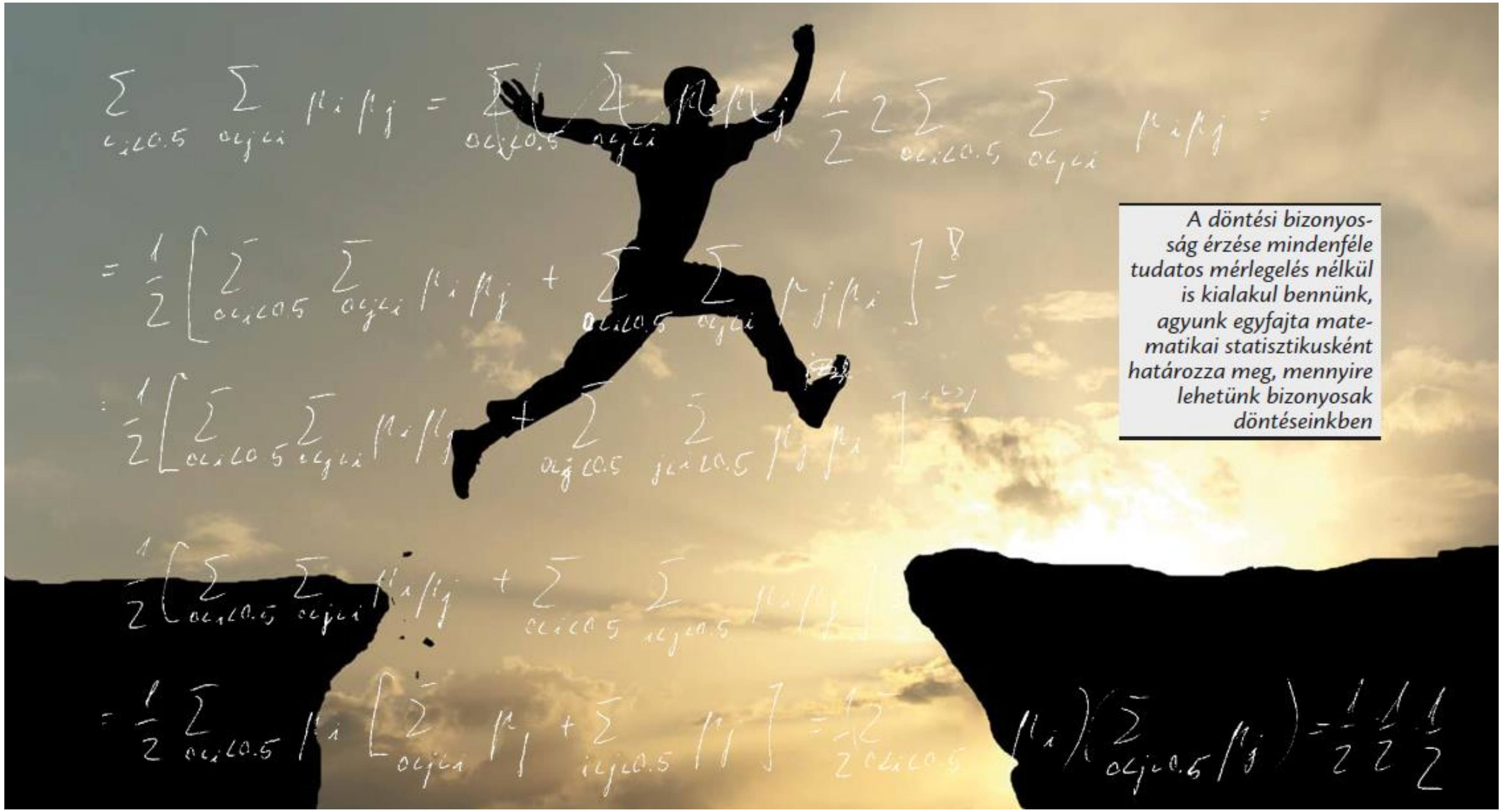
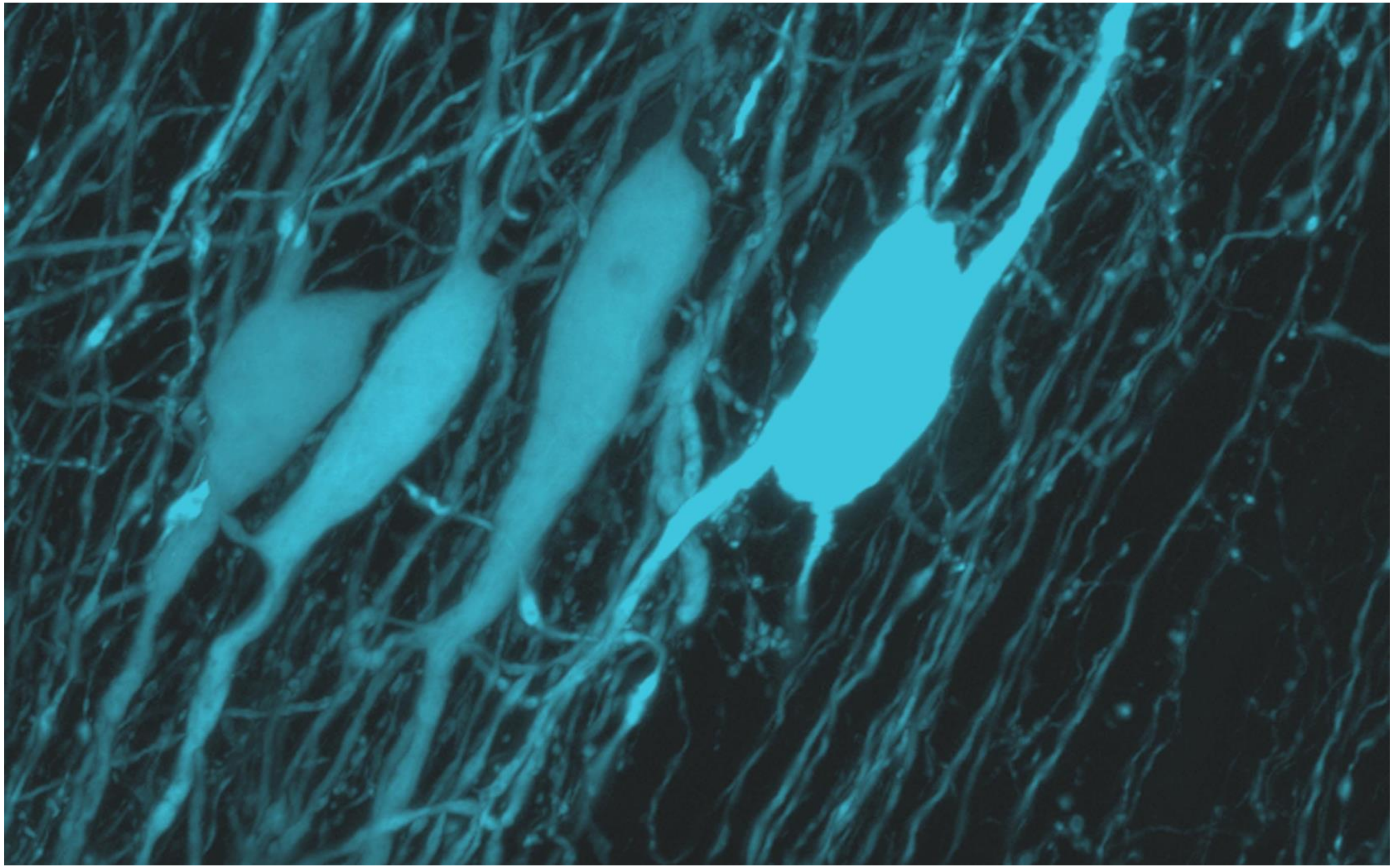
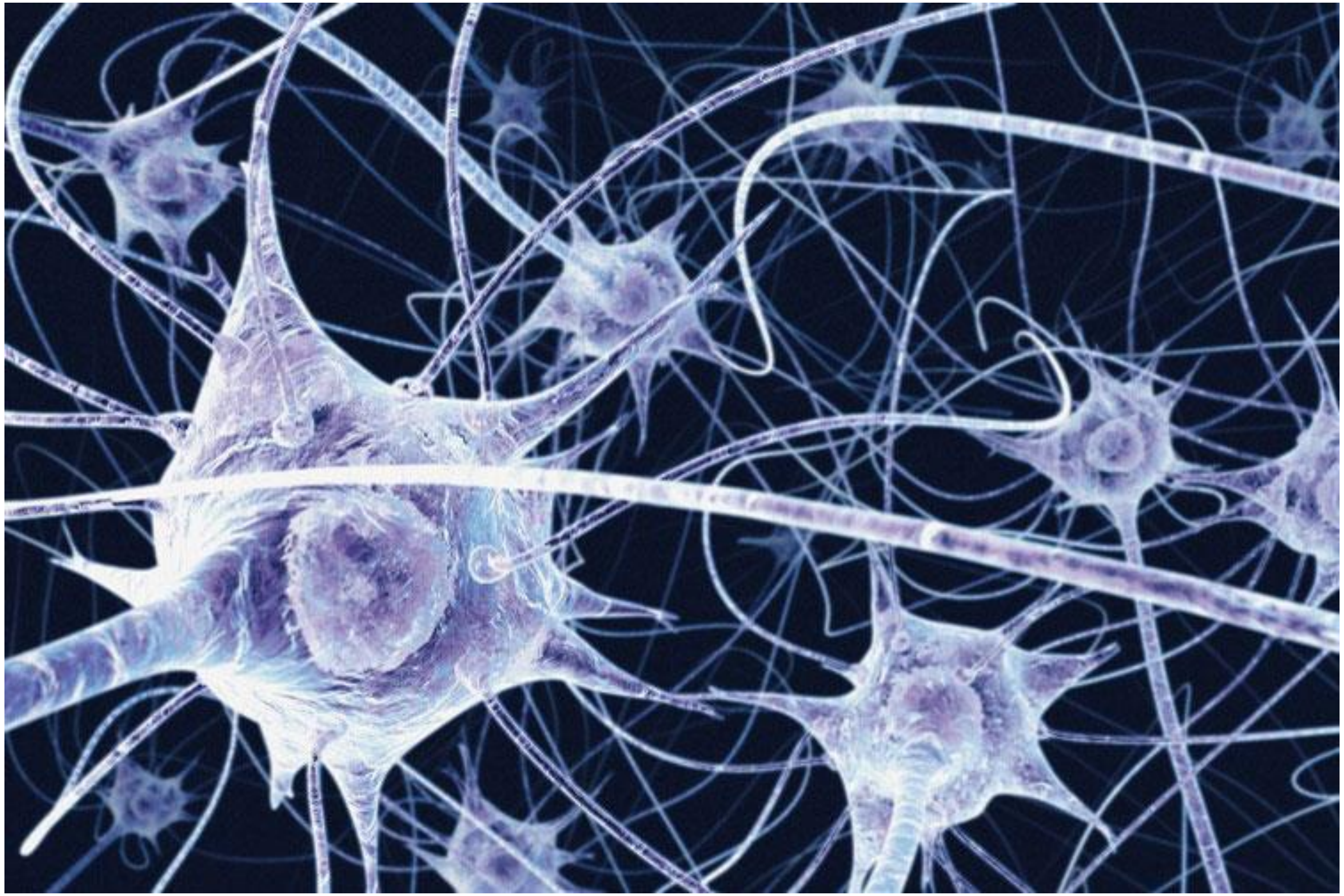
Levels of mathematical modelling in systems neuroscience

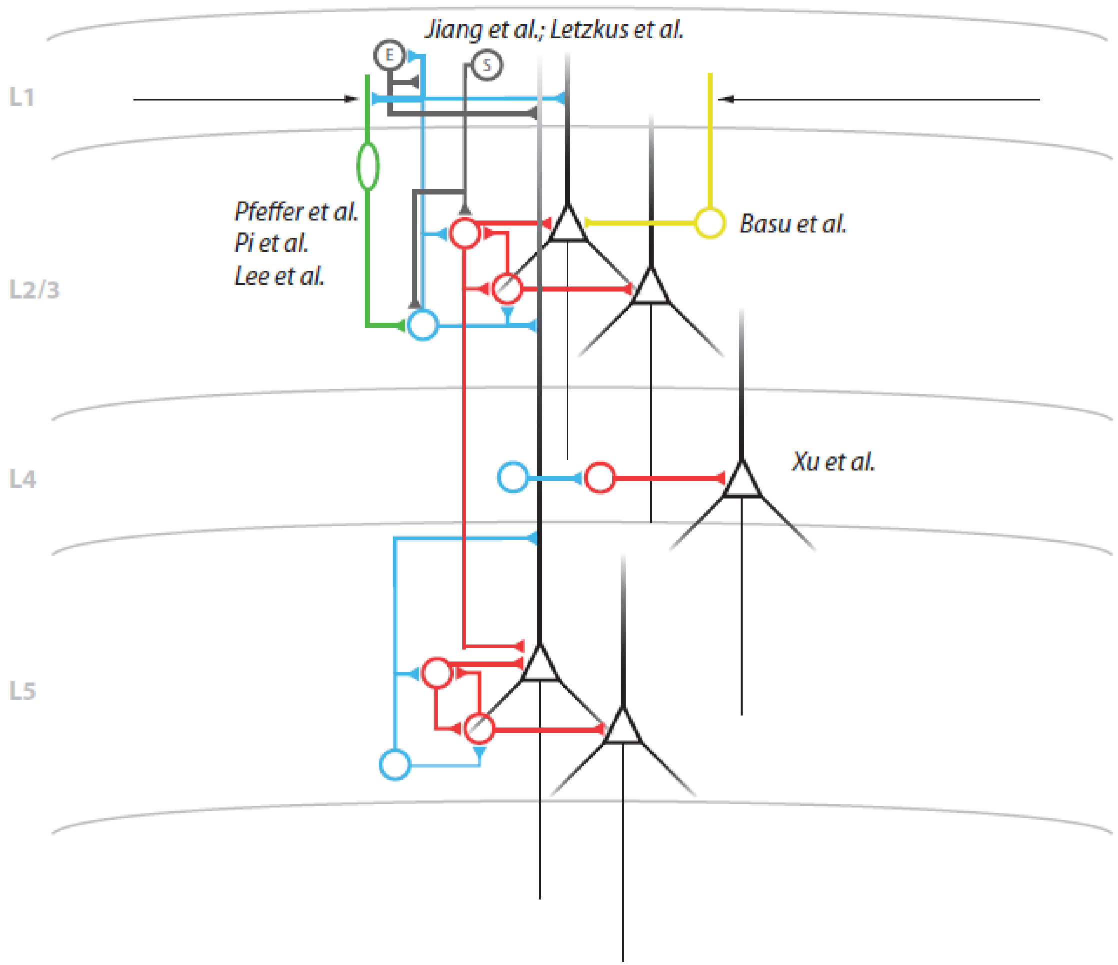


Balazs Hangya
Lendulet Laboratory of Systems Neuroscience
Institute of Experimental Medicine
Hungarian Academy of Sciences

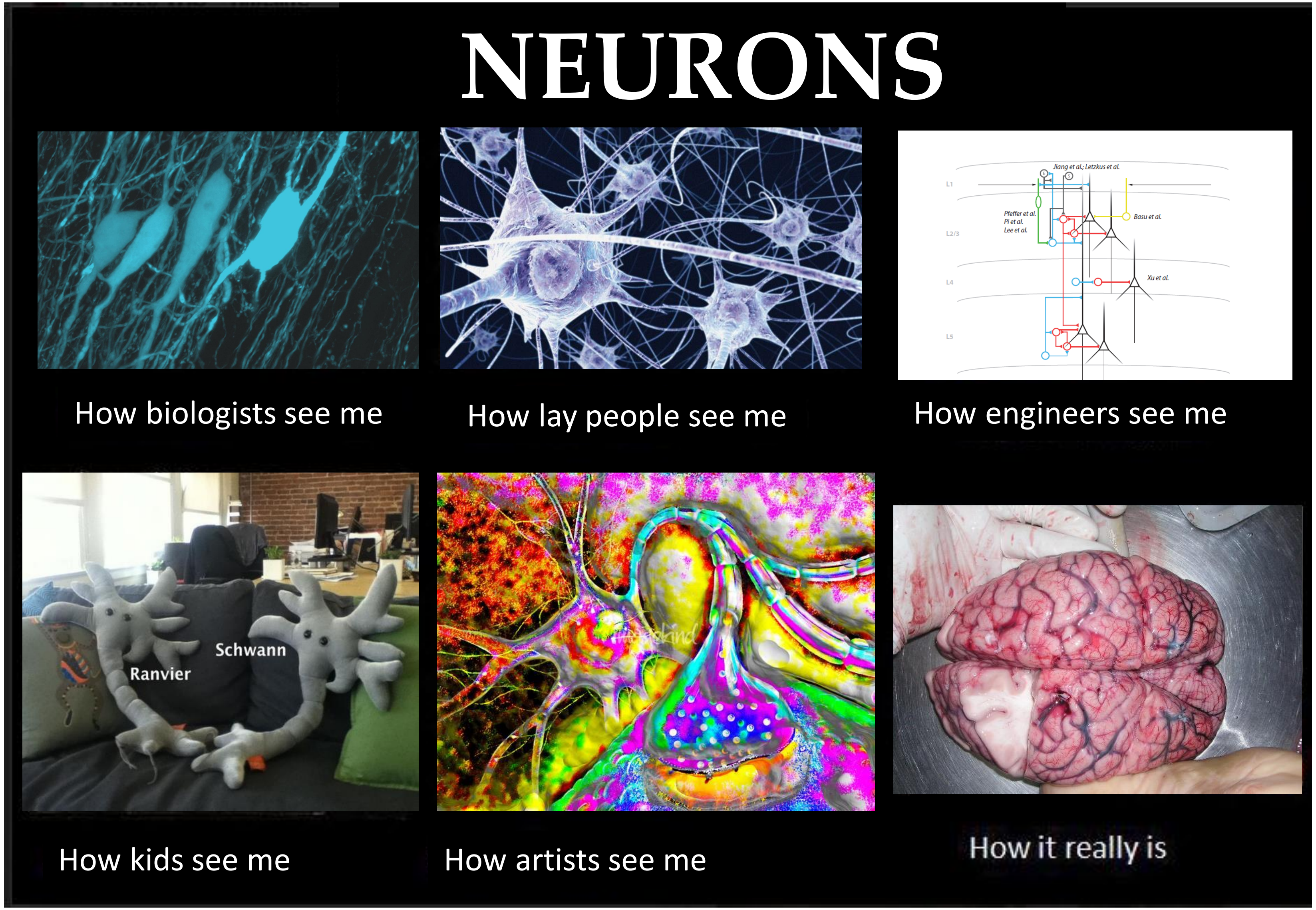
BME Matematikai Modellalkotás Szeminárium
10 Oct 2017, Budapest



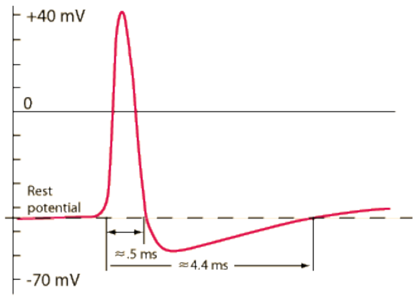


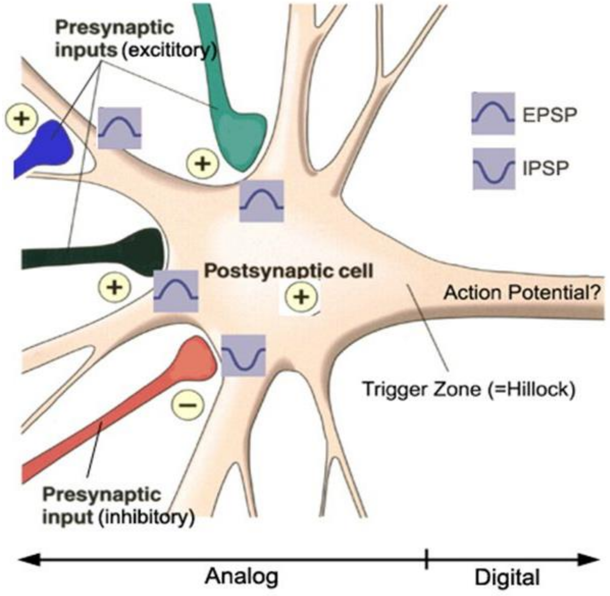


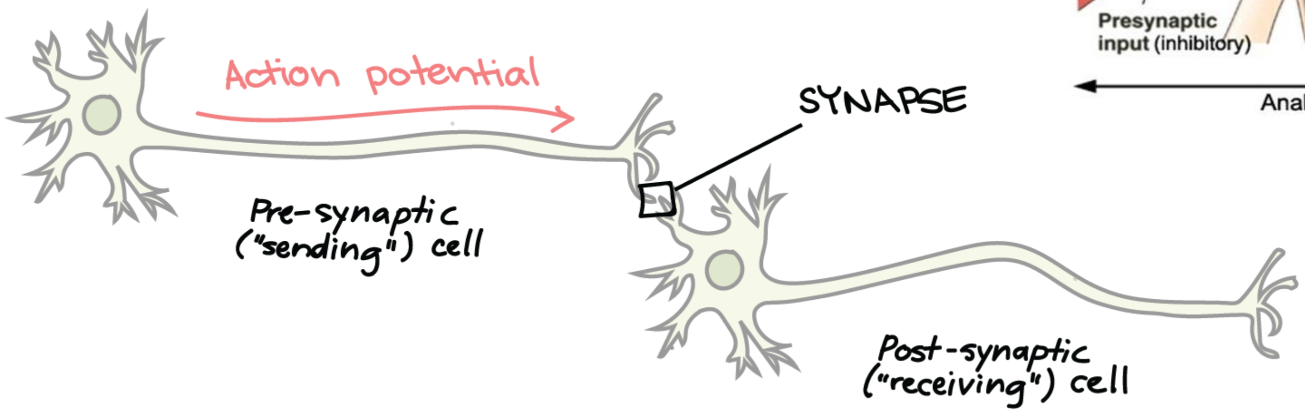
Jiang et al.; Letzkus et al.
E
S
L1
L2/3
L4
L5
Pfeffer et al.
Pi et al.
Lee et al.
Basu et al.
Xu et al.


NEURONS
How biologists see me
How lay people see me
How engineers see me
Schwann
Ranvier
How kids see me
How artists see me
How it really is

Neural communication







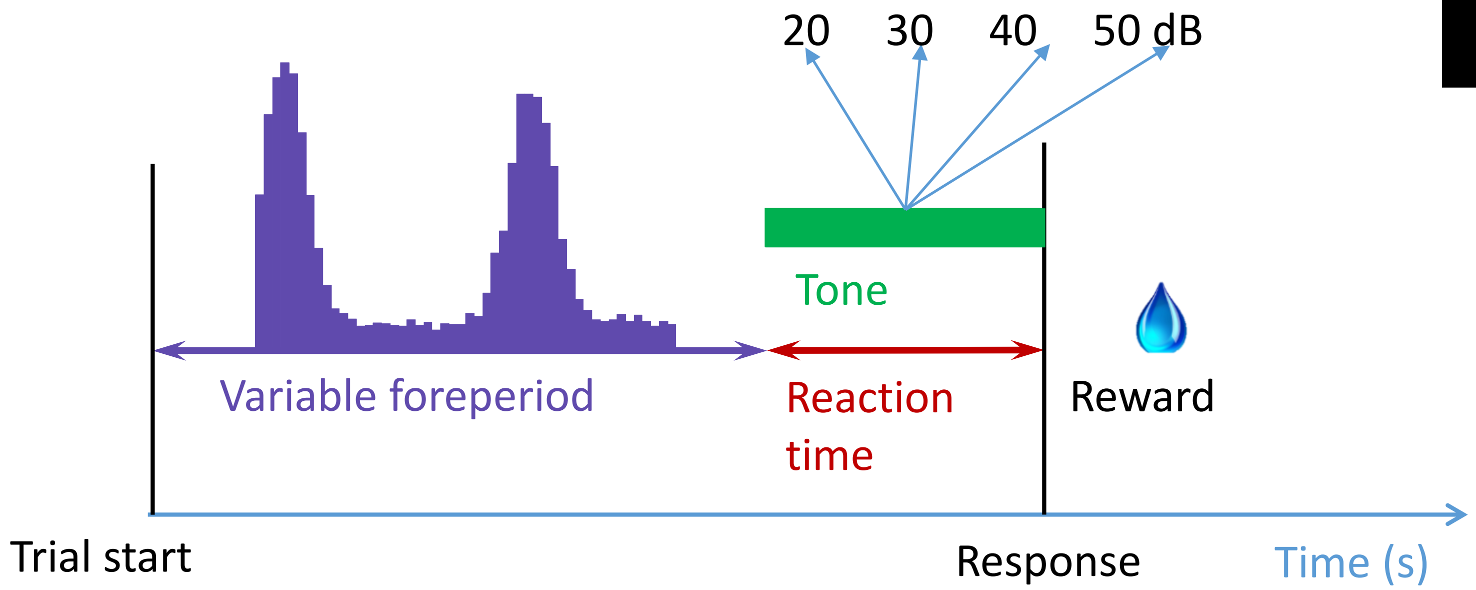
1110011110011110111110111100111011110001111011100011110111101011110011110111101111110111101111110011

Outline

- **Example #1: Temporal focus**
- Example #2: Sustained attention
- Example #3: Reinforcement learning
- Example #4: Decision confidence

Attention task in mice

- Sustained attention task
- Temporal focus (anticipation)



Subjective hazard rate as temporal attention

$$h(t) = \frac{f(t)}{1 - F(t)}$$

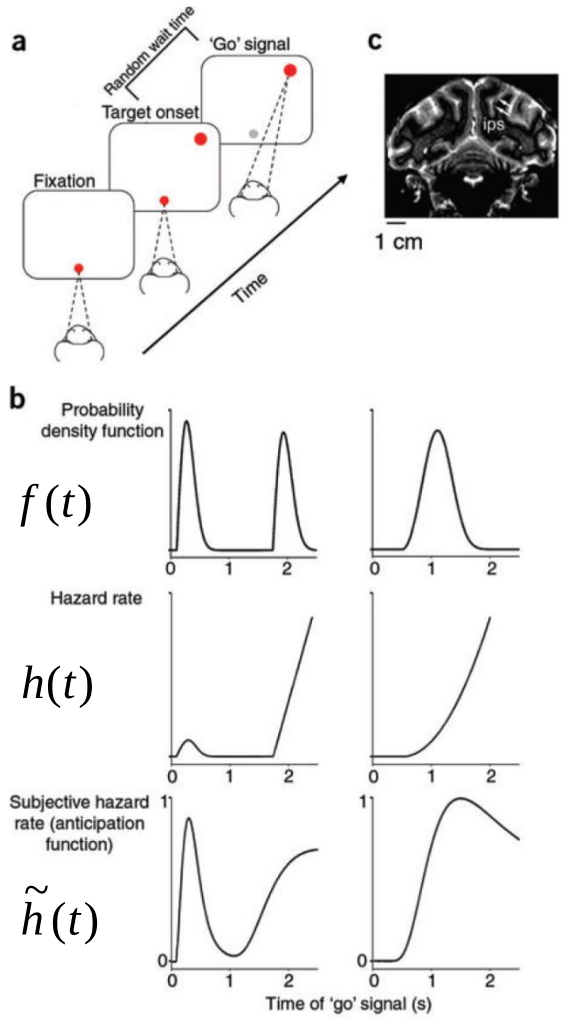
$$F(t) = \int_0^t f(s)ds$$

$$\tilde{f}(t) = \frac{1}{\Phi t\sqrt{2\pi}} \int_{-\infty}^{\infty} f(\tau)e^{-(\tau-t)^2/(2\Phi^2 t^2)}d\tau$$

$$\tilde{F}(t) = \int_0^t \tilde{f}(s)ds$$

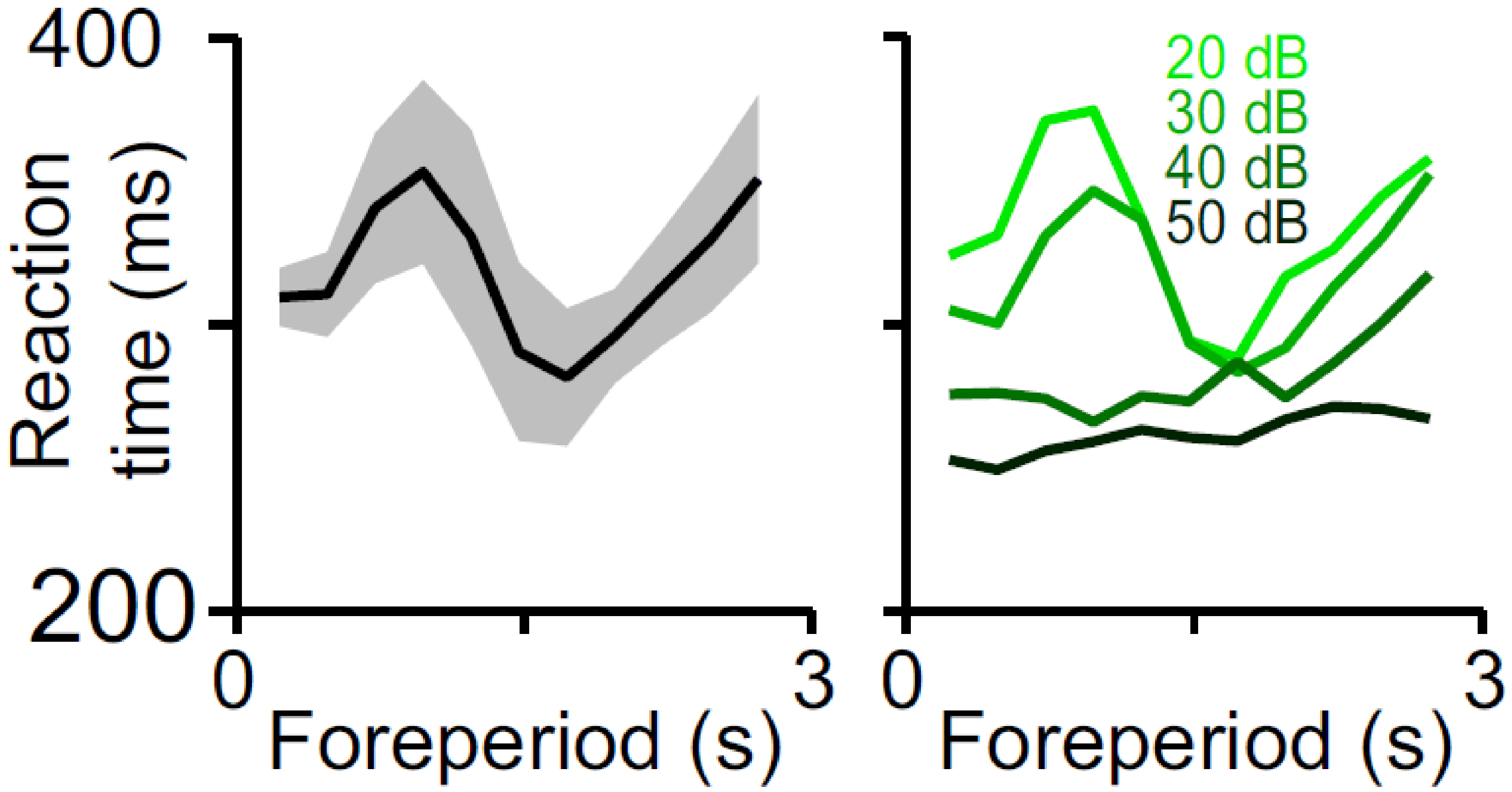
$$\tilde{h}(t) = \frac{\tilde{f}(t)}{1 - \tilde{F}(t)}$$

$$r(t) = w_e + w_u A_u(t-\tau) + w_b A_b(t-\tau) + \varepsilon$$

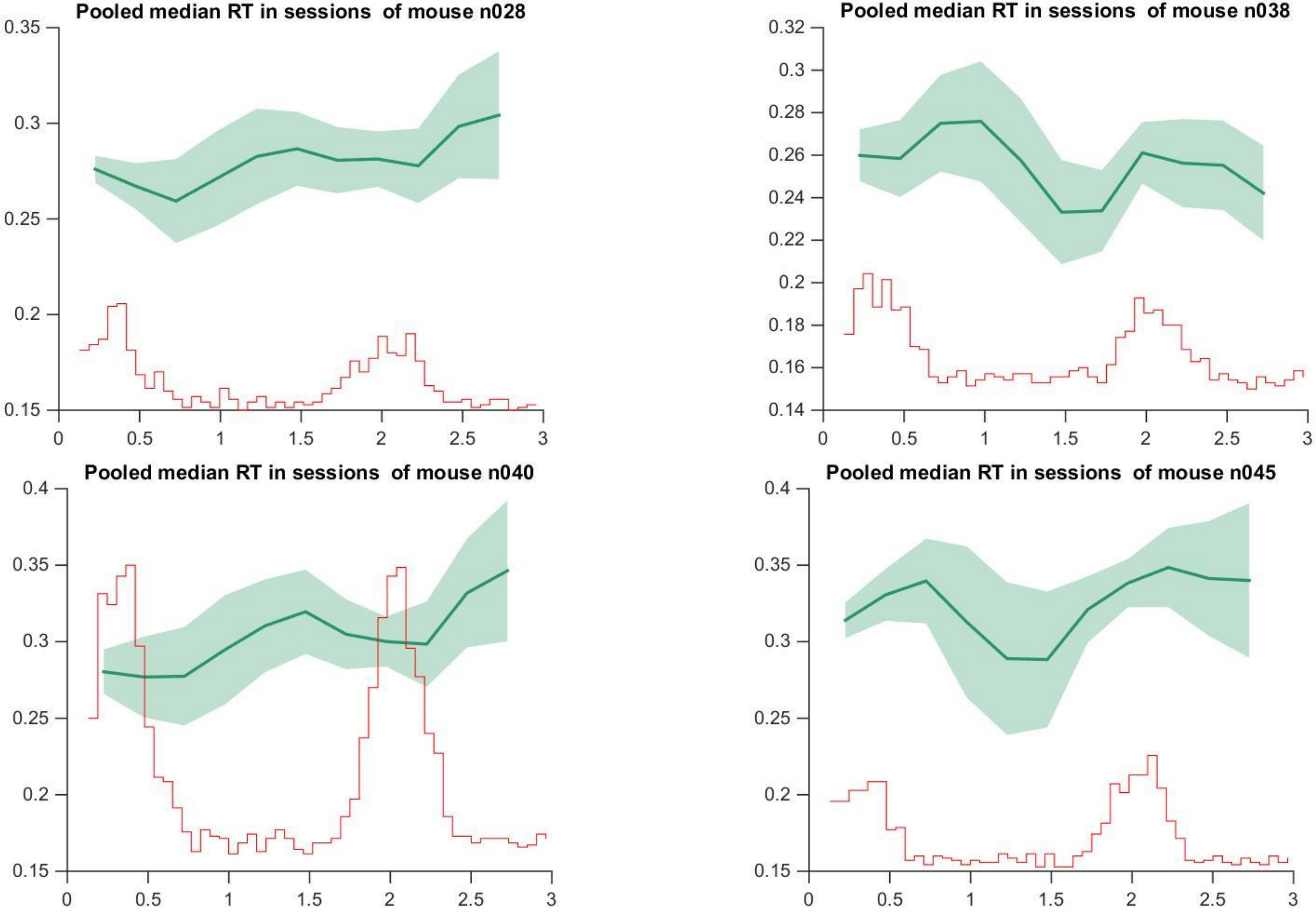


Janssen and Shadlen, 2005

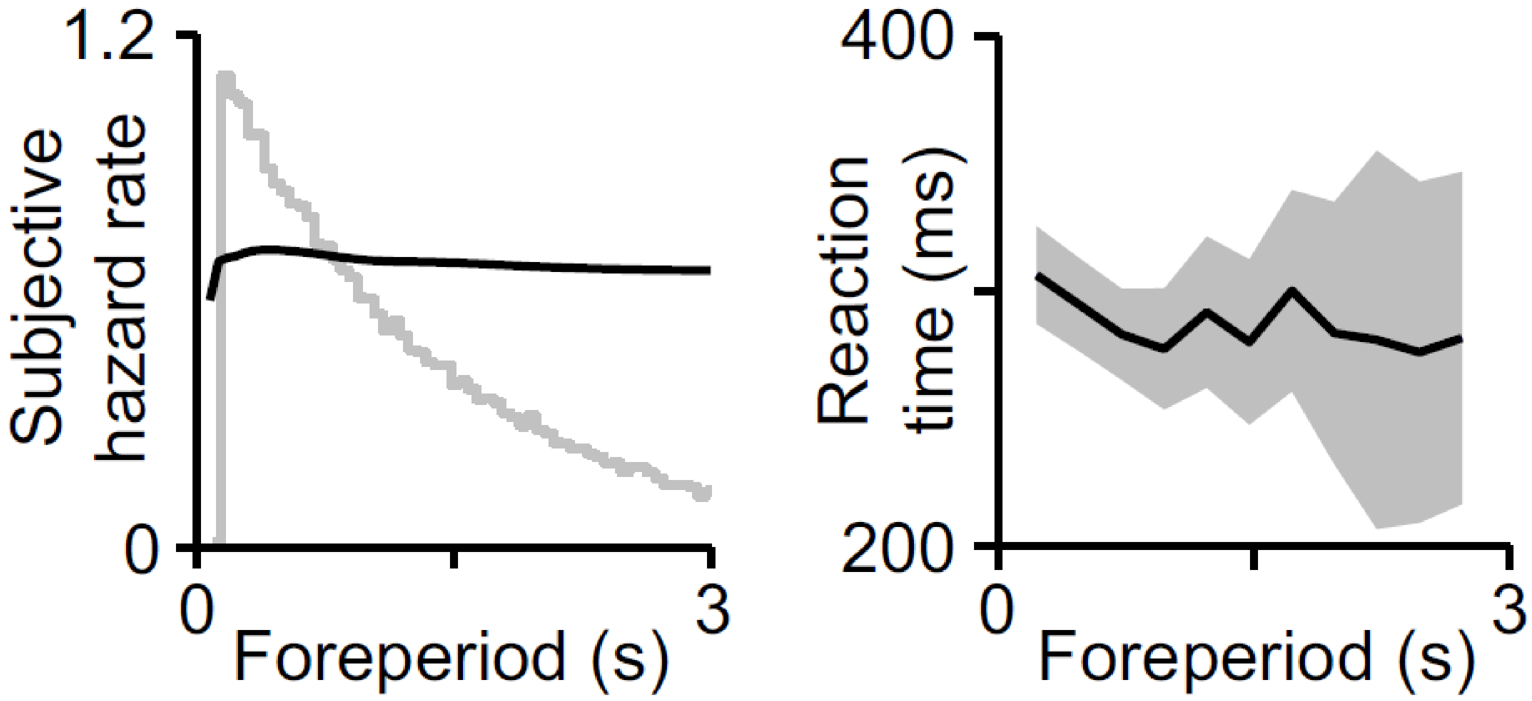
Reaction time as a function of foreperiod



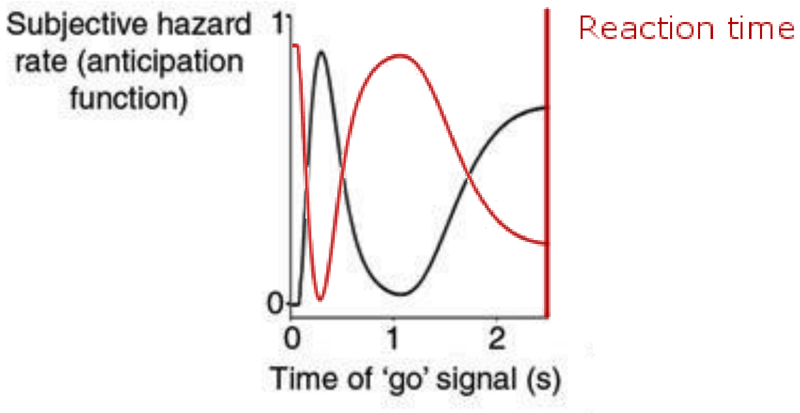
Reaction time as a function of foreperiod



Exponential foreperiod distribution



Different versions of the subjective hazard model



$$\tilde{f}(t) = \frac{1}{\Phi t\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x)e^{-(x-t)^2/(2\Phi^2 t^2)}dx$$

$$\tilde{F}(t) = \int_0^t \tilde{f}(s)ds$$

$$A_b(t) = \frac{\tilde{f}(t)}{1-\tilde{F}(t)}$$

$$r(t) = w_e + w_b A_b(t-\tau)$$

Different versions of the subjective hazard model

4. $r(t) = w_e + w_b A_b(t - \tau)$ $\quad \tilde{f}(t) = \frac{1}{\Phi t\sqrt{2\pi}} \int_{-\infty} f(x) e^{-(x-t)^2/(2\Phi^2 t^2)} dx$

5. $r(t) = w_e + w_r e^{t-\tau} + w_b A_b(t - \tau)$ $\quad \tilde{f}(t) = \frac{1}{\Phi t\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-(x-t)^2/(2\Phi^2 t^2)} dx$

6. $r(t) = w_e + w_b A_b(t - \tau)$ $\quad \tilde{f}(t) = \frac{1}{(\Phi t + \psi)\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-(x-t)^2/(2(\Phi t + \psi)^2)} dx$

7. $r(t) = w_e + w_b A_b(t - \tau)$ $\quad \tilde{f}(t) = \frac{1}{\Phi(t^{\psi})\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-(x-t)^2/(2(\Phi t^{\psi})^2)} dx$

8. $r(t) = w_e + w_r e^{t-\tau} + w_b A_b(t - \tau)$ $\quad \tilde{f}(t) = \frac{1}{(\Phi t + \psi)\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-(x-t)^2/(2(\Phi t + \psi)^2)} dx$

9. $r(t) = w_e + w_r e^{t-\tau} + w_b A_b(t - \tau)$ $\quad \tilde{f}(t) = \frac{1}{\Phi(t^{\psi})\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-(x-t)^2/(2(\Phi t)^2)} dx$

Different versions of the subjective hazard model

10.

$$r(t) = w_e + w_r e^{k(t-\tau)} + w_b A_b(t-\tau)$$

$$\tilde{f}(t) = \frac{1}{\Phi t\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-(x-t)^2/(2\Phi^2 t^2)} dx$$

11.

$$r(t) = w_e + w_r e^{k(t-\tau)} + w_b A_b(t-\tau)$$

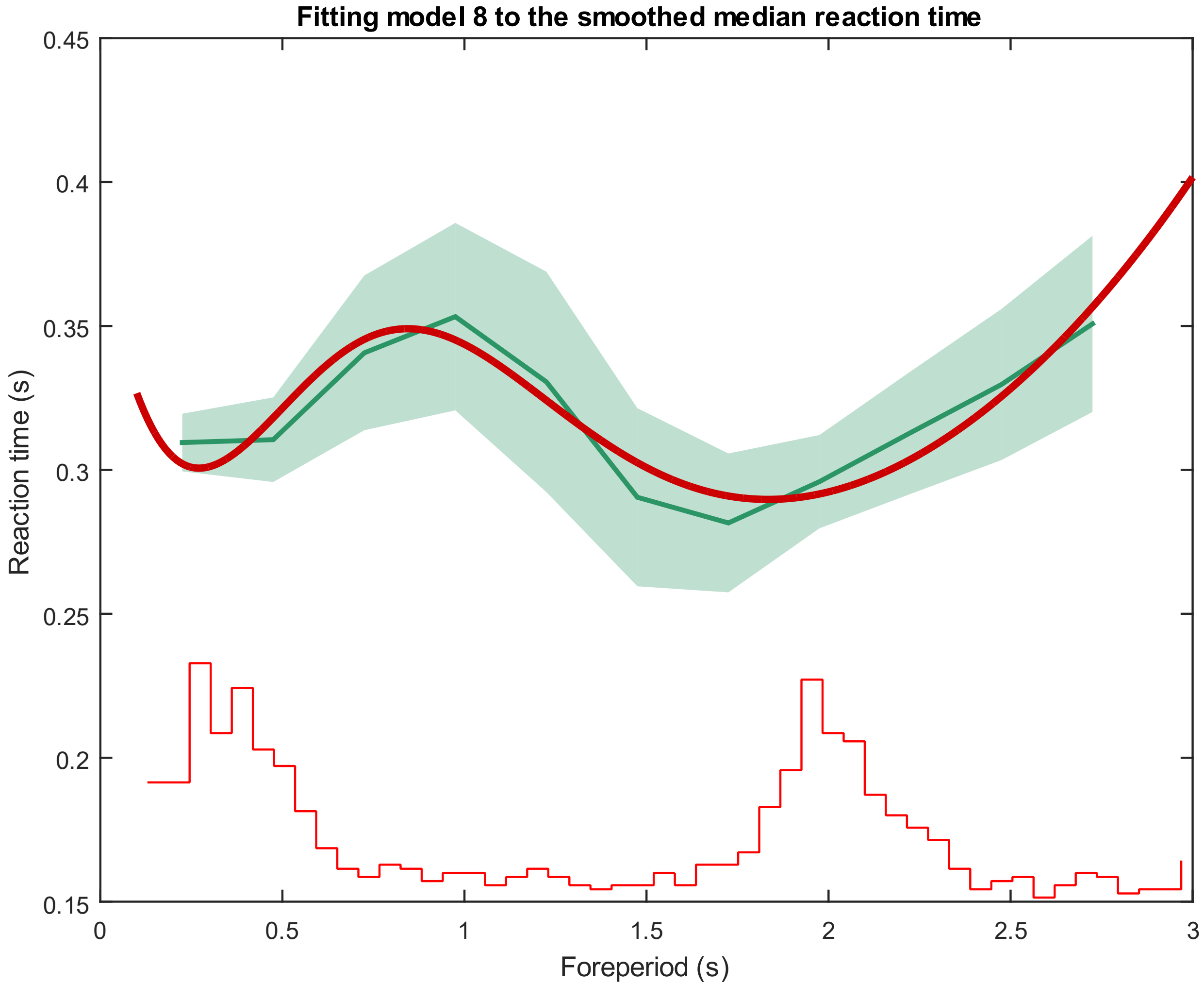
$$\tilde{f}(t) = \frac{1}{(\Phi t + \psi)\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-(x-t)^2/(2(\Phi t+\psi)^2)} dx$$

12.

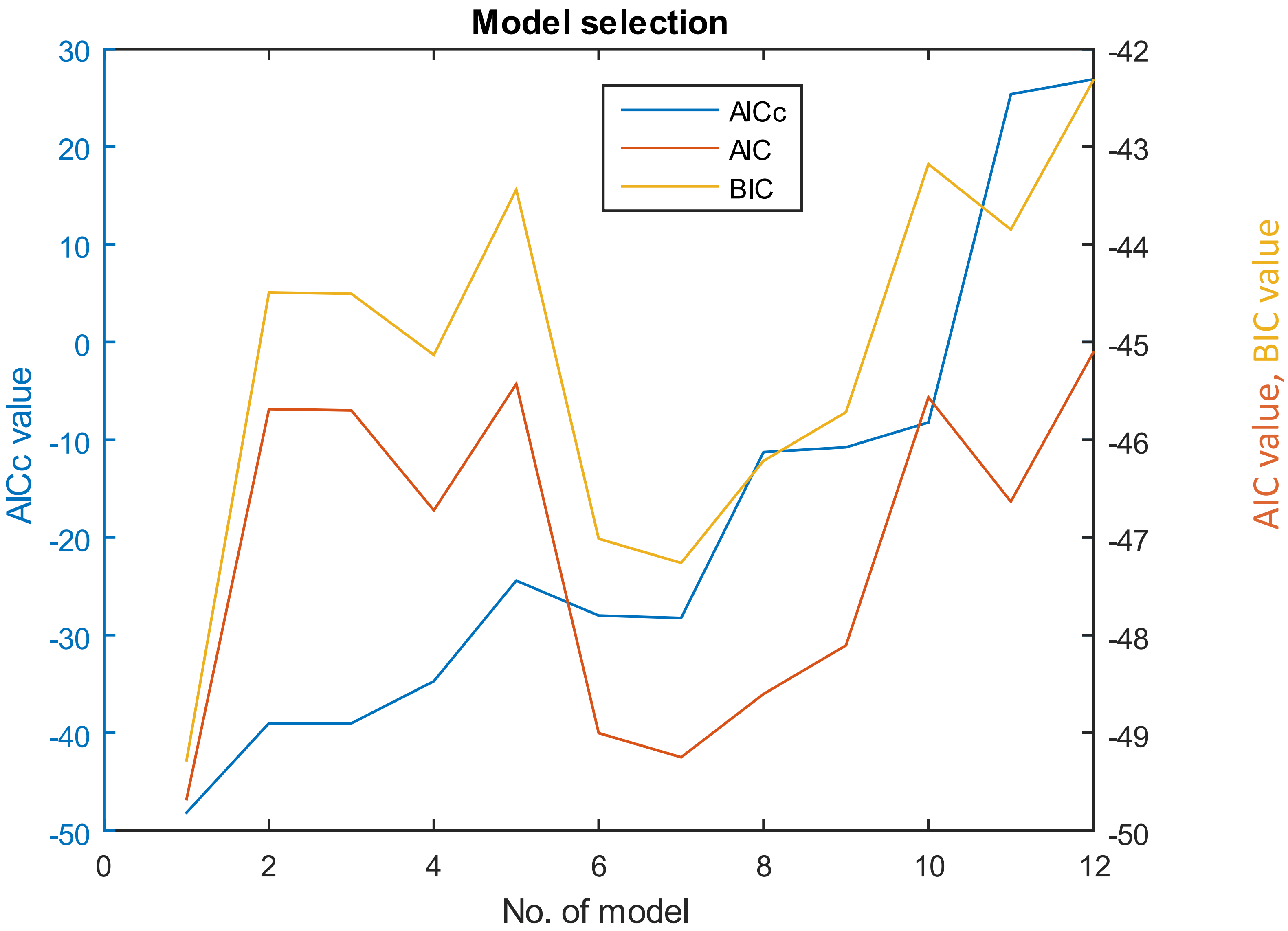
$$r(t) = w_e + w_r e^{k(t-\tau)} + w_b A_b(t-\tau)$$

$$\tilde{f}(t) = \frac{1}{\Phi(t^{\psi})\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-(x-t)^2/(2(\Phi t)^2)} dx$$

Model fitting



Model selection

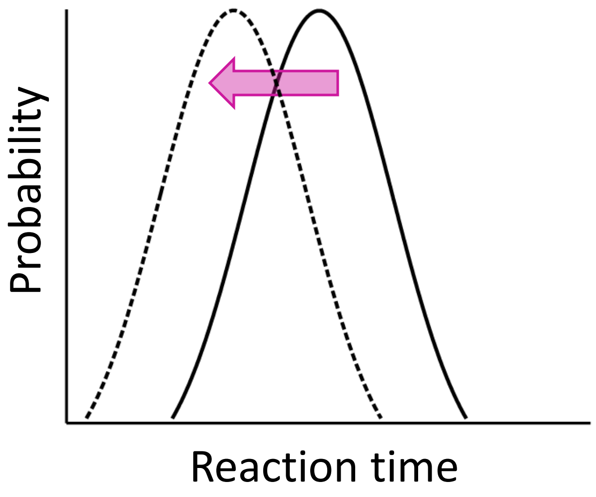


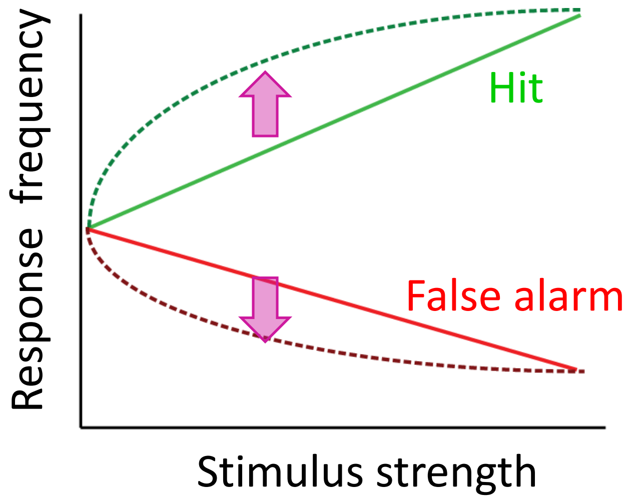
Outline

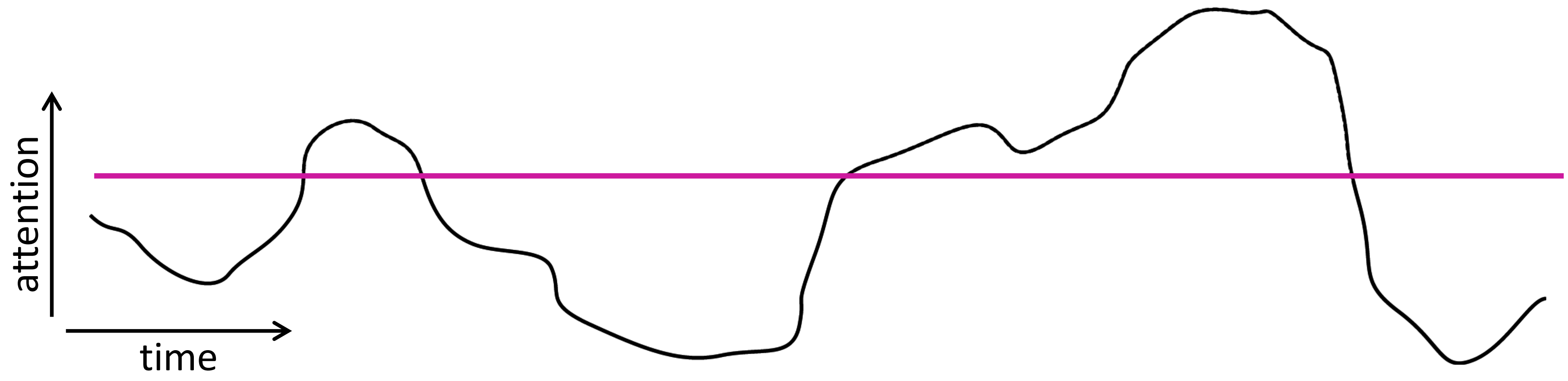
- Example #1: Temporal focus
- **Example #2: Sustained attention**
- Example #3: Reinforcement learning
- Example #4: Decision confidence

Sustained attention operationalized by RT and performance

Behavior varies...

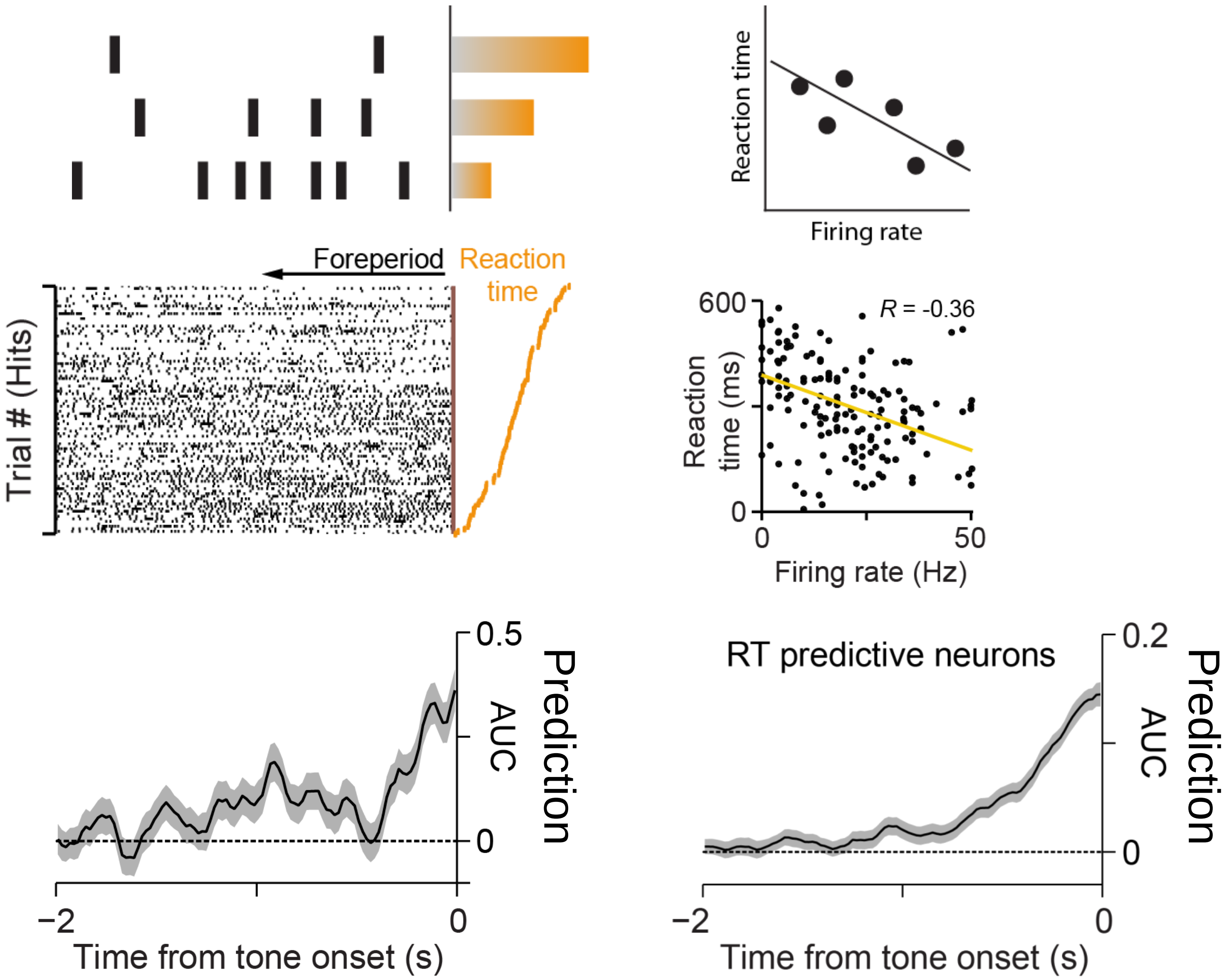




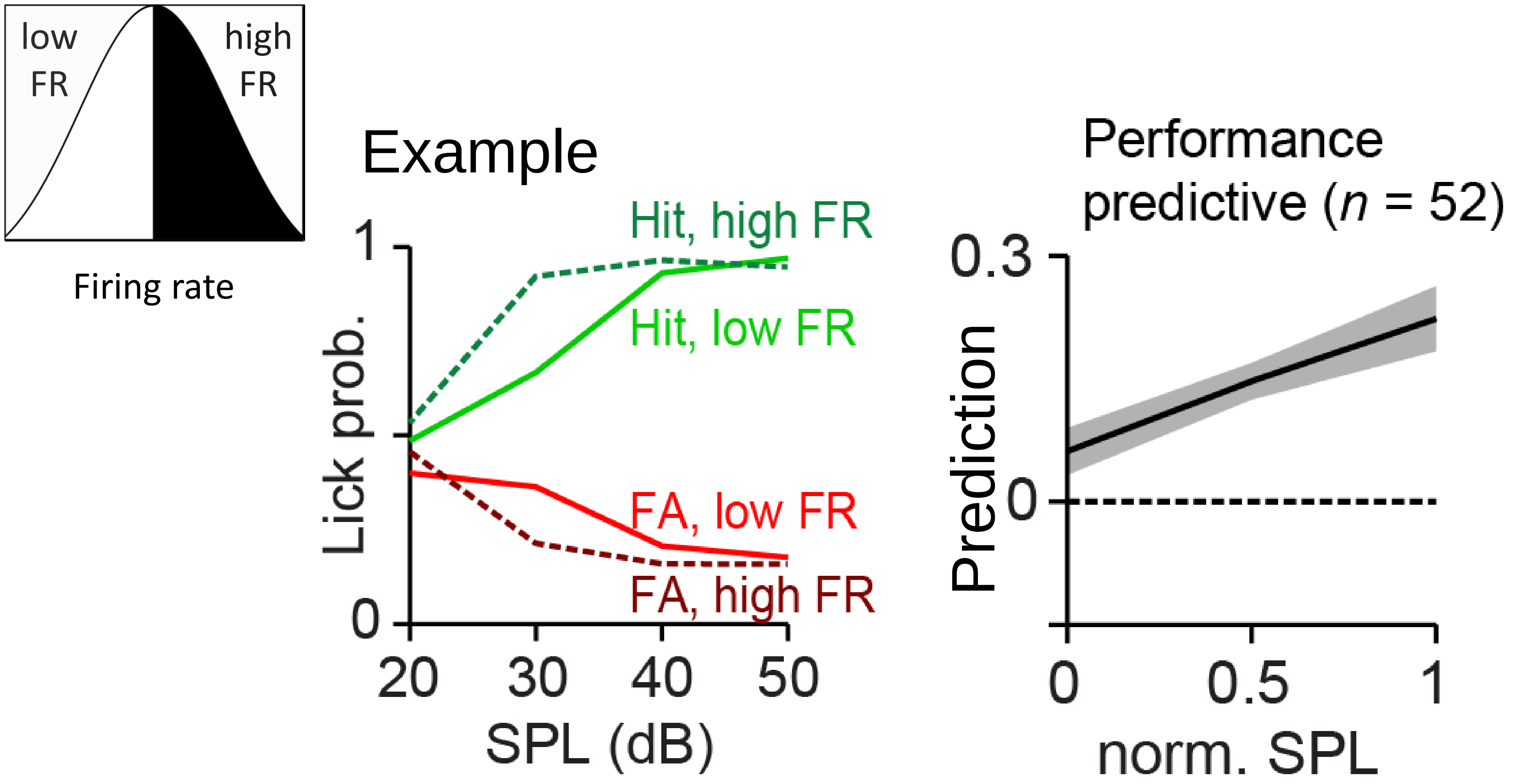


Attention wanders...

Some cells predict reaction time



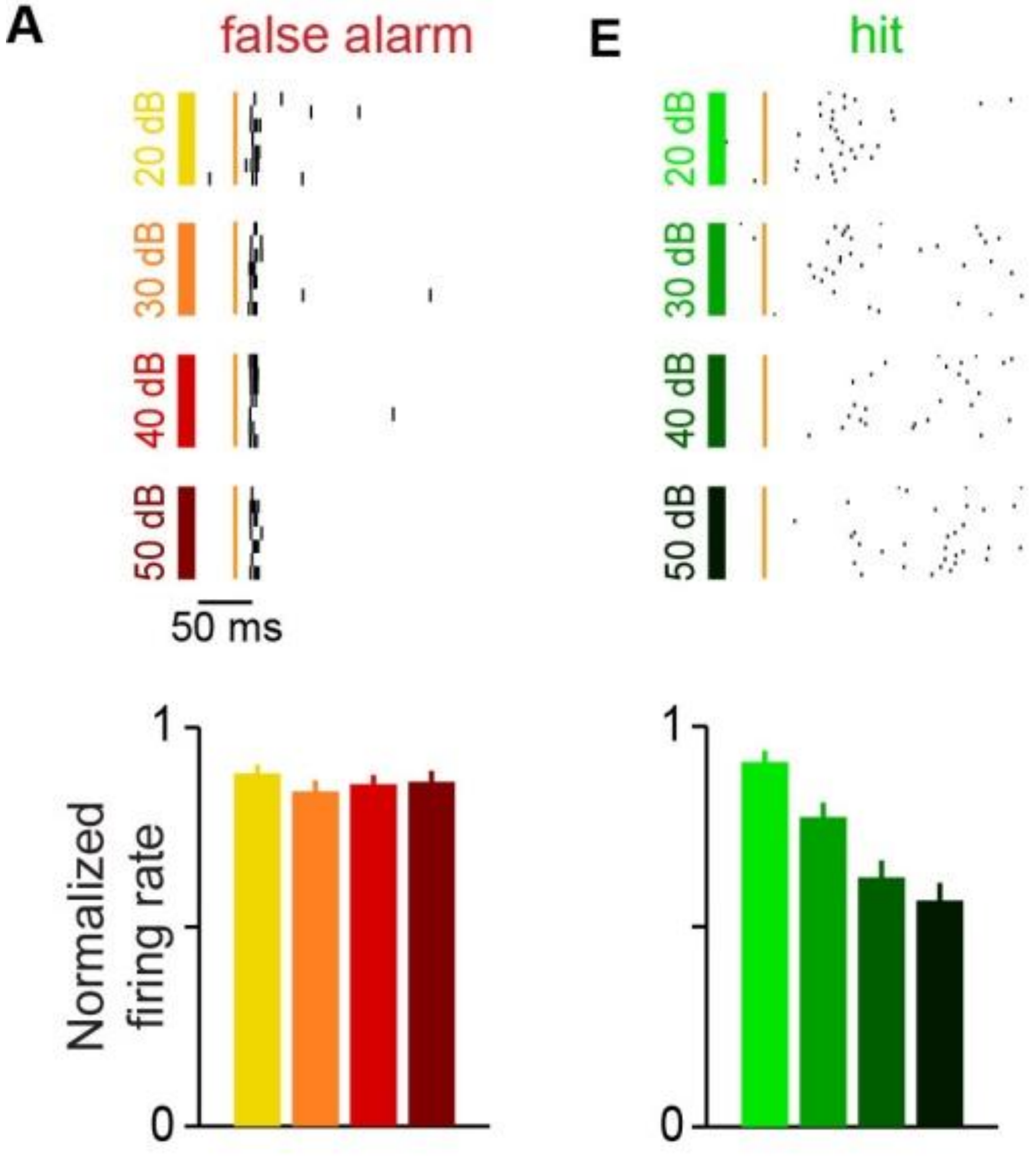
Some cells predict performance



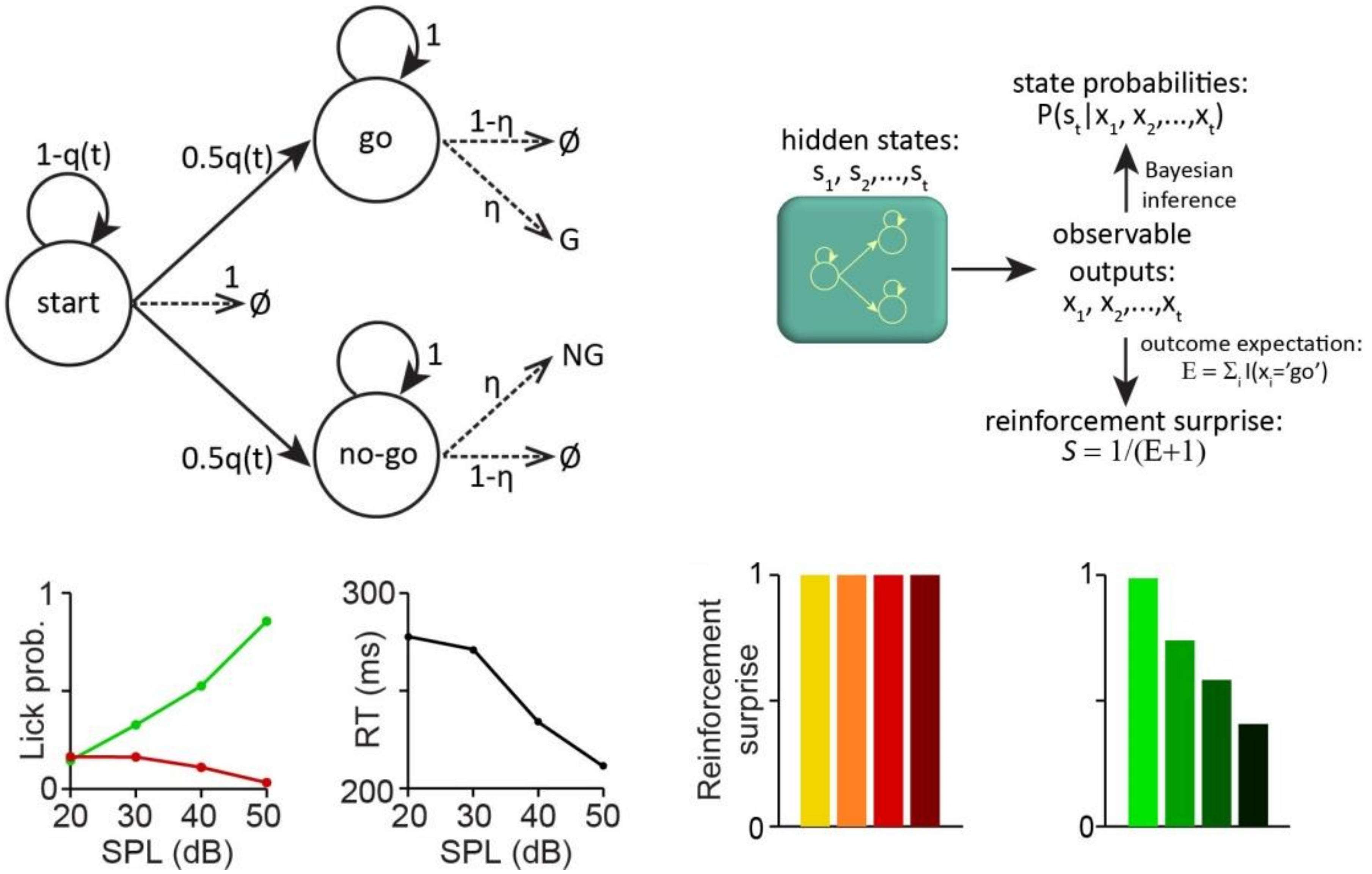
Outline

- Example #1: Temporal focus
- Example #2: Sustained attention
- **Example #3: Reinforcement learning**
- Example #4: Decision confidence

Activation by reward correlates with expectations



Activation by reward correlates with surprise



Outline

- Example #1: Temporal focus
- Example #2: Sustained attention
- Example #3: Reinforcement learning
- **Example #4: Decision confidence**

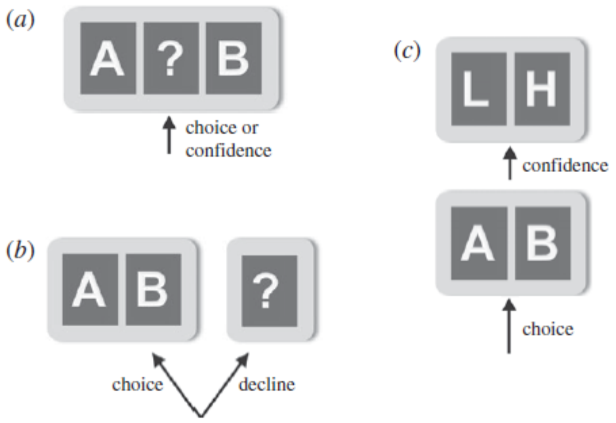
Are you sure?


Decisions are
accompanied by a
“feeling” of decision
confidence

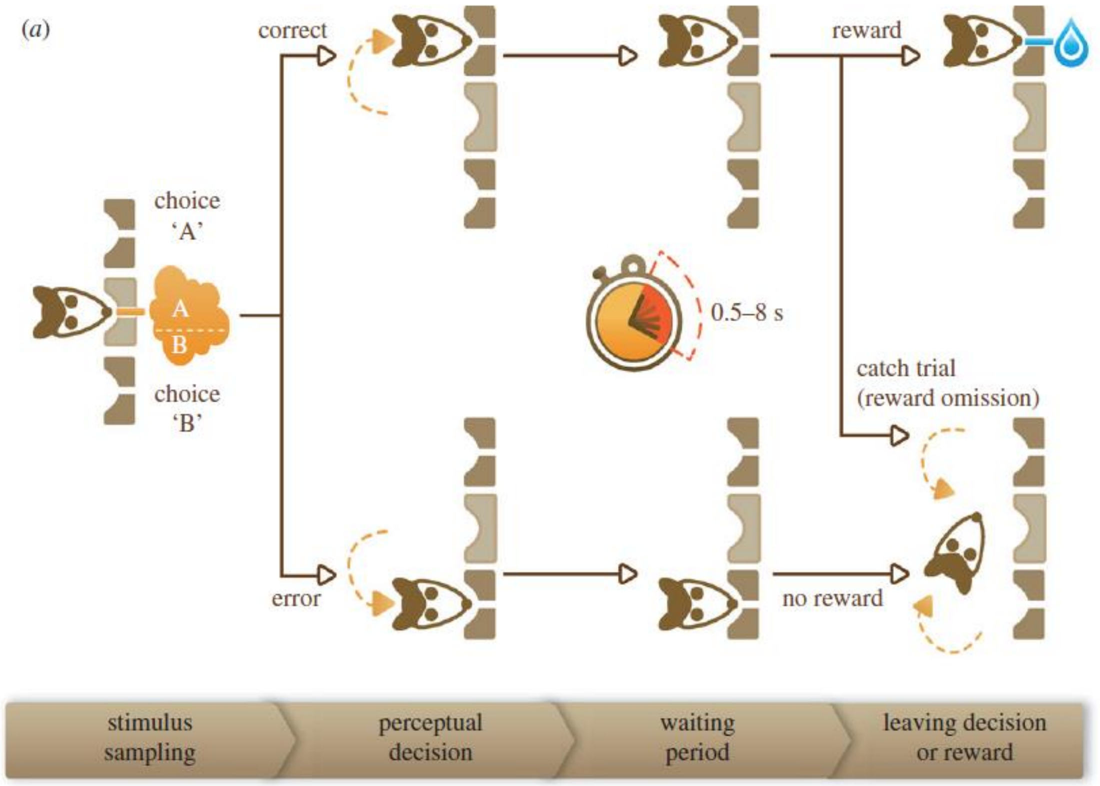
How to probe decision confidence?

Humans: just ask.... – easy (or is it?)

Non-human primates: uncertainty option, opt-out tasks, post-decision wager

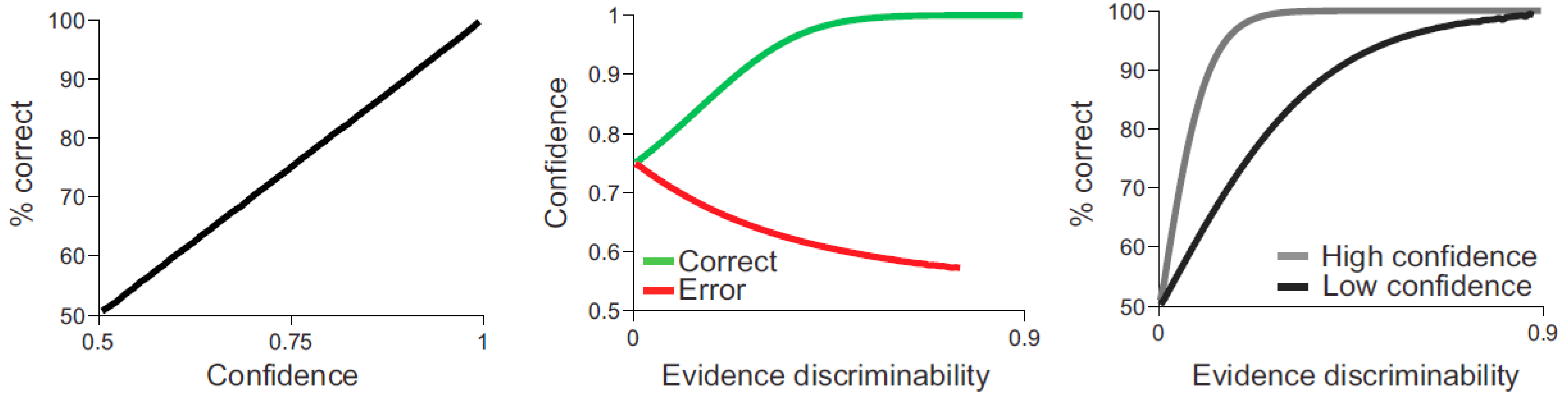


Rodents: the waiting time task



Kepecs & Mainen

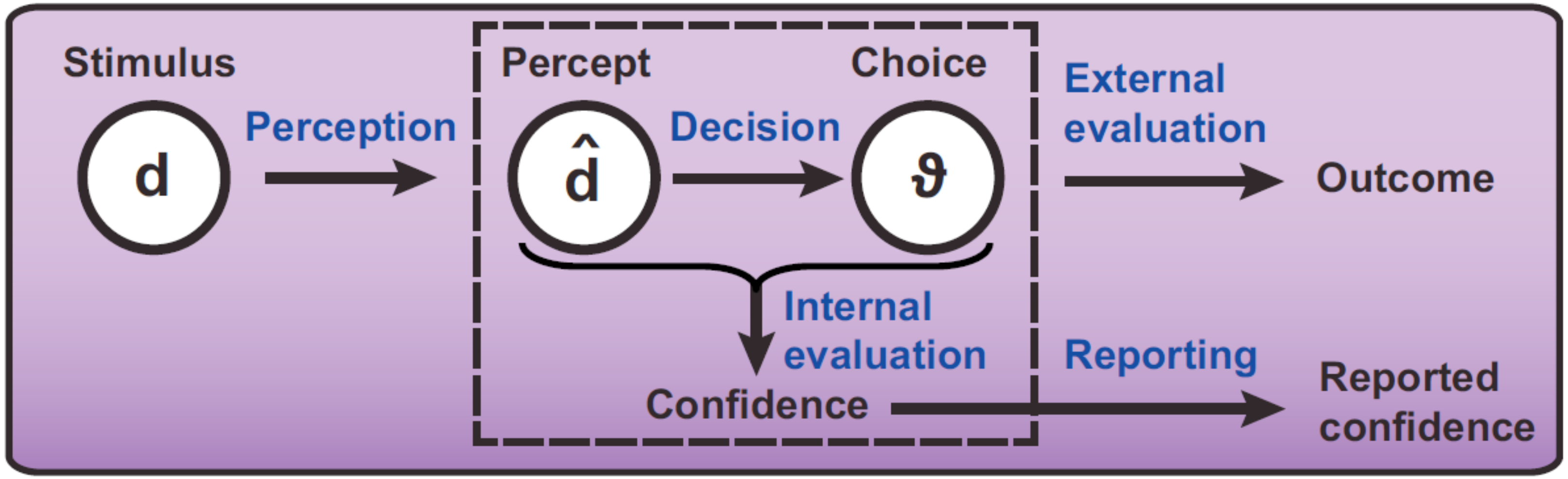
We find recurring patterns of confidence...

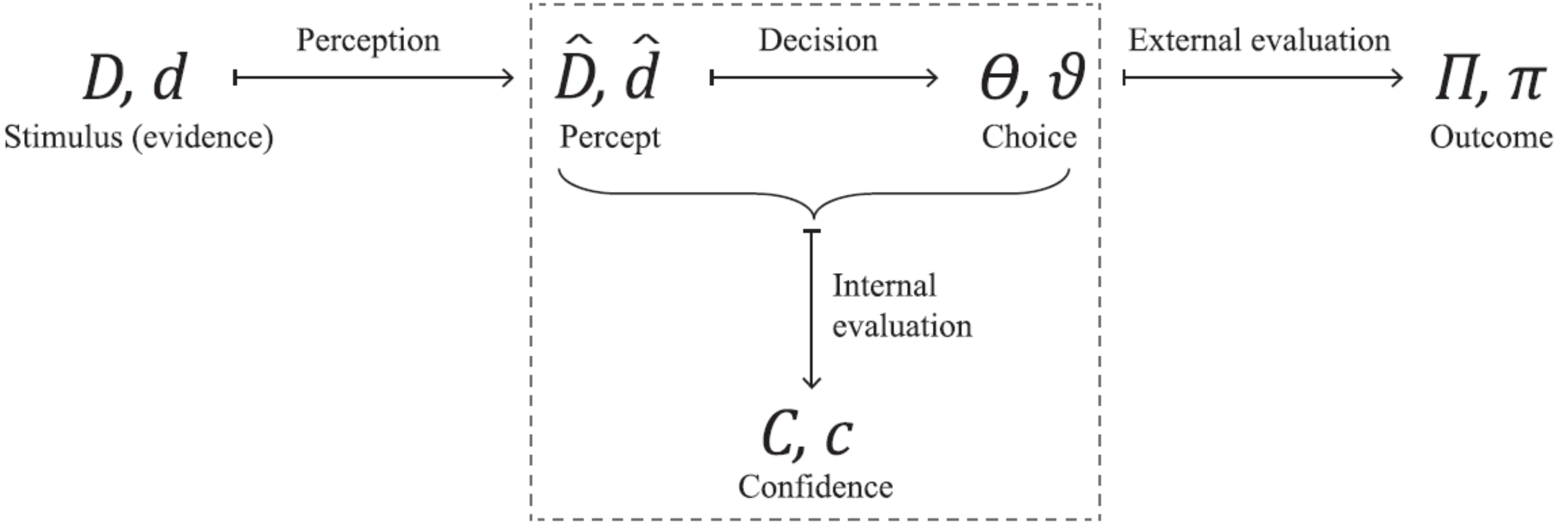


...Is there a reason for that? What are the mathematical laws that describe decision confidence? How general are they?

Hangya, Sanders, Kepecs 2016, Neural Computation

Fully stochastic model of decision making





Hangya, Sanders, Kepecs 2016, Neural Computation

Statistical decision confidence

The choice can be evaluated in terms of a hypothesis testing problem:

1. Null hypothesis (H_0): The choice $\vartheta = \theta(\hat{d})$ is incorrect;
2. Alternative hypothesis (H_1): The choice $\vartheta = \theta(\hat{d})$ is correct.

Definition 2. *Define confidence as*

$$c = P(H_1 | \hat{d}, \vartheta).$$

Equivalently,

$$c = P(\Pi(\theta) = 1 | \hat{d}, \vartheta).$$

Hangya, Sanders, Kepecs 2016, Neural Computation

The belief function

Definition 3. *Define the belief function* $\xi : \mathcal{R}(\hat{D}) \times \mathcal{R}(\theta) \to [0, 1]$ *as*

$$\xi(\hat{d}, \vartheta) = P(H_1 | \hat{d}, \vartheta) = P(\Pi(\theta) = 1 | \hat{d}, \vartheta), \tag{2.3}$$

where $\mathcal{R}(\hat{D})$ *denotes percept space and* $\mathcal{R}(\theta)$ *denotes the range of all possible choices (i.e., the choice space).*

Definition 4. *Accuracy is the expected proportion of correct choices:*

$$A = E[\Pi(\theta)]. \tag{2.4}$$

Hangya, Sanders, Kepecs 2016, Neural Computation

First theorem: Confidence predicts accuracy

We seek to determine the following function: $f : [0, 1] \to [0, 1]$, $f : c \mapsto A_c$, where A_c is the accuracy for choices with a given confidence. Our claim is that this function is the identity.

Theorem 1. *Accuracy equals confidence:*

$$A_c = c.$$

Proof: Fairly easy.

Hangya, Sanders, Kepecs 2016, Neural Computation

Define difficulty

Definition 5. *Define evidence discriminability as a (deterministic) function of the evidence distribution:*

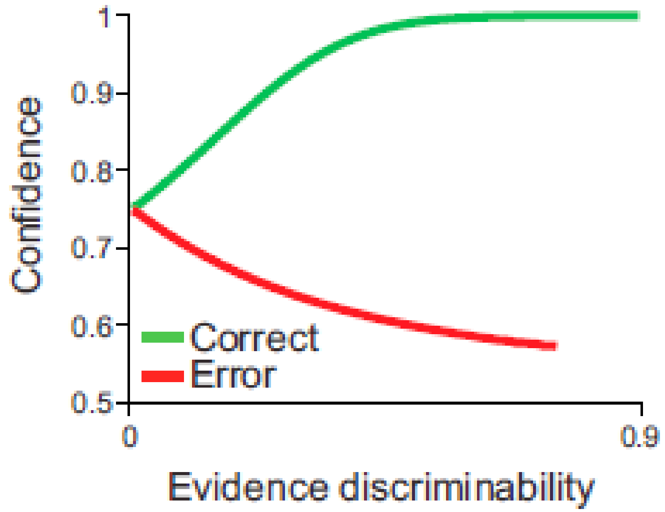
$$\Delta = \Delta(P(D)). \tag{2.6}$$

The evidence discriminability function has to fulfill the following property:

$$\begin{aligned}\Delta(P_1(D)) > \Delta(P_2(D)) &\iff P(H_1|P_1(D)) > P(H_1|P_2(D)) \\ &\iff P(\Pi(\theta) = 1|P_1(D)) > P(\Pi(\theta) = 1|P_2(D)) \\ &\iff E(\Pi(\theta)|P_1(D)) > E(\Pi(\theta)|P_2(D)),\end{aligned} \tag{2.7}$$

Hangya, Sanders, Kepecs 2016, Neural Computation

Second theorem: Confidence increases for correct and decreases for incorrect choices with decreasing difficulty

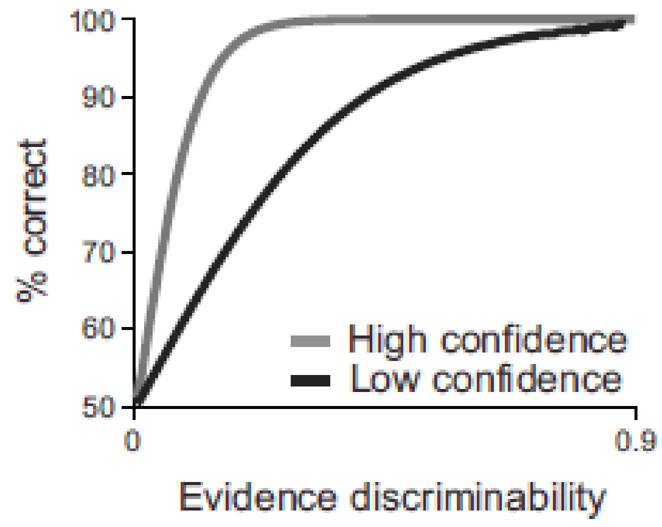


Theorem 2. *Let us assume that:*

- *Belief independence: the belief function (ξ) is independent of evidence discriminability*
- *Percept monotonicity: for any given confidence c, the relative frequency of percepts mapping to c by ξ changes monotonically with evidence discriminability for any fixed choice.*

Under these assumptions, confidence increases for correct choices and decreases for incorrect choices with increasing evidence discriminability.

Third theorem: Confidence predicts outcome beyond difficulty



Theorem 3. *For any given evidence discriminability, accuracy for low-confidence choices is not larger than that of high-confidence choices (splitting the confidence distribution at any particular value). A strict inequality holds in all cases when accuracy is dependent on the percept.*

Proof: Fairly straightforward consequence of the first theorem.

Hangya, Sanders, Kepecs 2016, Neural Computation

Fourth theorem: Average confidence in neutral evidence

Theorem 4. *Assuming*

- *The percept is determined by a symmetric distribution centered on the evidence ("symmetric noise model"),*
- *The evidence is distributed uniformly over the evidence space,*
- *The choice is deterministic,*

the average confidence for neutral evidence is precisely 0.75.

Lemma 1. *Integrating the product of the probability density function and the distribution function of any probability distribution symmetric to zero over the positive half-line results in 3/8:*

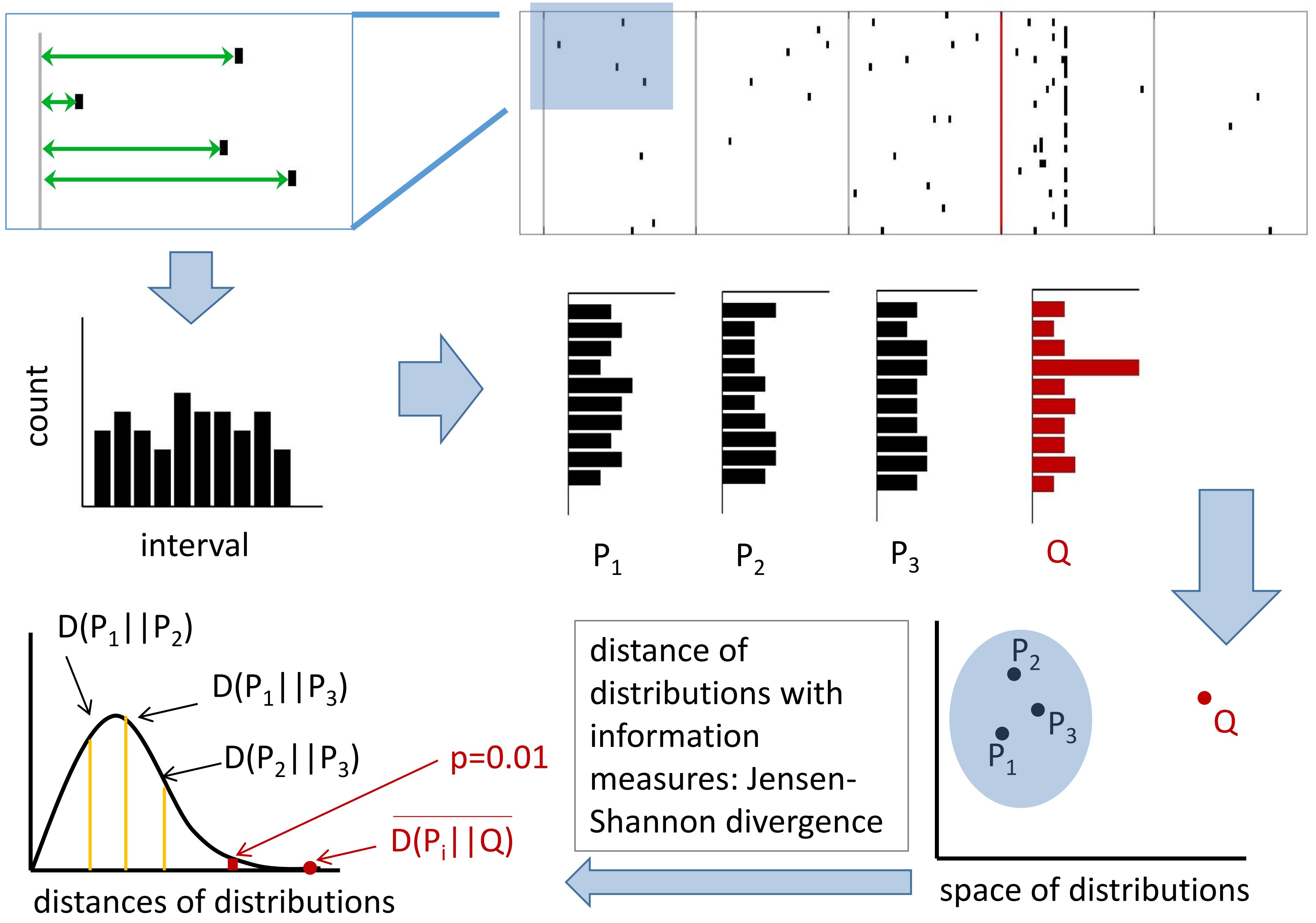
$$\int_0^\infty f(t)F(t)\mathrm{d}t = \frac{3}{8}. \tag{2.8}$$

Hangya, Sanders, Kepecs 2016, Neural Computation

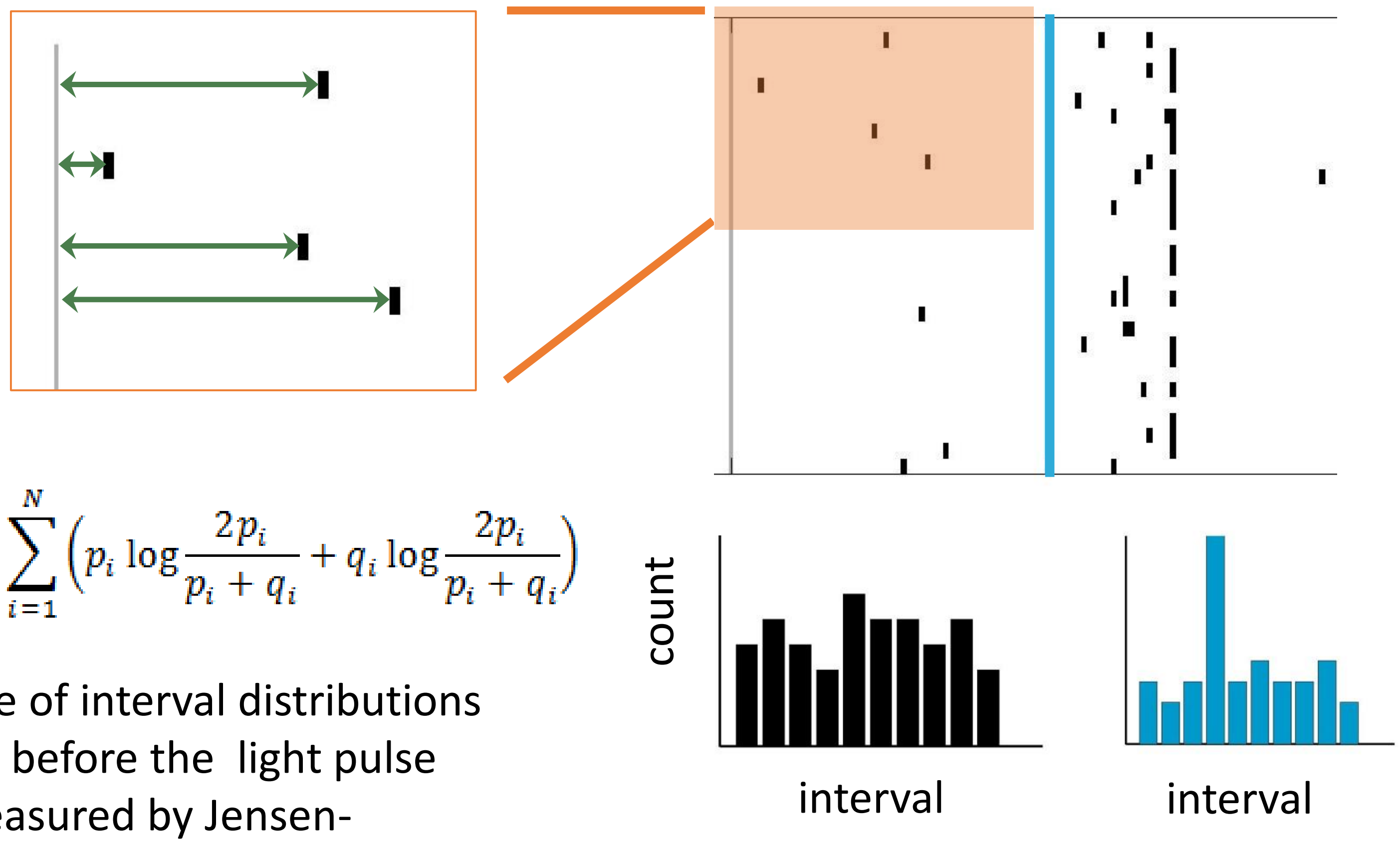
Outline

- Example #1: Temporal focus
- Example #2: Sustained attention
- Example #3: Reinforcement learning
- Example #4: Decision confidence
- **Bonus: Hypothesis testing for action potential timing**

Stimulus-Associated Spike Latency Test



Stimulus-Associated Spike Latency Test



$$D_{JS}^2(P||Q) = \sum_{i=1}^{N} \left(p_i \log \frac{2p_i}{p_i + q_i} + q_i \log \frac{2p_i}{p_i + q_i} \right)$$

distance of interval distributions after vs before the light pulse was measured by Jensen-Shannon information divergence

Acknowledgement

Temporal focus

Tamas Tardos

Sustained attention and reinforcement learning

Sachin P. Ranade

Maja Lorenc

Decision confidence

Joshua I.Sanders

SALT

Duda Kvitsiani

Adam Kepecs

Acknowledgement

kepecslab, Cold Spring Harbor

Duda Kvitsiani
Sachin P. Ranade
Hyun-Jae Pi
Maja Lorenc
Adam Kepecs

hangyalab, Hungarian Academy of Sciences

Panna Hegedüs
Nicola Solari
Diána Balázsfi
Katalin Sviatkó
Barnabás Kocsis

Tamás Laszlovszky
Bálint Király
Flóra Bús
Eszter Ujvári
Katalin Lengyel

The John Merck Fund
National Institute of Neurological Disorders and Stroke
Marie Curie Actions
Magyar Tudományos Akadémia 1825
Lendület program
erc