i]}) = \mathcal{N}(g^{[i]}(x^{[i]},v^{[i]}),\Pi_y^{[i]})$$

$$p(x^{[i+1]} \mid x^{[i]},v^{[i]}) = \mathcal{N}(f^{[i]}(x^{[i]},v^{[i]}),\Pi_x^{[i]})$$

$$p(v^{[i]}) = \mathcal{N}(\eta^{[i]},\Pi_v^{[i]})$$

Bayesian message passing

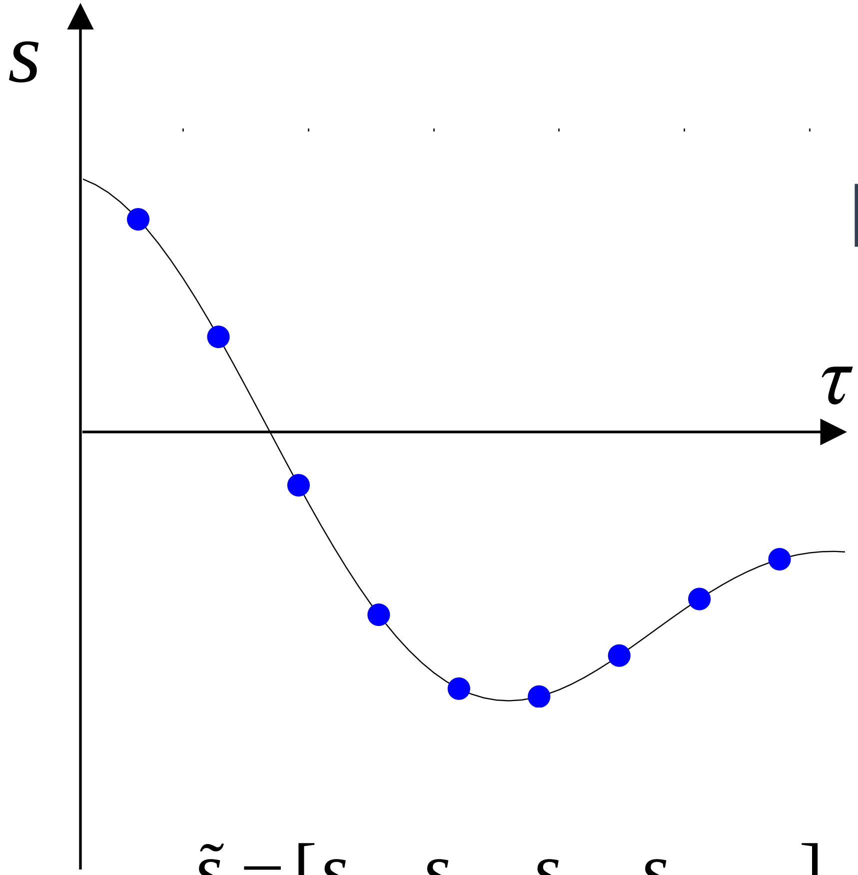
$$\varepsilon_y^{[i]} = y^{[i]} - g^{[i]}(\mu_x^{[i]},\mu_v^{[i]})$$

$$\varepsilon_x^{[i]} = \mu_x^{[i+1]} - f^{[i]}(\mu_x^{[i]},\mu_v^{[i]})$$

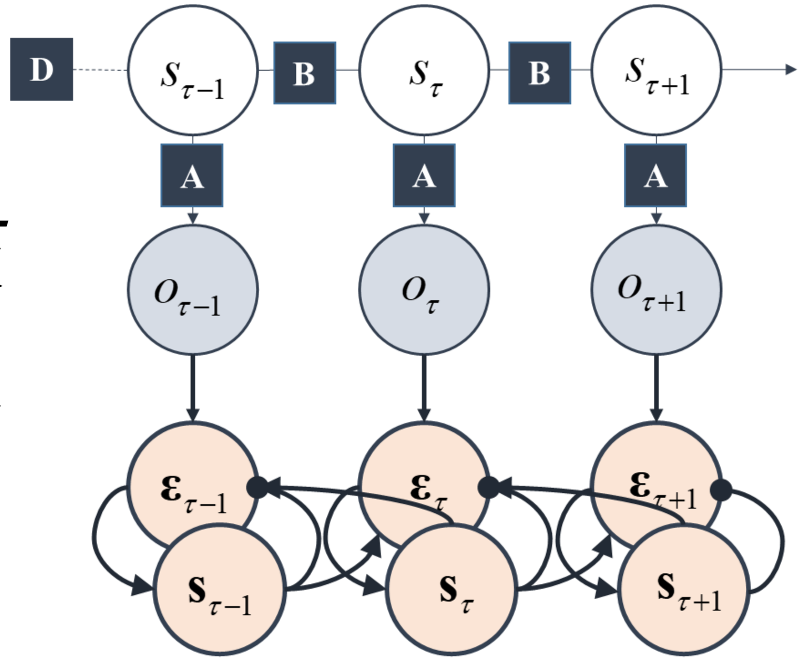
$$\varepsilon_v^{[i]} = \mu_v^{[i]} - \eta^{[i]}$$

$$\begin{aligned}\dot{\mu}_x^{[i]} &= \mu_x^{[i+1]} \\ &\quad + \partial_{\mu_x^{[i]}} g^{[i]} \cdot \Pi_y^{[i]} \varepsilon_y^{[i]} - \Pi_x^{[i-1]} \varepsilon_x^{[i-1]} + \partial_{\mu_x^{[i]}} f^{[i]} \cdot \Pi_x^{[i]} \varepsilon_x^{[i]} \\ \dot{\mu}_v^{[i]} &= \mu_v^{[i+1]} \\ &\quad + \partial_{\mu_v^{[i]}} g^{[i]} \cdot \Pi_y^{[i]} \varepsilon_y^{[i]} + \partial_{\mu_v^{[i]}} f^{[i]} \cdot \Pi_x^{[i]} \varepsilon_x^{[i]} - \Pi_v^{[i]} \varepsilon_v^{[i]}\end{aligned}$$

Discrete time models



$$\tilde{s} = [s_1,\ s_2,\ s_3,\ s_4,\ \ldots]$$



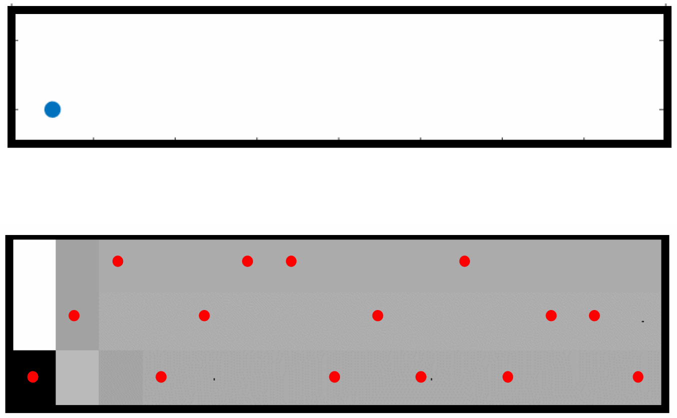
Generative model

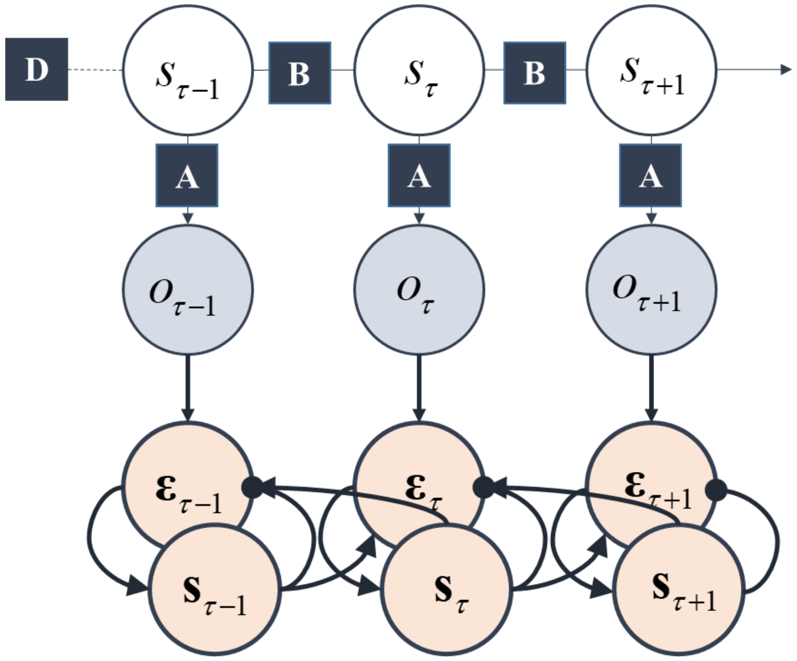
$$P(\tilde{o},\tilde{s}) = P(s_1)\prod_{\tau} P(s_{\tau+1} \mid s_\tau)P(o_\tau \mid s_\tau)$$

$$P(o_\tau \mid s_\tau) = Cat(\mathbf{A})$$
$$P(s_\tau \mid s_{\tau-1}) = Cat(\mathbf{B})$$
$$P(s_1) = Cat(\mathbf{D})$$

Bayesian message passing

$$\mathbf{s}_\tau = \sigma(\mathbf{v}_\tau);\ \ \dot{\mathbf{v}}_\tau = \boldsymbol{\varepsilon}_\tau$$
$$\boldsymbol{\varepsilon}_\tau = \ln \mathbf{A} \cdot o_\tau + \tfrac{1}{2}\ln(\mathbf{B}_\tau \mathbf{s}_\tau) + \tfrac{1}{2}\ln(\mathbf{B}^{\dagger}_{\tau+1}\mathbf{s}_{\tau+1}) - \ln \mathbf{s}_\tau$$





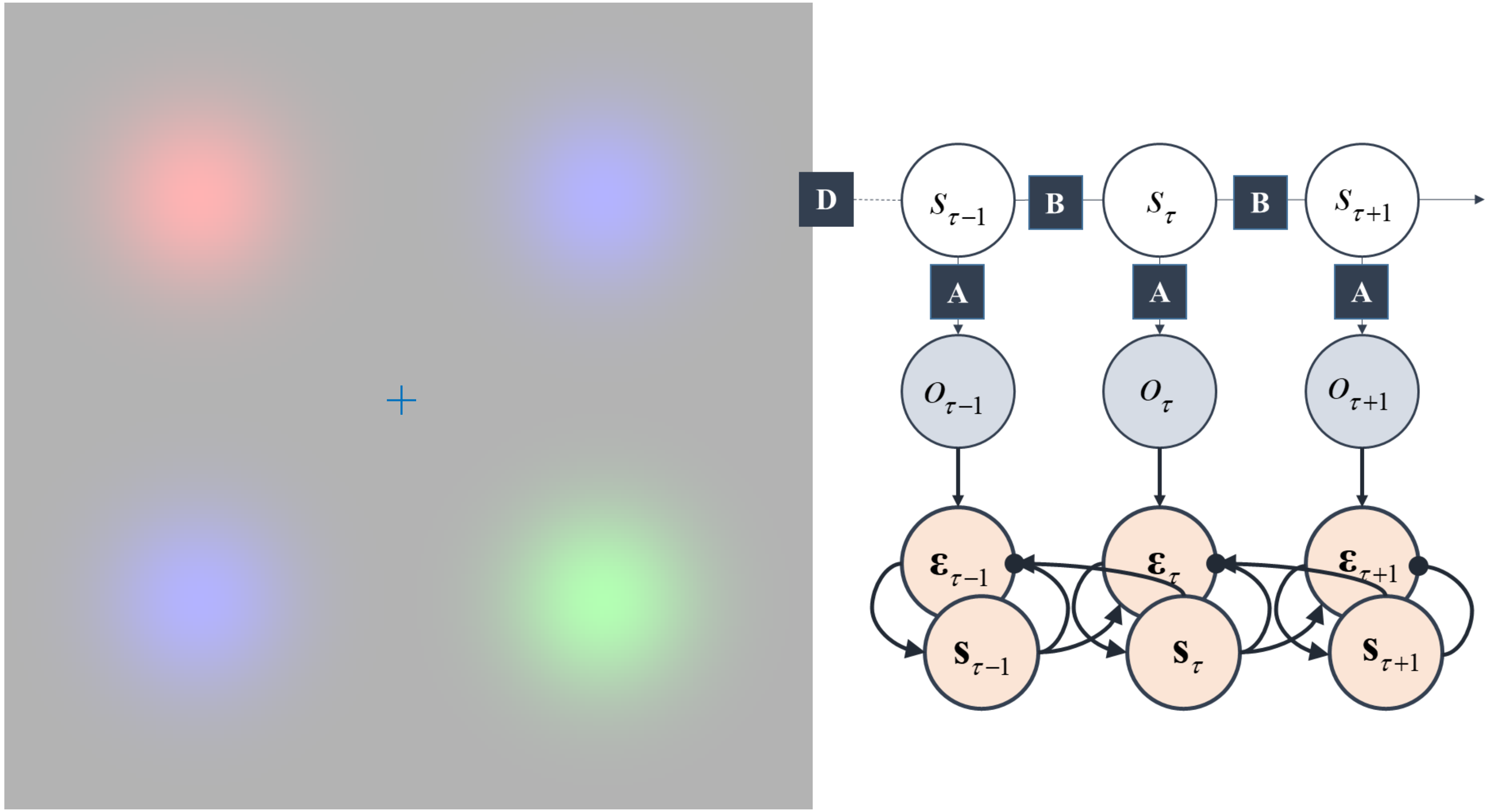
Generative model

$$P(\tilde{o},\tilde{s}) = P(s_1)\prod_{\tau} P(s_{\tau+1} \mid s_\tau)P(o_\tau \mid s_\tau)$$

$$P(o_\tau \mid s_\tau) = Cat(\mathbf{A})$$
$$P(s_\tau \mid s_{\tau-1}) = Cat(\mathbf{B})$$
$$P(s_1) = Cat(\mathbf{D})$$

Bayesian message passing

$$\mathbf{s}_\tau = \sigma(\mathbf{v}_\tau); \;\; \dot{\mathbf{v}}_\tau = \boldsymbol{\varepsilon}_\tau$$
$$\boldsymbol{\varepsilon}_\tau = \ln \mathbf{A} \cdot o_\tau + \tfrac{1}{2}\ln(\mathbf{B}_\tau \mathbf{s}_\tau) + \tfrac{1}{2}\ln(\mathbf{B}^{\dagger}_{\tau+1}\mathbf{s}_{\tau+1}) - \ln \mathbf{s}_\tau$$



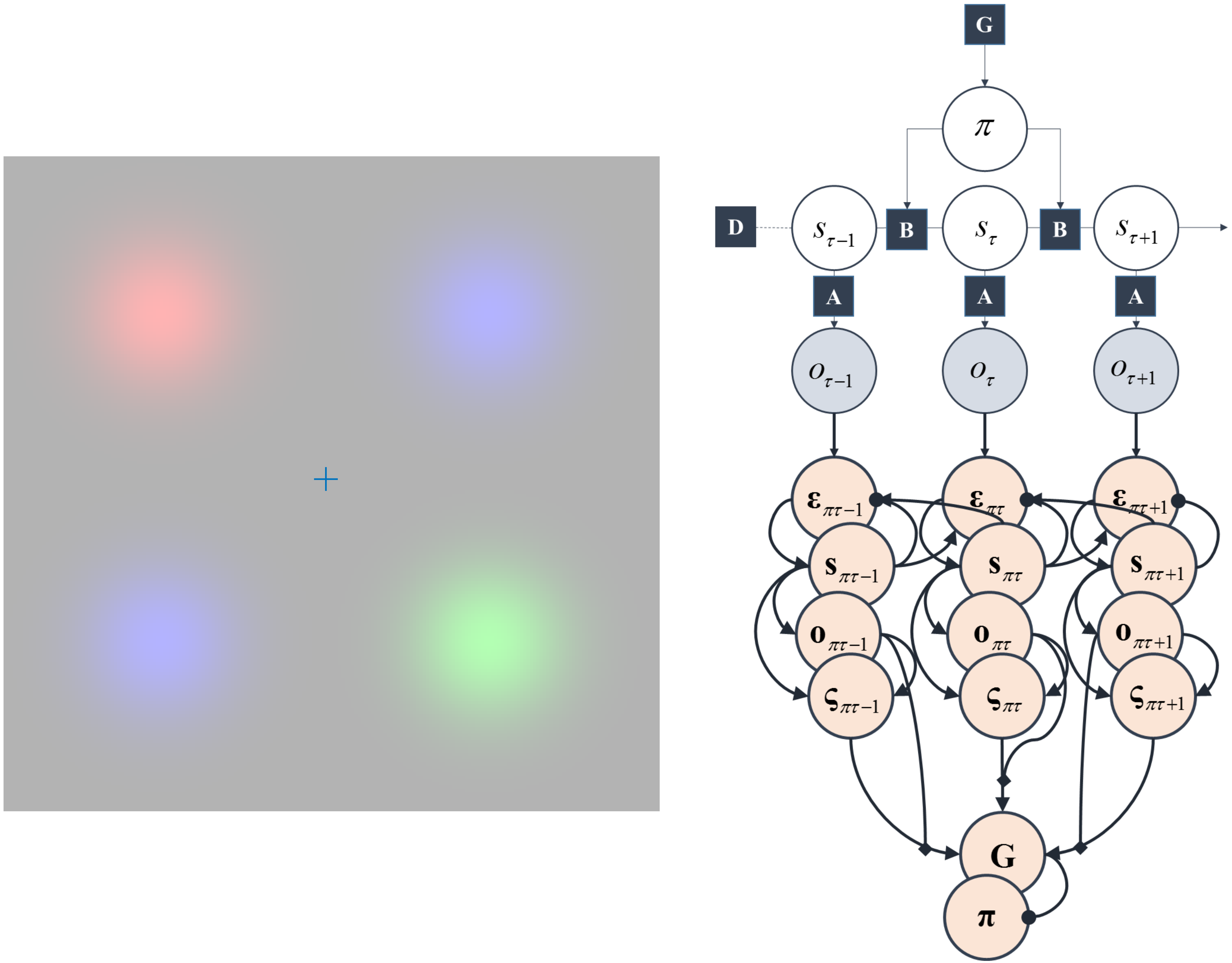
Generative model

$$P(\tilde{o},\tilde{s}) = P(s_1)\prod_{\tau} P(s_{\tau+1} \mid s_\tau) P(o_\tau \mid s_\tau)$$

$$P(o_\tau \mid s_\tau) = Cat(\mathbf{A})$$
$$P(s_\tau \mid s_{\tau-1}) = Cat(\mathbf{B})$$
$$P(s_1) = Cat(\mathbf{D})$$

Bayesian message passing

$$\mathbf{s}_\tau = \sigma(\mathbf{v}_\tau);\ \ \dot{\mathbf{v}}_\tau = \boldsymbol{\varepsilon}_\tau$$
$$\boldsymbol{\varepsilon}_\tau = \ln \mathbf{A} \cdot o_\tau + \tfrac{1}{2}\ln(\mathbf{B}_\tau \mathbf{s}_\tau) + \tfrac{1}{2}\ln(\mathbf{B}^{\dagger}_{\tau+1}\mathbf{s}_{\tau+1}) - \ln \mathbf{s}_\tau$$



Generative model

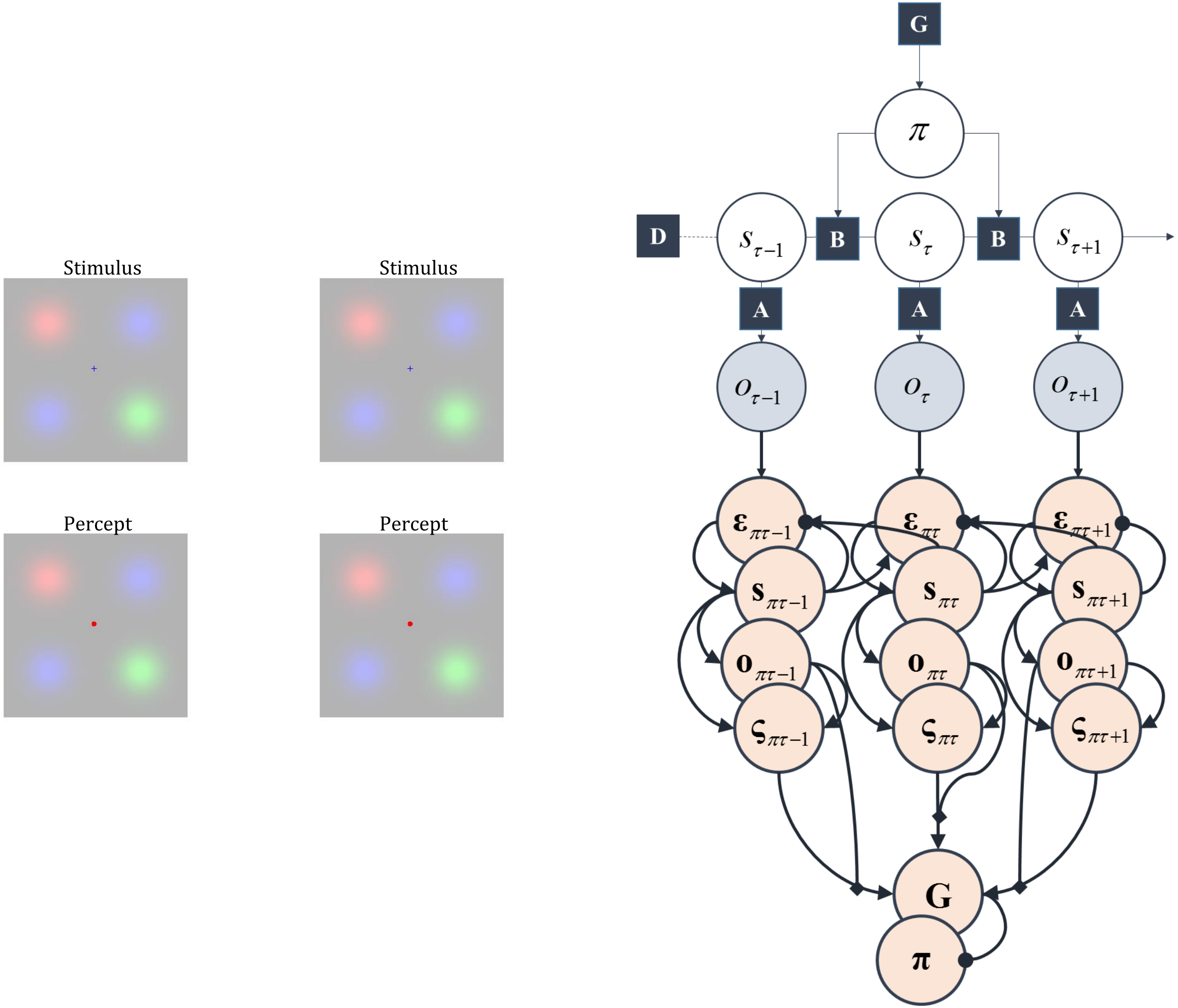
$$P(\tilde{o},\tilde{s},\pi)=P(s_1)P(\pi)\prod_{\tau}P(s_{\tau+1}\mid s_\tau,\pi)P(o_\tau\mid s_\tau)$$

$$P(o_\tau\mid s_\tau)=Cat(\mathbf{A})$$
$$P(s_\tau\mid s_{\tau-1},\pi)=Cat(\mathbf{B}_{\pi\tau})$$
$$P(o_\tau)=Cat(\mathbf{C})$$
$$P(s_1)=Cat(\mathbf{D})$$
$$P(\pi)=\sigma(-\mathbf{G})$$

Bayesian message passing

$$\mathbf{s}_\tau=\boldsymbol{\pi}\cdot\mathbf{s}_{\pi\tau}$$
$$\mathbf{s}_{\pi\tau}=\sigma(\mathbf{v}_{\pi\tau});\ \dot{\mathbf{v}}_{\pi\tau}=\boldsymbol{\varepsilon}_{\pi\tau}$$
$$\boldsymbol{\varepsilon}_{\pi\tau}=\ln\mathbf{A}\cdot o_\tau+\tfrac{1}{2}\ln(\mathbf{B}_{\pi\tau}\mathbf{s}_{\pi\tau})+\tfrac{1}{2}\ln(\mathbf{B}^{\dagger}_{\pi\tau+1}\mathbf{s}_{\pi\tau+1})-\ln\mathbf{s}_{\pi\tau}$$

$$\mathbf{o}_{\pi\tau}=\mathbf{A}\mathbf{s}_{\pi\tau}$$
$$\boldsymbol{\varsigma}_{\pi\tau}=\ln\mathbf{o}_{\pi\tau}-\ln\mathbf{C}+\mathbf{H}\cdot\mathbf{s}_{\pi\tau}$$
$$\mathbf{H}=-diag(\mathbf{A}\cdot\ln\mathbf{A})$$
$$\mathbf{G}_\pi=\mathbf{o}_{\pi\tau}\cdot\boldsymbol{\varsigma}_{\pi\tau}$$
$$\boldsymbol{\pi}=\sigma(-\mathbf{G})$$



Generative model

$$P(\tilde{o},\tilde{s},\pi)=P(s_1)P(\pi)\prod_{\tau}P(s_{\tau+1}\mid s_\tau,\pi)P(o_\tau\mid s_\tau)$$

$$P(o_\tau\mid s_\tau)=Cat(\mathbf{A})$$
$$P(s_\tau\mid s_{\tau-1},\pi)=Cat(\mathbf{B}_{\pi\tau})$$
$$P(o_\tau)=Cat(\mathbf{C})$$
$$P(s_1)=Cat(\mathbf{D})$$
$$P(\pi)=\sigma(-\mathbf{G})$$

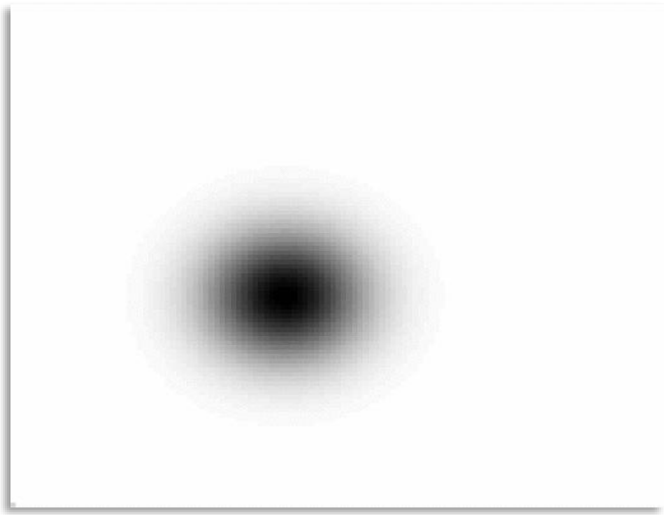
Bayesian message passing

$$\mathbf{s}_\tau=\boldsymbol{\pi}\cdot\mathbf{s}_{\pi\tau}$$
$$\mathbf{s}_{\pi\tau}=\sigma(\mathbf{v}_{\pi\tau});\ \ \dot{\mathbf{v}}_{\pi\tau}=\boldsymbol{\varepsilon}_{\pi\tau}$$
$$\boldsymbol{\varepsilon}_{\pi\tau}=\ln\mathbf{A}\cdot o_\tau+\tfrac{1}{2}\ln(\mathbf{B}_{\pi\tau}\mathbf{s}_{\pi\tau})+\tfrac{1}{2}\ln(\mathbf{B}^{\dagger}_{\pi\tau+1}\mathbf{s}_{\pi\tau+1})-\ln\mathbf{s}_{\pi\tau}$$

$$\mathbf{o}_{\pi\tau}=\mathbf{A}\mathbf{s}_{\pi\tau}$$
$$\boldsymbol{\varsigma}_{\pi\tau}=\ln\mathbf{o}_{\pi\tau}-\ln\mathbf{C}+\mathbf{H}\cdot\mathbf{s}_{\pi\tau}$$
$$\mathbf{H}=-diag(\mathbf{A}\cdot\ln\mathbf{A})$$
$$\mathbf{G}_\pi=\mathbf{o}_{\pi\tau}\cdot\boldsymbol{\varsigma}_{\pi\tau}$$
$$\boldsymbol{\pi}=\sigma(-\mathbf{G})$$

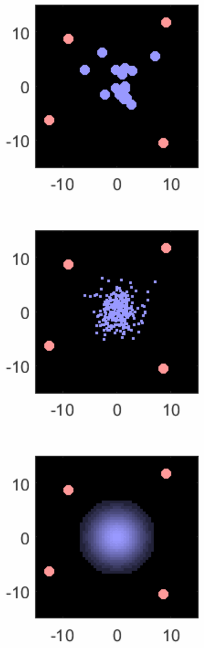
Active inference
Generative models
Exploitation
Exploration
Movement
Hierarchy

Goals and steady state



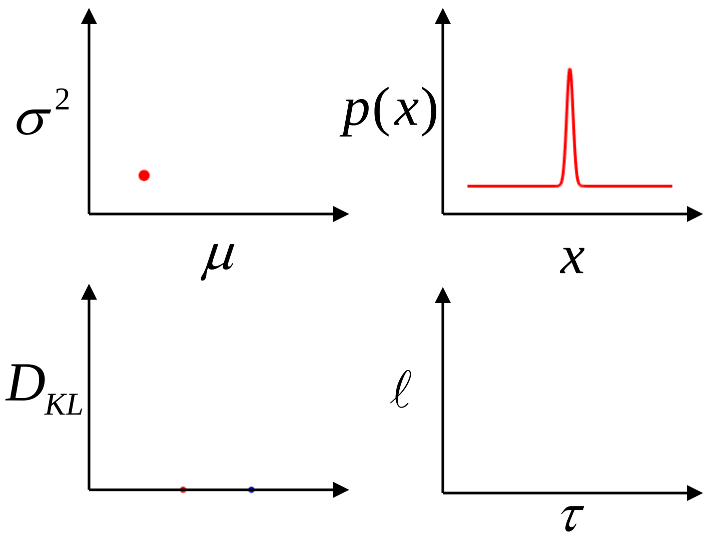
$$D_{KL}\left[P(o \mid \pi) \parallel P(o \mid C)\right]$$

Goals and steady state



$$D_{KL}\left[P(o\mid\pi)\parallel P(o\mid C)\right]$$

Goals and steady state



$$D_{KL}\left[P(o\,|\,\pi)\,\|\,P(o\,|\,C)\right]$$

Goals and steady state

$$-H\left[P(o\,|\,\pi)\right]-\mathrm{E}_{P(o|\pi)}\left[\ln P(o\,|\,C)\right]$$

$$\Updownarrow$$

$$D_{KL}\left[P(o\,|\,\pi)\,\|\,P(o\,|\,C)\right]$$

Active inference
Generative models
Exploitation
Exploration
Movement
Hierarchy

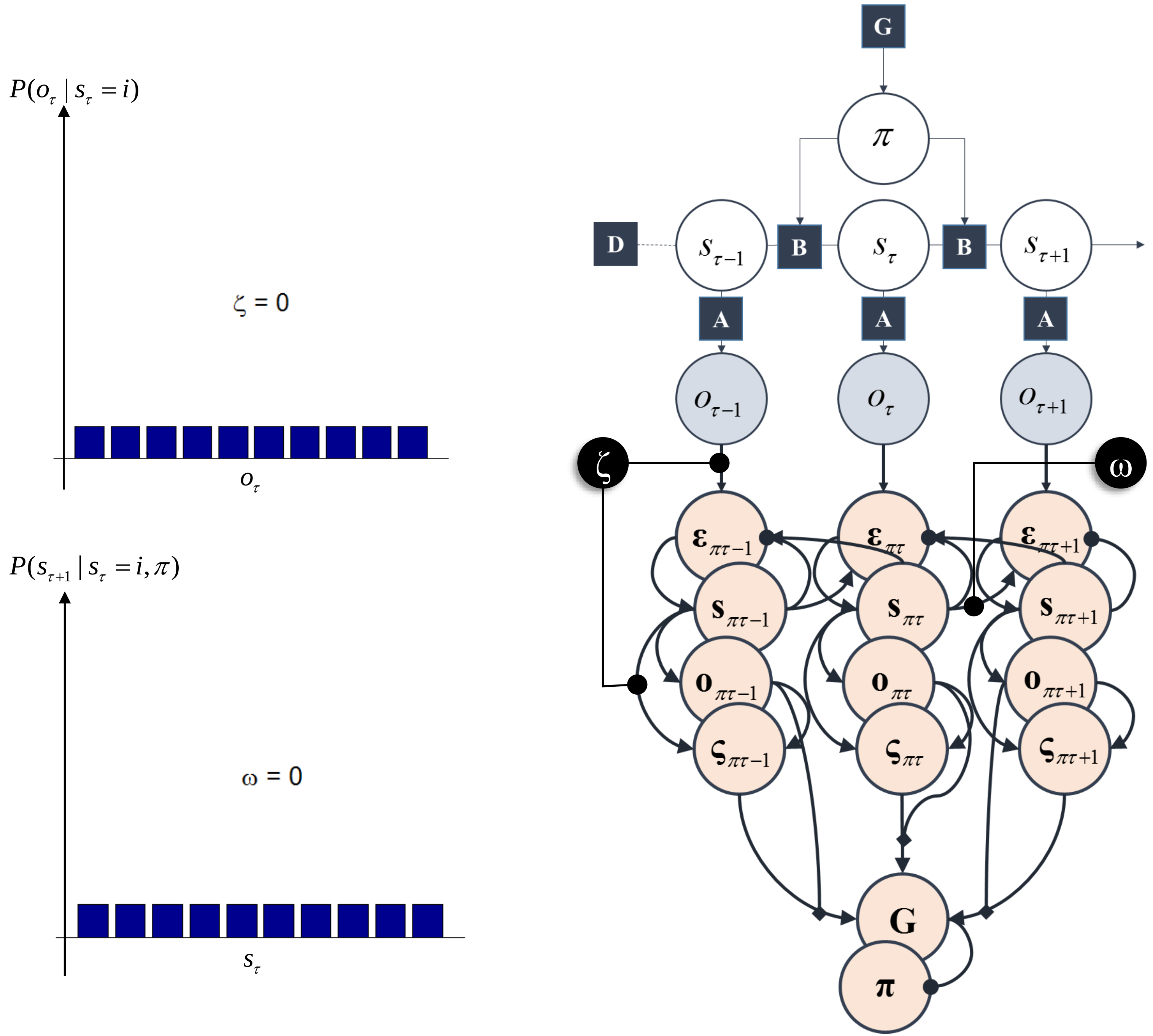
Information gain

$$\mathcal{I}(\pi) = D_{KL}\left[P(o,s\,|\,\pi)\,\|\,P(o\,|\,\pi)P(s\,|\,\pi)\right]$$

$$= \mathrm{E}_{P(o|\pi)}\left[D_{KL}\left[P(s\,|\,o,\pi)\,\|\,P(s\,|\,\pi)\right]\right]$$

$$= H\left[P(o\,|\,\pi)\right] - \mathrm{E}_{P(s|\pi)}\left[H\left[P(o\,|\,s)\right]\right]$$

$$P(o,s\,|\,\pi) = P(s\,|\,o,\pi)P(o\,|\,\pi) = P(o\,|\,s)P(s\,|\,\pi)$$



Generative model

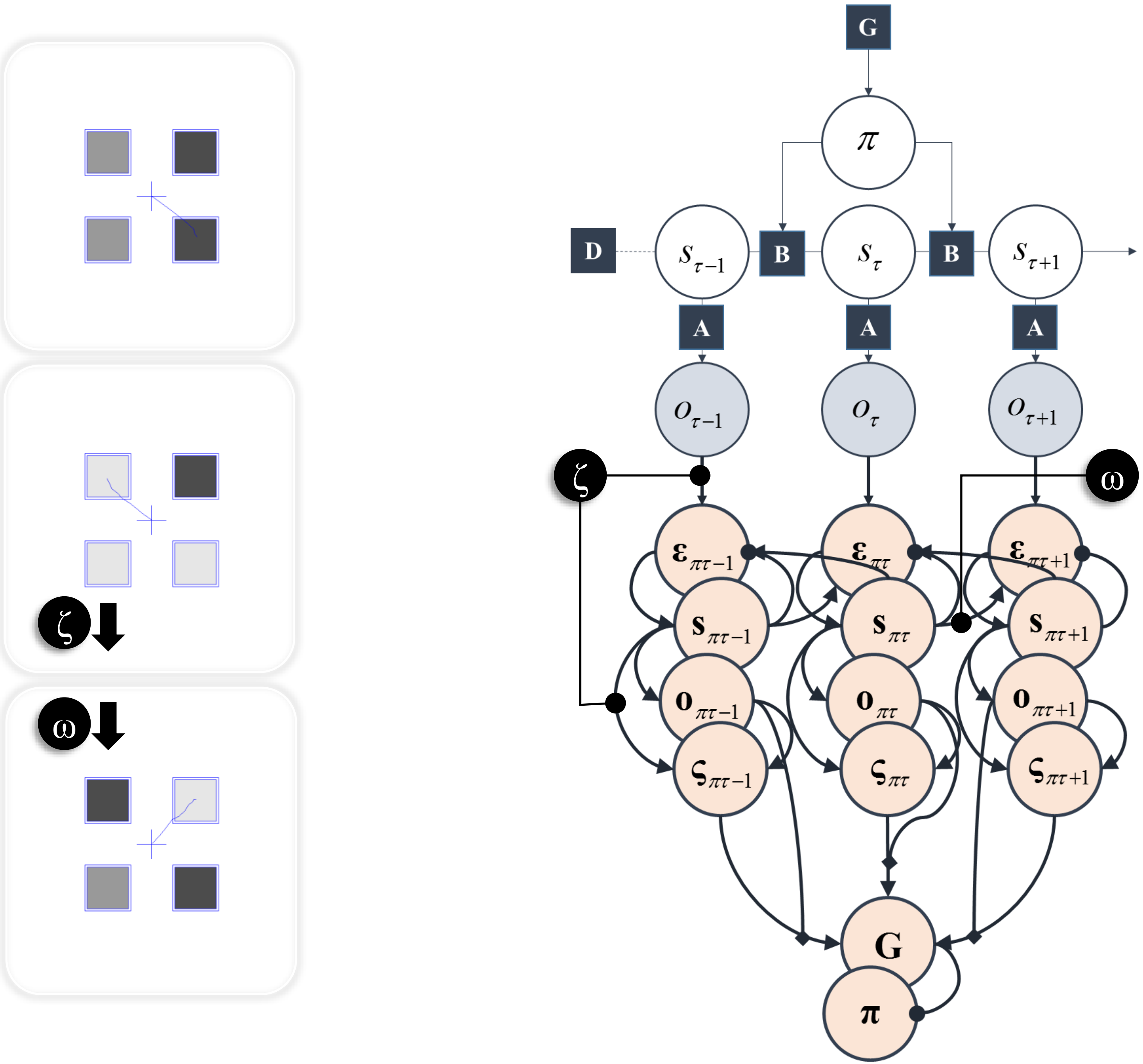
$$P(\tilde{o},\tilde{s},\pi) = P(s_1)P(\pi)\prod_\tau P(s_{\tau+1} \mid s_\tau,\pi)P(o_\tau \mid s_\tau)$$

$$P(o_\tau \mid s_\tau) = Cat(\mathbf{A})$$
$$P(s_\tau \mid s_{\tau-1},\pi) = Cat(\mathbf{B}_{\pi\tau})$$
$$P(o_\tau) = Cat(\mathbf{C})$$
$$P(s_1) = Cat(\mathbf{D})$$
$$P(\pi) = \sigma(-\mathbf{G})$$

Bayesian message passing

$$\mathbf{s}_\tau = \boldsymbol{\pi}\cdot\mathbf{s}_{\pi\tau}$$
$$\mathbf{s}_{\pi\tau} = \sigma(\mathbf{v}_{\pi\tau});\ \ \dot{\mathbf{v}}_{\pi\tau} = \boldsymbol{\varepsilon}_{\pi\tau}$$
$$\boldsymbol{\varepsilon}_{\pi\tau} = \ln\mathbf{A}\cdot o_\tau + \tfrac{1}{2}\ln(\mathbf{B}_{\pi\tau}\mathbf{s}_{\pi\tau}) + \tfrac{1}{2}\ln(\mathbf{B}^\dagger_{\pi\tau+1}\mathbf{s}_{\pi\tau+1}) - \ln\mathbf{s}_{\pi\tau}$$

$$\mathbf{o}_{\pi\tau} = \mathbf{A}\mathbf{s}_{\pi\tau}$$
$$\varsigma_{\pi\tau} = \ln\mathbf{o}_{\pi\tau} - \ln\mathbf{C} + \mathbf{H}\cdot\mathbf{s}_{\pi\tau}$$
$$\mathbf{H} = -diag(\mathbf{A}\cdot\ln\mathbf{A})$$
$$\mathbf{G}_\pi = \mathbf{o}_{\pi\tau}\cdot\varsigma_{\pi\tau}$$
$$\boldsymbol{\pi} = \sigma(-\mathbf{G})$$



Generative model

$$P(\tilde{o},\tilde{s},\pi) = P(s_1)P(\pi)\prod_{\tau} P(s_{\tau+1} \mid s_\tau,\pi)P(o_\tau \mid s_\tau)$$

$$P(o_\tau \mid s_\tau) = Cat(\mathbf{A})$$
$$P(s_\tau \mid s_{\tau-1},\pi) = Cat(\mathbf{B}_{\pi\tau})$$
$$P(o_\tau) = Cat(\mathbf{C})$$
$$P(s_1) = Cat(\mathbf{D})$$
$$P(\pi) = \sigma(-\mathbf{G})$$

Bayesian message passing

$$\mathbf{s}_\tau = \boldsymbol{\pi}\cdot\mathbf{s}_{\pi\tau}$$
$$\mathbf{s}_{\pi\tau} = \sigma(\mathbf{v}_{\pi\tau});\ \ \dot{\mathbf{v}}_{\pi\tau} = \boldsymbol{\varepsilon}_{\pi\tau}$$
$$\boldsymbol{\varepsilon}_{\pi\tau} = \ln\mathbf{A}\cdot o_\tau + \tfrac{1}{2}\ln(\mathbf{B}_{\pi\tau}\mathbf{s}_{\pi\tau}) + \tfrac{1}{2}\ln(\mathbf{B}^{\dagger}_{\pi\tau+1}\mathbf{s}_{\pi\tau+1}) - \ln\mathbf{s}_{\pi\tau}$$

$$\mathbf{o}_{\pi\tau} = \mathbf{A}\mathbf{s}_{\pi\tau}$$
$$\boldsymbol{\varsigma}_{\pi\tau} = \ln\mathbf{o}_{\pi\tau} - \ln\mathbf{C} + \mathbf{H}\cdot\mathbf{s}_{\pi\tau}$$
$$\mathbf{H} = -diag(\mathbf{A}\cdot\ln\mathbf{A})$$
$$\mathbf{G}_\pi = \mathbf{o}_{\pi\tau}\cdot\boldsymbol{\varsigma}_{\pi\tau}$$
$$\boldsymbol{\pi} = \sigma(-\mathbf{G})$$

Expected free energy

Explore

$$H\left[P(o \mid \pi)\right] - \mathrm{E}_{P(s|\pi)}\left[H\left[P(o \mid s)\right]\right]$$

$$\mathrm{E}_{P(o|\pi)}\left[\ln P(o \mid C)\right] + H\left[P(o \mid \pi)\right] - \mathrm{E}_{P(s|\pi)}\left[H\left[P(o \mid s)\right]\right]$$

$$\mathrm{E}_{P(o|\pi)}\left[\ln P(o \mid C)\right] + H\left[P(o \mid \pi)\right]$$

Exploit

Expected free energy

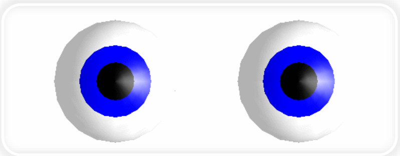
$$P(\pi) = \sigma[-G(\pi)]$$

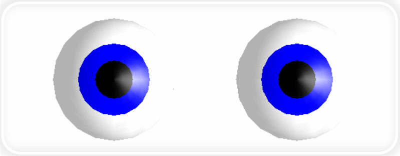
$$-G(\pi) = \mathrm{E}_{P(o|\pi)}\left[\ln P(o|C)\right] + H\left[P(o|\pi)\right] - \mathrm{E}_{P(s|\pi)}\left[H\left[P(o|s)\right]\right]$$

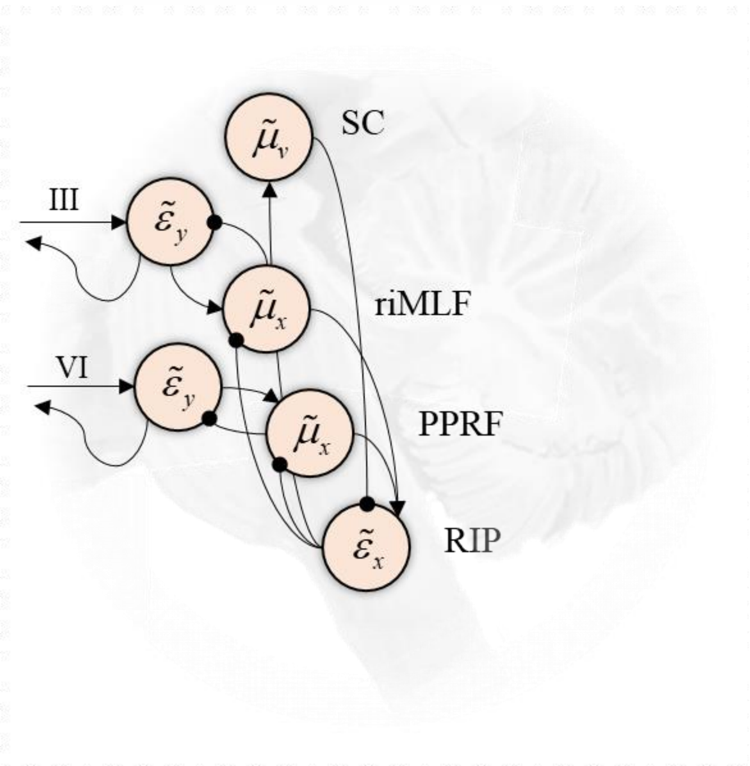
$$= \mathrm{E}_{P(o|\pi)}\left[\ln P(o|C)\right] + \mathrm{E}_{P(o,s|\pi)}\left[\ln P(s|\pi,o) - \ln P(s|\pi)\right]$$

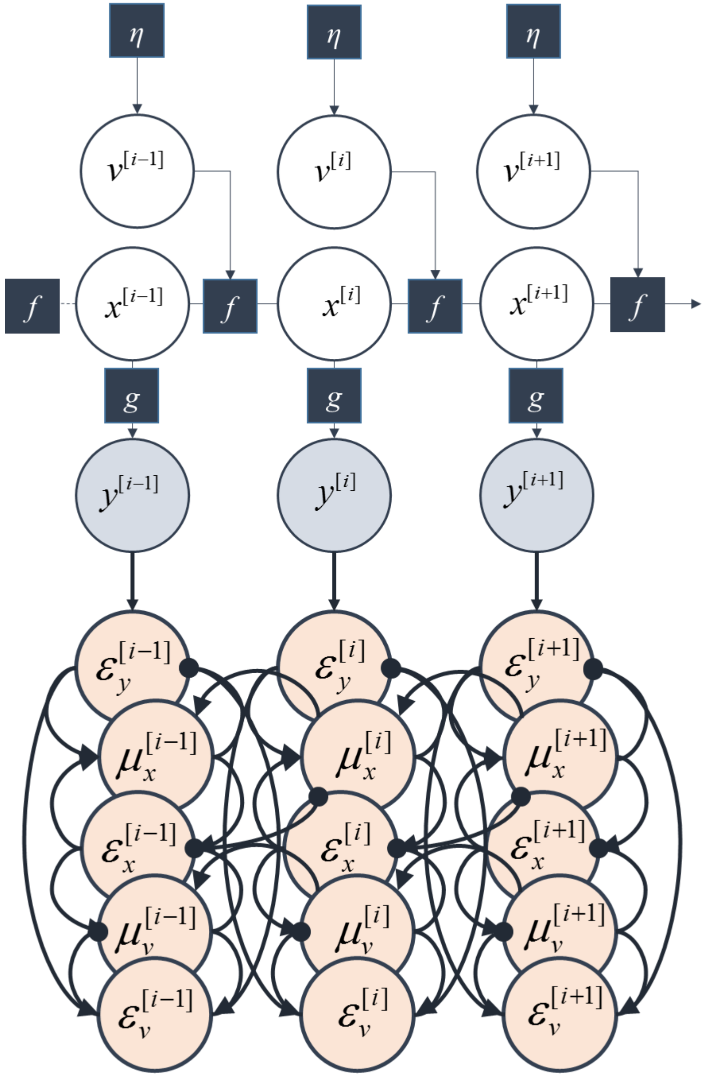
$$-F(\pi) = \ln P(o) + \mathrm{E}_{Q(s|\pi)}\left[\ln P(s|\pi,o) - \ln Q(s|\pi)\right]$$

Active inference
Generative models
Exploitation
Exploration
Movement
Hierarchy









Generative model

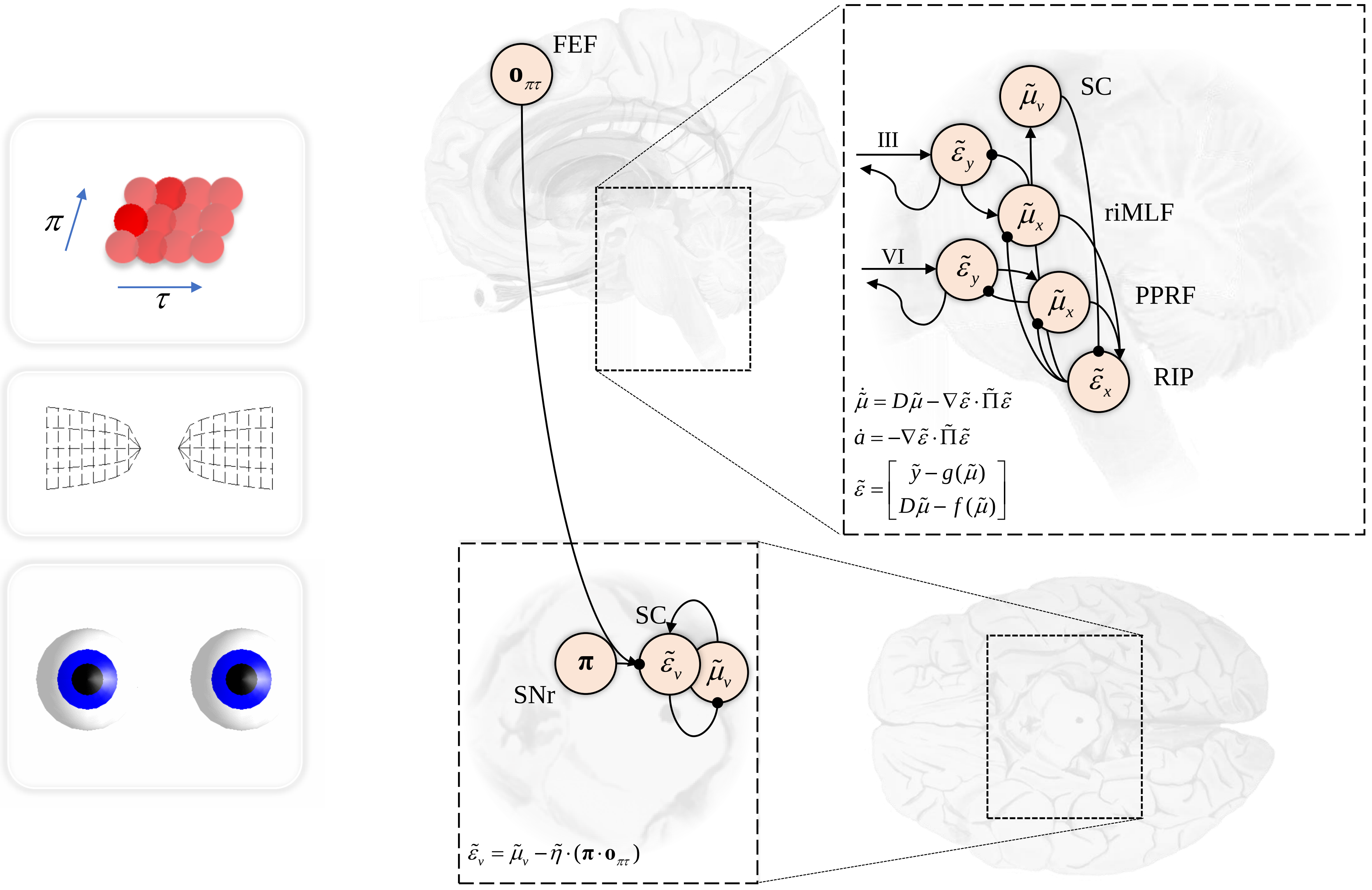
$$p(\tilde{y},\tilde{x},\tilde{v})=\prod_i p(v^{[i]})\,p(x^{[i+1]}\mid x^{[i]},v^{[i]})\,p(y^{[i]}\mid x^{[i]},v^{[i]})$$

$$p(y^{[i]}\mid x^{[i]},v^{[i]})=\mathcal{N}(g^{[i]}(x^{[i]},v^{[i]}),\Pi_y^{[i]})$$
$$p(x^{[i+1]}\mid x^{[i]},v^{[i]})=\mathcal{N}(f^{[i]}(x^{[i]},v^{[i]}),\Pi_x^{[i]})$$
$$p(v^{[i]})=\mathcal{N}(\eta^{[i]},\Pi_v^{[i]})$$

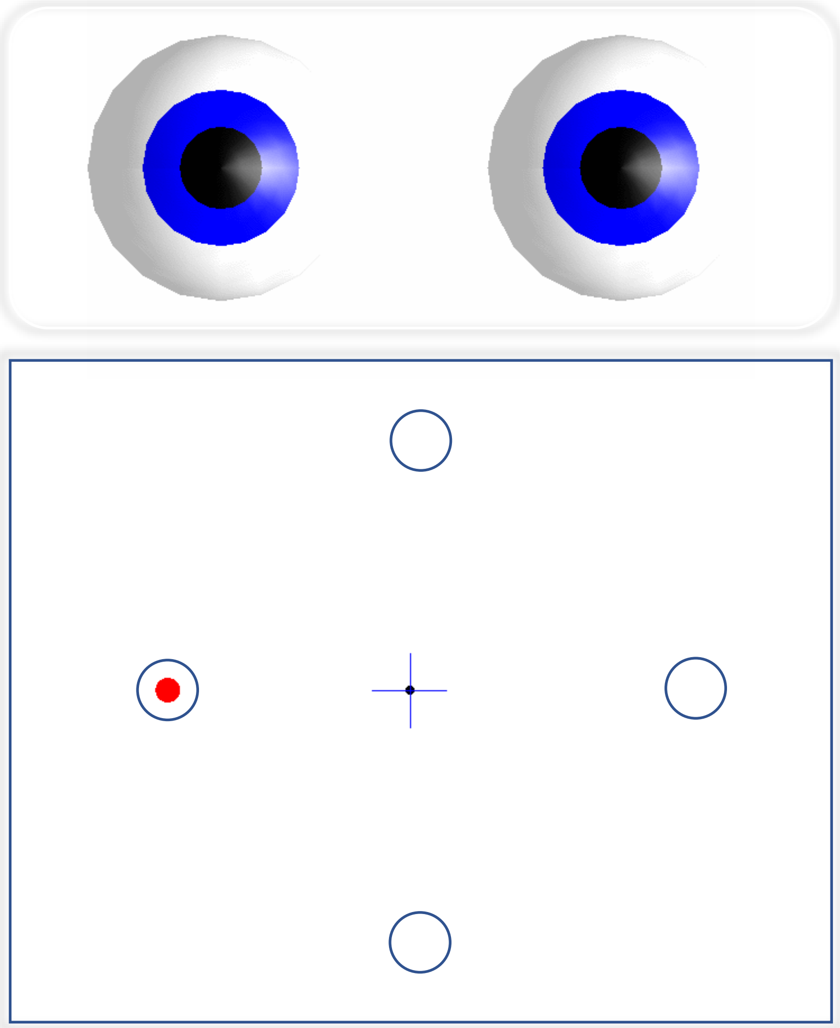
Bayesian message passing

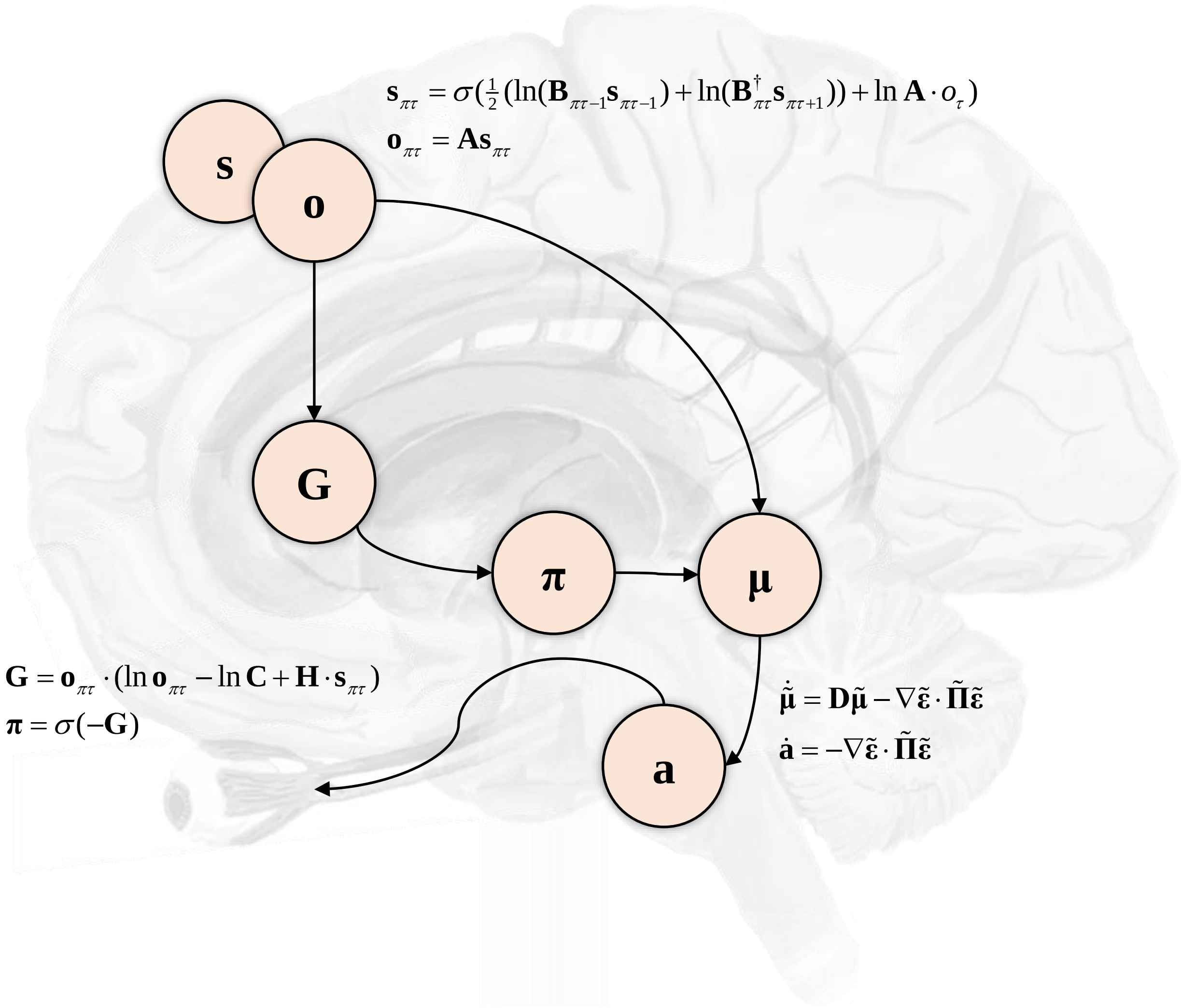
$$\varepsilon_y^{[i]}=y^{[i]}-g^{[i]}(\mu_x^{[i]},\mu_v^{[i]})$$
$$\varepsilon_x^{[i]}=\mu_x^{[i+1]}-f^{[i]}(\mu_x^{[i]},\mu_v^{[i]})$$
$$\varepsilon_v^{[i]}=\mu_v^{[i]}-\eta^{[i]}$$

$$\begin{aligned}\dot{\mu}_x^{[i]}&=\mu_x^{[i+1]}\\&\quad+\partial_{\mu_x^{[i]}}g^{[i]}\cdot\Pi_y^{[i]}\varepsilon_y^{[i]}-\Pi_x^{[i-1]}\varepsilon_x^{[i-1]}+\partial_{\mu_x^{[i]}}f^{[i]}\cdot\Pi_x^{[i]}\varepsilon_x^{[i]}\\\dot{\mu}_v^{[i]}&=\mu_v^{[i+1]}\\&\quad+\partial_{\mu_v^{[i]}}g^{[i]}\cdot\Pi_y^{[i]}\varepsilon_y^{[i]}+\partial_{\mu_v^{[i]}}f^{[i]}\cdot\Pi_x^{[i]}\varepsilon_x^{[i]}-\Pi_v^{[i]}\varepsilon_v^{[i]}\end{aligned}$$

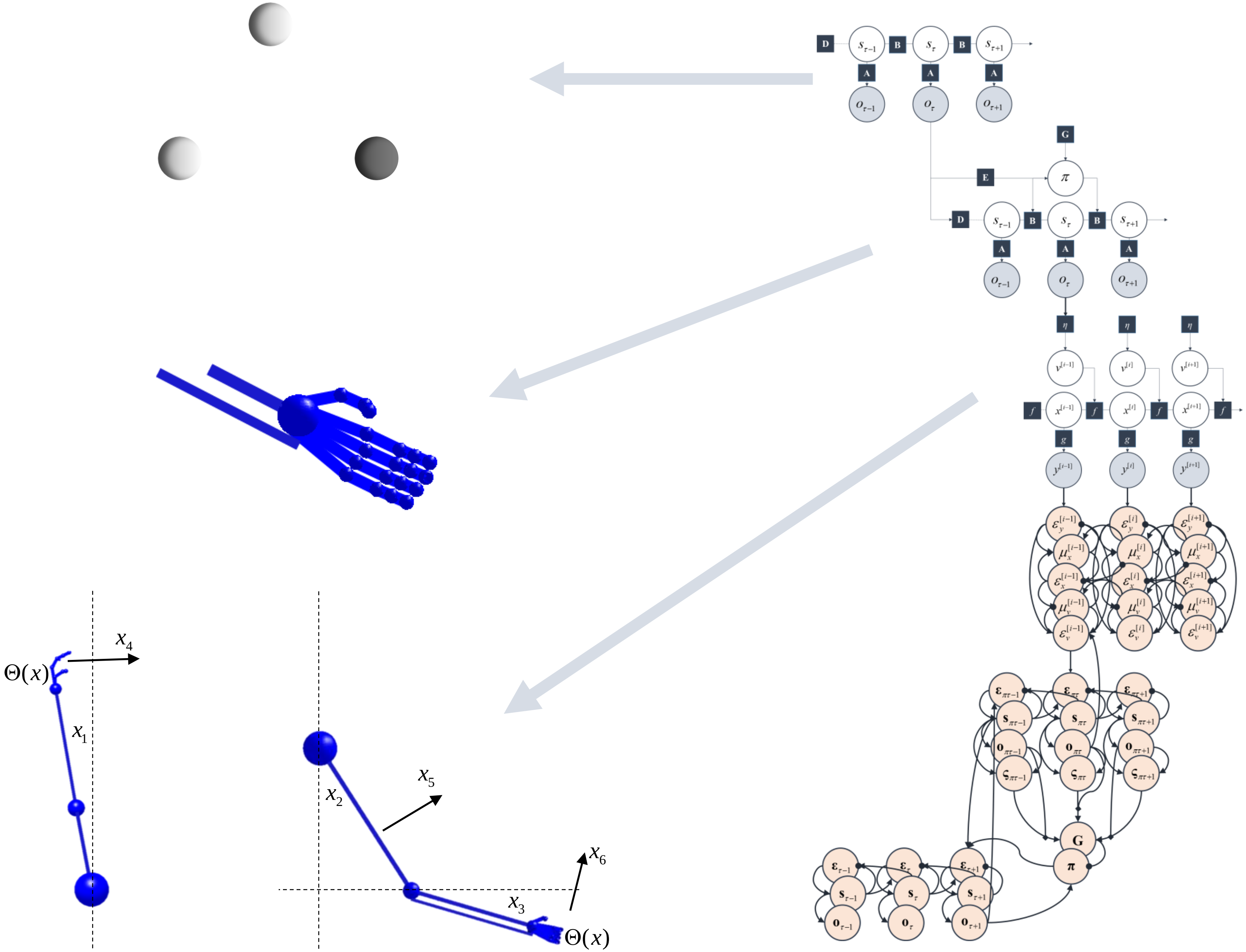


π
τ
FEF
$\mathbf{o}_{\pi\tau}$
SC
III
VI
riMLF
PPRF
RIP
$\tilde{\mu}_v$
$\tilde{\varepsilon}_y$
$\tilde{\mu}_x$
$\tilde{\varepsilon}_x$
$\dot{\tilde{\mu}} = D\tilde{\mu} - \nabla\tilde{\varepsilon} \cdot \tilde{\Pi}\tilde{\varepsilon}$
$\dot{a} = -\nabla\tilde{\varepsilon} \cdot \tilde{\Pi}\tilde{\varepsilon}$
$\tilde{\varepsilon} = \begin{bmatrix} \tilde{y} - g(\tilde{\mu}) \\ D\tilde{\mu} - f(\tilde{\mu}) \end{bmatrix}$
SNr
$\boldsymbol{\pi}$
$\tilde{\varepsilon}_v$
$\tilde{\varepsilon}_v = \tilde{\mu}_v - \tilde{\eta} \cdot (\boldsymbol{\pi} \cdot \mathbf{o}_{\pi\tau})$

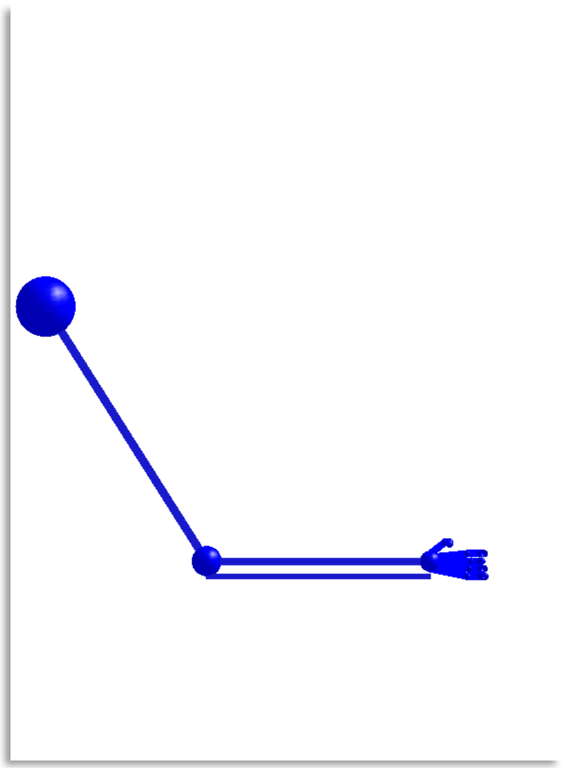


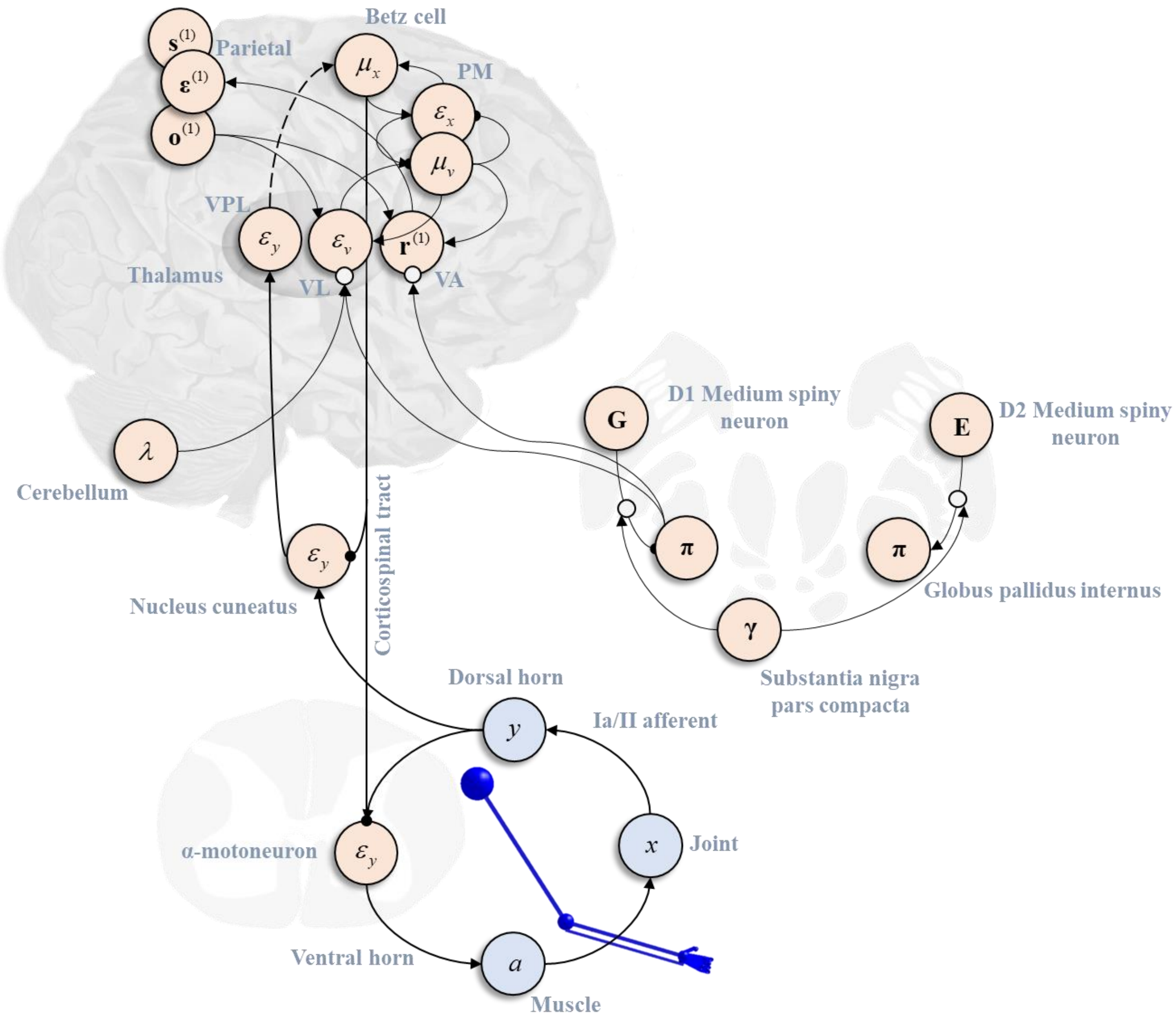

$\mathbf{s}_{\pi\tau} = \sigma(\tfrac{1}{2}(\ln(\mathbf{B}_{\pi\tau-1}\mathbf{s}_{\pi\tau-1}) + \ln(\mathbf{B}^{\dagger}_{\pi\tau}\mathbf{s}_{\pi\tau+1})) + \ln \mathbf{A}\cdot o_{\tau})$
$\mathbf{o}_{\pi\tau} = \mathbf{A}\mathbf{s}_{\pi\tau}$
s
o
G
$\boldsymbol{\pi}$
$\boldsymbol{\mu}$
a
$\mathbf{G} = \mathbf{o}_{\pi\tau}\cdot(\ln \mathbf{o}_{\pi\tau} - \ln \mathbf{C} + \mathbf{H}\cdot \mathbf{s}_{\pi\tau})$
$\boldsymbol{\pi} = \sigma(-\mathbf{G})$
$\dot{\tilde{\boldsymbol{\mu}}} = \mathbf{D}\tilde{\boldsymbol{\mu}} - \nabla\tilde{\boldsymbol{\varepsilon}}\cdot\tilde{\boldsymbol{\Pi}}\tilde{\boldsymbol{\varepsilon}}$
$\dot{\mathbf{a}} = -\nabla\tilde{\boldsymbol{\varepsilon}}\cdot\tilde{\boldsymbol{\Pi}}\tilde{\boldsymbol{\varepsilon}}$

Active inference
Generative models
Exploitation
Exploration
Movement
Hierarchy

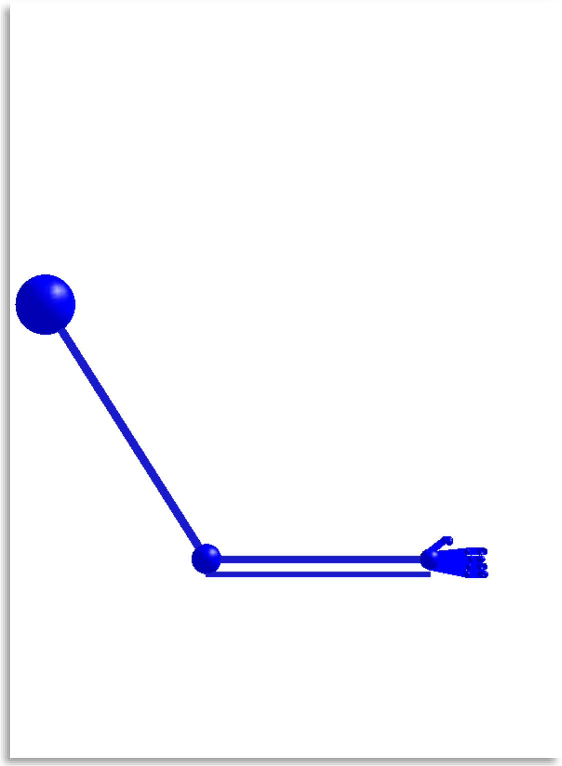


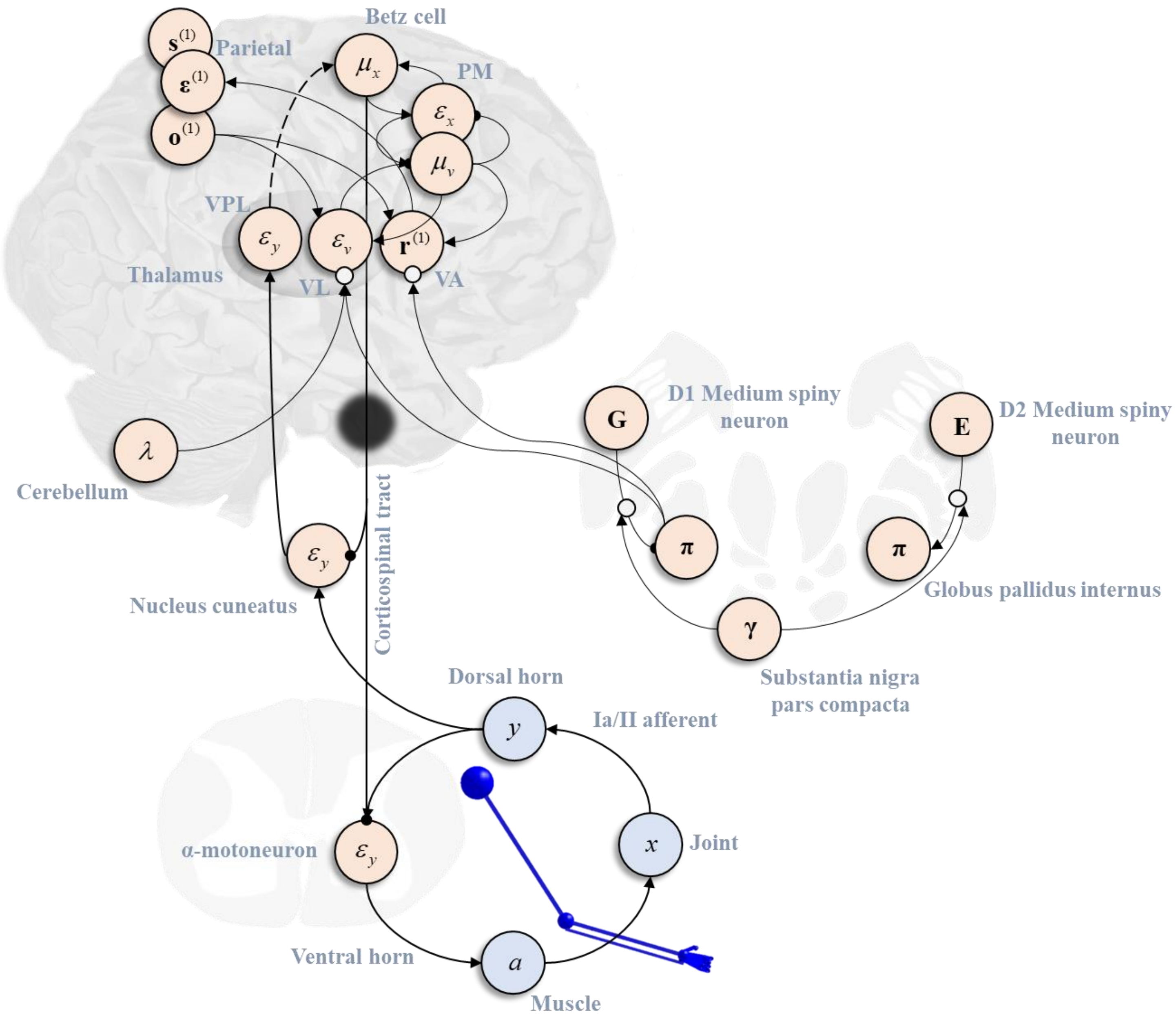
$\Theta(x)$
x_4
x_1
x_2
x_5
x_3
x_6
$\Theta(x)$



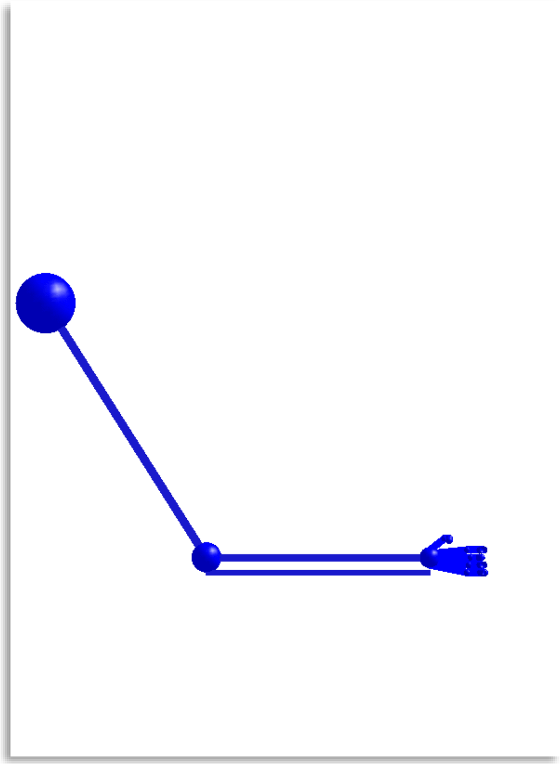


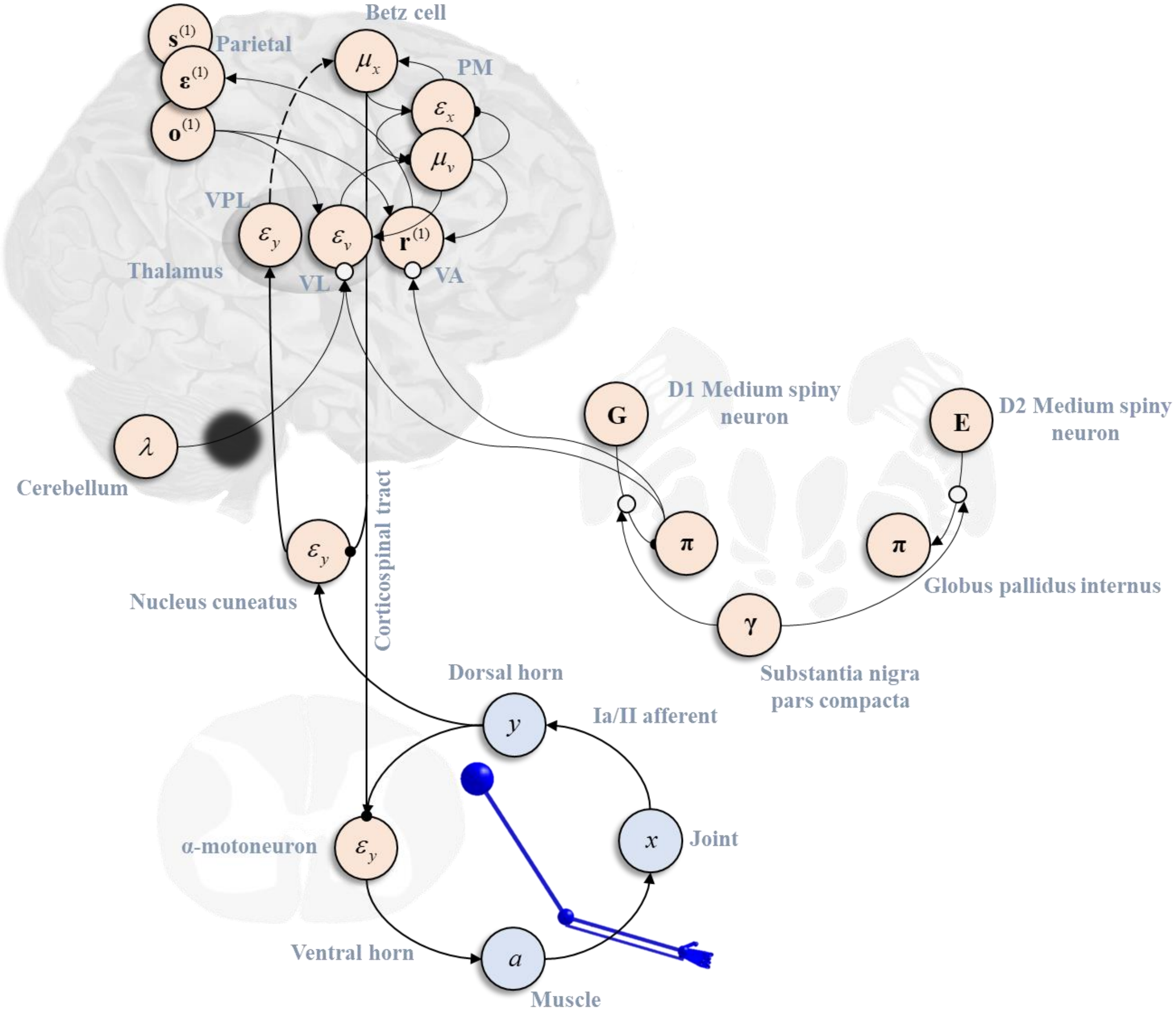
Betz cell
Parietal
PM
VPL
Thalamus
VL
VA
D1 Medium spiny neuron
D2 Medium spiny neuron
Cerebellum
Corticospinal tract
Globus pallidus internus
Nucleus cuneatus
Dorsal horn
Substantia nigra pars compacta
Ia/II afferent
Joint
α-motoneuron
Ventral horn
Muscle



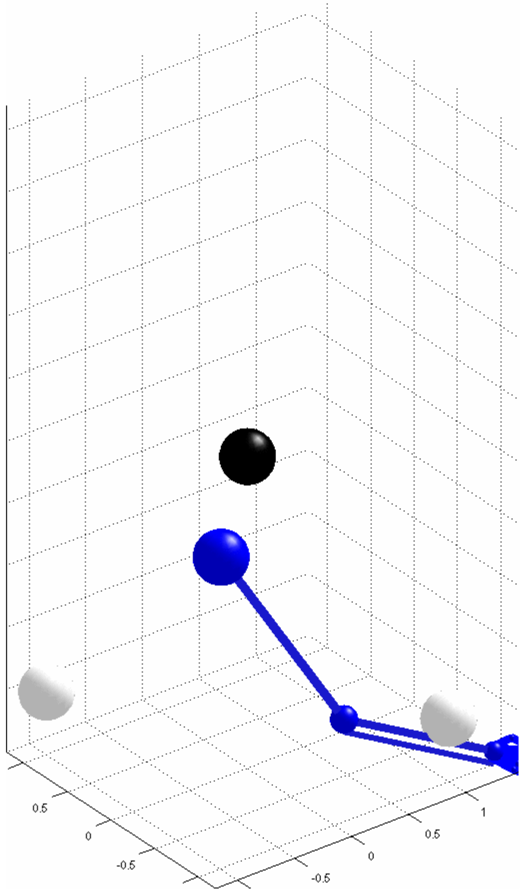


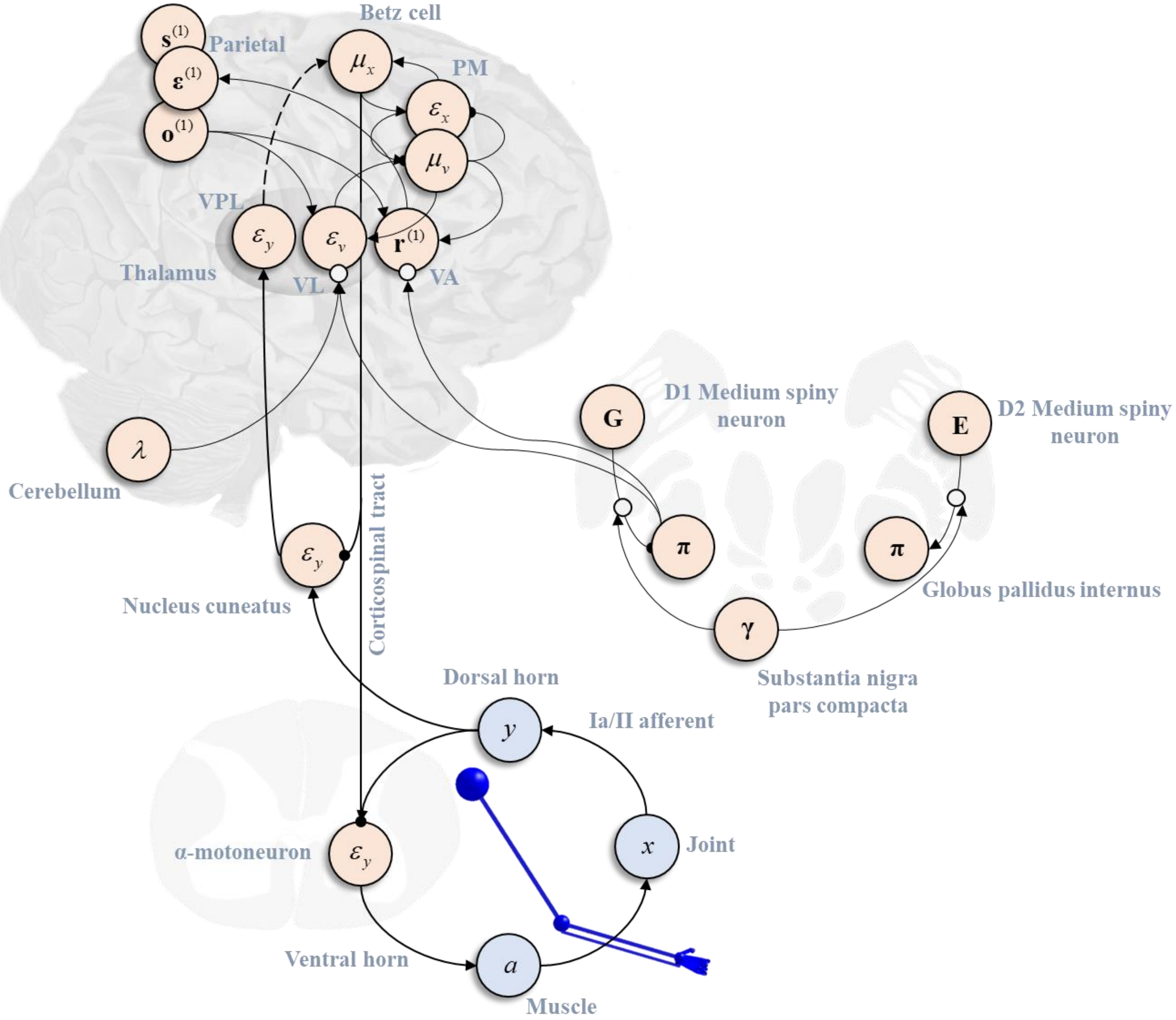
Betz cell
Parietal
PM
VPL
Thalamus
VL
VA
D1 Medium spiny neuron
D2 Medium spiny neuron
Cerebellum
Corticospinal tract
Nucleus cuneatus
Globus pallidus internus
Substantia nigra pars compacta
Dorsal horn
Ia/II afferent
Joint
α-motoneuron
Ventral horn
Muscle



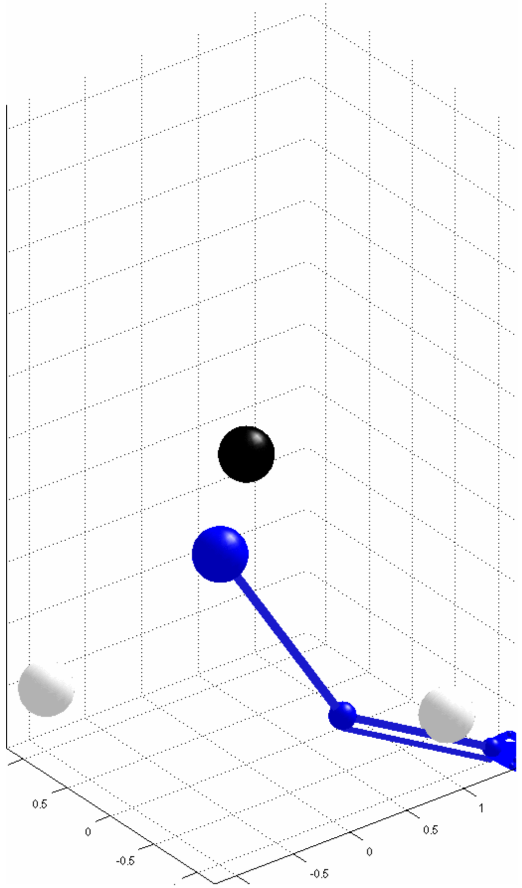


Betz cell
Parietal
PM
VPL
Thalamus
VL
VA
Cerebellum
D1 Medium spiny neuron
D2 Medium spiny neuron
Globus pallidus internus
Nucleus cuneatus
Corticospinal tract
Substantia nigra pars compacta
Dorsal horn
Ia/II afferent
Joint
α-motoneuron
Ventral horn
Muscle

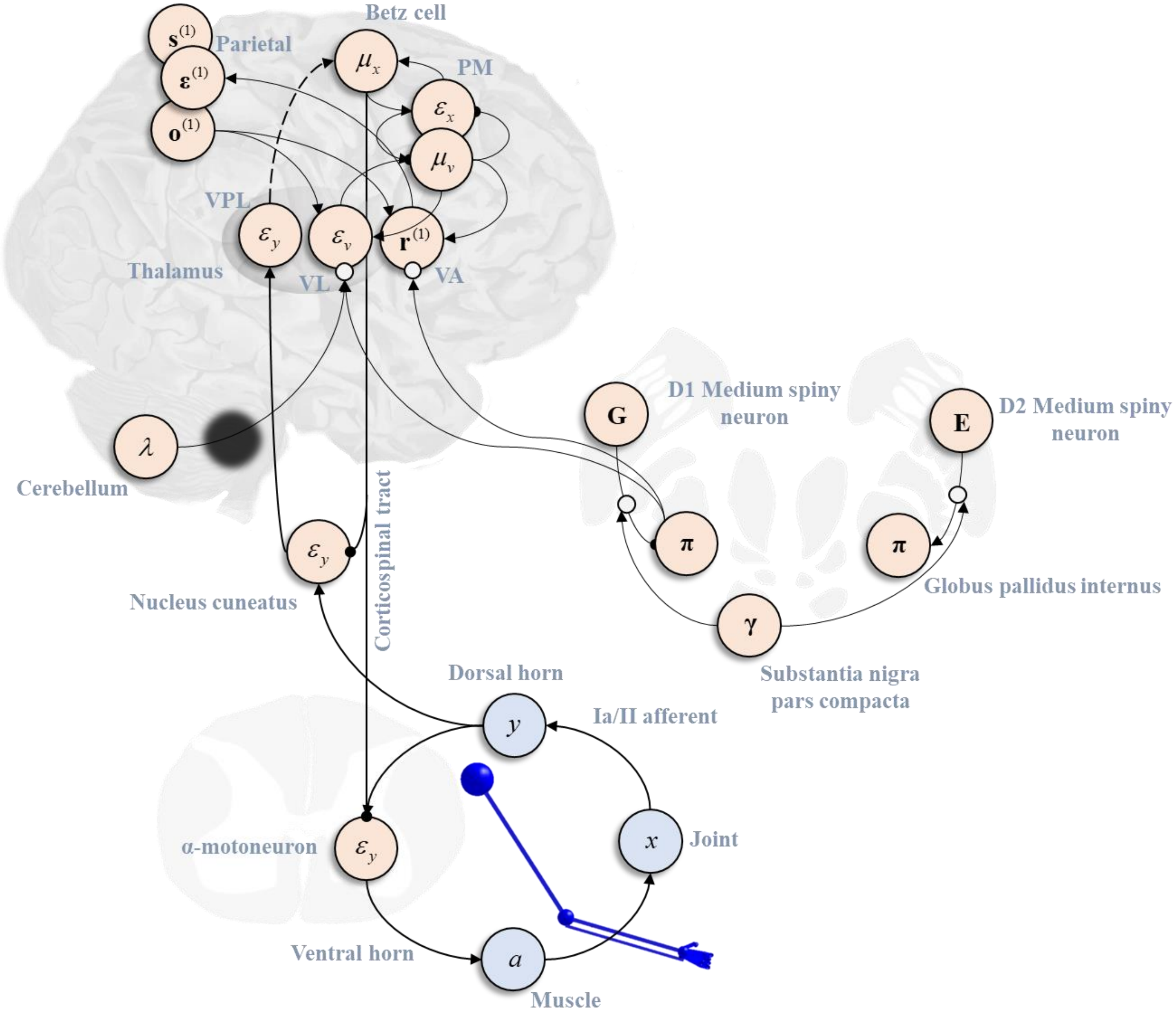




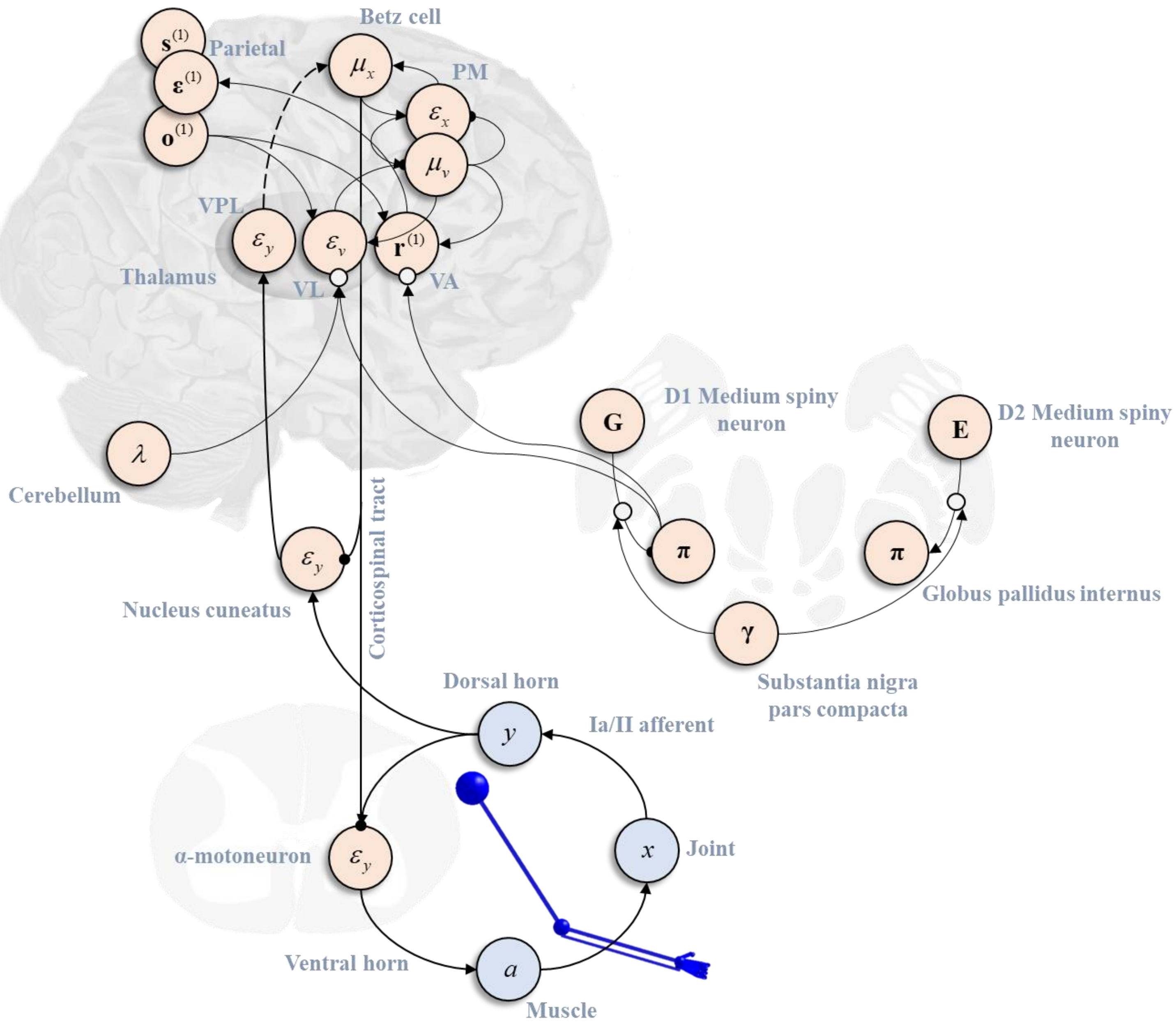
Betz cell
Parietal
PM
VPL
Thalamus
VL
VA
D1 Medium spiny neuron
D2 Medium spiny neuron
Cerebellum
Corticospinal tract
Nucleus cuneatus
Globus pallidus internus
Substantia nigra pars compacta
Dorsal horn
Ia/II afferent
α-motoneuron
Joint
Ventral horn
Muscle



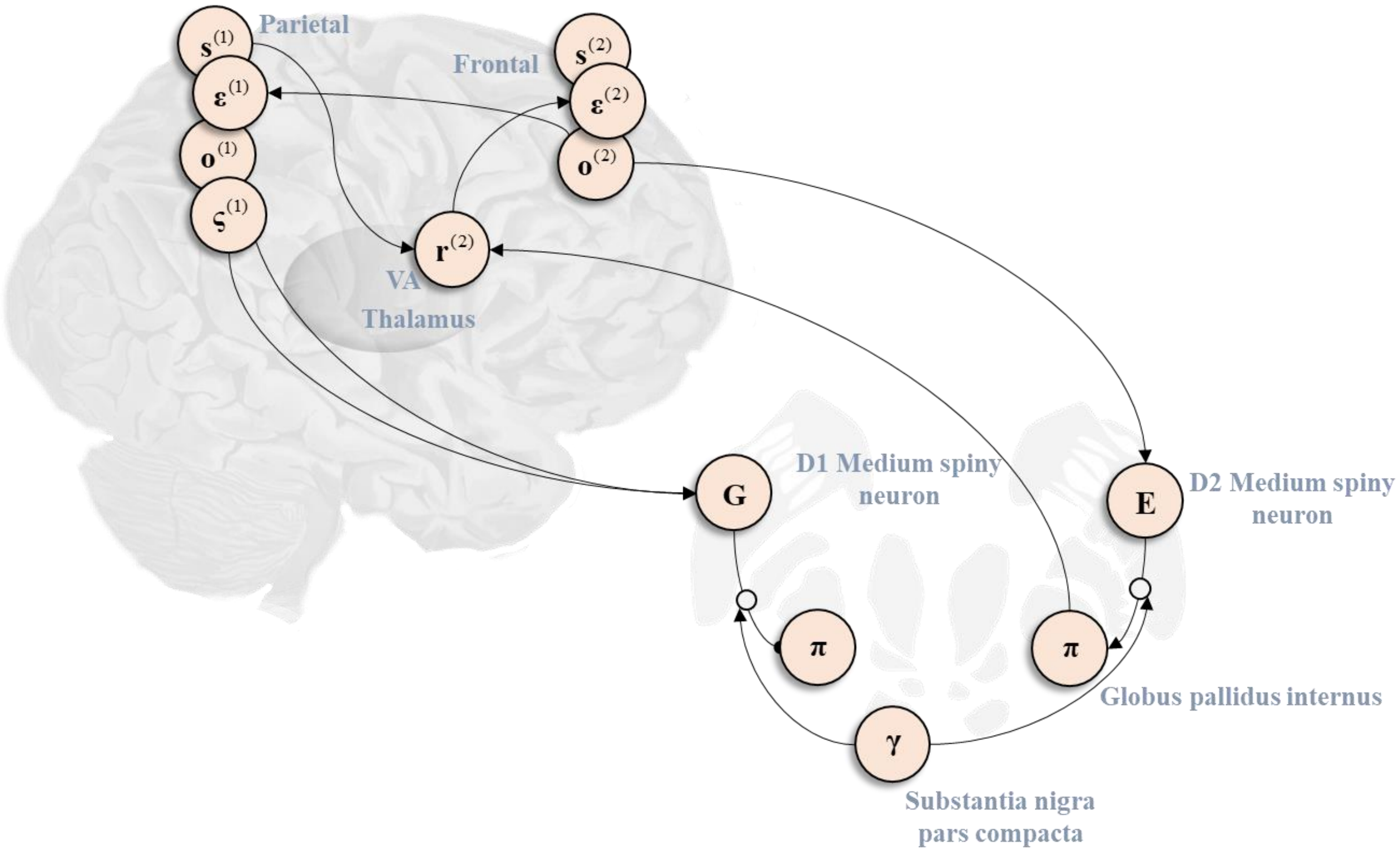
0.5
0
-0.5
-0.5
0
0.5
1



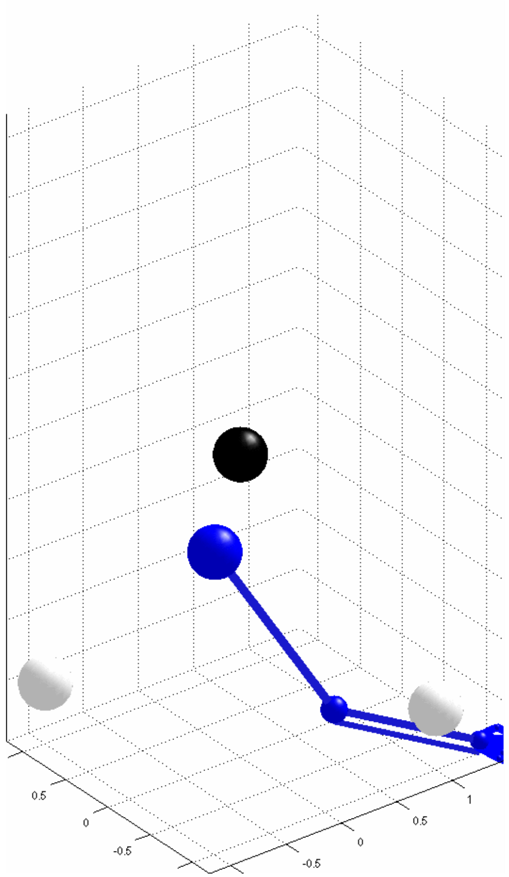
Betz cell
$\mathbf{s}^{(1)}$
Parietal
$\boldsymbol{\varepsilon}^{(1)}$
$\mathbf{o}^{(1)}$
μ_x
PM
ε_x
μ_v
VPL
ε_y
ε_v
$\mathbf{r}^{(1)}$
Thalamus
VL
VA
λ
Cerebellum
ε_y
Nucleus cuneatus
Corticospinal tract
G
D1 Medium spiny neuron
E
D2 Medium spiny neuron
$\boldsymbol{\pi}$
$\boldsymbol{\pi}$
Globus pallidus internus
$\boldsymbol{\gamma}$
Substantia nigra pars compacta
Dorsal horn
y
Ia/II afferent
α-motoneuron
ε_y
x
Joint
Ventral horn
a
Muscle

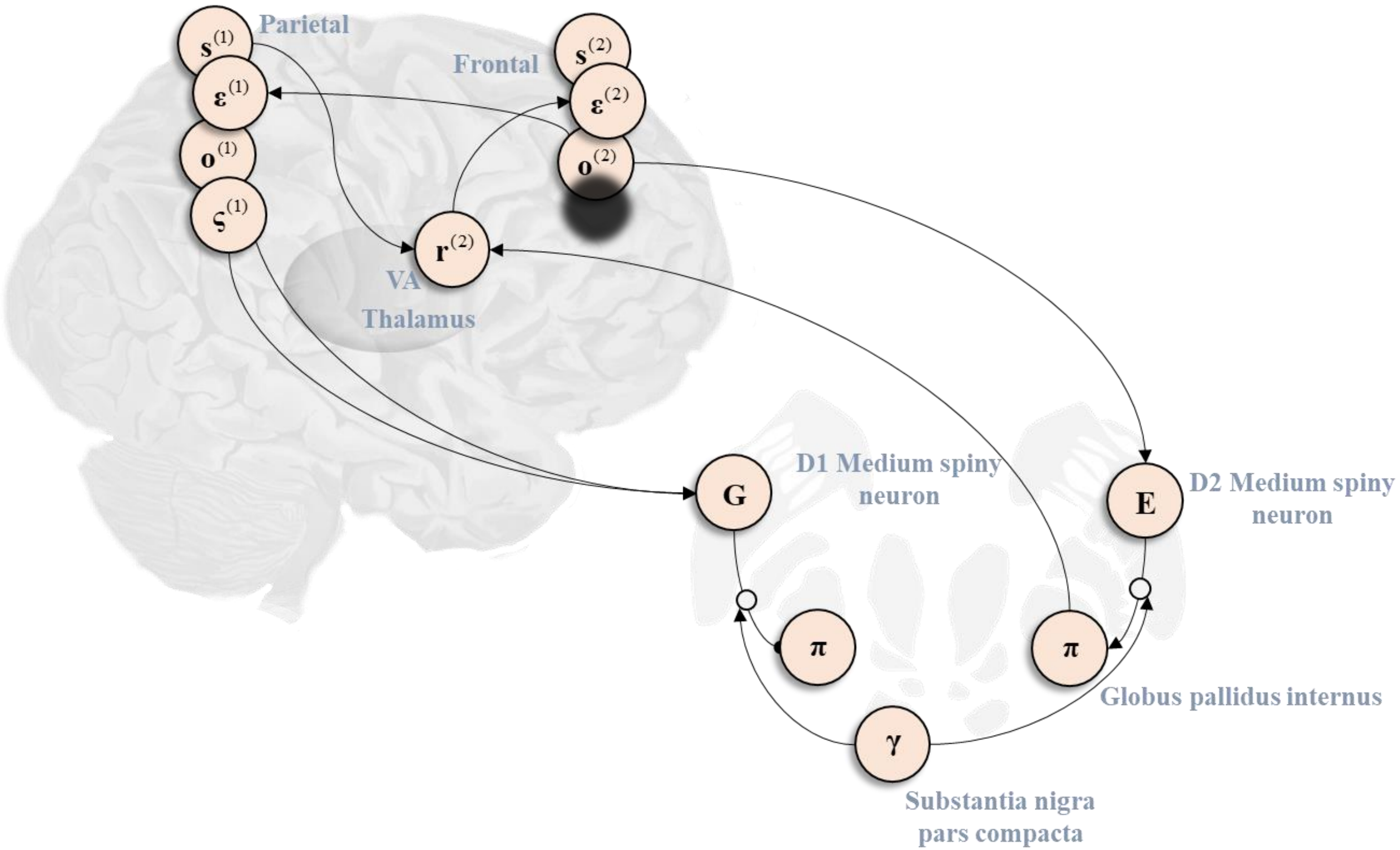


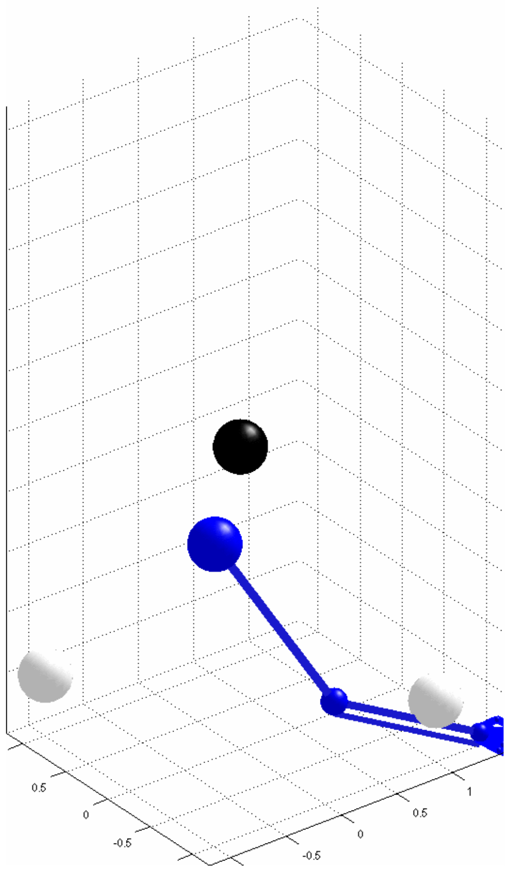
Betz cell
Parietal
PM
VPL
Thalamus
VL
VA
Cerebellum
Nucleus cuneatus
Corticospinal tract
D1 Medium spiny neuron
D2 Medium spiny neuron
Globus pallidus internus
Substantia nigra pars compacta
Dorsal horn
Ia/II afferent
Joint
α-motoneuron
Ventral horn
Muscle

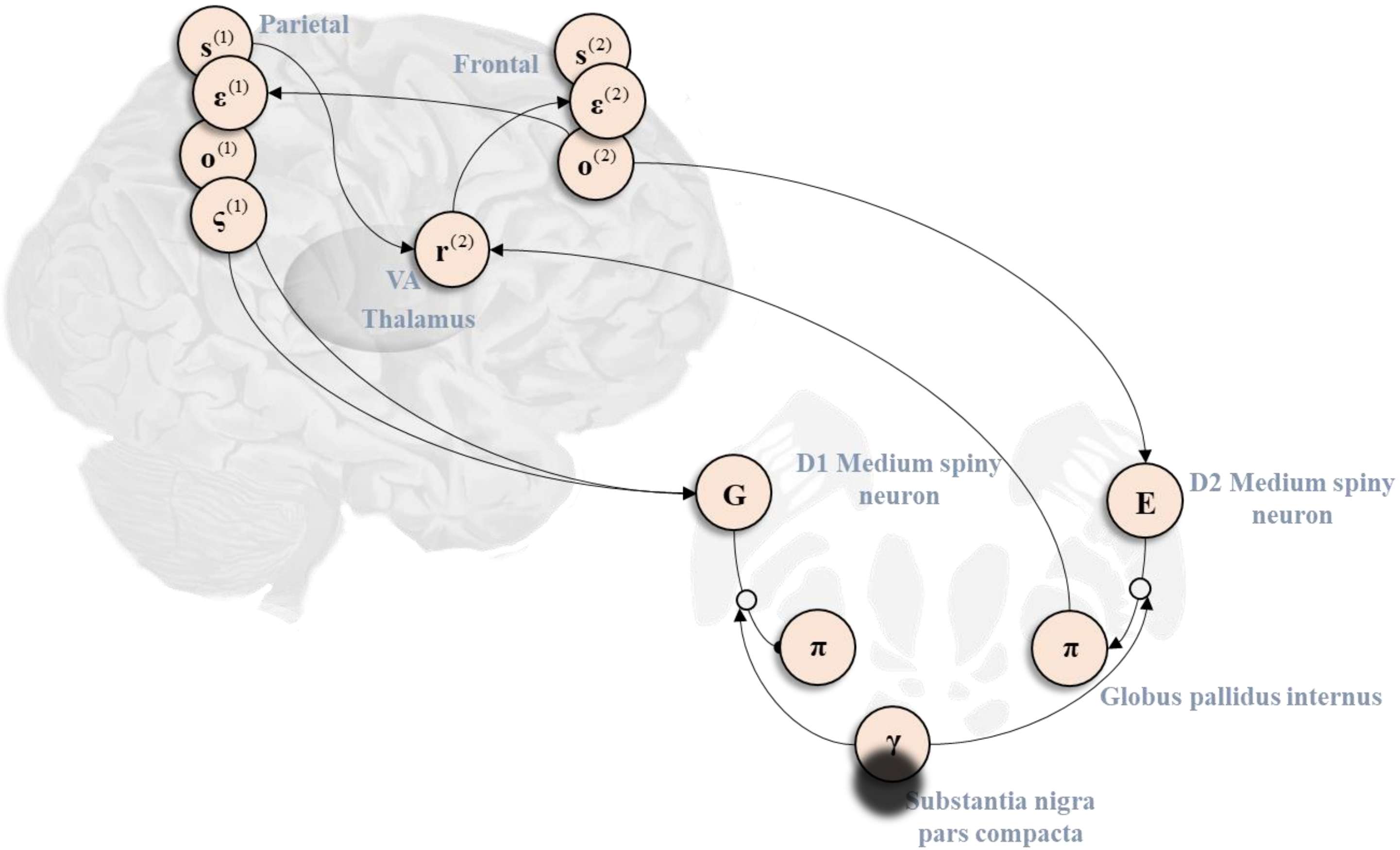


Parietal
Frontal
VA
Thalamus
D1 Medium spiny neuron
D2 Medium spiny neuron
Globus pallidus internus
Substantia nigra pars compacta
$s^{(1)}$
$\varepsilon^{(1)}$
$o^{(1)}$
$\varsigma^{(1)}$
$s^{(2)}$
$\varepsilon^{(2)}$
$o^{(2)}$
$r^{(2)}$
G
E
π
π
γ

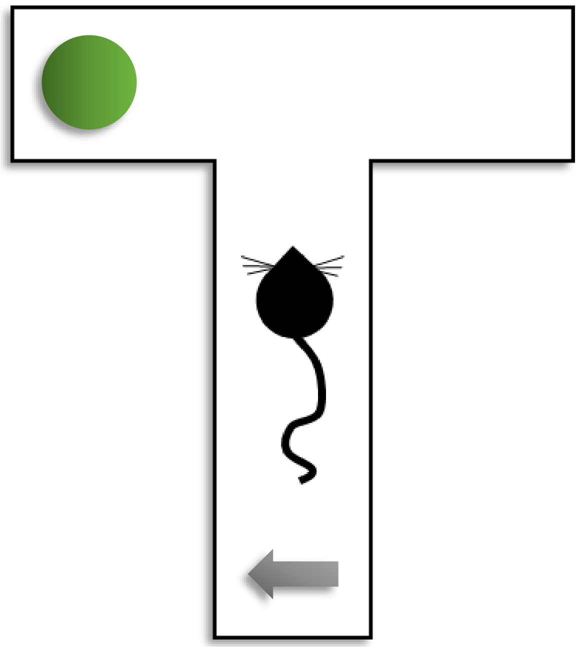

0.5
0
-0.5
-0.5
0
0.5
1

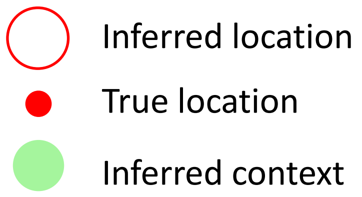

Parietal
Frontal
$s^{(1)}$
$\varepsilon^{(1)}$
$o^{(1)}$
$\varsigma^{(1)}$
$s^{(2)}$
$\varepsilon^{(2)}$
$o^{(2)}$
$r^{(2)}$
VA
Thalamus
G
D1 Medium spiny
neuron
E
D2 Medium spiny
neuron
π
π
Globus pallidus internus
γ
Substantia nigra
pars compacta

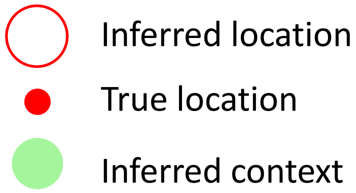

0.5
0
-0.5
-0.5
0
0.5
1

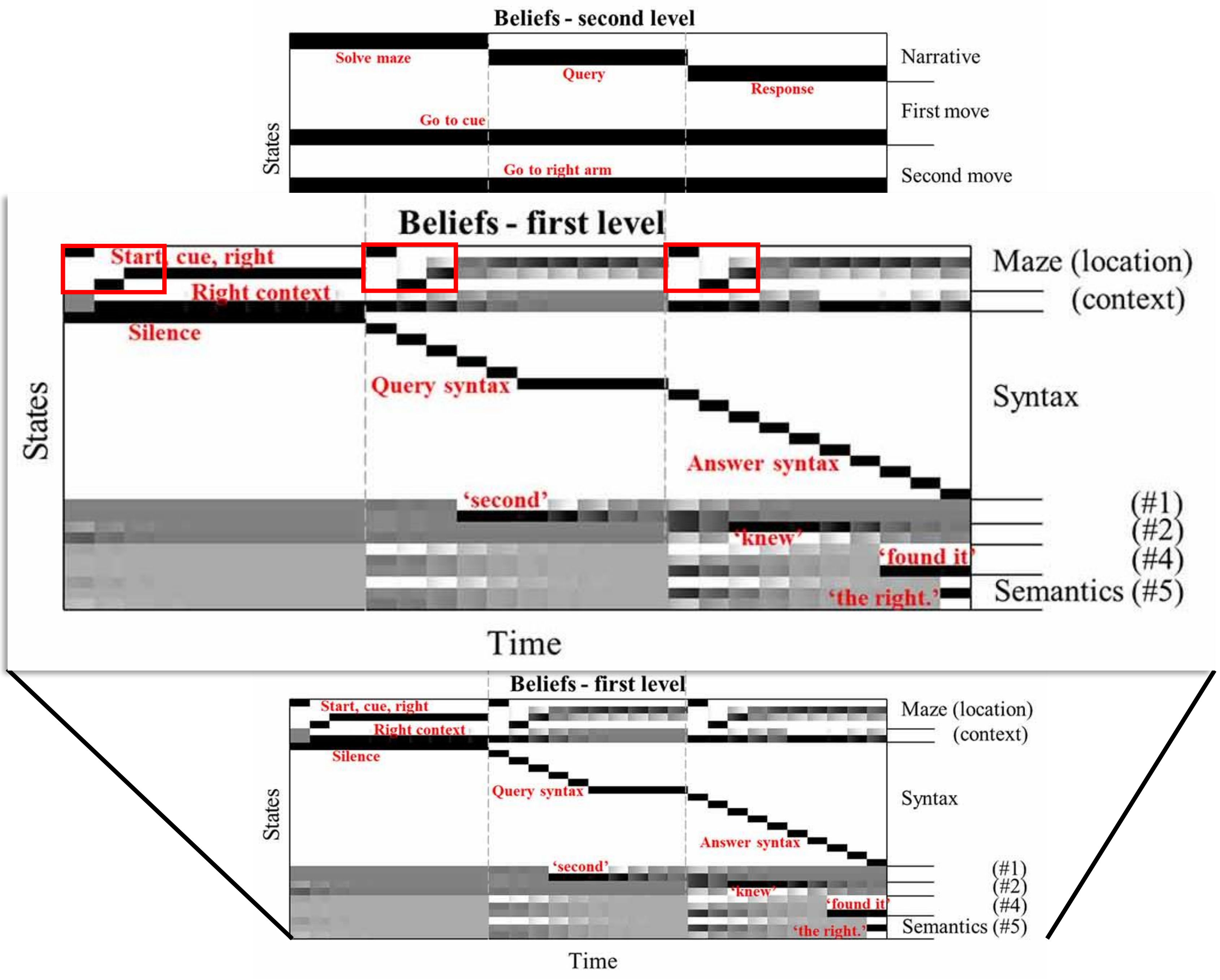

Parietal
Frontal
VA
Thalamus
D1 Medium spiny
neuron
D2 Medium spiny
neuron
Globus pallidus internus
Substantia nigra
pars compacta
G
E
π
π
γ

Navigating a T-maze



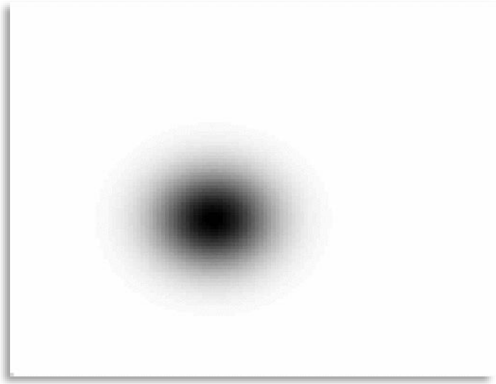

Inferred location
True location
Inferred context

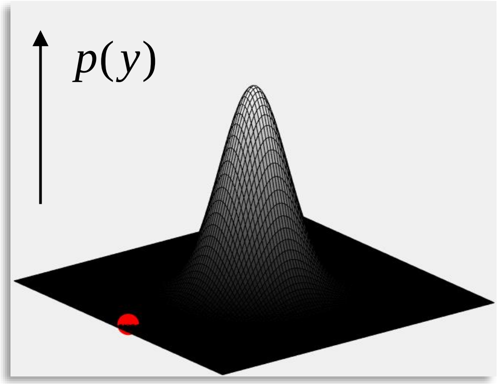

Inferred location
True location
Inferred context


Beliefs - second level
Solve maze
Query
Response
Go to cue
Go to right arm
Narrative
First move
Second move
States
Beliefs - first level
Start, cue, right
Right context
Silence
Query syntax
Answer syntax
'second'
'knew'
'found it'
'the right.'
Maze (location)
(context)
Syntax
(#1)
(#2)
(#4)
Semantics (#5)
States
Time
Beliefs - first level
Start, cue, right
Right context
Silence
Query syntax
Answer syntax
'second'
'knew'
'found it'
'the right.'
Maze (location)
(context)
Syntax
(#1)
(#2)
(#4)
Semantics (#5)
States
Time

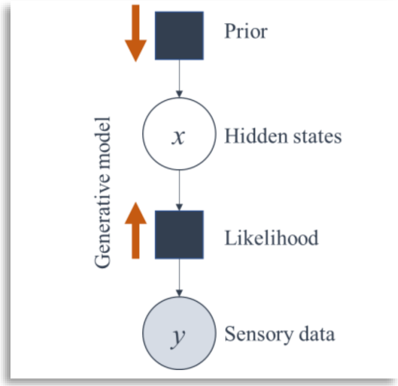
Inferred location
True location
Inferred context

Confabulation

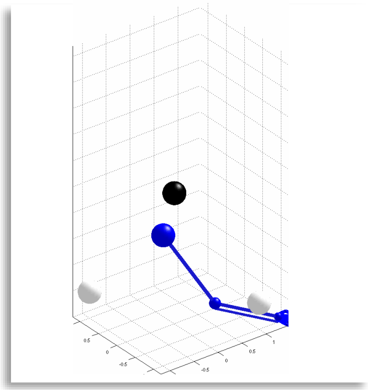




$p(y)$



Generative model
Prior
x
Hidden states
Likelihood
y
Sensory data



Thanks

Berk Mirza
David Benrimoh
Dimitrije Markovic
Emma Holmes
Giovanni Pezzulo
Jakub Limanowski
Jakob Hohwy
Jelle Bruineberg
Karl Friston
Lance Da Costa
Noor Sajid
Peter Vincent
Rick Adams
Stefan Kiebel
Vishal Rawji
And many others

Rosetrees Trust
Supporting the best in medical research

UCL

wellcome centre human neuroimaging