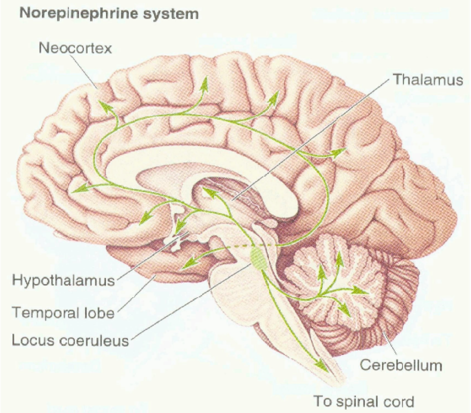
The Explore / Exploit Tradeoff
and
Locus Coeruleus / Norepinephrine Neuromodulation

LC-NE System

LC-NE System

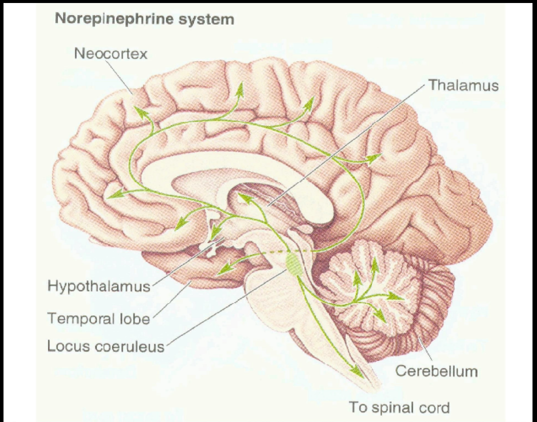
- **Locus Coeruleus:**
  - **A small nucleus of cells of in the rostral pontine tegmentum** *(upper brainstem)*



from *Neuroscience: Exploring the Brain* by M.F. Bear, B.W. Connors, and M.A. Paradiso, 2001

LC-NE System

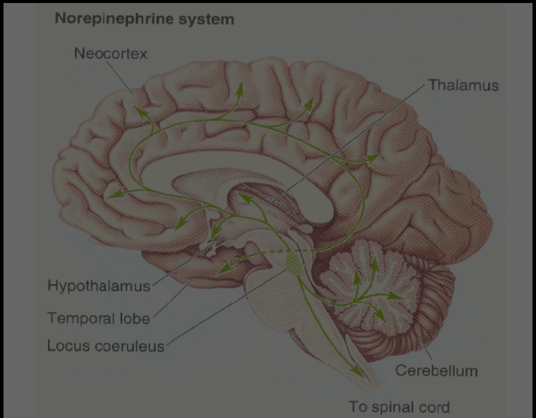
- **Locus Coeruleus:**
  - A small nucleus of cells of in the rostral pontine tegmentum *(upper brainstem)*
  - **Innervates all levels of neuraxis, source of 99% of norepinephrine in neocortex**



from *Neuroscience: Exploring the Brain* by M.F. Bear, B.W. Connors, and M.A. Paradiso, 2001

LC-NE System

- **Locus Coeruleus:**
  - **A small nucleus of cells of in the rostral pontine tegmentum** *(upper brainstem)*
  - **Innervates all levels of neuraxis, source of 99% of norepinephrine in neocortex**

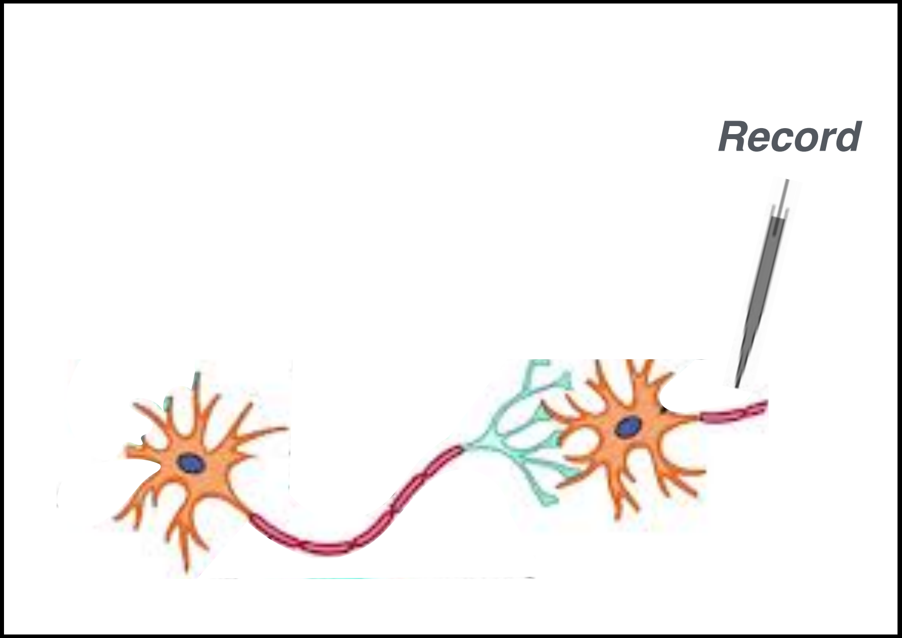


from *Neuroscience: Exploring the Brain* by M.F. Bear, B.W. Connors, and M.A. Paradiso, 2001

- **NE is a *neuromodulator…***
  (like dopamine)

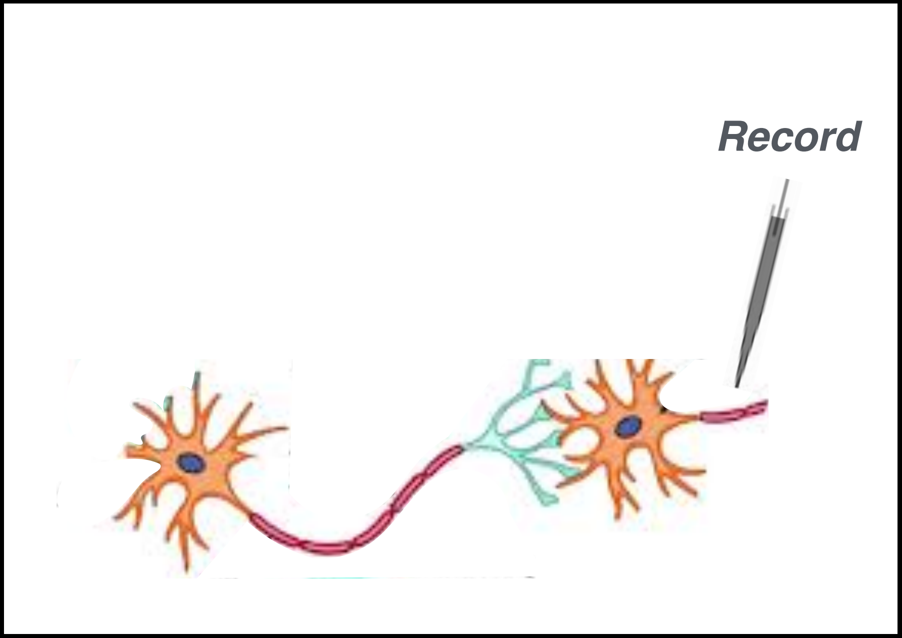
Neuromodulation

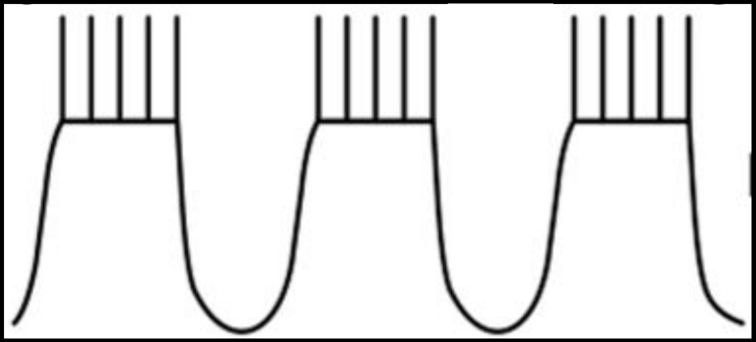
Chiodo & Berger, 1986; Waterhouse et al.,1998) Seamans & Yang, 2004)



Neuromodulation

Chiodo & Berger, 1986; Waterhouse et al.,1998) Seamans & Yang, 2004)

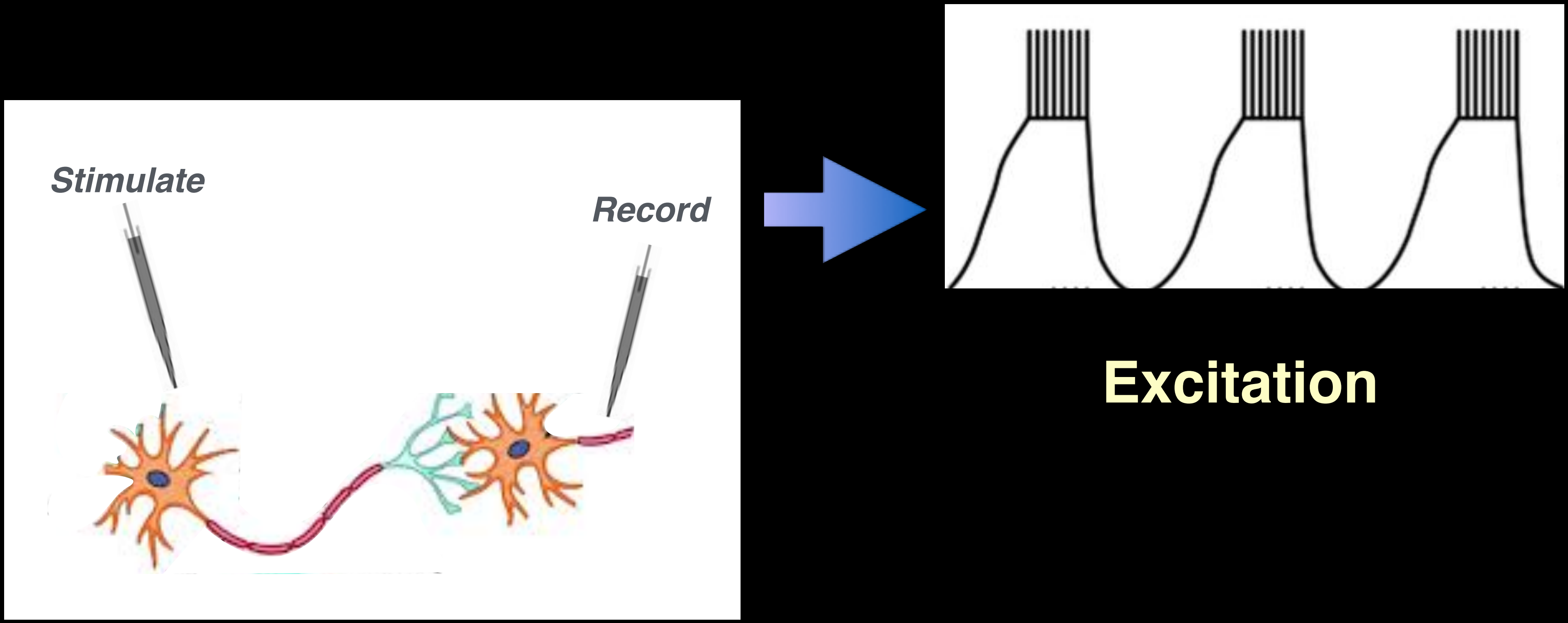




Baseline

Neuromodulation

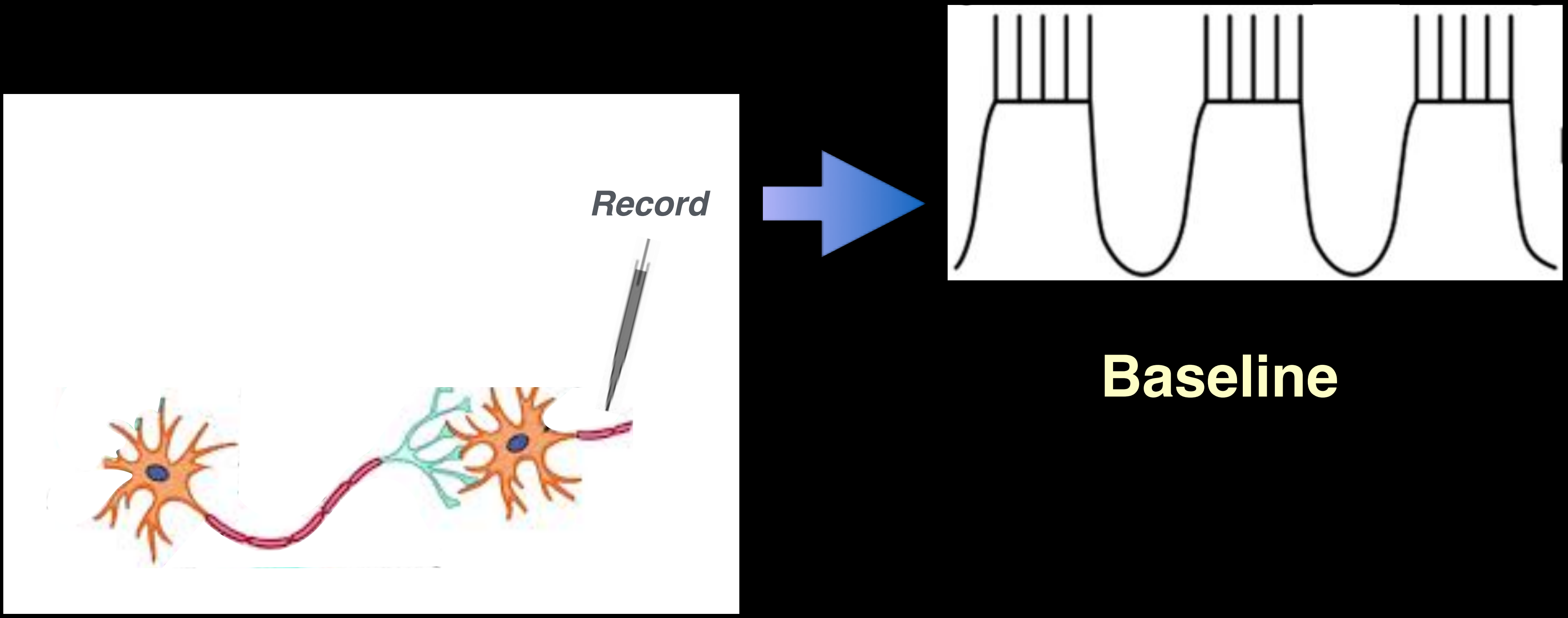
Chiodo & Berger, 1986; Waterhouse et al.,1998) Seamans & Yang, 2004)



Excitation

Neuromodulation

Chiodo & Berger, 1986; Waterhouse et al.,1998) Seamans & Yang, 2004)

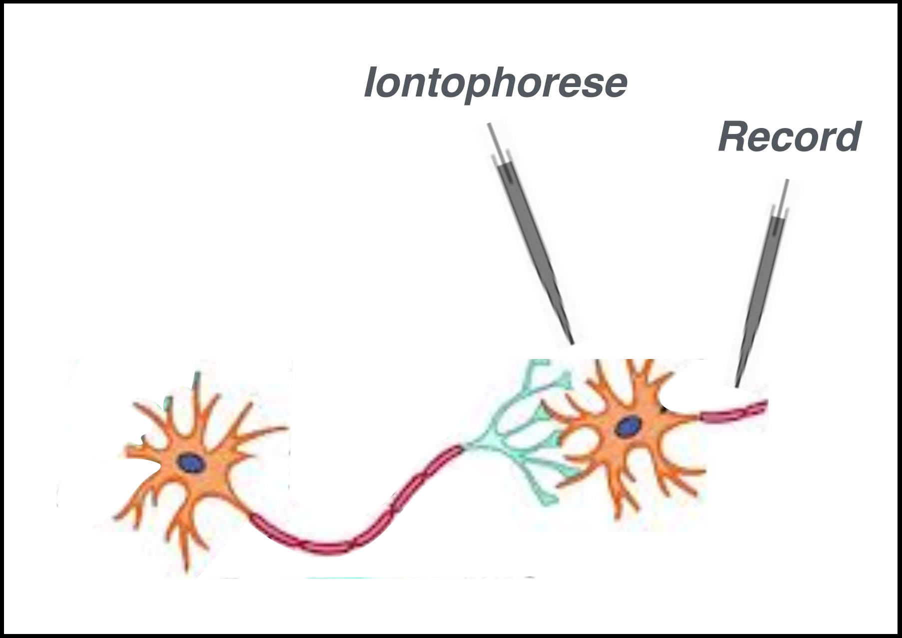


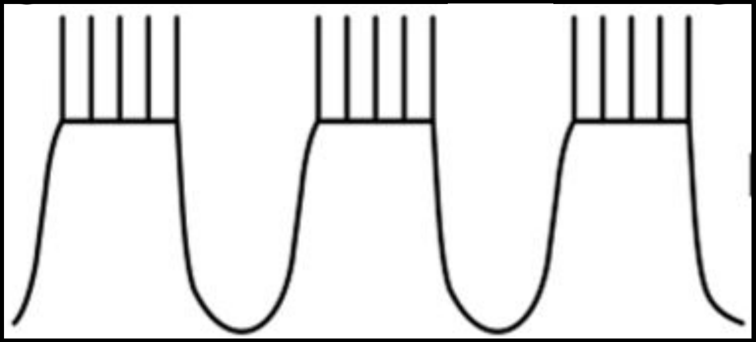
Baseline

Neuromodulation

Chiodo & Berger, 1986; Waterhouse et al.,1998) Seamans & Yang, 2004)

CA



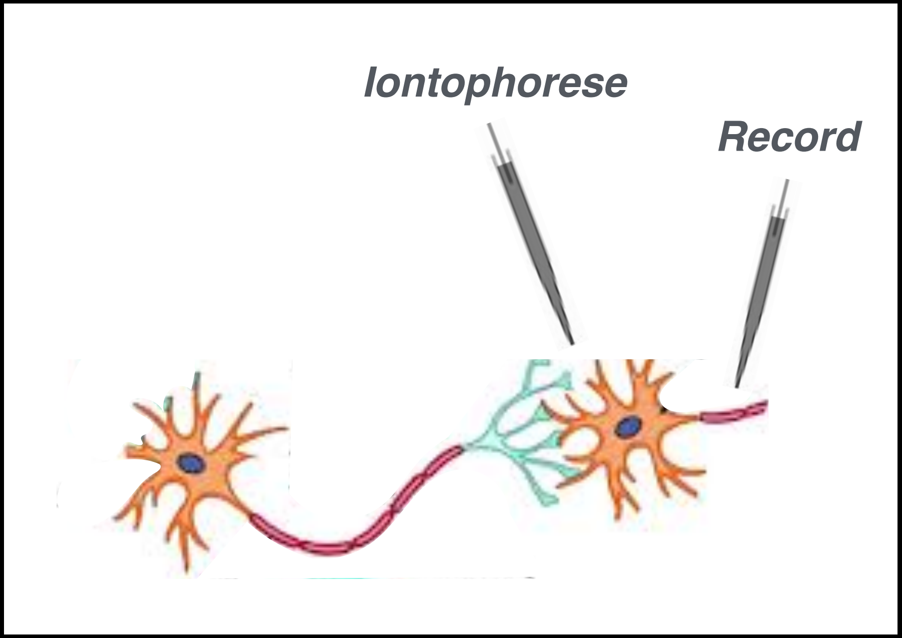


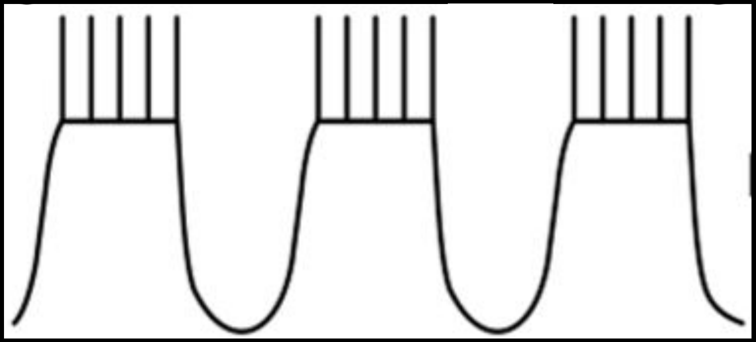
Baseline

Neuromodulation

Chiodo & Berger, 1986; Waterhouse et al.,1998) Seamans & Yang, 2004)

CA



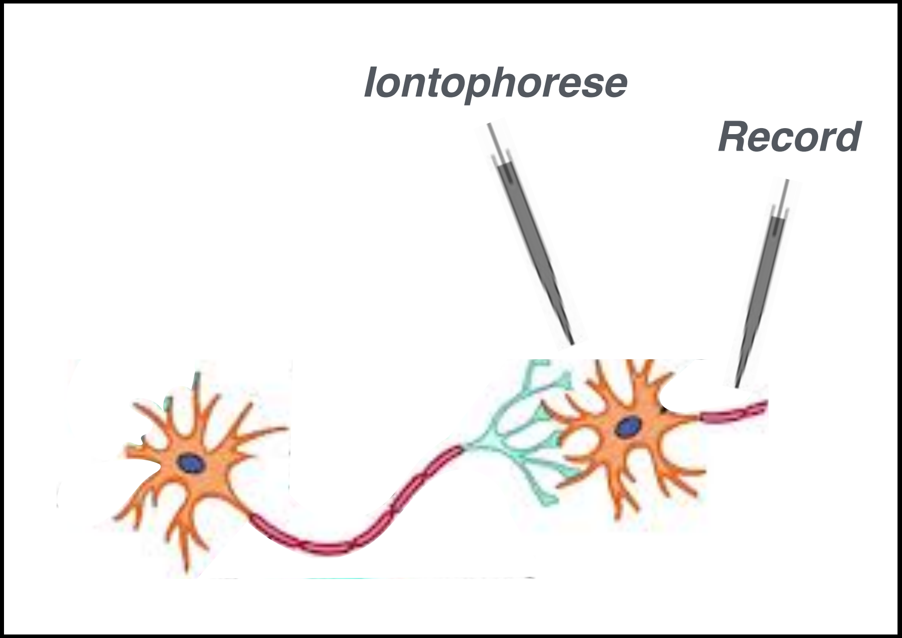


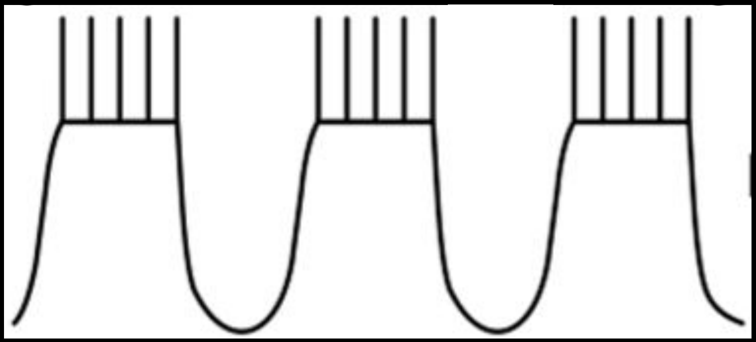
Baseline

Neuromodulation

Chiodo & Berger, 1986; Waterhouse et al.,1998) Seamans & Yang, 2004)

CA



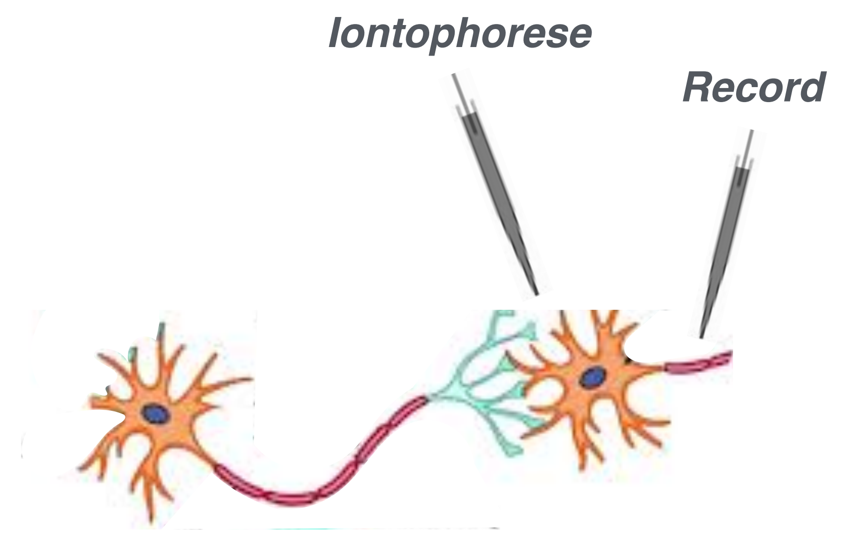


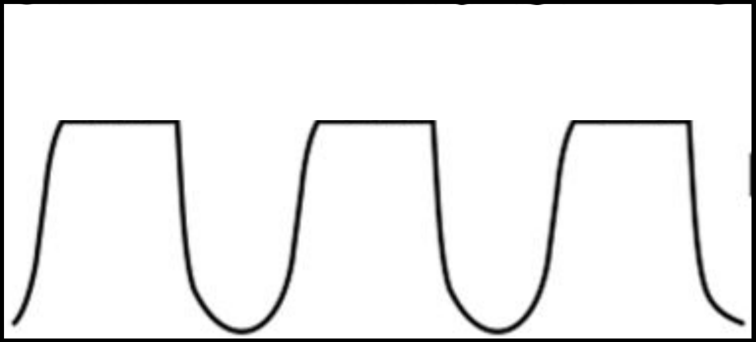
Baseline

Neuromodulation

Chiodo & Berger, 1986; Waterhouse et al.,1998) Seamans & Yang, 2004)

CA

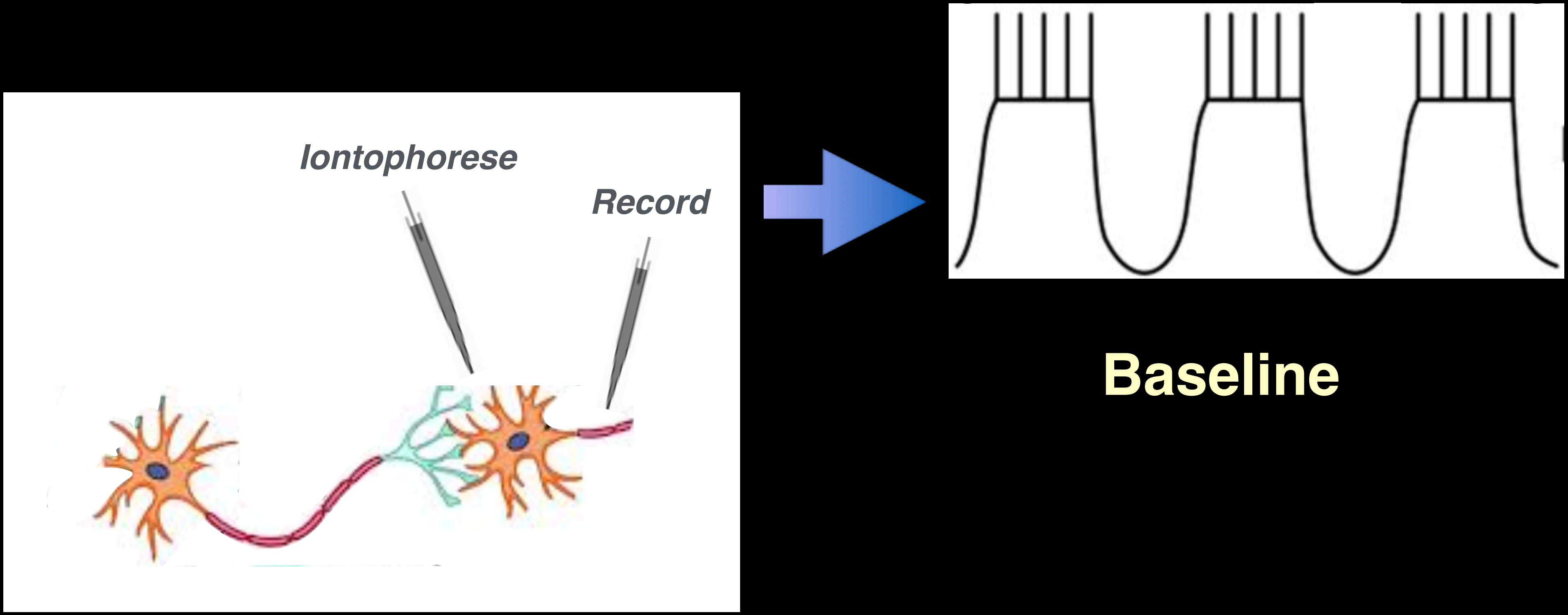




Inhibition

Neuromodulation

Chiodo & Berger, 1986; Waterhouse et al.,1998) Seamans & Yang, 2004)

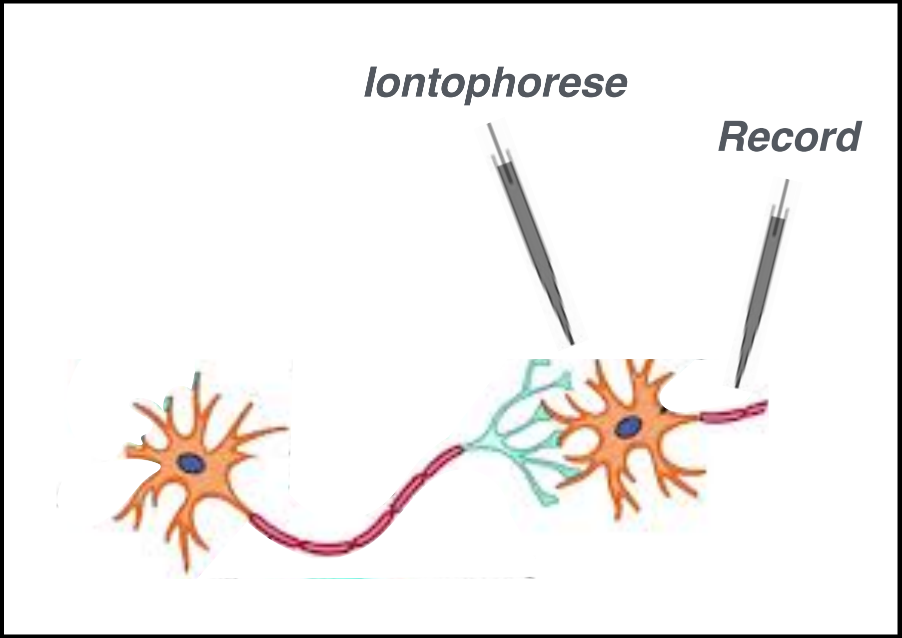


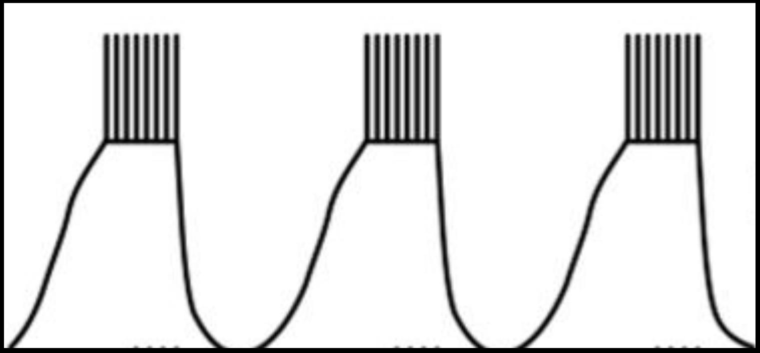
Baseline

Neuromodulation

Chiodo & Berger, 1986; Waterhouse et al.,1998) Seamans & Yang, 2004)

Glutamate



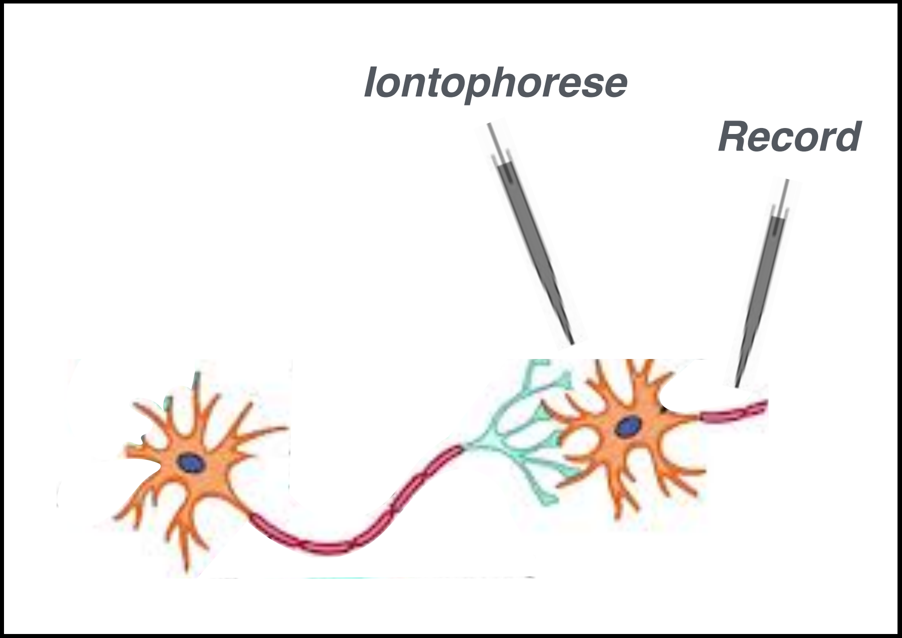


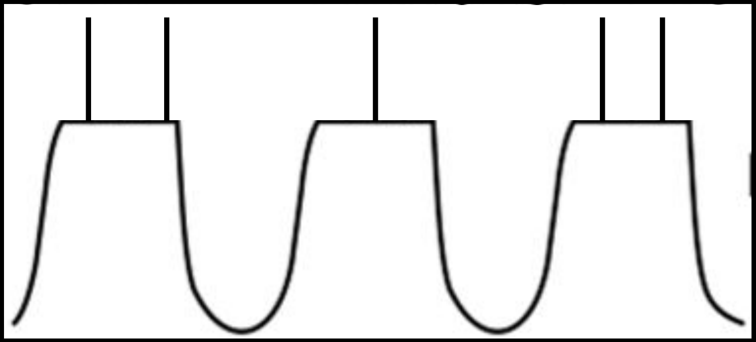
Excitation

Neuromodulation

Chiodo & Berger, 1986; Waterhouse et al.,1998) Seamans & Yang, 2004)

GABA



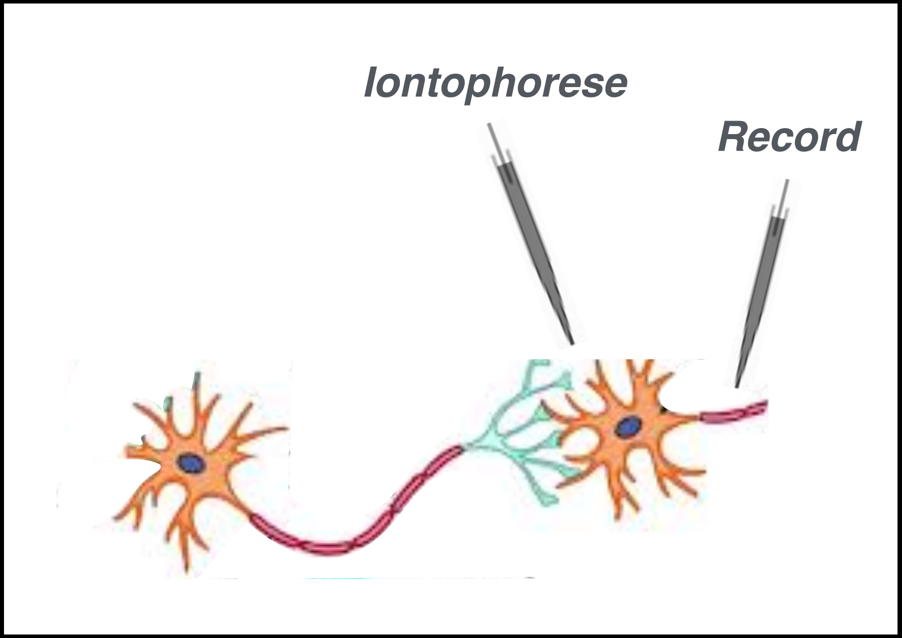


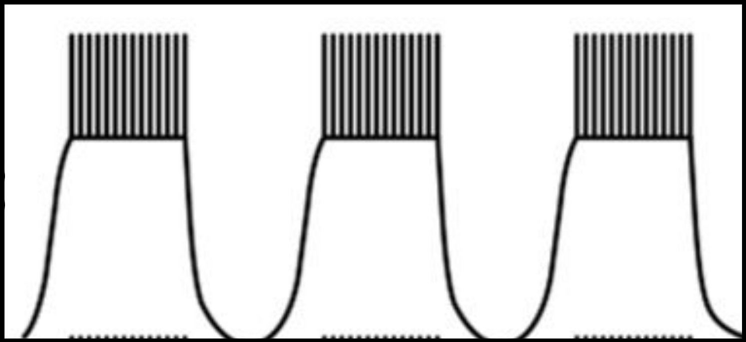
Inhibition

Neuromodulation

Chiodo & Berger, 1986; Waterhouse et al.,1998) Seamans & Yang, 2004)

CA + Glutamate



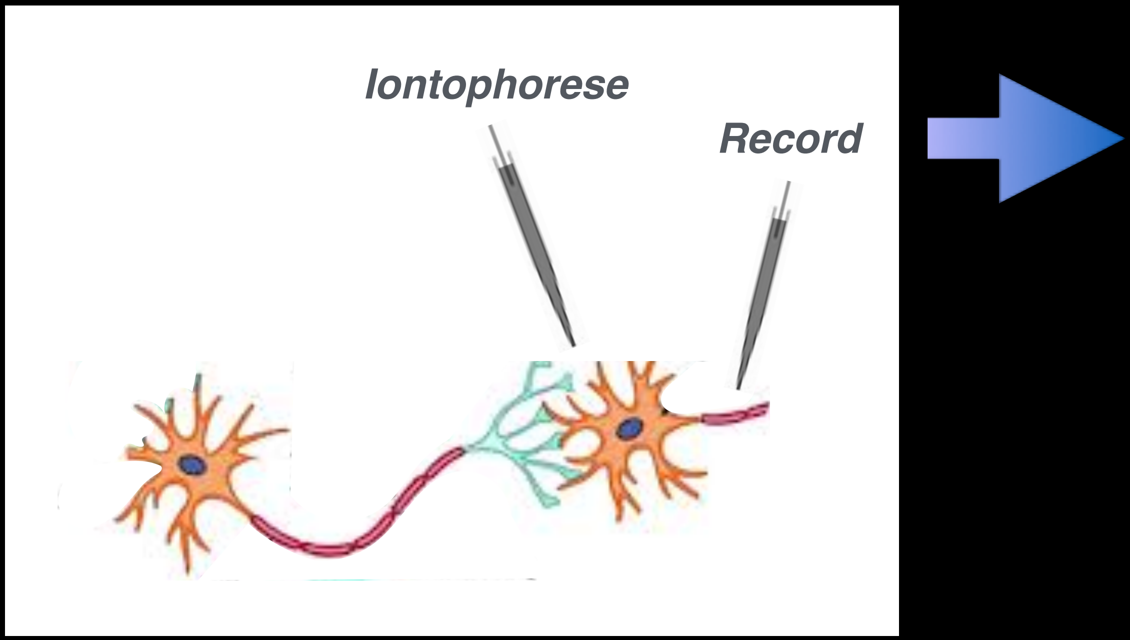


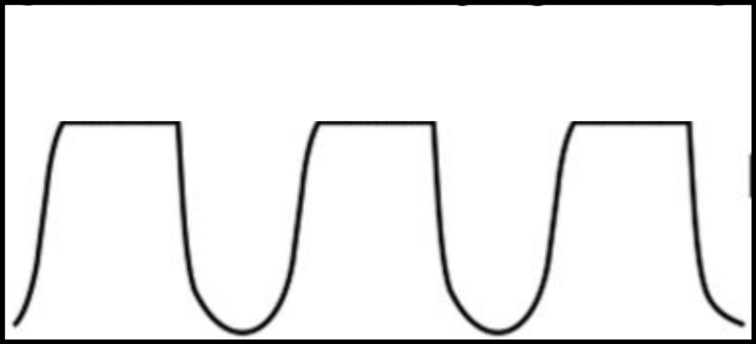
Potentiation

Neuromodulation

Chiodo & Berger, 1986; Waterhouse et al.,1998) Seamans & Yang, 2004)

CA + GABA



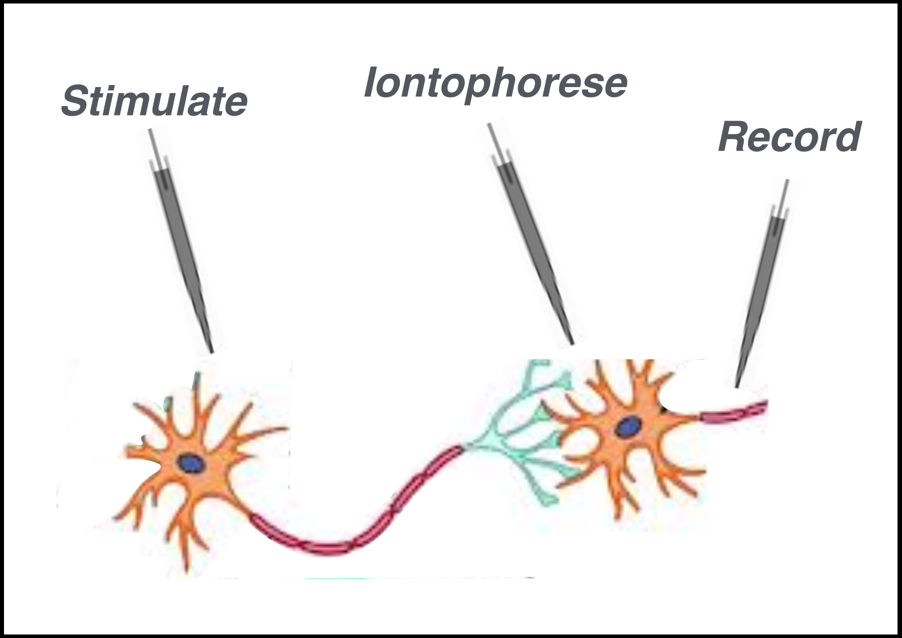


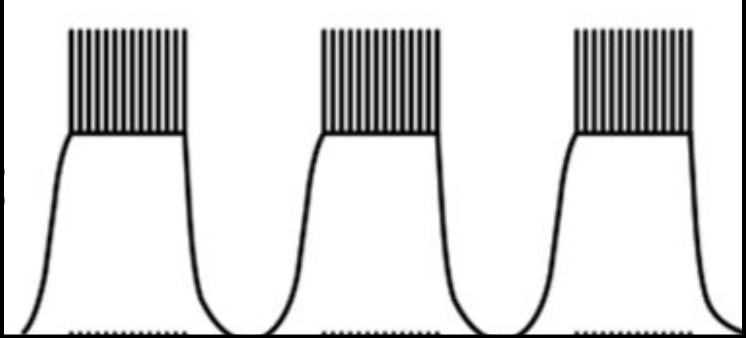
Potentiation

Neuromodulation

Chiodo & Berger, 1986; Waterhouse et al.,1998) Seamans & Yang, 2004)

CA + Stimulation



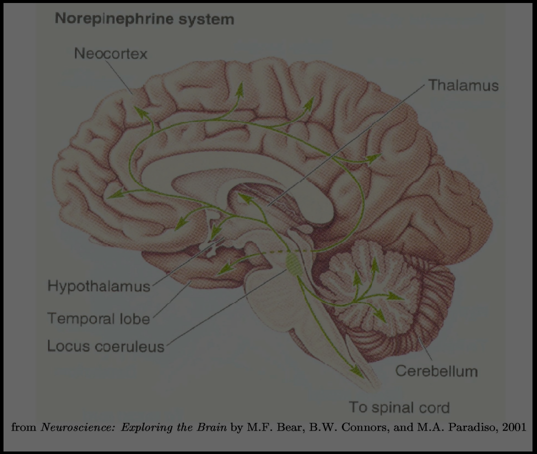


Potentiation

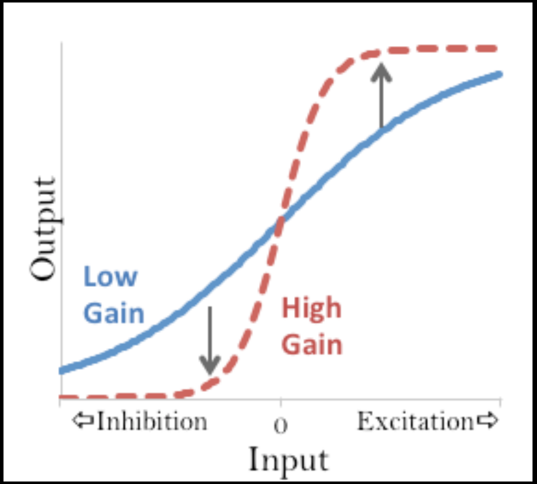
LC-NE System

- **Locus Coeruleus:**
  - A small nucleus of cells of in the rostral pontine tegmentum *(upper brainstem)*
  - Innervates all levels of neuraxis, source of 99% of norepinephrine in neocortex

- **NE is a *neuromodulator…*** *(like dopamine)*:
  - **modulates *gain* of activation function**
    Servan-Schreiber et al. (Science, 1990)



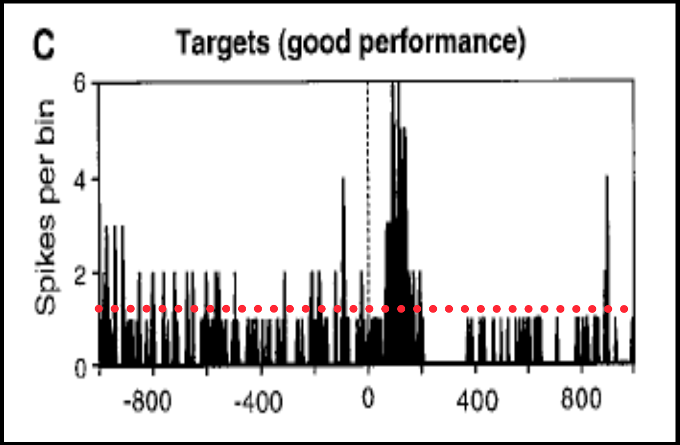
from *Neuroscience: Exploring the Brain* by M.F. Bear, B.W. Connors, and M.A. Paradiso, 2001



Two Modes of LC Function

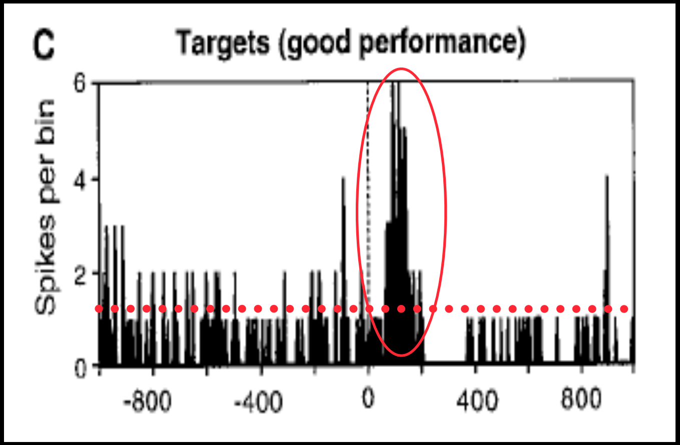
Two Modes of LC Function

- Phasic mode:
  - moderate baseline firing rate



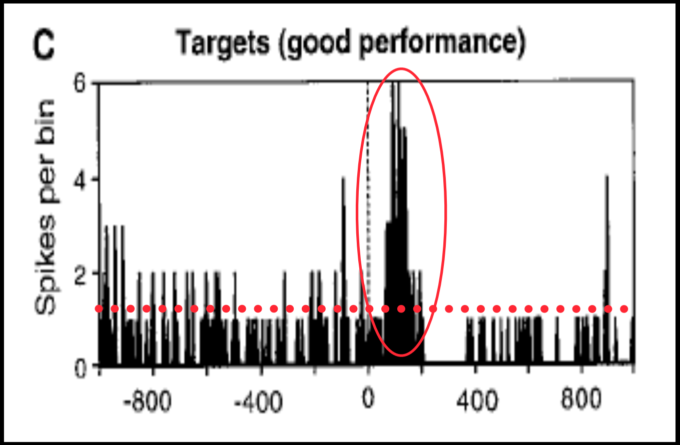
Two Modes of LC Function

- Phasic mode:
  - moderate baseline firing rate
  - phasic response to task-relevant events



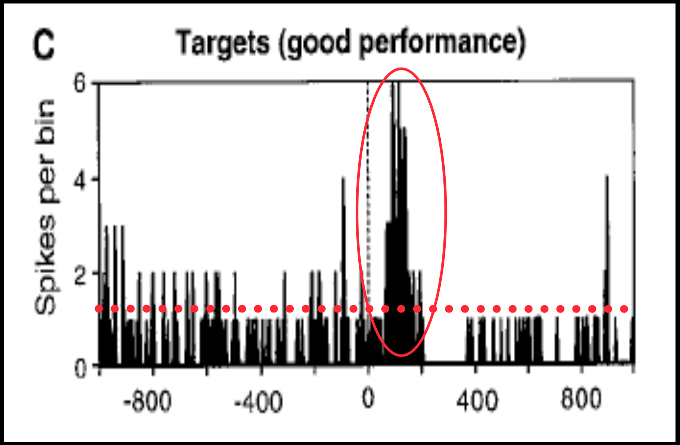
Two Modes of LC Function

- Phasic mode:
  - moderate baseline firing rate
  - phasic response to task-relevant events
  - transient increase gain (temporal filter)
    ⬆responsivity to task-relevant events



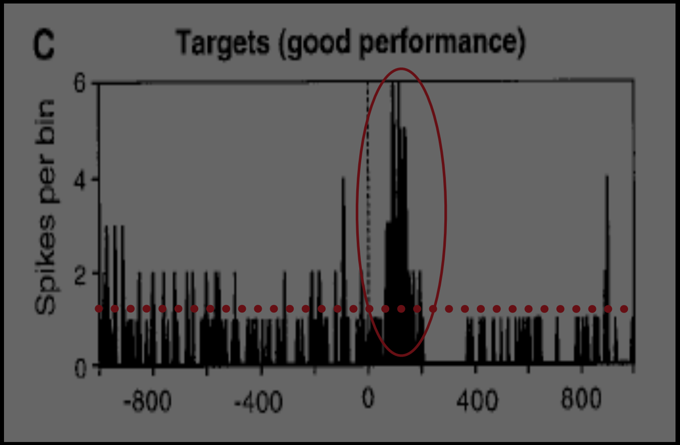
Two Modes of LC Function

- **Phasic mode:**
  - moderate baseline firing rate
  - phasic response to task-relevant events
  - transient increase gain (temporal filter)
    ⬆responsivity to task-relevant events
  - Behavior: task-focused ⇒ *exploitation*

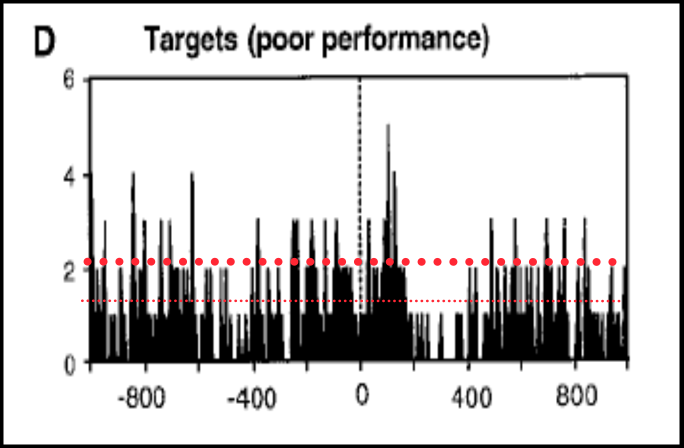


Two Modes of LC Function

- Phasic mode:
  - moderate baseline firing rate
  - phasic response to task-relevant events
  - transient increase gain (temporal filter)
    ⬆responsivity to task-relevant events
  - Behavior: task-focused ⇒ *exploitation*

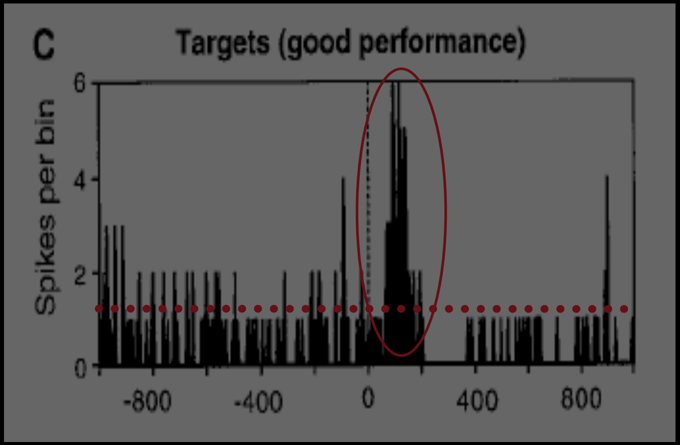


- Tonic Mode:
  - Higher baseline firing rate

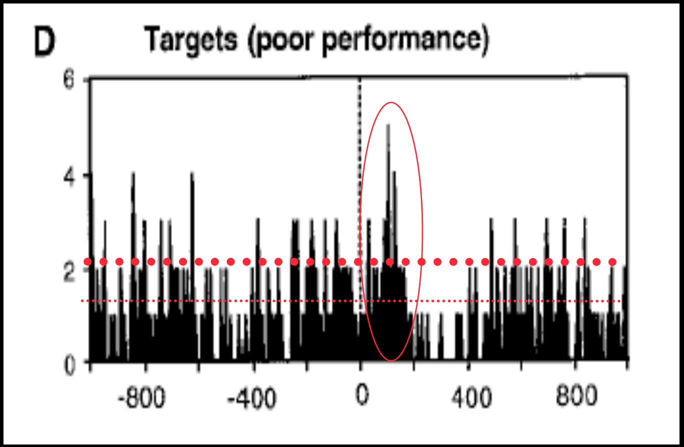


Two Modes of LC Function

- **Phasic mode:**
  - moderate baseline firing rate
  - phasic response to task-relevant events
  - transient increase gain (temporal filter) ⬆ responsivity to task-relevant events
  - Behavior: task-focused ⇒ *exploitation*

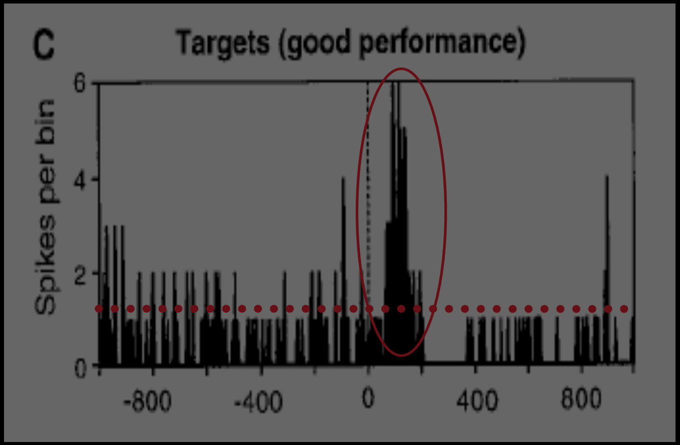


- **Tonic Mode:**
  - Higher baseline firing rate
  - diminished/absent phasic responses

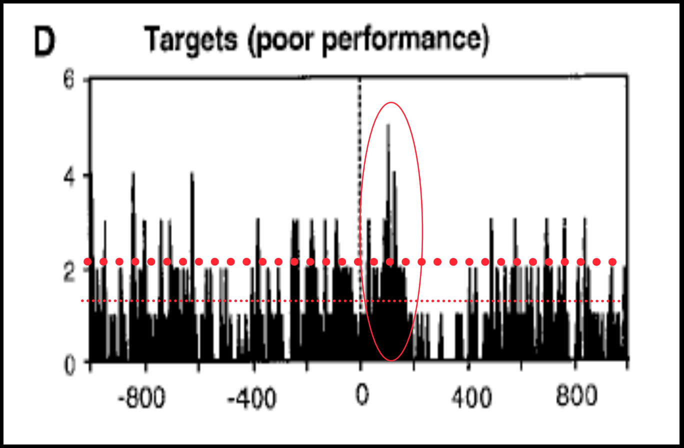


Two Modes of LC Function

- Phasic mode:
  - moderate baseline firing rate
  - phasic response to task-relevant events
  - transient increase gain (temporal filter)
    ⬆responsivity to task-relevant events
  - Behavior: task-focused ⇒ *exploitation*

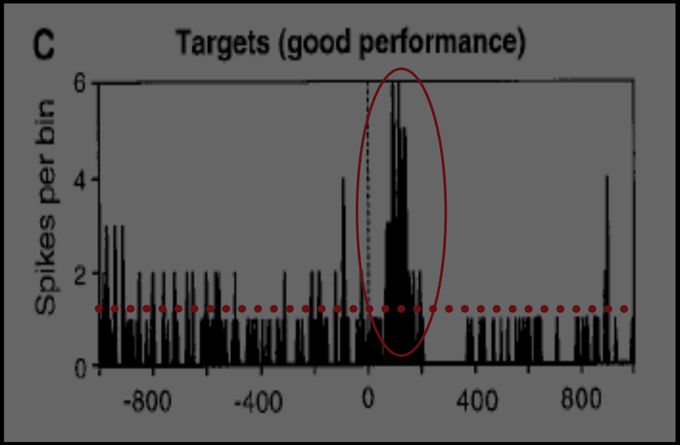


- Tonic Mode:
  - Higher baseline firing rate
  - diminished/absent phasic responses
  - indiscriminate increase in gain:
    ⬆responsivity to noise

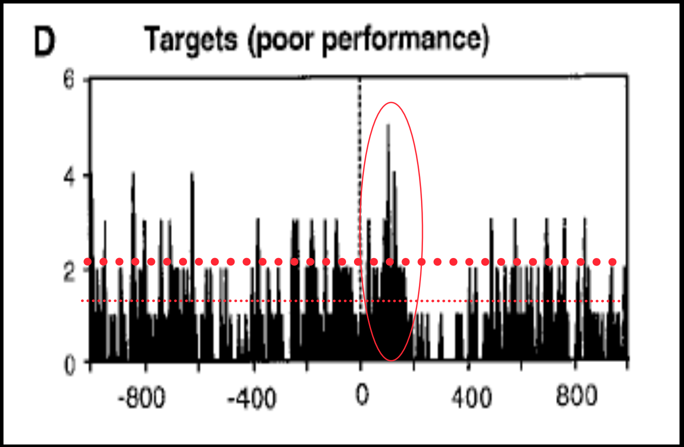


Two Modes of LC Function

- Phasic mode:
  - moderate baseline firing rate
  - phasic response to task-relevant events
  - transient increase gain (temporal filter)
    ↑responsivity to task-relevant events
  - Behavior: task-focused ⇒ *exploitation*

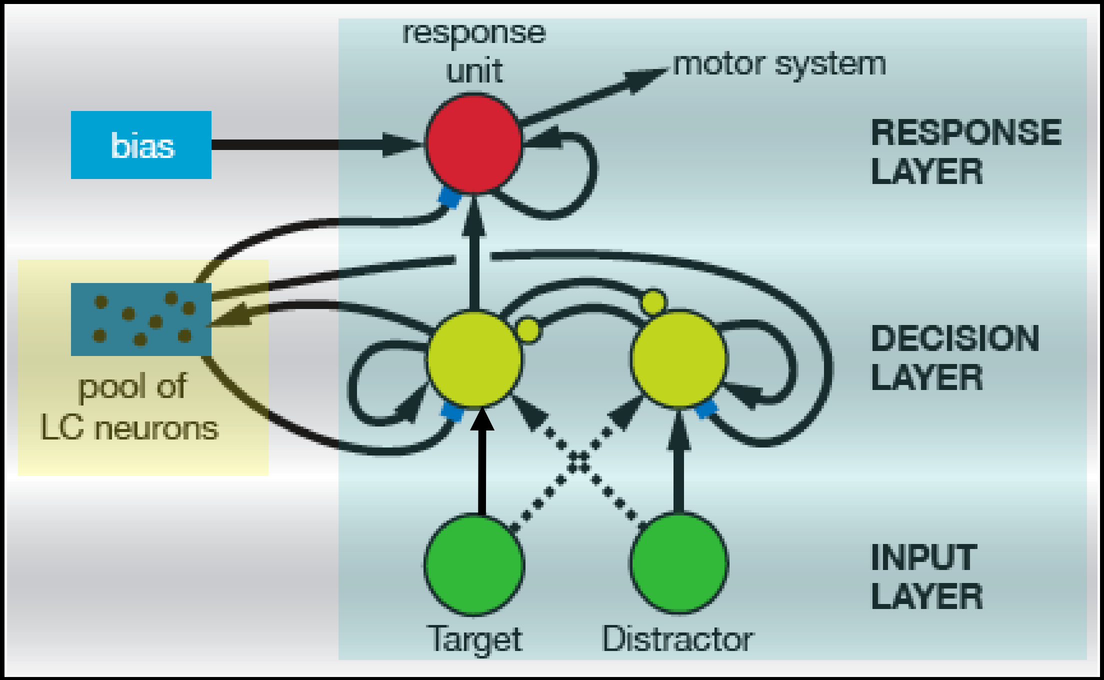


- Tonic Mode:
  - Higher baseline firing rate
  - diminished/absent phasic responses
  - indiscriminate increase in gain:
    ↑responsivity to noise
  - Behavior: distractable ⇒ *exploration*



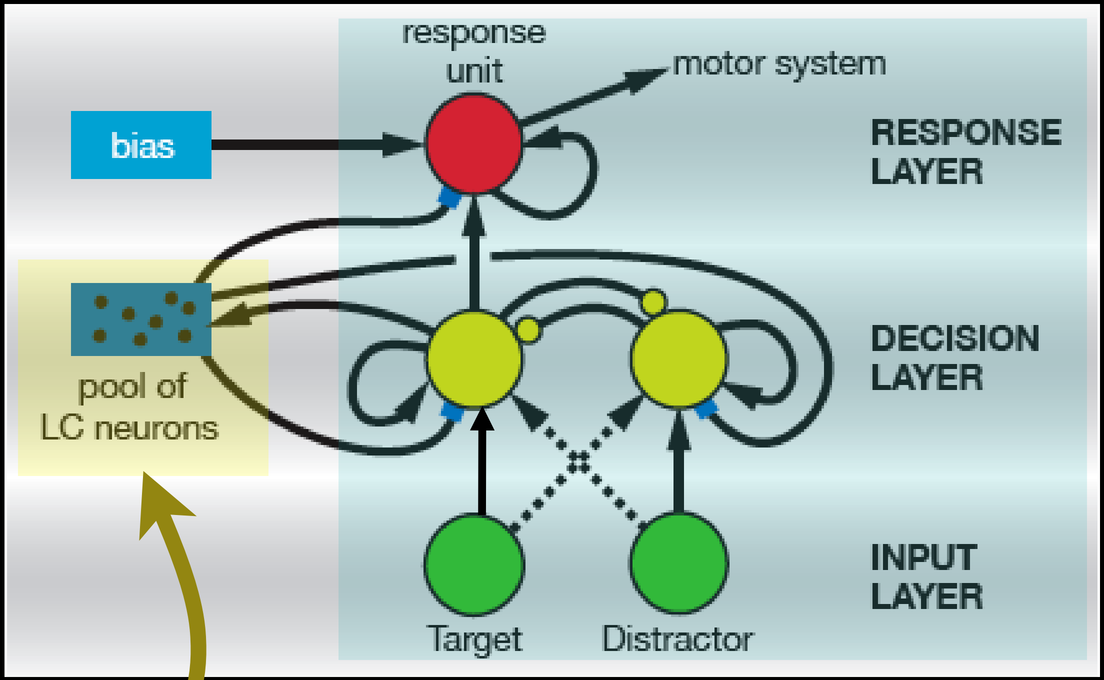
Model of the LC

Usher et al. (1999)



Model of the LC

Usher et al. (1999)



- **Neurophysiologically detailed model of LC:**
  - 250 units
  - response only to target input
  - electrotonic coupling among LC units; modulated to simulate tonic/phasic modes
  - LC units multiplicatively modulate input to decision and response layer units

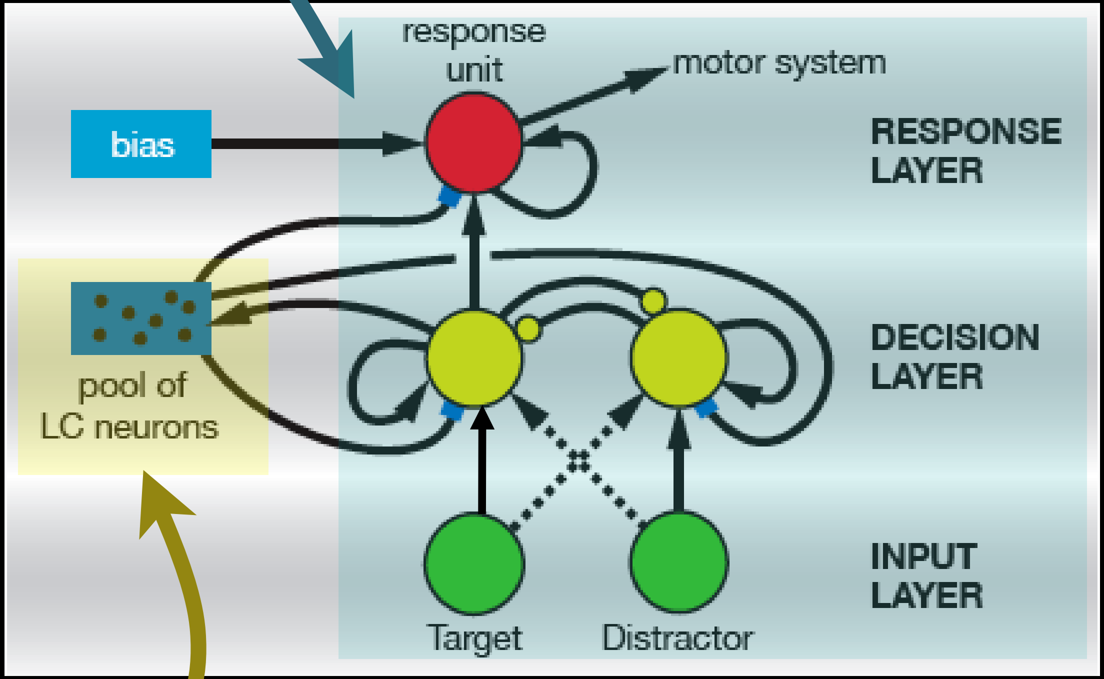
Can be abstracted using FitzHugh-Nagumo simplification (Gilzenrat et al, Neural Networks 2004)

Model of the LC

Usher et al. (1999)

- **Simple connectionist model of behavioral task:**
  - mutual inhibition between competing processing units
  - **distractor stimuli weakly activate target decision units**
  - **noise**

- **Neurophysiologically detailed model of LC:**
  - **250 units**
  - **response only to target input**
  - **electrotonic coupling among LC units; modulated to simulate tonic/phasic modes**
  - **LC units multiplicatively modulate input to decision and response layer units**

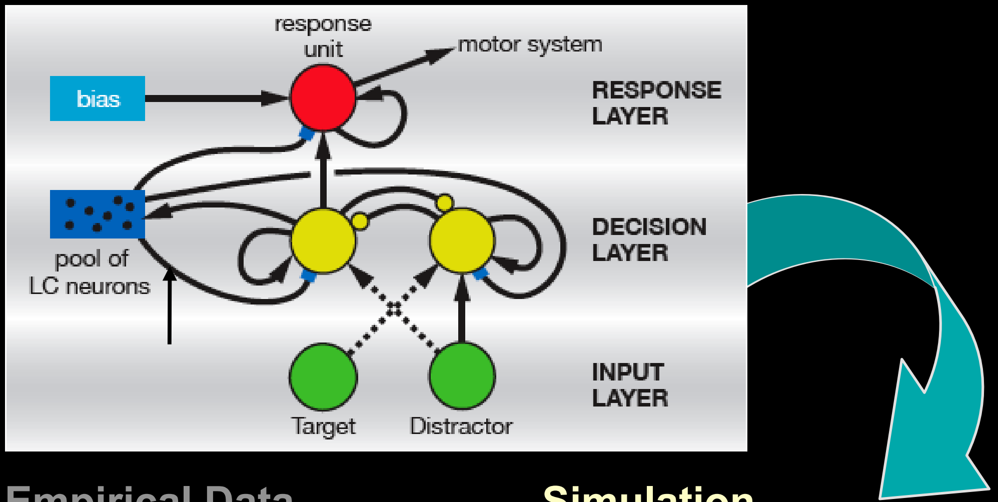


Can be abstracted using FitzHugh-Nagumo simplification (Gilzenrat et al, Neural Networks 2004)

Model Simulates Two Modes of LC Function

Usher et al. (1999)

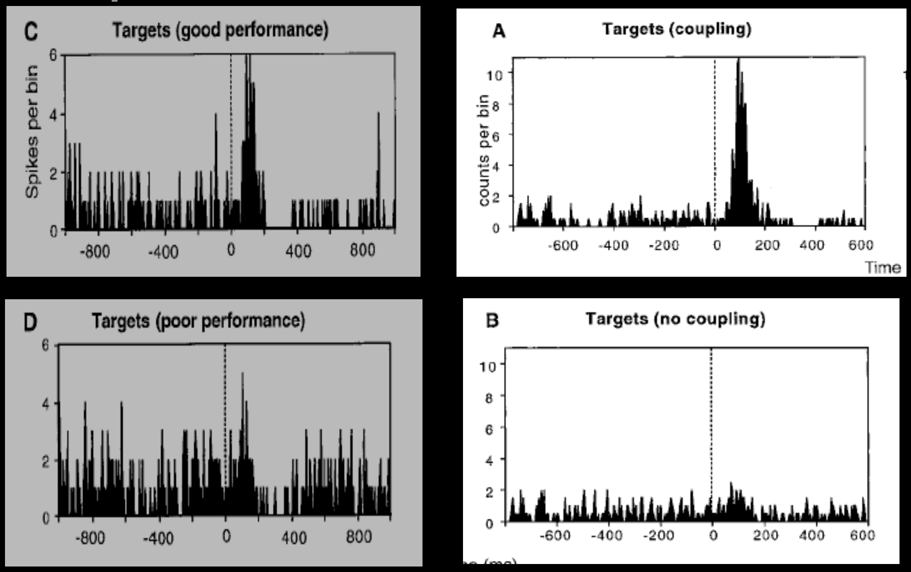
- Change in single parameter in LC *(electronic coupling)*
  - Increase in phasic response
  - Decrease in tonic activity



Empirical Data **Simulation**

Phasic Mode:

Tonic Mode:



Mechanisms of LC Modulation

- **Electrotonic coupling** *(Usher et al., 1999)*
- **External drive** *(Alvarez & Chow, 2001; Brown et al., 2004)*
  - **Biophysically realistic, coupled oscillators model…**

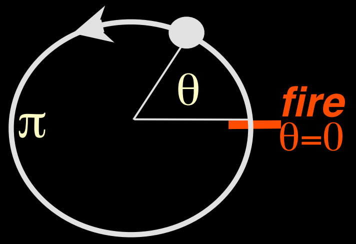
Simple Phase Oscillator Model

Ermentrout, 1996

phase sensitivity **sensory input** **coupling**

$$\frac{d\theta}{dt} = \omega + z(\theta)\,[I(t) + \eta(t) + \sum \ldots]$$

natural frequency **“noise”**



This is a formal reduction of the Hindmarsh-Rose conductance-based model of neuronal firing

(Rose & Hindmarsh, 1989)

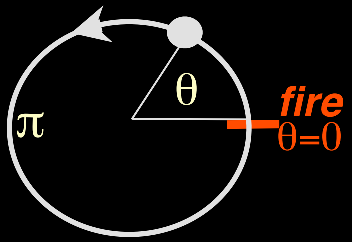
Simple Phase Oscillator Model

Ermentrout, 1996

phase sensitivity **sensory input** **coupling**

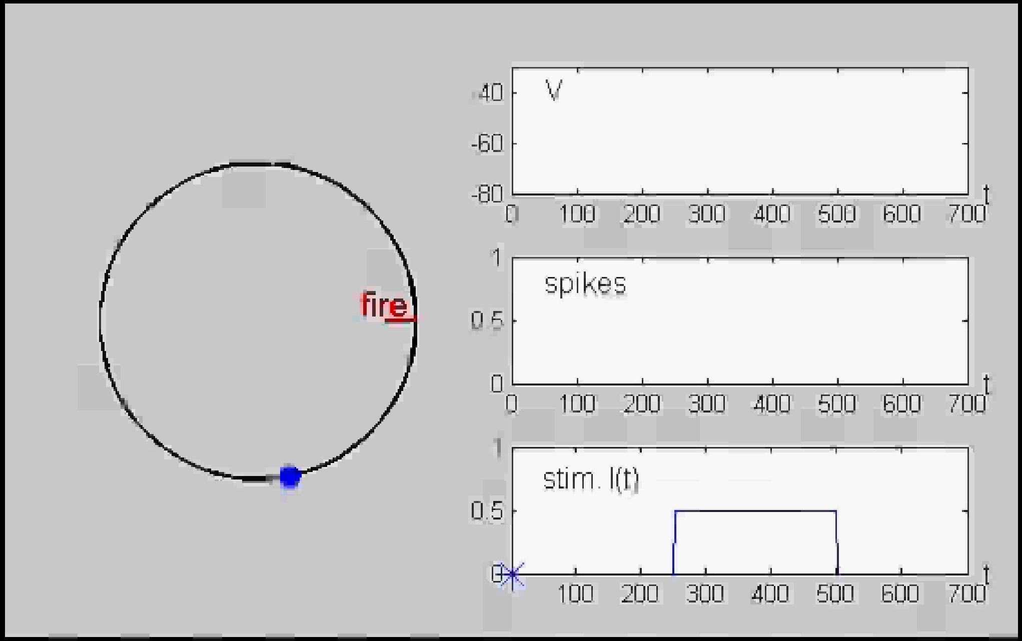
$$\frac{d\theta}{dt} = \omega + z(\theta)\,[I(t) + \eta(t) + \sum \ldots]$$

natural frequency **“noise”**



This is a formal reduction of the Hindmarsh-Rose conductance-based model of neuronal firing

(Rose & Hindmarsh, 1989)



Population of Coupled Phase Oscillators

Brown et al. (J. Neural Computation , 2004)

phase sensitivity

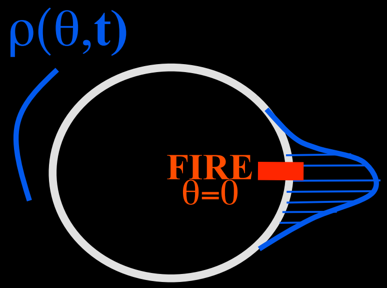
sensory input

coupling

$$\frac{d\theta}{dt} = \omega + z(\theta)\left[I(t) + \eta(t) + \sum \ldots\right]$$

natural frequency

“noise”



Tonic mode: high baseline drive (3 Hz)

Phasic mode: low baseline drive (1 Hz)

Population of Coupled Phase Oscillators

Brown et al. (J. Neural Computation , 2004)

phase sensitivity

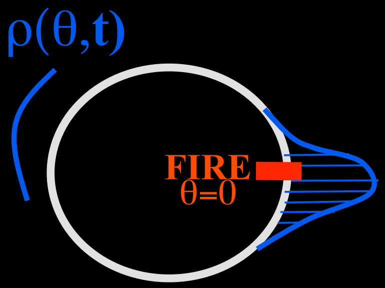
sensory input

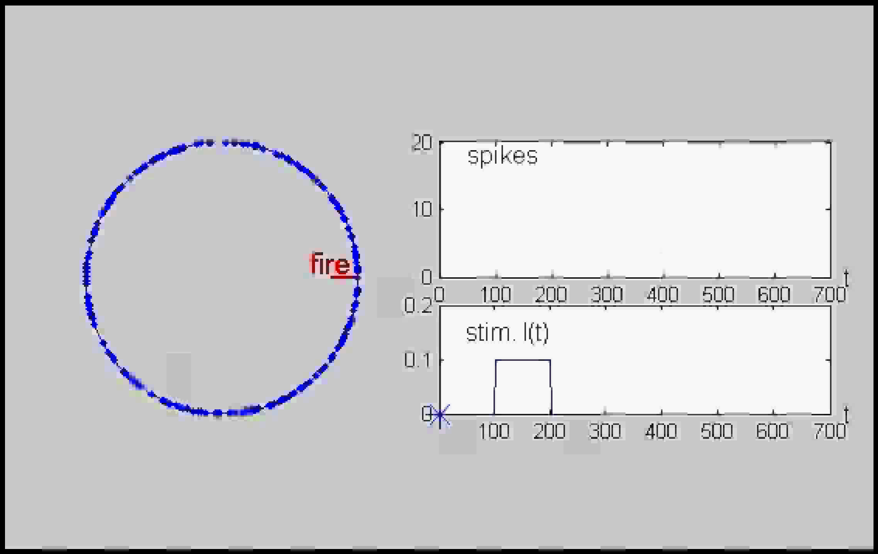
coupling

$$\frac{d\theta}{dt} = \omega + z(\theta)\,[I(t) + \eta(t) + \sum \ldots]$$

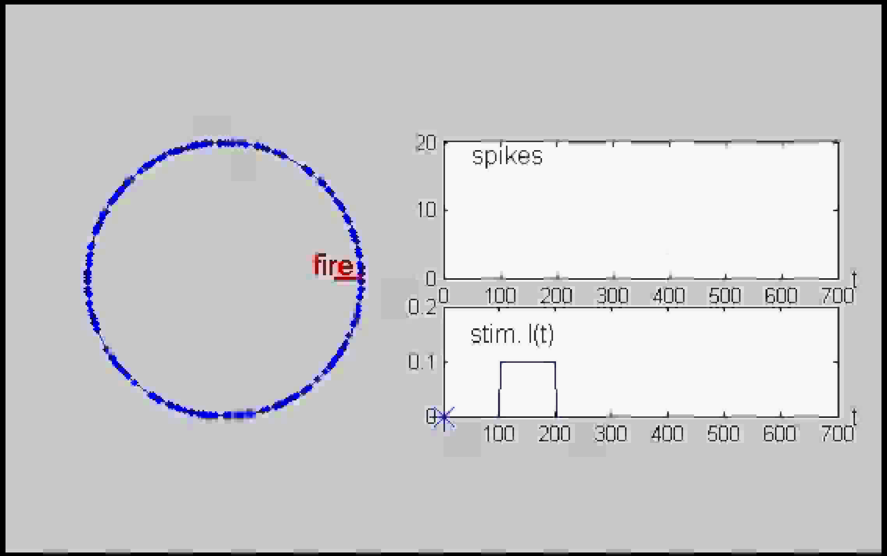
natural frequency

“noise”





Tonic mode: **high baseline drive (3 Hz)**



Phasic mode: **low baseline drive (1 Hz)**

LC Phasic Mode
& Optimization of Decision Making

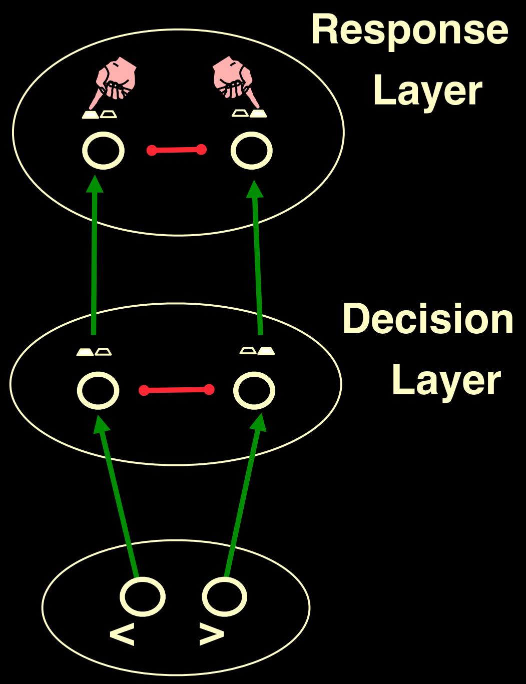
- LC *phasic response* transiently increases gain following completion of decision process:

  optimizes performance in multilayer systems

LC Phasic Mode
& Optimization of Decision Making

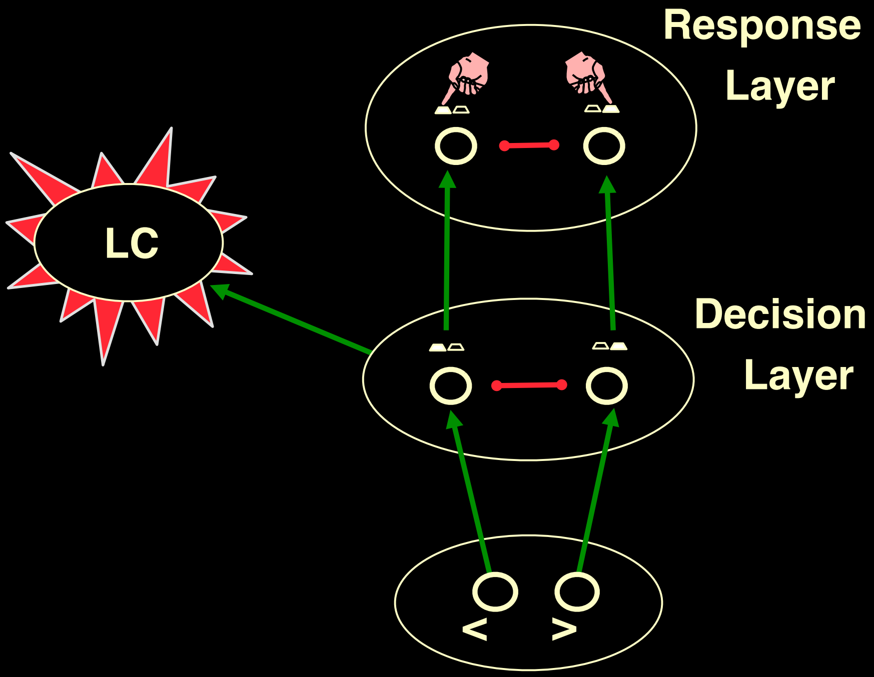
- LC *phasic response* transiently increases gain following completion of decision process

- Simplest multi-layered system:
  - Decision process occurs at one level
  - Response mechanism at a subsequent level executes decision

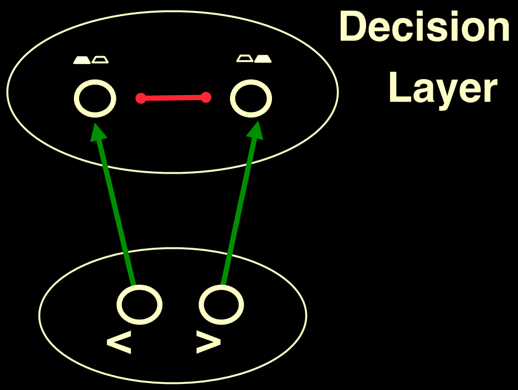


LC Phasic Mode & Optimization of Decision Making

- LC *phasic response* transiently increases gain following completion of decision process
- Simplest multi-layered system:
  - Decision process occurs at one level
  - Response mechanism at a subsequent level executes decision
  - LC driven by decision process



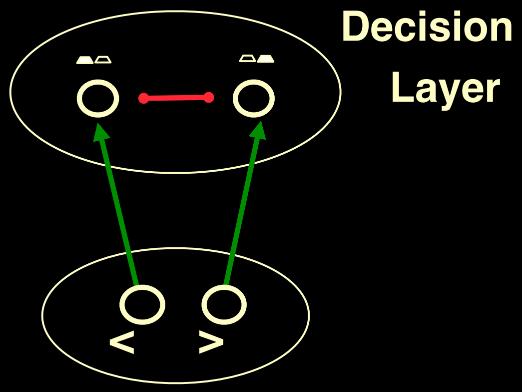
LC Phasic Mode
& Optimization of Decision Making



LC Phasic Mode
& Optimization of Decision Making

- LC *phasic response* transiently increases gain following completion of decision process

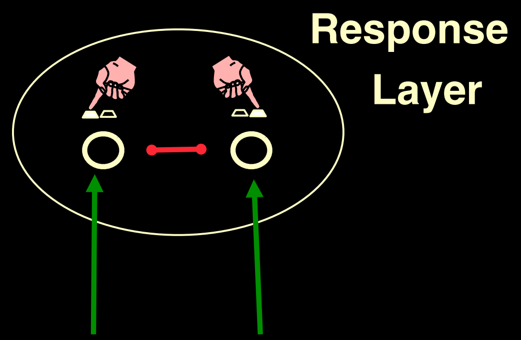
- Simplest multi-layered system:
  - Decision process occurs at one level



LC Phasic Mode
& Optimization of Decision Making

- LC *phasic response* transiently increases gain following completion of decision process

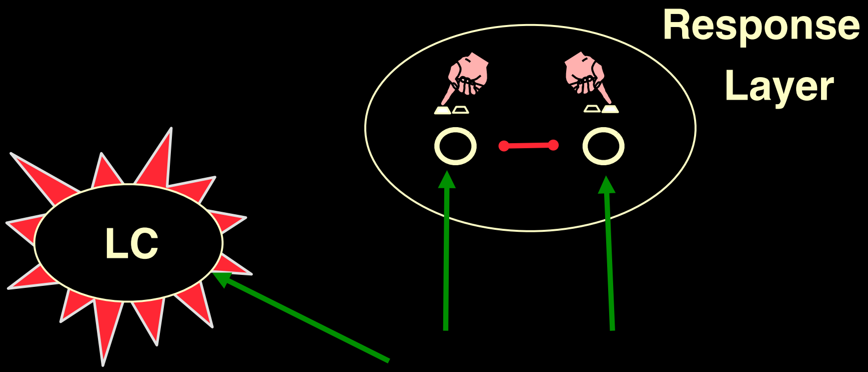
- Simplest multi-layered system:
  - Response mechanism at a subsequent level executes decision



LC Phasic Mode
& Optimization of Decision Making

- LC *phasic response* transiently increases gain following completion of decision process

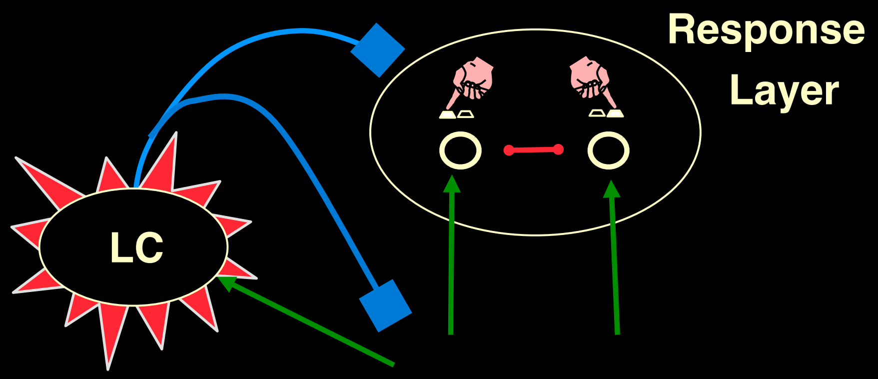
- Simplest multi-layered system:
  - LC driven by decision process



LC Phasic Mode
& Optimization of Decision Making

- LC *phasic response* transiently increases gain following completion of decision process

- **Simplest multi-layered system:**

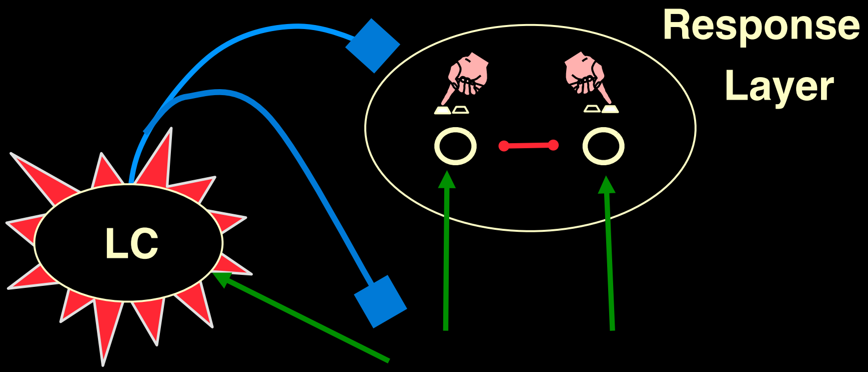


♦ increases gain globally throughout the system

LC Phasic Mode
& Optimization of Decision Making

- LC *phasic response* transiently increases gain following completion of decision process

- **Simplest multi-layered system:**

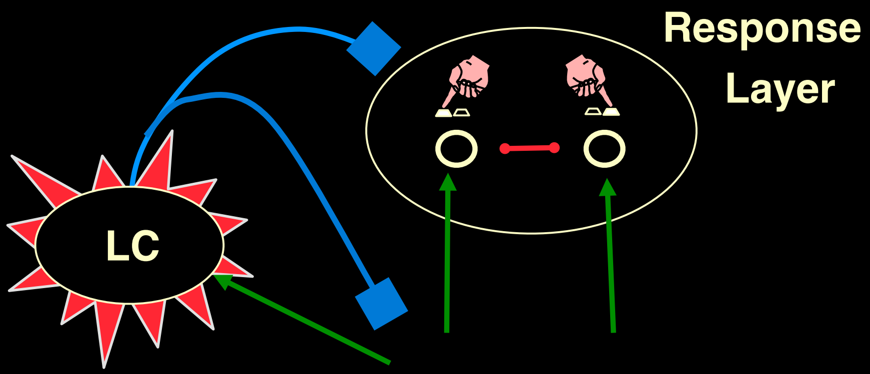


♦ forces “read-out” of response as soon as decision process has crossed threshold

LC Phasic Mode
& Optimization of Decision Making

- LC *phasic response* transiently increases gain following completion of decision process

- **Simplest multi-layered system:**

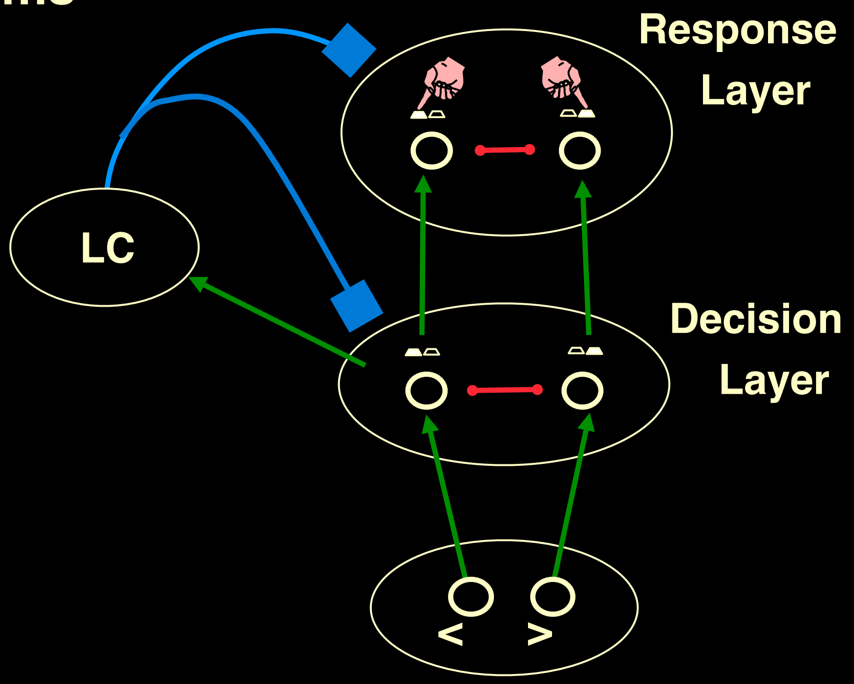


⇒“collapses” processing in system
around outcome of decision process

LC Tonic Mode & Exploration

- LC *phasic response* transiently increases gain following completion of decision process:

optimizes performance in multilayer systems

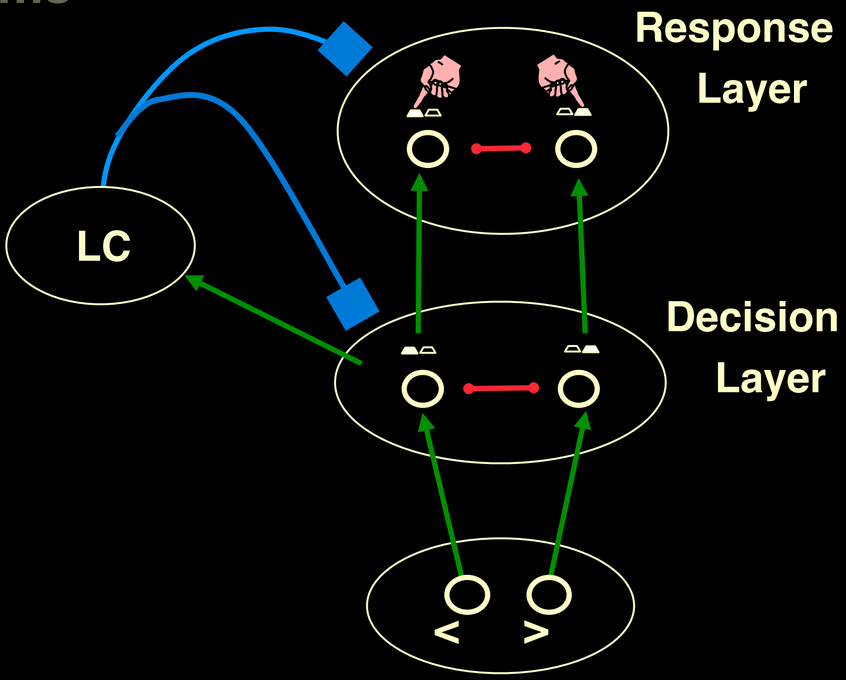


LC Tonic Mode & Exploration

- LC *phasic response* transiently increases gain following completion of decision process:

  optimizes performance in multilayer systems

  mediates tradeoff between
  complexity *(multilayer system)*
  and *efficiency*



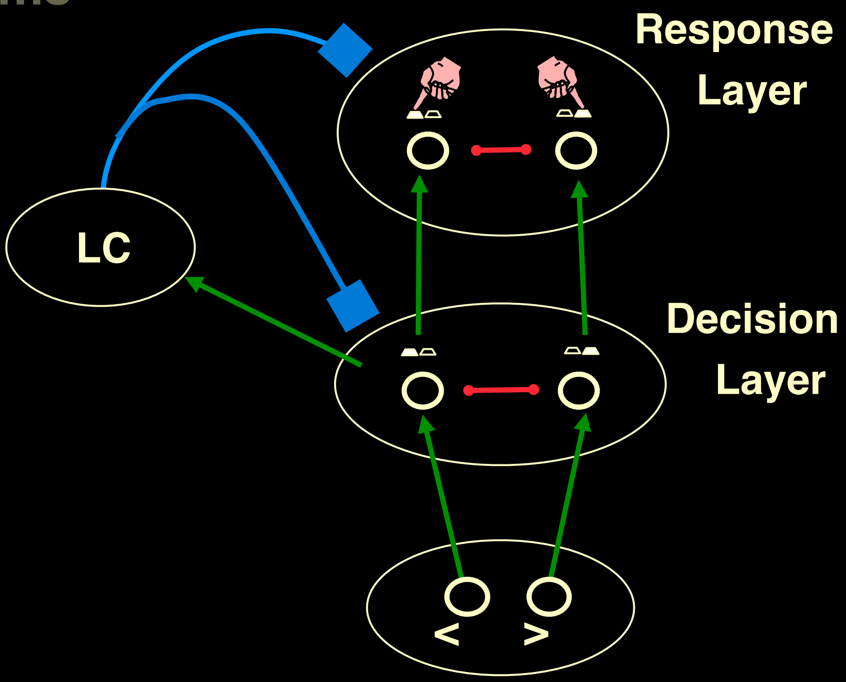
LC Tonic Mode & Exploration

- LC *phasic response* transiently increases gain following completion of decision process:

optimizes performance in multilayer systems

mediates tradeoff between
complexity *(multilayer system)*
and *efficiency*

⇒ *exploitation*



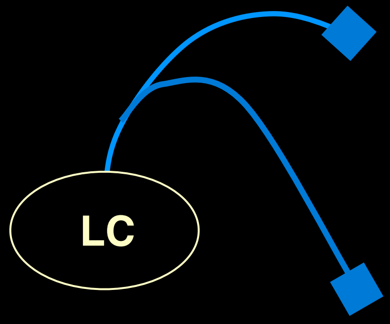
LC Tonic Mode & Exploration

- LC *phasic response* transiently increases gain following completion of decision process:

  optimizes performance in multilayer systems

  mediates tradeoff between
  complexity *(multilayer system)*
  and *efficiency*

  ⇒ *exploitation*



- LC *tonic response* produces a sustained, indiscriminate increase in gain throughout the system

tantamount to increasing noise

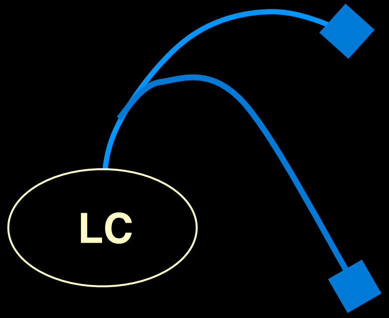
LC Tonic Mode & Exploration

- LC *phasic response* transiently increases gain following completion of decision process:

  optimizes performance in multilayer systems

  mediates tradeoff between
  complexity *(multilayer system)*
  and *efficiency*

  ⇒ *exploitation*

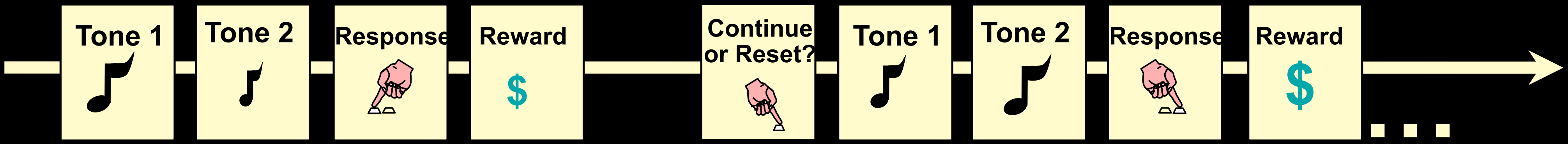


- LC *tonic response* produces a sustained, indiscriminate increase in gain throughout the system

  ⇒ *random exploration*

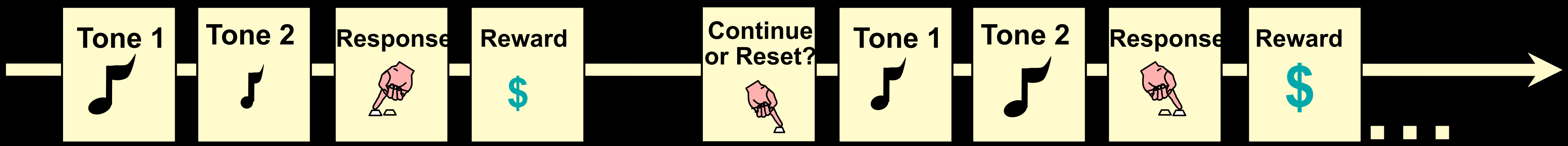
Diminishing Utility (Foraging) Task

Gilzenrat et al. (2010)



Diminishing Utility (Foraging) Task

Gilzenrat et al. (2010)



Tone 1

Tone 2

Response

Reward

Continue or Reset?

Tone 1

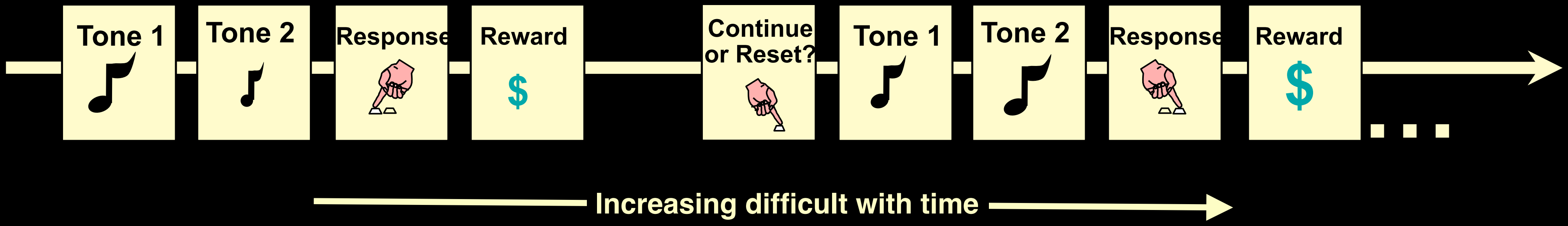
Tone 2

Response

Reward

Diminishing Utility (Foraging) Task

Gilzenrat et al. (2010)



Tone 1

Tone 2

Response

Reward

Continue or Reset?

Tone 1

Tone 2

Response

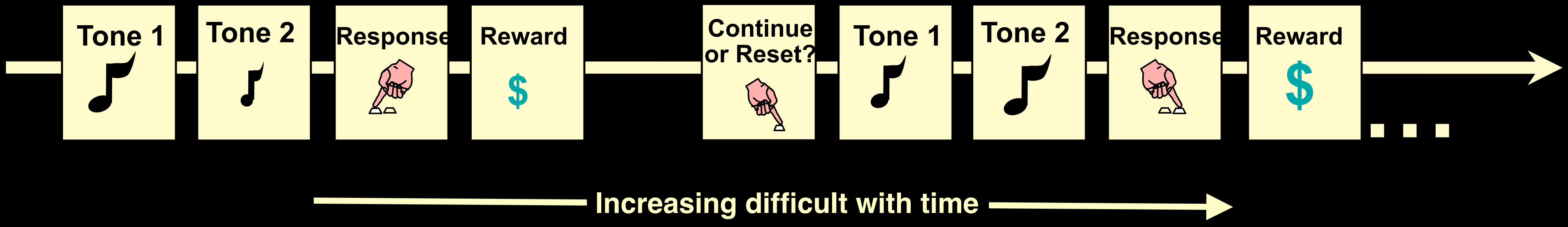
Reward

. . .

Increasing difficult with time

Diminishing Utility (Foraging) Task

Gilzenrat et al. (2010)



Tone 1

Tone 2

Response

Reward

Continue or Reset?

Tone 1

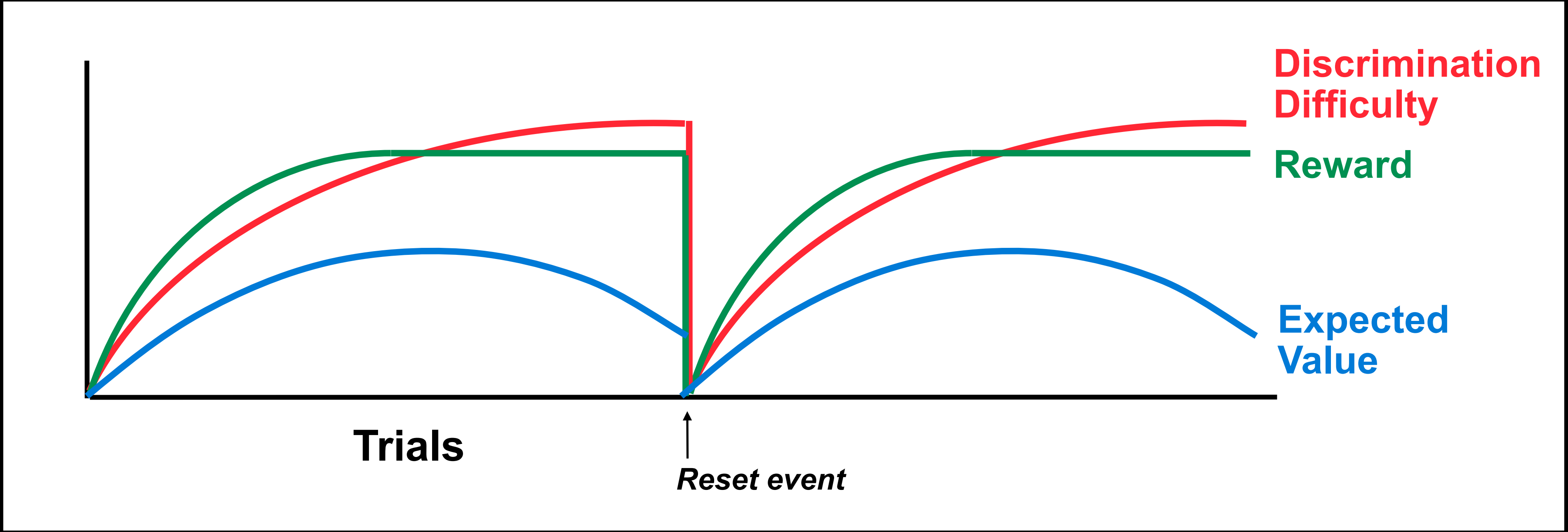
Tone 2

Response

Reward

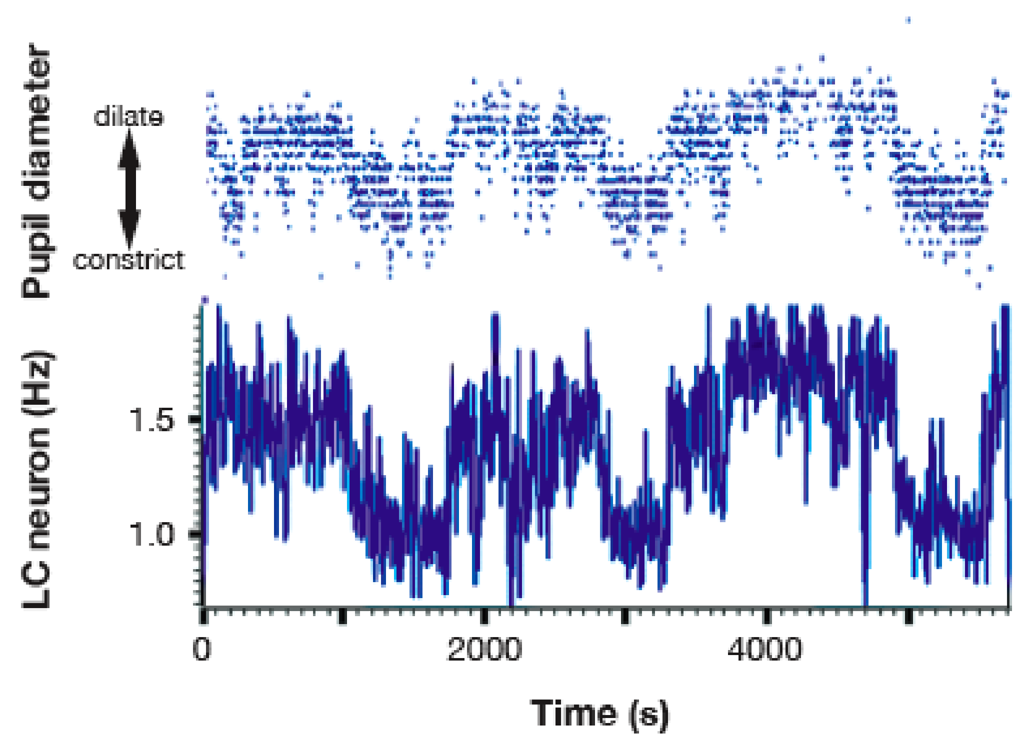
...

Increasing difficult with time



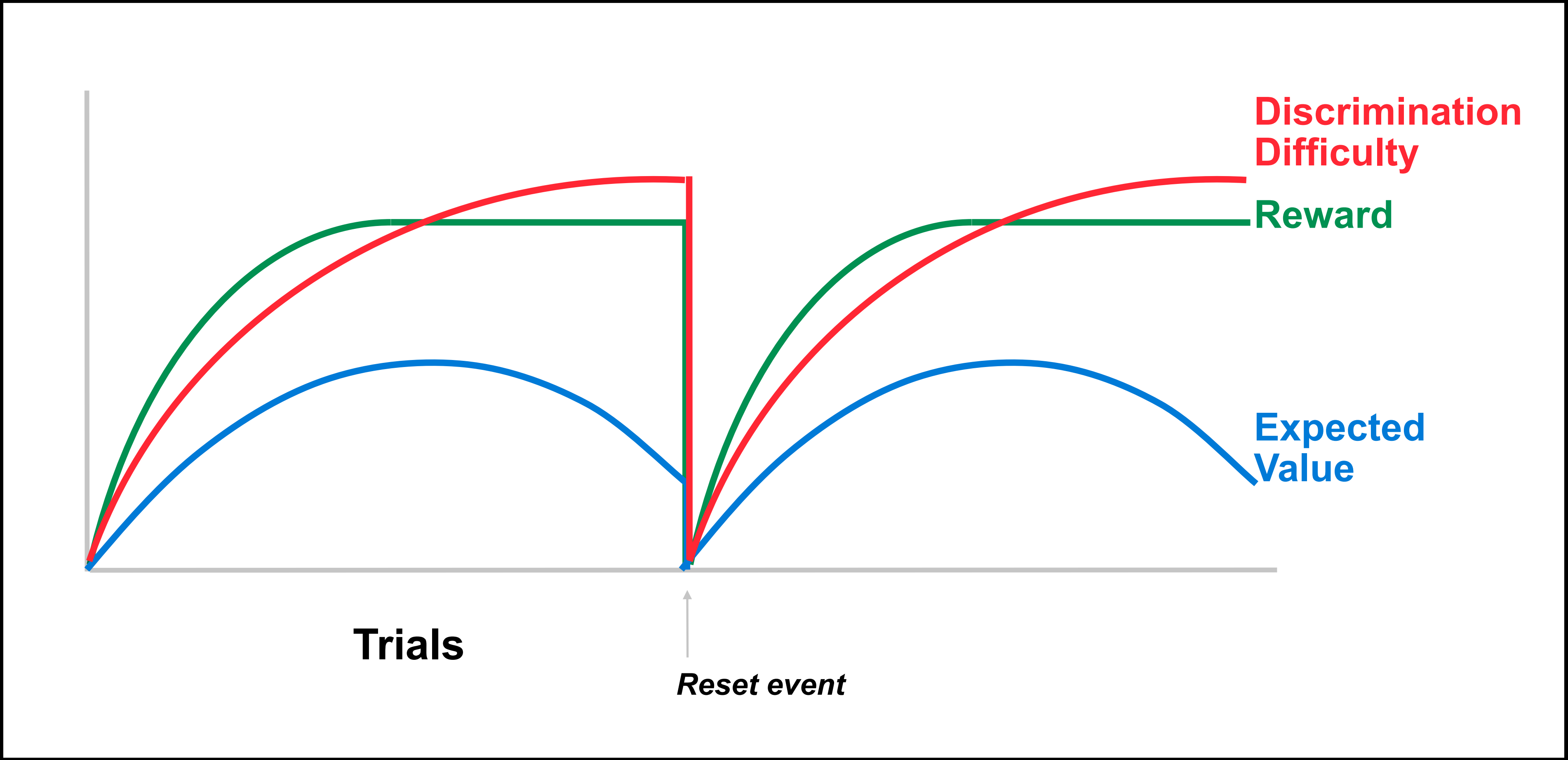
LC and the Pupil

Aston-Jones & Cohen (2005)



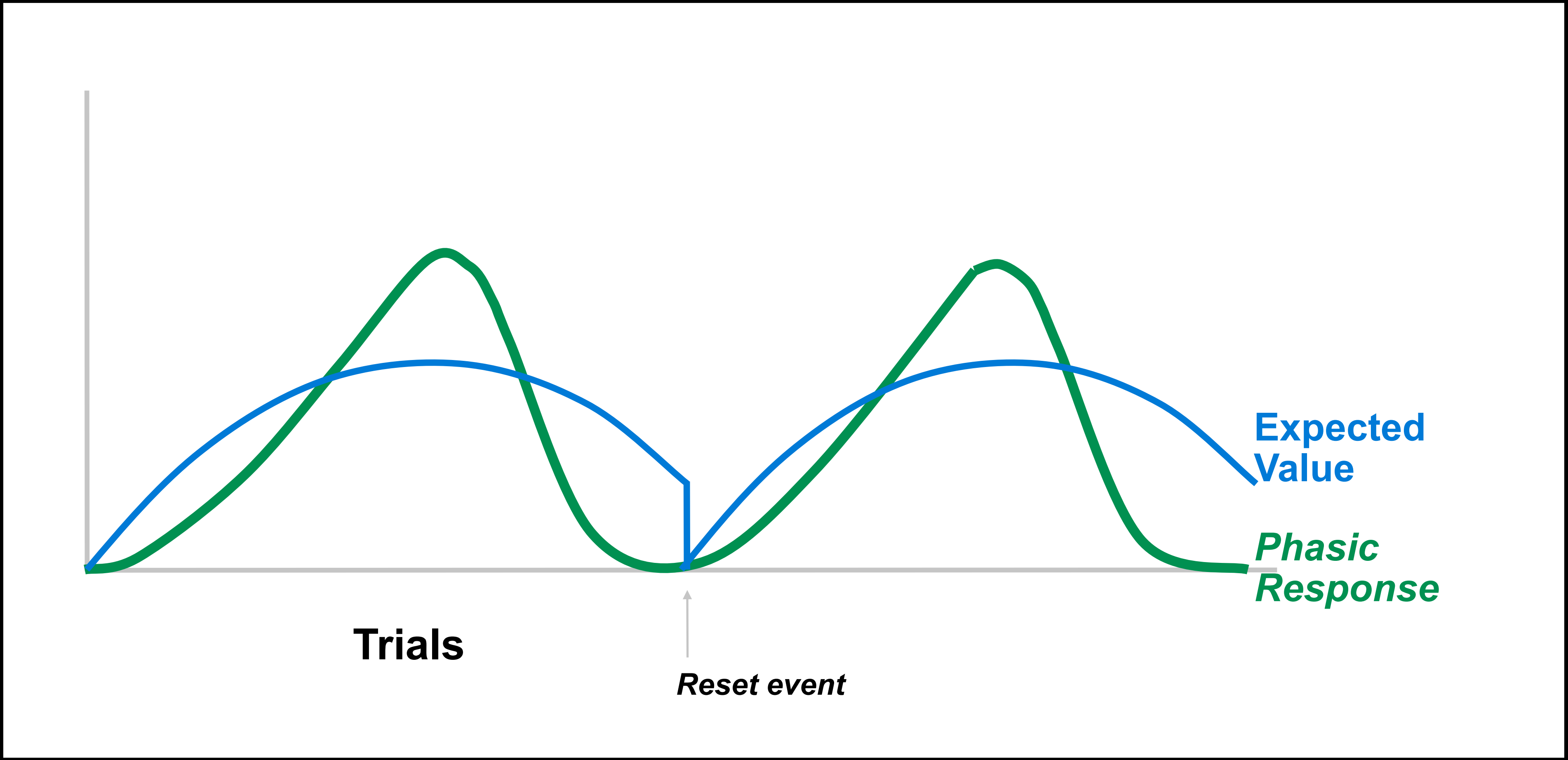
Diminishing Utility (Foraging) Task

Gilzenrat et al. (2010)



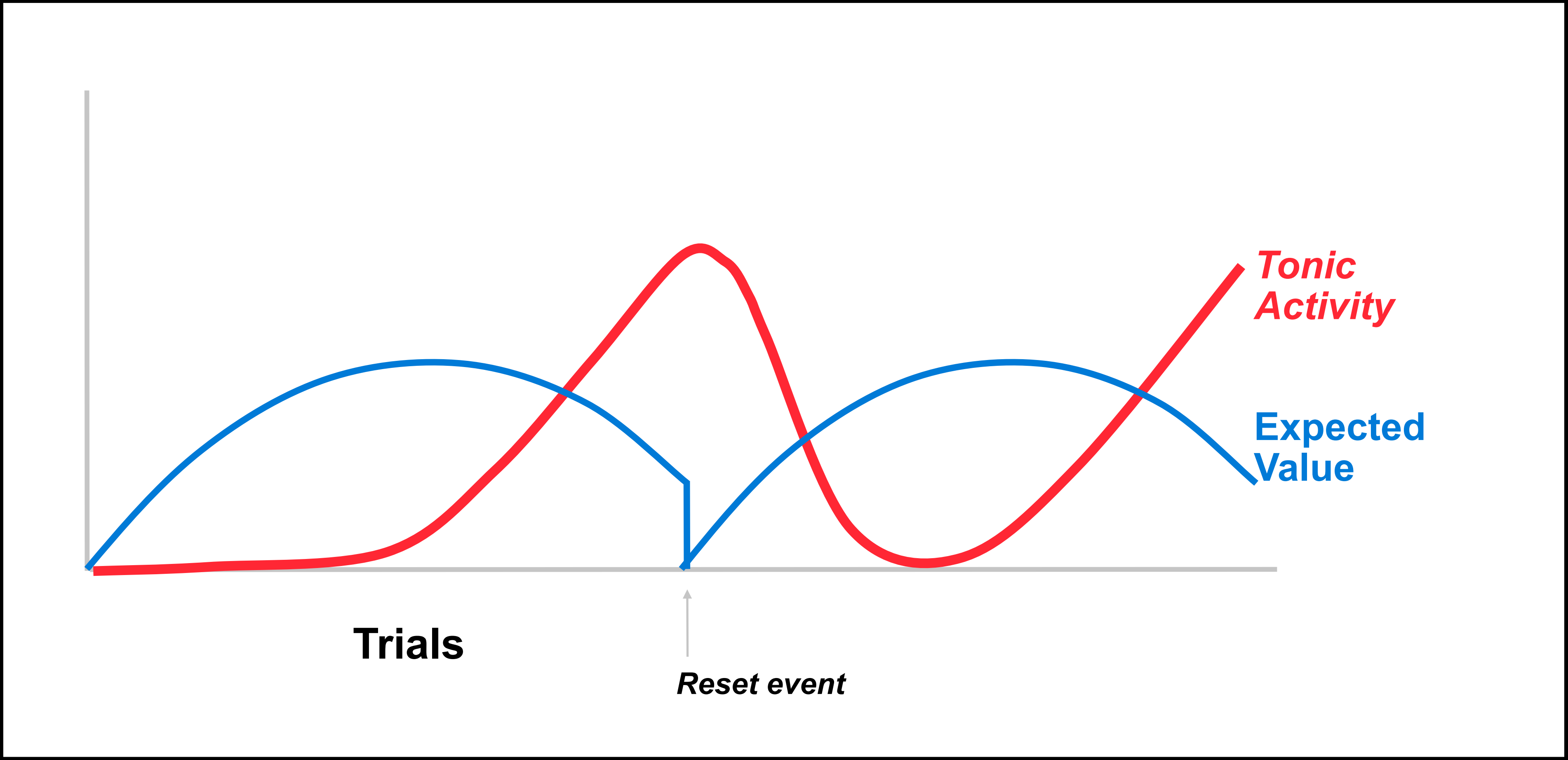
Diminishing Utility (Foraging) Task

Gilzenrat et al. (2010)



Diminishing Utility (Foraging) Task

Gilzenrat et al. (2010)



Diminishing Utility (Foraging) Task

Gilzenrat et al. (2010)

<u>Baseline</u> Pupil Diameter
(Tonic Response)

Expected Value

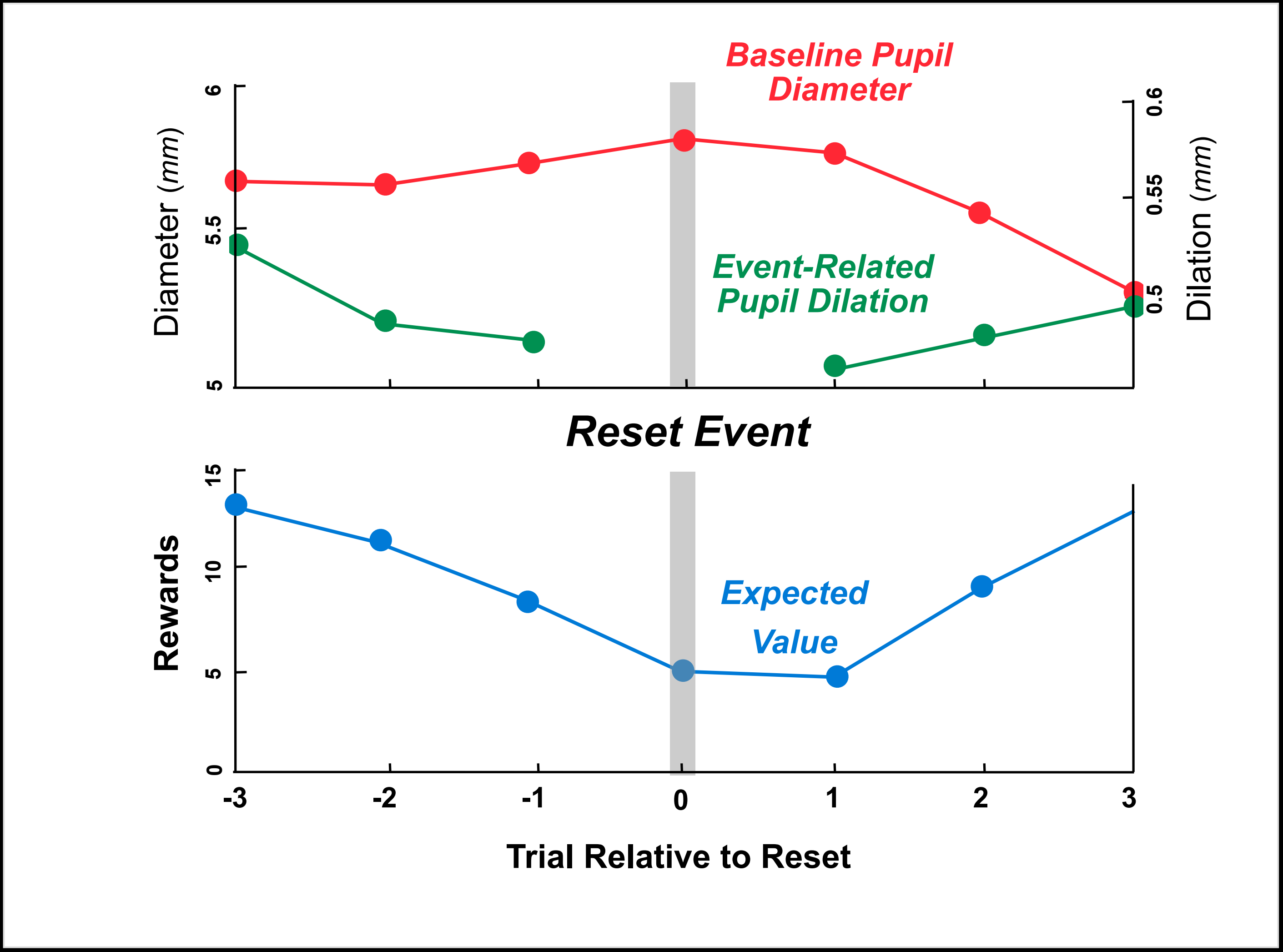
Event-Related <u>Pupil Dilation</u>
(Phasic Response)

Diminishing Utility (Foraging) Task

Gilzenrat et al. (2010)

Diminishing Utility (Foraging) Task

Gilzenrat et al. (2010)



The Explore/Exploit Tradeoff

The Explore/Exploit Tradeoff

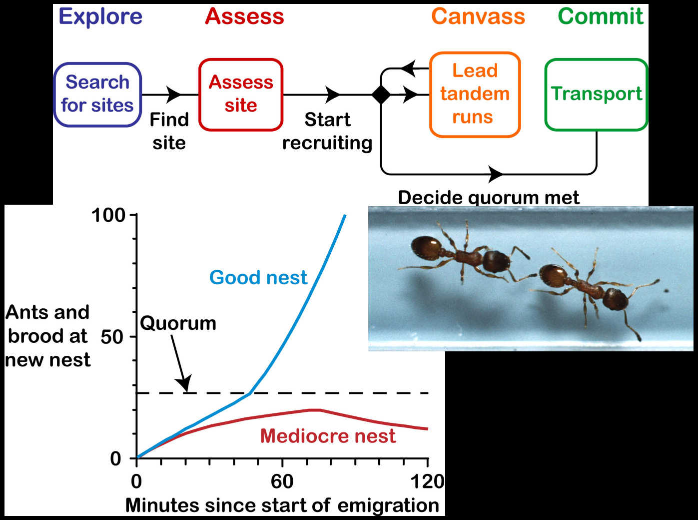
- **All species exhibit it**
  - **fungi** *(Watkinson et al. 2005)*

HYPHAE

These are the vegetative, filamentous, tube-like organs of most fungi. When a spore germinates, what emerges is a hypha (sometimes more than one hypha), which grows at its tip. The picture below shows hyphae emerging from spores and looking for food.

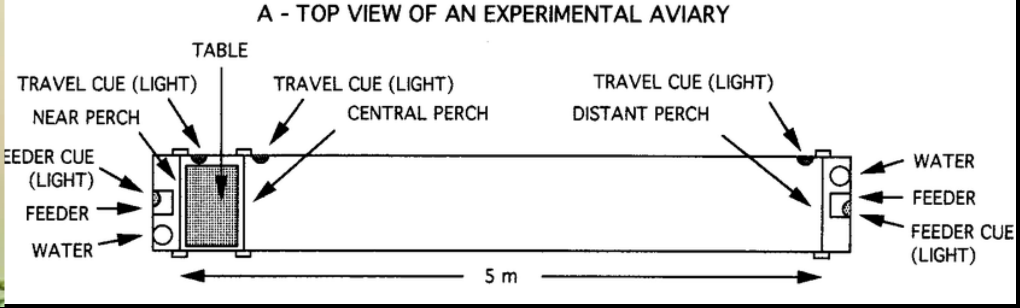
The Explore/Exploit Tradeoff

- **All species exhibit it**
  - **fungi** *(Watkinson et al. 2005)*
  - **ants** *(Pratt & Sumpter 2006)*



The Explore/Exploit Tradeoff

- **All species exhibit it**
  - **fungi** *(Watkinson et al. 2005)*
  - **ants** *(Pratt & Sumpter 2006)*
  - **birds** *(Kacelnik et al.)*



The Explore/Exploit Tradeoff

- **All species exhibit it**
  - – **fungi** *(Watkinson et al. 2005)*
  - – **ants** *(Pratt & Sumpter 2006)*
  - – **birds** *(Kacelnik et al.)*
  - – **people** *(Daw et al., 2006; Wilson et al., 2014)…*

STARTERS

Pizza Bianca (White Pizza) Sprinkled with Parmesan Cheese, Oregano, and Olive Oil.	$5.50
Mozzarella in Carozza Fried Mozzarella with Anchovies, Garlic and Cream Sauce.	$6.50
Calzone Made fresh on the premises.	$5.50

PASTA

Fettuccine Alfredo Egg Noodles in Cream Sauce and Parmesan Cheese.	$10.95
Seafood Scampi Shrimp, scallops, squid sautéed in garlic, olive oil. Served on fresh linguine.	$16.95
New **Ravioli di Ricotta** Pasta Stuffed with Ricotta and Parmesan Cheese. Served with Tomato Sauce.	$11.95
Linguine Vegetariani Sautéed in butter with mushrooms, zucchini and red pepper.	$13.00

PIZZA

	10"	14"
Cheese and Tomato based on the "Pines of Rome" classic	$11.75	$13.75
Fresh Herb a blend of herbs, topped with an olive oil & chèvre cheese	$14.75	$16.95
New **Thai Chicken** chicken marinated in a spicy peanut-ginger sesame sauce	$15.75	$17.95
Garden Veggie olives, onions, green peppers, & tomatoes	$13.75	$15.95
Greek Pizza mediterranean spiced grilled chicken & mozzarella	$14.75	$16.75
New **The Works** Canadian bacon, salami, pepperoni, sausage, beef, olives, pineapple, mushrooms, onions, green peppers, etc. etc.	$18.95	22.95
Hawaiian BBQ fresh pineapple, BBQ chicken, & smoked gouda	$16.75	$18.95

The Explore/Exploit Tradeoff

- **All species exhibit it**
  - **fungi** *(Watkinson et al. 2005)*
  - **ants** *(Pratt & Sumpter 2006)*
  - **birds** *(Kacelnik et al.)*
  - **people** *(Daw et al., 2006; Wilson et al., 2014)…*

STARTERS

Pizza Bianca (White Pizza) ... $5.50
Sprinkled with Parmesan Cheese, Oregano, and Olive Oil.

Mozzarella in Carozza $6.50
Fried Mozzarella with Anchovies, Garlic and Cream Sauce.

Calzone $5.50
Made fresh on the premises.

PASTA

Fettuccine Alfredo $10.95
Egg Noodles in Cream Sauce and Parmesan Cheese.

Seafood Scampi $16.95
Shrimp, scallops, squid sautéed in garlic, olive oil. Served on fresh linguine.

New **Ravioli di Ricotta** $11.95
Pasta Stuffed with Ricotta and Parmesan Cheese. Served with Tomato Sauce.

Linguine Vegetariani $13.00
Sautéed in butter with mushrooms, zucchini and red pepper.

PIZZA

	10"	14"
Cheese and Tomato based on the "Pines of Rome" classic	$11.75	$13.75
Fresh Herb a blend of herbs, topped with an olive oil & chèvre cheese	$14.75	$16.95
New **Thai Chicken** chicken marinated in a spicy peanut-ginger sesame sauce	$15.75	$17.95
Garden Veggie	$13.75	$15.95

exploit

Canadian bacon, salami, pepperoni, sausage, beef, olives, pineapple, mushrooms, onions, green peppers, etc. etc.

Hawaiian BBQ fresh pineapple, BBQ chicken, & smoked gouda	$16.75	$18.95

The Explore/Exploit Tradeoff

- All species exhibit it
  - **fungi** *(Watkinson et al. 2005)*
  - **ants** *(Pratt & Sumpter 2006)*
  - **birds** *(Kacelnik et al.)*
  - **people** *(Daw et al., 2006; Wilson et al., 2014)…*

STARTERS

PASTA

Fettuccine Alfredo $10.95
Egg Noodles in Cream Sauce and Parmesan Cheese.

Seafood Scampi $16.95
Shrimp, scallops, squid sautéed in garlic, olive oil. Served on fresh linguine.

New **Ravioli di Ricotta** $11.95

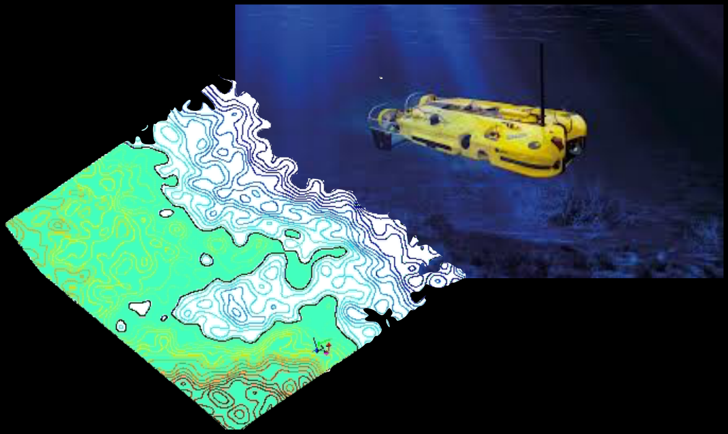
explore

PIZZA

	10"	14"
Cheese and Tomato based on the "Pines of Rome" classic	$11.75	$13.75
Fresh Herb a blend of herbs, topped with an olive oil & chèvre cheese	$14.75	$16.95
New **Thai Chicken** chicken marinated in a spicy peanut-ginger sesame sauce	$15.75	$17.95
Garden Veggie olives, onions, green peppers, & tomatoes	$13.75	$15.95
Greek Pizza mediterranean spiced grilled chicken & mozzarella	$14.75	$16.75
New **The Works** Canadian bacon, salami, pepperoni, sausage, beef, olives, pineapple, mushrooms, onions, green peppers, etc. etc.	$18.95	22.95
Hawaiian BBQ fresh pineapple, BBQ chicken, & smoked gouda	$16.75	$18.95

The Explore/Exploit Tradeoff

- **All species exhibit it**
  - **fungi** *(Watkinson et al. 2005)*
  - **ants** *(Pratt & Sumpter 2006)*
  - **birds** *(Kacelnik et al.)*
  - **people** *(Daw et al., 2006; Wilson et al., 2014)…*
  - **engineers** *(Kaelbling et al., 1996; Auer et al., 2002, Ogren et al., 2004)*



The Explore/Exploit Tradeoff

- **All species exhibit it**
  - **fungi** *(Watkinson et al. 2005)*
  - **ants** *(Pratt & Sumpter 2006)*
  - **birds** *(Kacelnik et al.)*
  - **people** *(Daw et al., 2006; Wilson et al., 2014)…*

In love:

Should I stay or should I go now?
If I go there will be trouble ← ***exploration***
And if I stay it may be double ← ***exploitation***
So come on and let me know
Should I stay or should I go?

The Clash

The Explore/Exploit Tradeoff

- **All species exhibit it**
  - **fungi** *(Watkinson et al. 2005)*
  - **ants** *(Pratt & Sumpter 2006)*
  - **birds** *(Kacelnik et al.)*
  - **people** *(Daw et al., 2006; Wilson et al., 2014)…*

and in war:

As we know, there are known knowns.
There are things we know we know.
We also know there are known unknowns.
That is to say we know there are some things we do not know.
But there are also unknown unknowns,
The ones we don't know we don't know.

Donald Rumsfeld
Department of Defense news briefing
Feb. 12, 2002
(courtesy of Peter Dayan and Angela Yu)

The Explore/Exploit Tradeoff

- **All species exhibit it**
  - **fungi** *(Watkinson et al. 2005)*
  - **ants** *(Pratt & Sumpter 2006)*
  - **birds** *(Kacelnik et al.)*
  - **people** *(Daw et al., 2006; Wilson et al., 2014)…*

and in war:

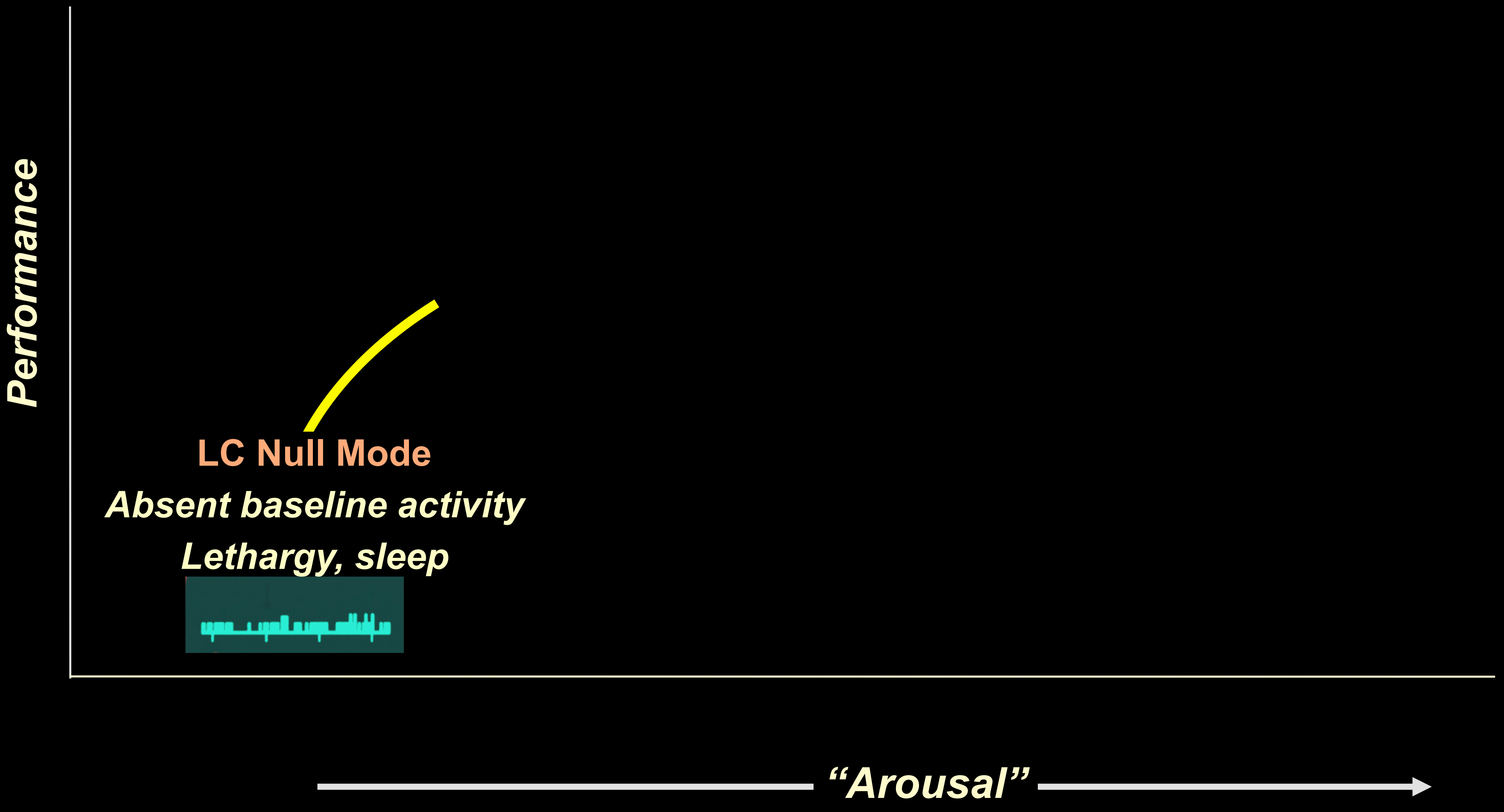
it was formulated during the war, and efforts to solve it so sapped the energies and minds of Allied analysts that the suggestion was made that the problem be dropped over Germany, as

the ultimate instrument of intellectual sabotage.

Peter Whittle, 1975

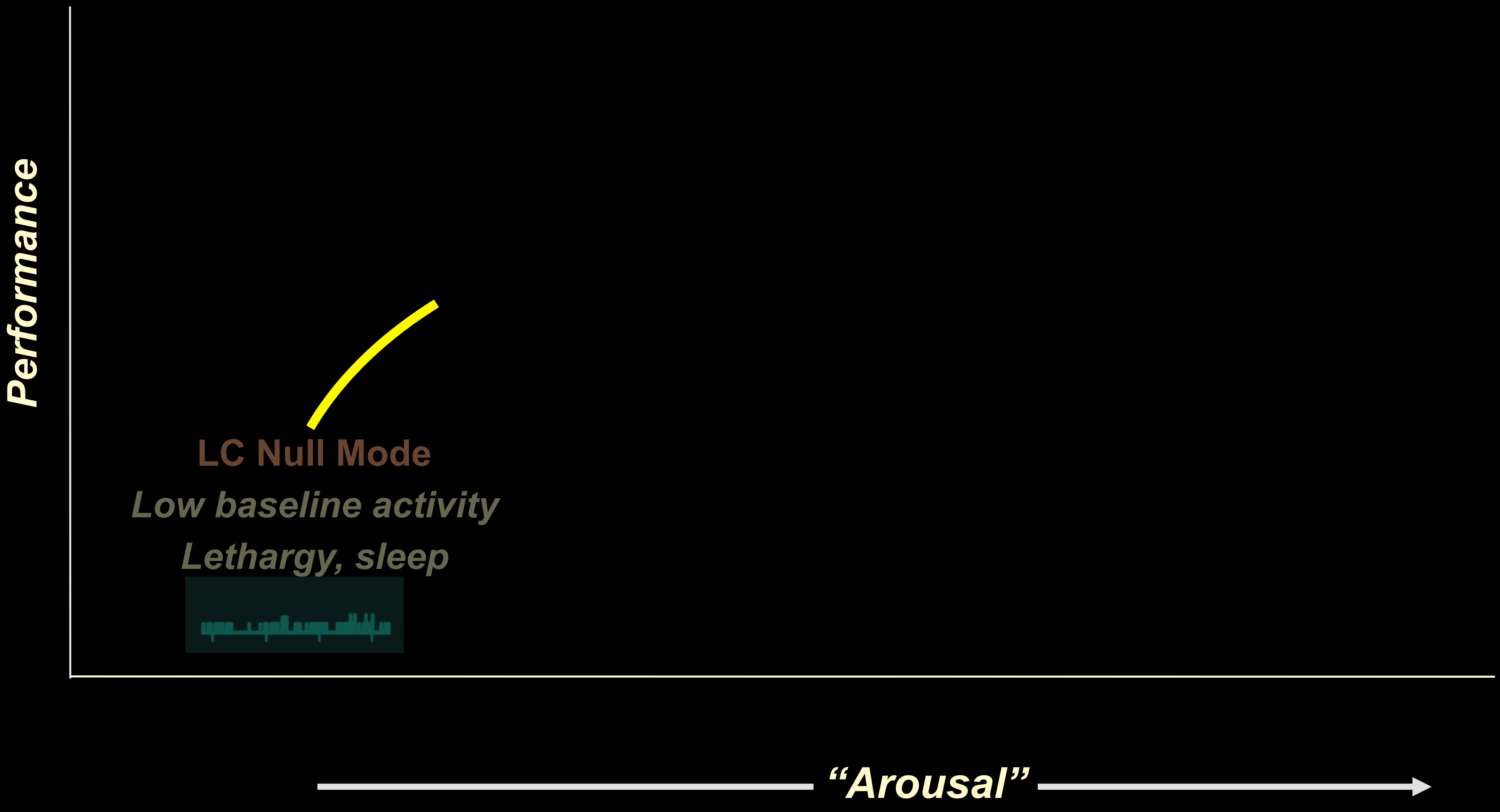
Yerkes-Dodson “Inverted U” Curve

Aston-Jones et al (J Neurosci 1994)



Yerkes-Dodson “Inverted U” Curve

Aston-Jones et al (J Neurosci 1994)



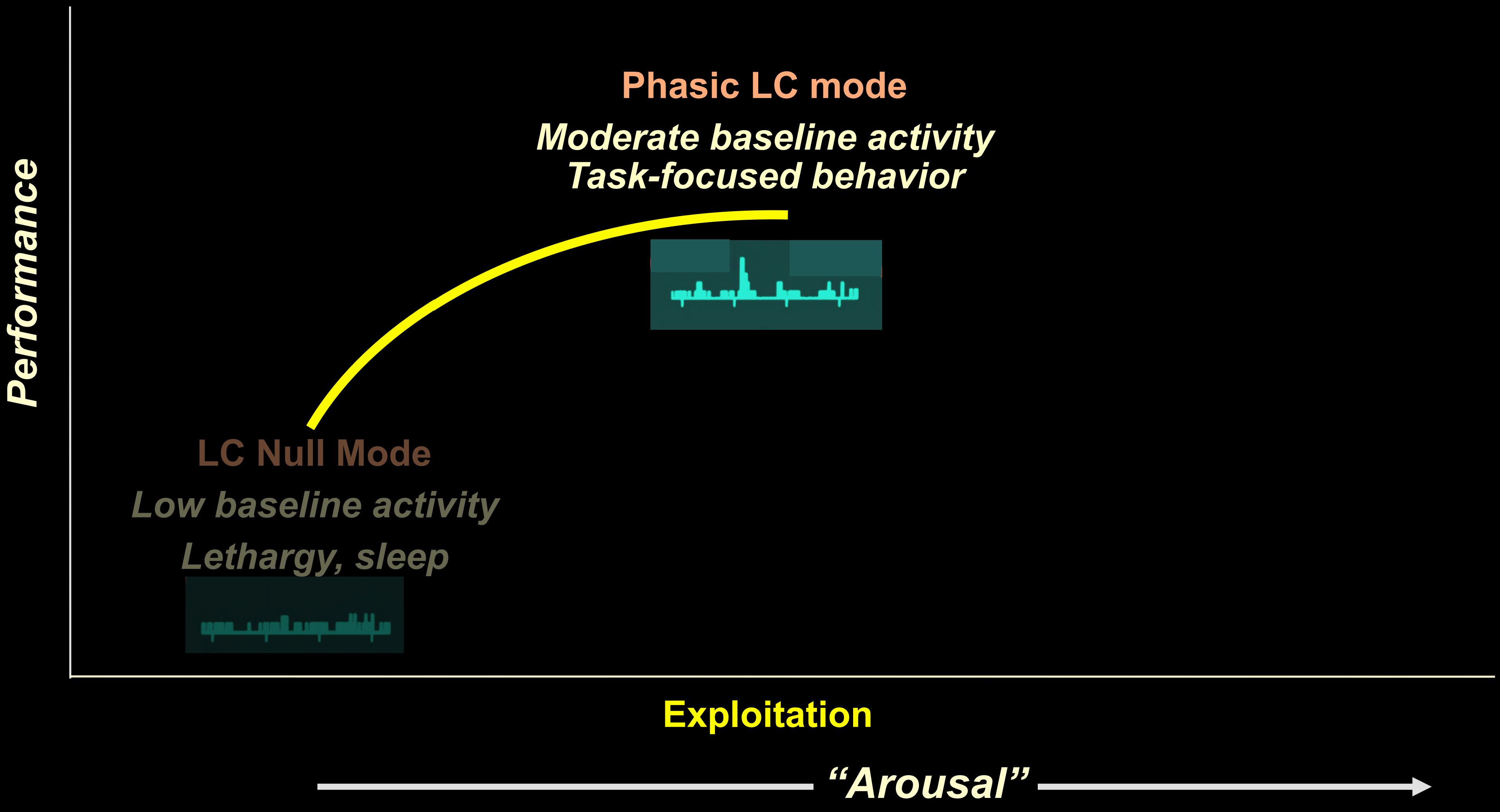
LC Null Mode

Low baseline activity

Lethargy, sleep

Yerkes-Dodson “Inverted U” Curve

Aston-Jones et al (J Neurosci 1994)



Phasic LC mode

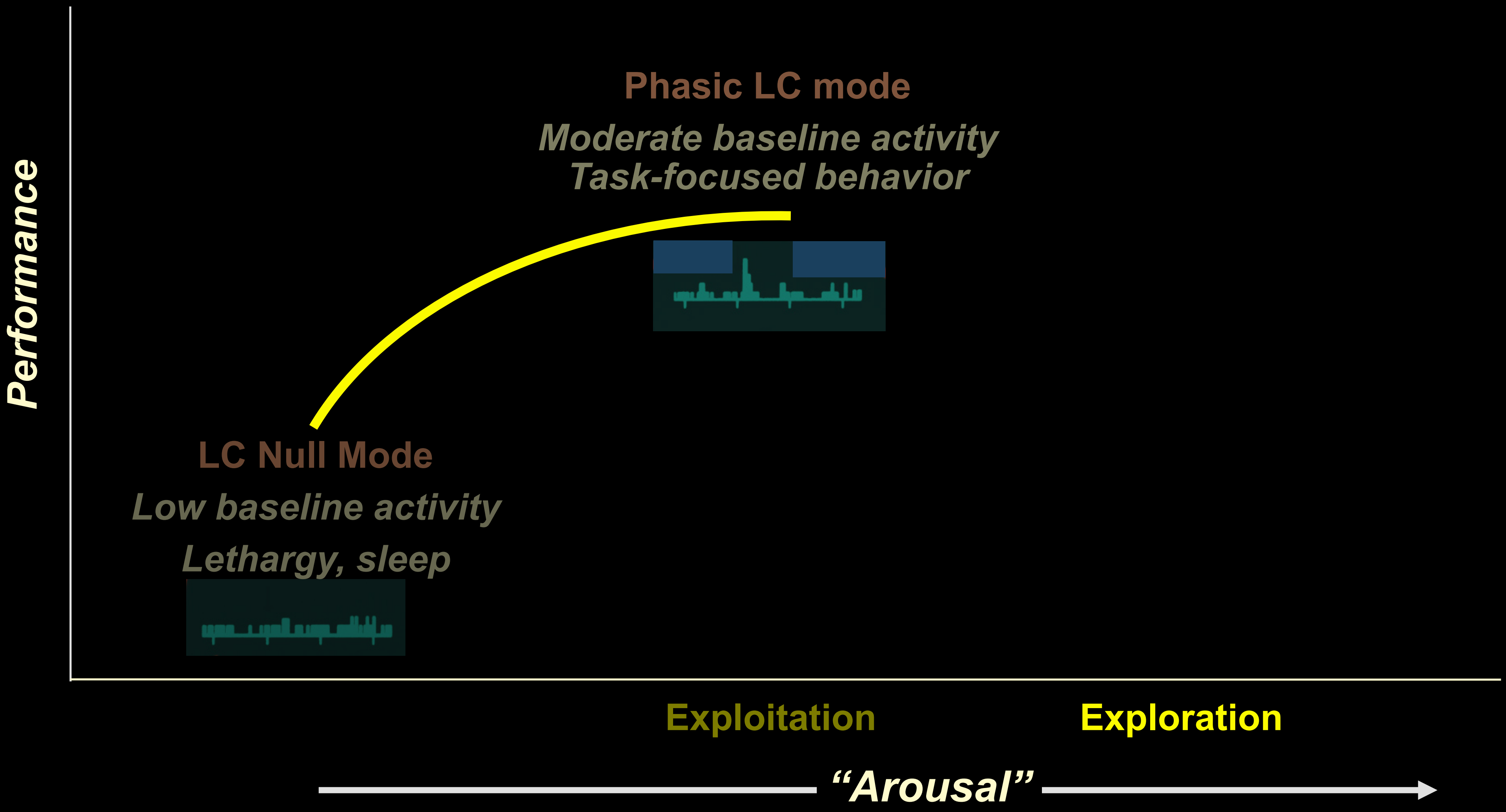
Moderate baseline activity
Task-focused behavior

LC Null Mode

Low baseline activity
Lethargy, sleep

Yerkes-Dodson “Inverted U” Curve

Aston-Jones et al (J Neurosci 1994)



Phasic LC mode

Moderate baseline activity
Task-focused behavior

LC Null Mode

Low baseline activity

Lethargy, sleep

Performance

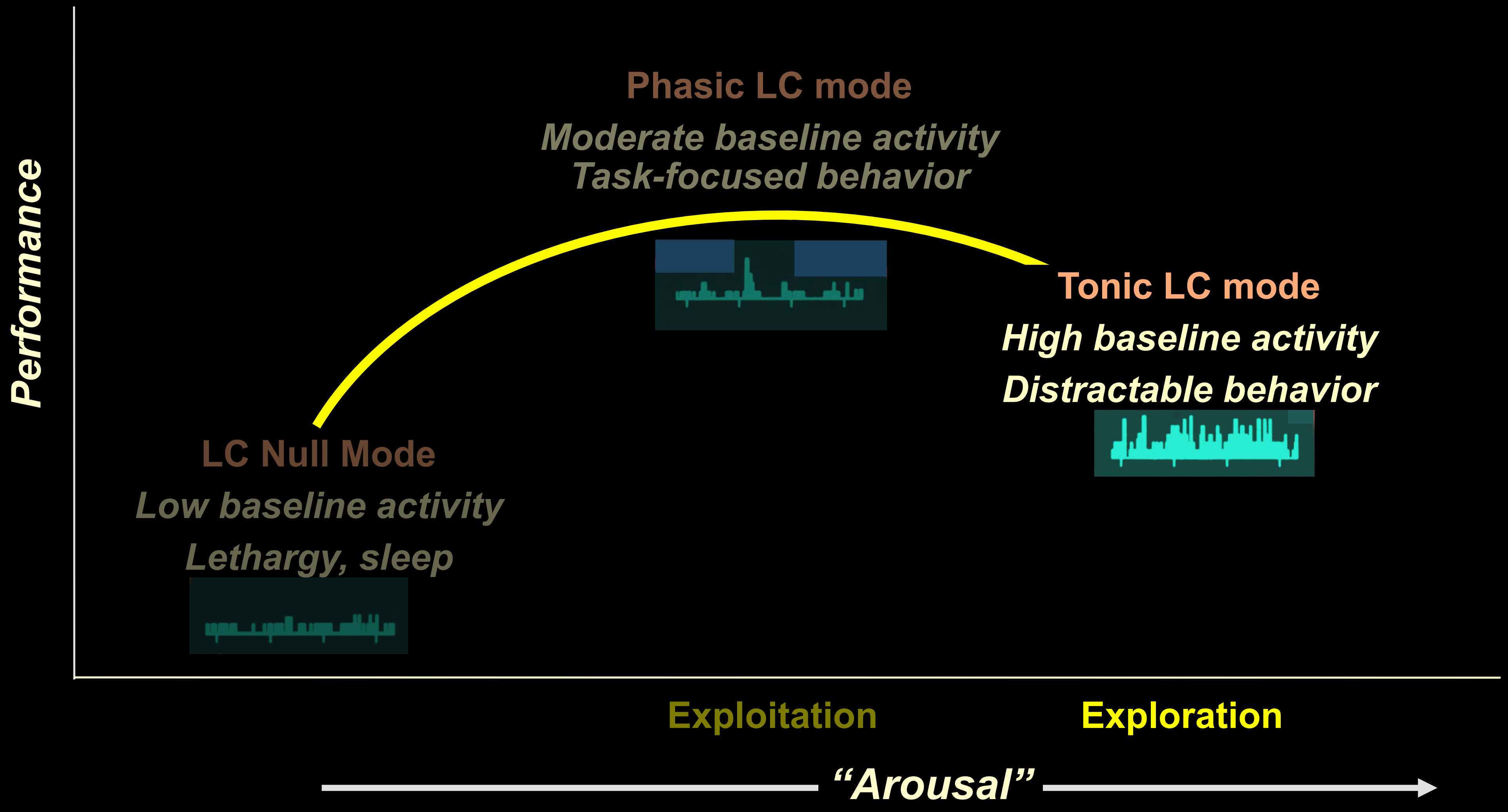
Exploitation

Exploration

“Arousal”

Yerkes-Dodson "Inverted U" Curve

Aston-Jones et al (J Neurosci 1994)



Phasic LC mode

Moderate baseline activity
Task-focused behavior

Tonic LC mode

High baseline activity
Distractable behavior

LC Null Mode

Low baseline activity
Lethargy, sleep

Performance

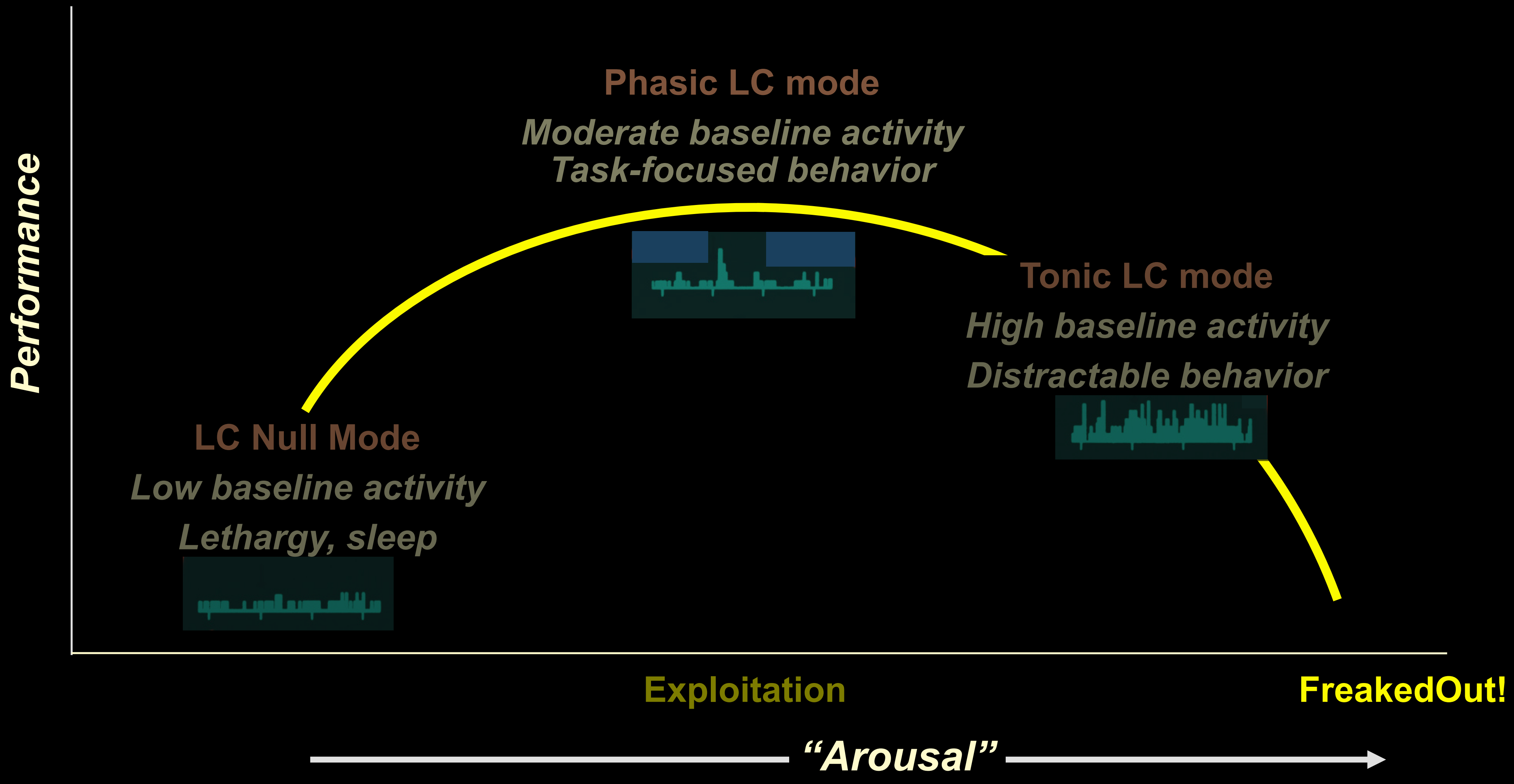
Exploitation

Exploration

"Arousal"

Yerkes-Dodson “Inverted U” Curve

Aston-Jones et al (J Neurosci 1994)



Phasic LC mode

Moderate baseline activity
Task-focused behavior

Tonic LC mode

High baseline activity
Distractable behavior

LC Null Mode

Low baseline activity
Lethargy, sleep

Summary

- LC phasic responses mediate exploration mitigate tradeoff between *complexity and efficiency*
  - Dynamic, event-related adjustment of gain optimizes performance

- LC phasic vs. tonic modes mediate tradeoff between *exploitation and exploration*
  - phasic release of NE: temporal filter — supports current task set
  - tonic NE release: indiscriminate increase in responsivity
  - modulation of processing "style:" focus vs. integration

Explore / Exploit and RL

Explore / Exploit and RL

- Challenge for reinforcement learning:
  - rewarding behaviors should be reinforced, others discouraged *(exploitation)*

Explore / Exploit and RL

- Challenge for reinforcement learning:
  - rewarding behaviors should be reinforced, others discouraged *(exploitation)*
    - reinforcement learning algorithms provably converge on optimal reward seeking behavior in a stable environment *(i.e., when contingencies don't change)*

Explore / Exploit and RL

- **Challenge for reinforcement learning:**
  - **rewarding behaviors should be reinforced, others discouraged *(exploitation)***
    - **reinforcement learning algorithms provably converge on optimal reward seeking behavior in a stable environment** *(i.e., when contingencies don't change)*
  - **however, a highly reinforced behavior will be resistant to change**

Explore / Exploit and RL

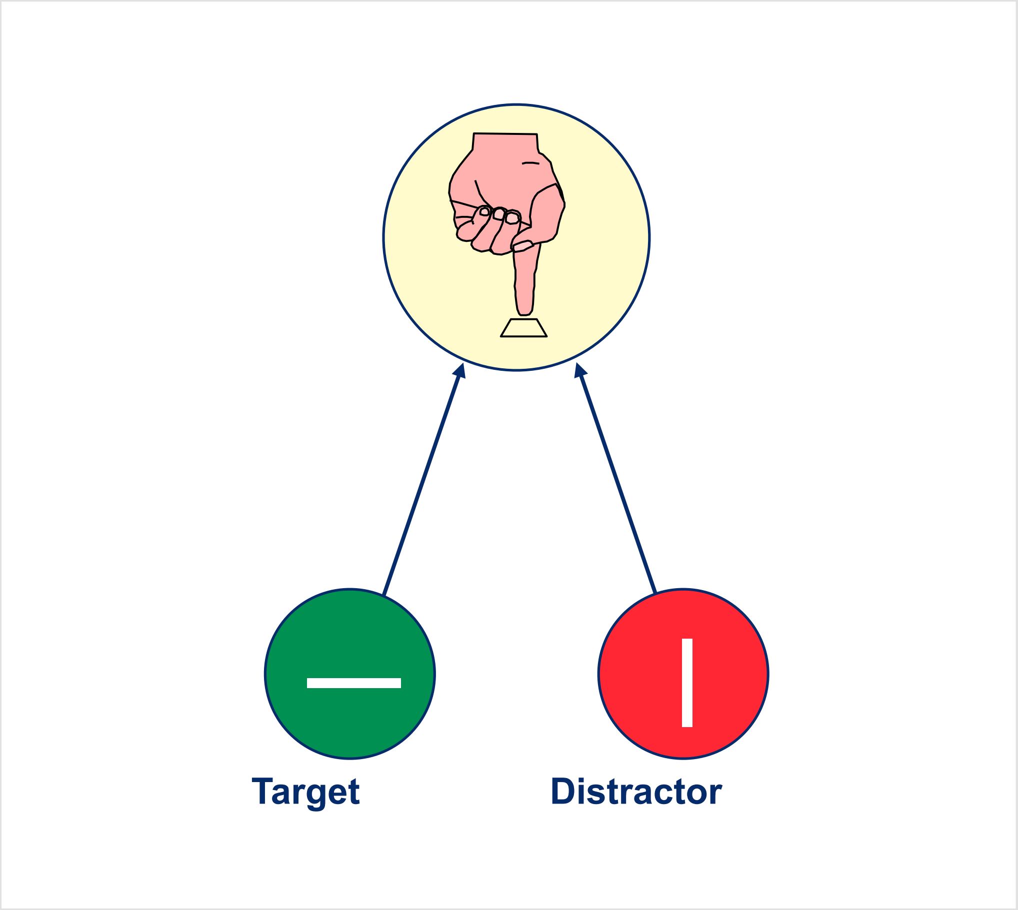
- **Challenge for reinforcement learning:**
  - rewarding behaviors should be reinforced, others discouraged *(exploitation)*
    - reinforcement learning algorithms provably converge on optimal reward seeking behavior in a stable environment *(i.e., when contingencies don't change)*
  - however, a highly reinforced behavior will be resistant to change
  - this makes it hard to adapt if/when the environment changes *(exploration)*

Explore / Exploit and RL

- Challenge for reinforcement learning:
  - rewarding behaviors should be reinforced, others discouraged *(exploitation)*
    - reinforcement learning algorithms provably converge on optimal reward seeking behavior in a stable environment *(i.e., when contingencies don't change)*
  - however, a highly reinforced behavior will be resistant to change
  - this makes it hard to adapt if/when the environment changes *(exploration)*
- Simple example…

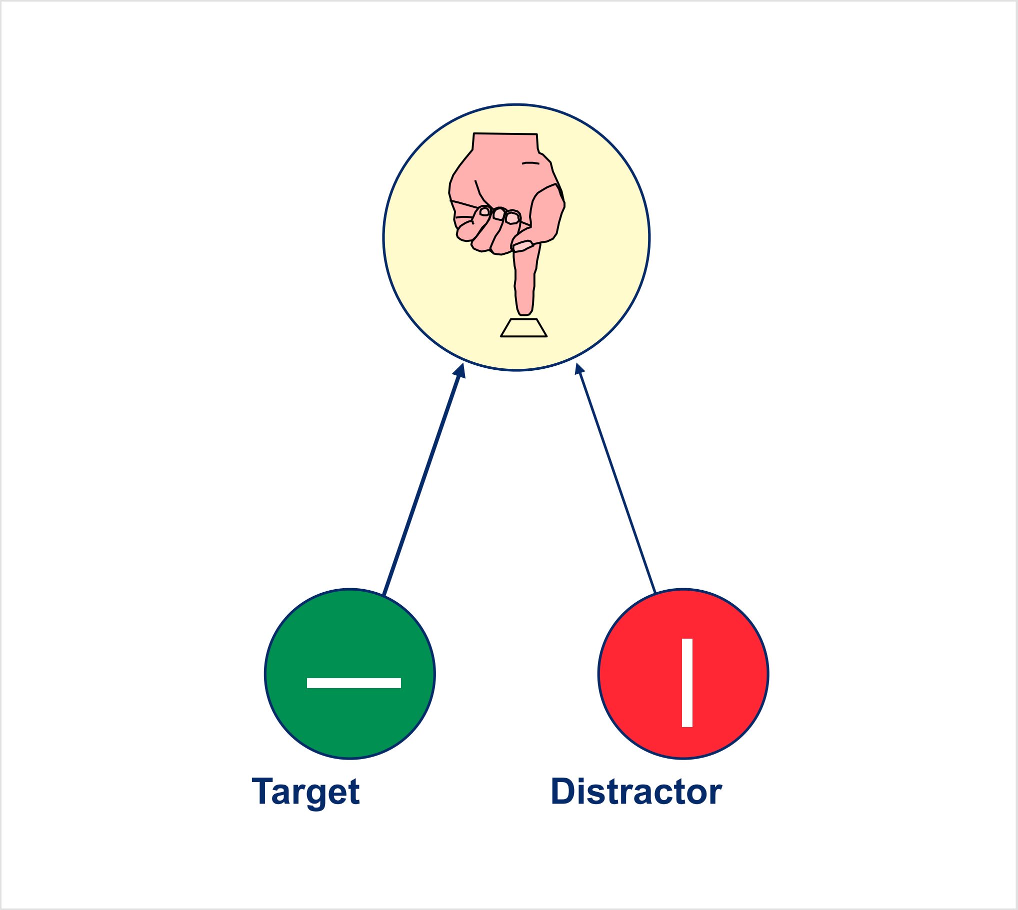
Reversal Conditioning

Before Learning



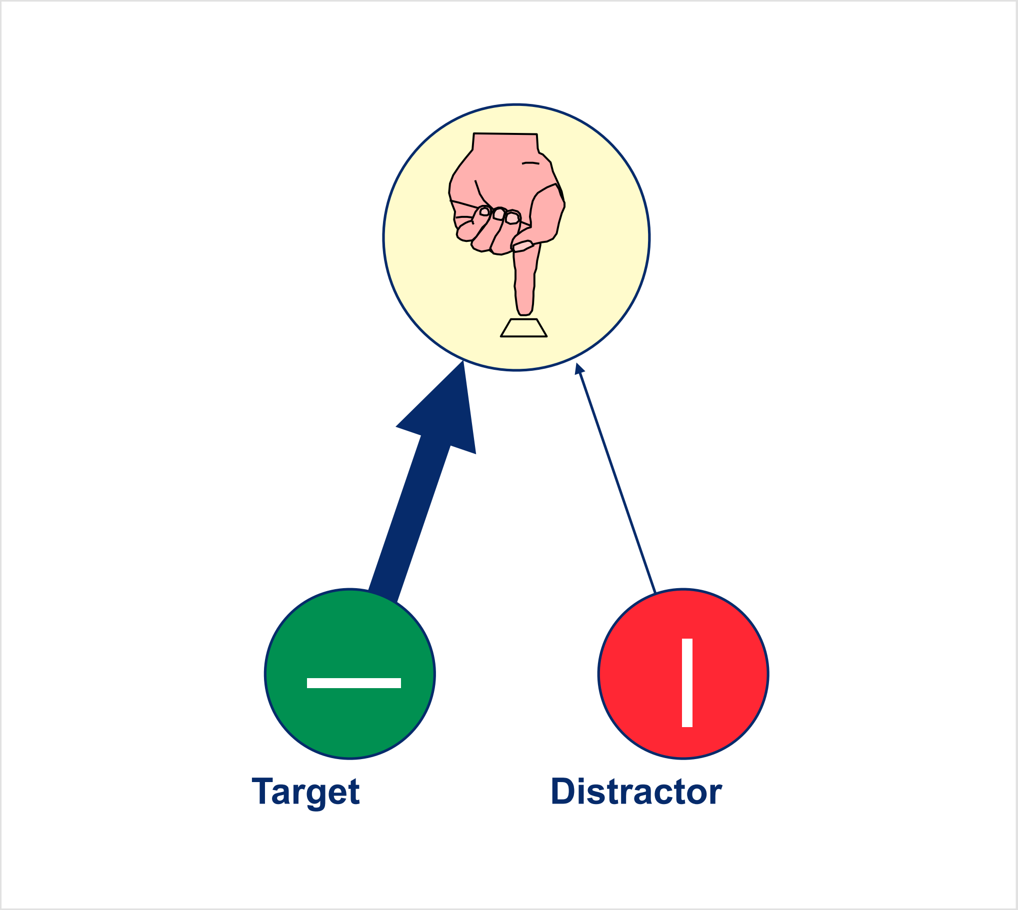
Reversal Conditioning

After Learning



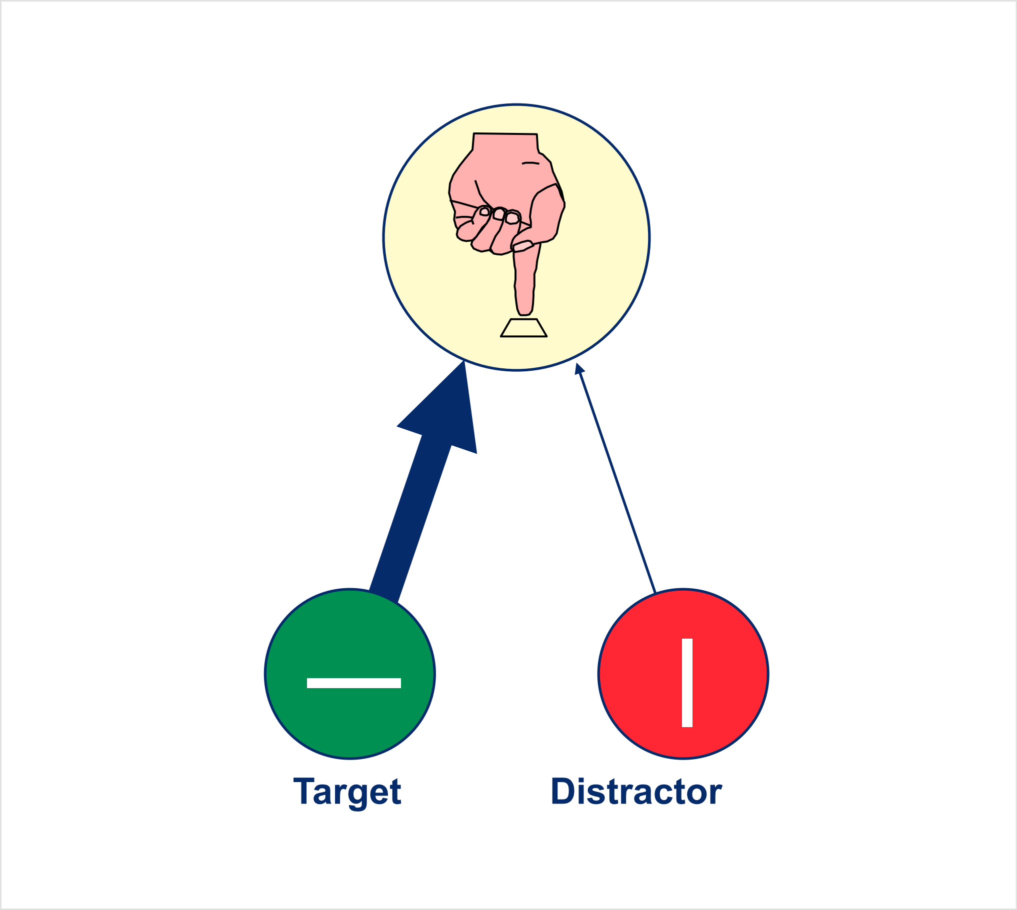
Reversal Conditioning

After Learning



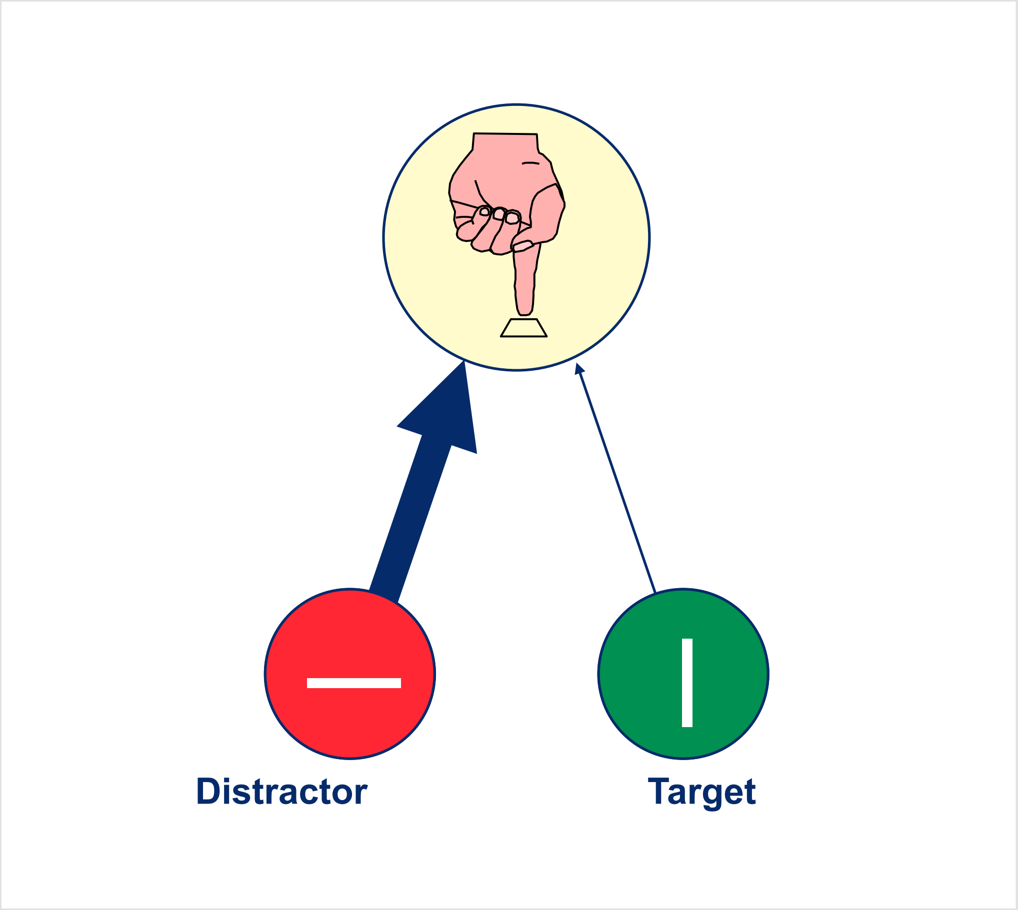
Reversal Conditioning

Reversal



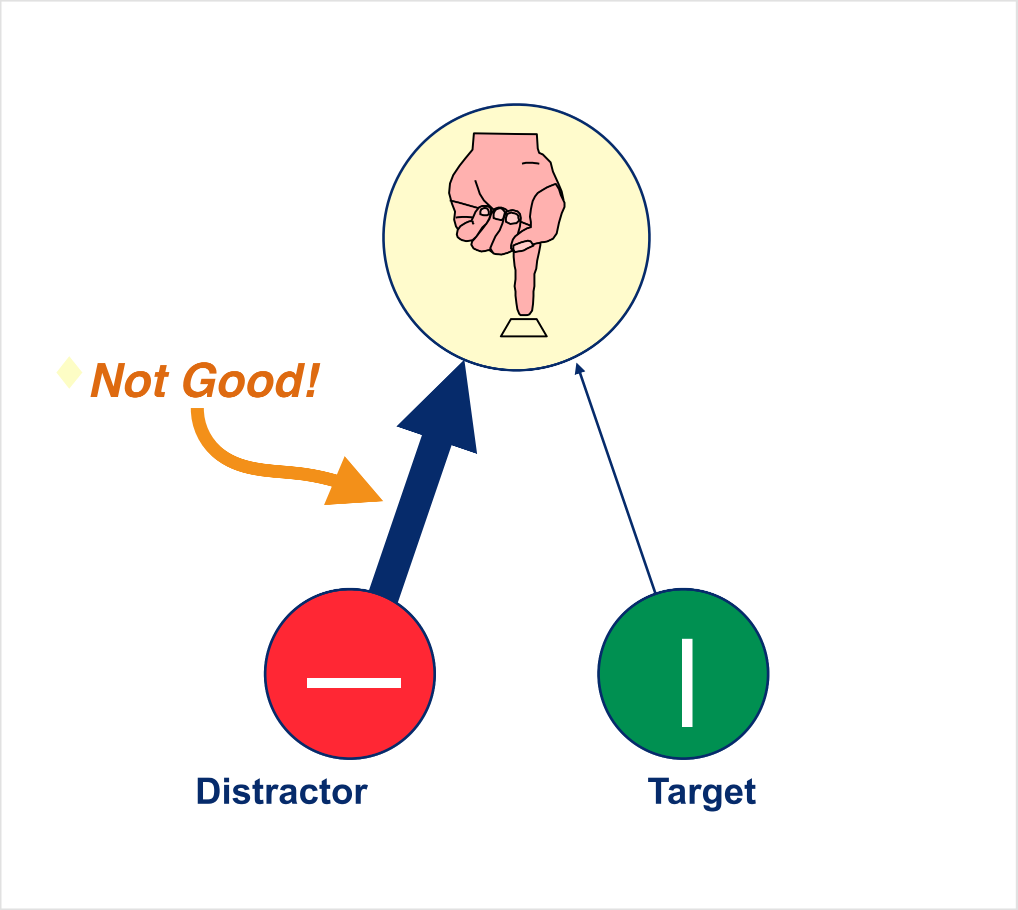
Reversal Conditioning

Reversal



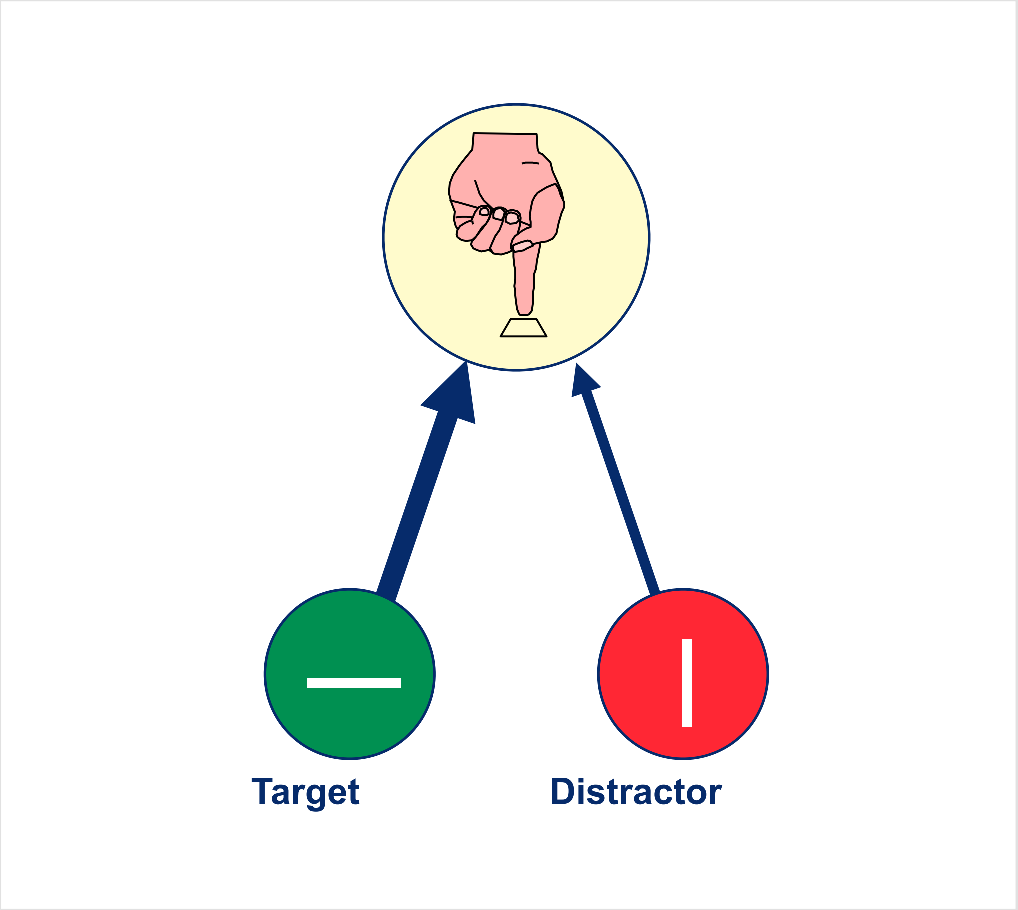
Reversal Conditioning

Reversal



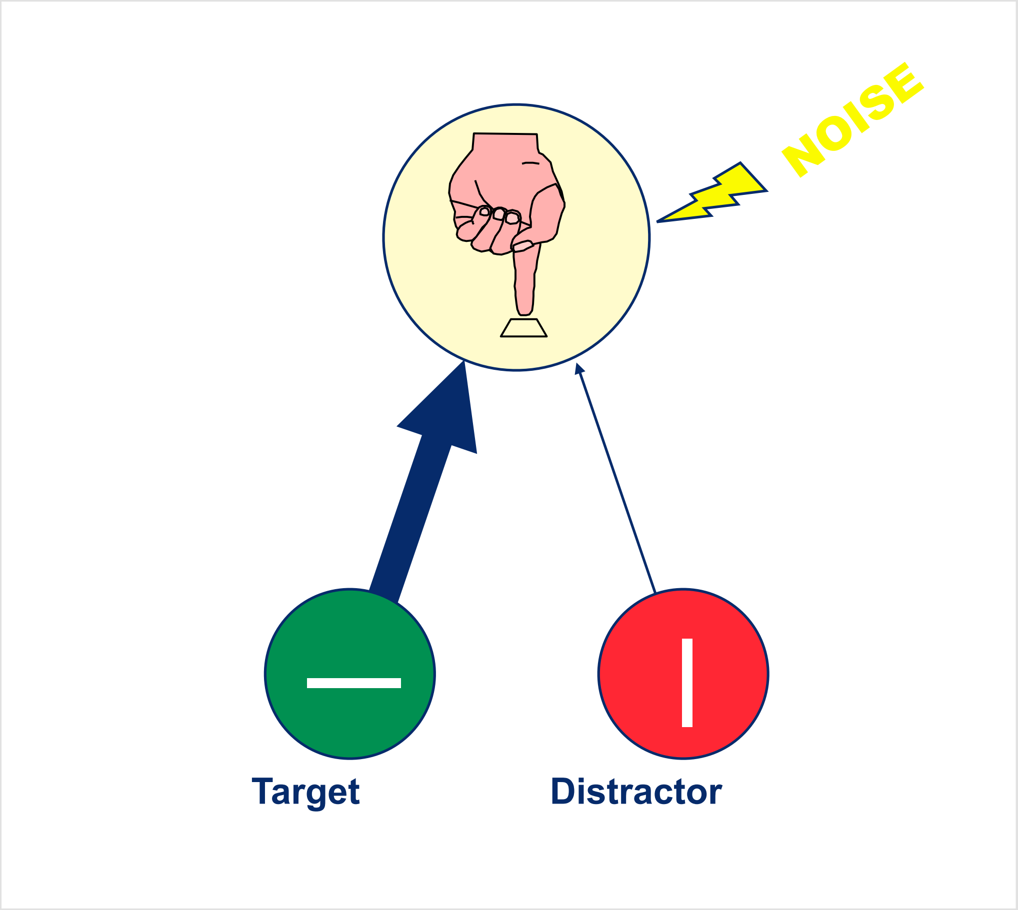
Reversal Conditioning

Solution: Weaker Learning?



Reversal Conditioning

Solution: *Adaptive Annealing*



Explore / Exploit and RL

- Challenge for reinforcement learning:
  - rewarding behaviors should be reinforced, others discouraged *(exploitation)*
    - reinforcement learning algorithms provably converge on optimal reward seeking behavior in a stable environment *(i.e., when contingencies don't change)*
  - however, a highly reinforced behavior will be resistant to change
  - this makes it hard to adapt if/when the environment changes *(exploration)*
- **Solution: LC/NE - Dopamine interactions…**

GAIN & PROCESSING STYLE

EXTRA SLIDES

ADAPTIVE GAIN HYPOTHESIS

Adaptive Gain Hypothesis

Adaptive Gain Hypothesis

- LC phasic activity
  - *transiently increases gain* in response to *task-relevant events*

Adaptive Gain Hypothesis

- LC phasic activity
  - *transiently increases gain* in response to *task-relevant events*
  - optimizes performance of *current* task

Adaptive Gain Hypothesis

- LC phasic activity
  - *transiently increases gain* in response to *task-relevant events*
  - optimizes performance of *current* task

⇨ ***exploitation***

Adaptive Gain Hypothesis

- LC phasic activity
  - *transiently increases gain* in response to *task-relevant events*
  - optimizes performance of *current* task

⇨ *exploitation*

- LC tonic activity

Adaptive Gain Hypothesis

- LC phasic activity
  - *transiently increases gain* in response to *task-relevant events*
  - optimizes performance of *current* task

⇨ *exploitation*

- LC tonic activity
  - produces *sustained, indiscriminate increases in gain* effectively increasing noise *(e.g., one vs. many amplifiers)*

Adaptive Gain Hypothesis

- **LC phasic activity**
  - ***transiently increases gain*** **in response to** ***task-relevant events***
  - **optimizes performance of** ***current*** **task**

    ⇨ ***exploitation***

- **LC tonic activity**
  - **produces** ***sustained, indiscriminate increases in gain***
    effectively increasing noise *(e.g., one vs. many amplifiers)*
  - **promotes opportunities to sample other behaviors /**
    identify other sources of reward:

Adaptive Gain Hypothesis

- LC phasic activity
  - *transiently increases gain* in response to *task-relevant events*
  - optimizes performance of *current* task

    ⇨ ***exploitation***

- LC tonic activity
  - produces ***sustained, indiscriminate increases in gain***
    effectively increasing noise *(e.g., one vs. many amplifiers)*
  - promotes opportunities to sample other behaviors /
    identify other sources of reward:

    ⇨ ***(random) exploration***

What Drives Changes in LC Mode?

What Drives Changes in LC Mode?

- Assessments of utility in frontal cortex?

What Drives Changes in LC Mode?

- Assessments of utility in frontal cortex?
    - Orbitofrontal cortex (OFC) exhibits both transient and sustained responses to positively valenced events

What Drives Changes in LC Mode?

- **Assessments of utility in frontal cortex?**
  - **Orbitofrontal cortex** (OFC) **exhibits both transient and sustained responses to positively valenced events**
  - **Anterior cingulate cortex** (ACC) **exhibits transient** (and sustained?) **responses to negatively valenced events** *(pain, fear, errors, conflict)*

What Drives Changes in LC Mode?

- **Assessments of utility in frontal cortex?**
  - **Orbitofrontal cortex** (OFC) **exhibits both transient and sustained responses to positively valenced events**
  - **Anterior cingulate cortex** (ACC) **exhibits transient** (and sustained?) **responses to negatively valenced events** *(pain, fear, errors, conflict)*
  - **Activity in ACC** (and OFC?) **is associated with trial-trial adjustments in performance** *(e.g., error and conflict-associated increases in threshold and/or attention)*

What Drives Changes in LC Mode?

- **Assessments of utility in frontal cortex?**
  - **Orbitofrontal cortex** (OFC) **exhibits both transient and sustained responses to positively valenced events**
  - **Anterior cingulate cortex** (ACC) **exhibits transient** (and sustained?) **responses to negatively valenced events** *(pain, fear, errors, conflict)*
  - **Activity in ACC** (and OFC?) **is associated with trial-trial adjustments in performance** *(e.g., error and conflict-associated increases in threshold and/or attention)*
- **Anatomical support:**

What Drives Changes in LC Mode?

- **Assessments of utility in frontal cortex?**
  - **Orbitofrontal cortex** (OFC) **exhibits both transient and sustained responses to positively valenced events**
  - **Anterior cingulate cortex** (ACC) **exhibits transient** (and sustained?) **responses to negatively valenced events** *(pain, fear, errors, conflict)*
  - **Activity in ACC** (and OFC?) **is associated with trial-trial adjustments in performance** *(e.g., error and conflict-associated increases in threshold and/or attention)*

- **Anatomical support:**
  - **OFC and ACC provide strongest cortical projections to LC**

What Drives Changes in LC Mode?

- **Assessments of utility in frontal cortex?**
  - **Orbitofrontal cortex** (OFC) **exhibits both transient and sustained responses to positively valenced events**
  - **Anterior cingulate cortex** (ACC) **exhibits transient** (and sustained?) **responses to negatively valenced events** *(pain, fear, errors, conflict)*
  - **Activity in ACC** (and OFC?) **is associated with trial-trial adjustments in performance** *(e.g., error and conflict-associated increases in threshold and/or attention)*

- **Anatomical support:**
  - **OFC and ACC provide strongest cortical projections to LC**
  - **Physiology?**

What Drives Changes in LC Mode?

What Drives Changes in LC Mode?

- **Formalization:**
  - **Short-term (transient) decreases in utility**

    (indicate need to support task performance)

What Drives Changes in LC Mode?

- **Formalization:**
  - **Short-term (transient) decreases in utility**

    (indicate need to support task performance)

    ⇒ **augment LC phasic mode (exploitation)**

What Drives Changes in LC Mode?

- **Formalization:**
  - **Short-term (transient) decreases in utility**

    (indicate need to support task performance)

    ⇒ **augment LC phasic mode (exploitation)**

  - **long-term (persistent) decreases in utility**

    (indicate potential value of changing behavior)

What Drives Changes in LC Mode?

- **Formalization:**
  - **Short-term (transient) decreases in utility**

    (indicate need to support task performance)

    ⇒ augment LC phasic mode (exploitation)

  - **long-term (persistent) decreases in utility**

    (indicate potential value of changing behavior)

    ⇒ favor LC tonic mode: (exploration)

What Drives Changes in LC Mode?

- **Formalization:**
  - **Short-term (transient) decreases in utility**

    (indicate need to support task performance)

    ⇒ **augment LC phasic mode (exploitation)**
  - **long-term (persistent) decreases in utility**

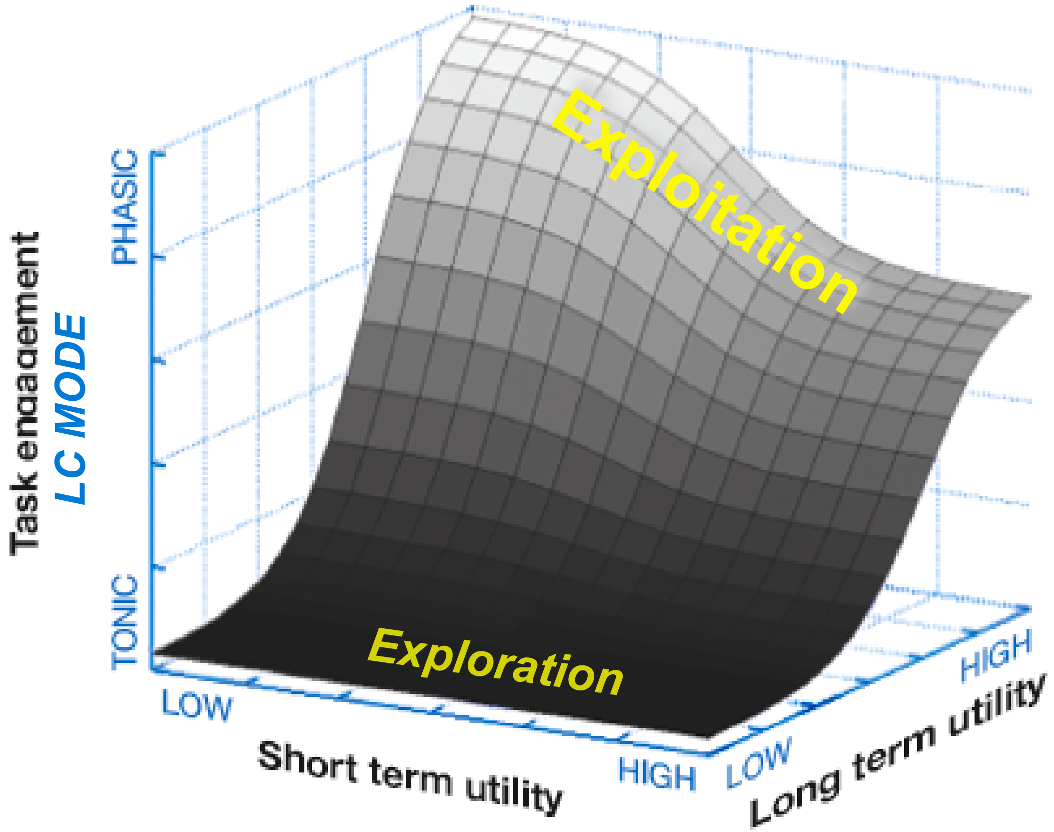
    (indicate potential value of changing behavior)

    ⇒ **favor LC tonic mode: (exploration)**
  - **Shifts mediated by changes in simple LC physiological parameters**

    (e.g., electronic coupling and/or baseline drive)

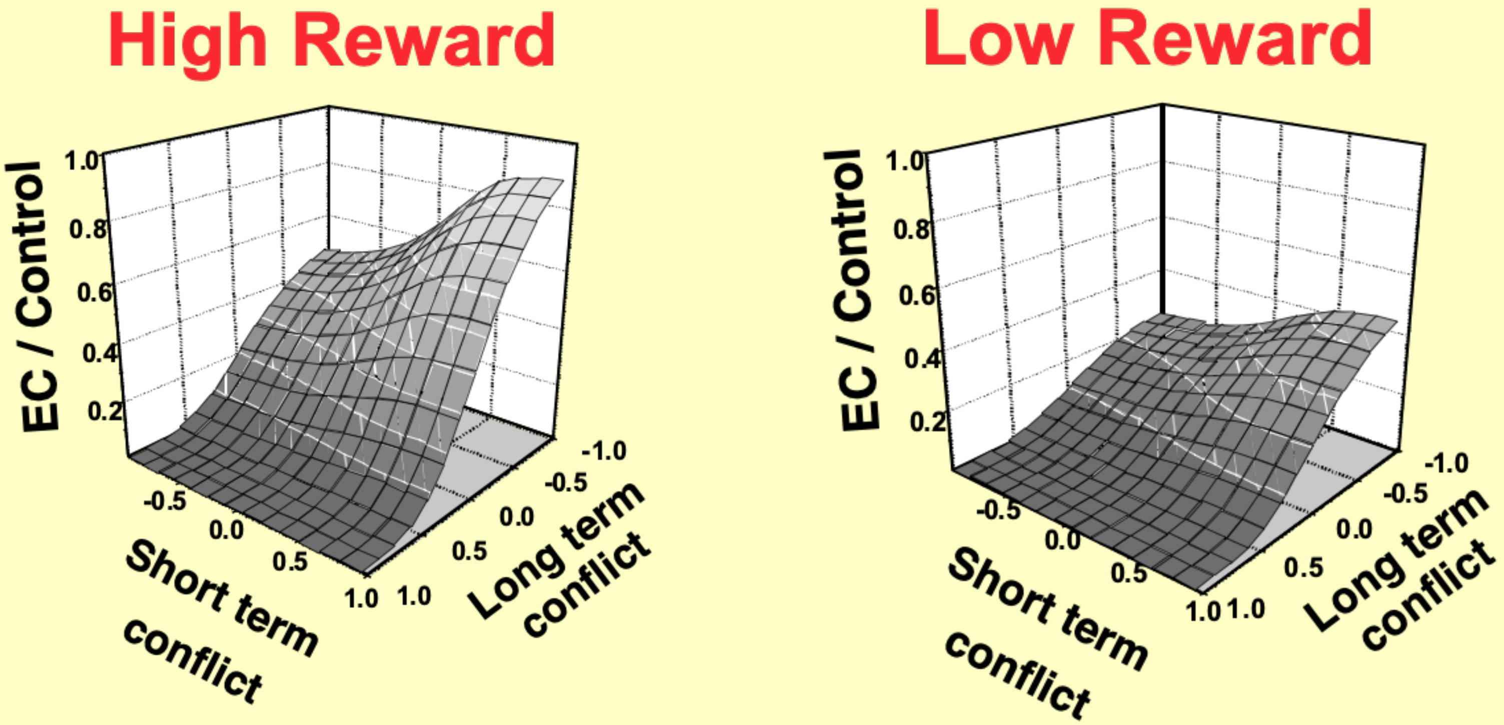
LC Mode and Utility

Aston-Jones & Cohen (Ann Rev of Neurosci, 2005)



LC coupling / drive = : [1-logistic(short term utility)] * [logistic (long term utility)

Control, Conflict and Reward


High Reward

Low Reward

$$EC = f\,(\mathrm{reward}) \times \frac{e^{-\textit{(long term conflict)}}}{1+e^{-\textit{short term conflict)}}}$$

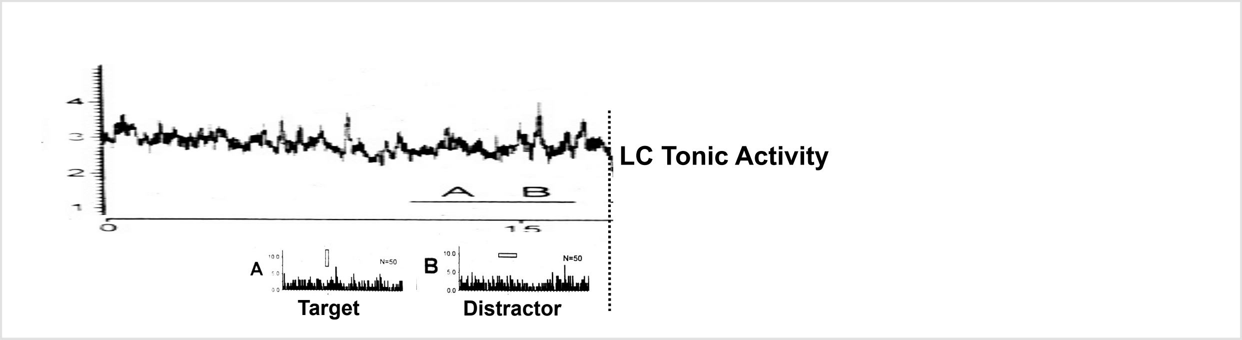
DA-NE &
REVERSAL CONDITIONING

Adaptive Gain Hypothesis & Exploration vs. Exploitation

- Theory:
  - Integrative utility function (OFC/ACC) + Adaptive gain control (LC-NE)

    = Auto-regulation of exploitation vs. exploration (DA)

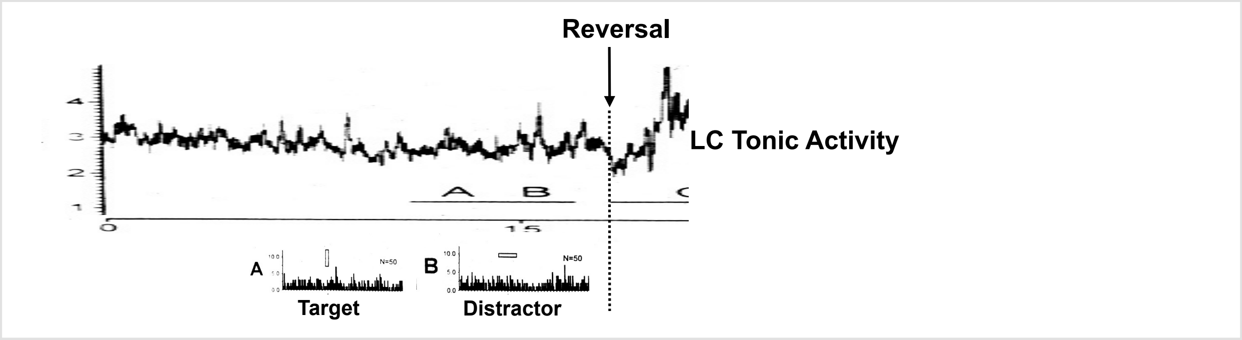
- **Reversal conditioning** (Aston-Jones et al, J. Neurosci. 1997)
  - Acquisition of initial association ➨ increased utility ➨ LC phasic mode



Adaptive Gain Hypothesis & Exploration vs. Exploitation

- Theory:
  - Integrative utility function (OFC/ACC) + Adaptive gain control (LC-NE)

    = Auto-regulation of exploitation vs. exploration (DA)
- **Reversal conditioning** (Aston-Jones et al, J. Neurosci. 1997)
  - Acquisition of initial association ➨ increased utility ➨ LC phasic mode
  - Contingency reversal ➨ reduced utility ➨ LC tonic mode

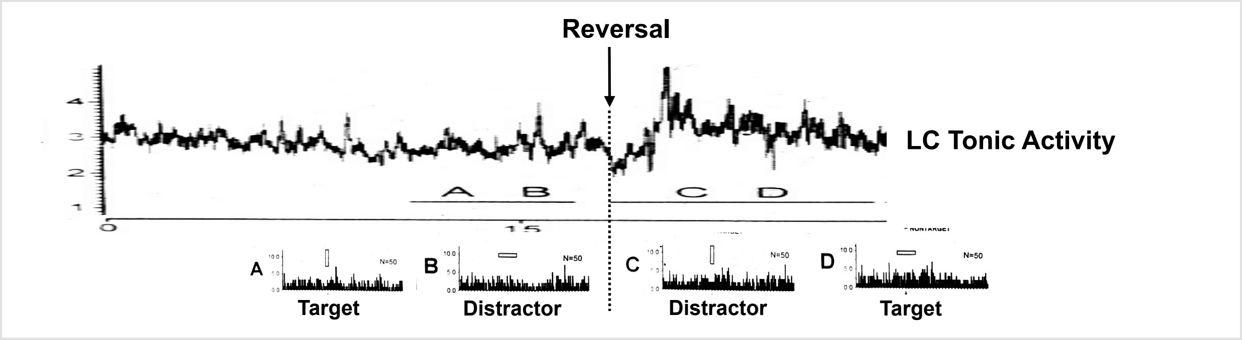


Adaptive Gain Hypothesis & Exploration vs. Exploitation

- Theory:
  - Integrative utility function (OFC/ACC) + Adaptive gain control (LC-NE)

    = Auto-regulation of exploitation vs. exploration (DA)

- **Reversal conditioning** (Aston-Jones et al, J. Neurosci. 1997)
  - Acquisition of initial association ➨ increased utility ➨ LC phasic mode
  - Contingency reversal ➨ reduced utility ➨ LC tonic mode
  - Acquisition of new association ➨ increased utility ➨ LC phasic mode



Adaptive Gain Hypothesis & Exploration vs. Exploitation

Adaptive Gain Hypothesis & Exploration vs. Exploitation

- **Theory:**
  - **Integrative utility function (OFC/ACC) + Adaptive gain control (LC-NE)**

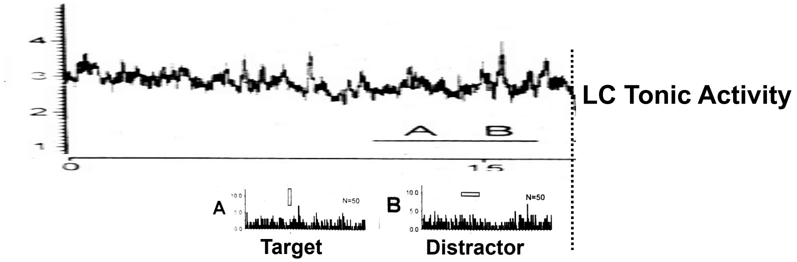
    = Auto-regulation of exploitation vs. exploration (DA)

Adaptive Gain Hypothesis & Exploration vs. Exploitation

- Theory:
  - Integrative utility function (OFC/ACC) + Adaptive gain control (LC-NE)

    = Auto-regulation of exploitation vs. exploration (DA)

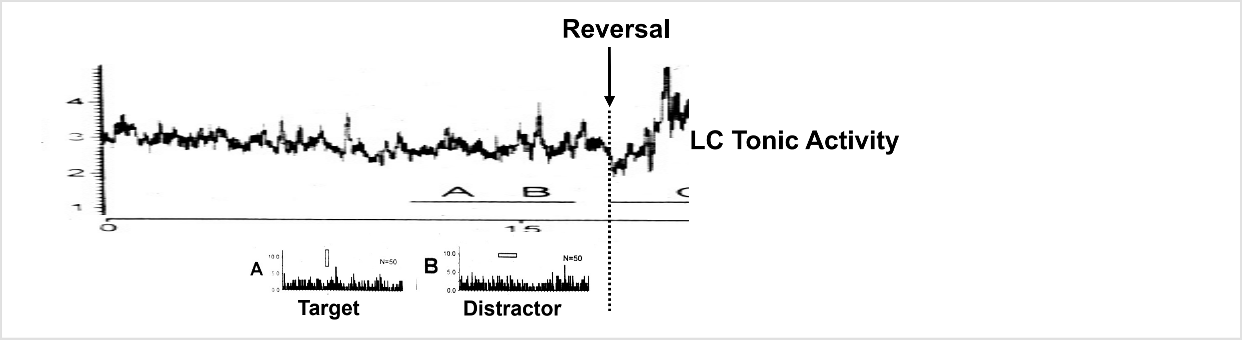
- **Reversal conditioning** (Aston-Jones et al, J. Neurosci. 1997)
  - **Acquisition of initial association ➨ increased utility ➨ LC phasic mode**



Adaptive Gain Hypothesis & Exploration vs. Exploitation

- Theory:
  - Integrative utility function (OFC/ACC) + Adaptive gain control (LC-NE)

    = Auto-regulation of exploitation vs. exploration (DA)
- **Reversal conditioning** (Aston-Jones et al, J. Neurosci. 1997)
  - Acquisition of initial association ➨ increased utility ➨ LC phasic mode
  - Contingency reversal ➨ reduced utility ➨ LC tonic mode

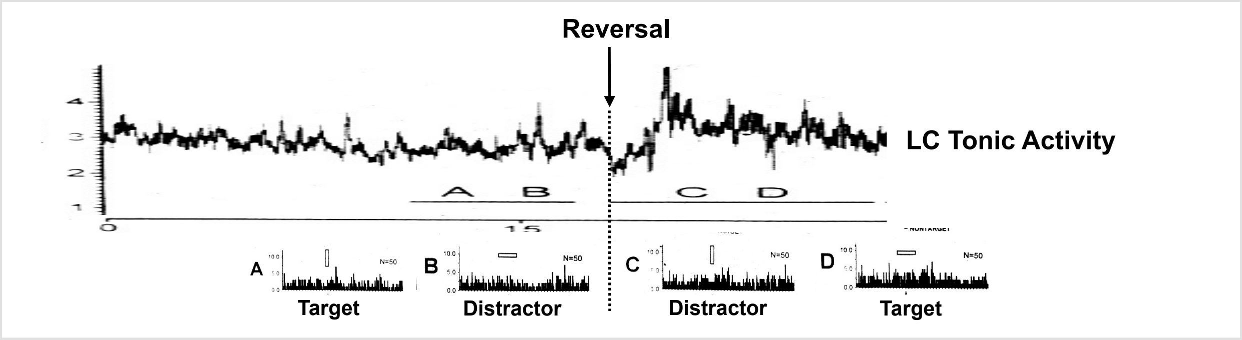


Adaptive Gain Hypothesis & Exploration vs. Exploitation

- Theory:
  - Integrative utility function (OFC/ACC) + Adaptive gain control (LC-NE)

    = Auto-regulation of exploitation vs. exploration (DA)

- **Reversal conditioning** (Aston-Jones et al, J. Neurosci. 1997)
  - Acquisition of initial association ➨ increased utility ➨ LC phasic mode
  - Contingency reversal ➨ reduced utility ➨ LC tonic mode
  - Acquisition of new association ➨ increased utility ➨ LC phasic mode



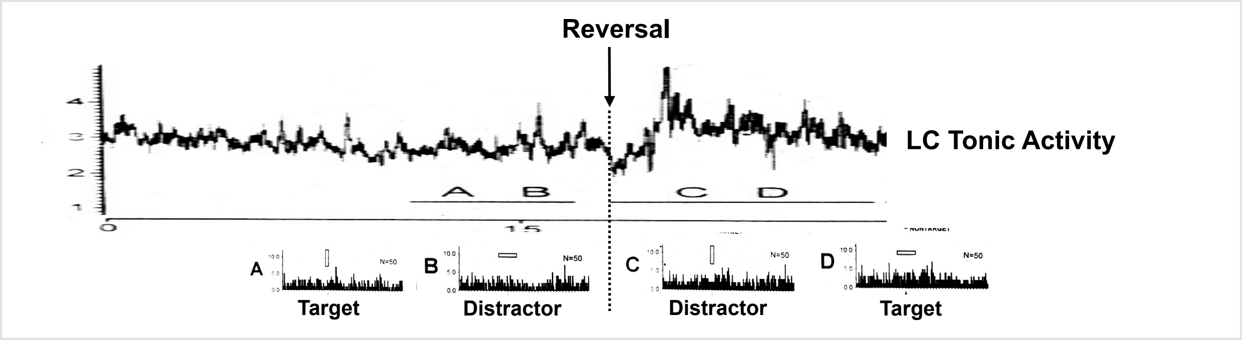
Adaptive Gain Hypothesis & Exploration vs. Exploitation

- Theory:
  - Integrative utility function (OFC/ACC) + Adaptive gain control (LC-NE)

    = Auto-regulation of exploitation vs. exploration (DA)

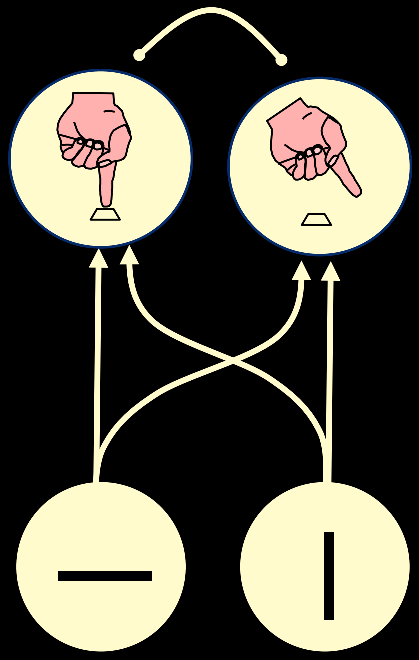
- **Reversal conditioning** (Aston-Jones et al, J. Neurosci. 1997)
  - Acquisition of initial association ➨ increased utility ➨ LC phasic mode
  - Contingency reversal ➨ reduced utility ➨ LC tonic mode
  - Acquisition of new association ➨ increased utility ➨ LC phasic mode

⇒ LC-NE system should augment performance in reversal conditioning



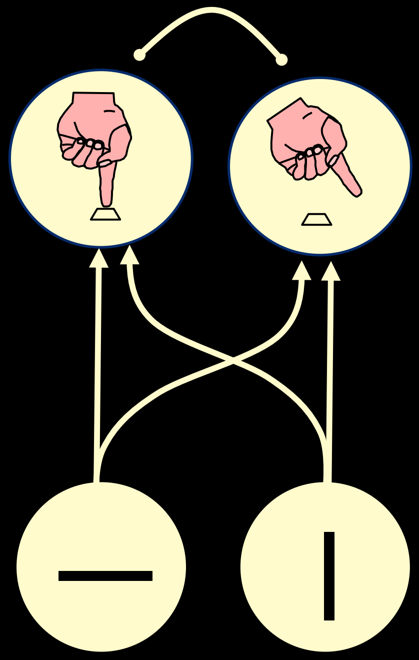
Model of DA-NE Interxns in Reversal Conditioning

McClure et al (NIPS, 2005)



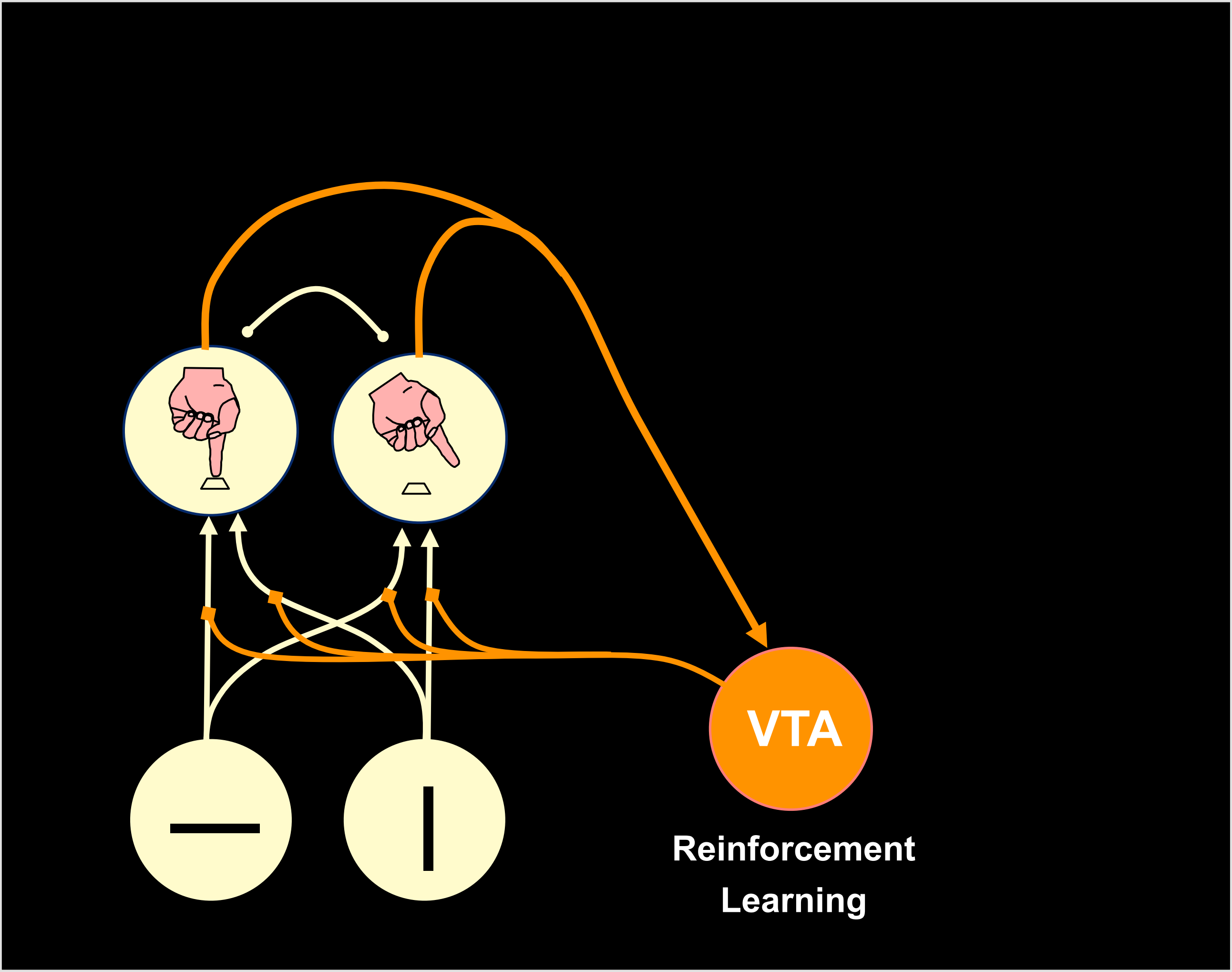
Model of DA-NE Interxns in Reversal Conditioning

McClure et al (NIPS, 2005)



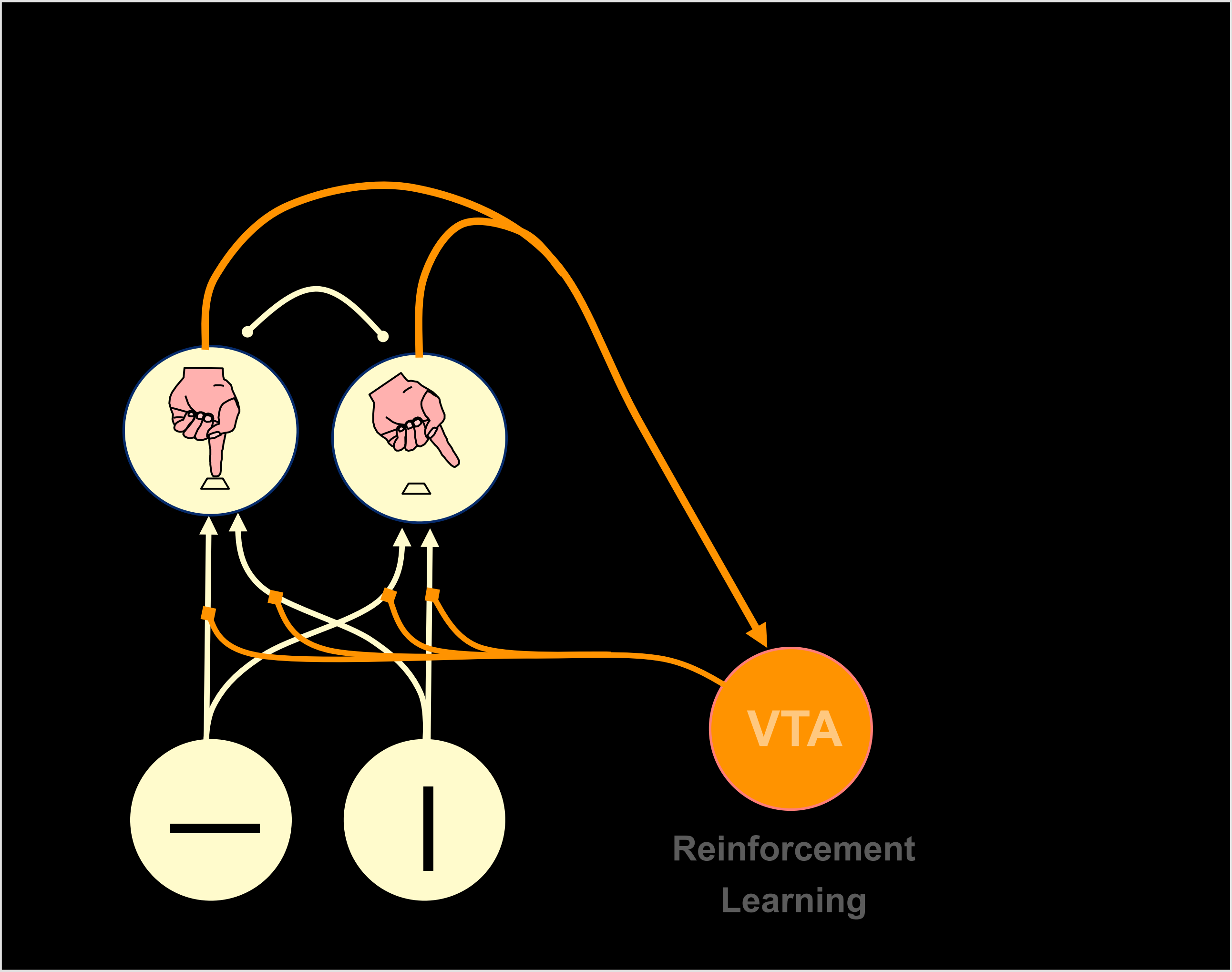
Model of DA-NE Interxns in Reversal Conditioning

McClure et al (NIPS, 2005)



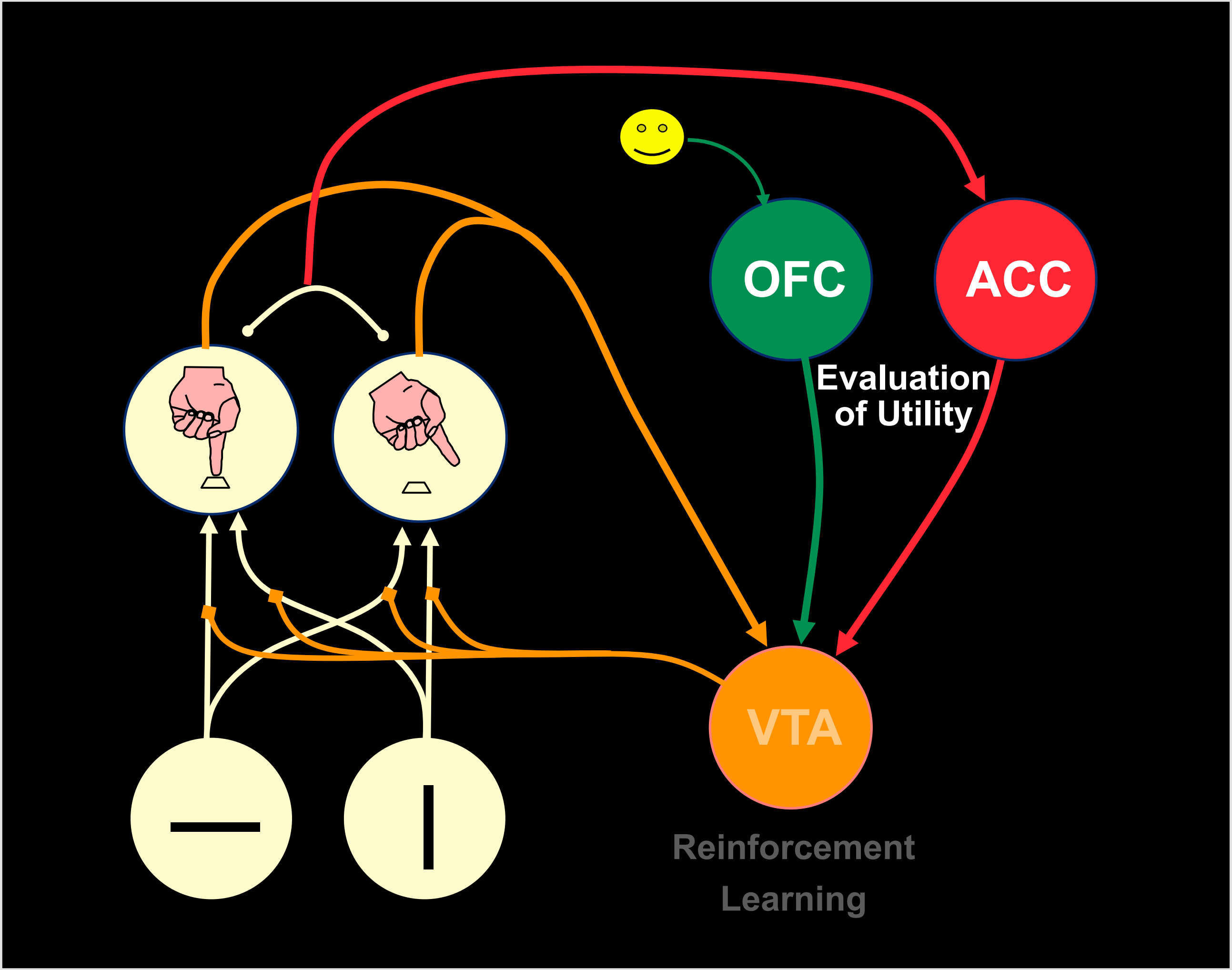
Model of DA-NE Interxns in Reversal Conditioning

McClure et al (NIPS, 2005)



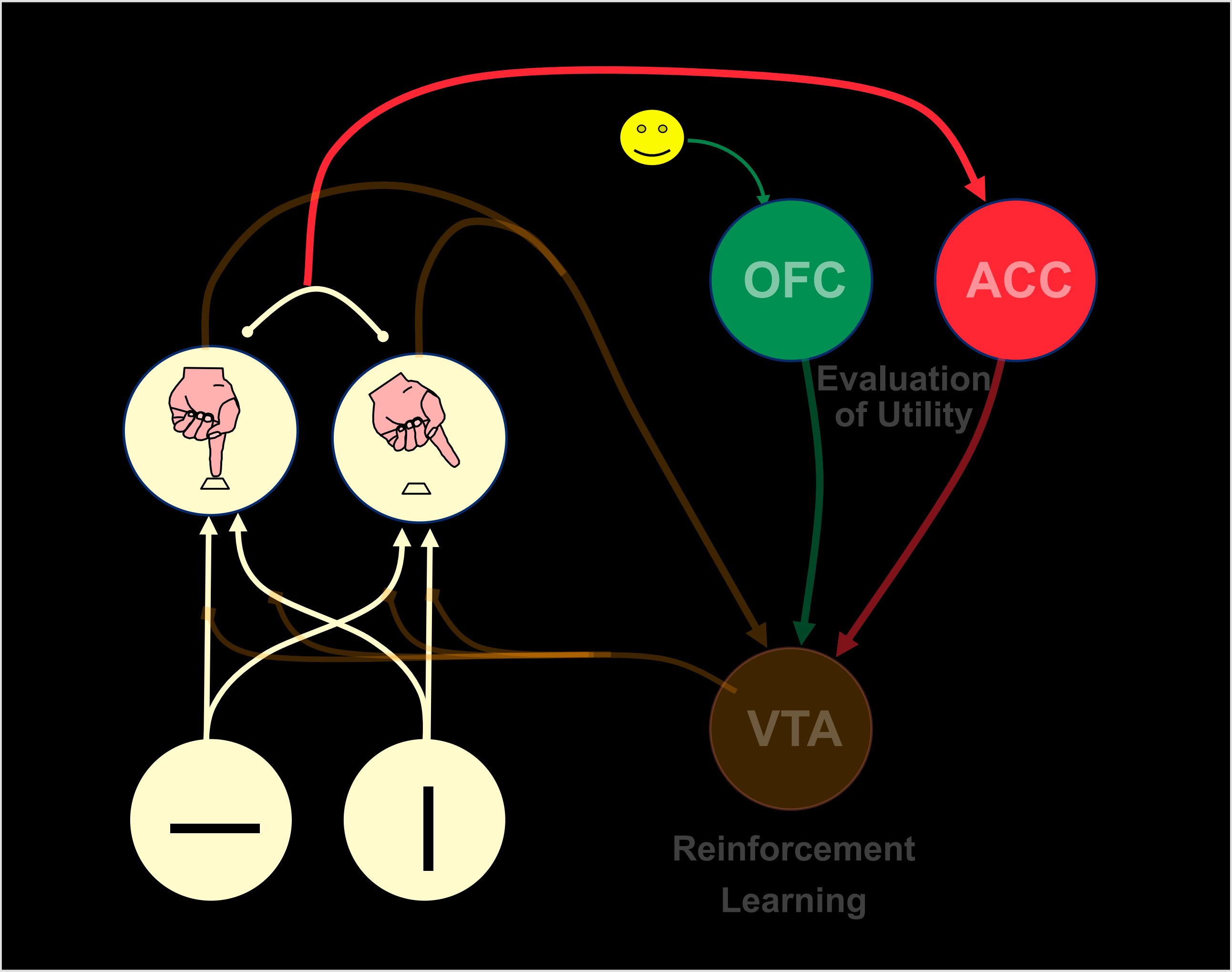
Model of DA-NE Interxns in Reversal Conditioning

McClure et al (NIPS, 2005)



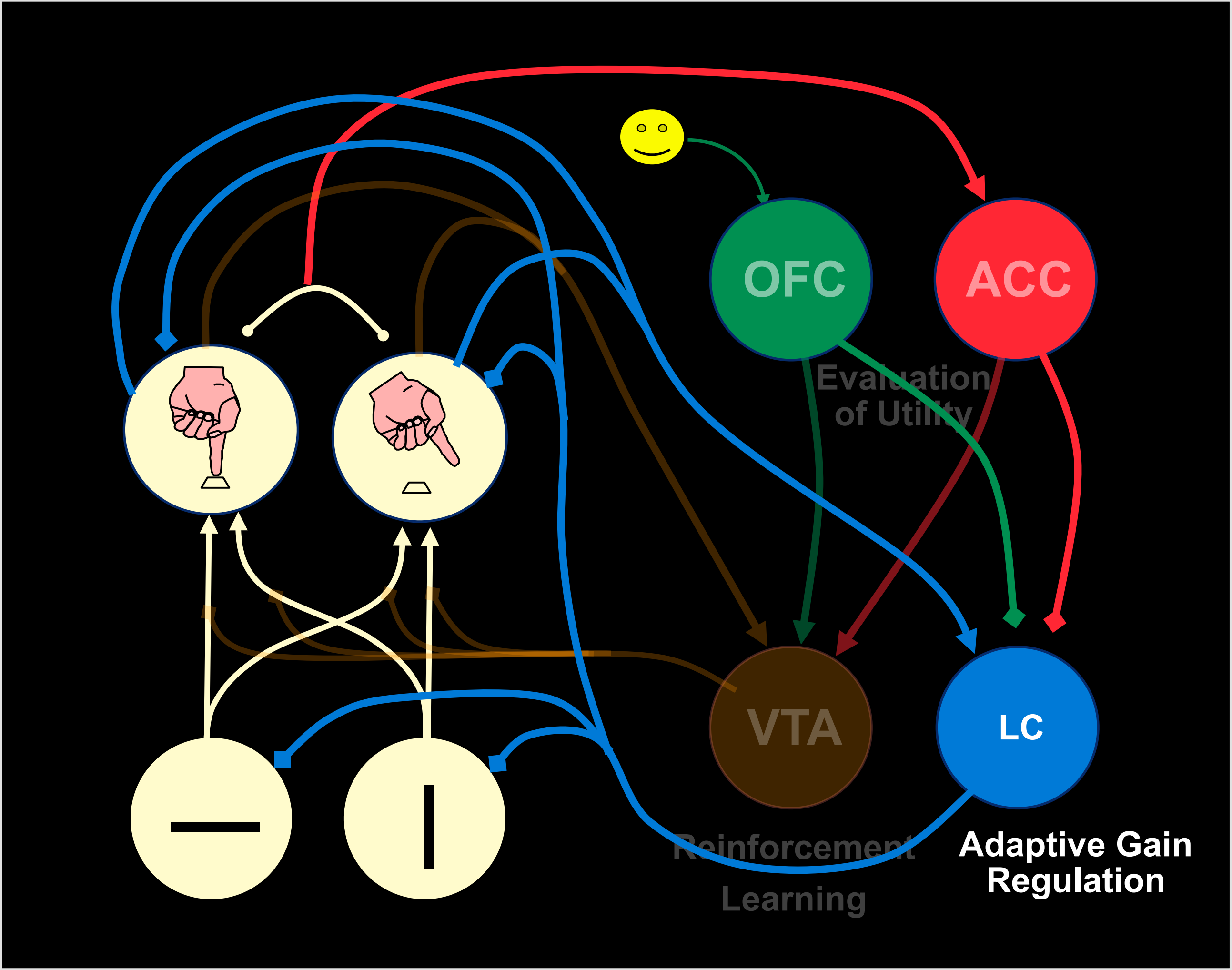
Model of DA-NE Interxns in Reversal Conditioning

McClure et al (NIPS, 2005)



Model of DA-NE Interxns in Reversal Conditioning

McClure et al (NIPS, 2005)



Reversal Conditioning Performance

McClure et al (NIPS, 2005)