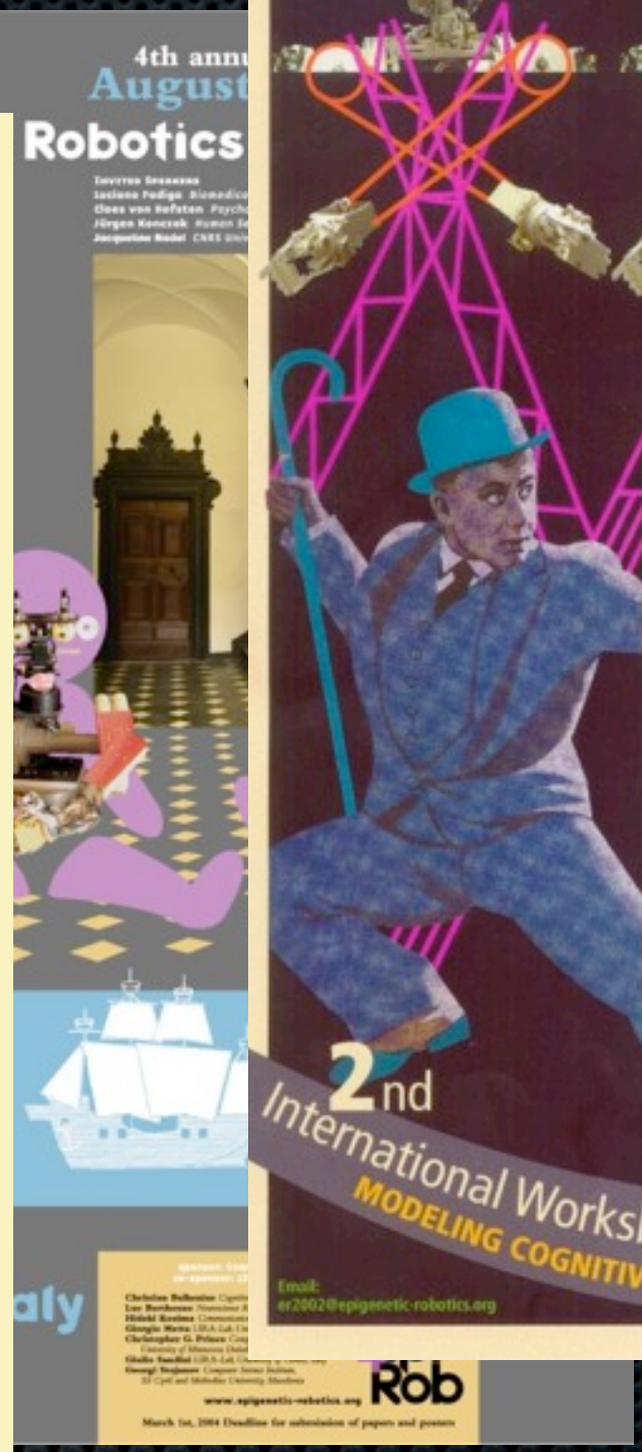
Learning Flexible Goal-Directed Behavior

Christian Balkenius

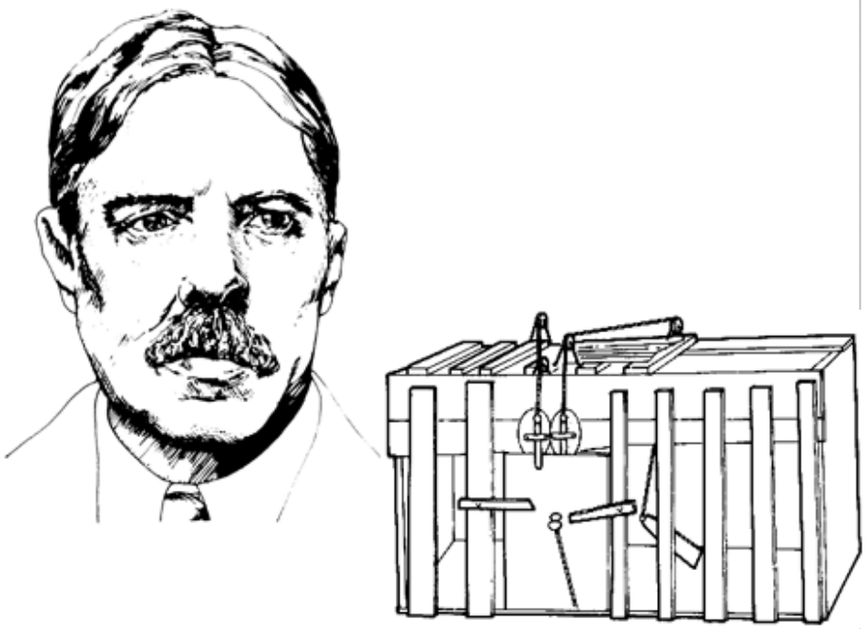
Lund University Cognitive Science


AUGUST 4 & 5 • 2003
3rd
ANNUAL
INTERNATIONAL
WORKSHOP
EPIGENETIC ROBOTICS
MODELING COGNITIVE DEVELOPMENT IN ROBOTIC SYSTEMS
SPEAKERS
GYÖRGY GERGELY
INSTITUTE FOR PSYCHOLOGICAL RESEARCH HUNGARIAN ACADEMY OF SCIENCES BUDAPEST, HUNGARY
ROD GRUPEN
LABORATORY FOR PERCEPTUAL ROBOTICS UNIVERSITY OF MASSACHUSETTS AMHERST, MA, USA
DEB ROY
MEDIA LAB MIT, USA
www.epigenetic-robotics.org
BOSTON UNIVERSITY BOSTON MA USA
BOSTON
EpiRob 2003 will be held after the Cognitive Science Society Meeting
Deadline for submission of papers and poster abstracts March 14, 2003

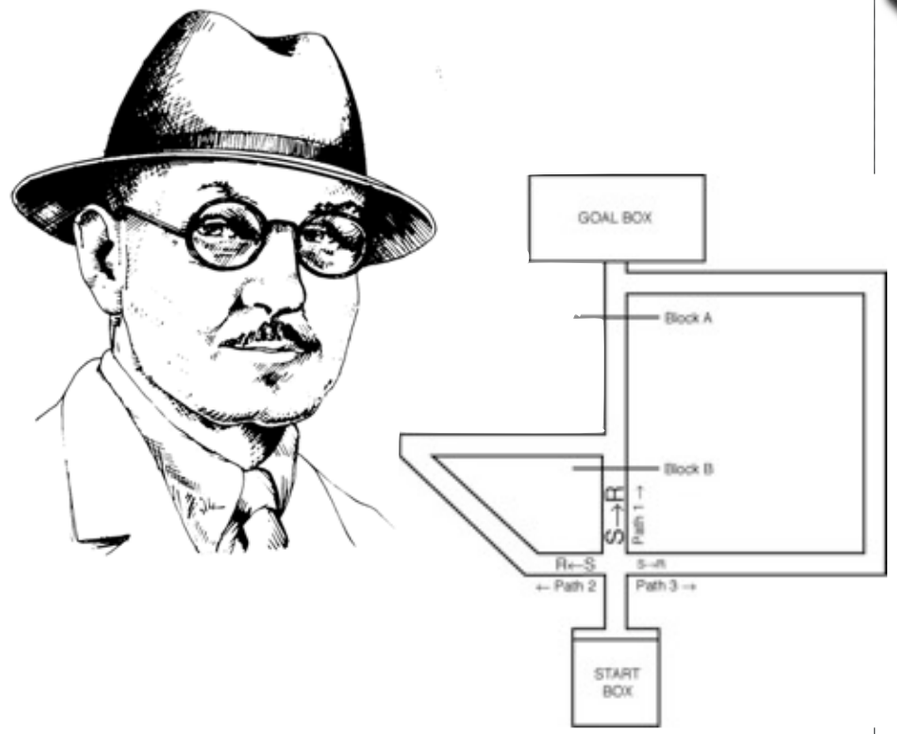

4th annu
August
Robotics
aly
Rob
March 1st, 2004 Deadline for submission of papers and posters


2002
August 10•11
University of
Edinburgh
Scotland
Luc Steels
Colwyn Trevarthen
John Weng
2nd
International Workshop on EPIGENETIC ROBOTICS
MODELING COGNITIVE DEVELOPMENT IN ROBOTIC SYSTEMS
Email:
er2002@epigenetic-robotics.org
www.epigenetic-robotics.org

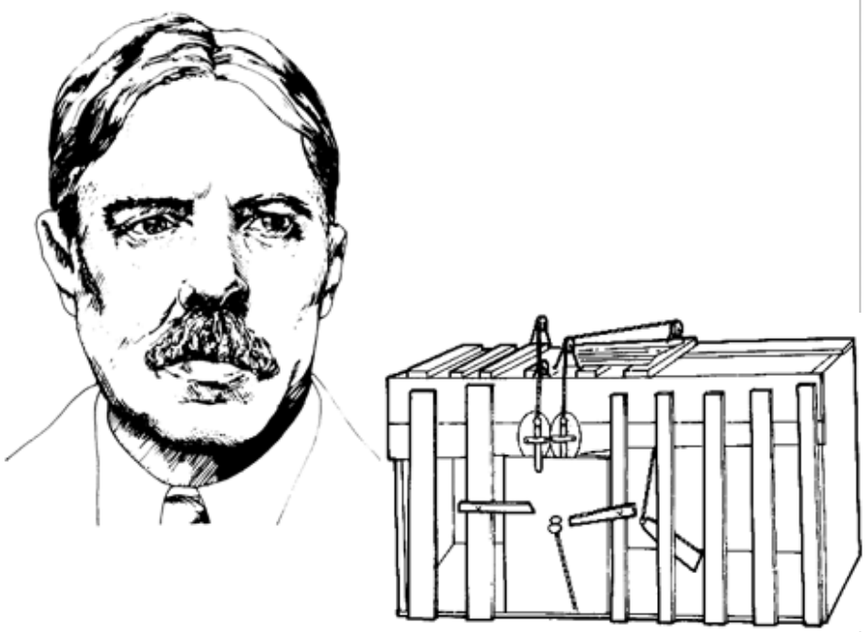
- Stimulus-Approach
- Stimulus-Response
- Contextual Inhibition



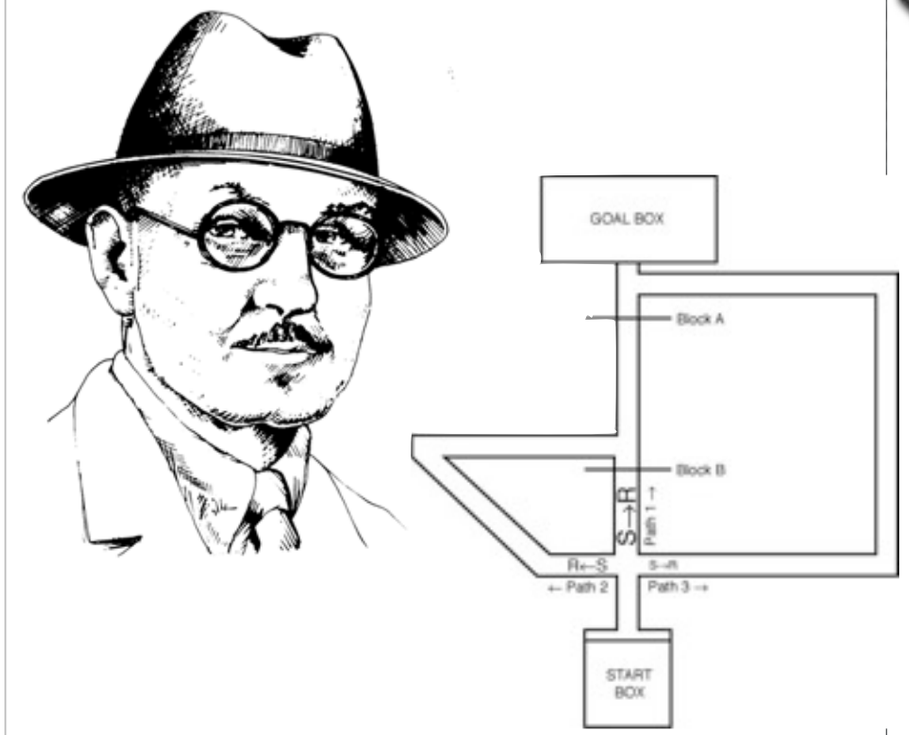
E. L. Thorndike
(1874-1949)



E. C. Tolman
(1886-1959)

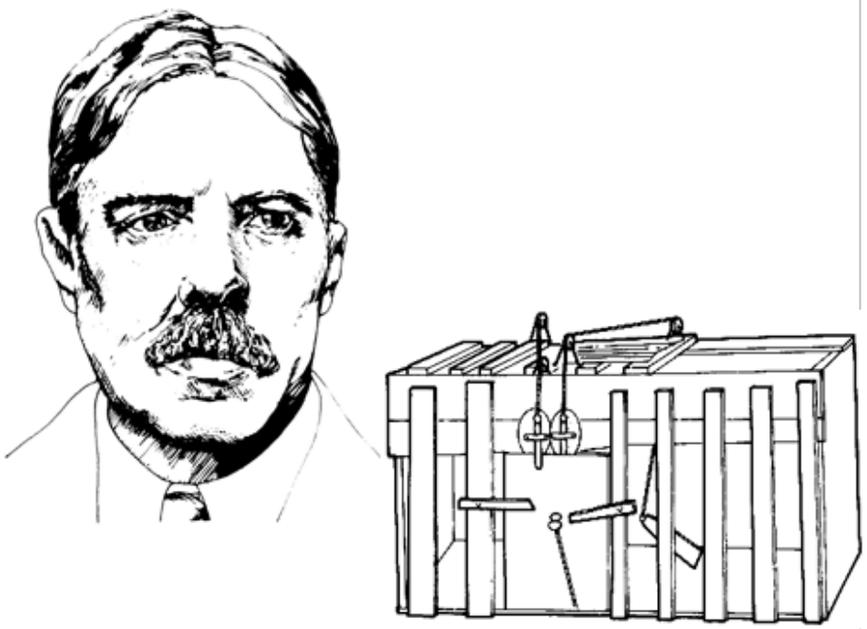


E. L. Thorndike
(1874-1949)

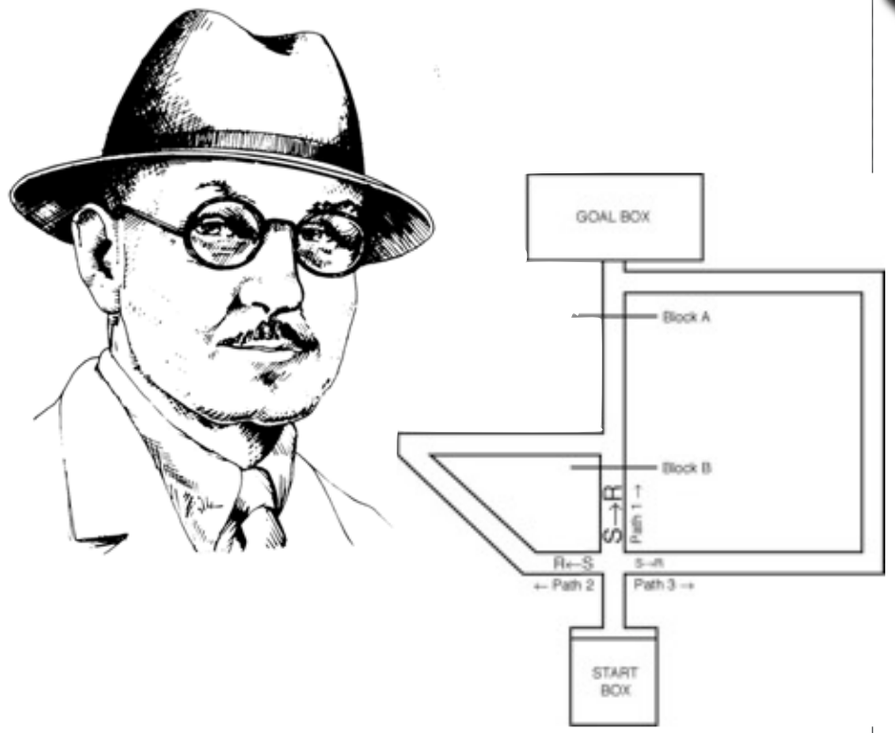


E. C. Tolman
(1886-1959)

Reactive Behavior



E. L. Thorndike
(1874-1949)

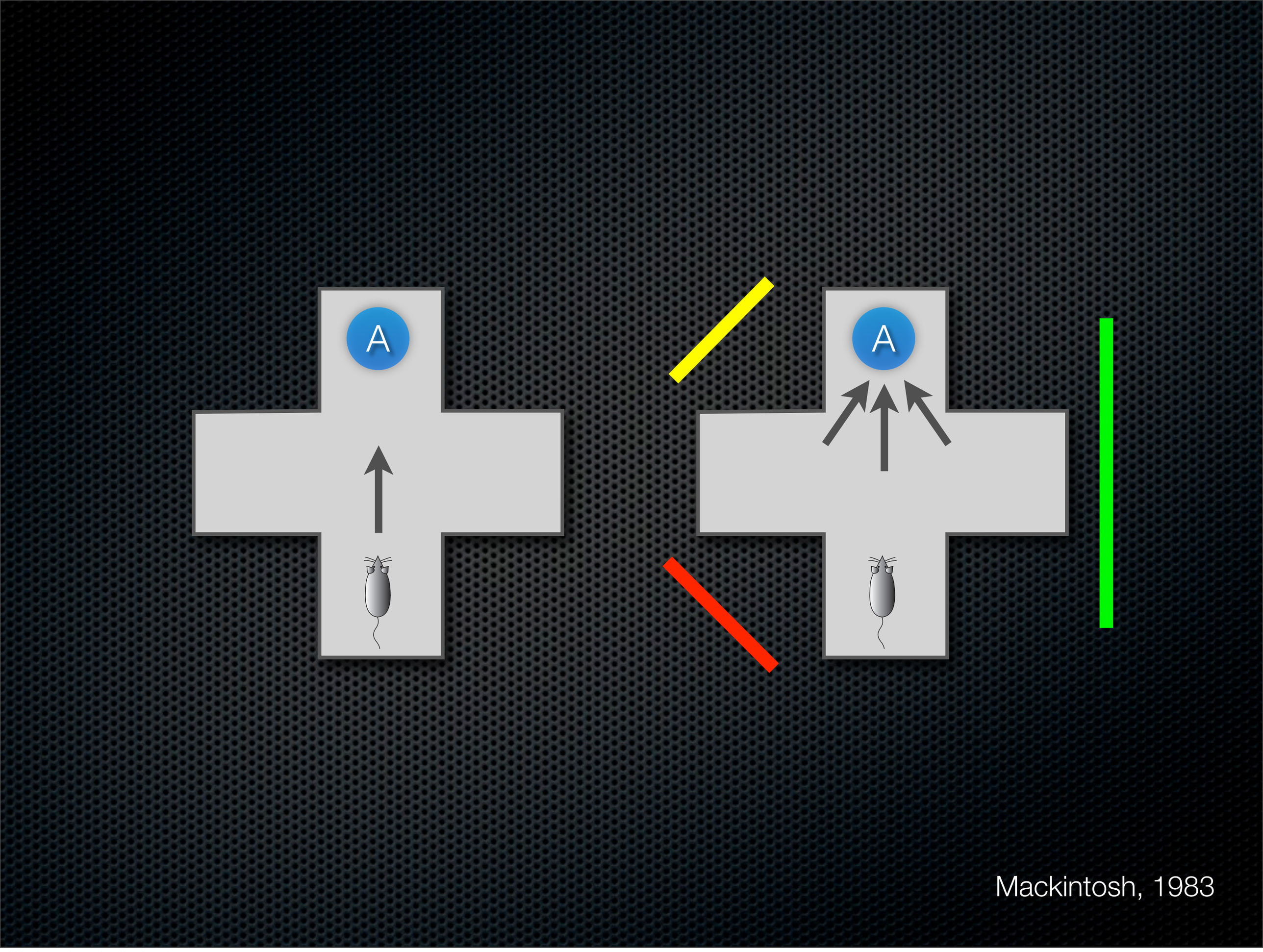


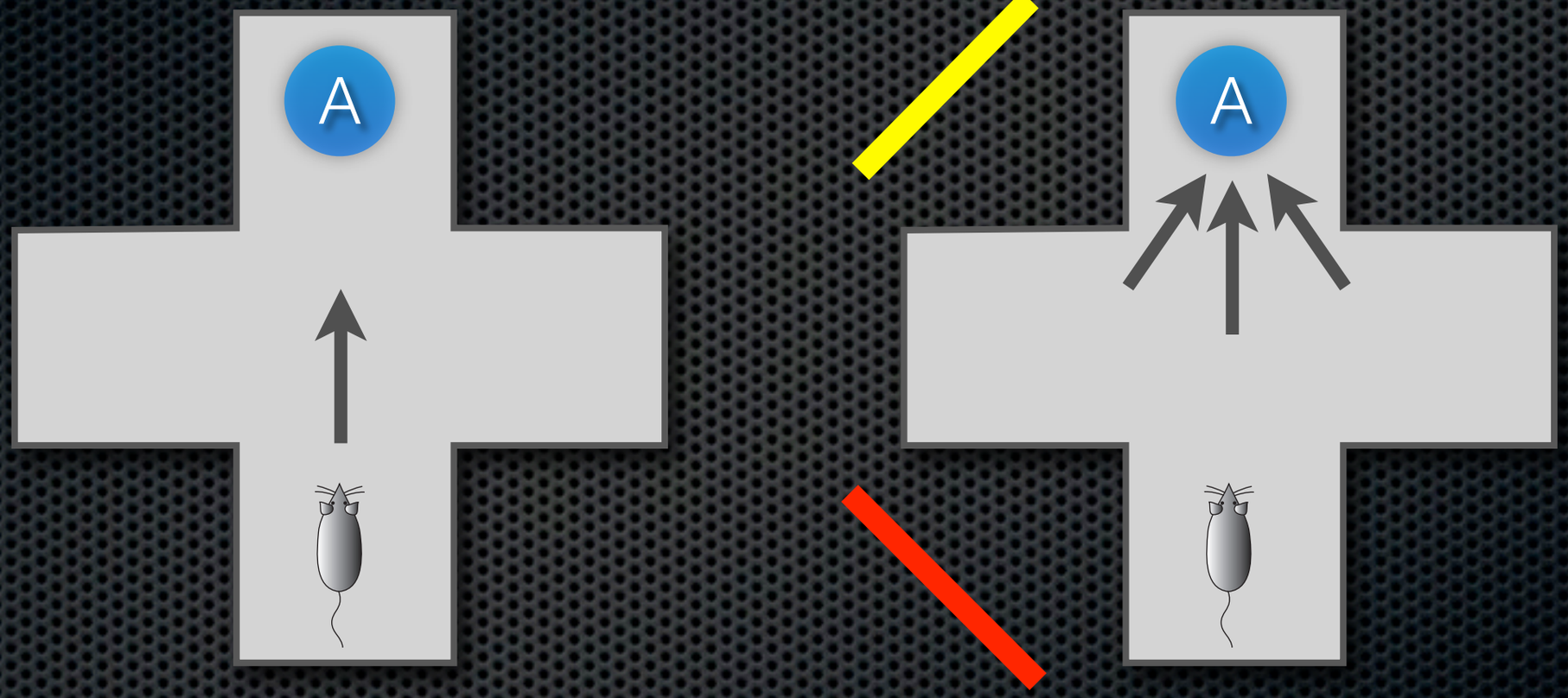
E. C. Tolman
(1886-1959)

Reactive Behavior

Purposive Behavior

Mackintosh, 1983

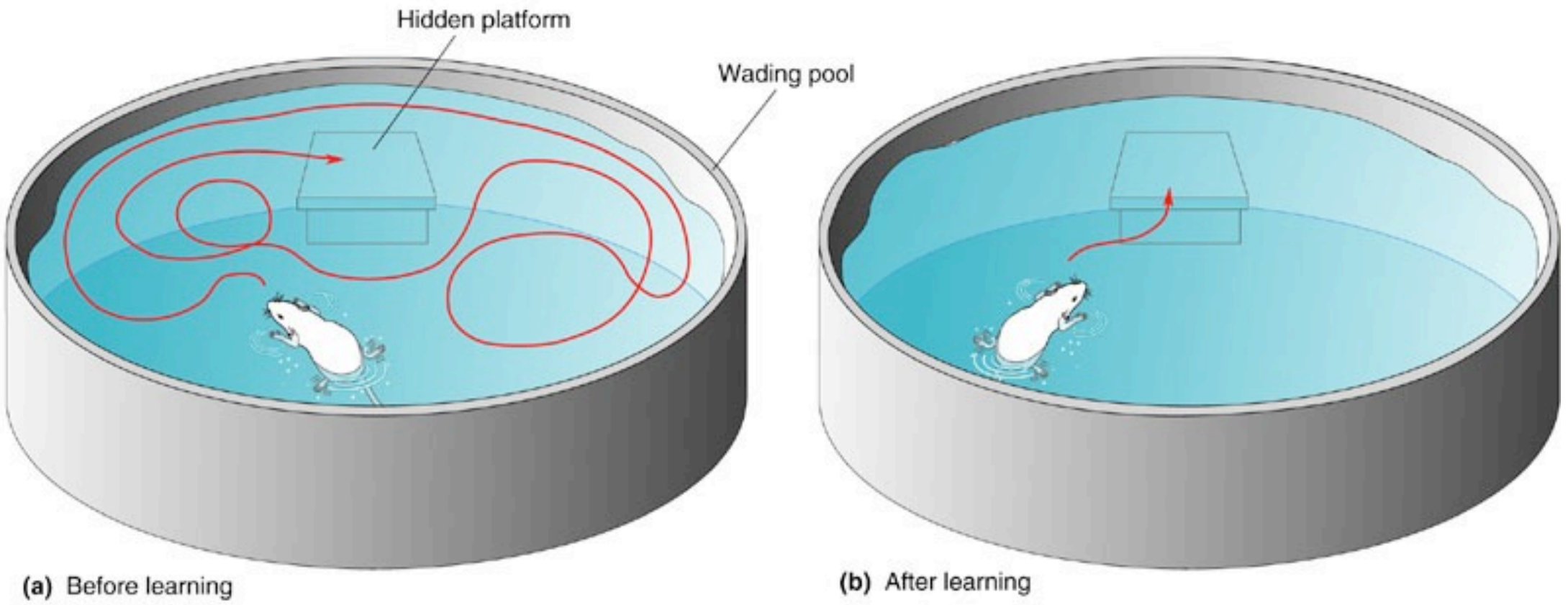

A
A
Mackintosh, 1983

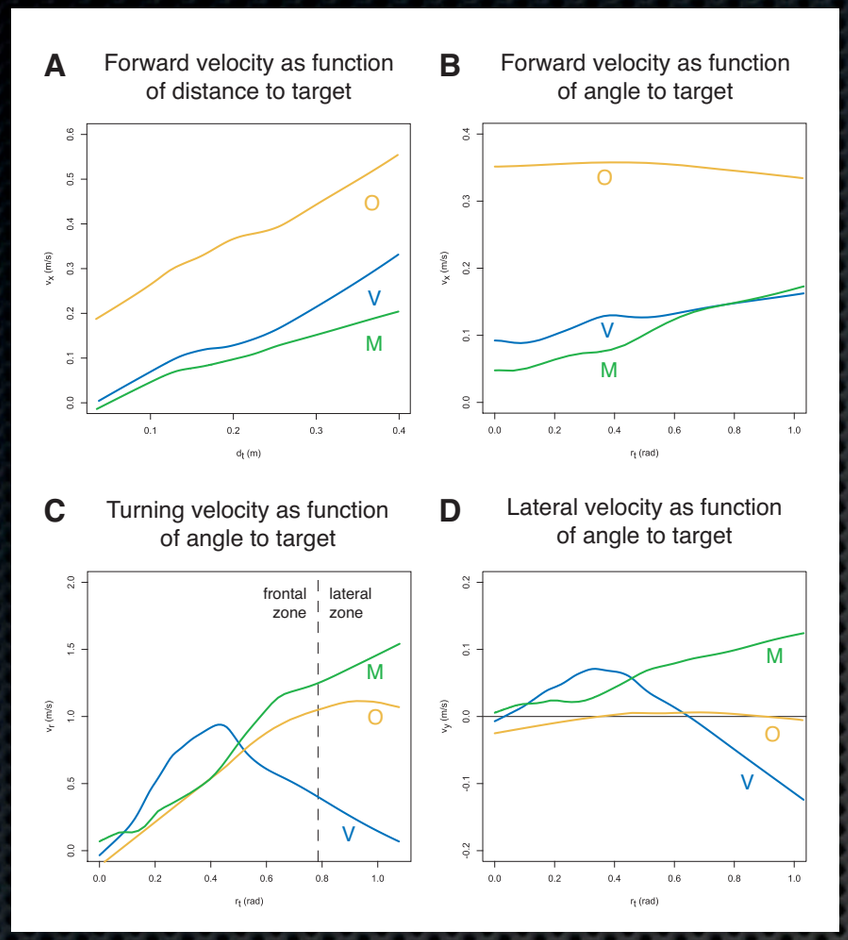


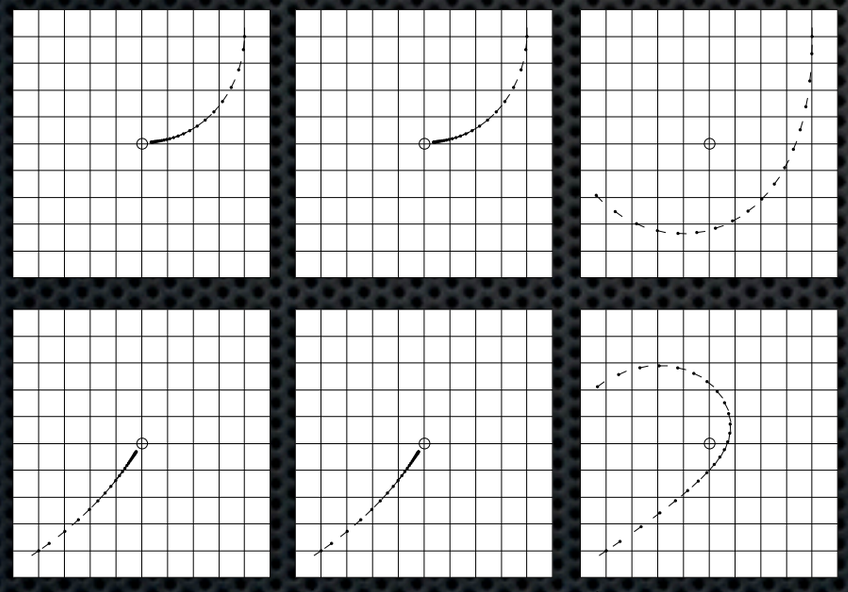
Stimulus-Response

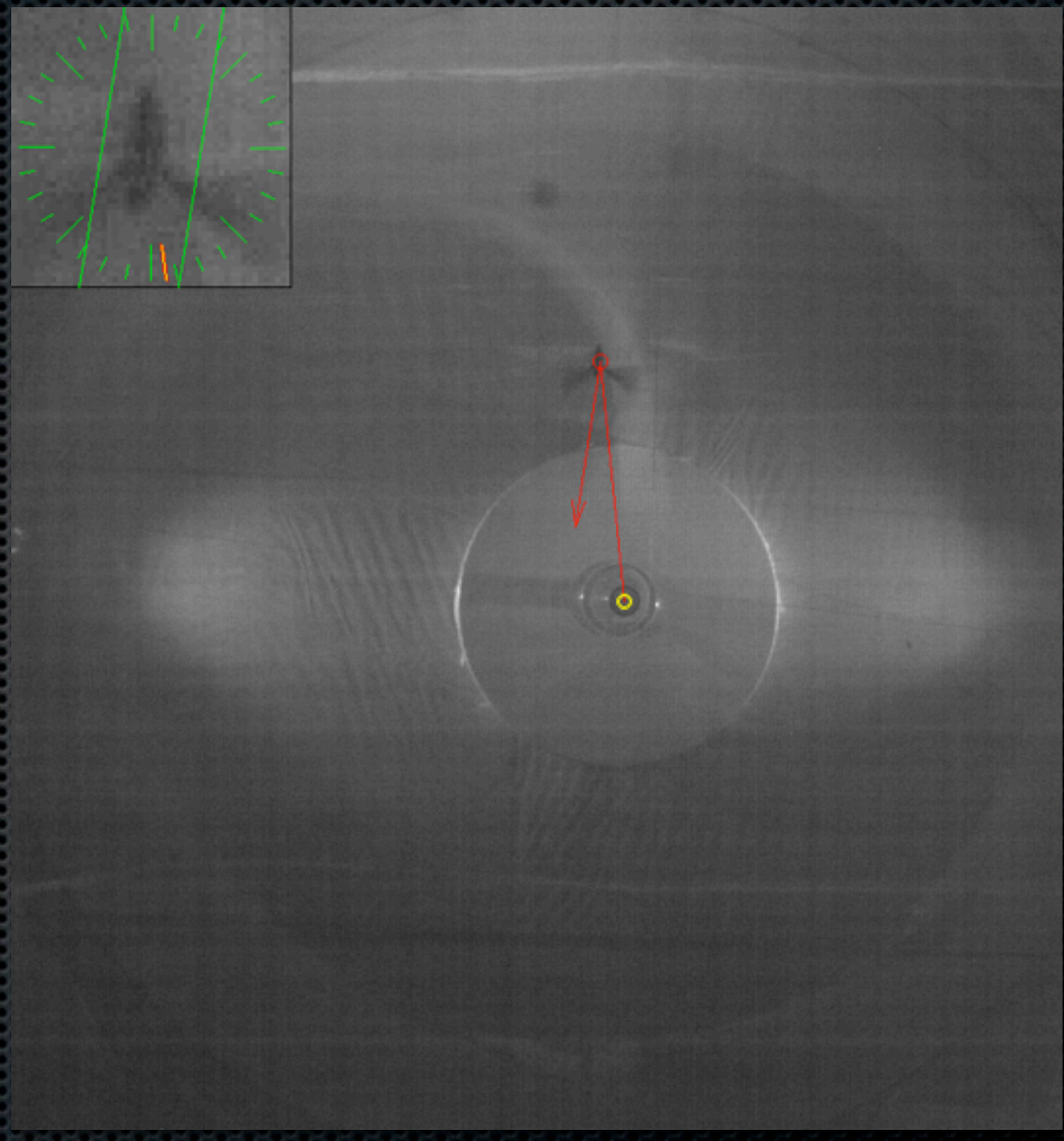
Stimulus-Approach

Mackintosh, 1983

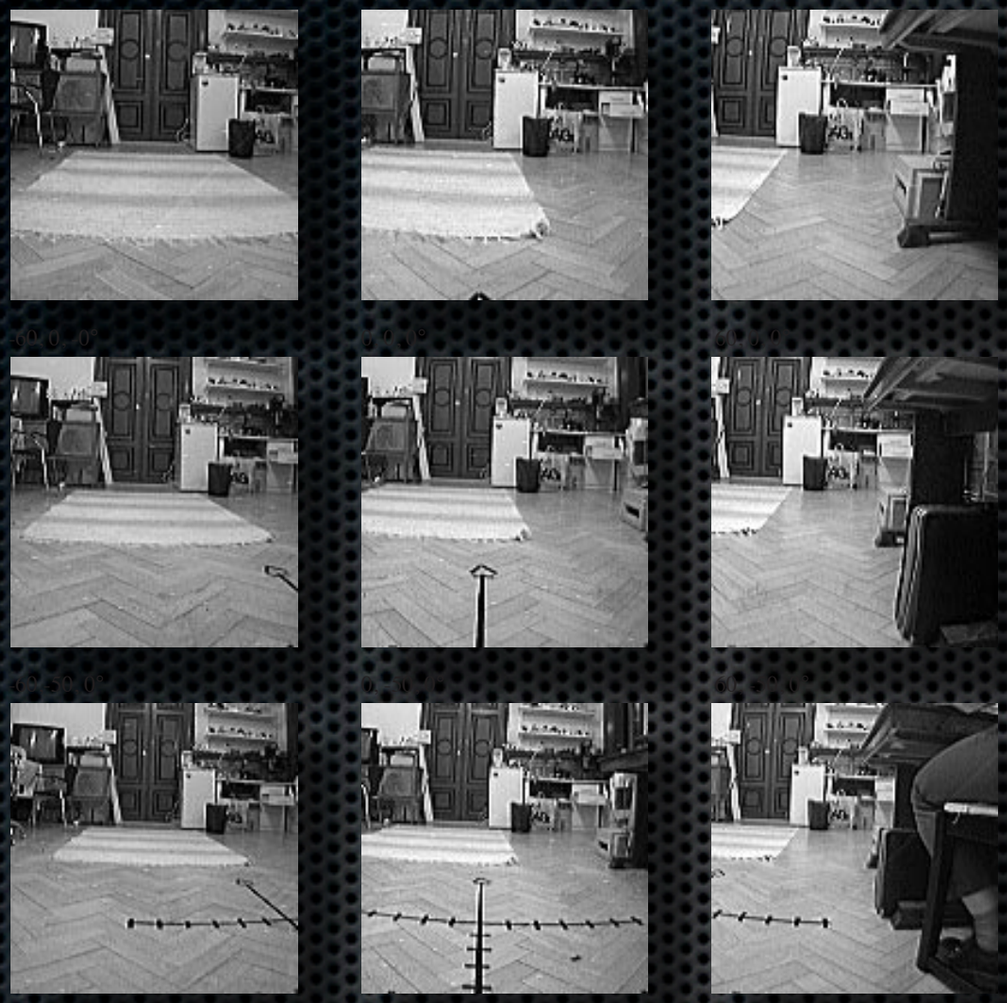

Hidden platform
Wading pool
(a) Before learning
(b) After learning

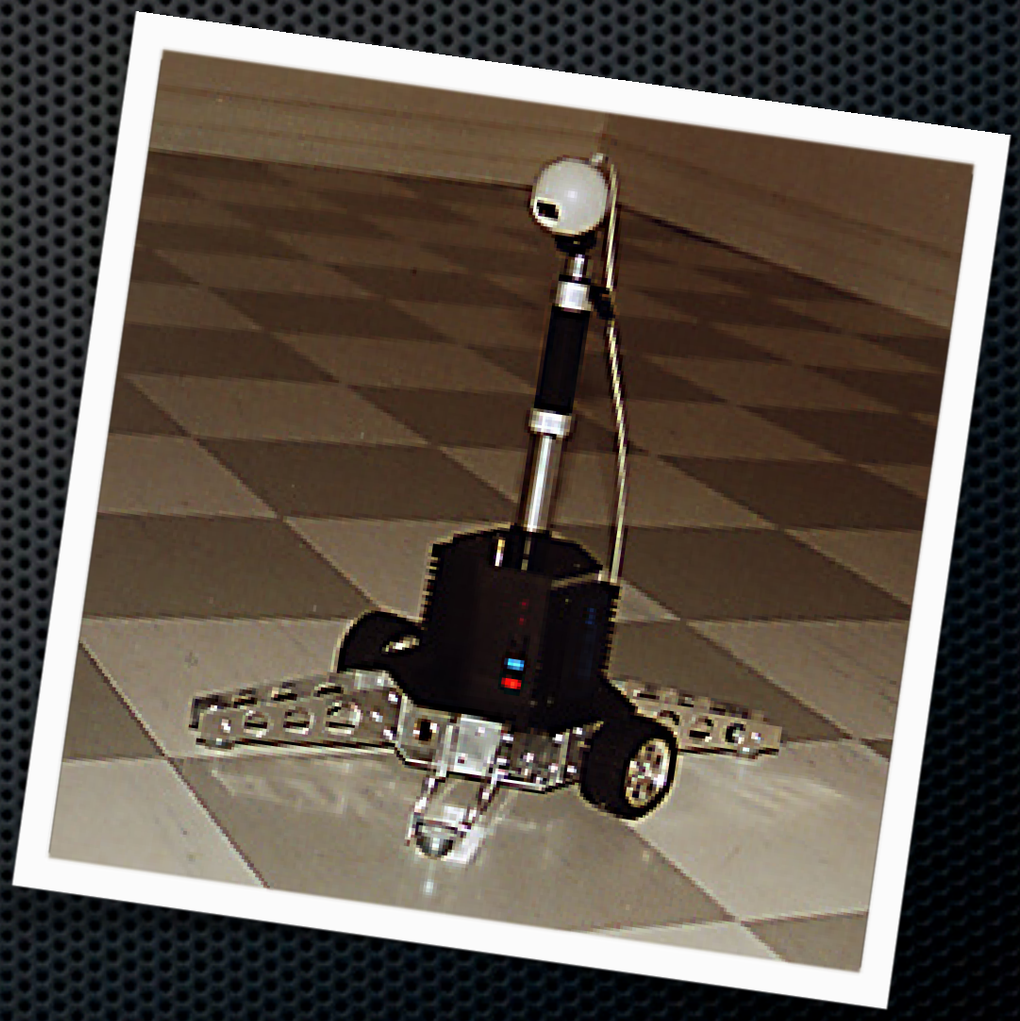




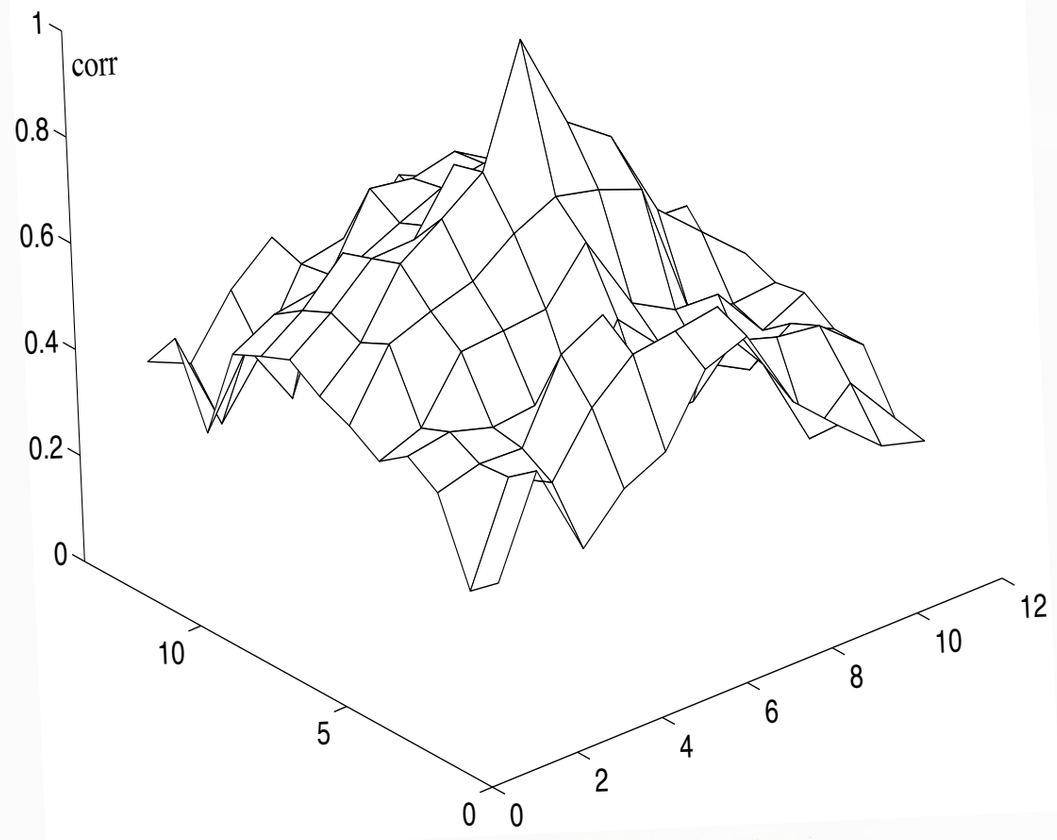


Balkenius, Dacke, Balkenius, 2010





Balkenius, Robotics and Autonomous Systems 1998



Balkenius, Robotics and Autonomous Systems 1998





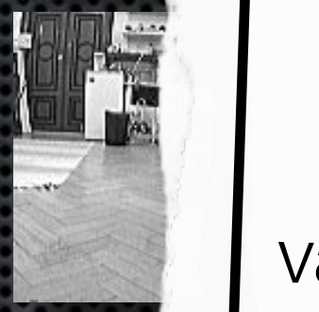




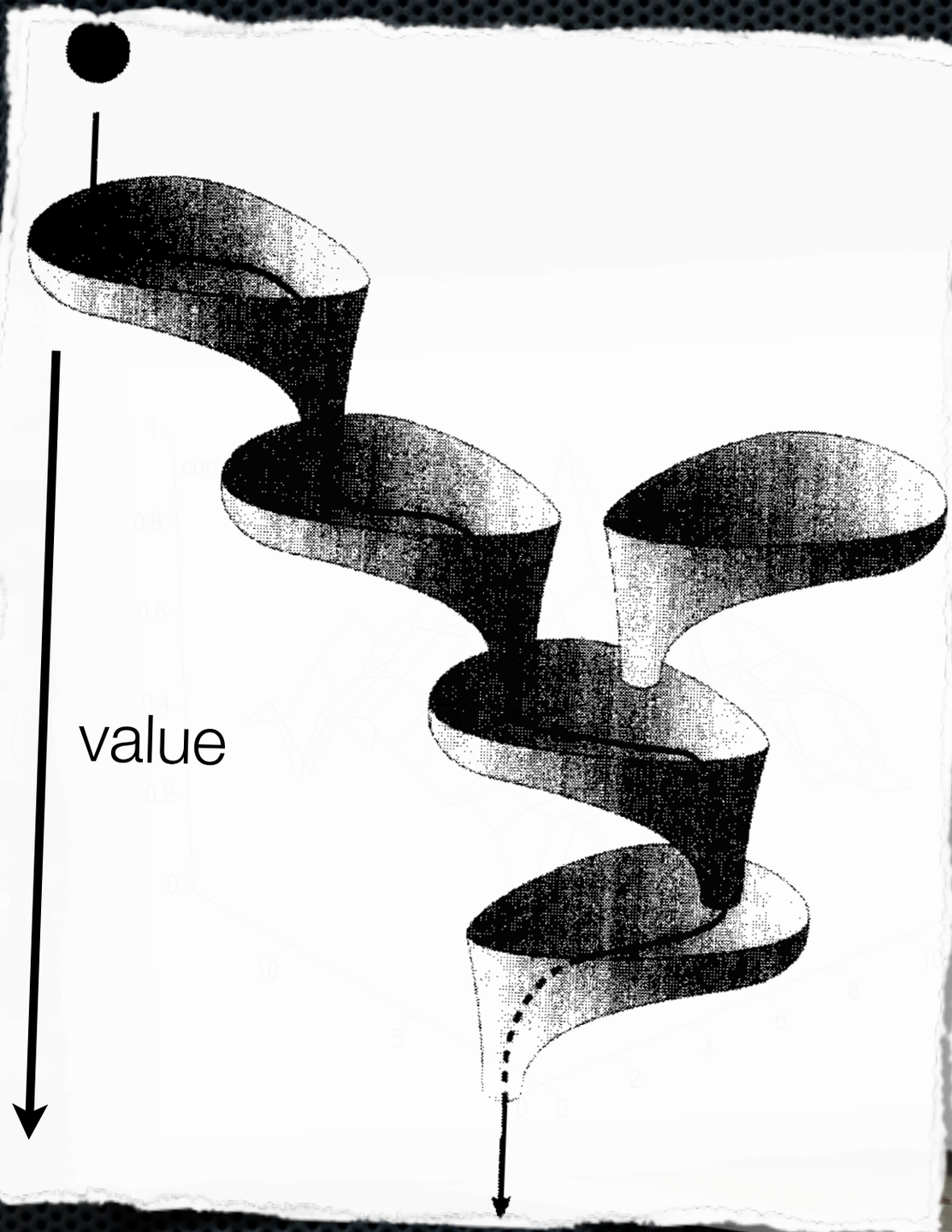

Balkenius, Robotics and Autonomous Systems 1998





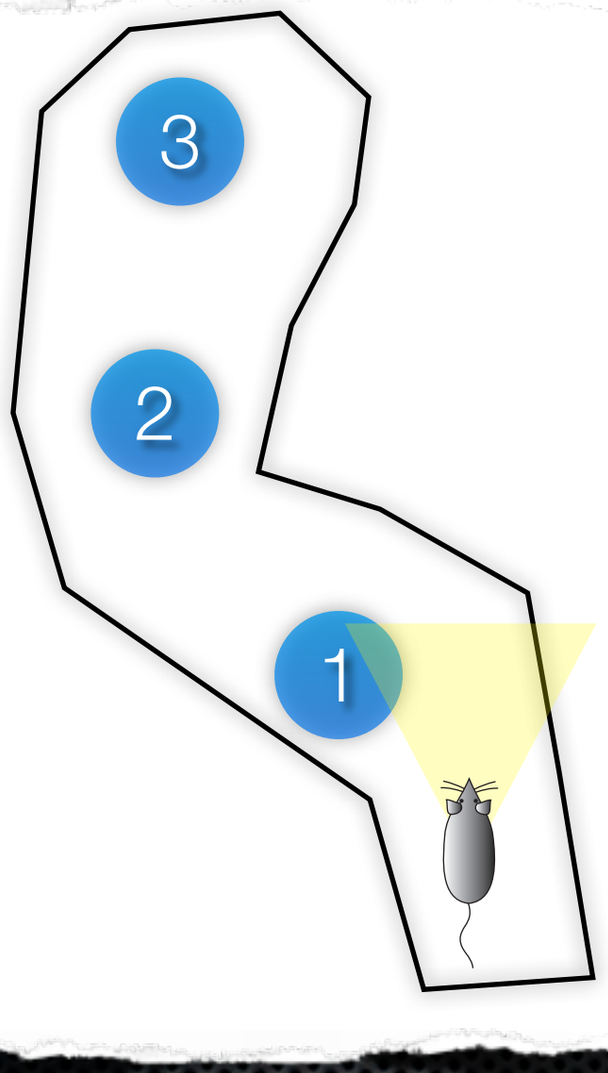




value

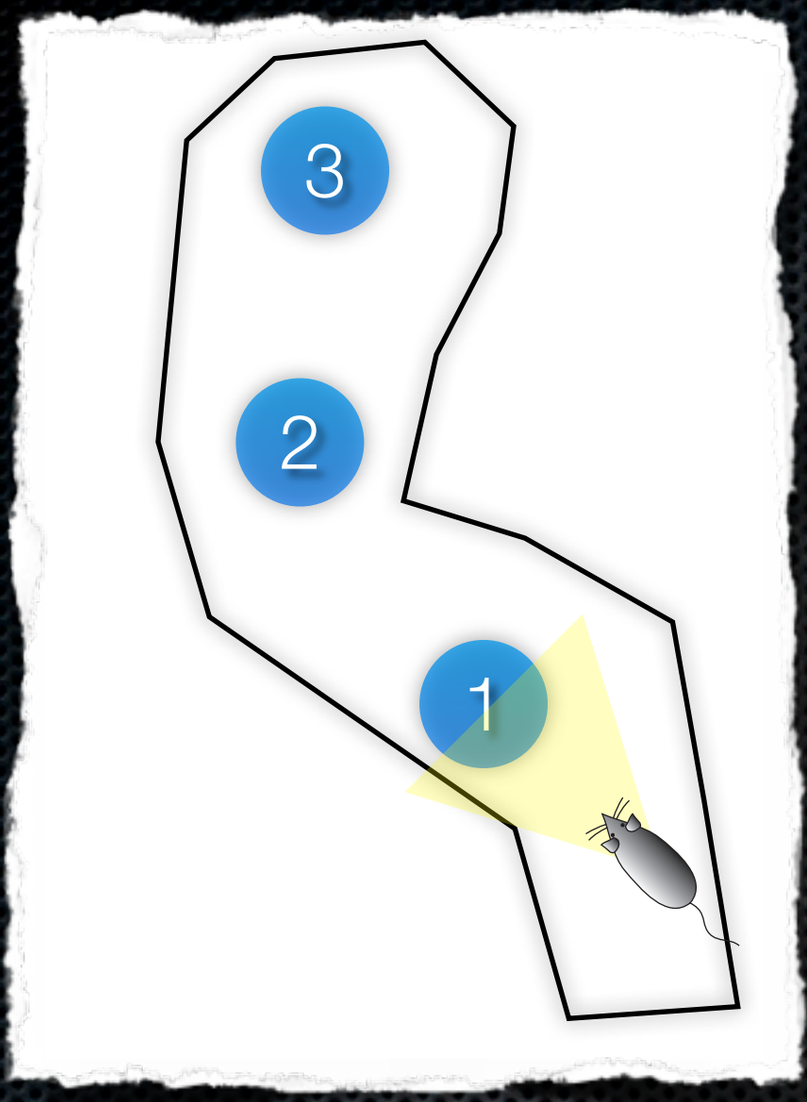


12

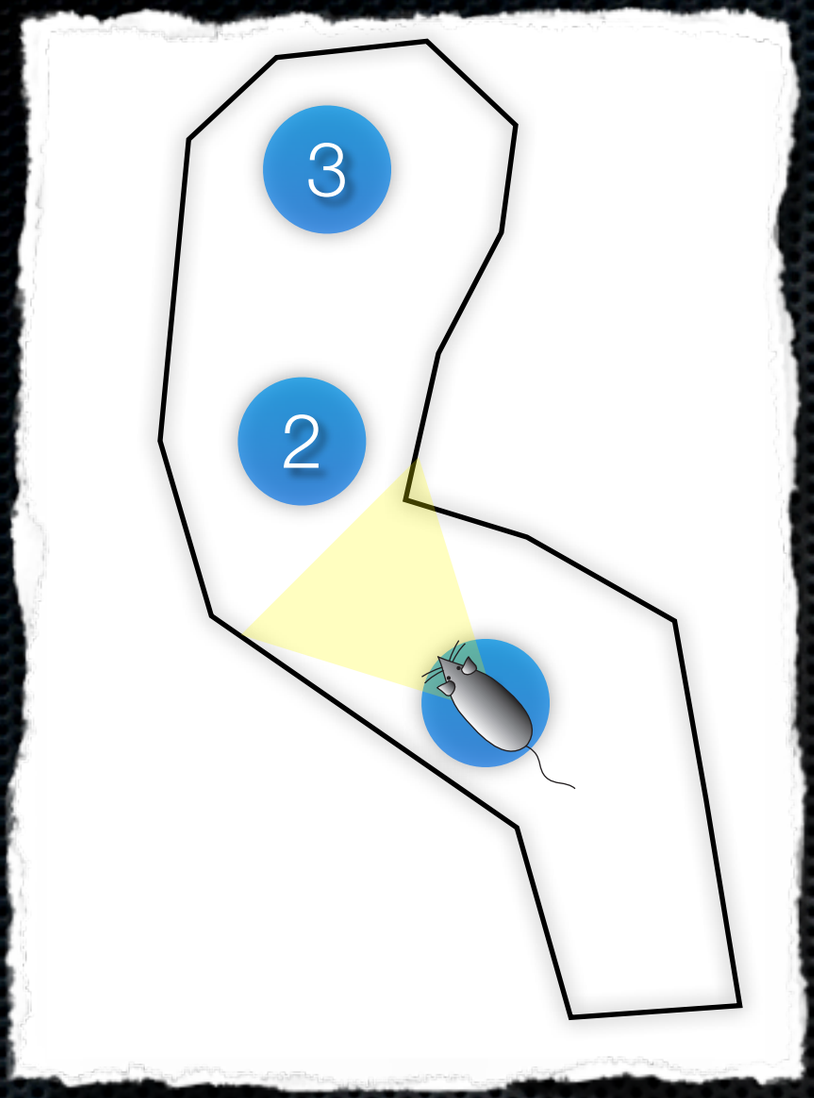
Balkenius, Robotics and Autonomous Systems 1998



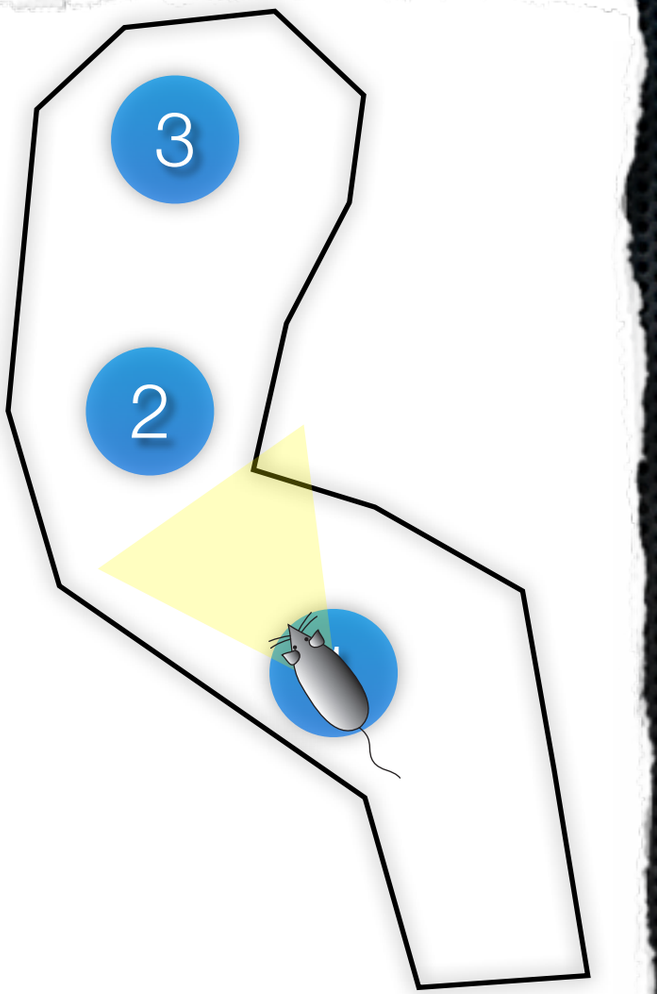
Stimulus-Approach



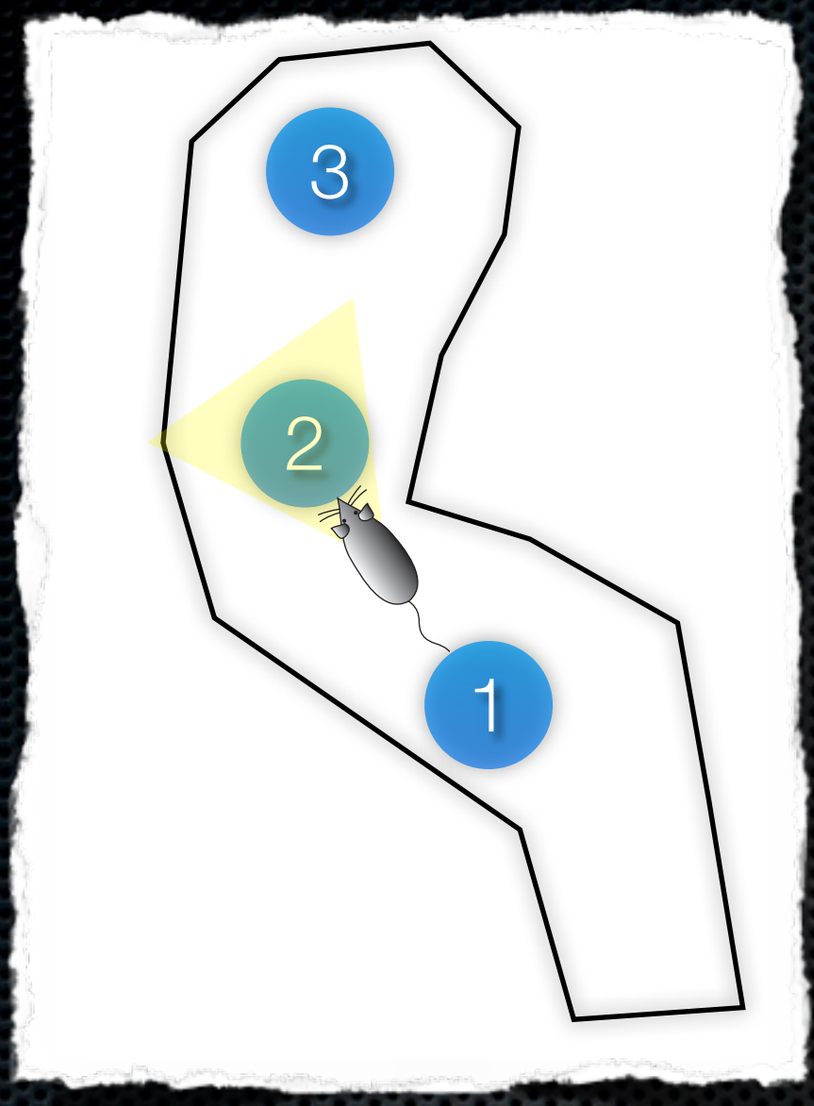
Stimulus-Approach



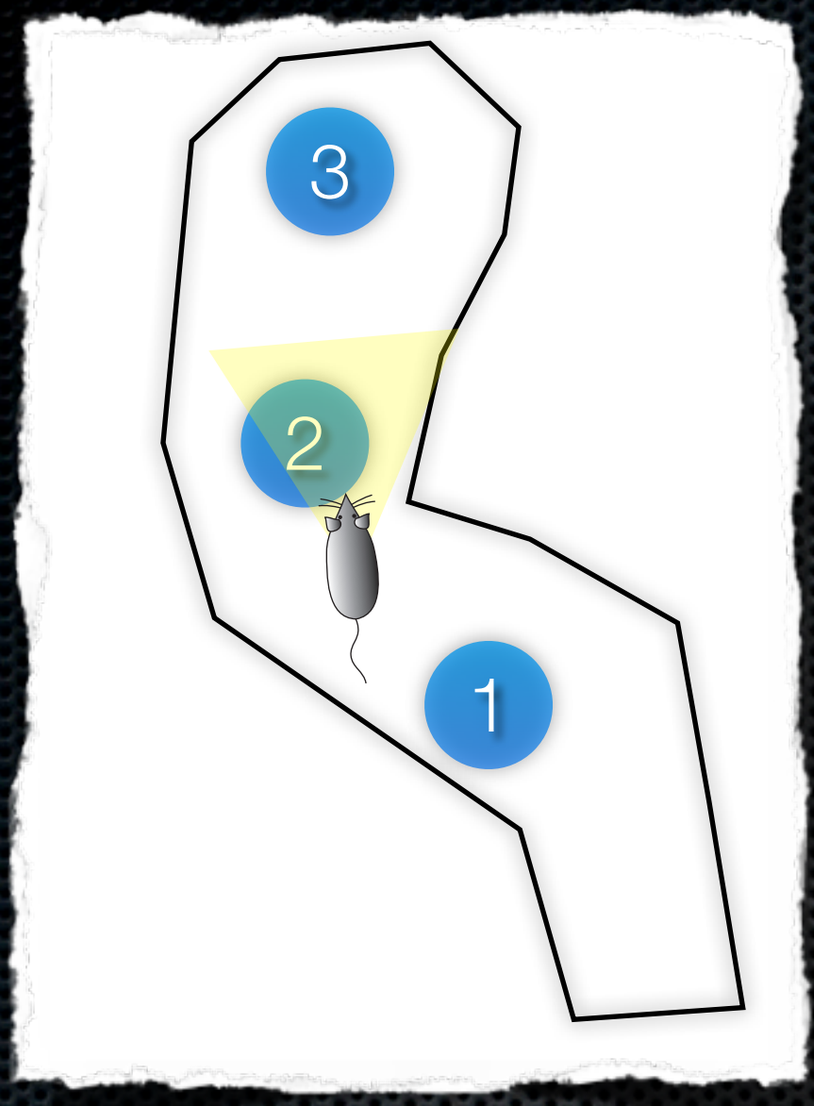
Stimulus-Approach



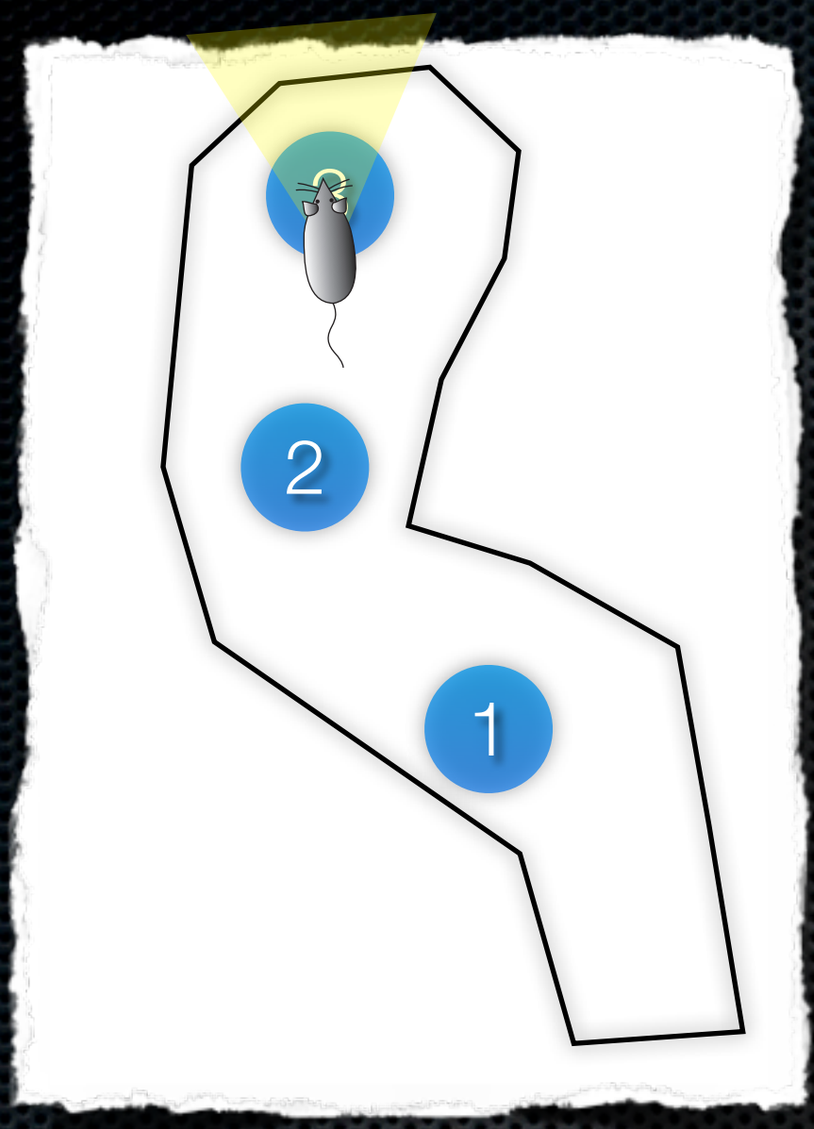
Stimulus-Approach



Stimulus-Approach



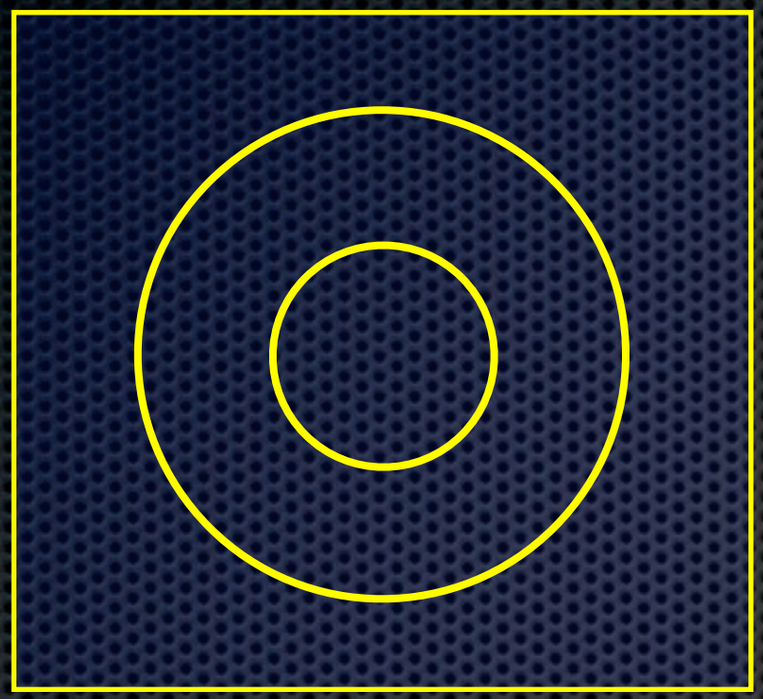
Stimulus-Approach



Stimulus-Approach

Switching Control

eye regions

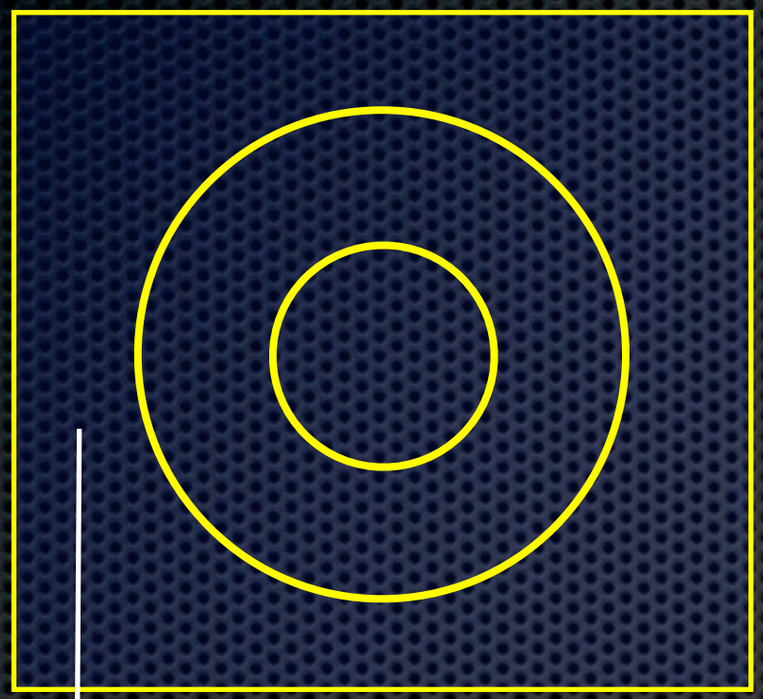


QuickTime™ and a
MPEG-4 Video decompressor
are needed to see this picture.

Balkenius & Kopp, 1997

Switching Control

eye regions

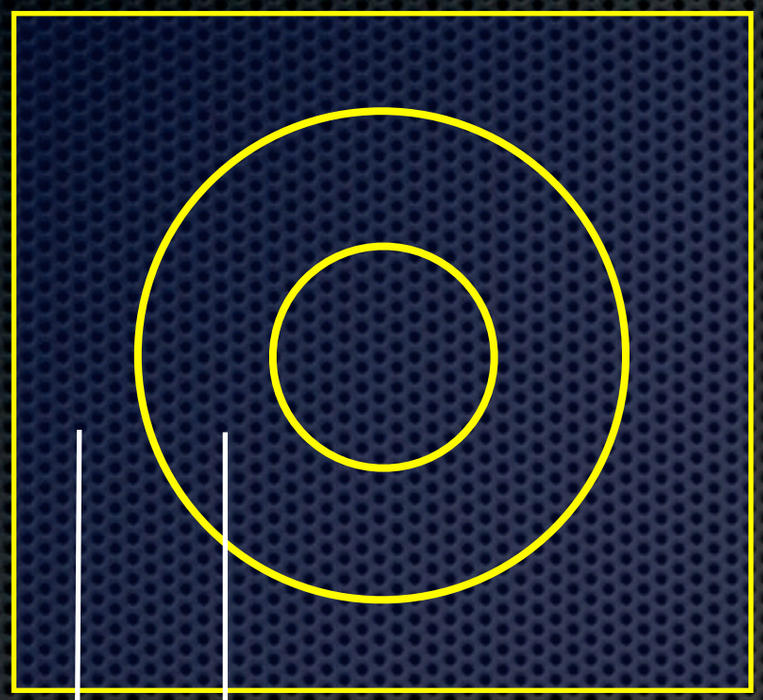


orienting

QuickTime™ and a
MPEG-4 Video decompressor
are needed to see this picture.

Balkenius & Kopp, 1997

Switching Control

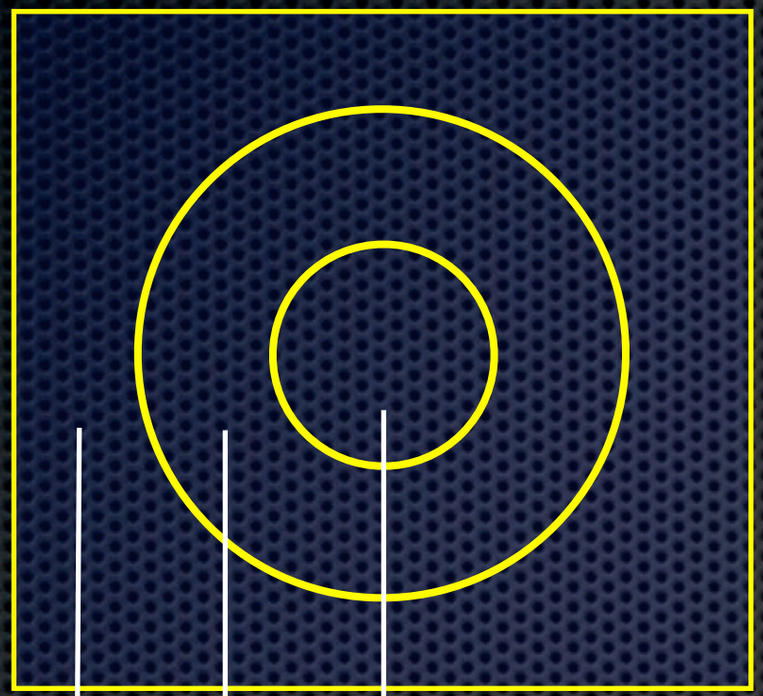


QuickTime™ and a
MPEG-4 Video decompressor
are needed to see this picture.

Balkenius & Kopp, 1997

Switching Control

eye regions



QuickTime™ and a
MPEG-4 Video decompressor
are needed to see this picture.

fixation / smooth pursuit

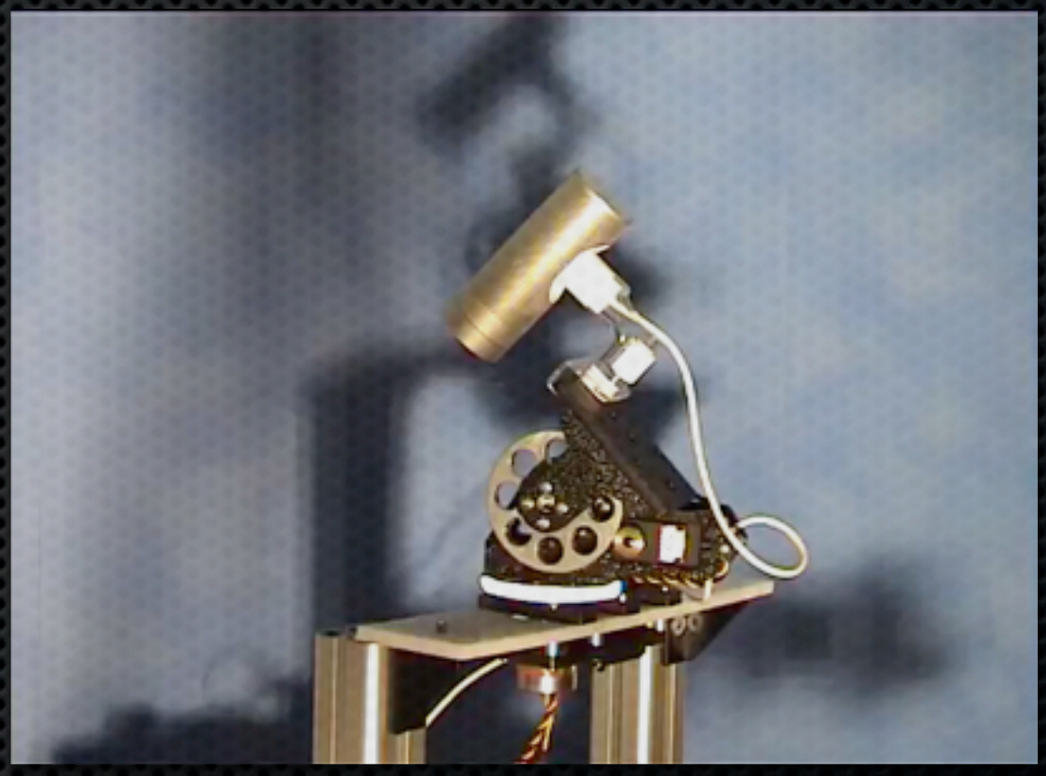
saccades

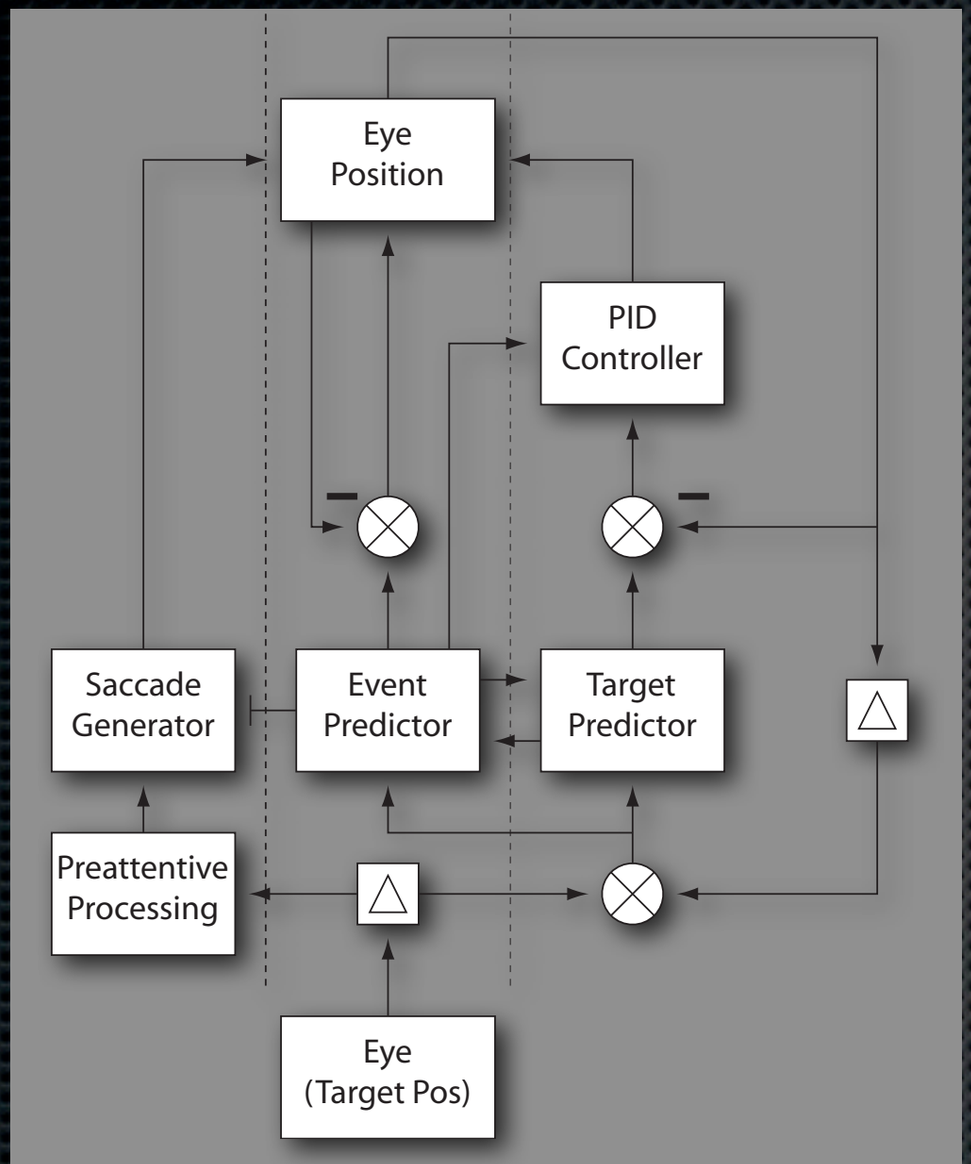
orienting

Balkenius & Kopp, 1997

BabYlab

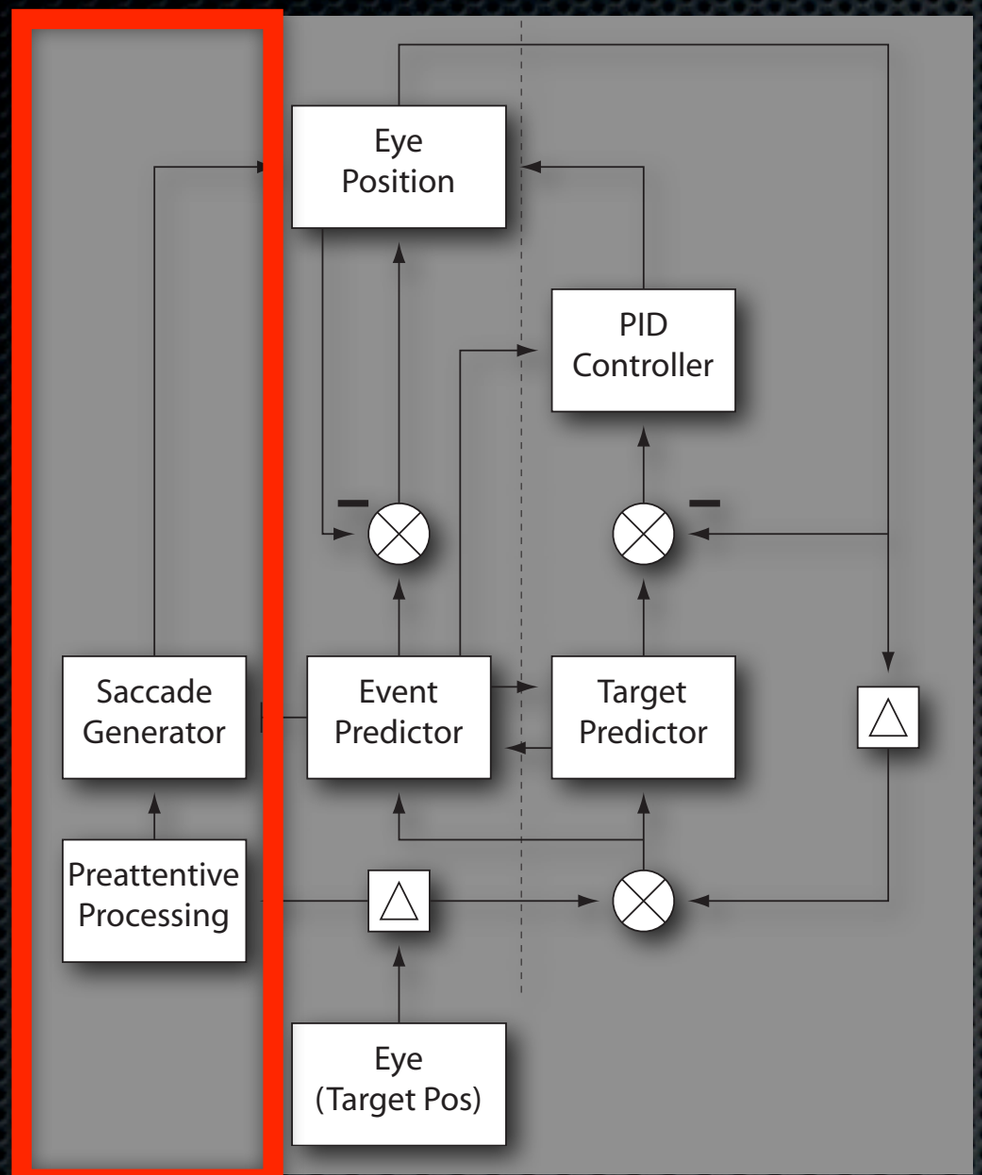
Smooth Pursuit Eye-Movements

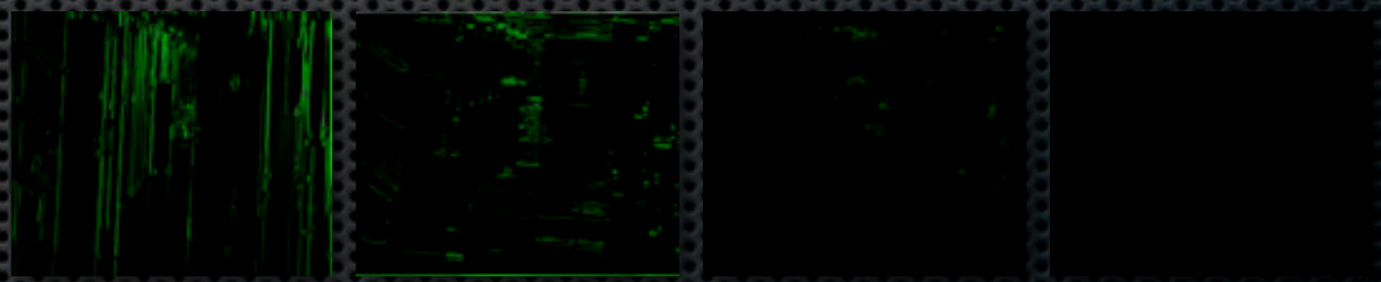




Balkenius & Johansson, 2005

Orienting

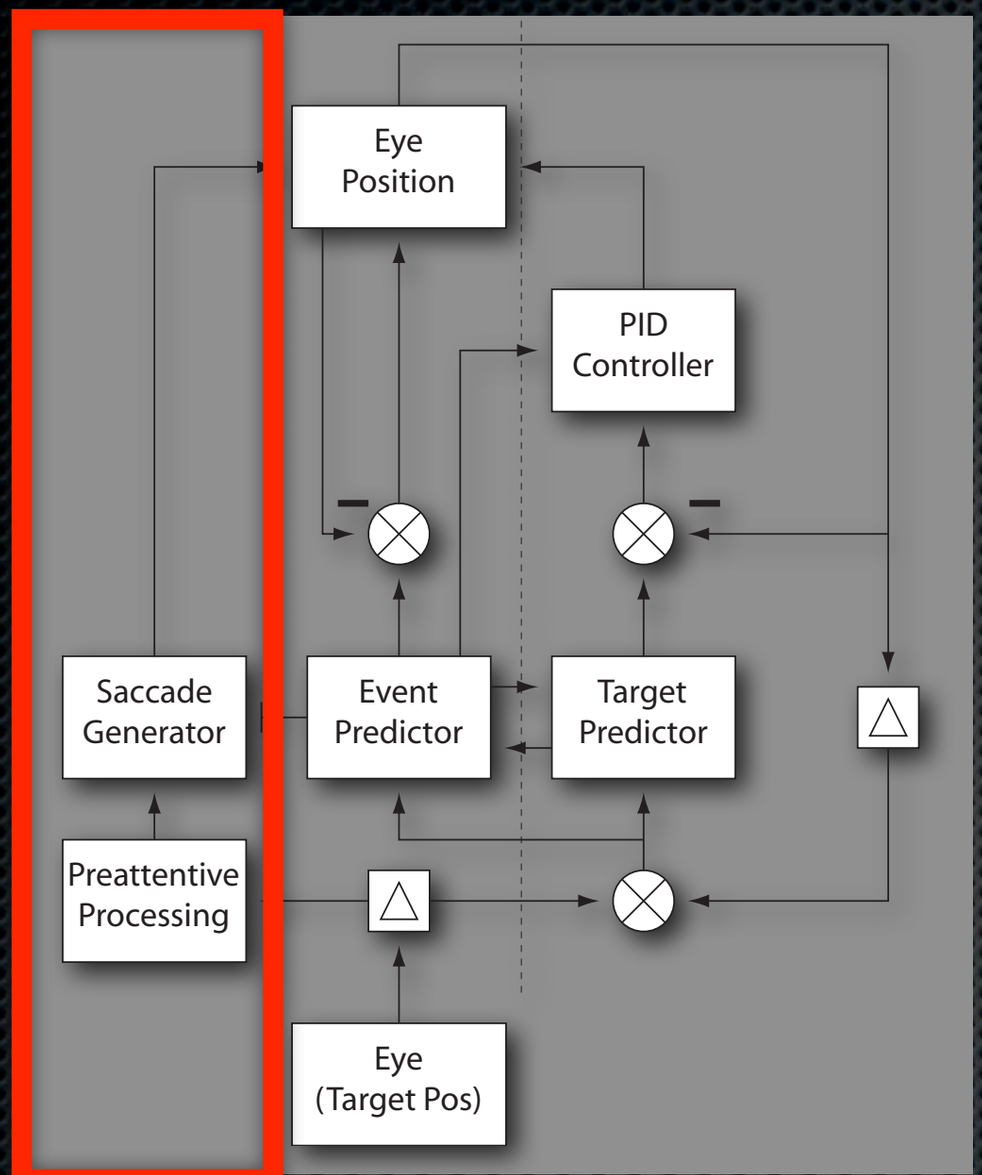


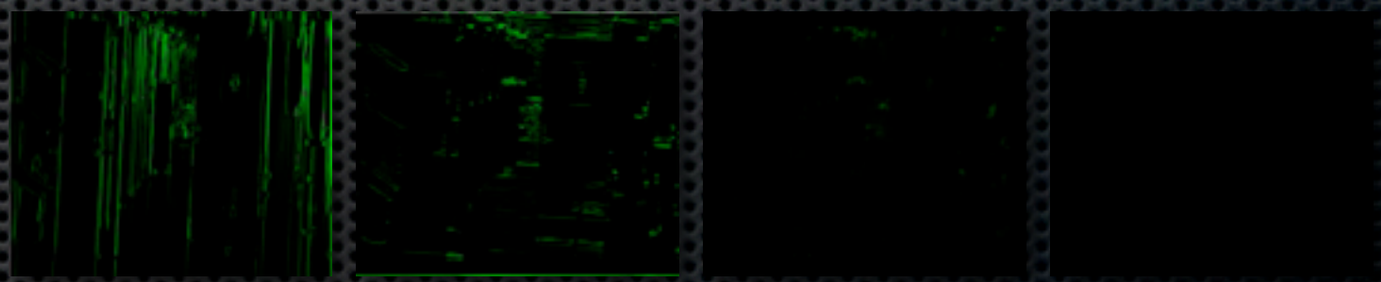


F_0 F_1 F_2 F_3

Balkenius, Åström, Eriksson, 2004

Orienting



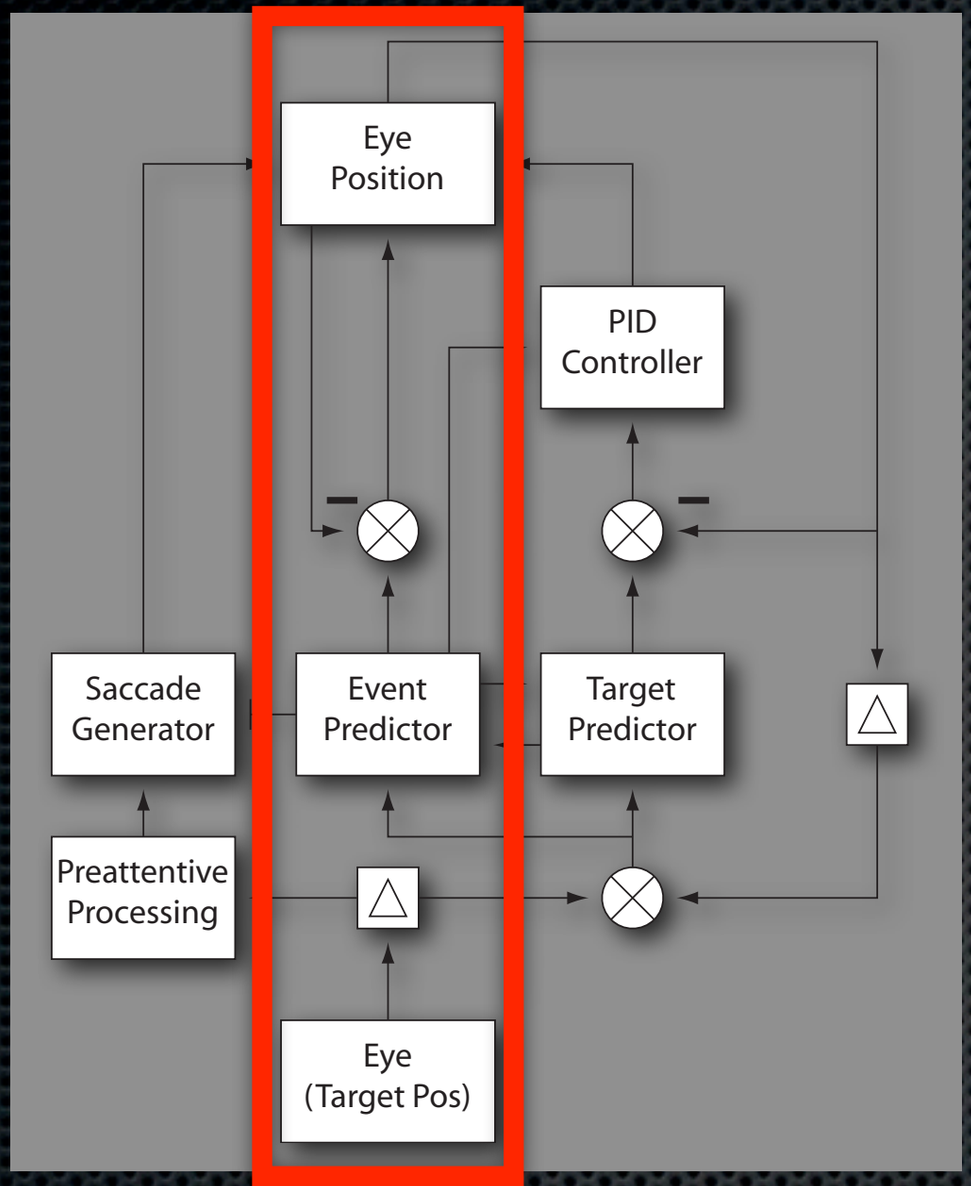


F_0 F_1 F_2 F_3

$$S(x,y) = G(x,y) * \sum_m \theta_m F_m(x,y)$$

Balkenius, Åström, Eriksson, 2004

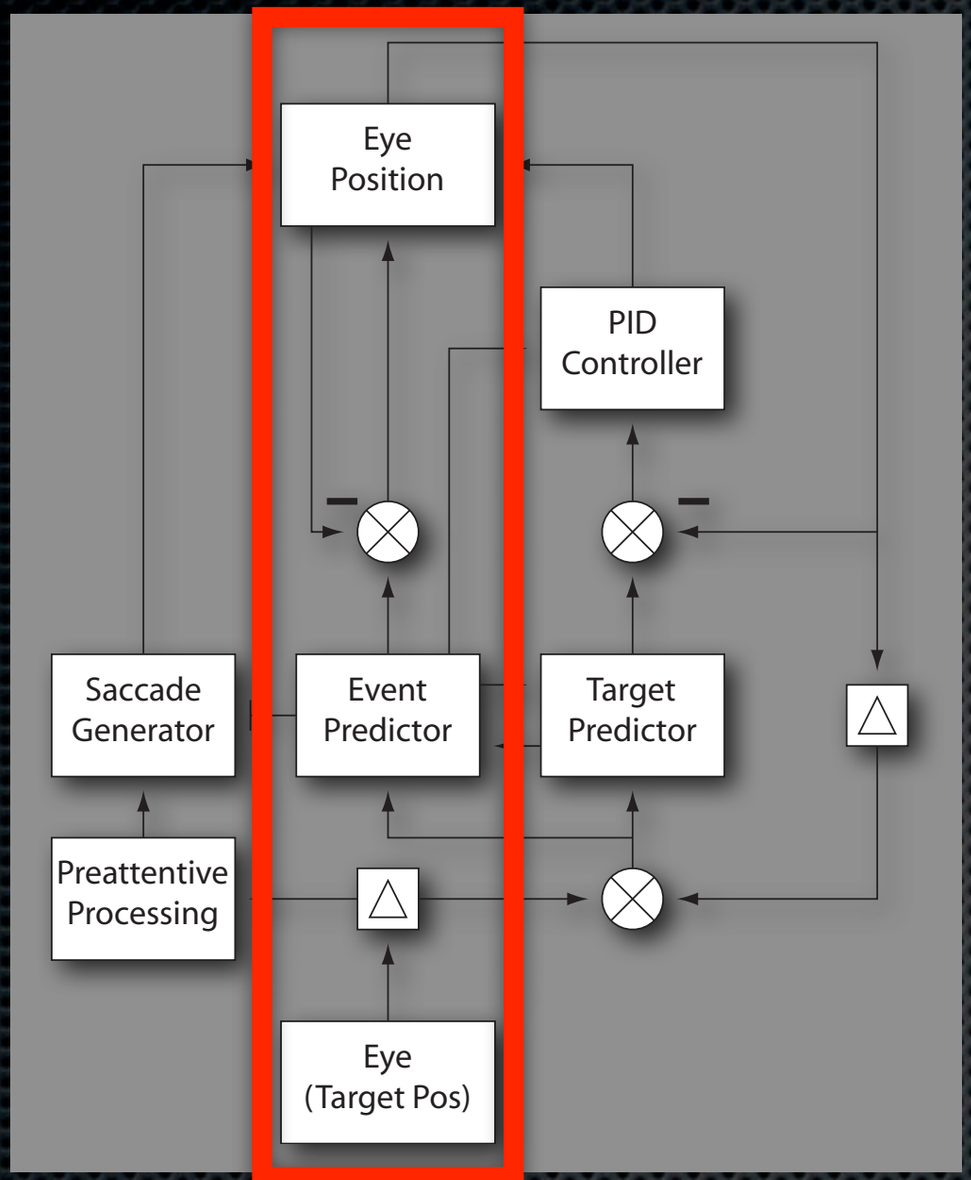
Saccades

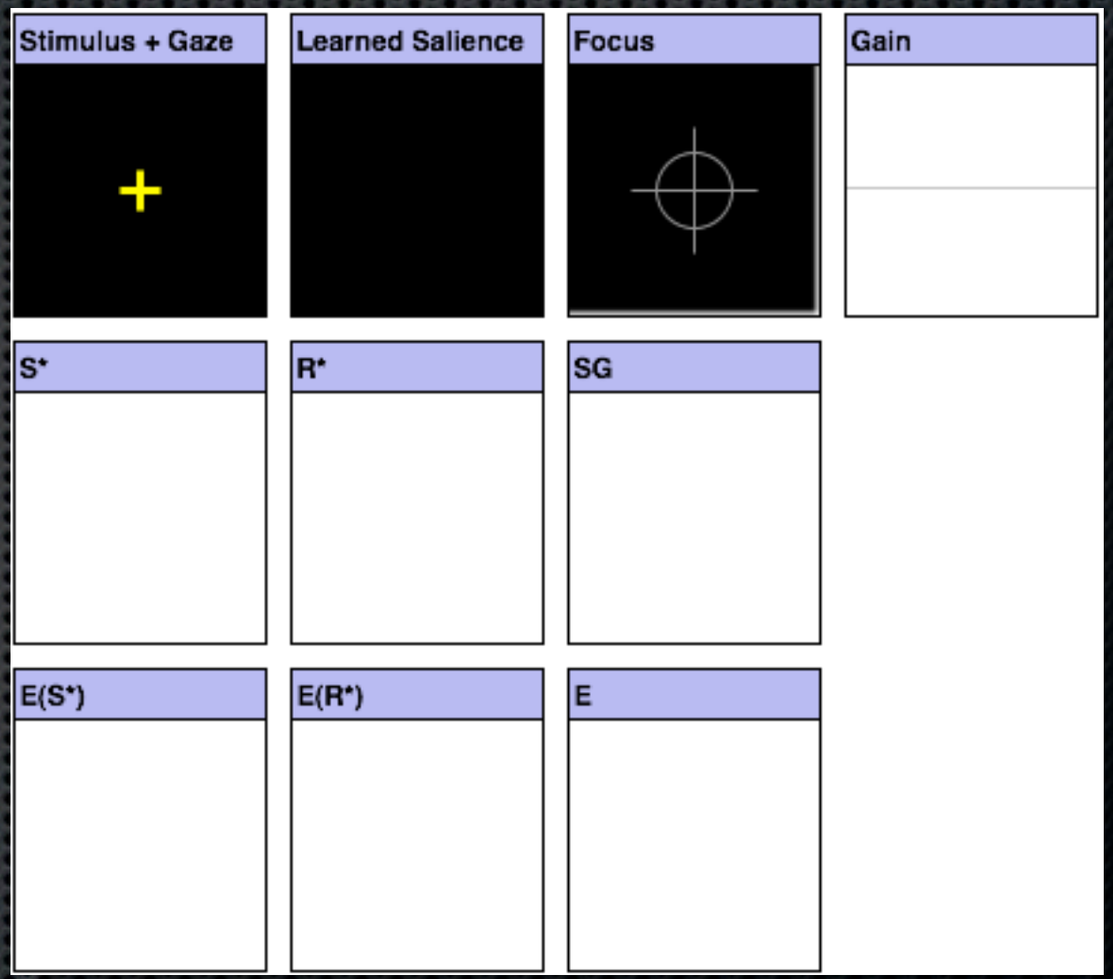


stimulus-response

Balkenius, 2000

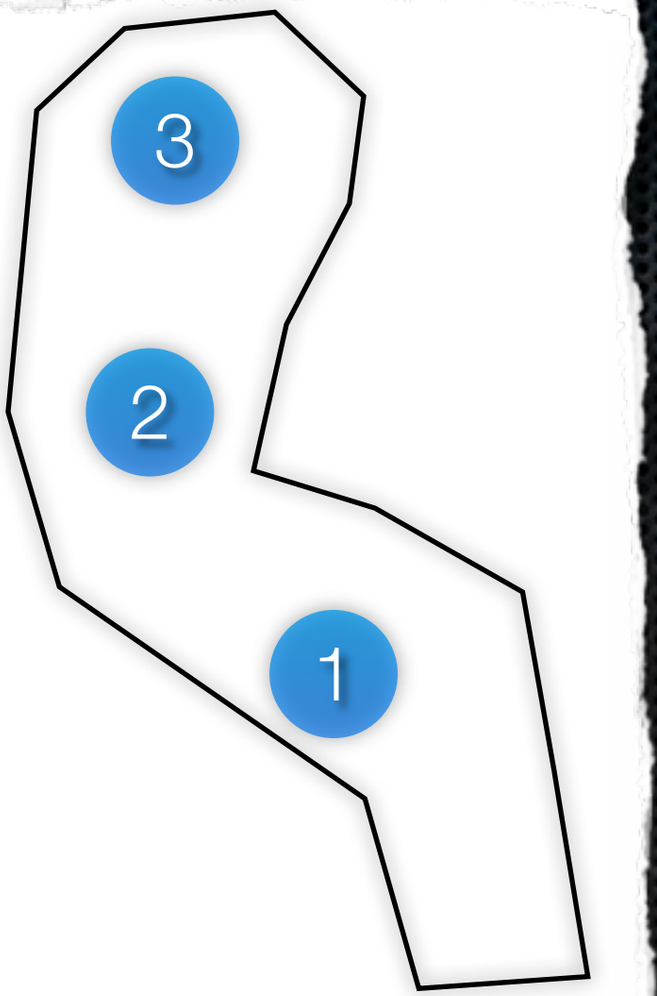
Saccades

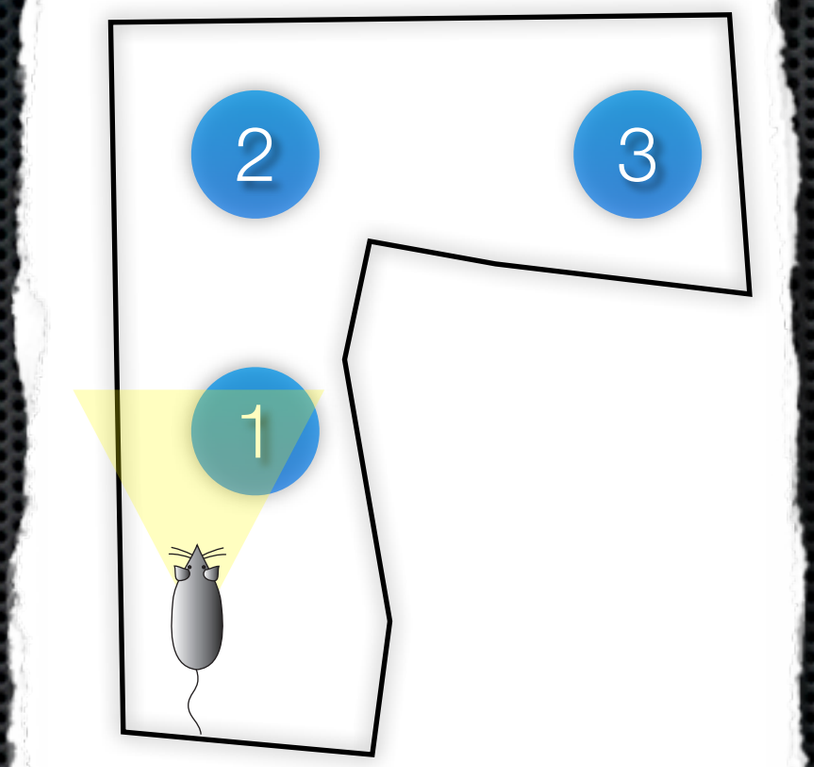




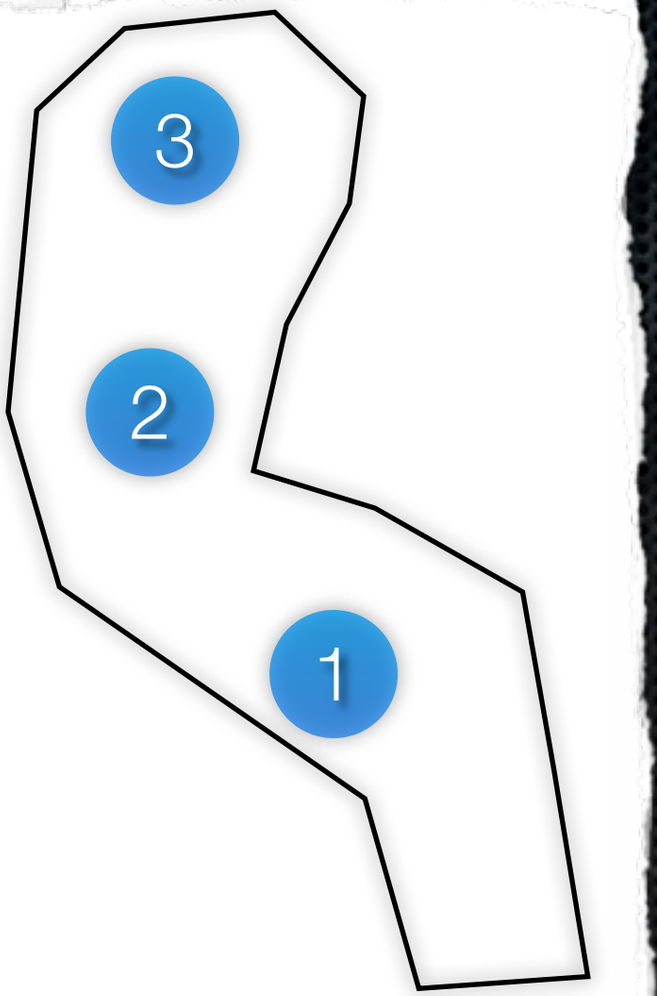
stimulus-response

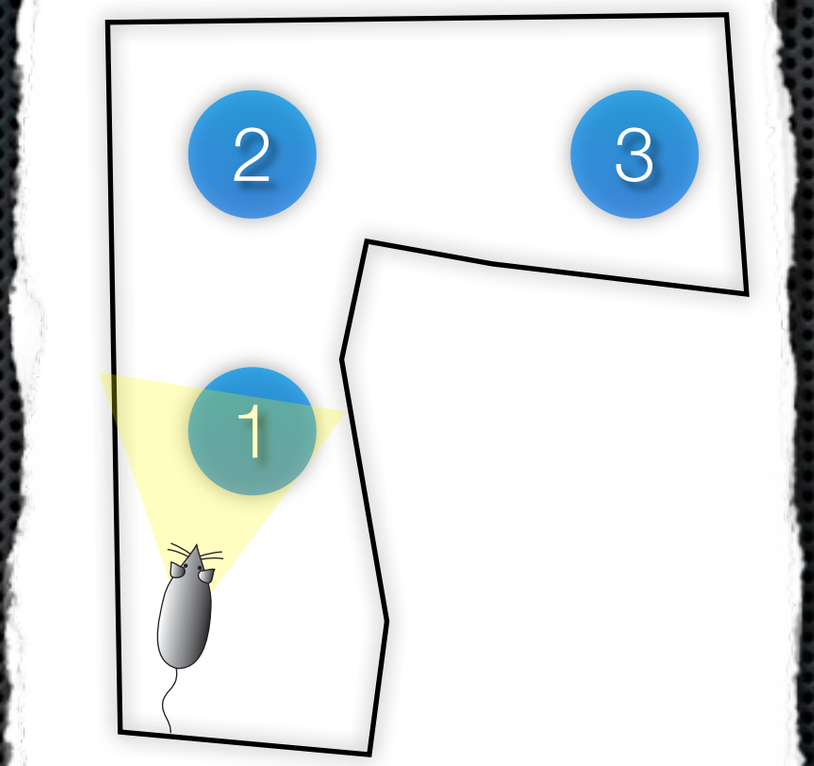
Balkenius, 2000



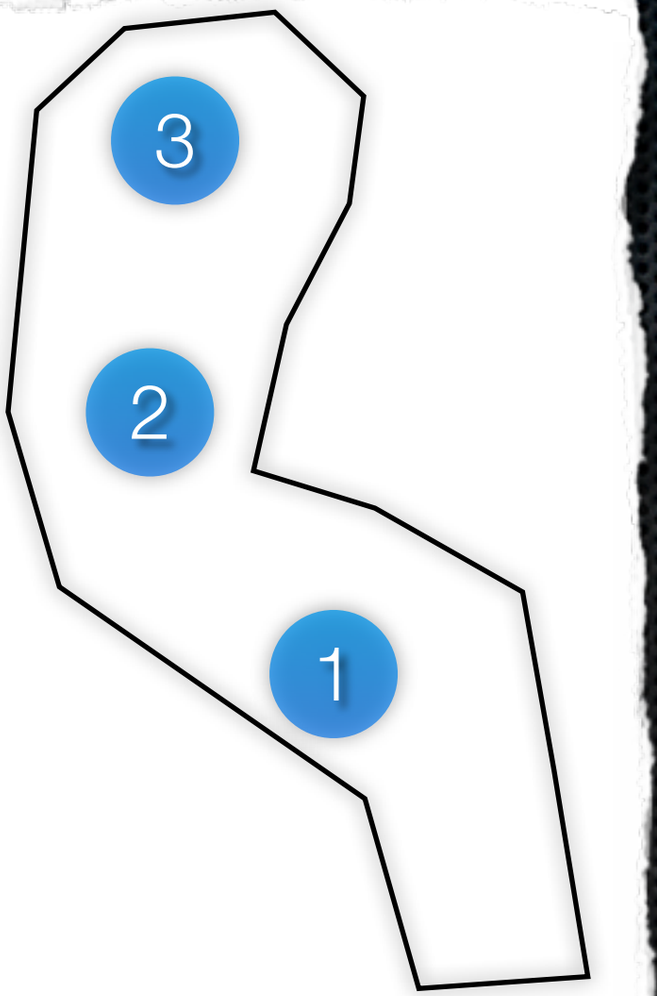


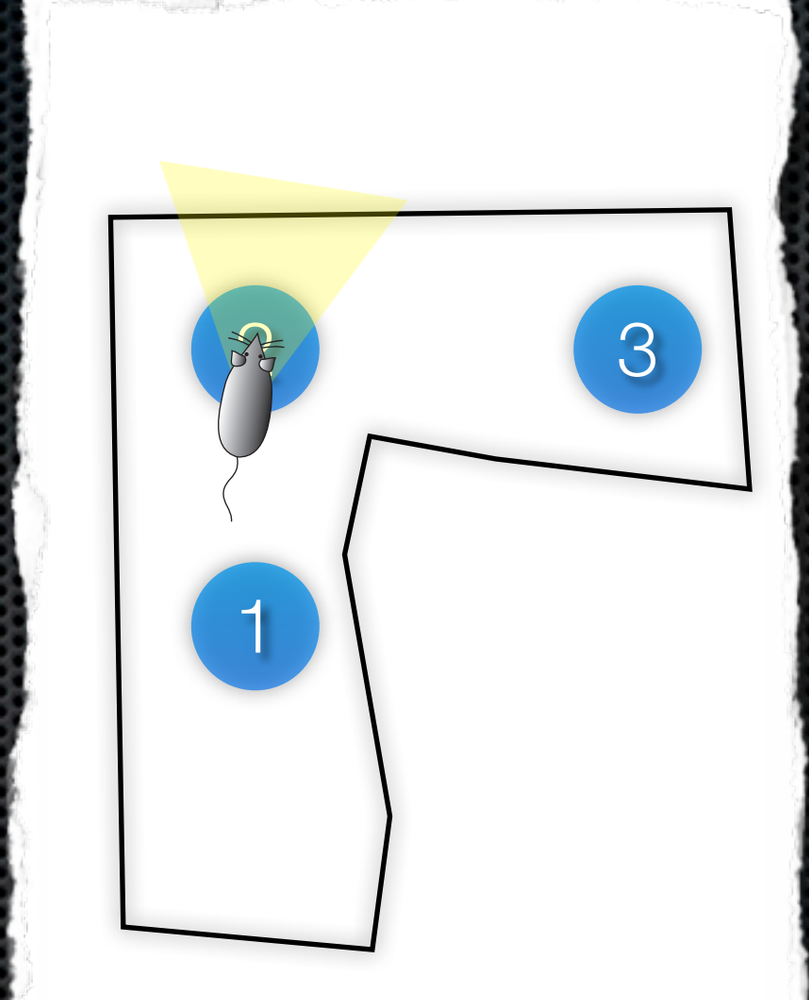
Stimulus-Approach



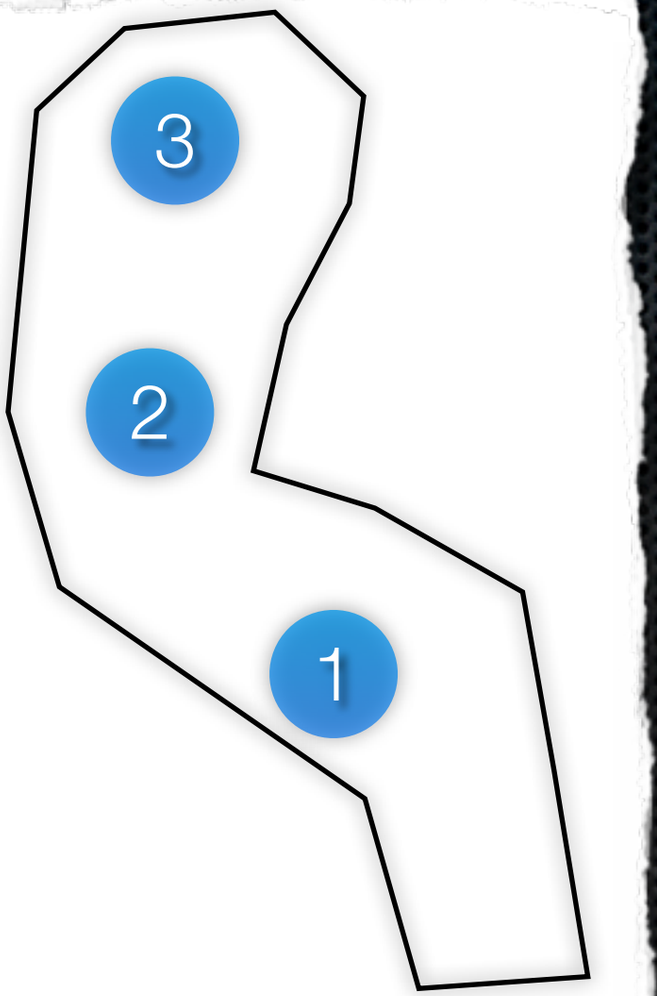


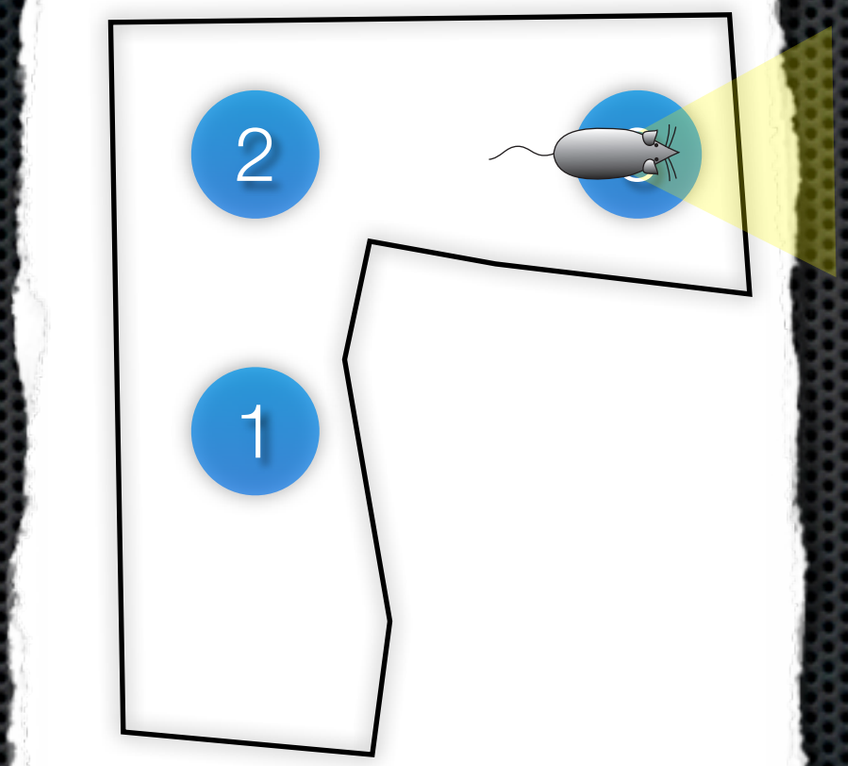
Stimulus-Approach



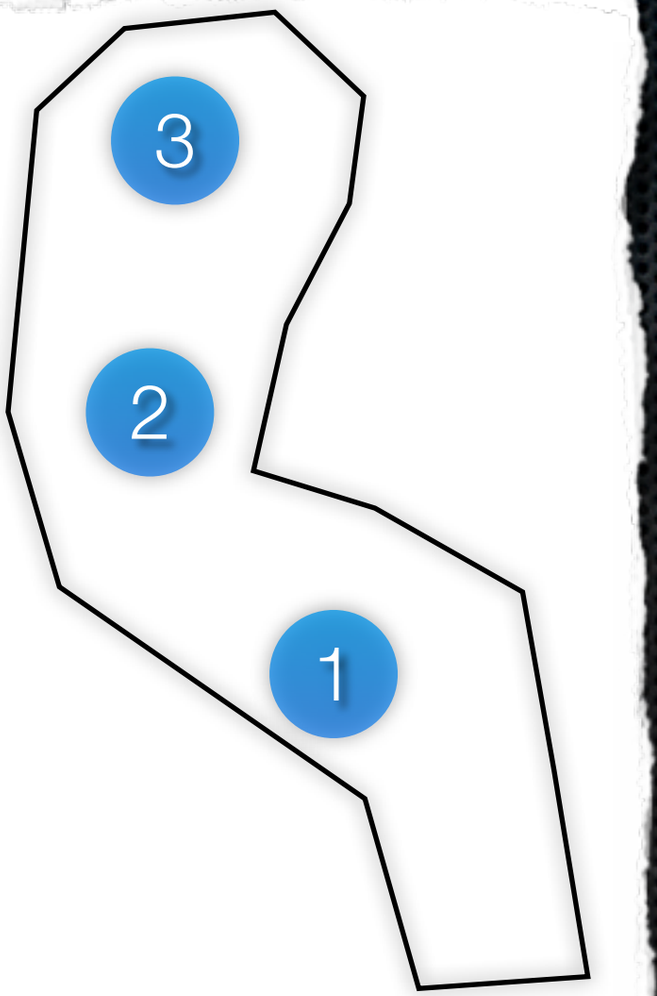


Stimulus-Approach

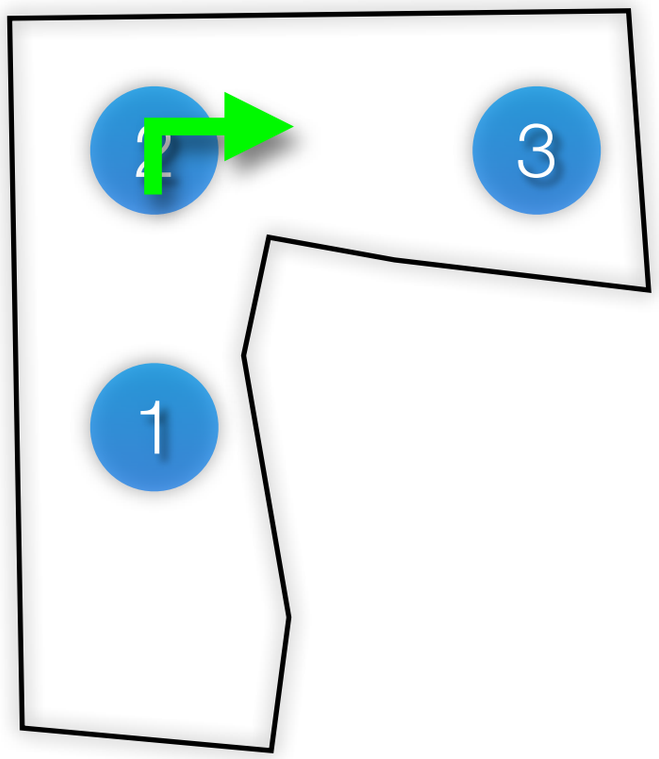




Stimulus-Approach

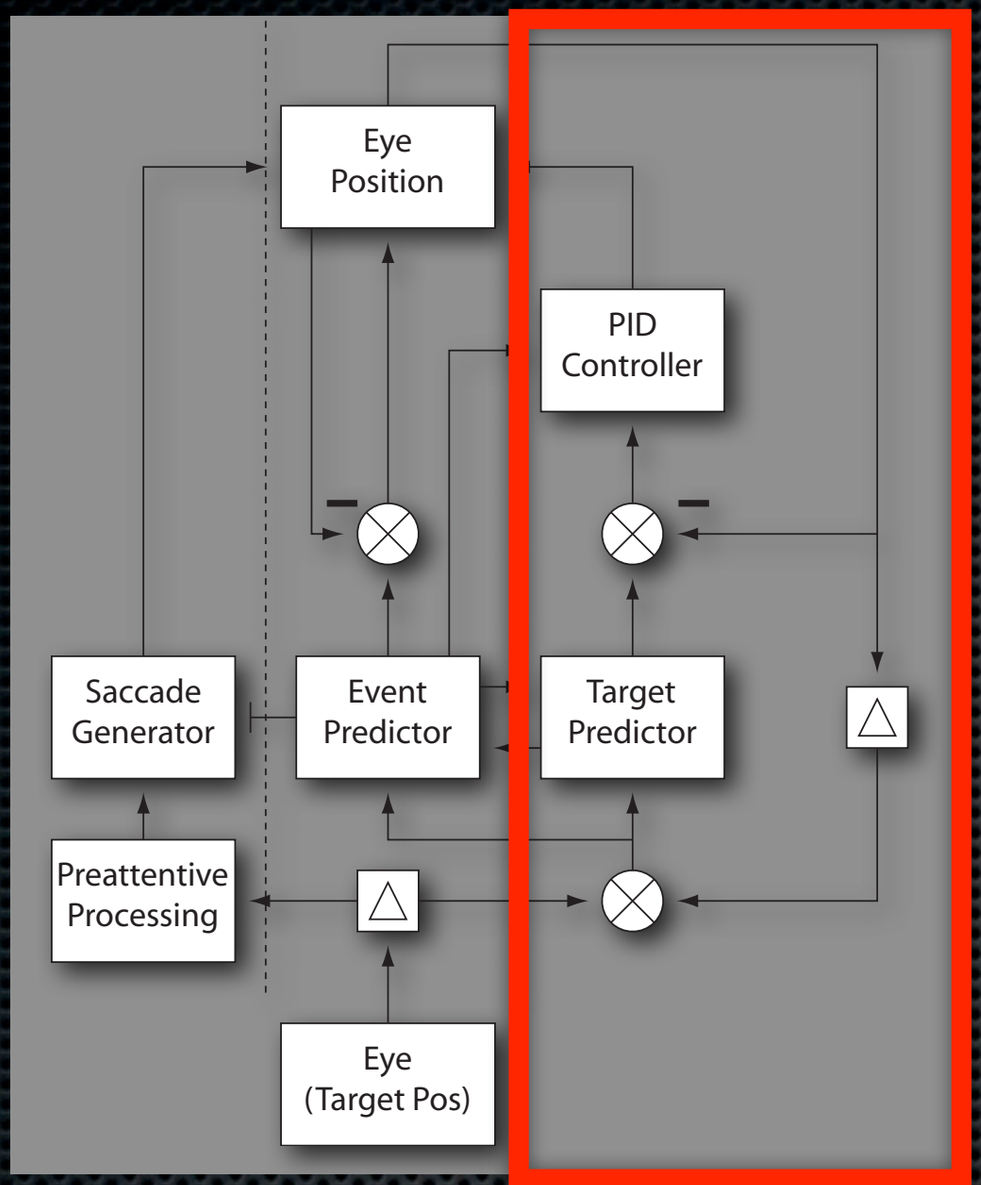


Stimulus-Approach



Stimulus-Response

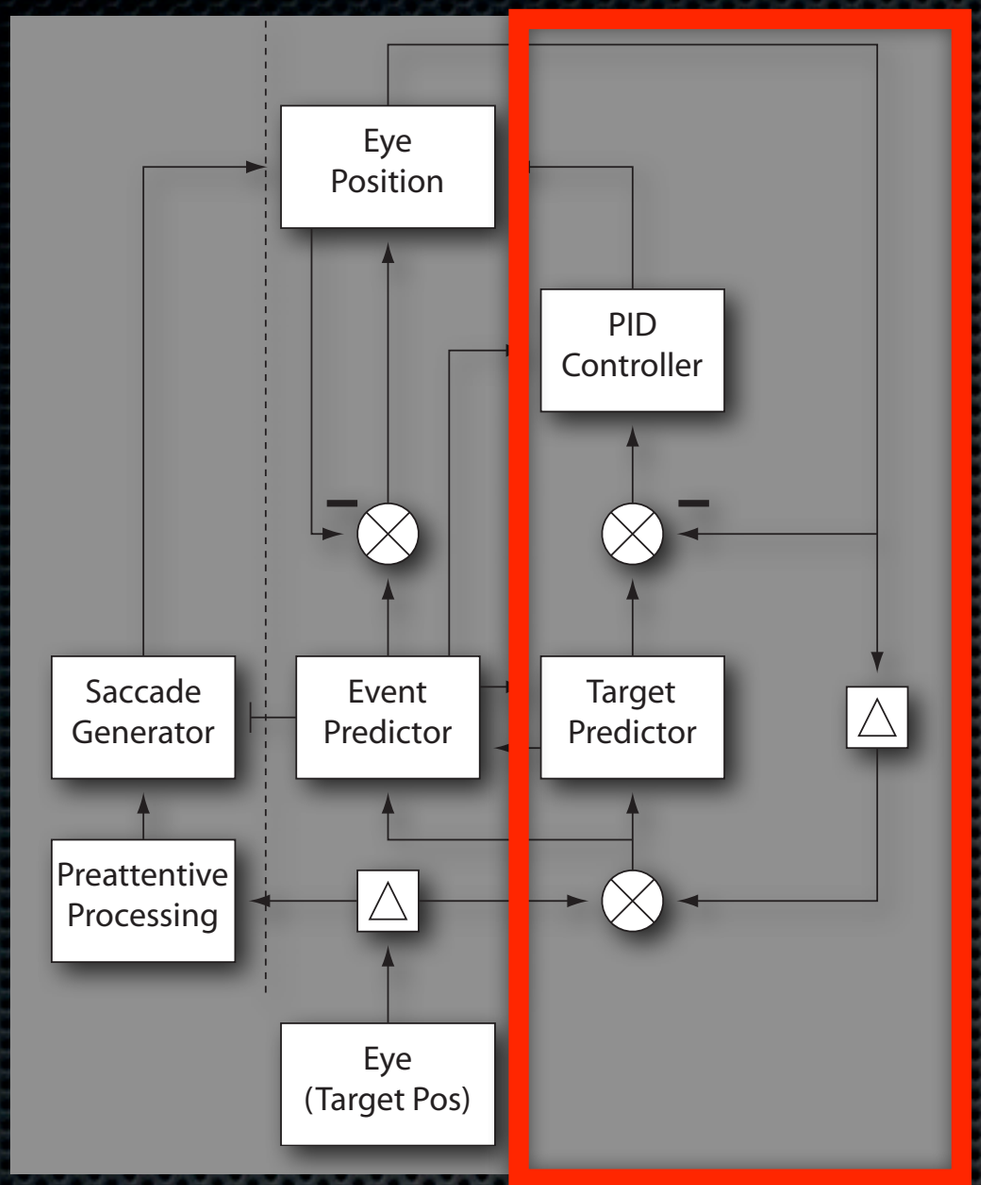
Smooth Pursuit

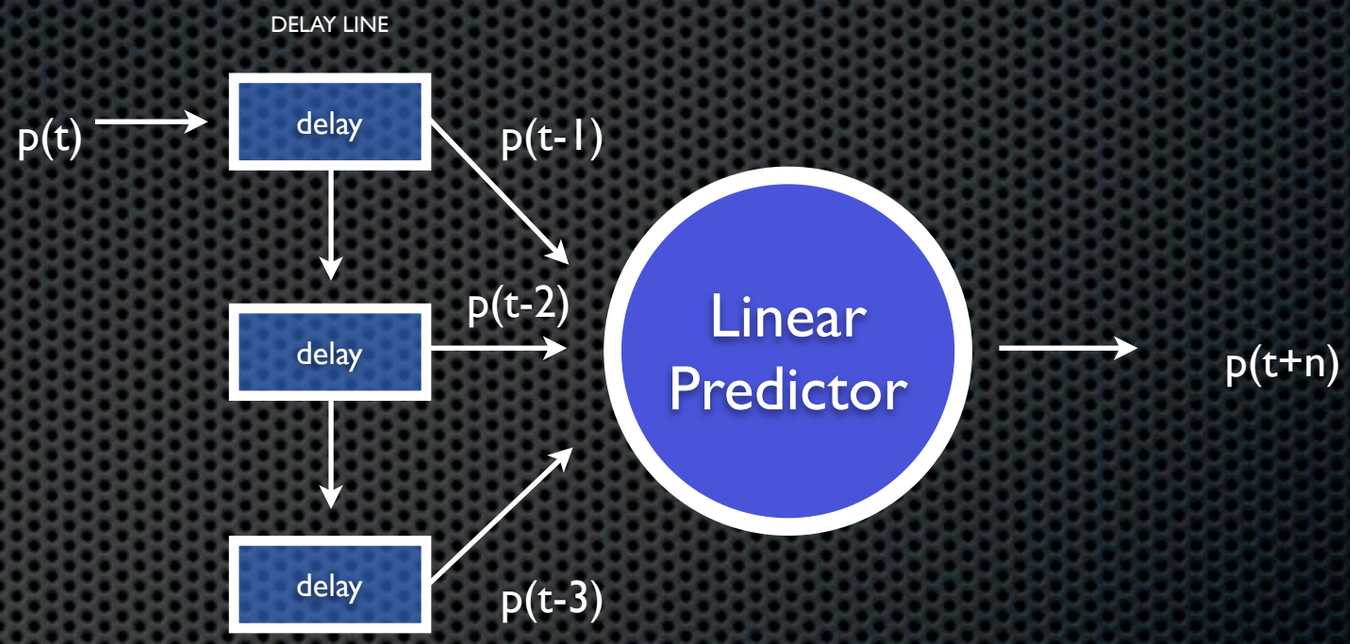


stimulus-approach

Balkenius & Johansson, 2005

Smooth Pursuit

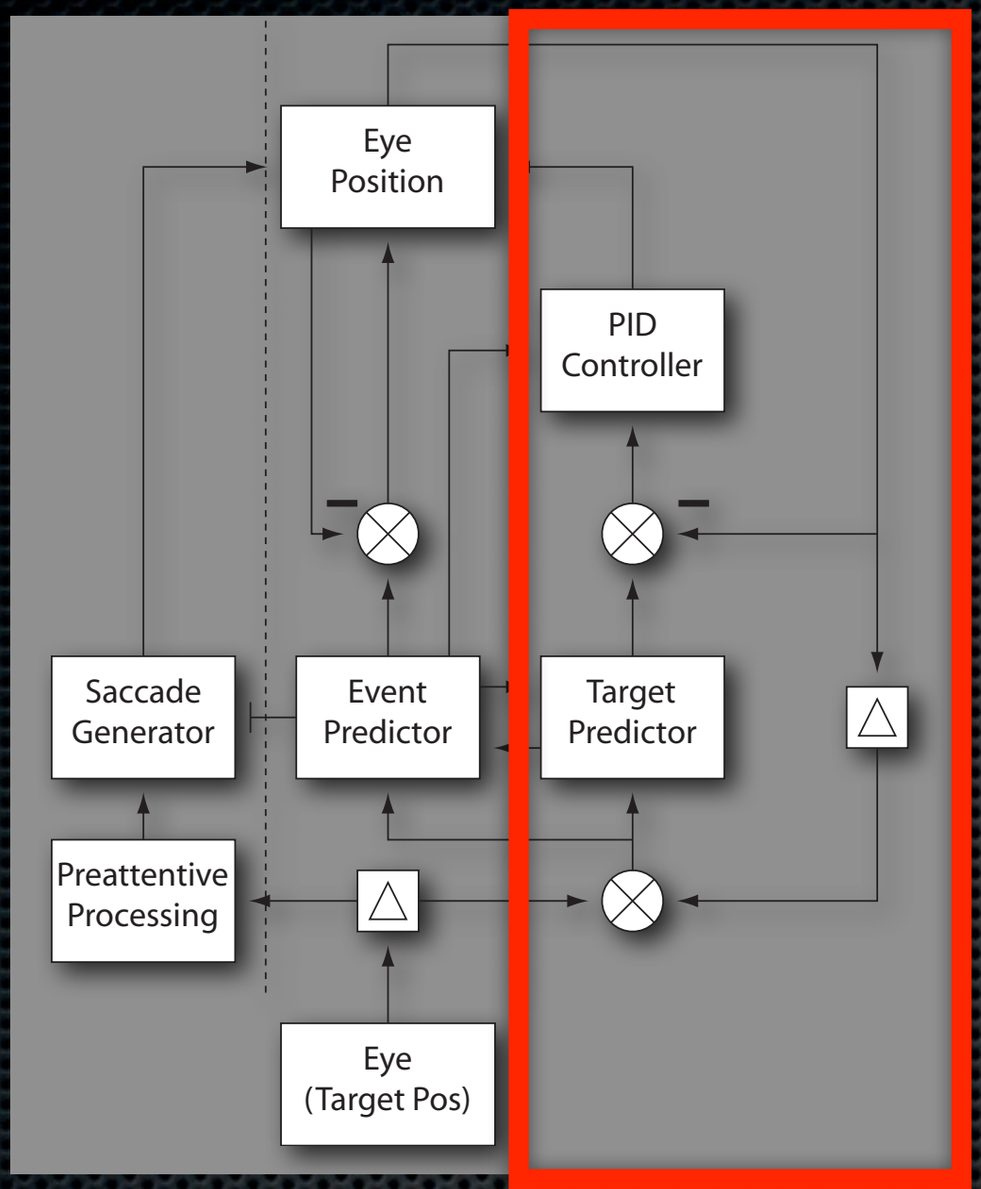




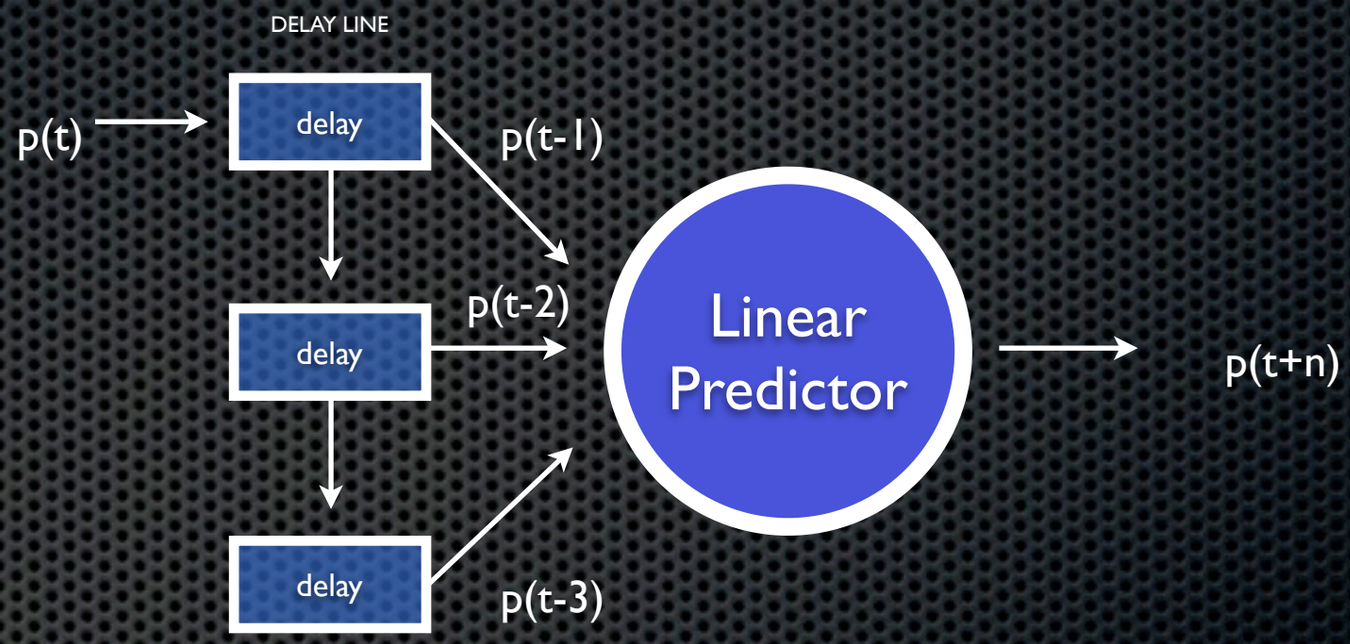
stimulus-approach

Balkenius & Johansson, 2005

Smooth Pursuit



stimulus-approach



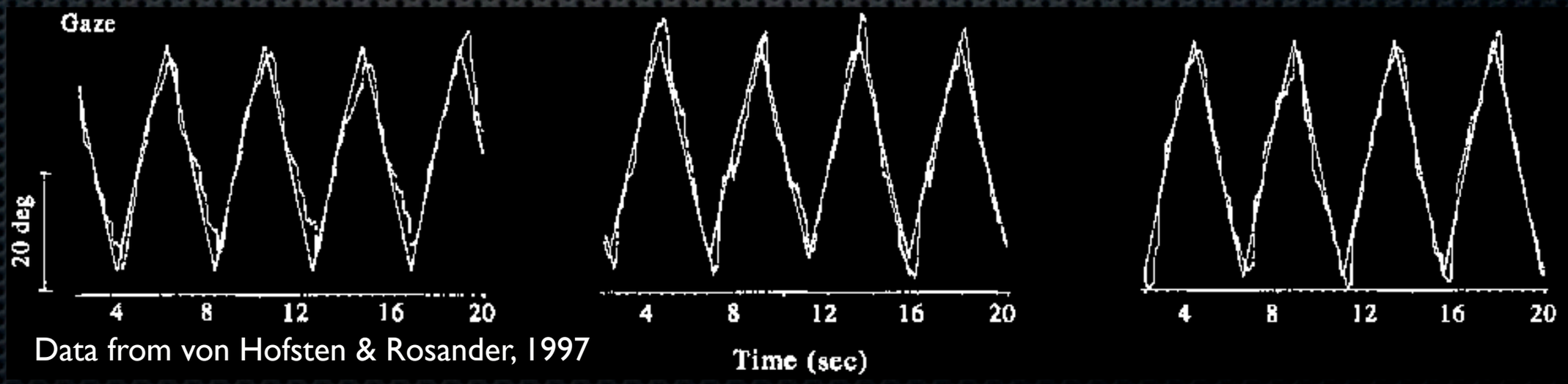
Prediction confidence sets the gain of the controller

Balkenius & Johansson, 2005

The Development of Smooth Pursuit

- Gradual development from catch up saccades to smooth pursuit from 0-4 month

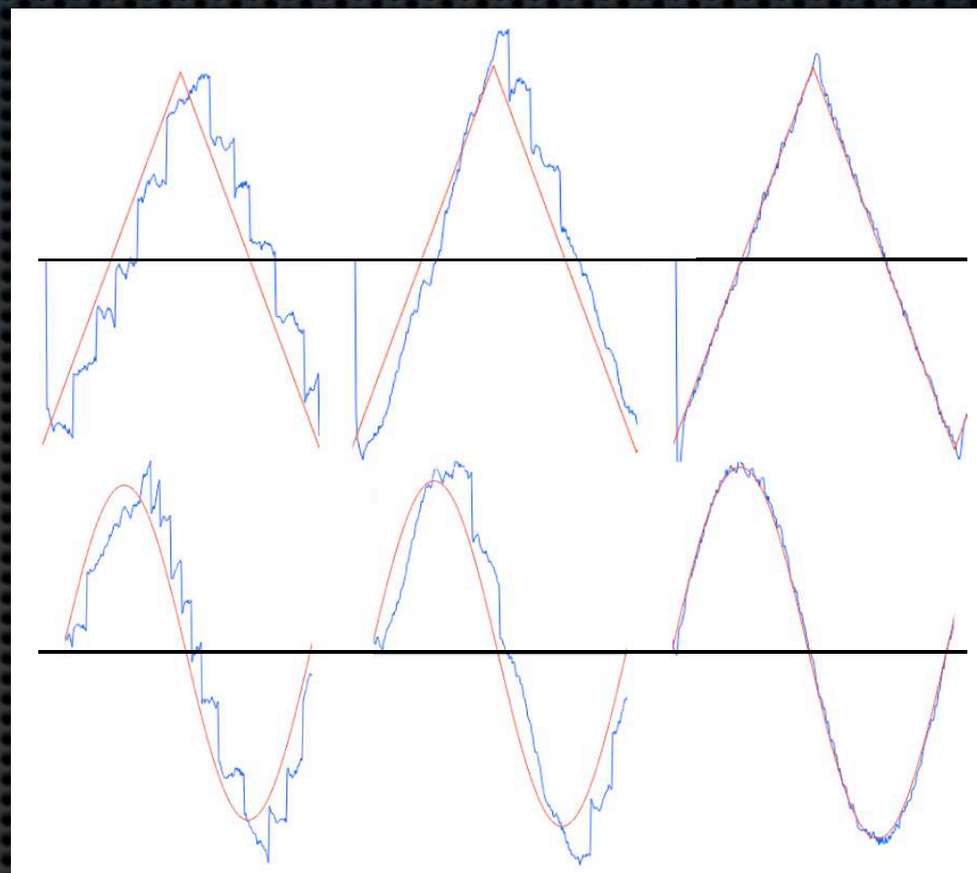
Simulation from Balkenius & Johansson, Epigenetic Robotics, 2005



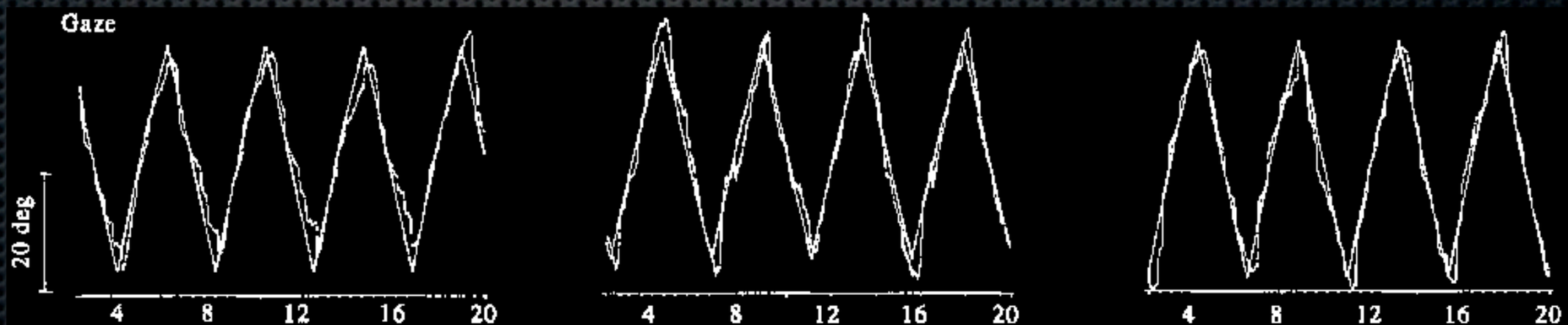
Data from von Hofsten & Rosander, 1997

The Development of Smooth Pursuit

- Gradual development from catch up saccades to smooth pursuit from 0-4 month

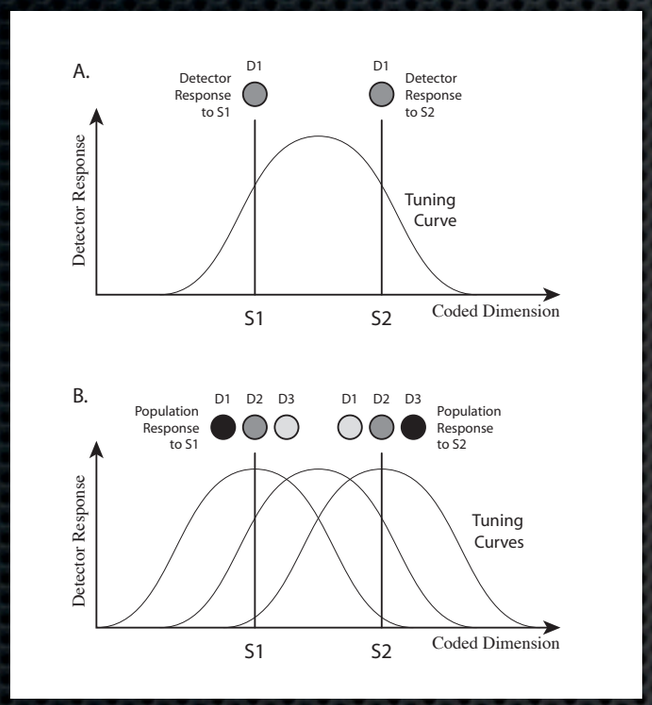


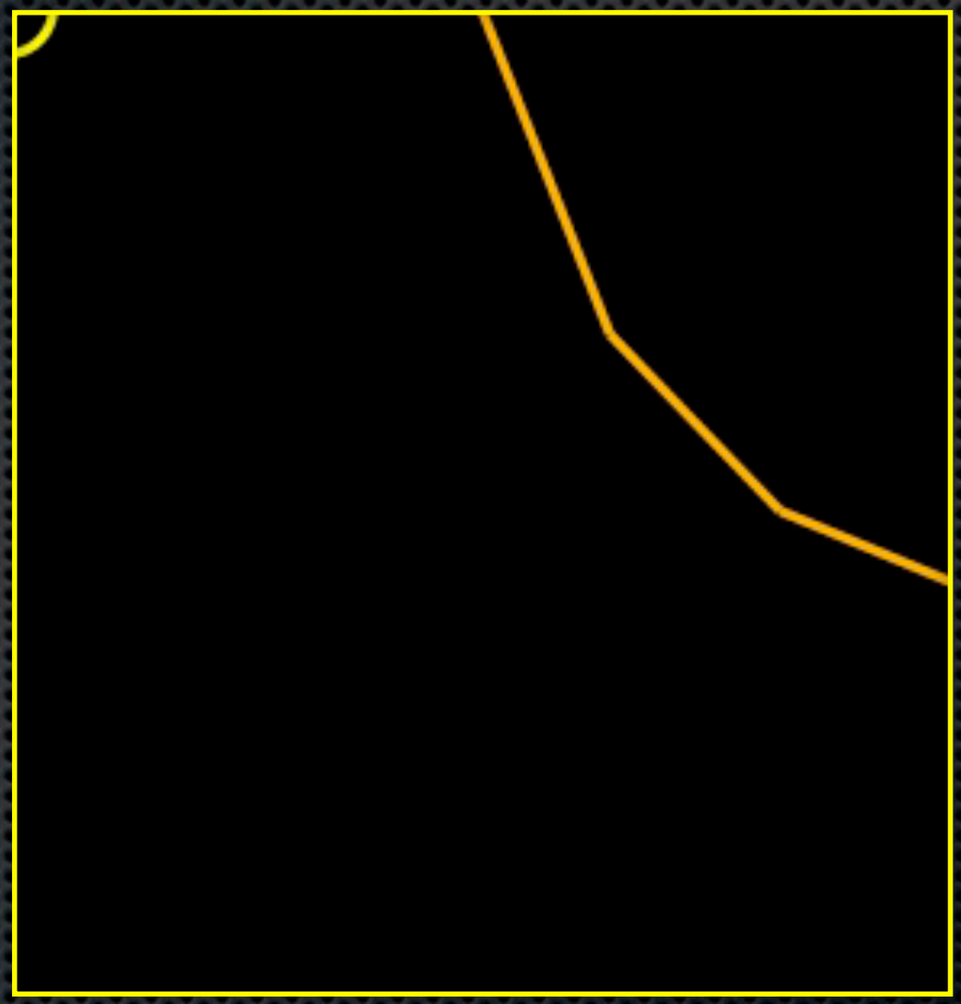
Simulation from Balkenius & Johansson, Epigenetic Robotics, 2005



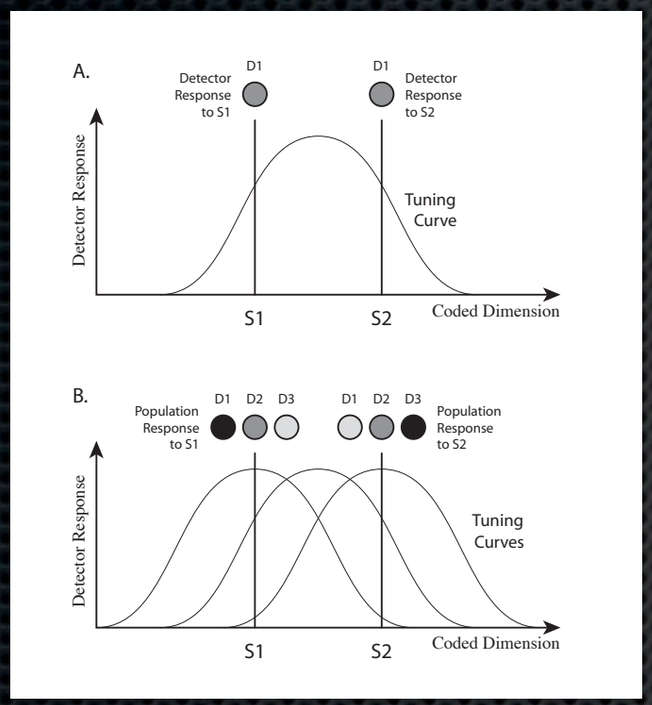
Data from von Hofsten & Rosander, 1997

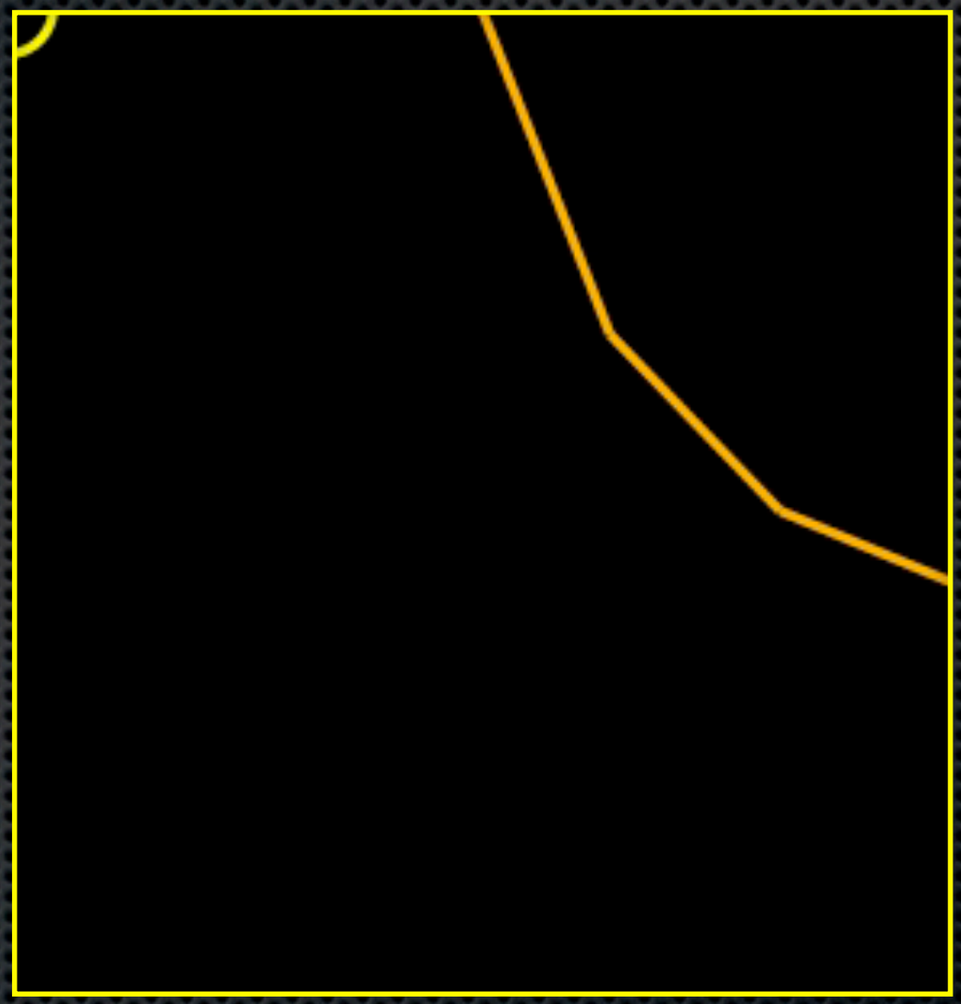
Learning to Reach





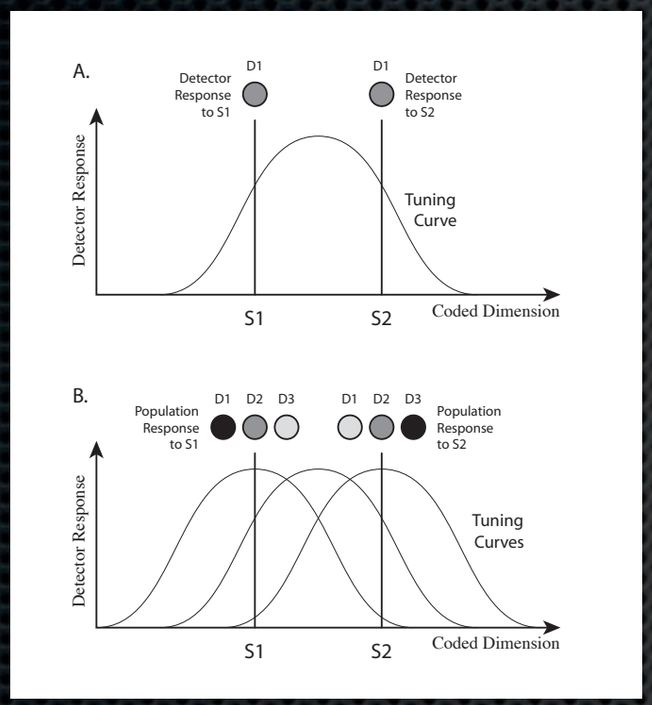
Learning to Reach

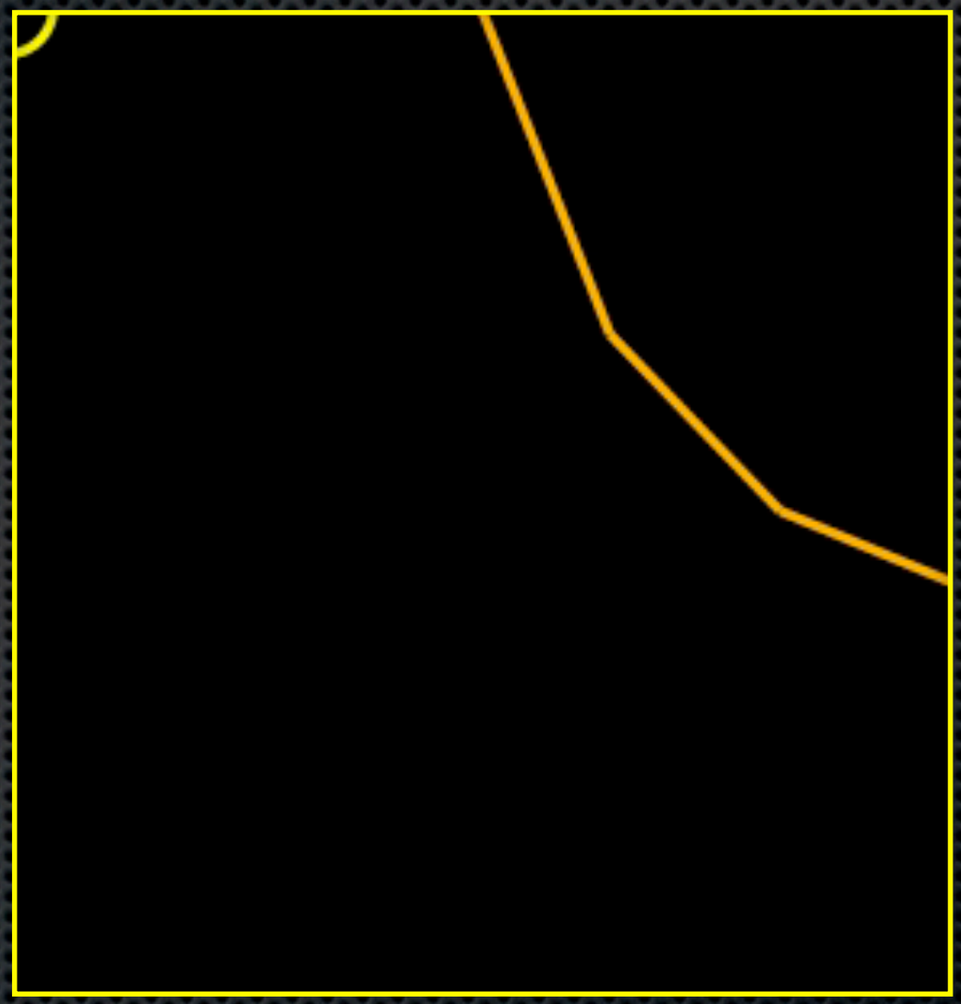




the system learns associations between retinal positions and joint angles

Learning to Reach

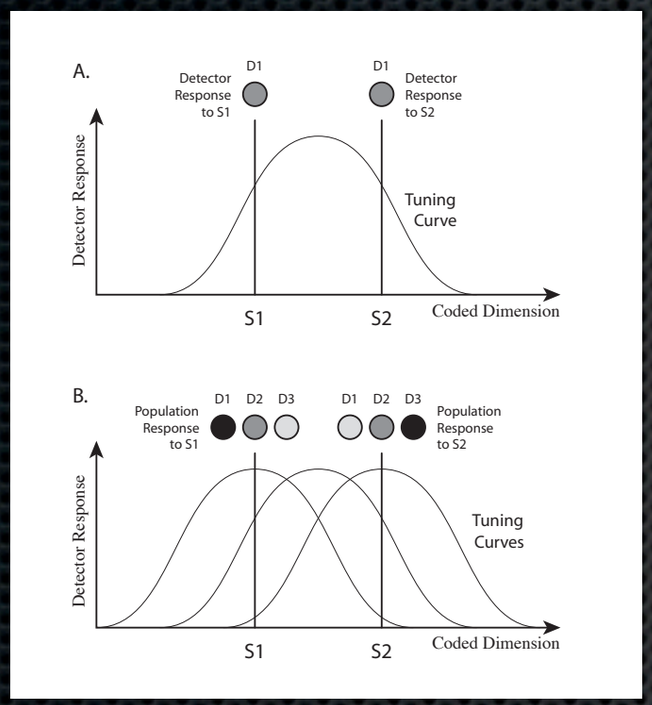


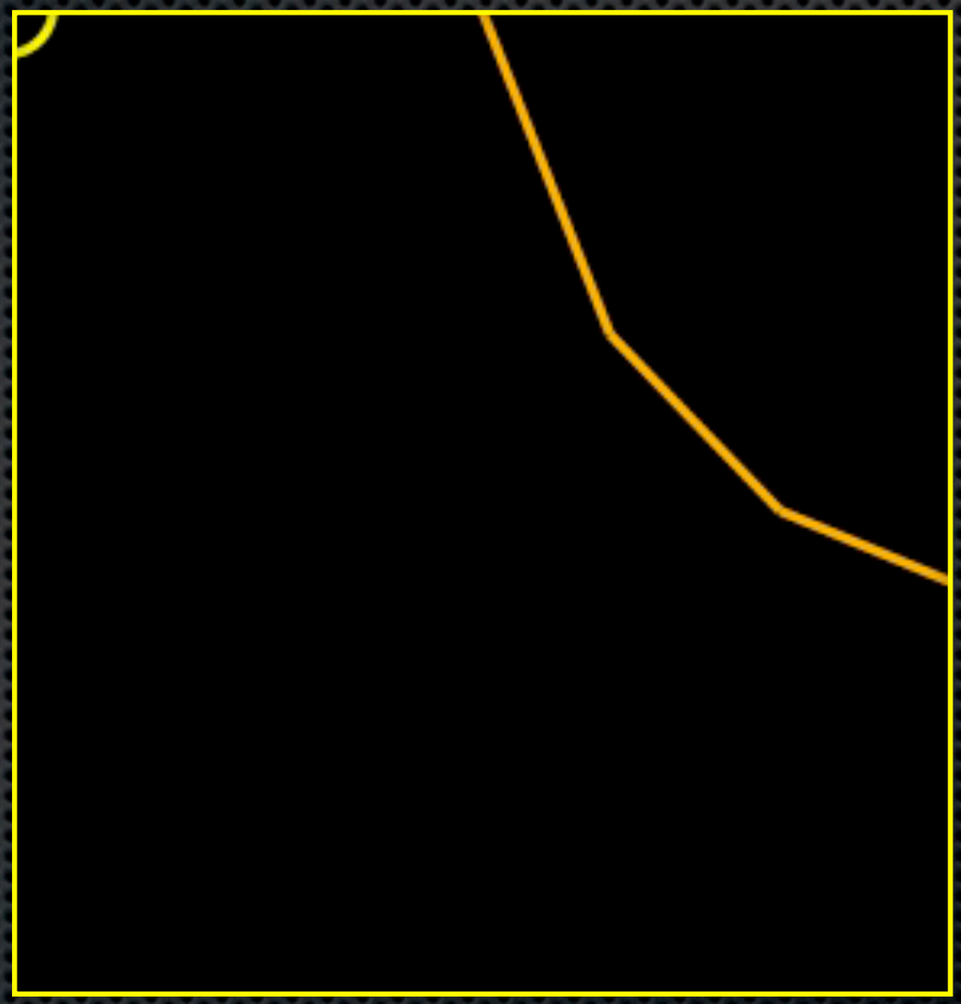


the system learns associations between retinal positions and joint angles

not driven by error between hand position and target

Learning to Reach





the system learns associations between retinal positions and joint angles

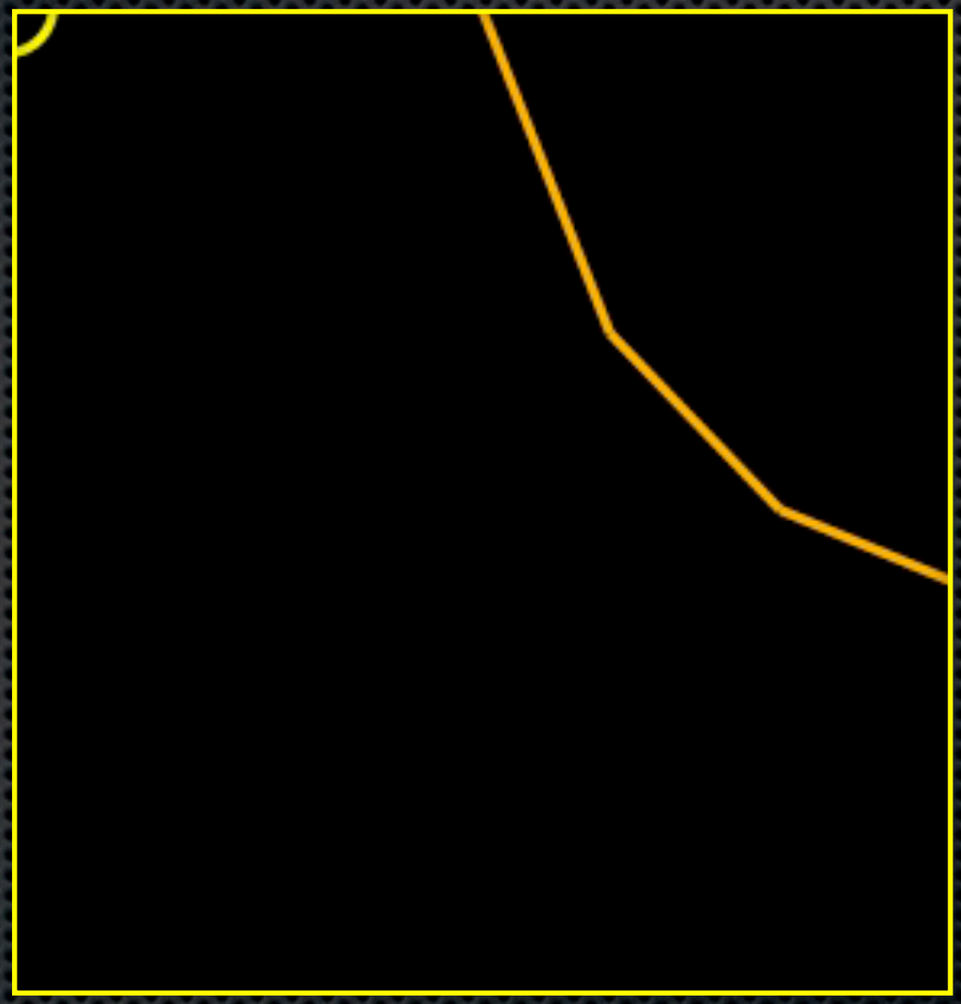
not driven by error between hand position and target

population coding supports interpolation and some extrapolation

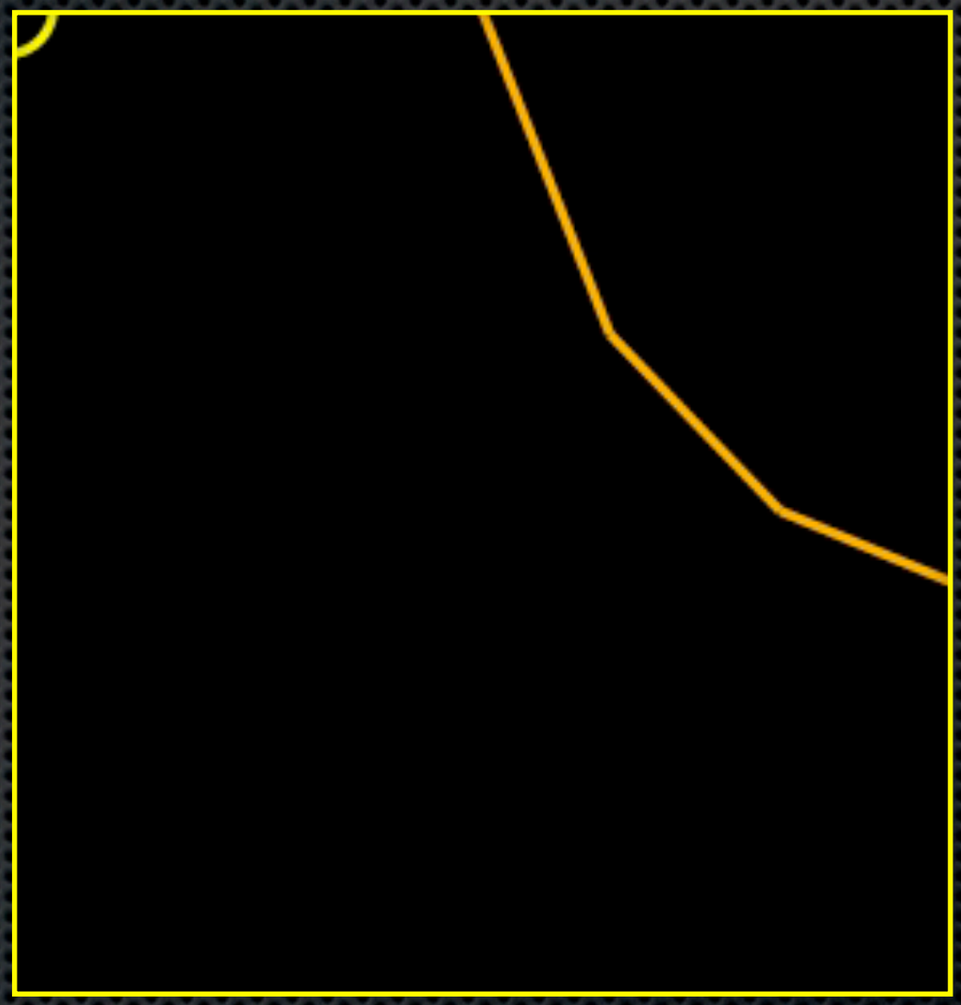
Learning to Reach

sensory prediction
+
inverse model

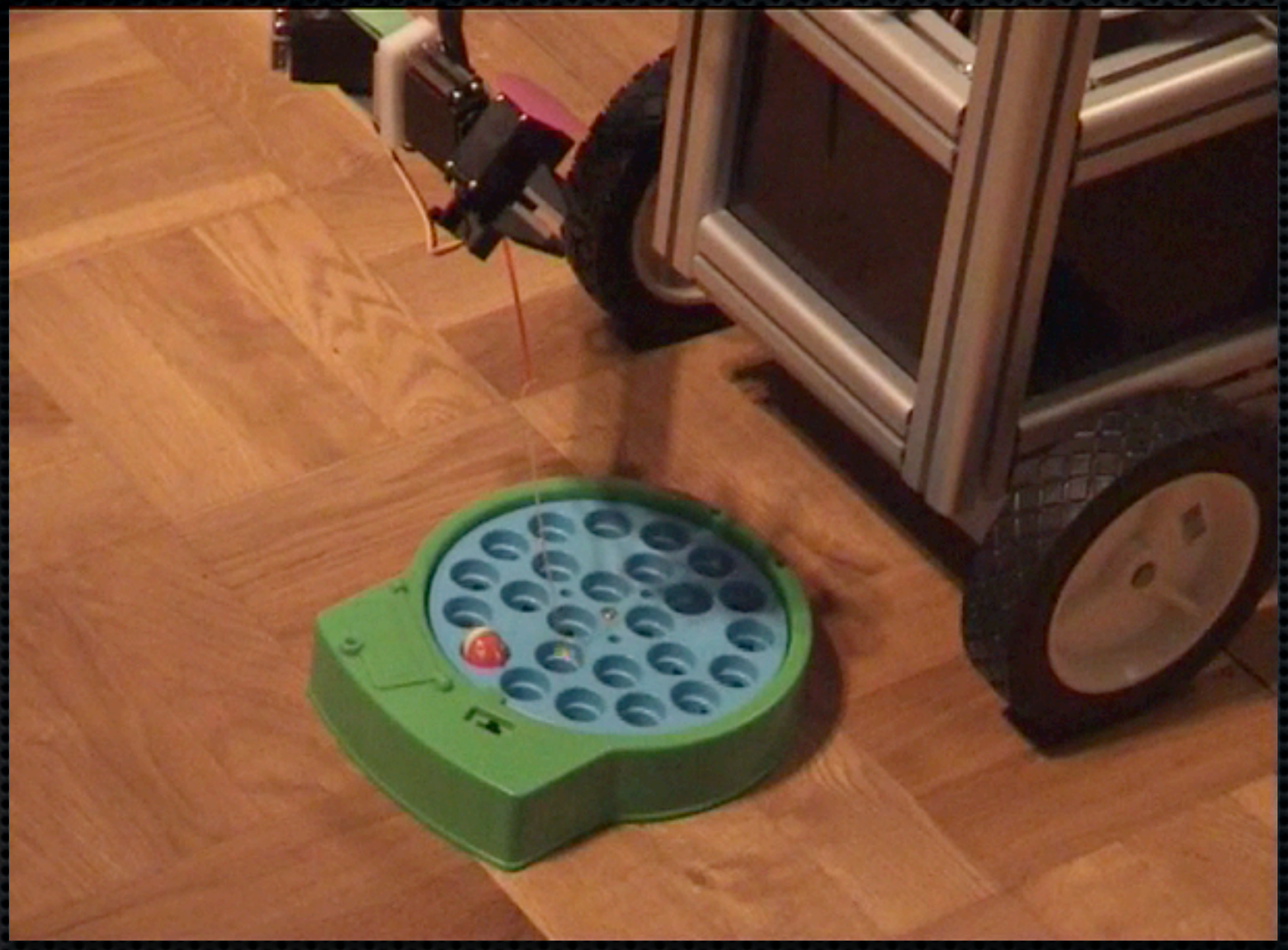
Learning to Reach



Learning to Reach



ongoing interaction



Fishing

0 min

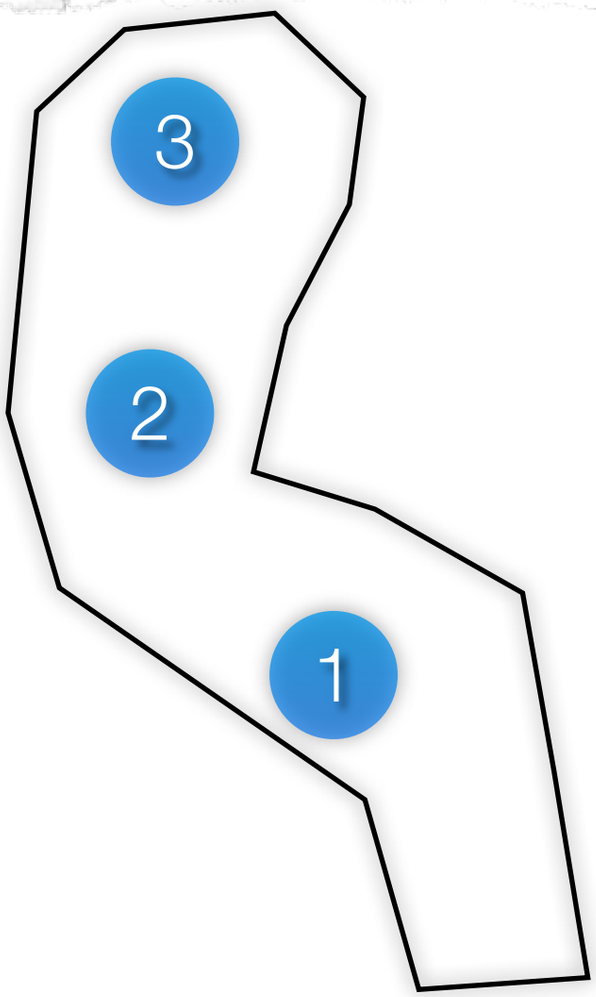
500 ms sensory-motor delay

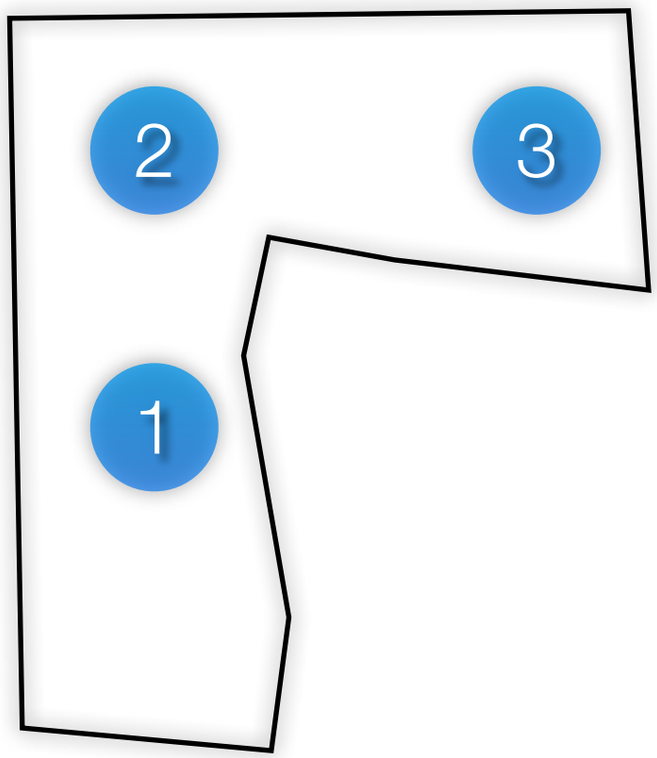
Fishing

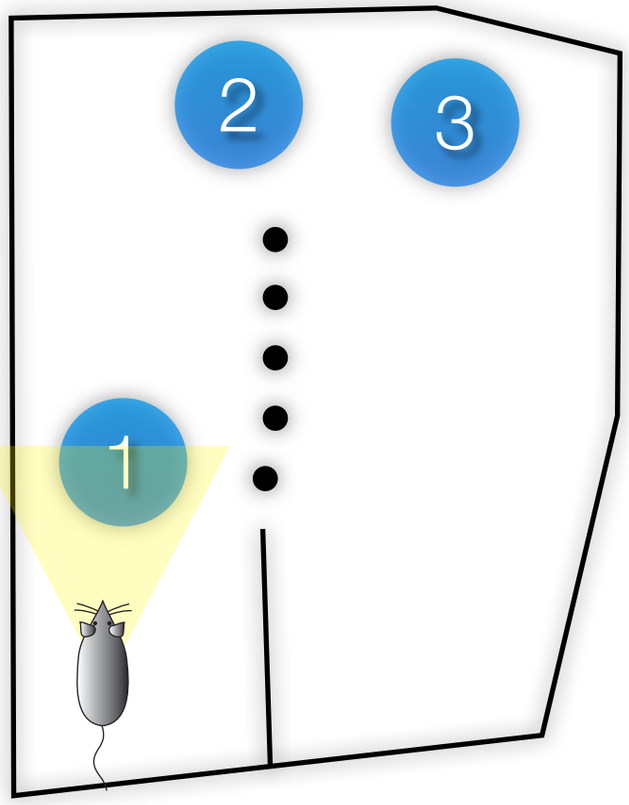
2 min

Fishing

5 min

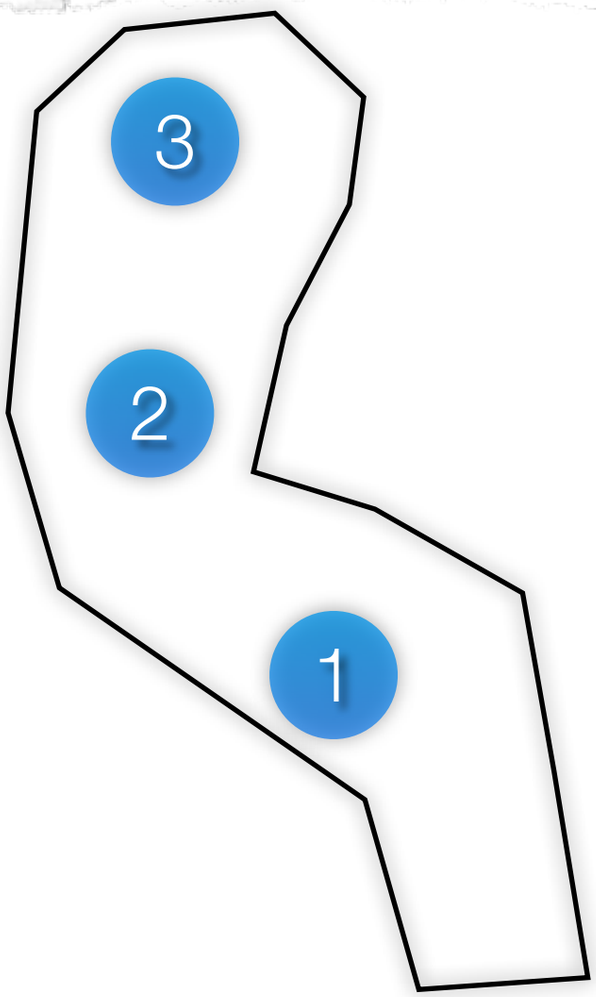




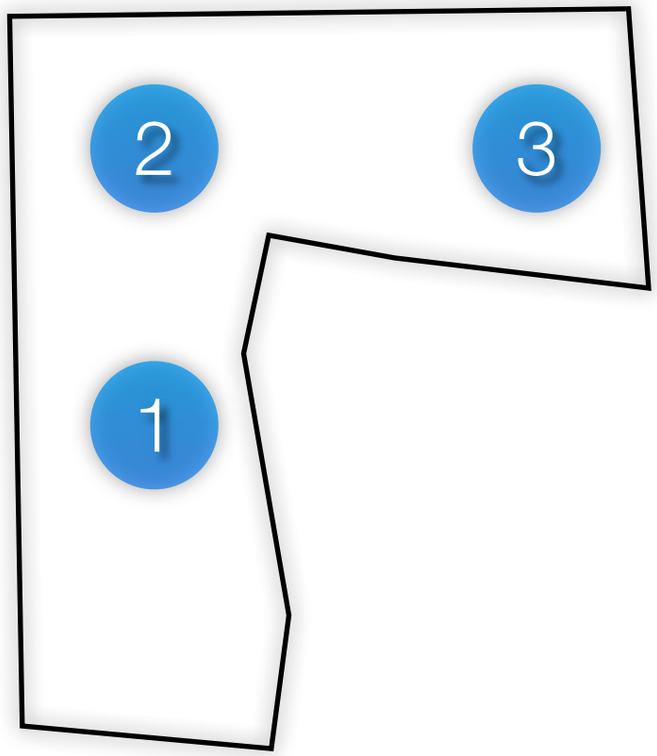


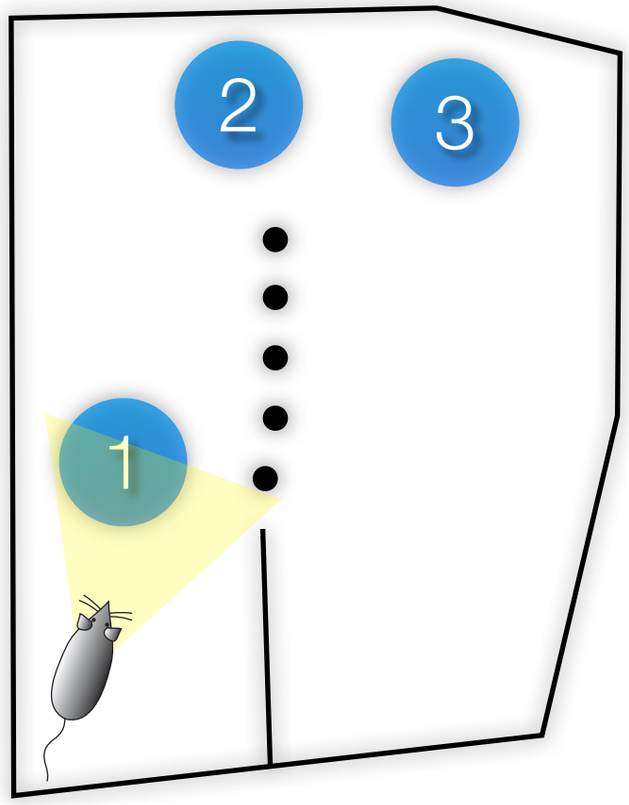
Stimulus-Approach

Stimulus-Response

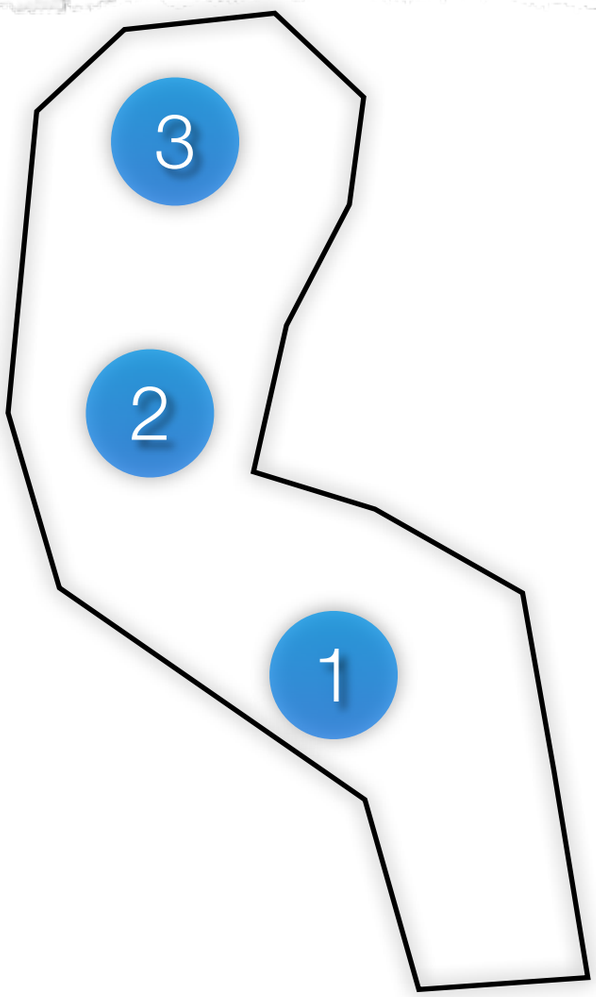


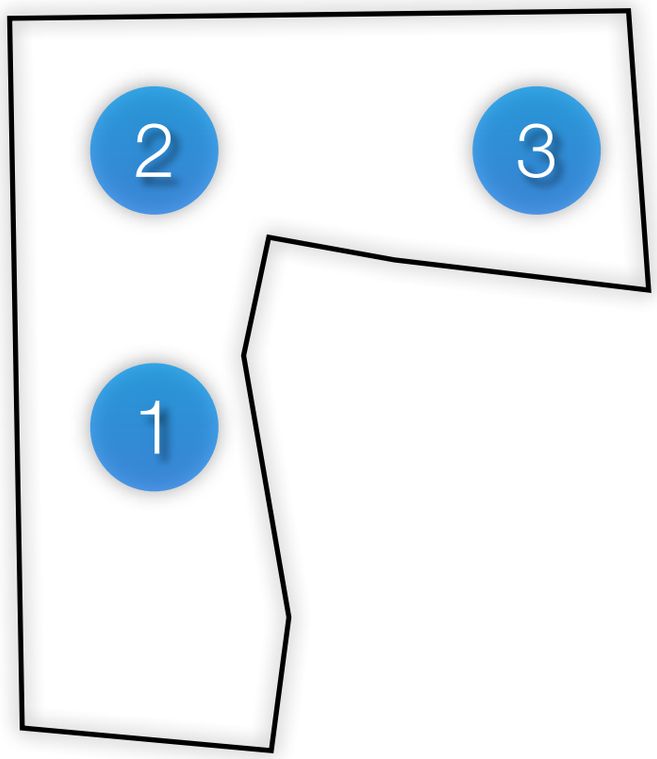
Stimulus-Approach

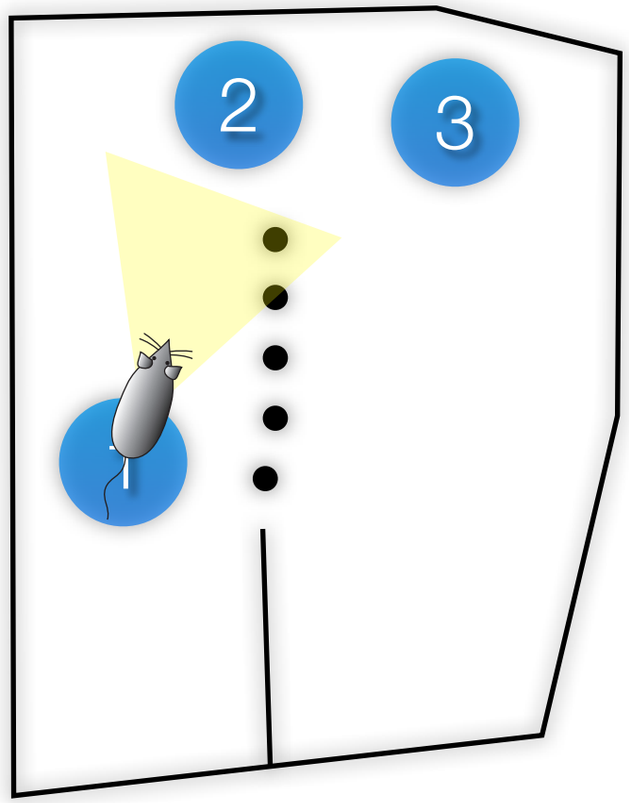




Stimulus-Response

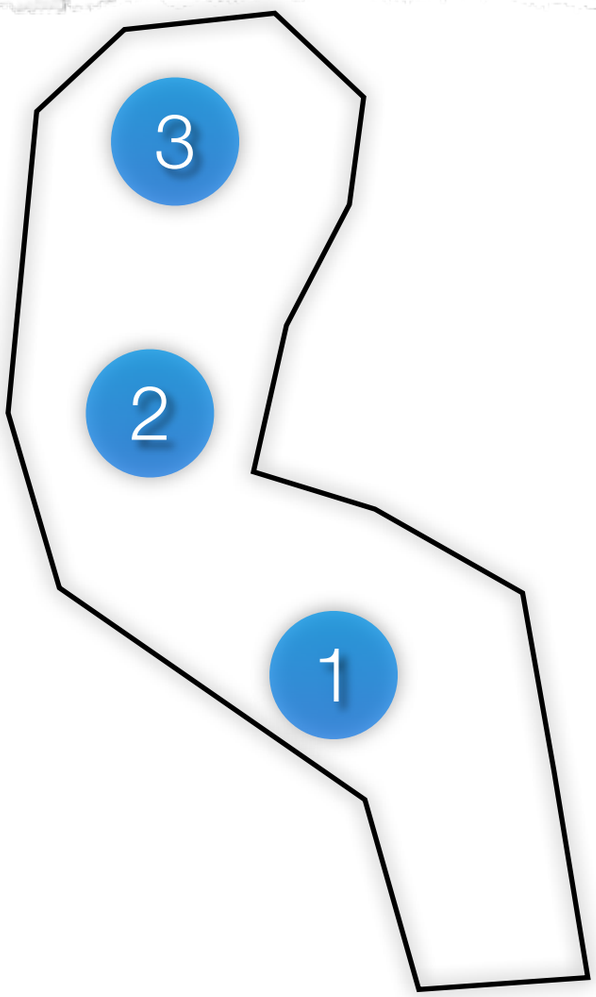


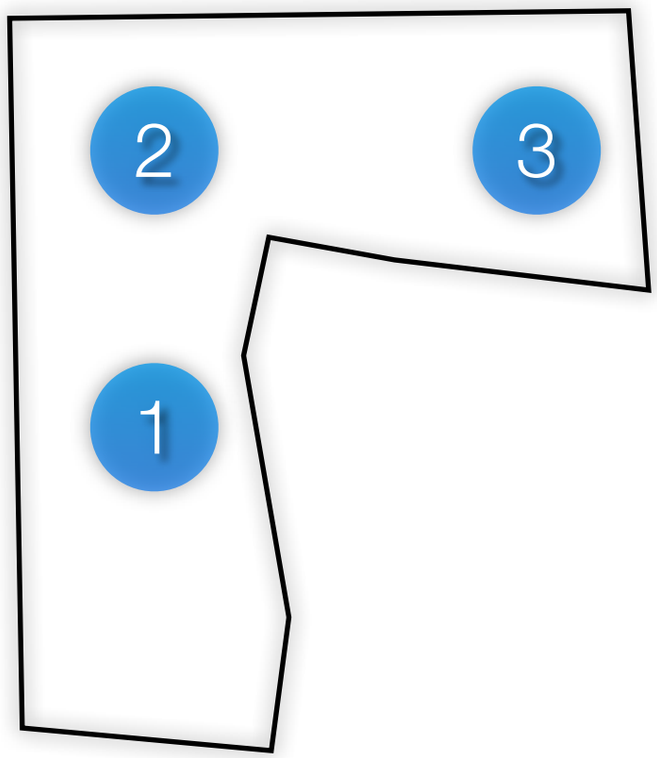


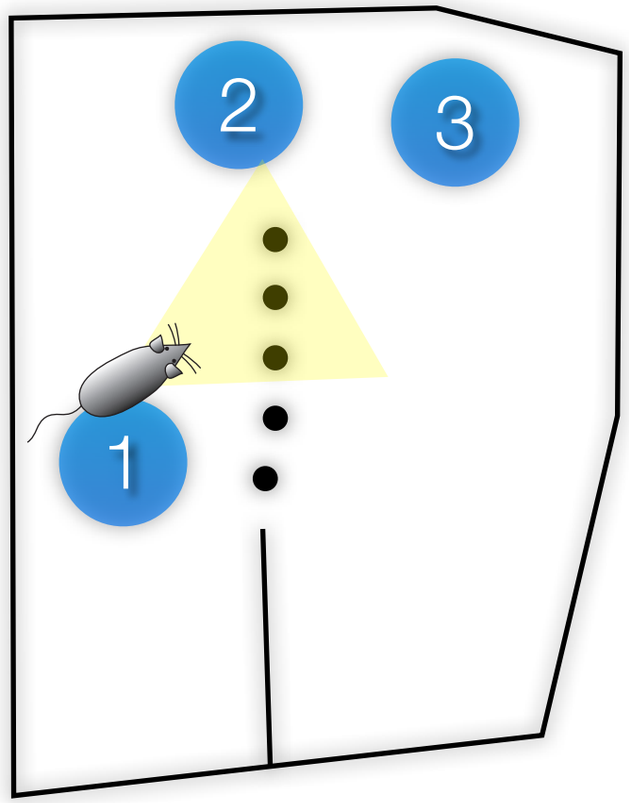


Stimulus-Approach

Stimulus-Response

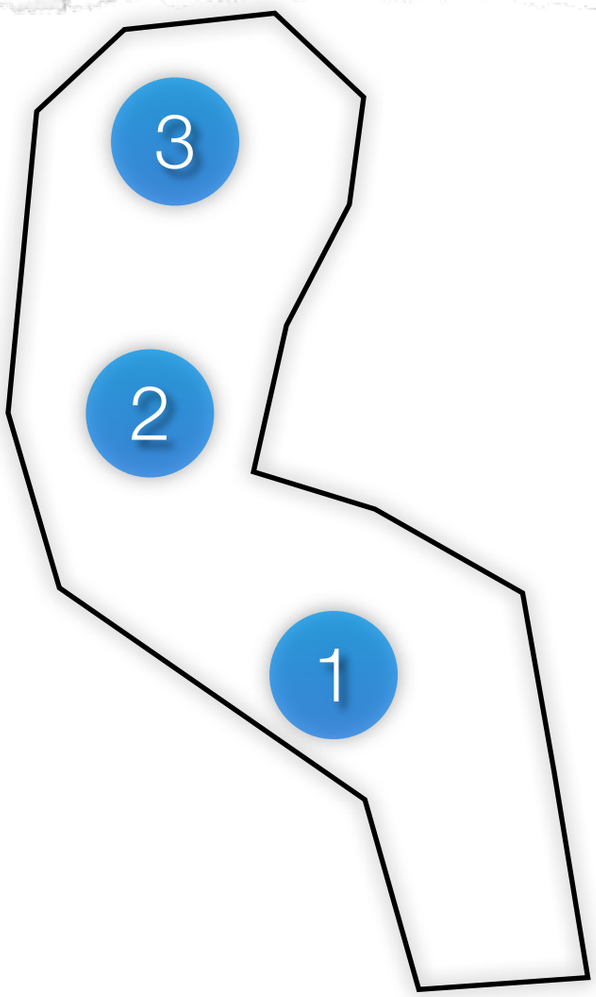




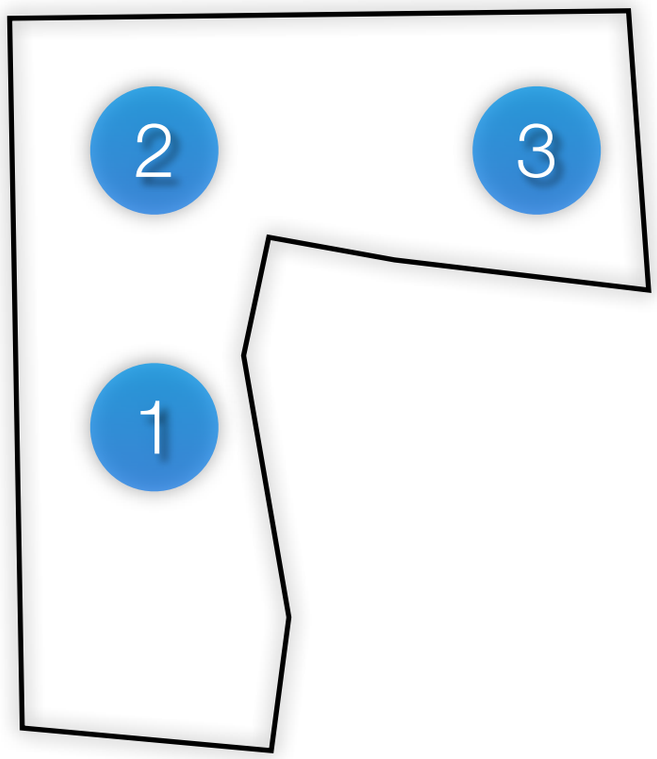


Stimulus-Approach

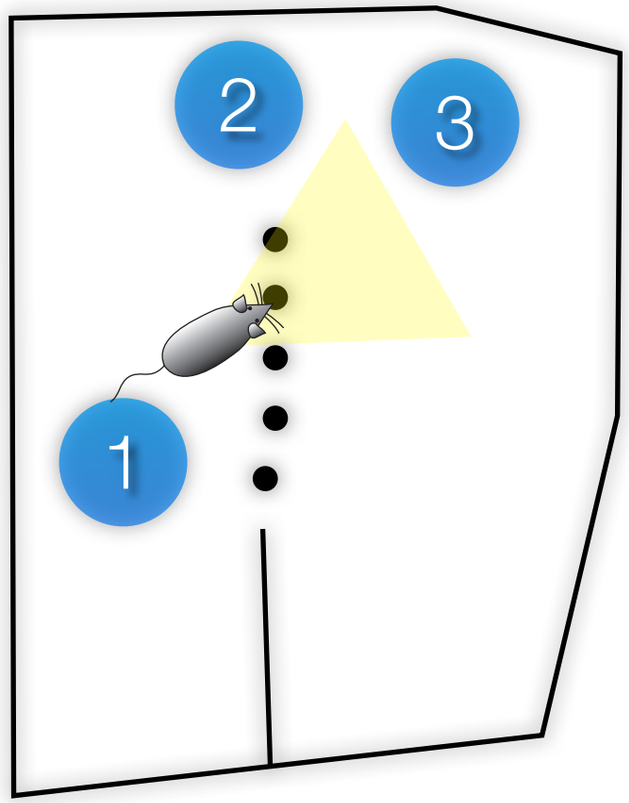
Stimulus-Response

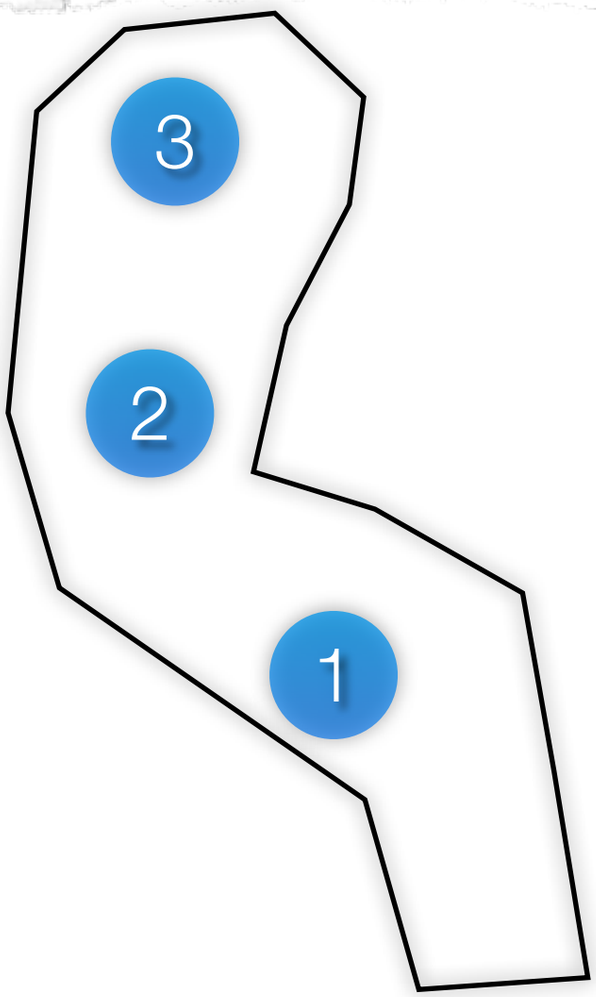


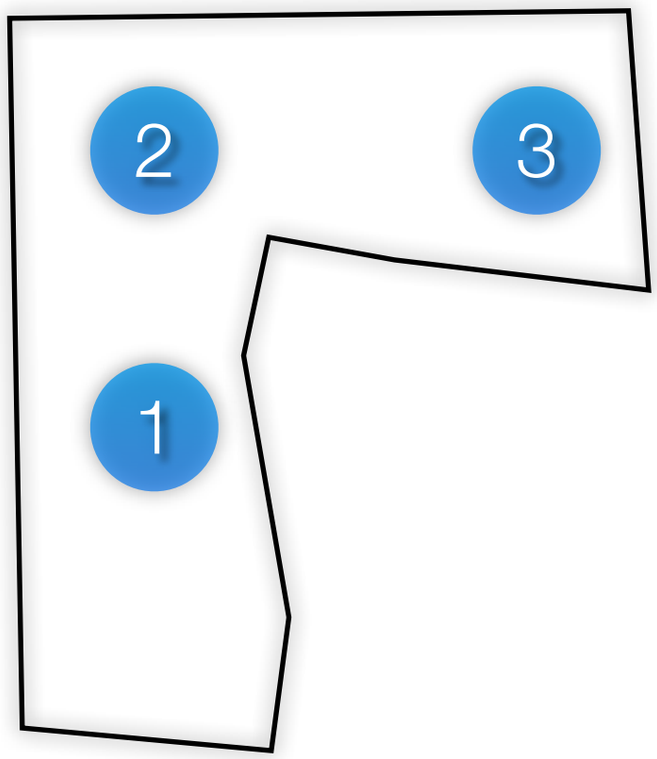
Stimulus-Approach

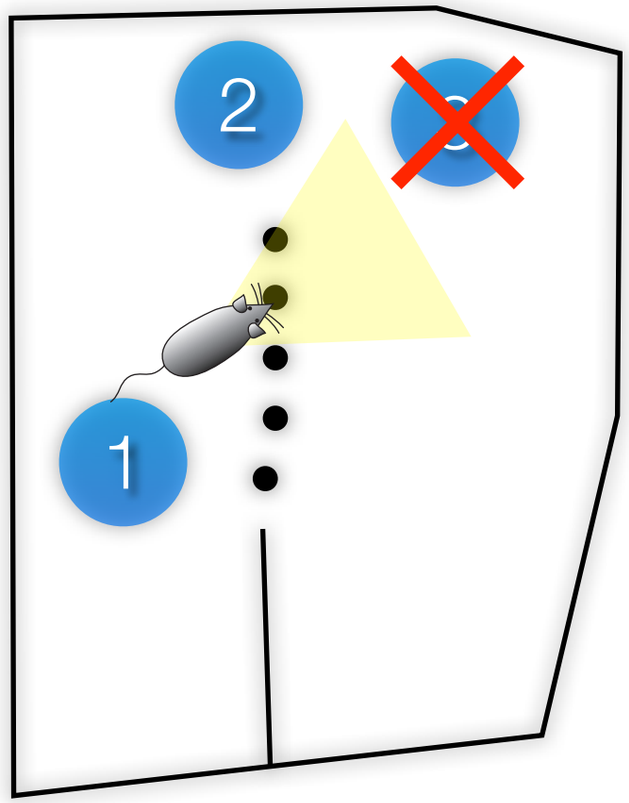


Stimulus-Response



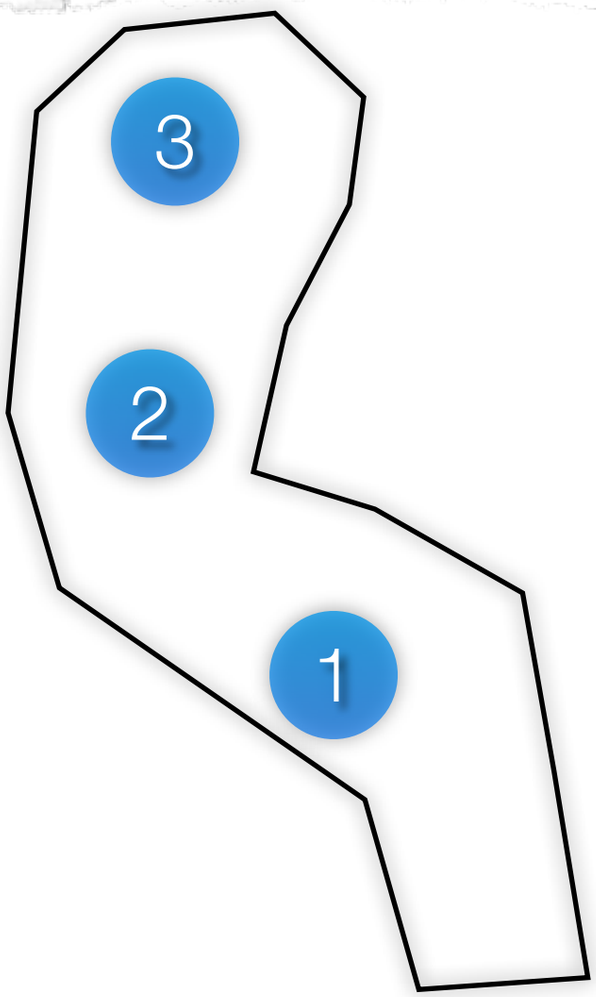


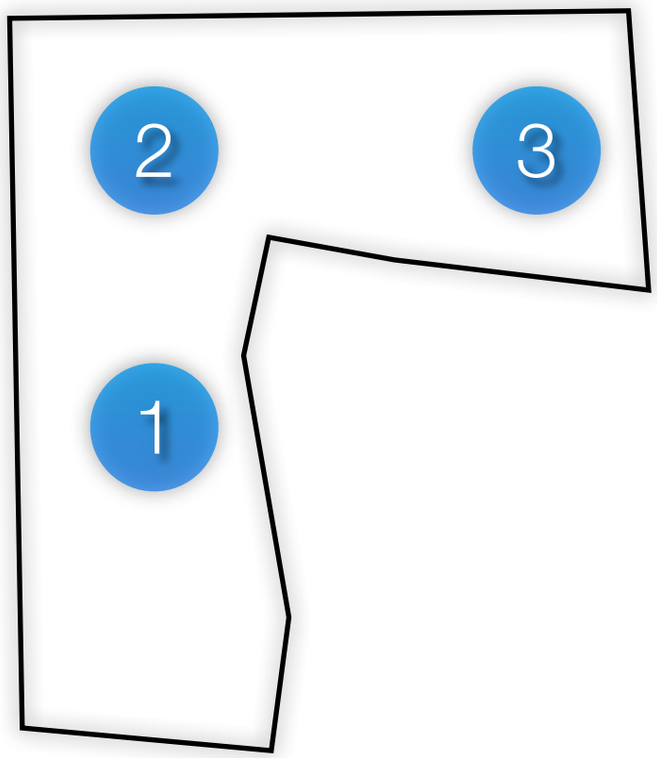


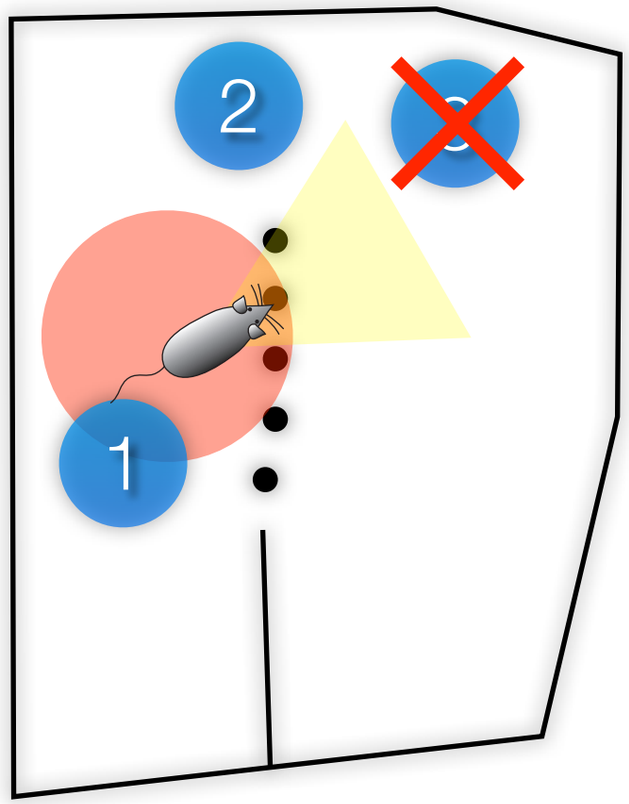


Stimulus-Approach

Stimulus-Response

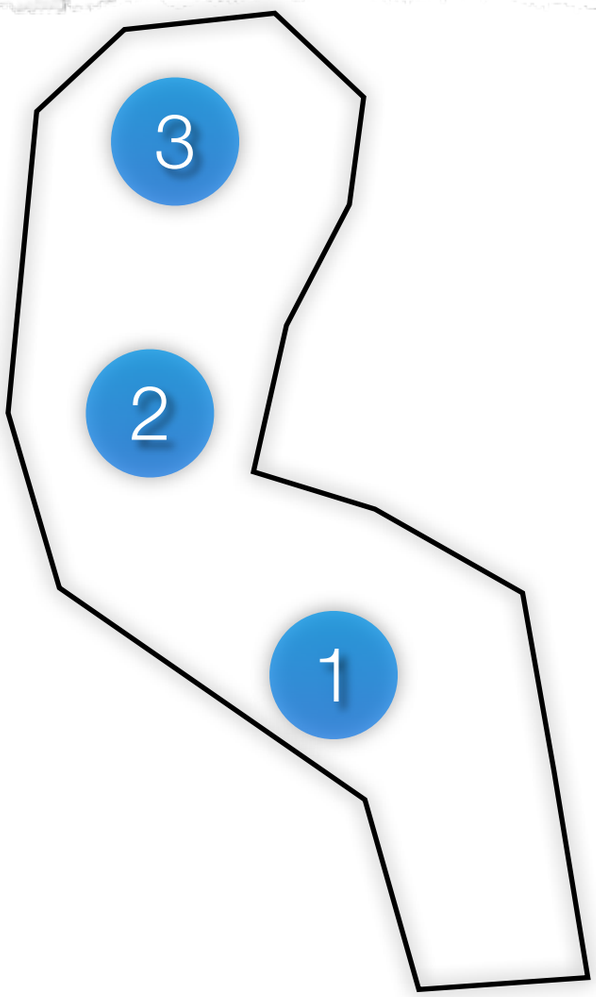




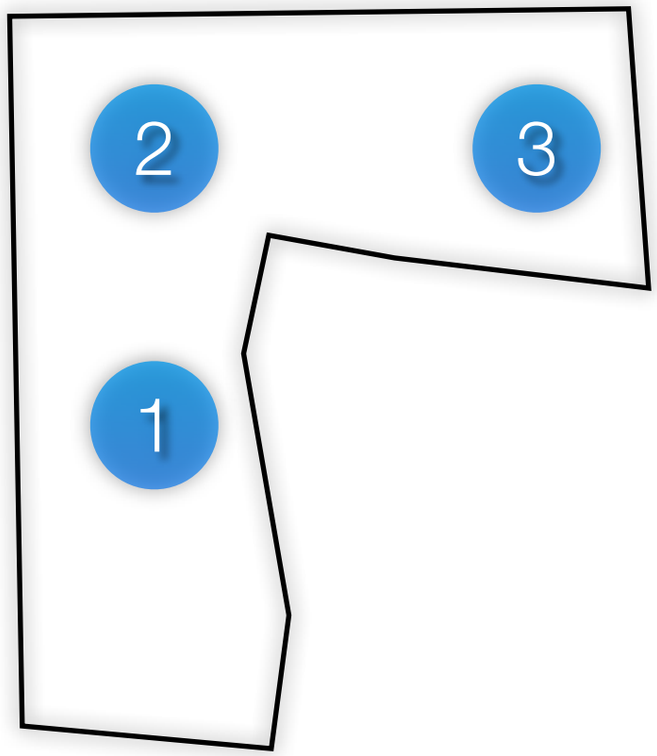


Stimulus-Approach

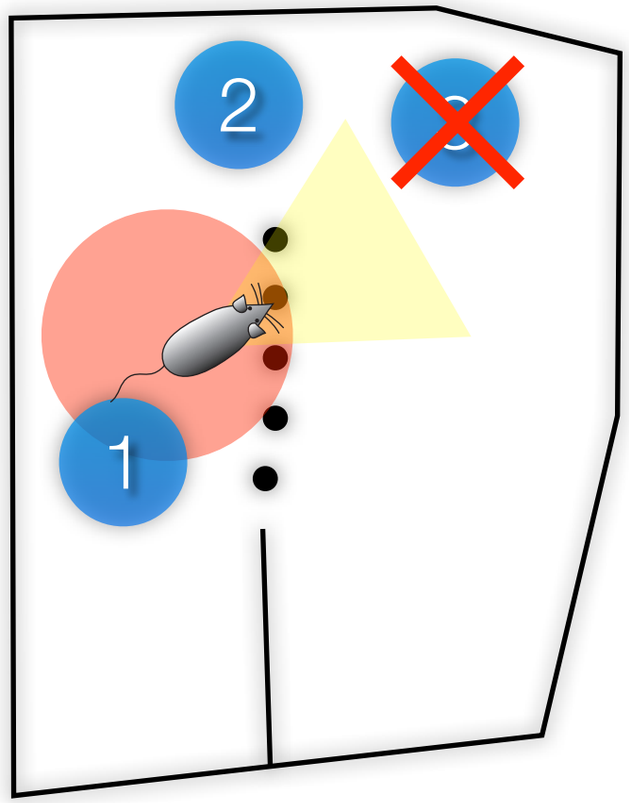
Stimulus-Response



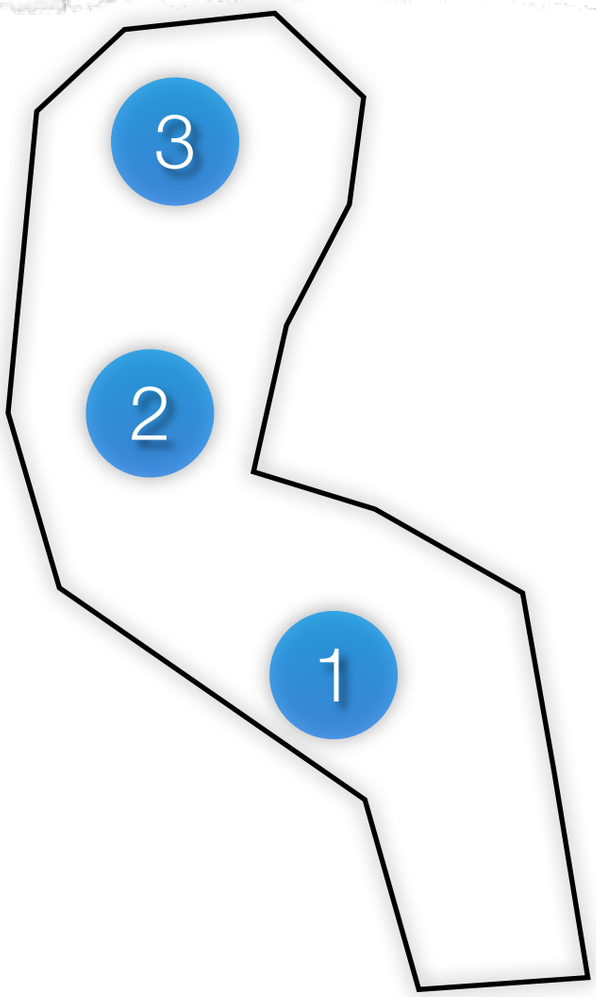
Stimulus-Approach



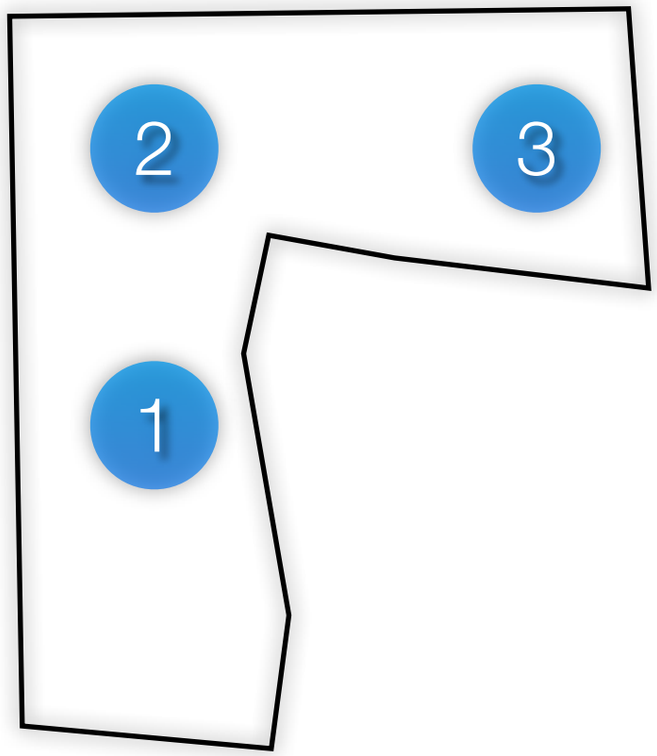
Stimulus-Response



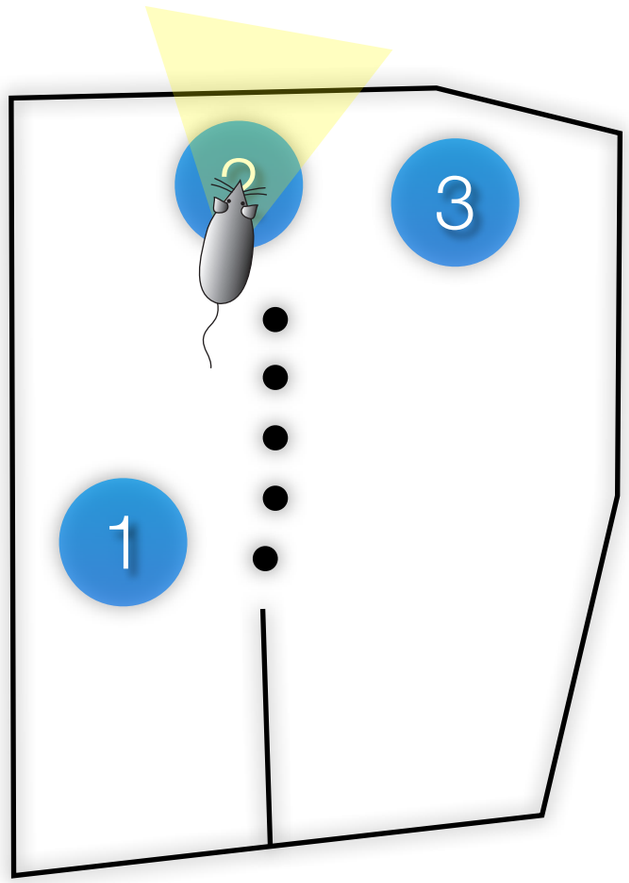
Contextual Inhibition



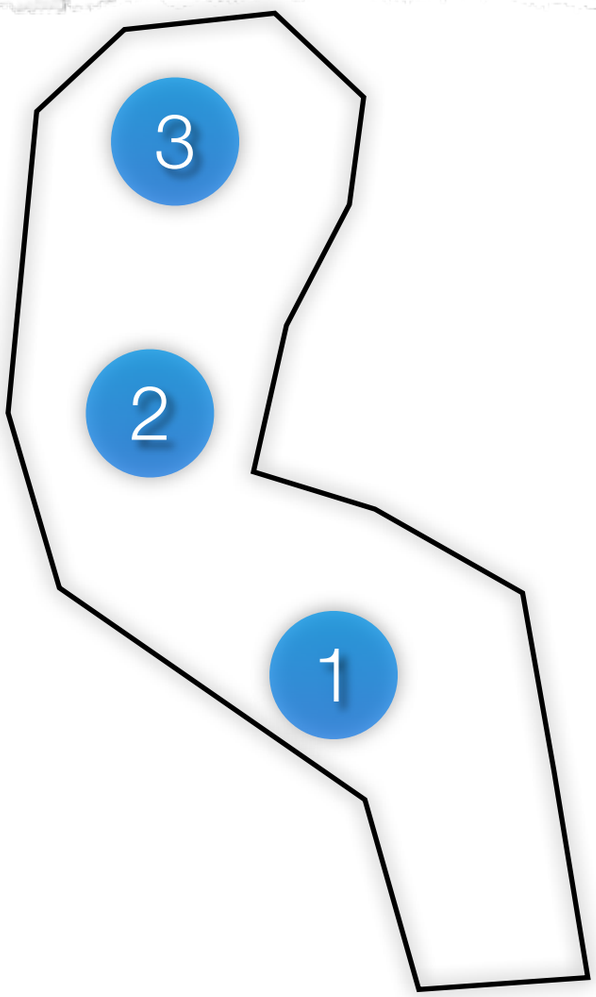
Stimulus-Approach



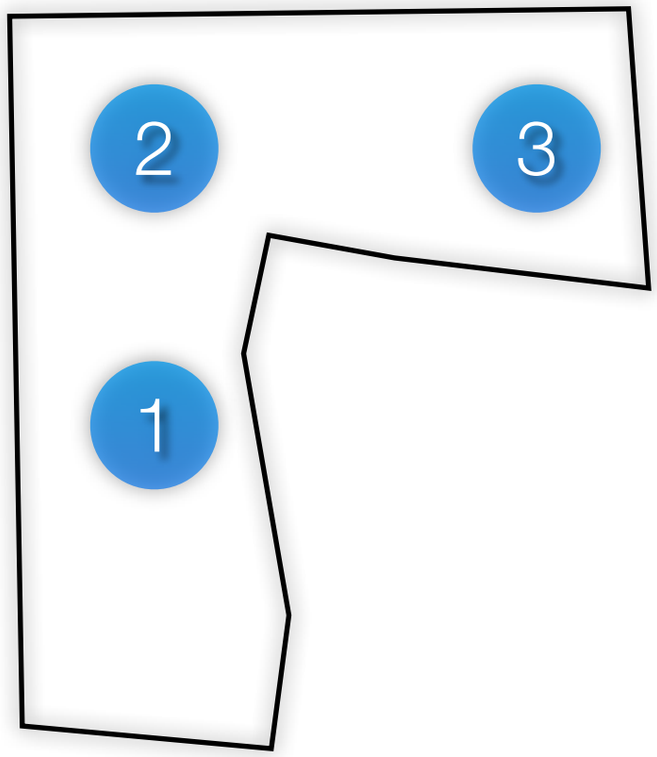
Stimulus-Response



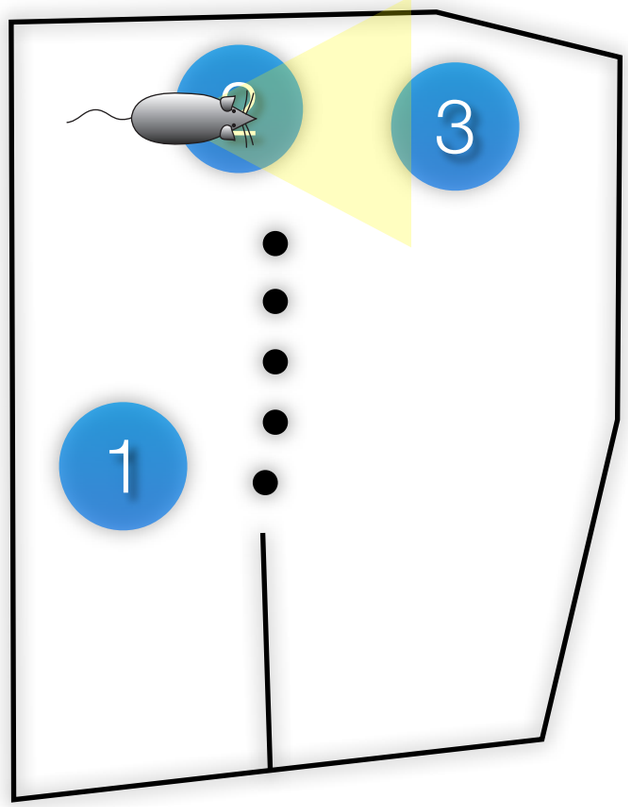
Contextual Inhibition



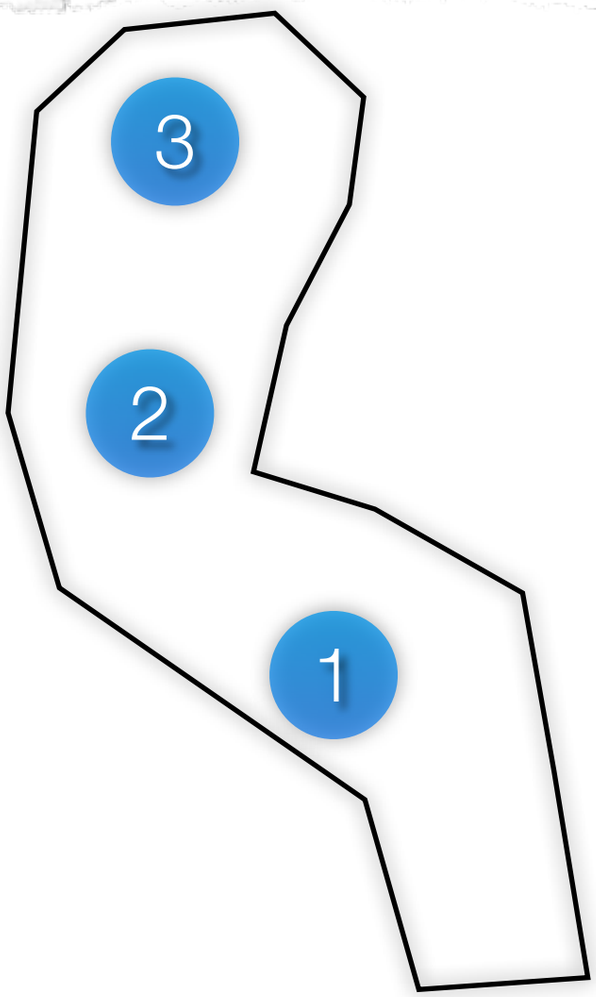
Stimulus-Approach



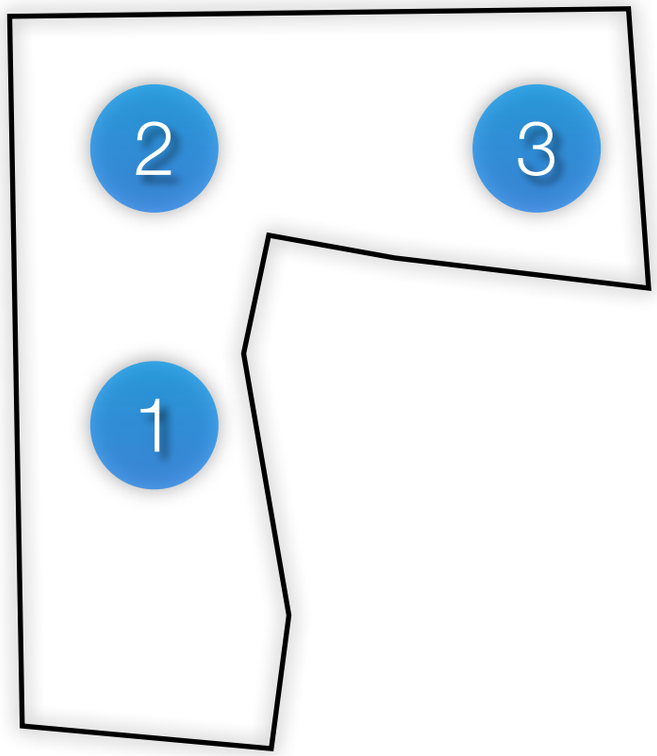
Stimulus-Response



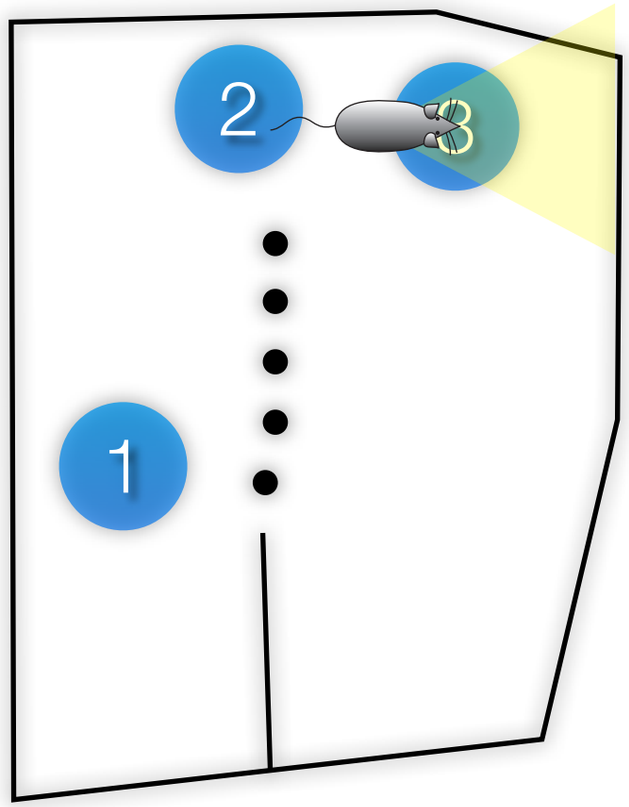
Contextual Inhibition



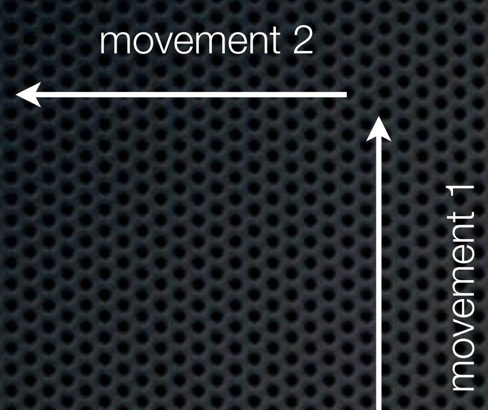
Stimulus-Approach

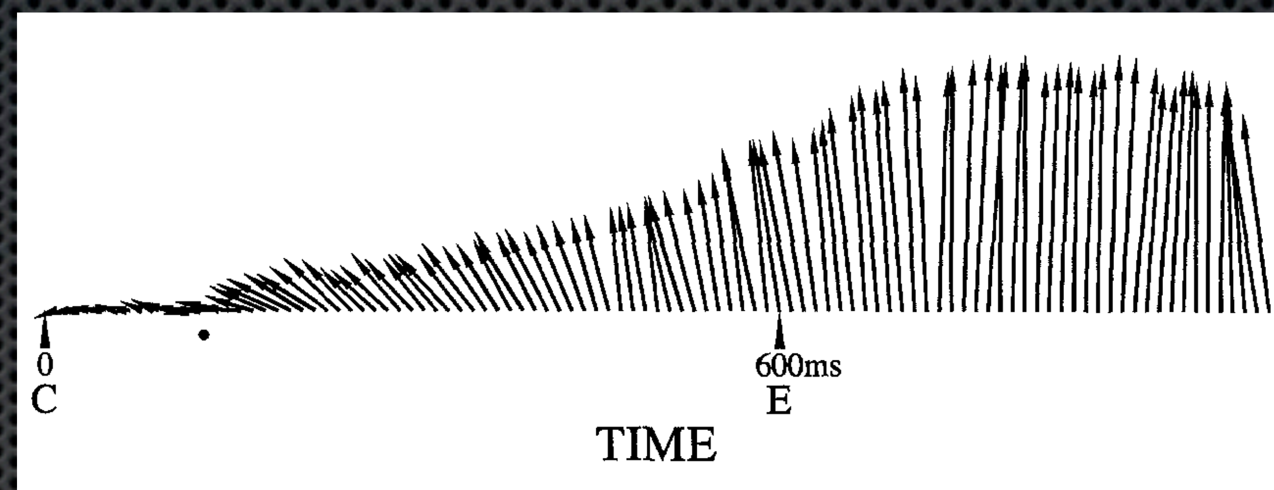


Stimulus-Response



Contextual Inhibition





Neuronal population vectors are plotted every 10 ms vs time. C, onset of the delay; E, end of the waiting period. The *filled circle* on the abscissa indicates the time after the beginning of the delay (130 ms) at which the population vector reached statistical significance

Ashe, et al. (1993) Exp Brain Res 95

Context Effects

Phase 1	CXA : CS + US
Phase 2	CXA : CS
Test A	CXA : CS → no-CR
Test B	CXB : CS → CR

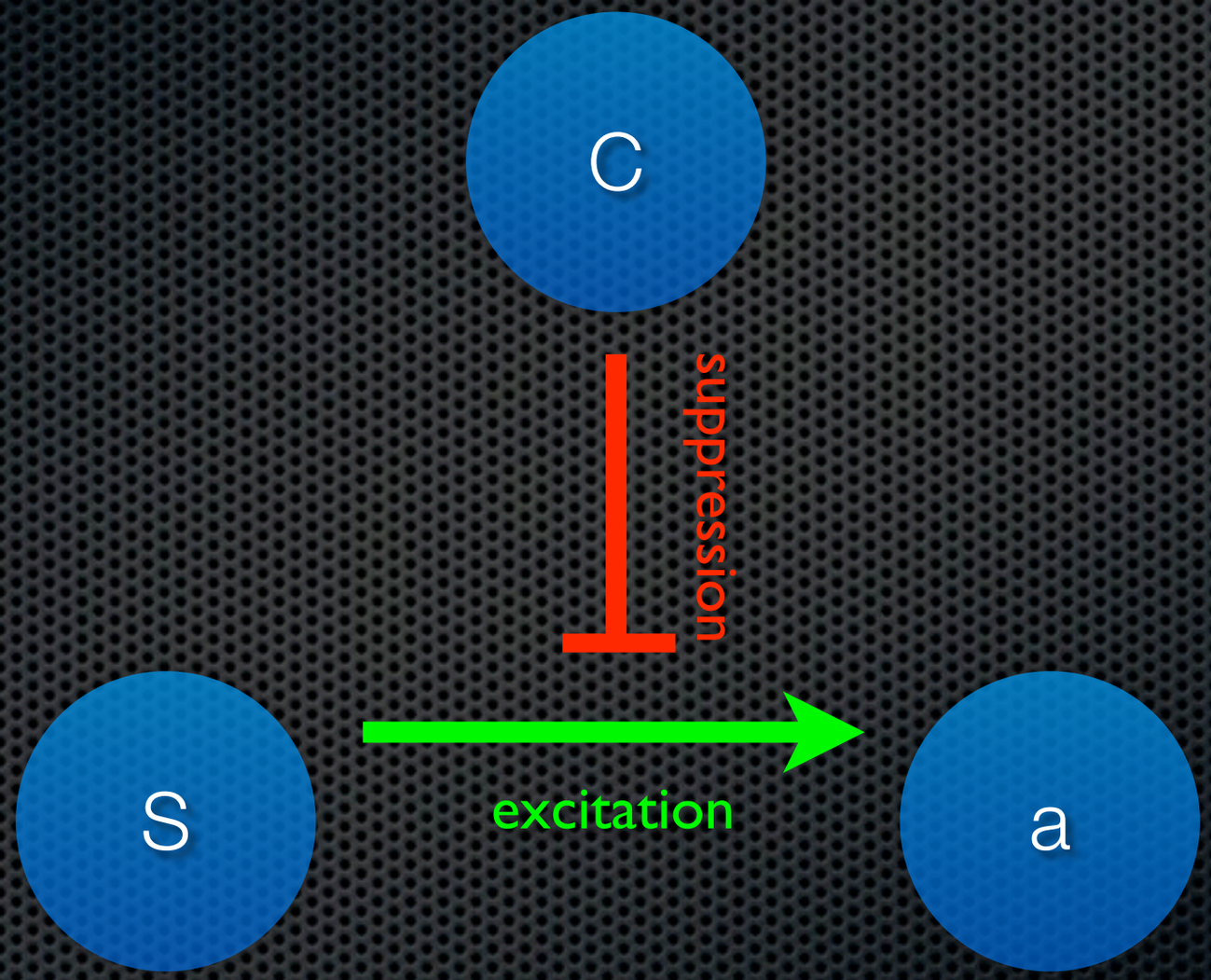
Extinction does not transfer to a new context (Bouton 1991, 1992)

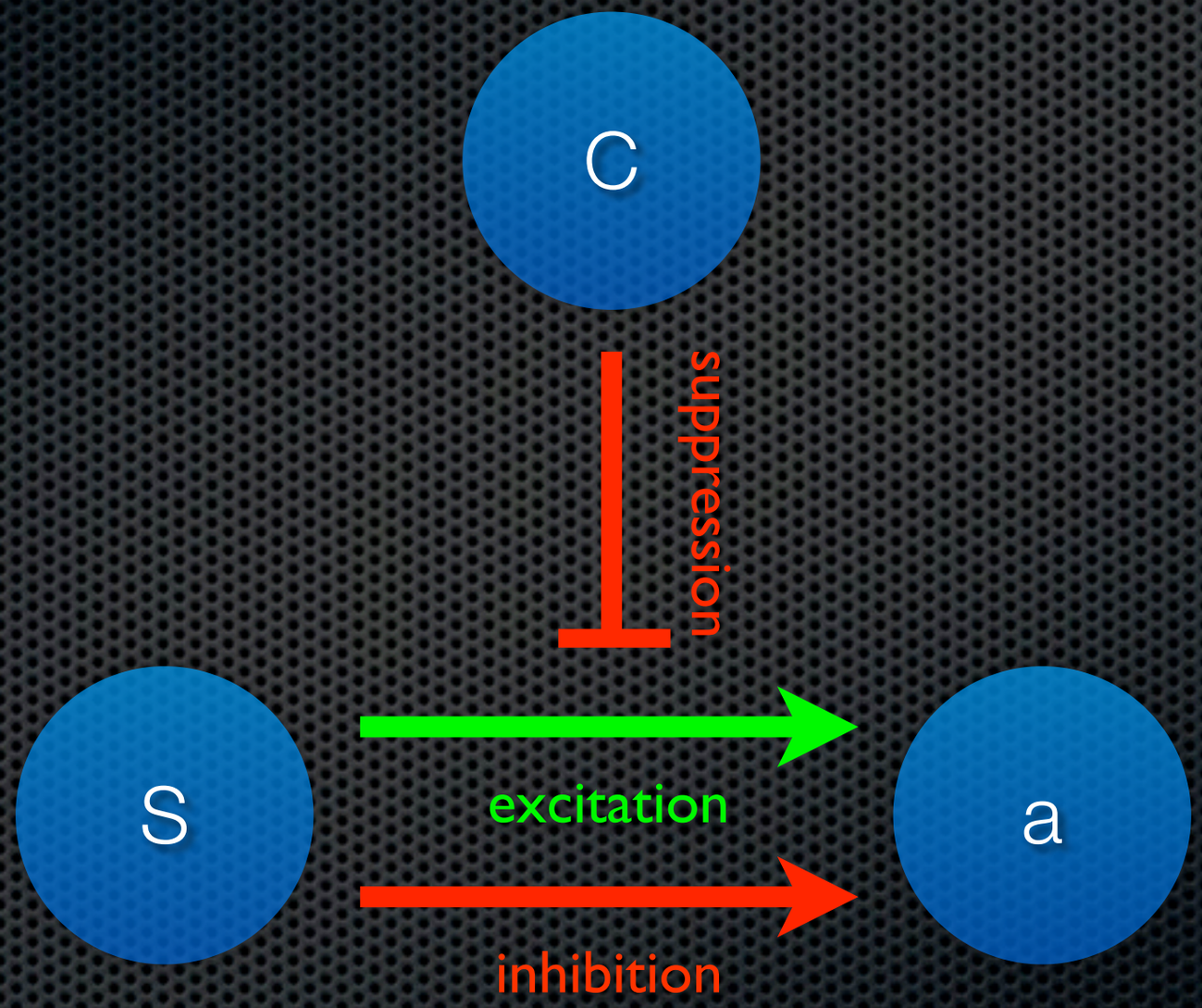
Contextual Inhibition

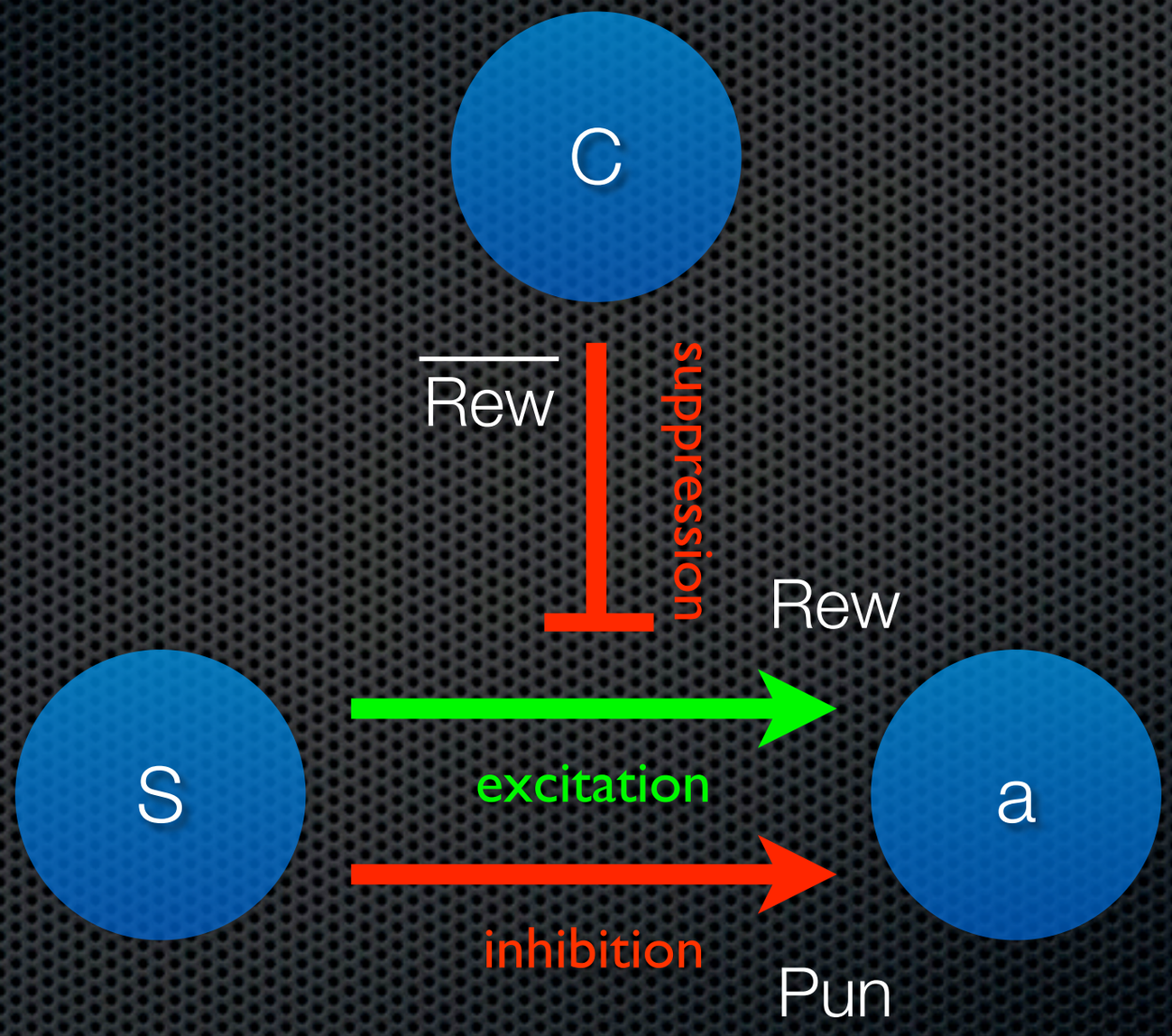
Three Learning Conditions

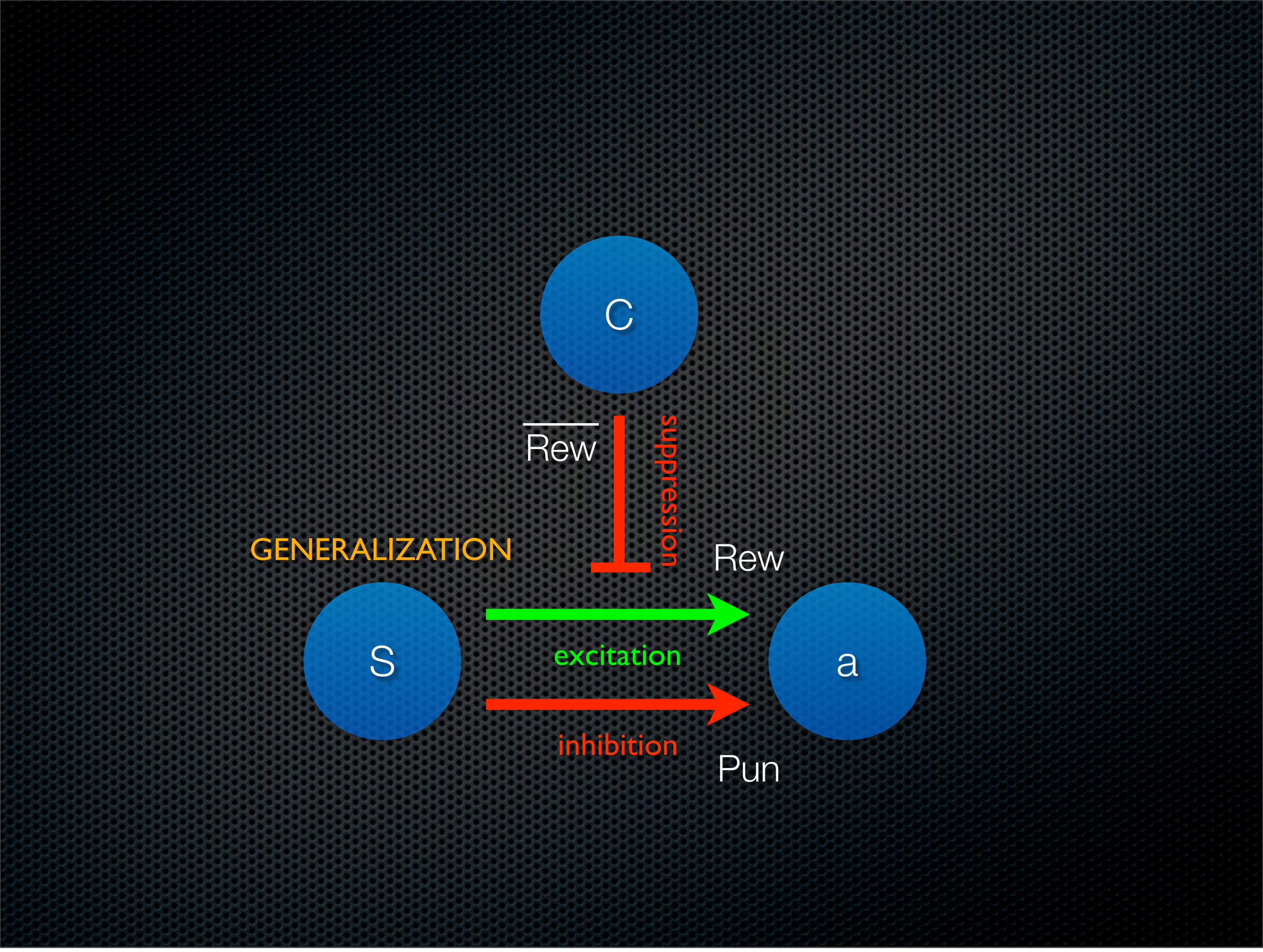
Rew	better than expected	maximal generalization
$\overline{\text{Rew}}$	worse than expected	contextual exception
Pun	bad	minimal generalization

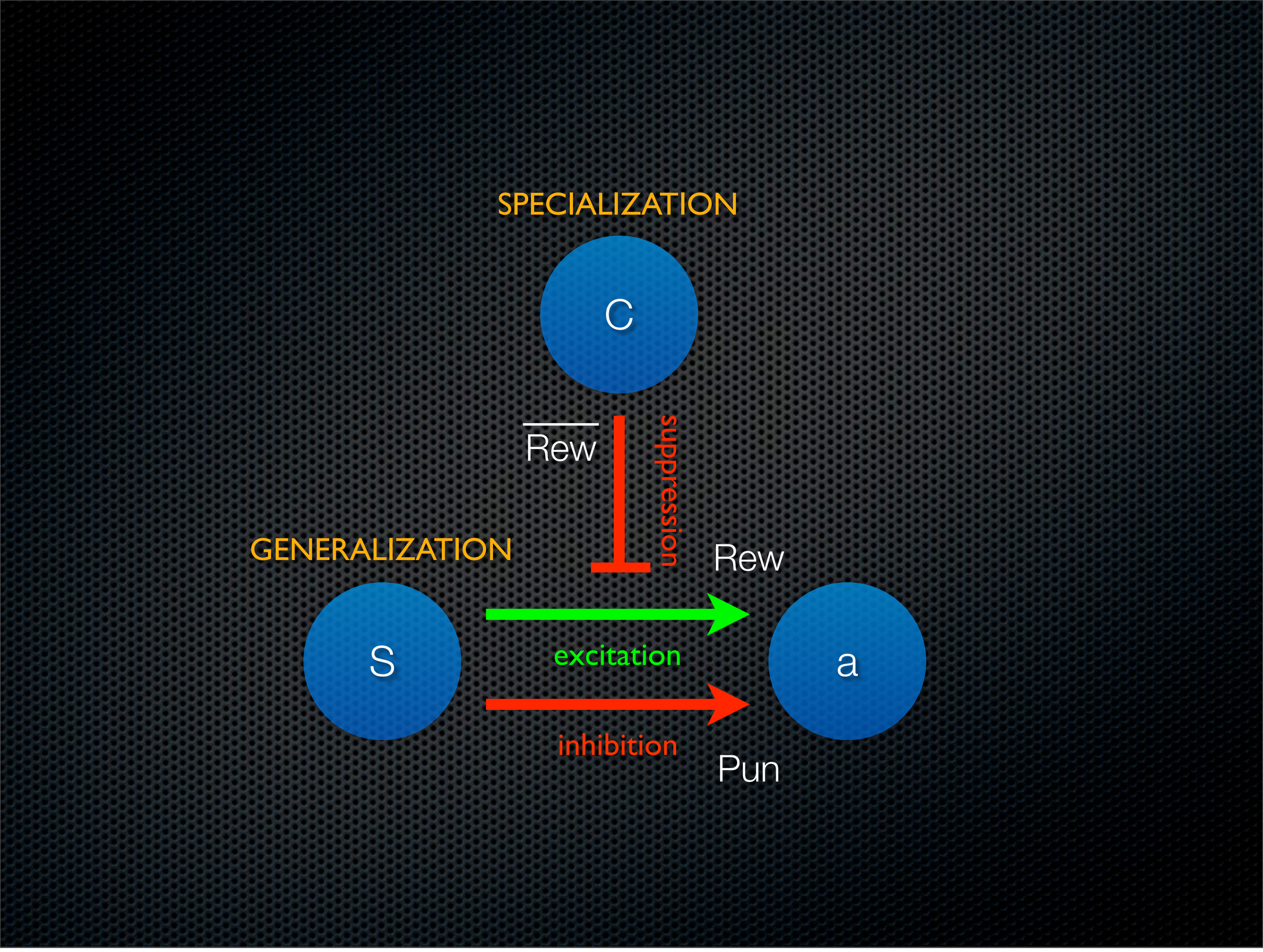

S
excitation
a

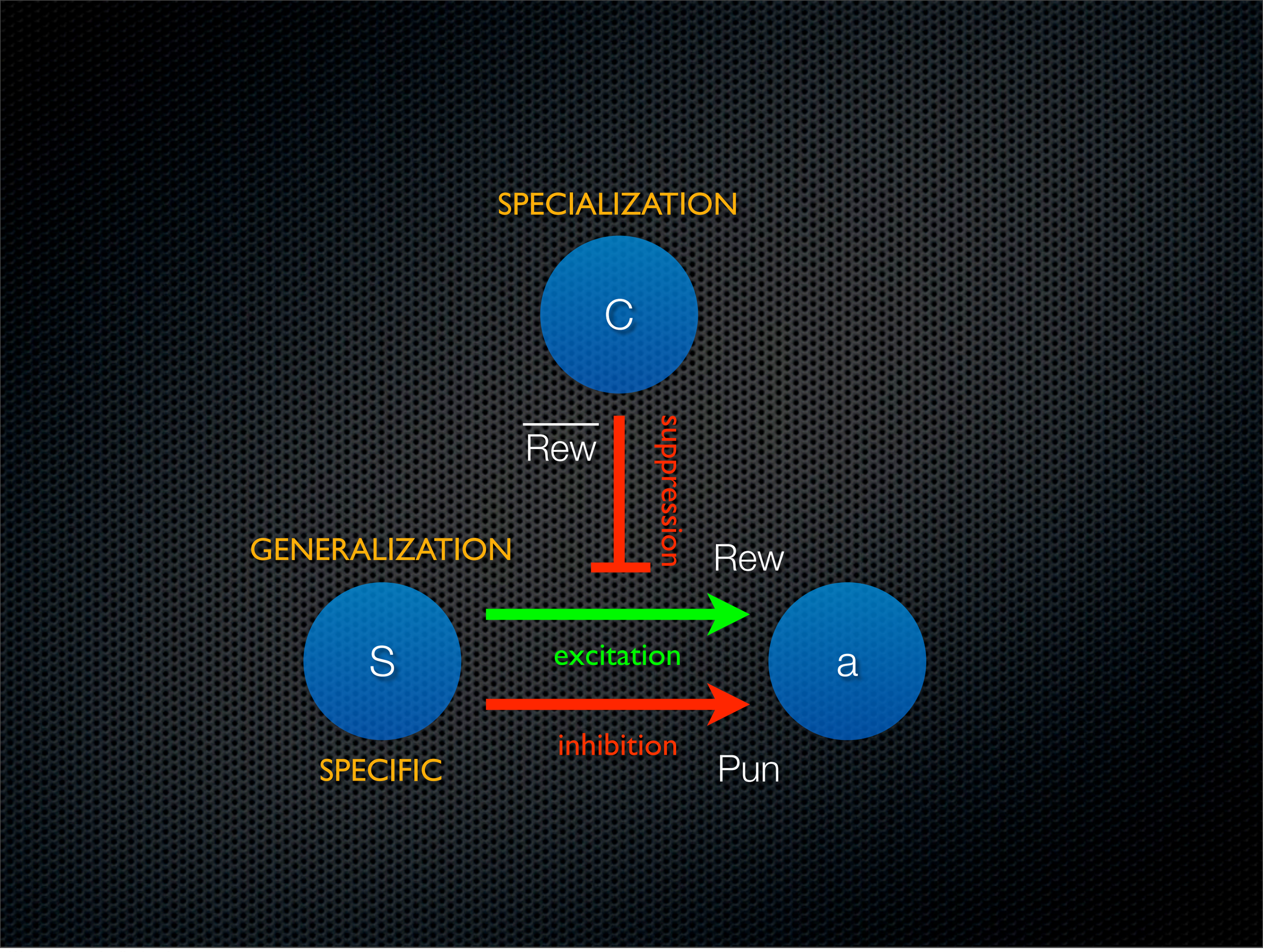

C
suppression
S
excitation
a


C
suppression
S
excitation
a
inhibition


C
Rew
suppression
Rew
S
excitation
a
inhibition
Pun


C
Rew
suppression
GENERALIZATION
Rew
S
excitation
a
inhibition
Pun

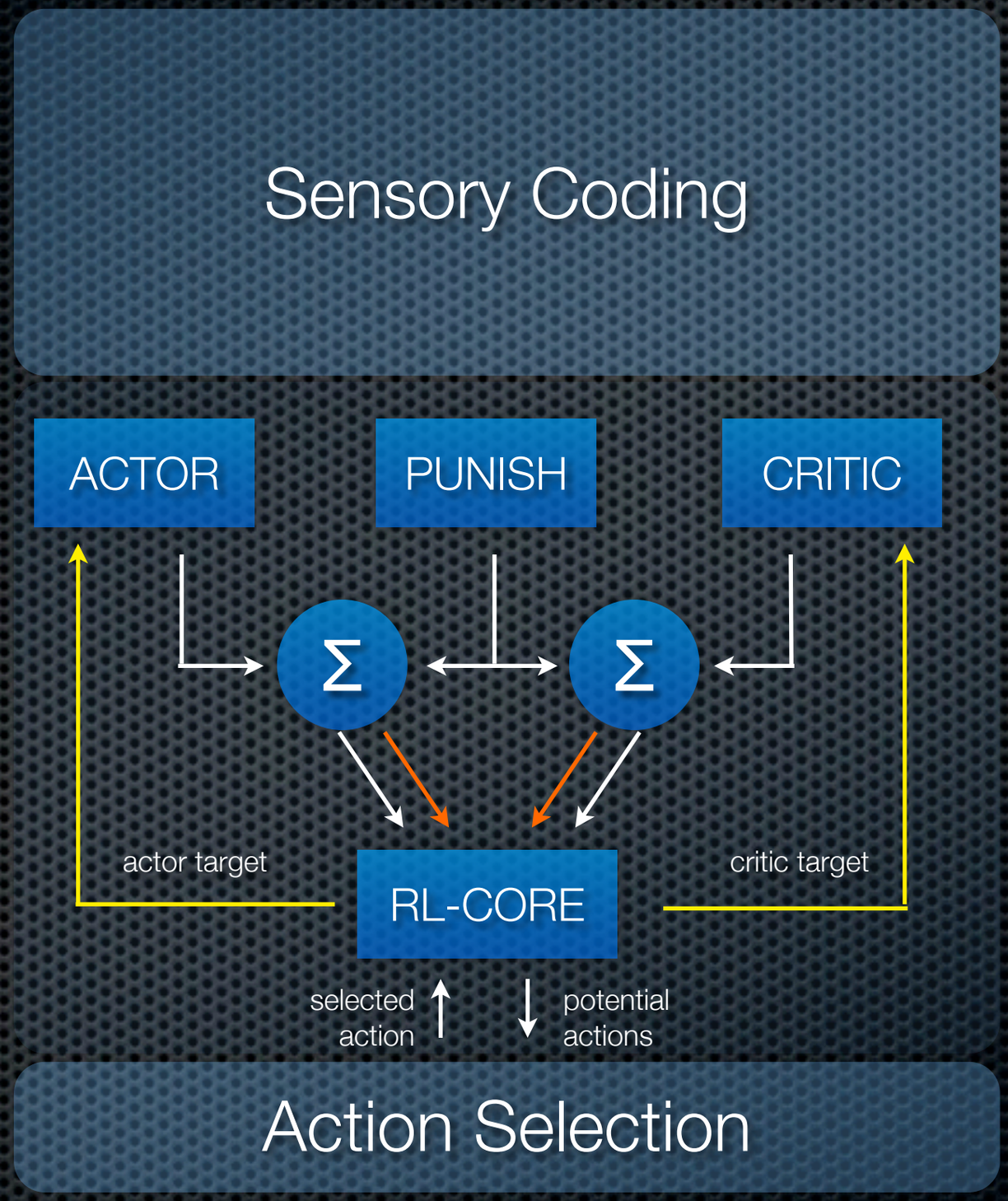

SPECIALIZATION
C
Rew
suppression
GENERALIZATION
Rew
S
excitation
a
inhibition
Pun

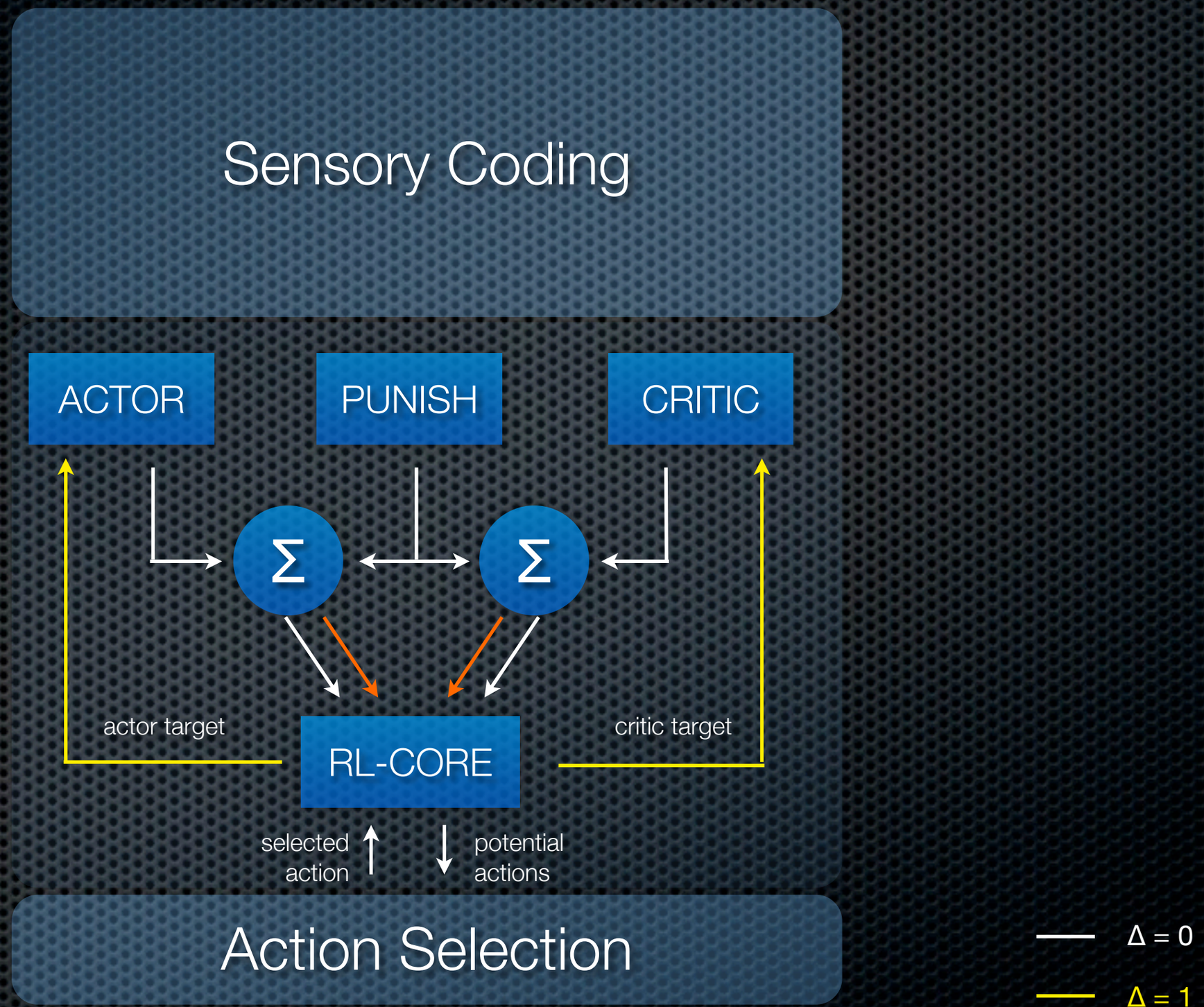

SPECIALIZATION
C
Rew
suppression
GENERALIZATION
Rew
S
excitation
a
inhibition
SPECIFIC
Pun

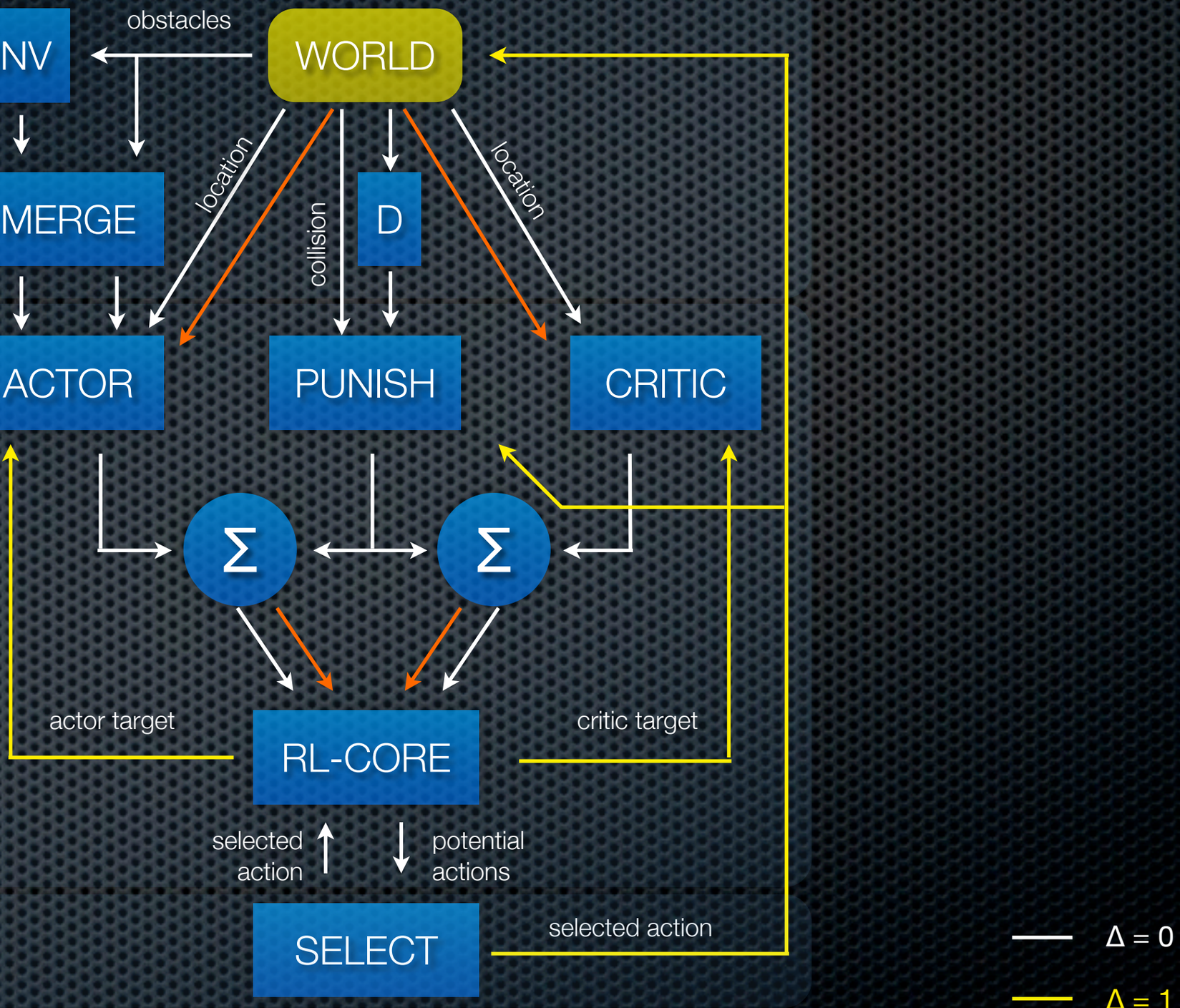
Sensory Coding

State & Action
Evaluation

Action Selection


Sensory Coding
ACTOR
PUNISH
CRITIC
Σ
Σ
actor target
critic target
RL-CORE
selected action
potential actions
Action Selection


Sensory Coding
ACTOR
PUNISH
CRITIC
Σ
Σ
actor target
RL-CORE
critic target
selected action
potential actions
Action Selection
Δ = 0
Δ = 1
Δ = 2


INV
obstacles
WORLD
MERGE
location
collision
D
location
ACTOR
PUNISH
CRITIC
Σ
Σ
actor target
RL-CORE
critic target
selected
action
potential
actions
SELECT
selected action
Δ = 0
Δ = 1
Δ = 2


HUCKEBY

Learning Algorithm

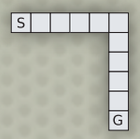
$$Q(c, s, a_j) = \sum_{i=0}^{n} s_i w_{ij} I_{ij},$$

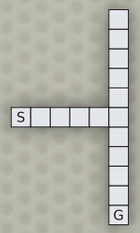
$$I_{ij} = \prod_{k=0}^{p} (1 - c_k u_{ijk})$$

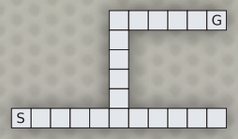
$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \alpha \frac{s_i a_j}{|s|} \Delta Q_t \qquad \Delta Q_t > 0$$

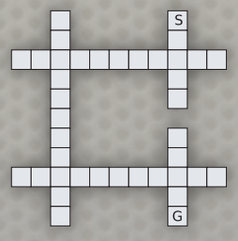
$$u_{ijk}^{(t+1)} = u_{ijk}^{(t)} - \beta (1 - u_{ijk}^{(t)}) \frac{s_i a_j c_k}{|s| w_{ij}} \Delta Q_t \qquad \Delta Q_t < 0$$

Maze

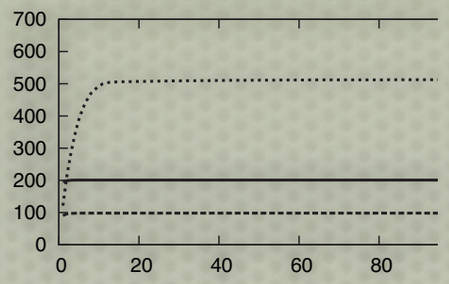


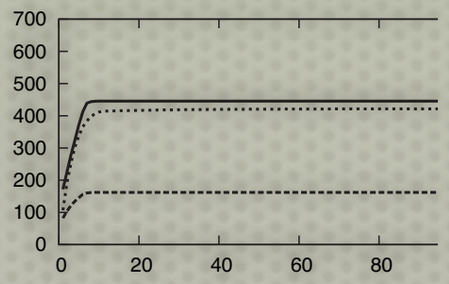


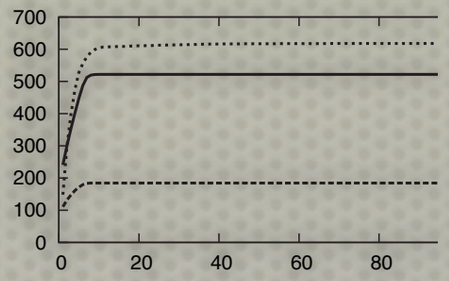


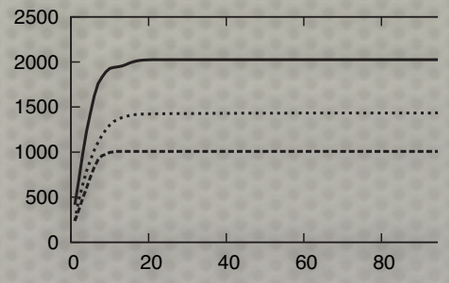


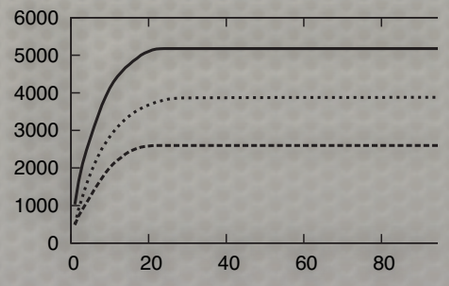
Accumulated Extra Steps





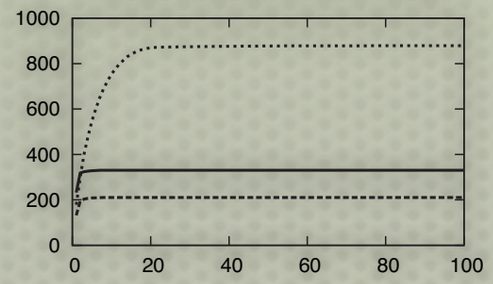


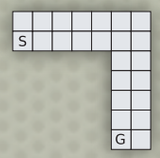


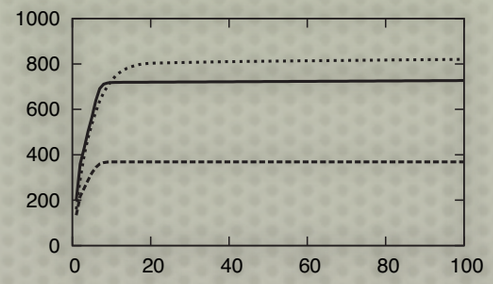


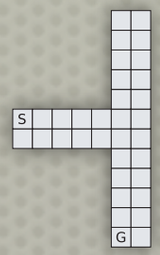
Maze

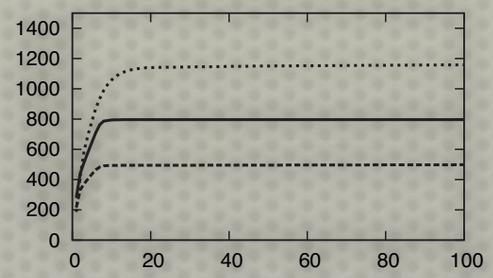
Accumulated Extra Steps

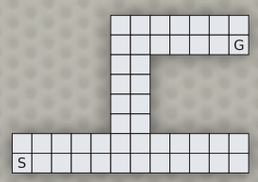


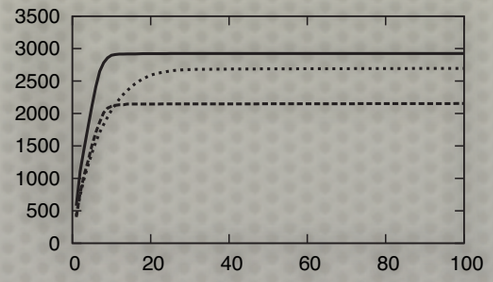


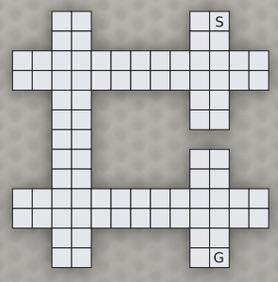


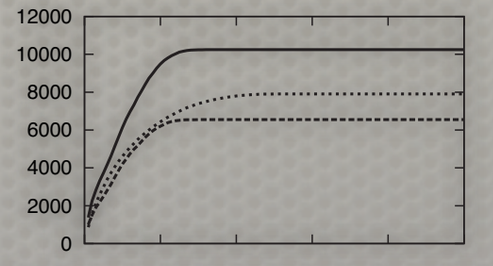




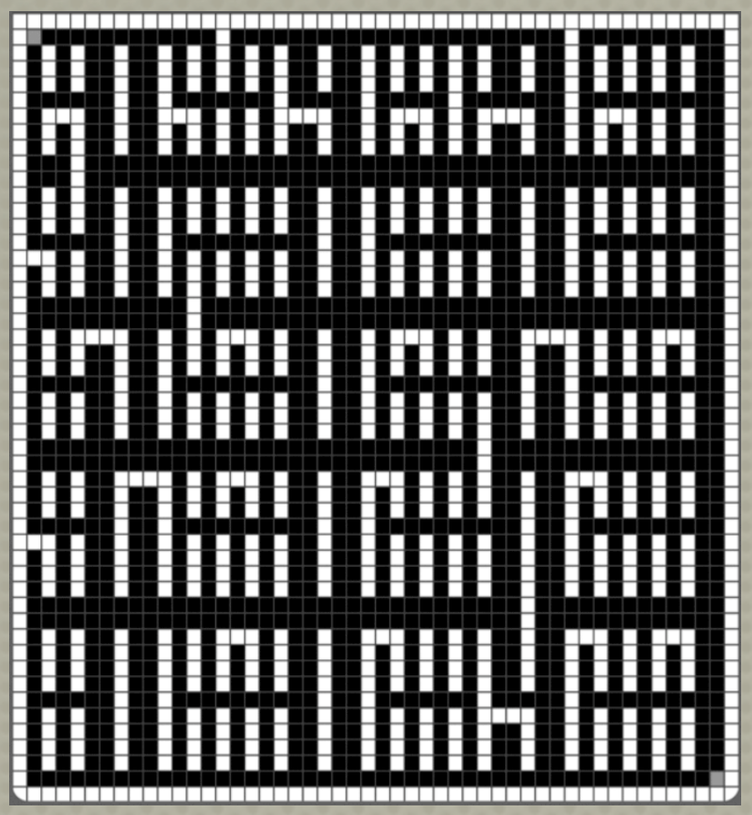


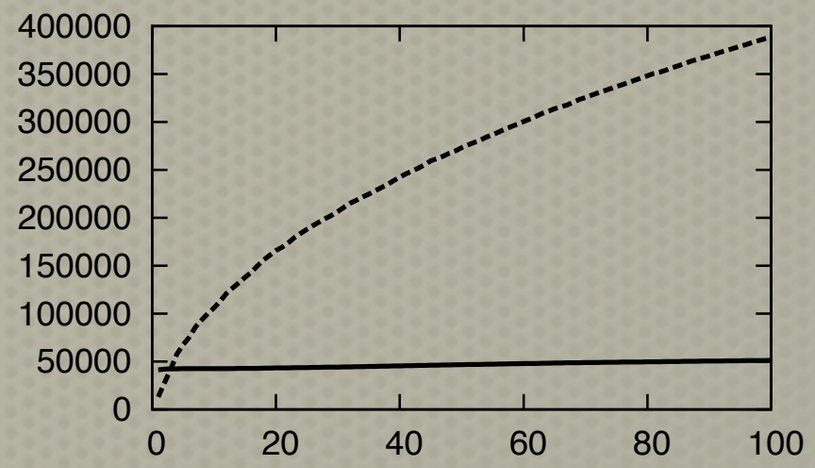




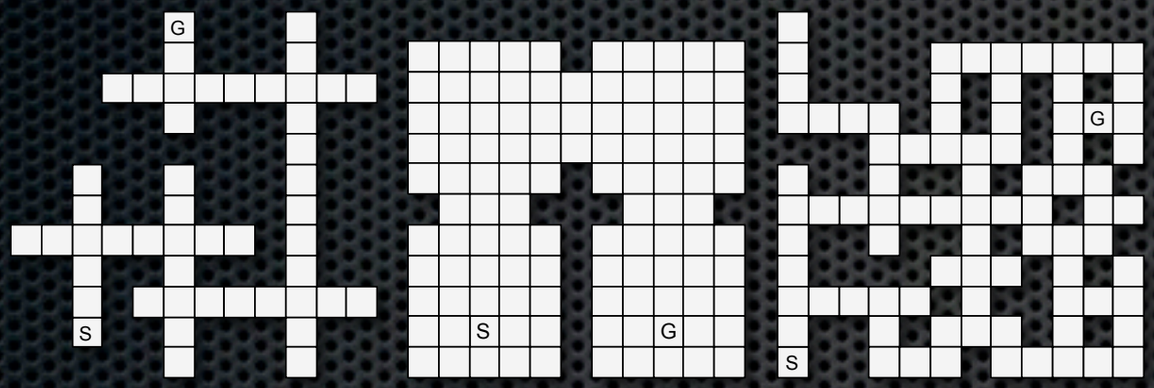


A More Complex Example

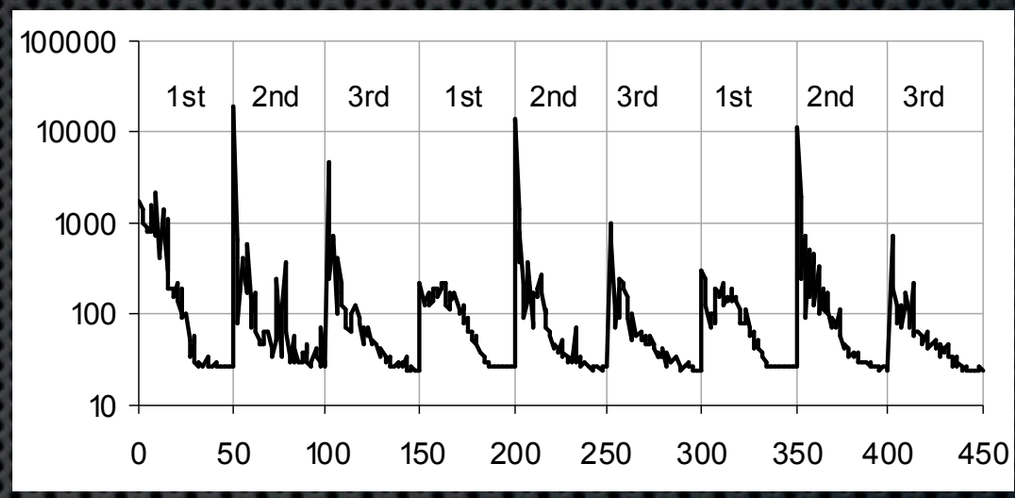




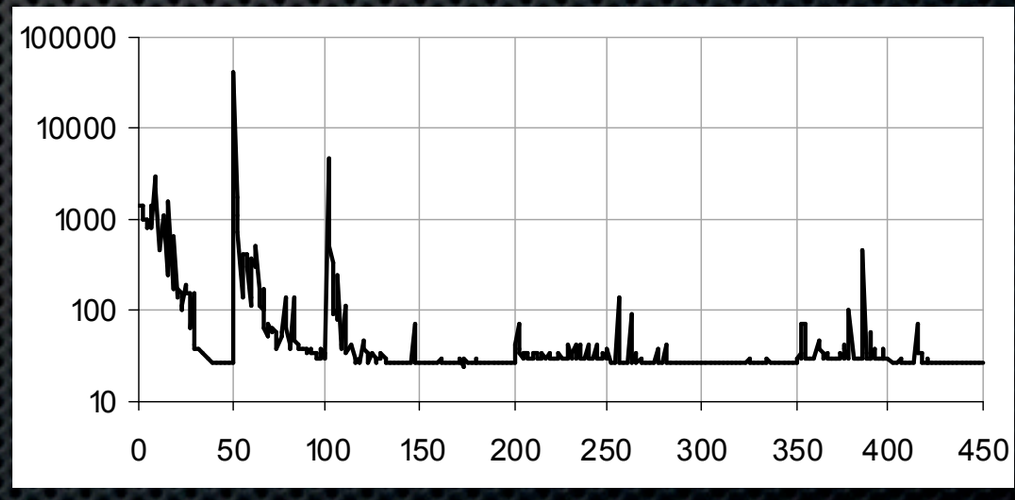
Context Prevents Catastrophic Forgetting



Q-Learning with a regular linear network



Context-Q



Winberg, 2005

Four Algorithms

Q	ContextQ	ContextAC	ContextACP
Q(s, a)	Q(c, s, a)	Q(c, s, a) V(s)	Q(c, s, a) V(s) P(s, a)
‘stimulus generalization’	contextual specialization	progress separate from state that controls action	learns to avoid doing bad things

Four Algorithms

Q	ContextQ	ContextAC	ContextACP
Q(s, a)	Q(c, s, a)	Q(c, s, a) V(s)	Q(c, s, a) V(s) P(s, a)
‘stimulus generalization’	contextual specialization	progress separate from state that controls action	learns to avoid doing bad things

general → less general

- Stimulus-Approach
- Stimulus-Response
- Contextual Inhibition