Stimulus Representation in Temporal-difference Models of the Dopamine System

Rich Sutton
Department of Computing Science
University of Alberta

joint work with Elliot Ludvig and Jim Kehoe

Outline

- General comments about modeling
- Overview of TD dopamine models
- Stimulus representations
- Core TD algorithm
- Reprise of CSC vs microstimuli comparison
- Conclusion

R L
&
A I

Reinforcement Learning and Artificial Intelligence

PIs:
Rich Sutton
Michael Bowling
Dale Schuurmans
Csaba Szepesvari

University of Alberta

Alberta Ingenuity Centre for Machine Learning

Informatics iCORE Circle of Research Excellence

Reinforcement learning and temporal-difference learning

- Algorithms have been validated within four research communities
  - Artificial intelligence
  - Psychology
  - Operations research
  - Neuroscience

Marr's three levels of explanation for information-processing systems

- Computation theory
  - *What* is computed?
- Algorithms
  - *How* is it computed?
- Implementation
  - Really, how is it done?

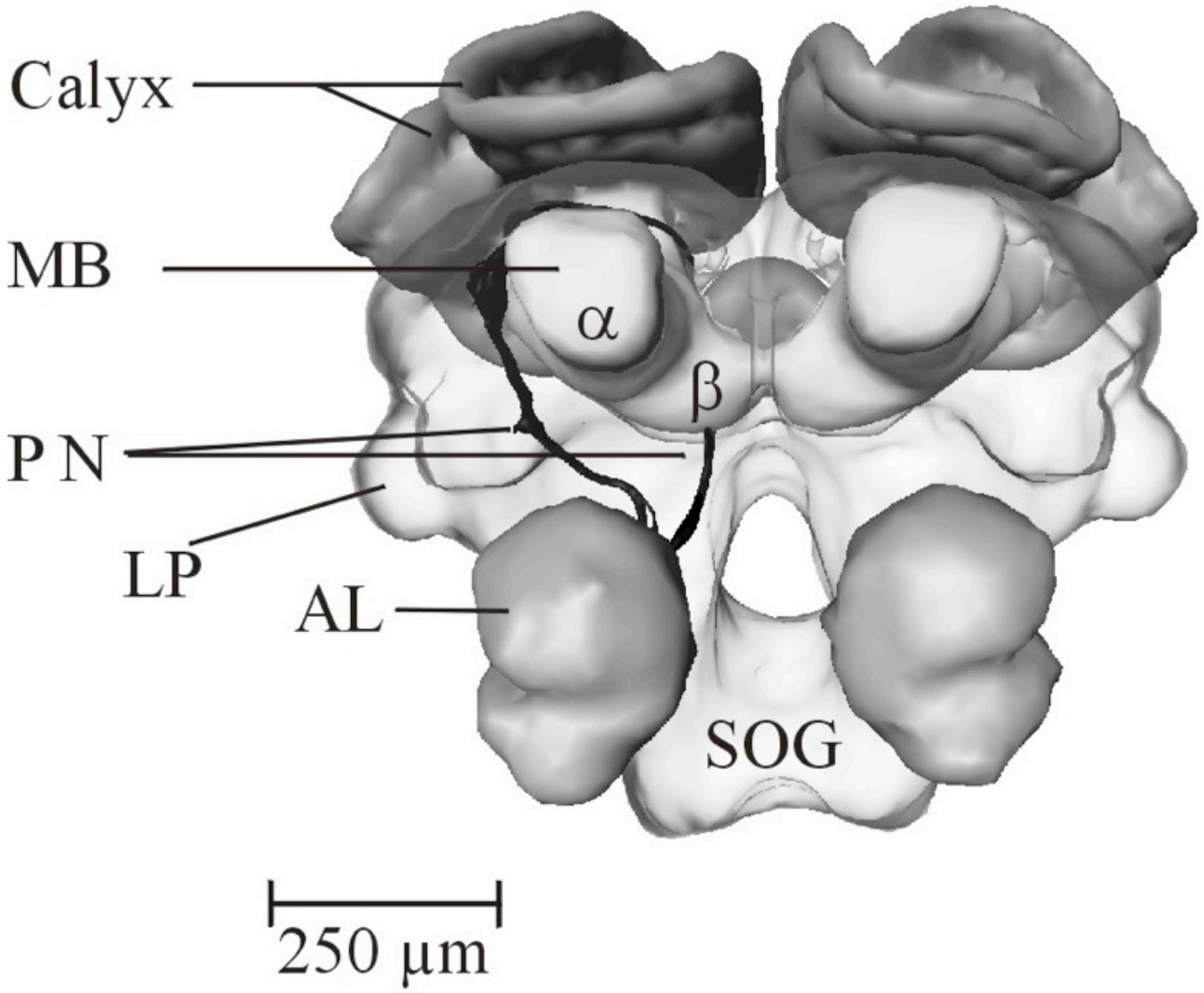
Reward pred. (TD) error

TD models

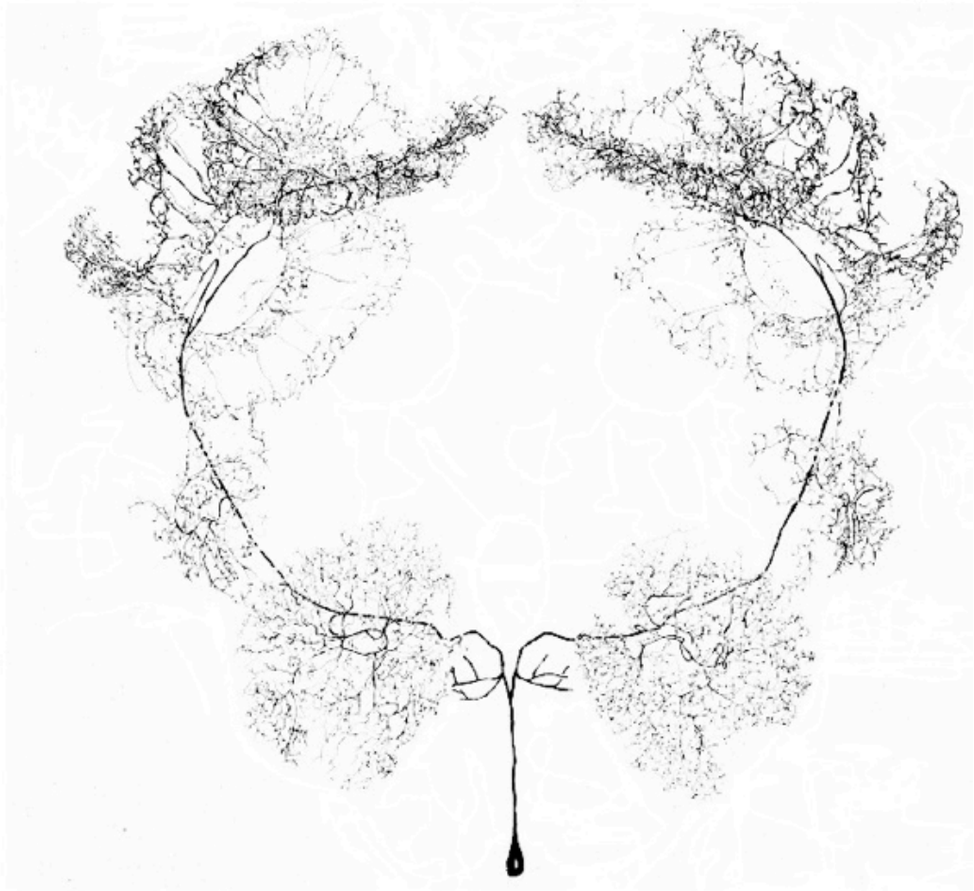
TD error = Dopamine

Levels can be validated independently

Brain Reward Systems

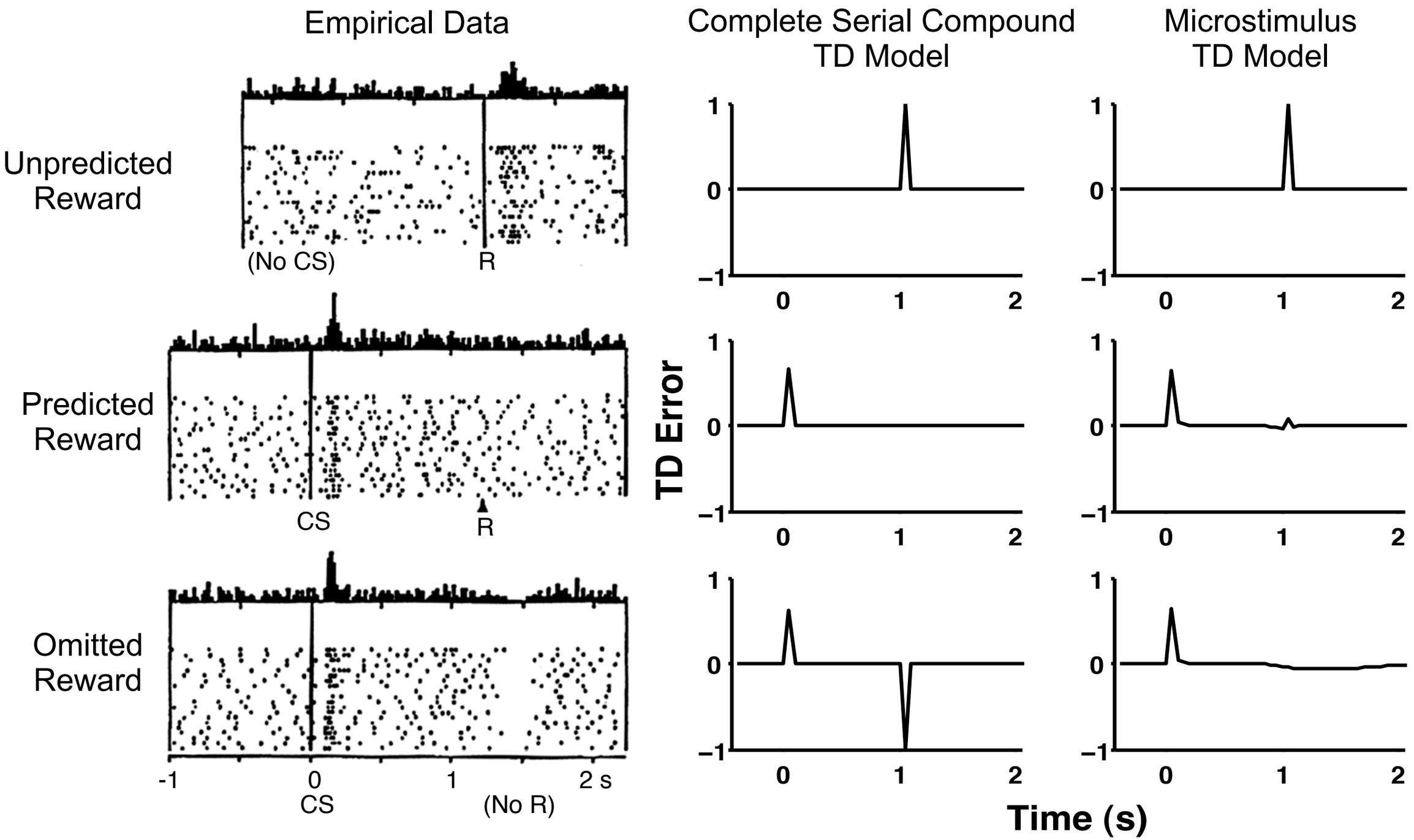


Honeybee brain



VUM neuron

R. Menzel, M. Hammer



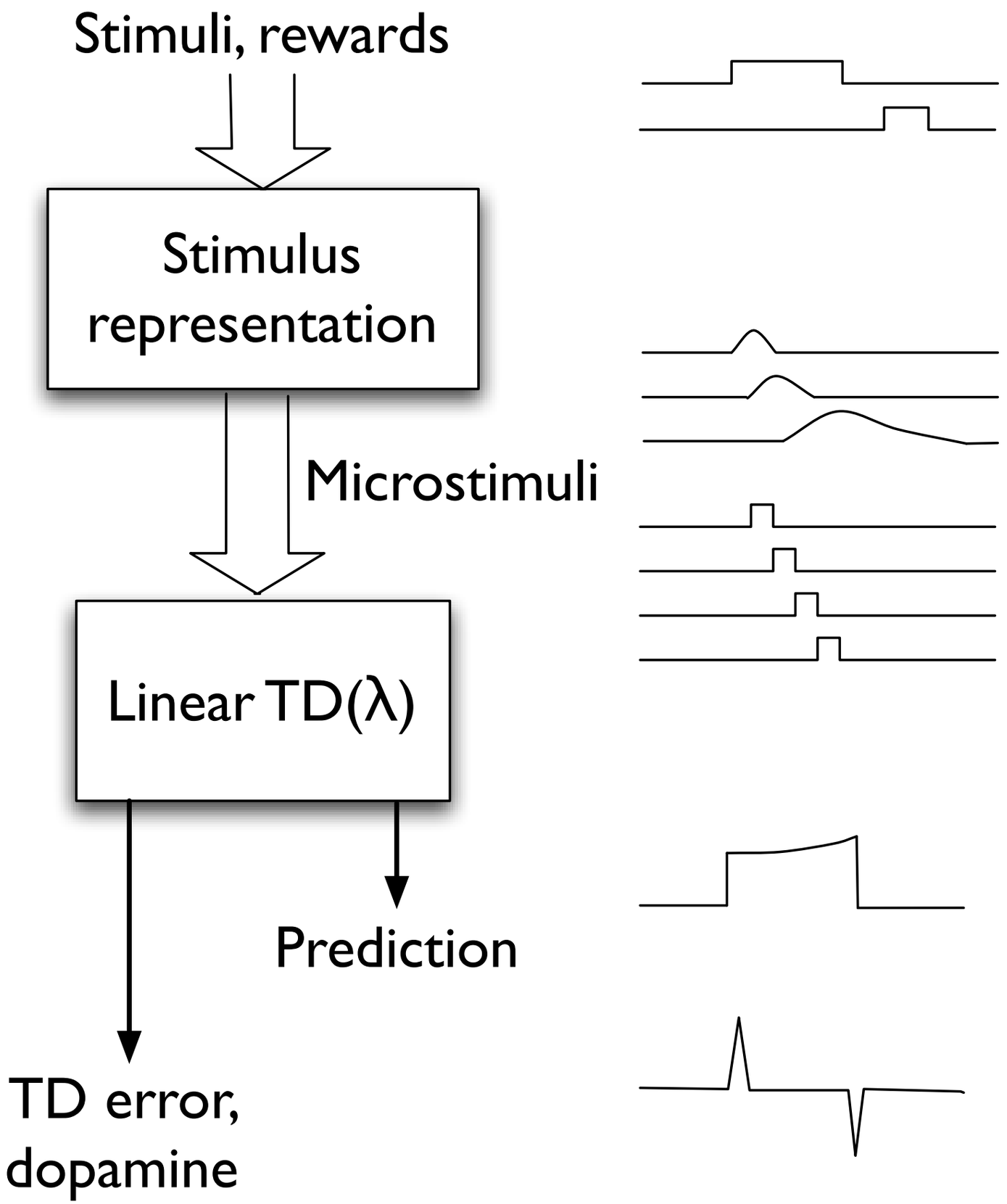
Schultz et al., 1997

Problems with current TD models of dopamine

- Implausible clock-like mechanism
- Poor handling of variations in reward timing
- Predicts large negative error on reward omission
- Needs large (non-physiological) negative errors
- Complexity
- Changes in learning algorithm

Daw, 2006
Niv et al., 2005
Bayer & Glimcher, 2005

TD models


Stimuli, rewards
Stimulus representation
Microstimuli
Linear TD(λ)
Prediction
TD error, dopamine

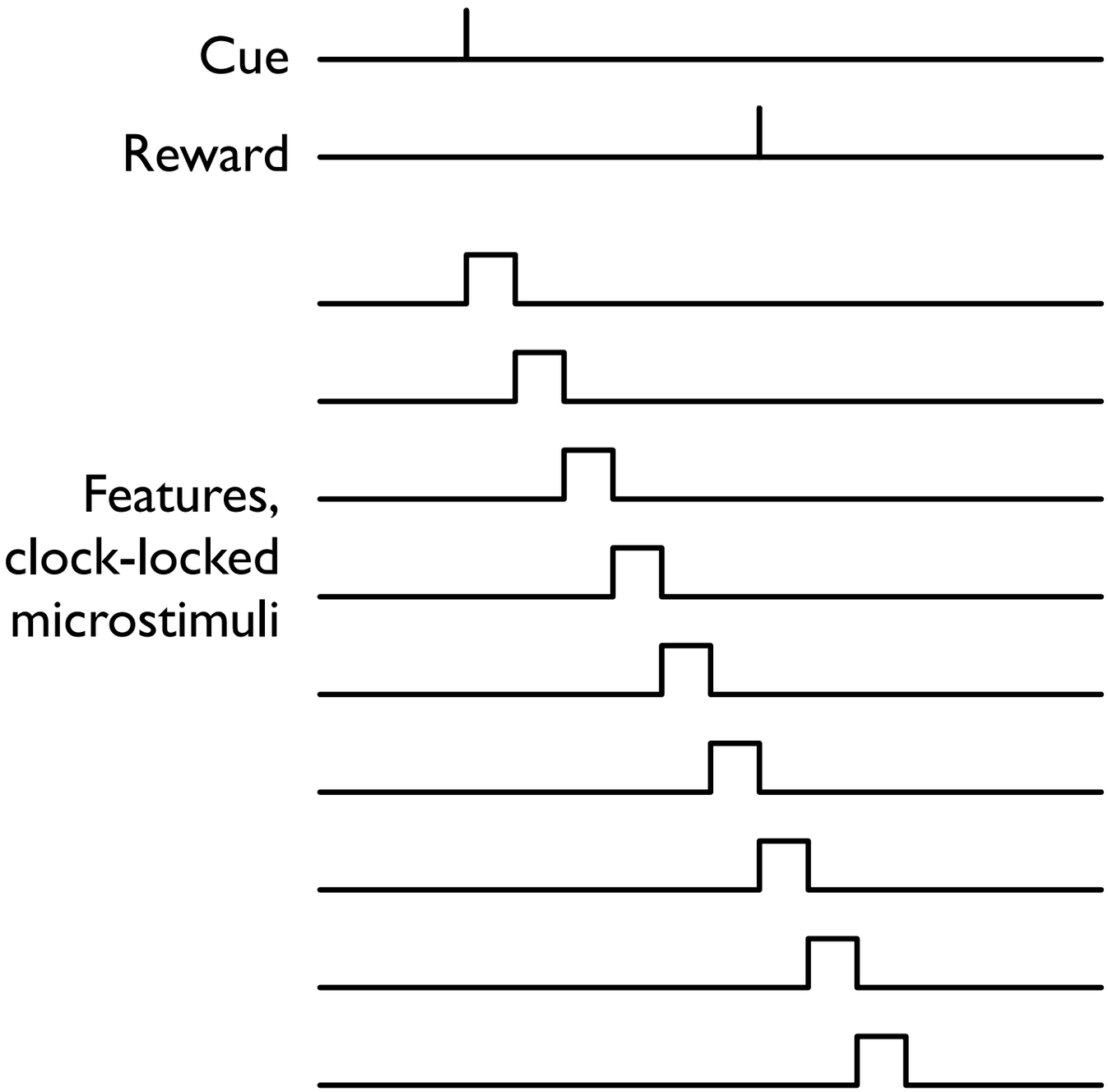
New model

- Temporal generalization via internal microstimulus representation of overt stimuli
- Cueing role for rewards
- Underlying learning algorithm unchanged
  - Retains abilities of previous TD models
- Extended eligibility traces

Stimulus representations

- Trial level
- Real time
  - Presence/absence
  - CSC (current standard TD model)
  - Microstimuli

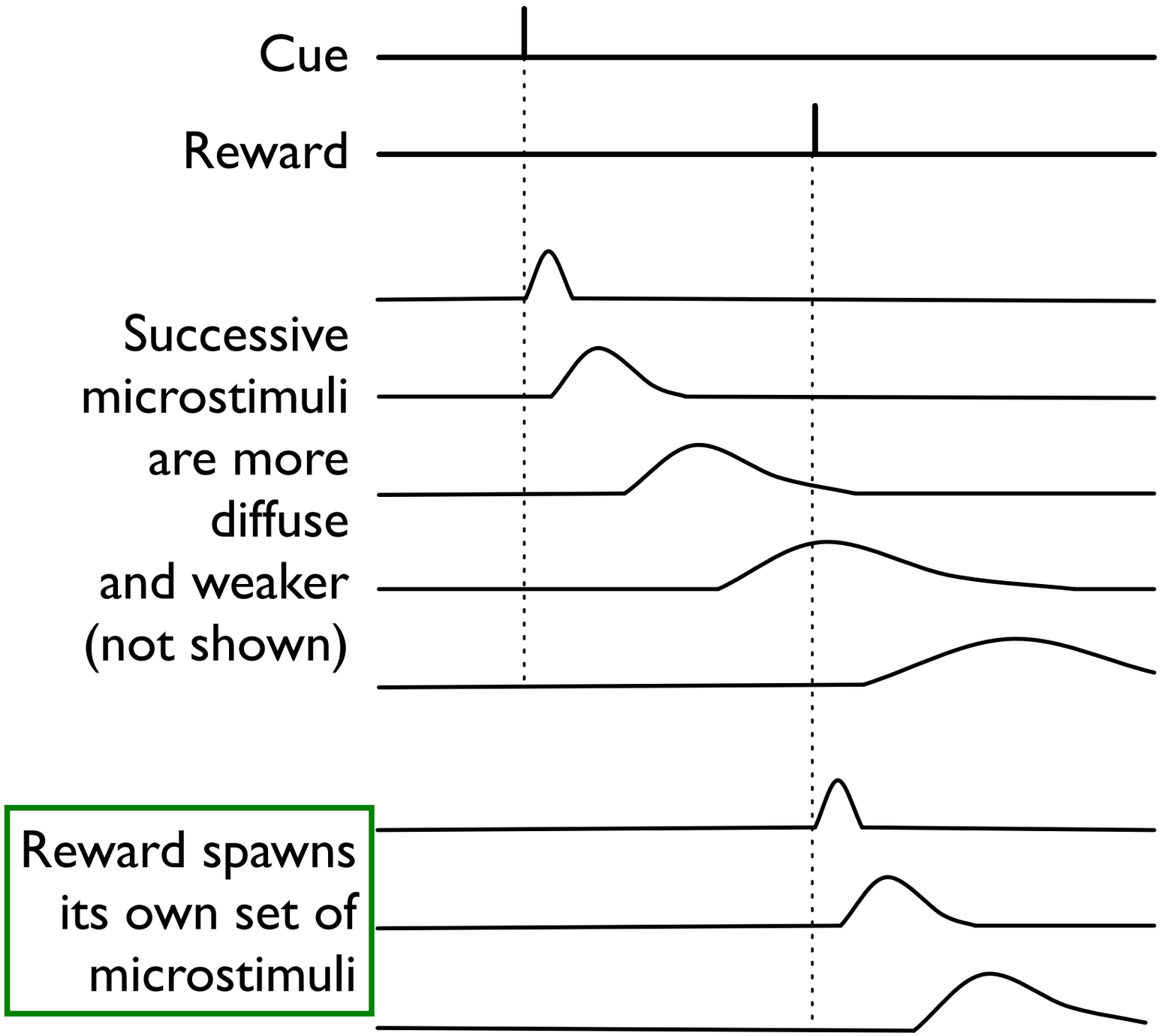
Complete Serial Compound (CSC) stimulus rep'n



No generalization between time instants

No independent effect of reward

Temporally-extended microstimuli

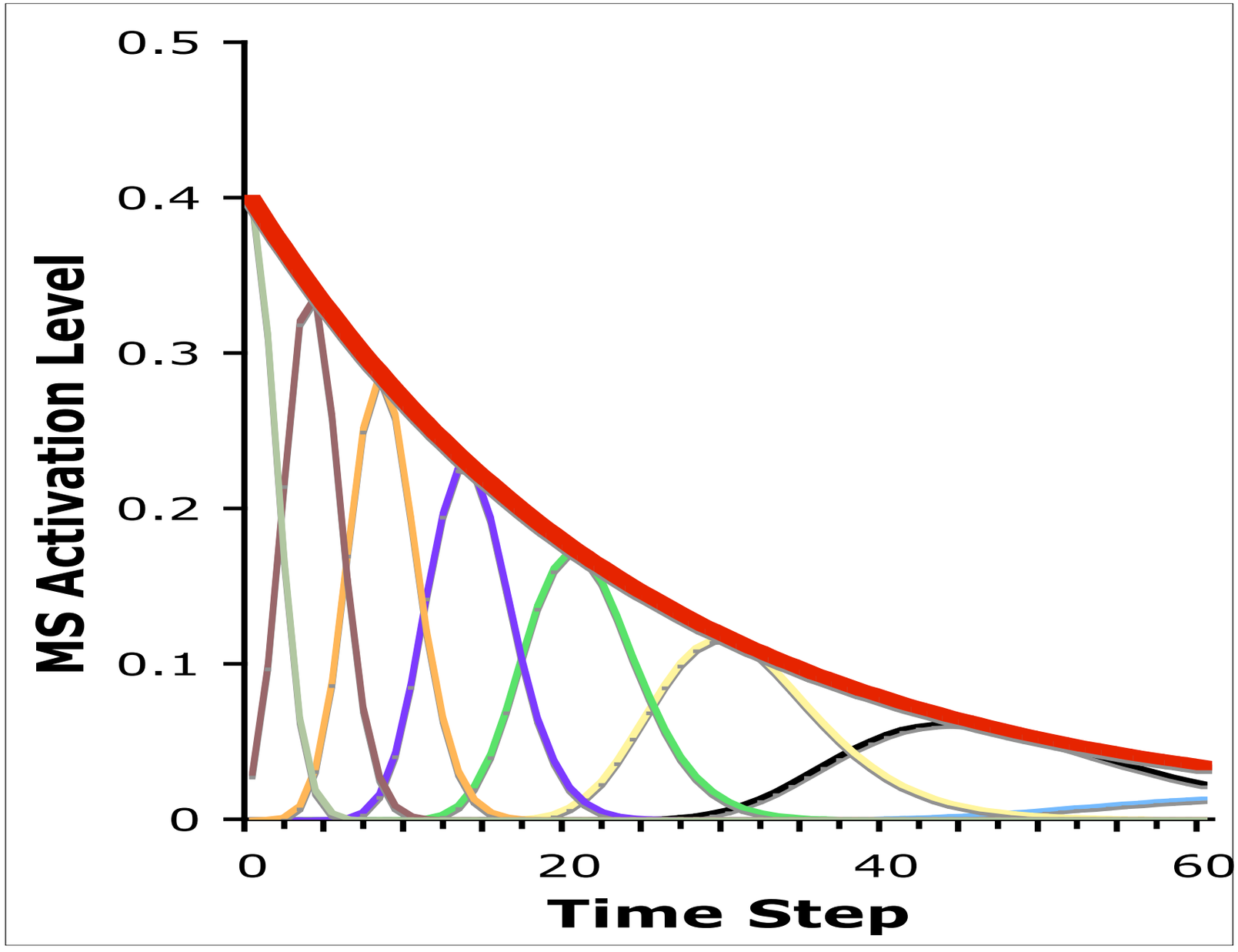


Generalization across nearby time instants

Temporal uncertainty

cf. Machado, 1997
Grossberg & Schmayuk, 1989
Suri & Schultz, 1999

Successive microstimuli get weaker



you can understand everything in this talk at this level:

- Linear TD models can only add things to produce their predictions
  - therefore, stimulus representations determine what can be learned
- TD(λ) is a magic thing that wants to predict the discounted future reward
- Dopamine is TD error is Reward + ΔPrediction

- Linear models can only add things

$$\text{Prediction} = \mathbf{w}_t^T \mathbf{x}_t = \sum_{i=1}^{n} w_t(i) x_t(i)$$

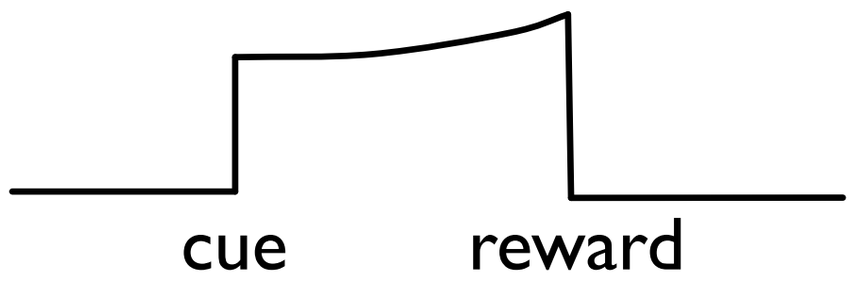
microstimuli

weights

- But they want to predict discounted reward

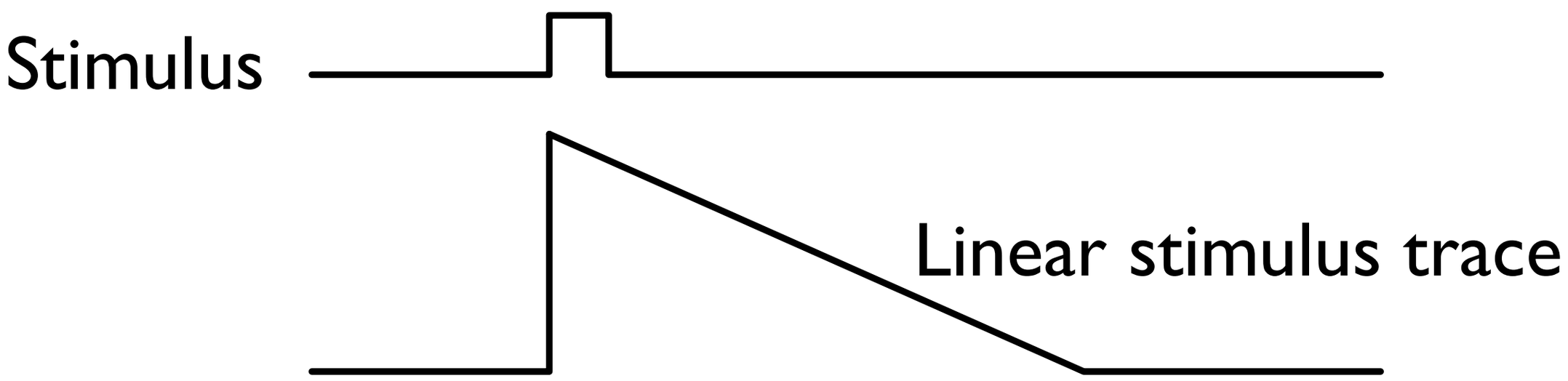
$$\text{Ideal prediction} = E\left[\sum_{k=1}^{\infty} \gamma^{k-1} r_{t+k}\right]$$

Ideal prediction profile



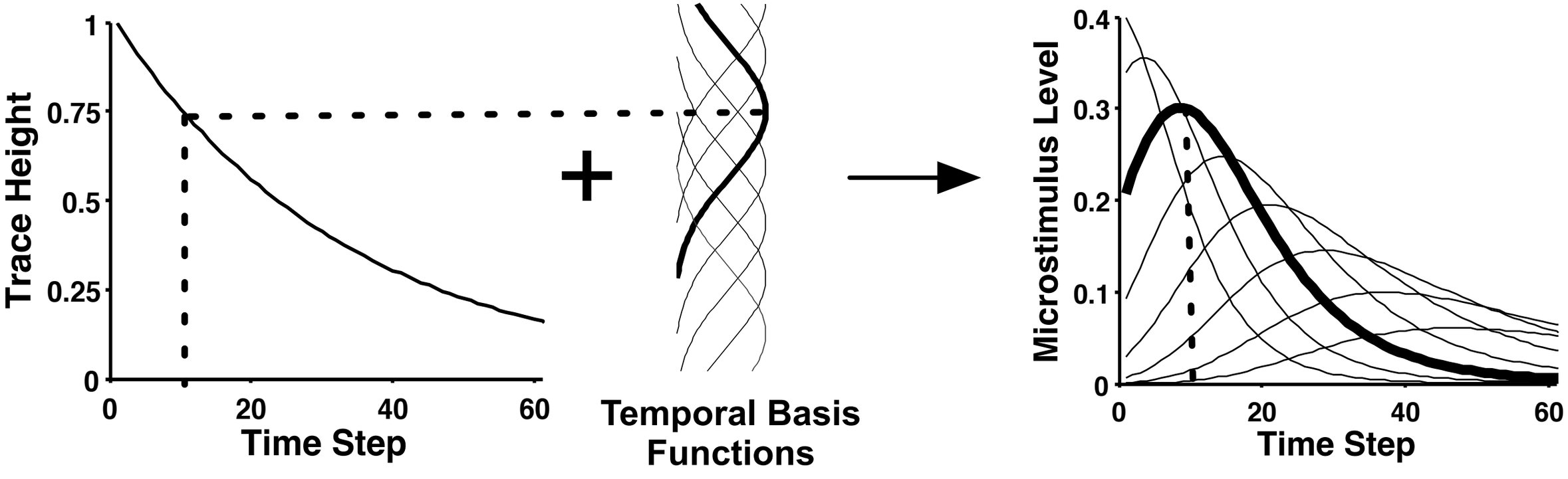
Making microstimuli with stimulus traces (1)

- Start with classical *stimulus trace* (Ebbinghaus, 1888)
  - Stimuli leave behind a short term (seconds) representation of themselves

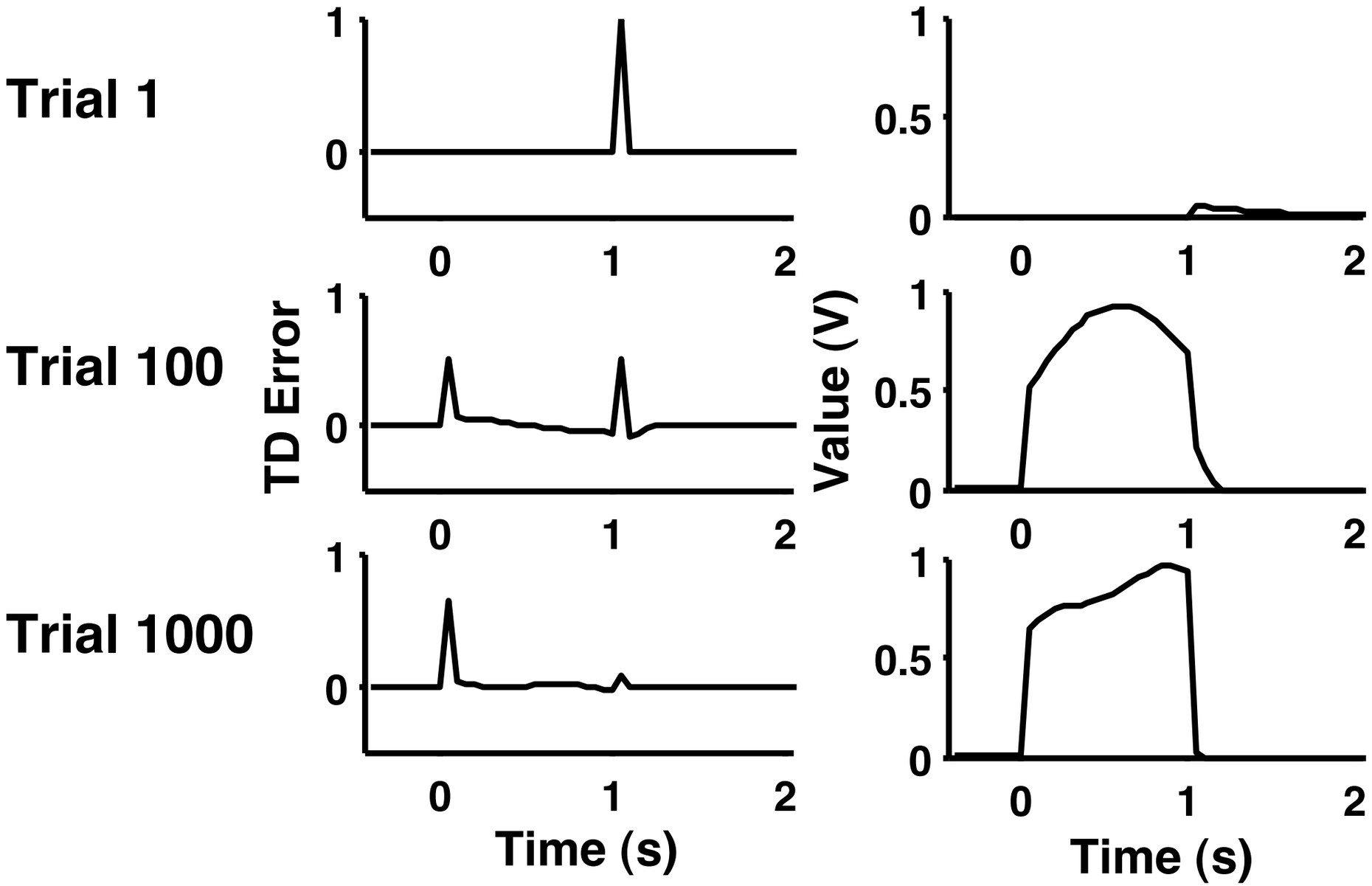


Making microstimuli with stimulus traces (2)

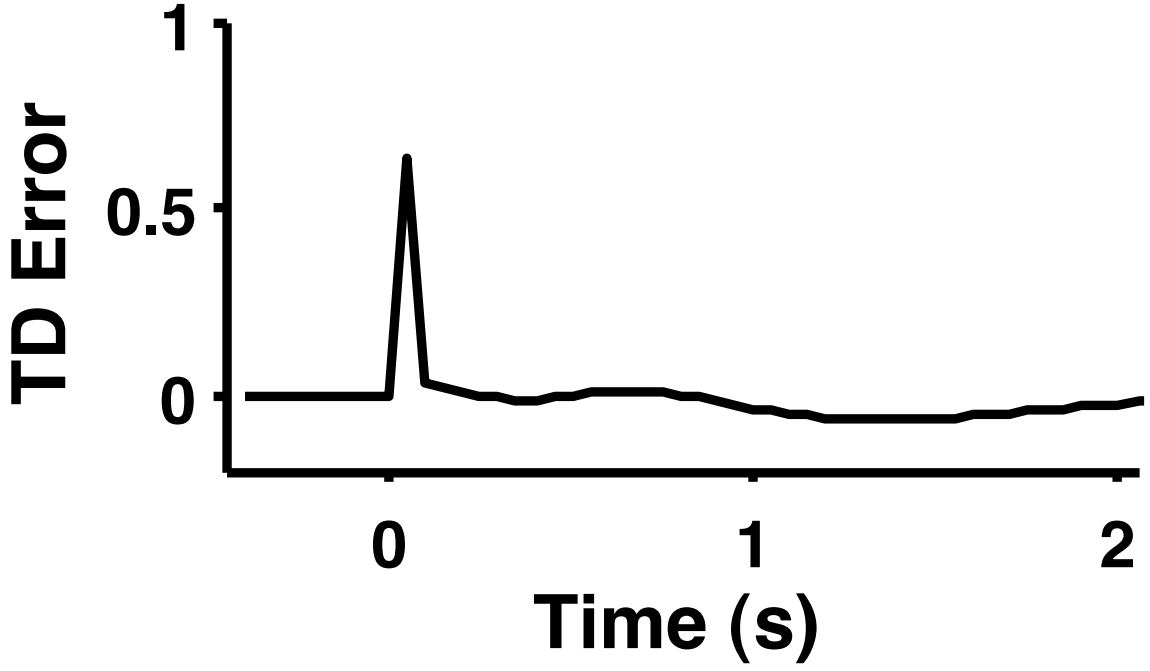
- Microstimuli represent the trace's height (coarsely)

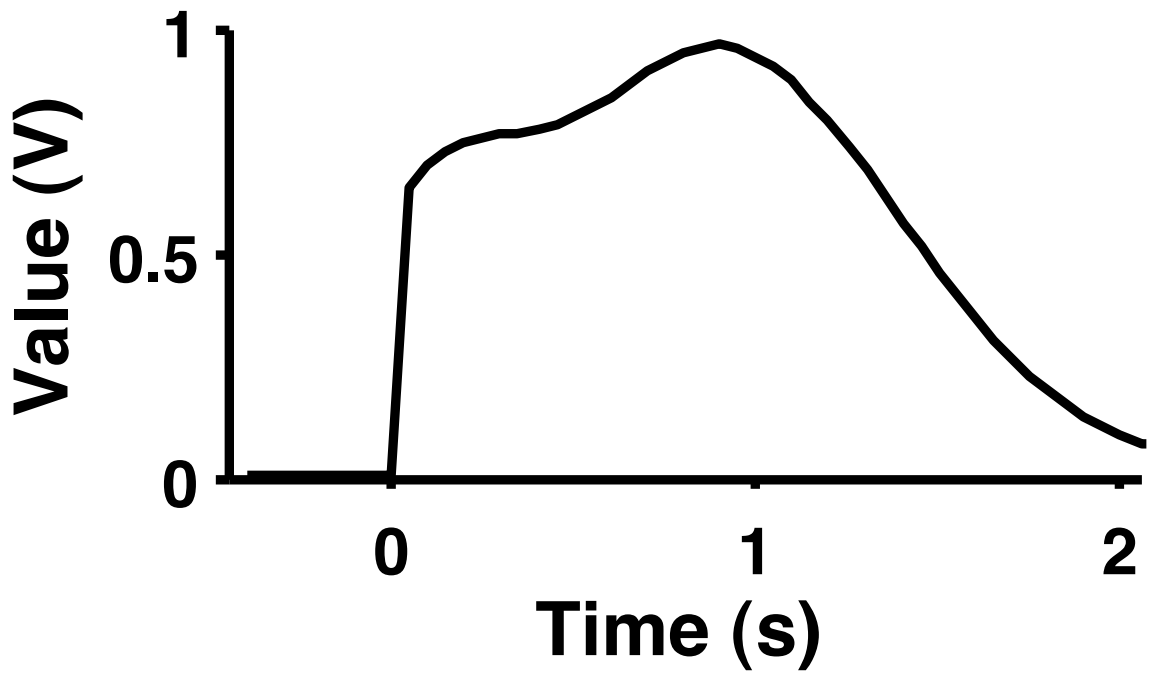


MS model - acquisition

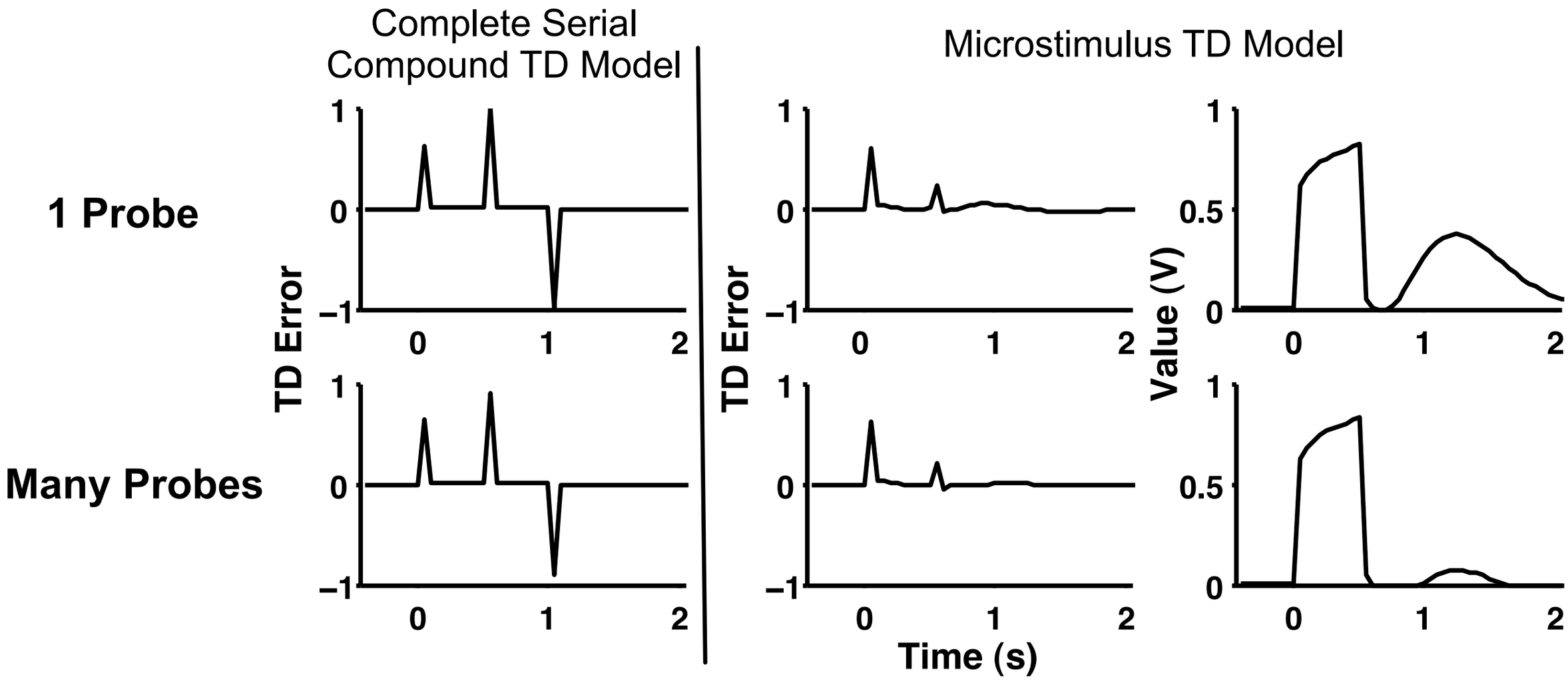


MS model - omission





MS model - reward early



Hollerman and Schultz, 1998

Conclusion regards microstimuli + reward cues

- Benefits
  - more realistic, plausible, natural
    - handles variations in reward timing better
    - does not produce large negative TD errors
- Not tweaks to the TD model, not extensions
  - should be thought of not as adding something, but as taking away two artificial assumptions

Ongoing work

- Microstimuli + presence/absence
- Assessing effect of extended eligibility trace
- Response generation in classical conditioning
- Experimental work with Jim Kehoe, UNSW
  - Rabbit NMR
  - Detailed timing effects