Learning and the partial observability of continuous time

François Rivest

April 7, 2010

Barbados

Plan

- Context & motivation
- What's particular about time?
- Traditional temporal representations
- Learning a temporal representation under reinforcement learning: A brain model (Rivest & al. J.C.N., 2009)
- Problems with recurrent neural networks and temporal series prediction
- Are we asking the right question?
- Basis of a new approach (if time)
- Take home message

Motivation

- How is the brain developing/constructing representations under reinforcement learning?
- While the dynamic of the environment is important, “models” often avoid time!

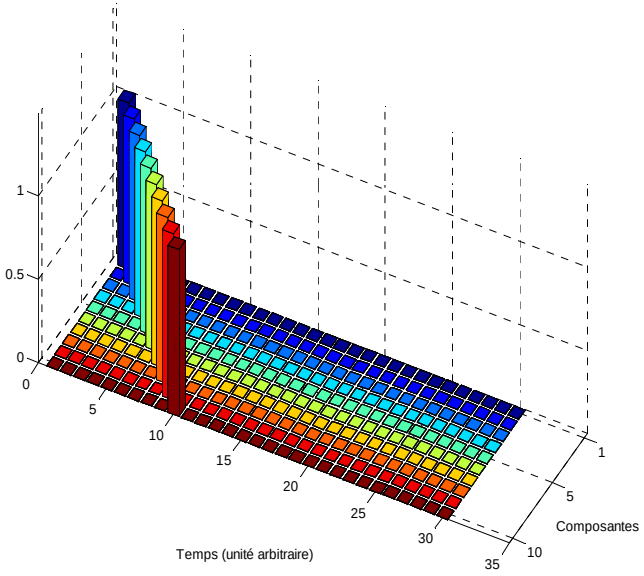
- Goal: *Modeling* the learning of temporal representations (environment dynamics representations) under reinforcement learning.

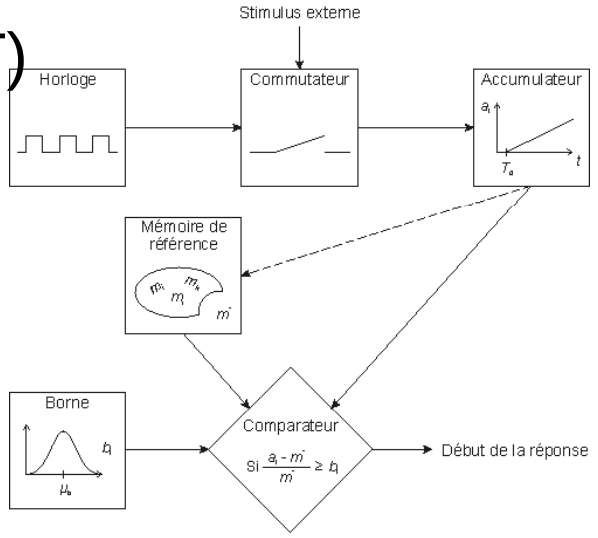
What's particular about time?

- Time is only partially observable
  - There is no time sensor in the multi-seconds range!
  - (Unless you have a watch!)
- Time is never directly observable
  - Time is a constant!
  - (Unless you have a watch!)
- Yet, the timing of events shapes our responses
  - If there is a temporal relationship, you will learn it!
  - (Note that temporal relationship is neither necessary nor sufficient to learn relationships in general.)

Traditional temporal representations

- Use of delay lines
  - Unrealistic for delays in the order of seconds (time-delay networks)
- Providing a full semi-markov model of the task.
  - This is what we would like to learn automatically
- Using a clock & accumulator (SET)
  - When should we start the accumulator, is there a clock?
- ***Maybe temporal representation can be learned…***

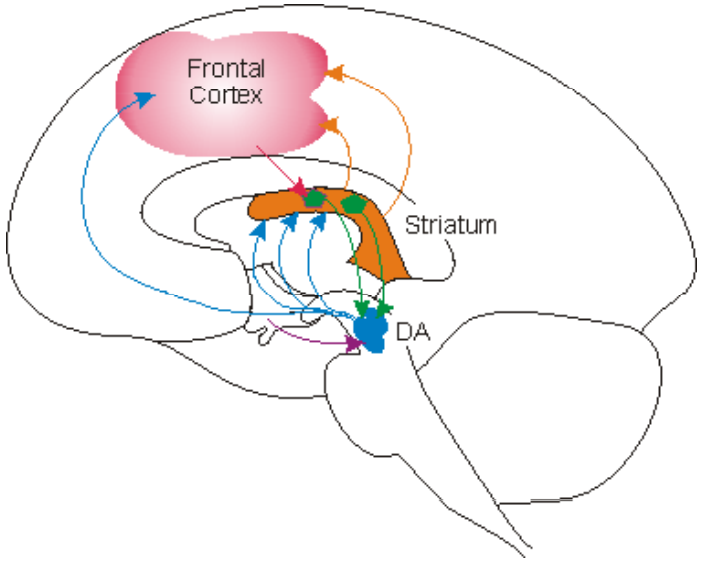




Learning a temporal representation under reinforcement learning: A brain model

Rivest, Kalaska, & Bengio (2009) Alternative Time Representation in Dopamine Models. *J. Comp. Neurosci.*

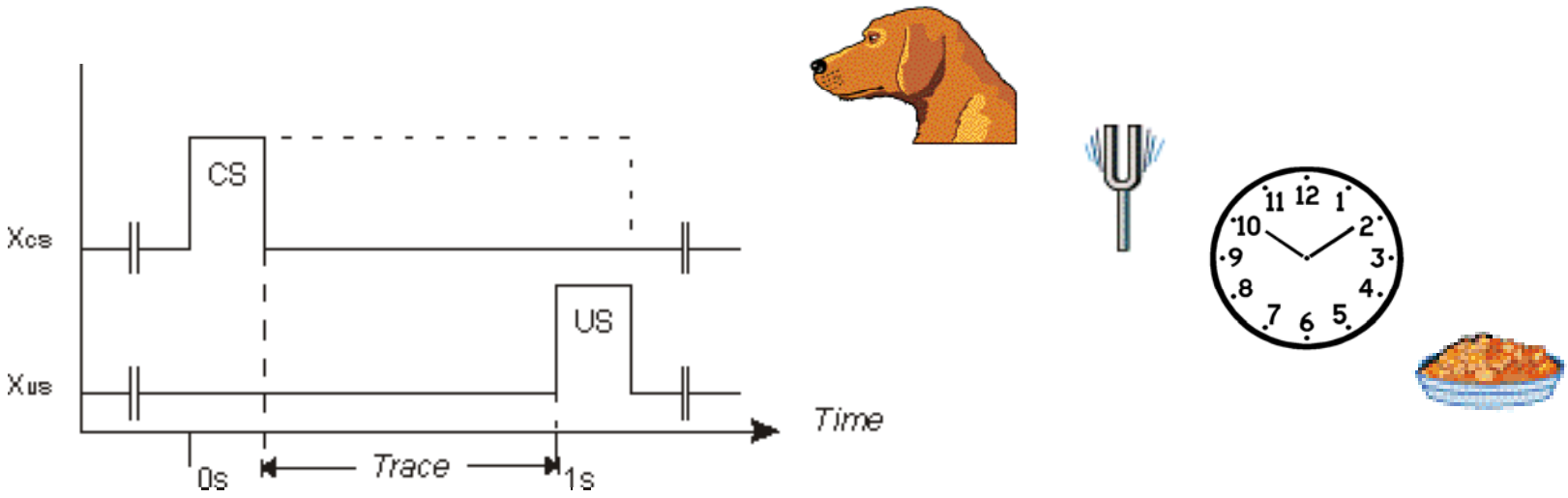
Goals and Hypothesis



- Hypothesis:
  - If the cortex learns the environment dynamics (unsupervised), then the resulting representation could be sufficient, in conjunction with TD, to explain the observed dopaminergic data.
- Other goals:
  - Develop a model that could help understanding how time could be represented within an artificial neural working memory.
  - Evaluate hypothesis about the cortico-basal interactions, such as the possible role of dopamine in cortical learning.

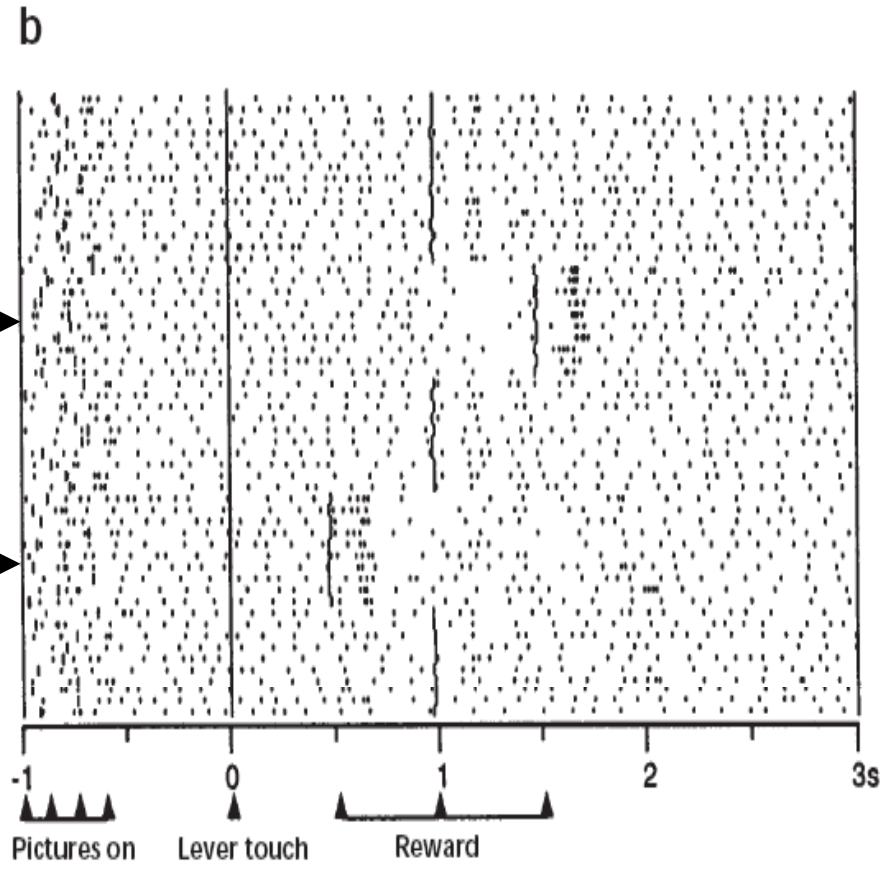
Learning a temporal representation

- Looking for the simplest problem in which a temporal representation is learned.
- Classical/trace conditioning with a fixed ISI.



Dopaminergic neurons evidences

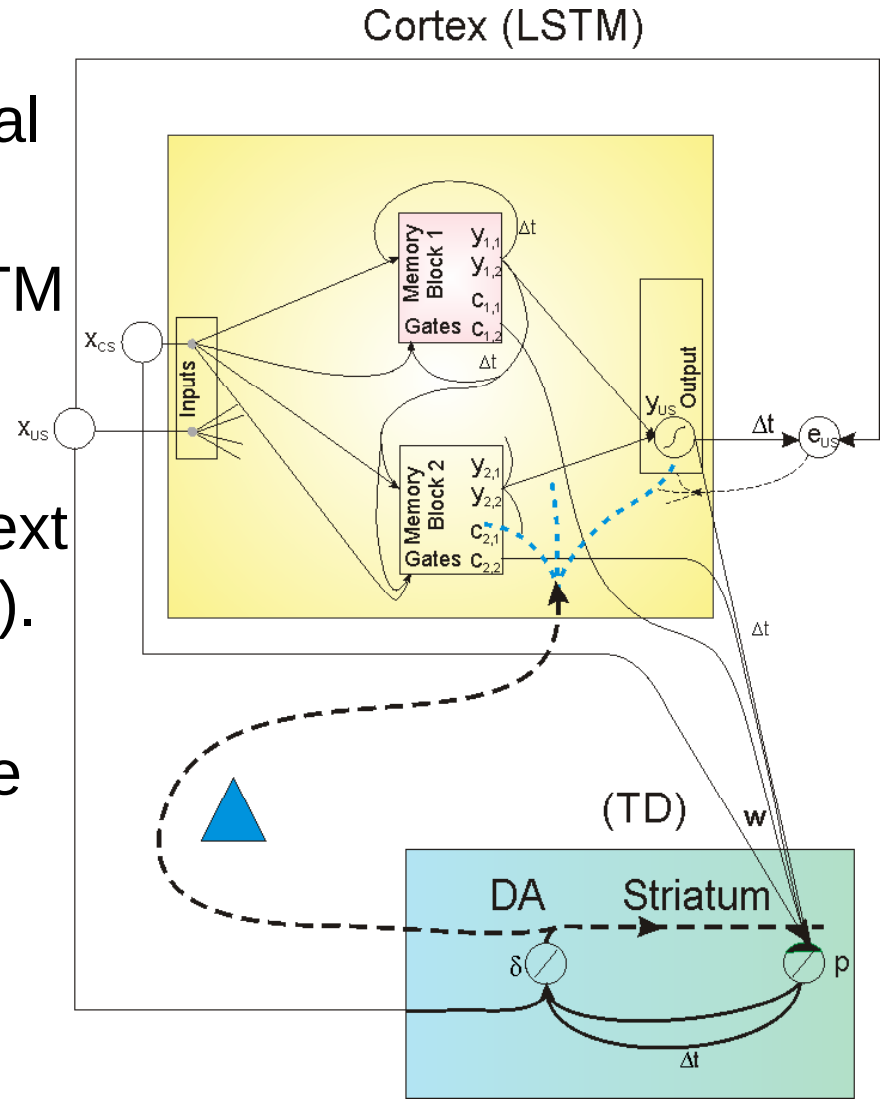
- Dopaminergic neurons show that temporal knowledge of the task exists in the brain.
  - TD error ≈ dopaminergic phasic signal
- Dopaminergic activity after conditioning:
  - When reward is given late (2nd)
  - When reward is given early (4th)



Hollerman & Schultz (1998)

The model

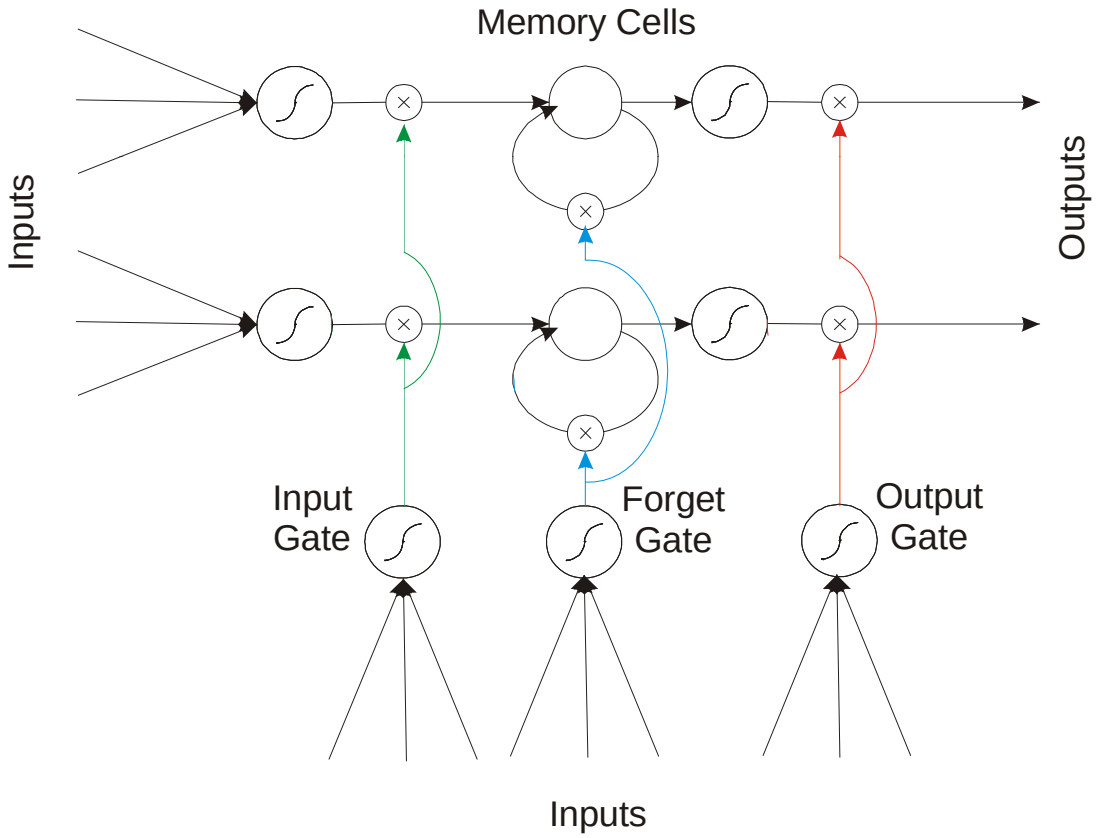
- Unsupervised recurrent neural network to model the cortex.
- Long short term memory LSTM as a working memory model.
  - (Hochreiter & Schmidhuber 97)
- LSTM learns to predict the next inputs (next observable state).
- LSTM internal activities (cortex) serve as inputs to the RL system (TD, in the basal ganglia).



Working Memory (memory block)

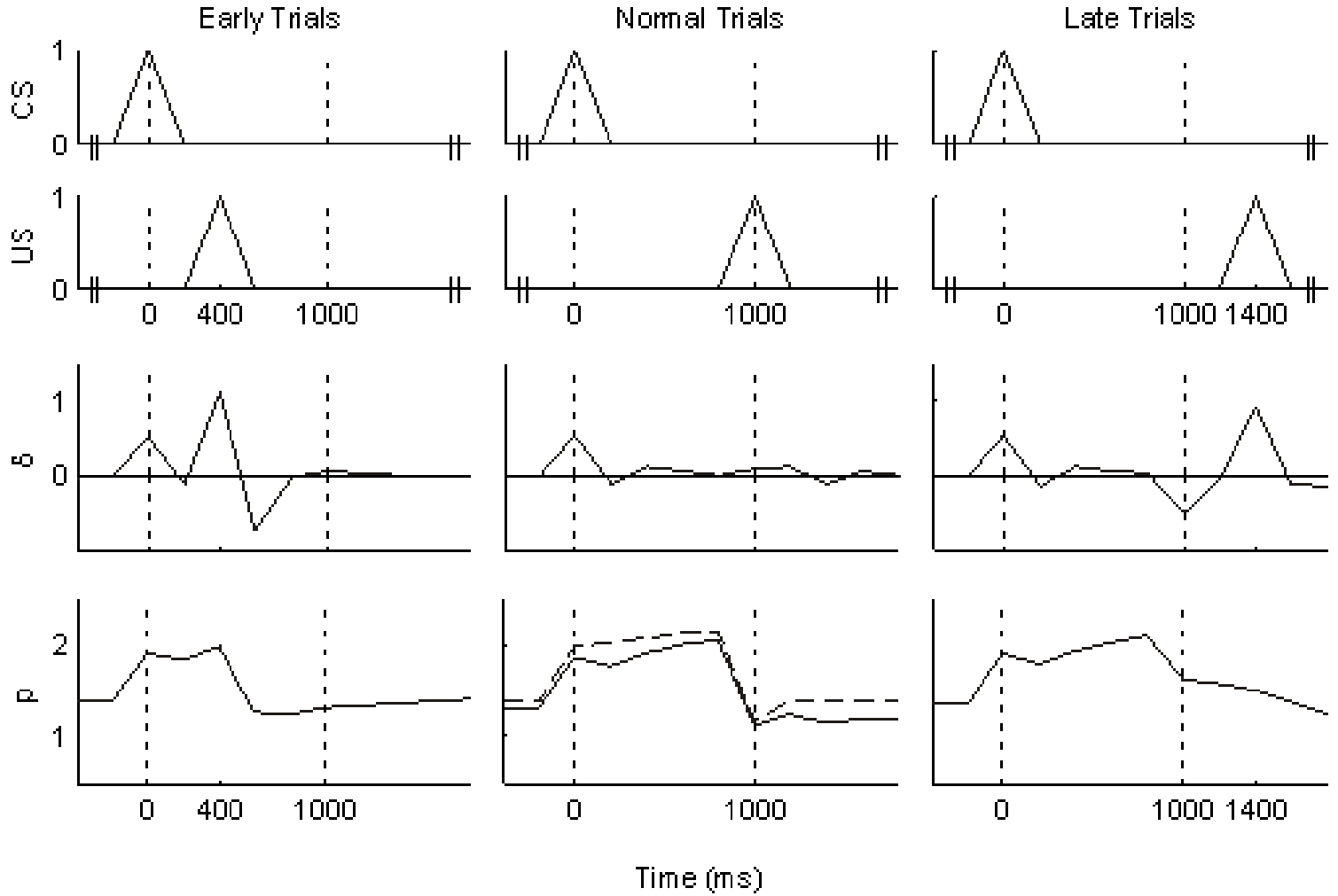
- Frontal cortex is often considered to implement working memory
- Working memory must contain some form of gating

- Trained using backprop.
- The linear units allow error to back-propagate error further in time.
- The algorithm takes O(1) time and memory.



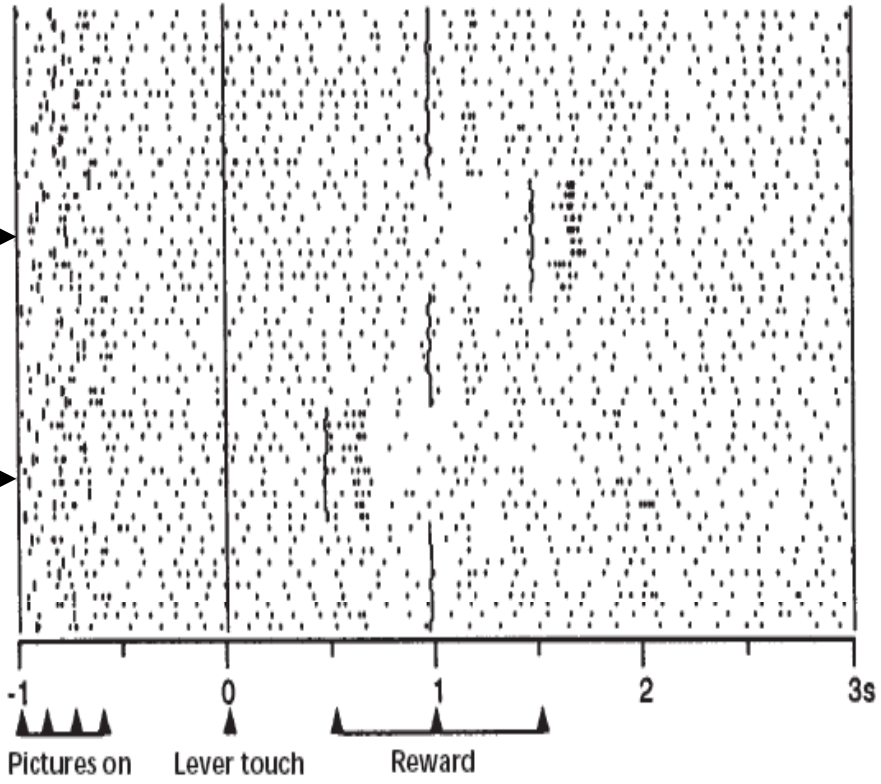
Dopaminergic data

- The model reproduces dopaminergic activity



Dopaminergic neurons evidences

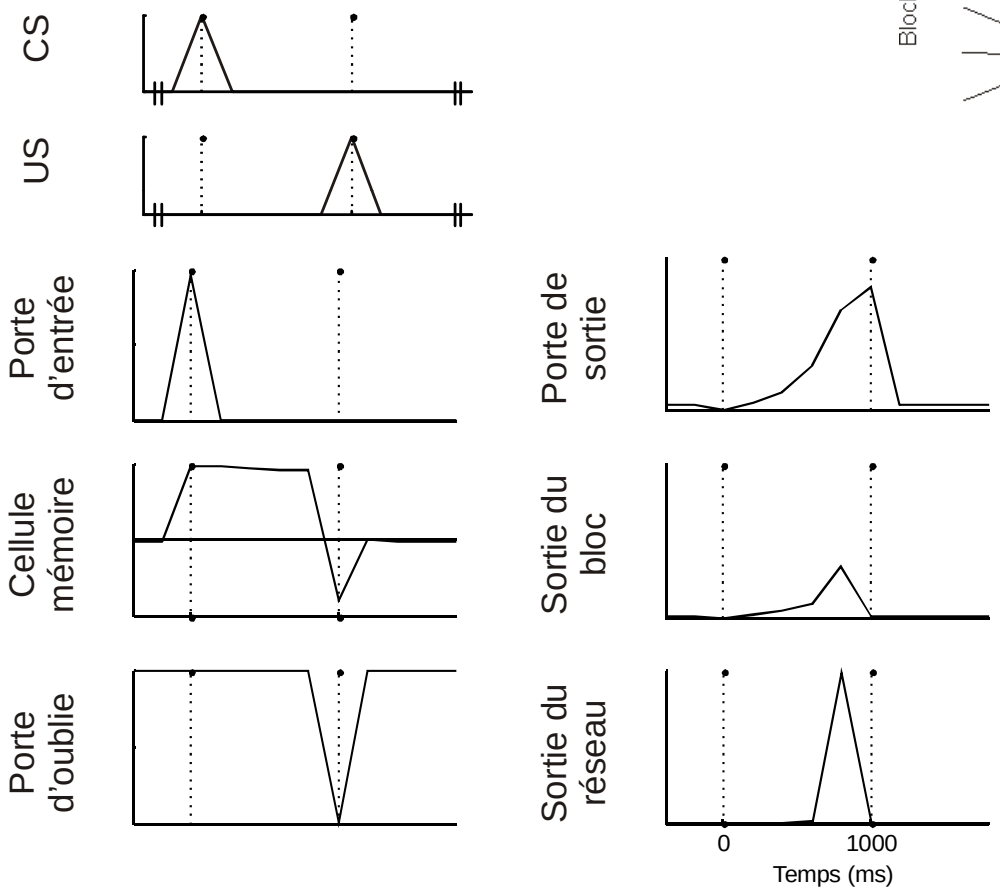
- Dopaminergic neurons show that temporal knowledge of the task exists in the brain.
  - TD error ≈ dopaminergic phasic signal
- Dopaminergic activity after conditioning:
  - When reward is given late (2nd)
  - When reward is given early (4th)

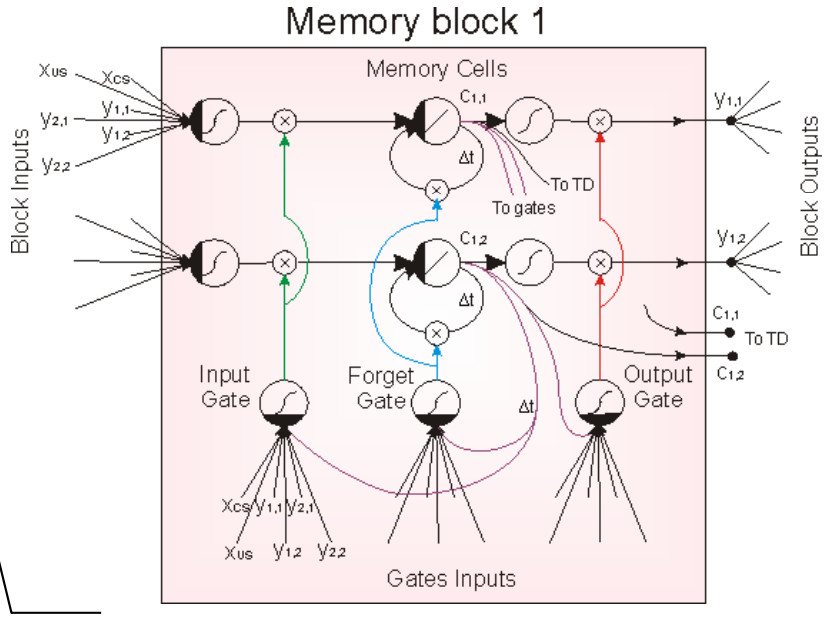


Hollerman & Schultz (1998)

Time representation in LSTM

Trace networks (1)



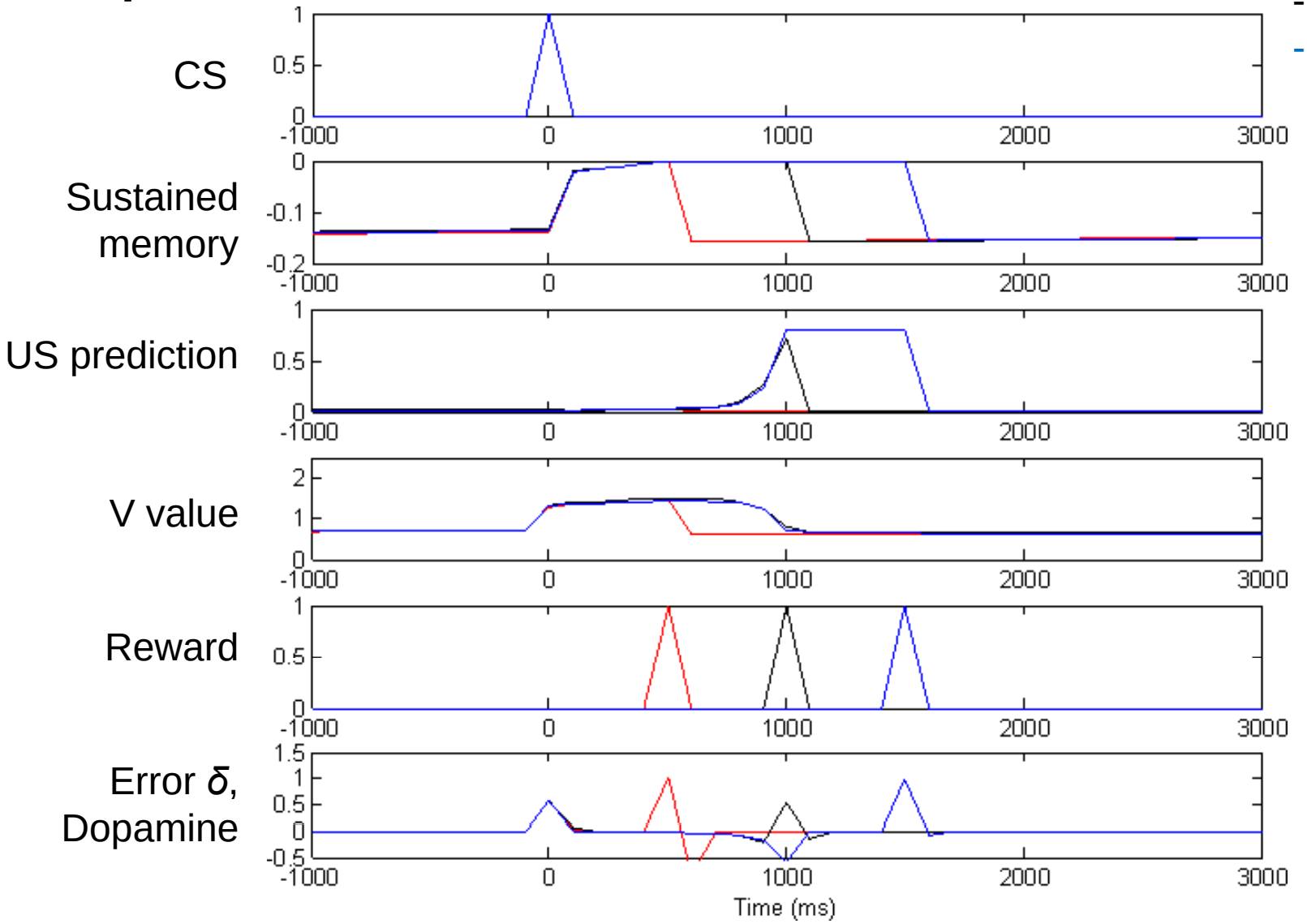


- Training history effect (curriculm)

- Classical versus trace conditioning

Representation and TD error

-Early
-Normal
-Late



Summary

- In agreement with the hypothesis:
  - If the cortex learns the environment dynamics (unsupervised), then the resulting representation could be sufficient, in conjunction with TD, to explain the observed dopaminergic data.
- Other goals:
  - Develop a model that could help understand how temporal representation can develop within working memory.
    - In trace conditioning, the working memory is used to remember that a CS was observed and therefore, that a US is to be expected.
    - In delay conditioning, linear build-up of activity in memory cell.
    - The model also shows how each signal is combined to form the prediction V.
  - Evaluate the possible roles of dopamine in cortical learning.
    - We showed that the error signal could be used to speed up cortical learning!

Problems with recurrent neural networks and temporal series prediction

- They have a finite amount of working memory, so it seems hard to learn long-term dependencies without knowing what to kept in working memory
- Slow learner
  - Gradient descent
  - Sigmoidal activation
  - Time-step dependence
  - They are not a nice adaptive dynamical system!
- Don’t deal with timing error, they deal with output error

Are we asking the right question?

- Most current learning algorithms are asking what?
  - What event should append at time *t*
  - What value should we predict for time *t*
  - Etc…
- A different and important question might be when!
  - When should event *x* append?
  - At which rate should reward *r* come?
  - Etc…

Useful properties to seek

- That timing could be learned while learning an association, not only after
- That learning timing would require no more than a linear (or constant) amount of time and memory with respect to the interval time length to learn
- That timing precision could be proportional with the interval length (Weber's law for time)
- That one can correct for timing error, not only for action error (adapting the timing of action, not just which action)

Take home messages

- Learning in the brain is not a single learning rule.
  - The brain great learning ability emerges from the interaction of different systems specialized in different aspects of learning.
- Time is an important factor.
  - The brain is slow and the world is not a discrete sequence of events (it has a dynamic).
  - There is an enormous amount of constancy between *t* and *t+1.*
  - Knowing the environment dynamics allows one to filter out a lot of predictable information.
  - The brain rarely takes a decision based on a single snap-shot of the inputs. It accumulate evidences using attention and time.

Acknowledgements

- Organisers:
  - Rich Sutton
  - Doina Precup
  - Elliot Ludvig
- Collaborators:
  - Yoshua Bengio (UMontréal)
  - John F. Kalaska (UMontréal)
  - Doina Precup (McGill)
  - Thomas R. Shultz (McGill)
  - Frédéric Dandurand (CNRS)
  - Marc Bellemare (UAlberta)
  - Elliot Ludvig (UAlberta)
- Funding: