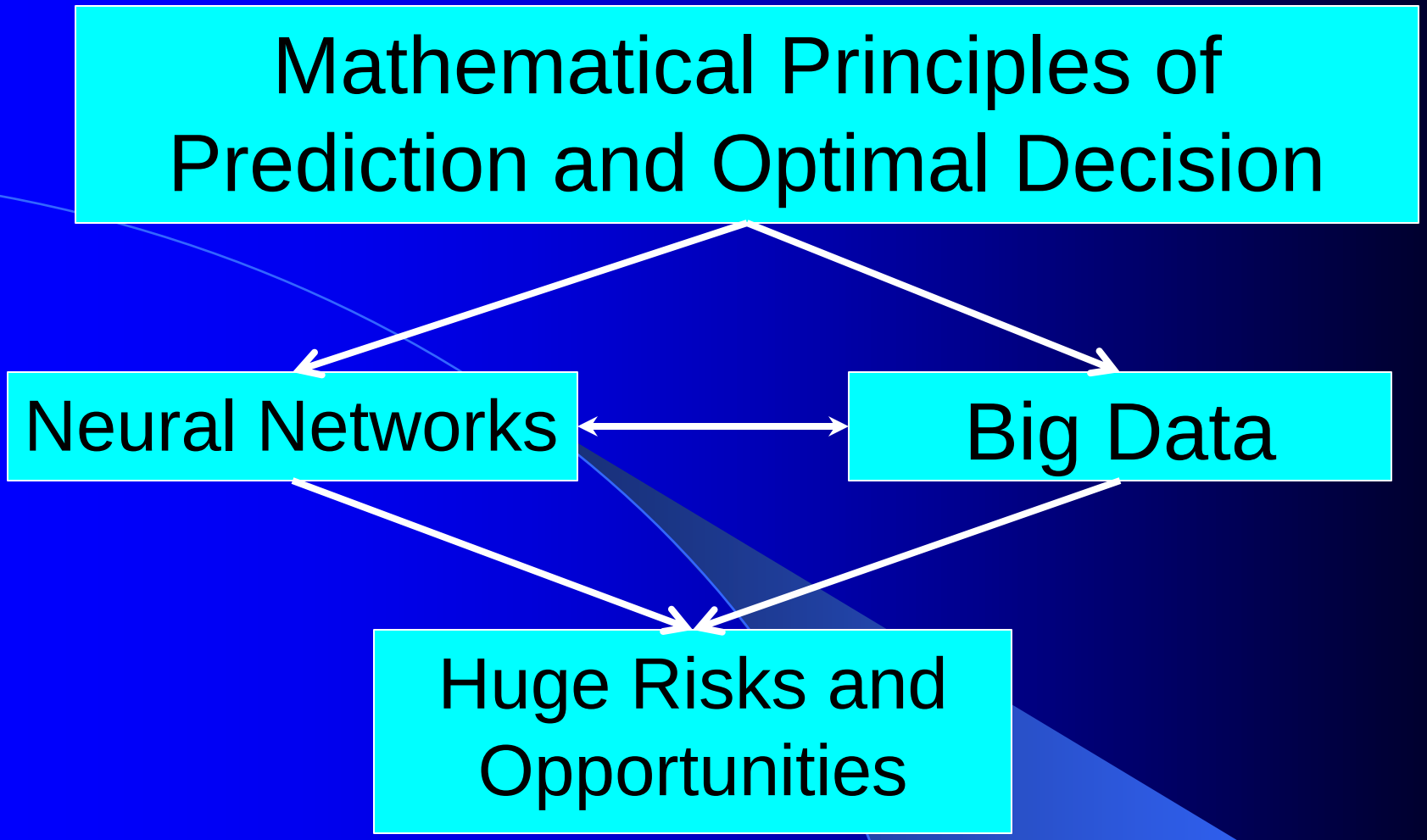
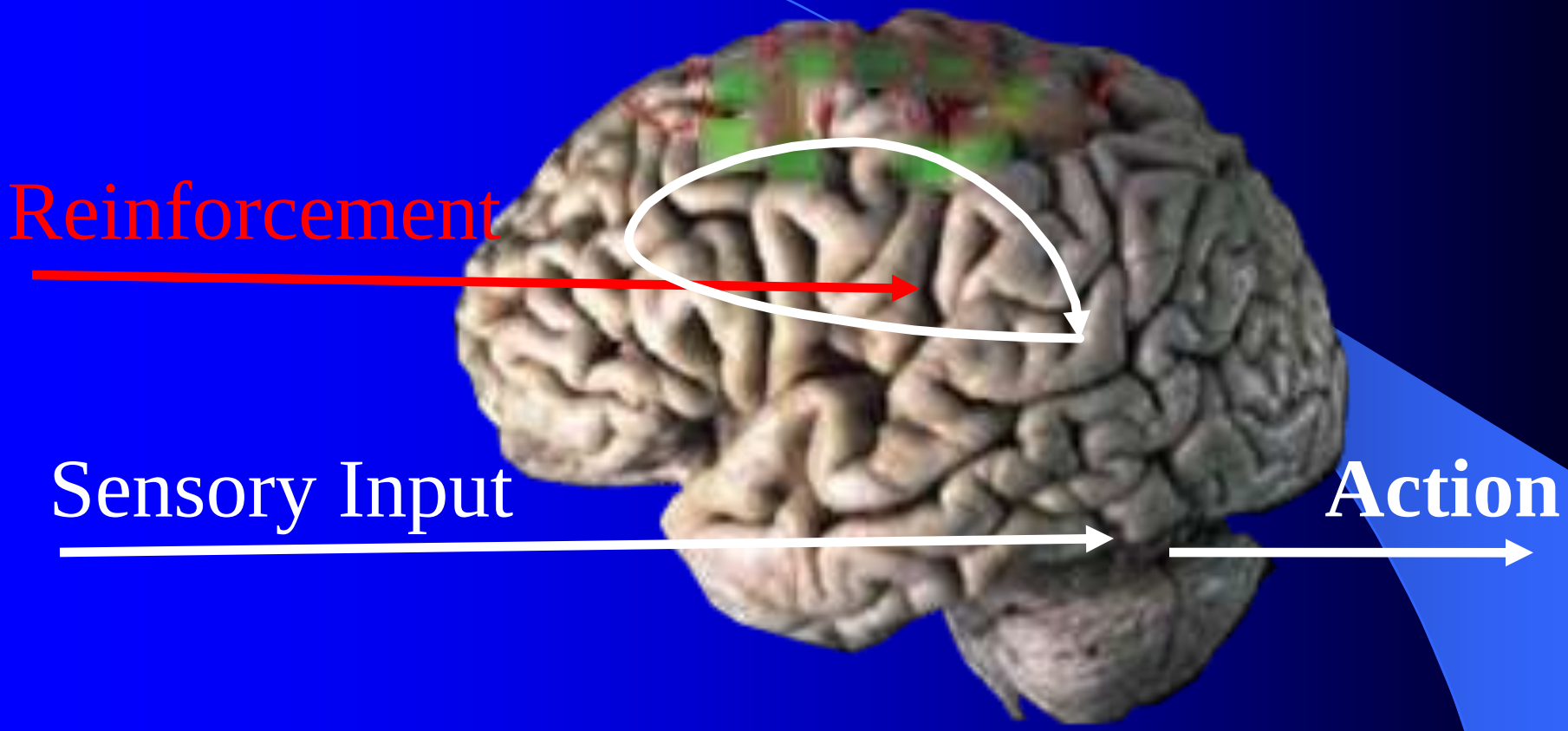


(1) Bigger Risks and Opportunities
(2) Mathematics of Prediction
WATCH THE URLs for Details and Optimal Decision!

Never forget this existence proof!



Brain As Whole System Is an Intelligent Controller

-- Mouse maximize probability of survival among other things

-- Lots of animal behavior research

-- Lots of recent motor control research (UCSD…)

5 Grand Challenges for Adaptive and Intelligent Systems – General-purpose massively parallel designs to learn….

Pr(A|B)=Pr(B|A)* Pr(A)/Pr(B)

COPN 2008

Prediction

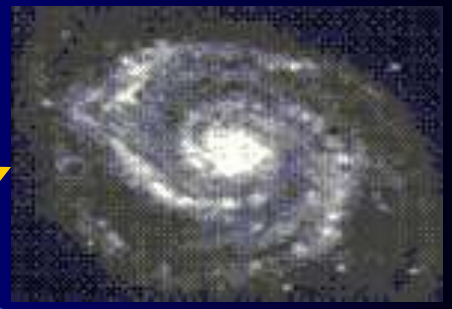
Memory

. . .

Clustering

Optimization

J(t)=Max<J(t+1)+U>



$$\frac{\partial^{+} z_n}{\partial z_i} = \frac{\partial z_n}{\partial z_i} + \sum_{j=i+1}^{n-1} \frac{\partial^{+} z_n}{\partial z_j} \frac{\partial z_j}{\partial z_i}$$

About 10 Independent Systems Operators (ISOs) Run the US Grid

ISOs Decide:

- Unit Commitment (Contracts to Generators a Day Ahead) & Advance Planning
- Economic Dispatch (Generators Used and Loads & Prices every 15 minutes)
- Some Regulation Functions (every 2 seconds): stability, V, υ, also linked to old balancing authorities

Transco like IOU

Big Loads

?

Distribution Companies, Discos

Small Loads

See www.ferc.gov, event calendar, June 2010

"NSF is currently supporting research to develop a '4th generation intelligent grid' that would use intelligent system-wide optimization to allow up to 80% of electricity to come from renewable sources and 80% of cars to be pluggable electric vehicles (PEV) without compromising reliability , and at minimum cost to the Nation (Werbos 2011)."

Werbos 2011: IEEE Computational Intelligence Magazine, August 2011

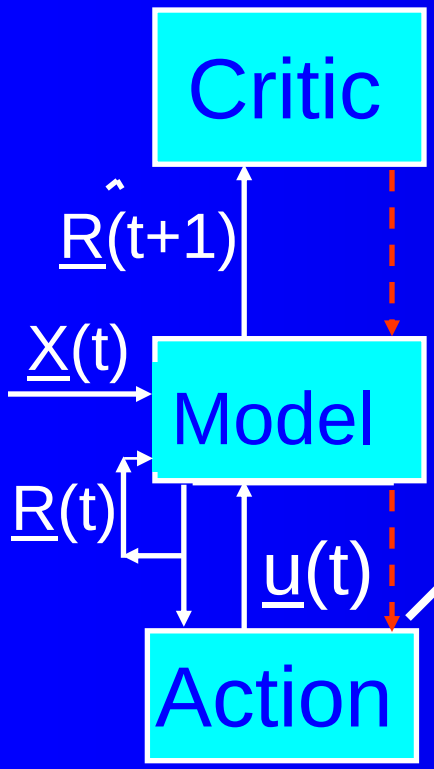
A Grand Challenge : (Control/Decide) Simple How To Manage Thousands of Space Robots to AssembleThis Structure

Links from nss.org/EU:

- **NIAC Report:** New Design for 9¢/kwh if launch costs down to $500/kg-LEO

See review in Ad Astra Summer 2014

RLADP From Vector to Mammal: see http:arxiv.org 2014 MLCI



0. Vector Intelligence – HDP, DHP, GDHP, etc.

Add new spatial complexity logic (ObjectNets +…, Suitable for CNNs)

1. First ever system which learned master class chess Fogel, Proc IEEE 2004

Add ability to make Decisions, plays (Modified Bellman eqs for Multiscale t.)

2. reptile

Add Creativity System (Cognitive map of space of possible decisions)

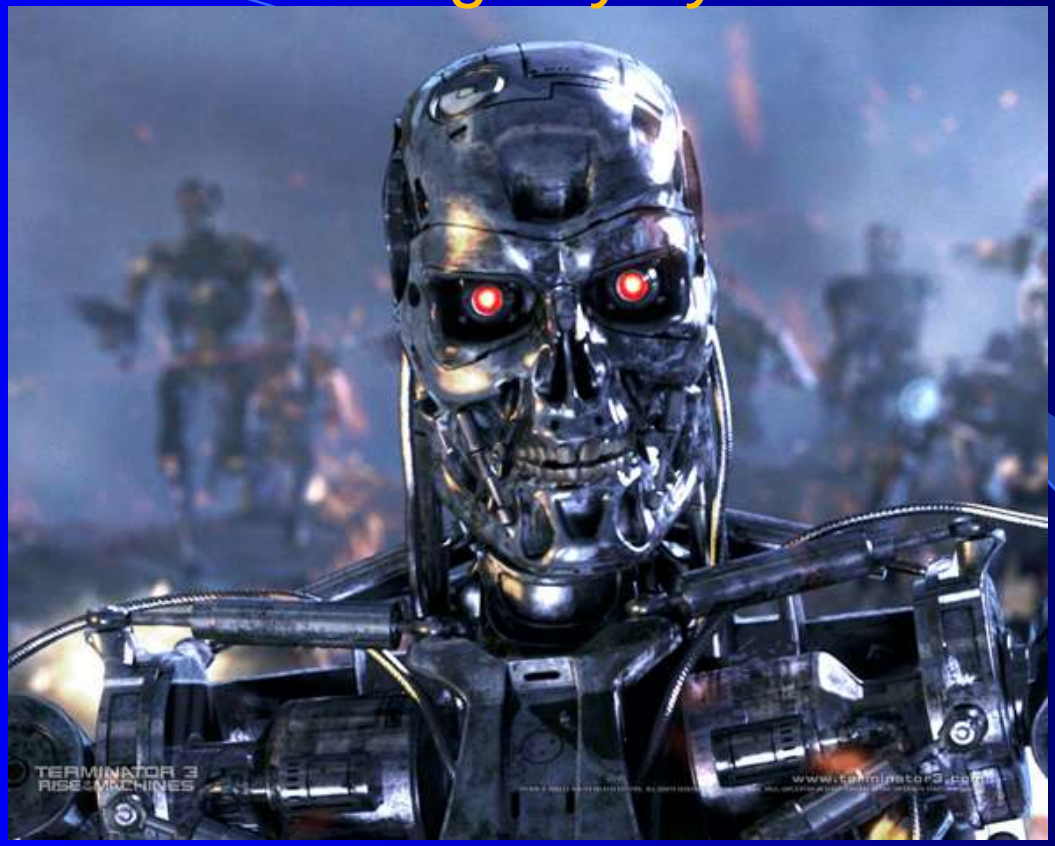
3. Mouse

S.N. Balakrishnan: Using DHP, Reduced Error in Hit to Kill Missile Interception more than order of magnitude vs. all previous methods

* First proven in comparative study by Cottrell for BMDO across hundreds of methods, including his own

See the SPIE slide show link at top of www.werbos.com/Mind.htm

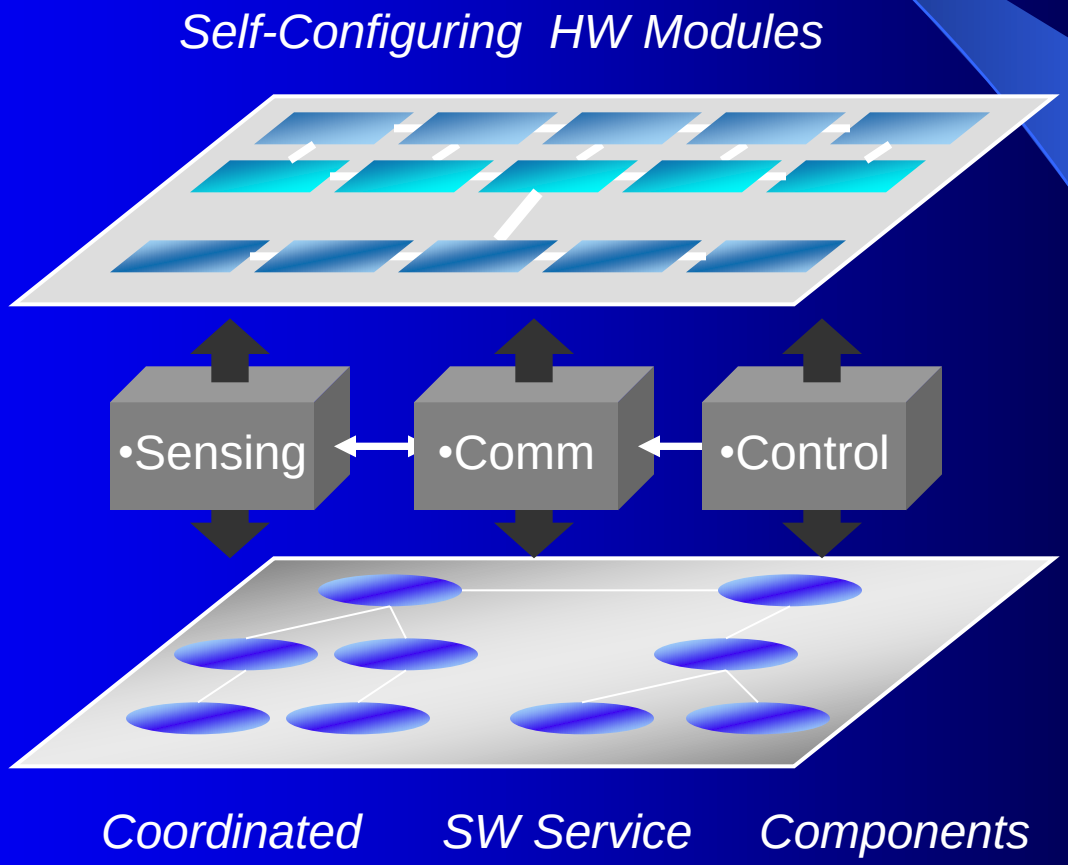
IJCNN91 Seattle: Boeing says you MUST see Terminator



Actual company Cyberdyne/Neurodyne funded by me that week
Bad nano guy a morph between me and Neurodyne
Starts with NN Theater Missile Interception (as in Seattle!)
2 key items – robot arm (award that week) but what of chip?
Schwartzennegger voice – briefing on the ship that week
Movie explains information can be sent backwards through time

NSF 2004: A New Vision Which Later Became Cyberphysical Systems (CPS) and Internet of Things (IOT)

Cyberinfrastructure: The Entire Web From Sensors To Decisions/Actions/Control For Max Performance



“A New Business Plan for IOT. IOT will control every car, every Pacemaker, every household, factory, generator, drone as one system. It must be intelligent and secure, so it should use an expert system like Watson, running US government efficiently from two guarded hard server farms and rest of world from others.”

Q1: “Efficiently? By what metric? What VALUES/UTILITY will this optimize?” A:”Values?? Our programmers can take of that. If any meatheads object, our security can take care of them.”

Q2: “Where are PEOPLE in this IOT?” A: “Easy. We will turn people into things, with BCI we now have …”

Winter Soldier: Another Warning

Will IBM Watson Save us from misuse of real algorithms to serve an emerging cabal of a few? (Orson Scott Card, Empire)

Or is faith and wide use of artificial intelligence a worse threat than artificial intelligence? Will we kill ourselves by stupidity?

Will control of brains by folks who do not understand them lead to really gross loss of freedom, as in this guy (or in "Clone Armies") even if nonsurgical stimulation?

Three Paths Forward

- Artificial Intelligence (AI)
  - Build/Train computers to be more like humans
  - **NATURAL** AI (CI) grounded in **real mathematics** offers more hope to actually reach the goal
- Artificial Robotics/Artificial Stupidity (AS)
  - Trains humans to be more like robots/slaves. Which candidates grow beyond the plastic molds they were trained to fit? Upton Sinclair: "It is difficult to get a man to understand something, when his salary depends on his not understanding it."
- Natural Intelligence – old but new
  - Train humans to be more human, more sane, to live up to more of their full potential. NN math allows more self-understanding, a prerequisite to full self-consciousness. (final slides, coming…)
  - Tools to help humans do this (Google to SAS, beyond)
  - Grid: values from humans, market DESIGN, energy, spirit
  - SPS: teleautonomy (Baiden), use just vector intelligence to automate subtasks, coordinate human operators by VR/markets not BCI

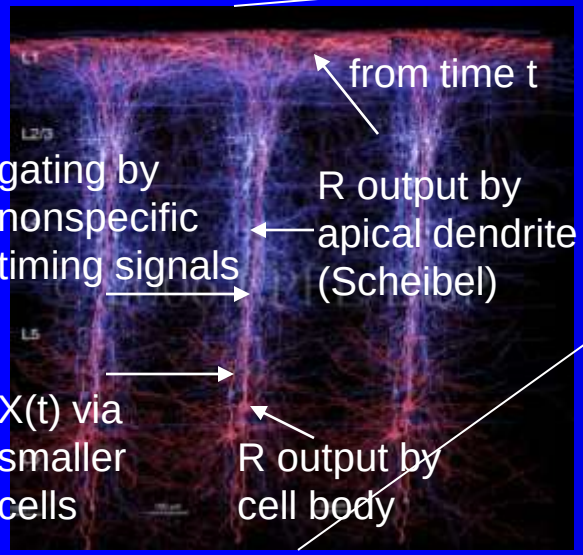
Mathematical Foundations of Prediction Under Complexity

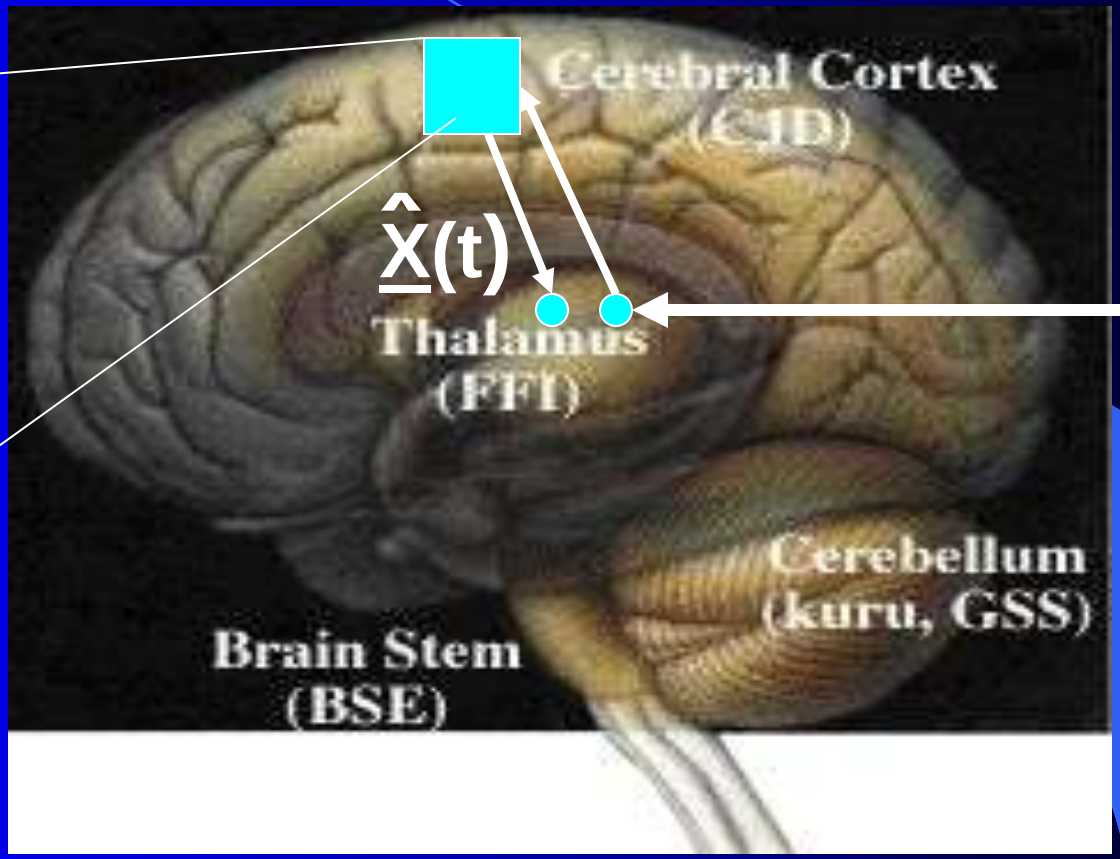
Paul J. Werbos, pwerbos@gmail.com

www.werbos.com/Erdos.pdf

- Why this is a crucial and timely piece of a larger problem
- Roadmap and definitions from vector prediction to grid and graph prediction and beyond
- Why it is not easy and not yet solved
- What must be built upon and improved

Ability to learn to "Predict Anything" Found in the Brain (Nicolelis, Chapin)





$X(t)$

(Richmond): "t+1" – t is .12 seconds. Each cycle has a forwards pass to predict, and a backwards pass to adapt

(Bliss, Spruston): found "reverse nMDA" synapse and backpropagation along dendrites

BUT: needs demonstration for more than just rat whiskers! We need "COPN2"!

What the Brain Teaches Us About Prediction

- **One universal system can learn to "predict everything**." No need for 125 different methods in 32 chapters. But "who pays for lunch"? How can it be possible?
- Can take full advantage of **massive parallel hardware** like CNN chips.
- **All predictions – including pattern recognition and memory – are in service to action**. What is true versus what is useful? It is always about "prediction of the future."
- **Incredible complexity** – learns nonlinear dynamic relations among millions of variables, based on only 10 data frames per second (300 million per year).

Definition of (Offline) Vector Prediction Task

- Assume a time-series database of a vector $x \in R^n$ and the existence of another time-series vector $r \in R^m$, obeying the dynamics

$$x(t) = h(r(t), e_1(t))$$

$$r(t) = f(r(t-1), e_2(t))$$

where e_1 and e_2 are random vectors. Try to estimate h and f or Pr(h,f) as accurately as possible, so as to be able to predict future values of x or Pr(x) or known functions U(x) as accurately as possible.

Question to Census Statistical Advisory Council (1978): What Principles Most Important in Building Or Understanding Such a Prediction System?

All said: They do not exist. It is impossible.
I would never use such a machine
even if I had it for free in my own lab.

Why It Was Seen As Impossible: 4 Schools of Thought in Statistics

- Probabilism ("We don't do inference. We just prove stuff.")
- Maximum Likelihood (Simplified from Jeffreys and Carnap)

  $\Pr(f, h \mid \text{Data}) \approx \Pr(\text{Data} \mid f, h)$

- Bayesian (e.g. Raiffa)
  - Most popular: $\Pr(f, h \mid \text{Data})$

    $= \Pr(\text{Data} \mid f, h) * \Pr(f, h) / \Pr(\text{Data})$
  - Sometimes minimize utility-based loss function
- Robust statistics (Tukey, Mosteller): try to get useful results without assuming model must be true for some value of weights W. (Also used by Raiffa, Werbos, and Vapnik.)

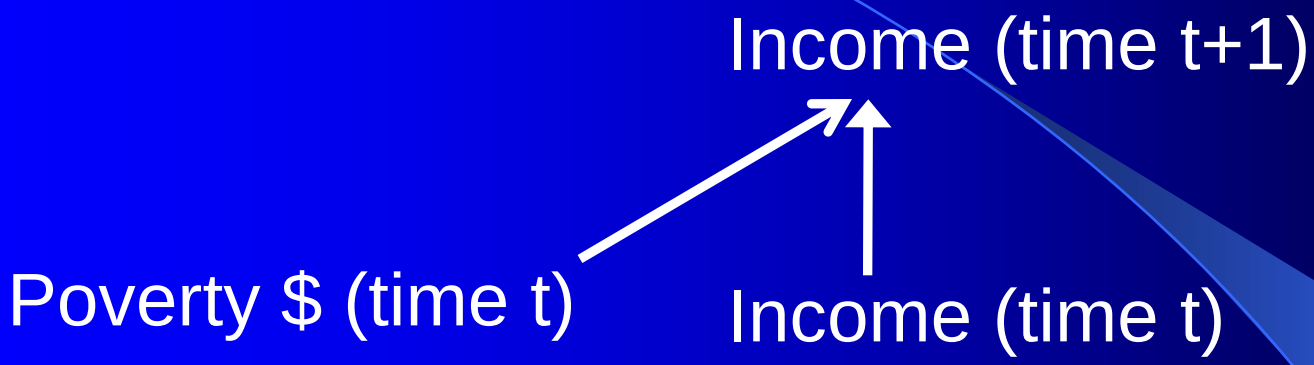
Correlation Versus Causality – Why Most Data Mining is Bogus and How We can Infer Causality

	Provinces Getting Poverty $	Provinces Not Getting Poverty $
Low Income	30	2
High Income	3	20

Human intuition or statistics for data at one time seem to say this poverty program causes low income! But think. This chart only tells you that poverty money goes to places with poverty. It does not cause the poverty

Even if you use very fancy data mining or statistical methods, you can still make huge mistakes by this kind of analysis. Vietnam War, 1967…

How to Avoid Such Mistakes

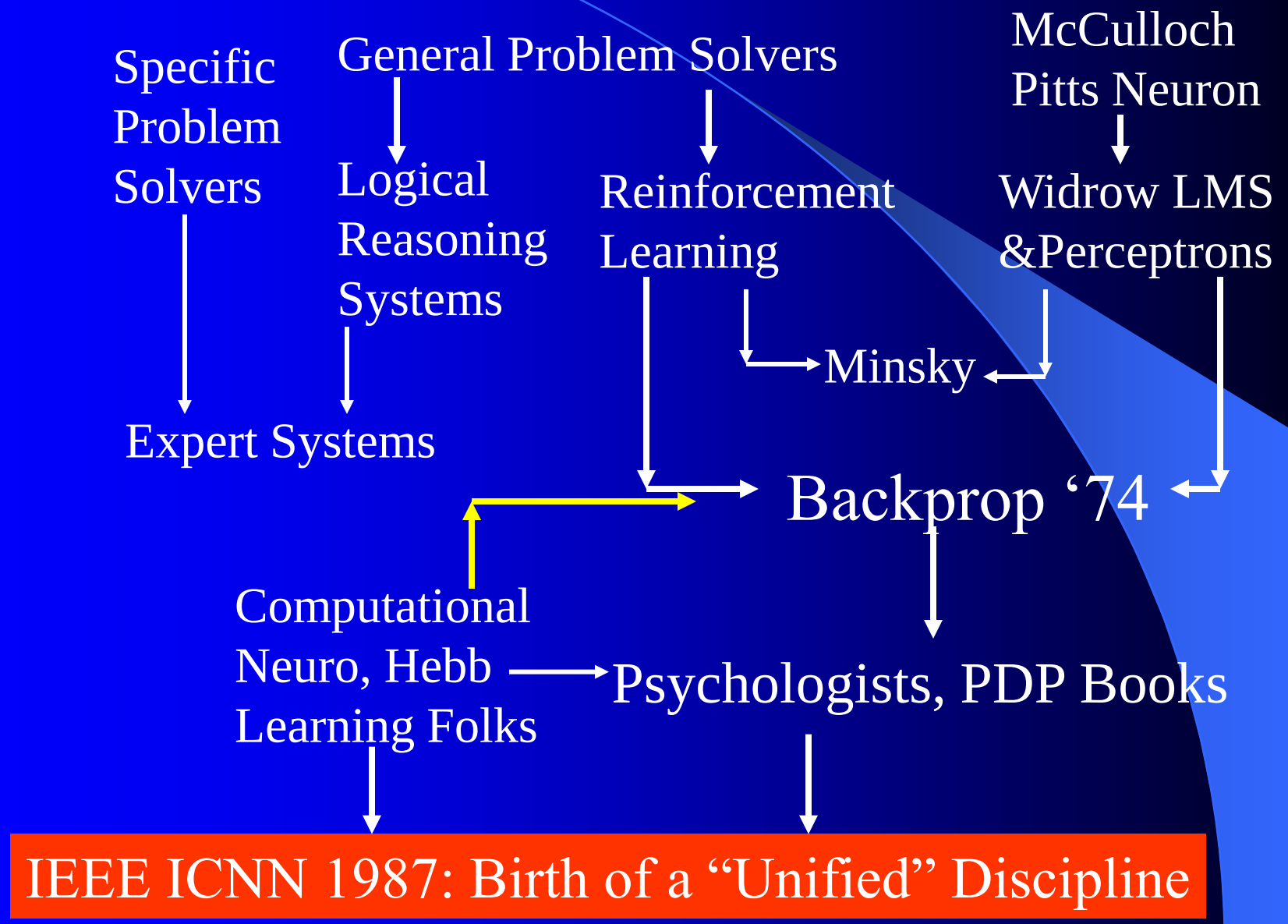


- Predict how your actions will change the other variables from one time to the next
- People call this "better statistical controls." But better statistical control really means ever better prediction of changes over time. This is never perfect; our ability to act correctly is always limited by our knowledge of how the world works. (TLRN does this automatically.)

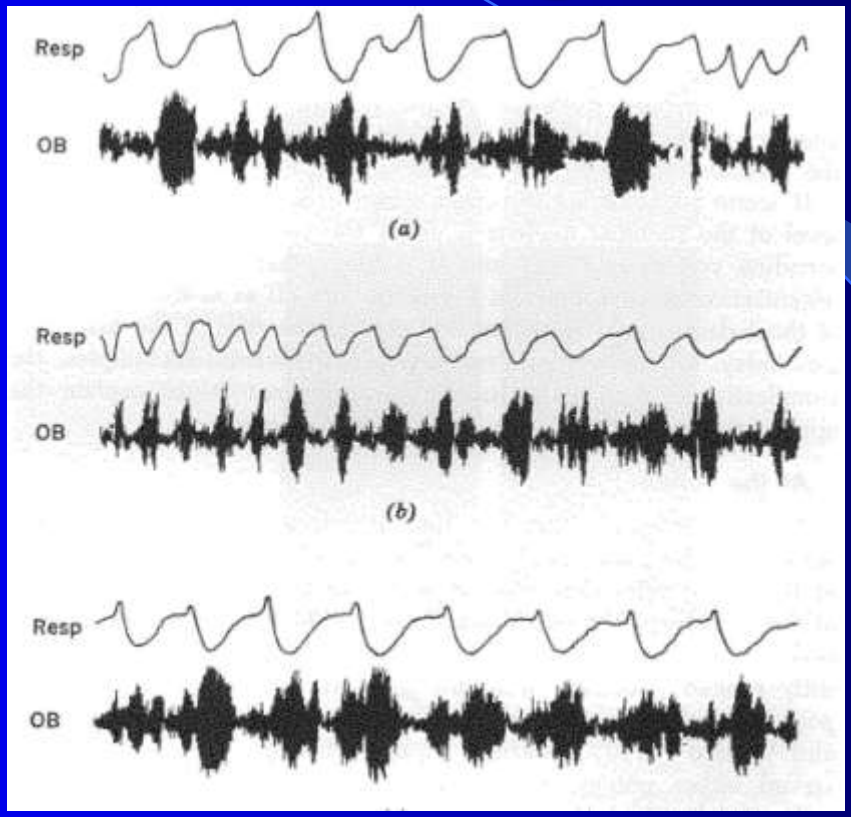
Maximum Likelihood (ML) Approach – e.g. Regression

- $Y(t+1) = b_0 + b_1 Pov(t) + b_2 Y(t) + e(t)$
- Log $Pr(e(t)) = k - ce^2(t)$ Normal/Gaussian
- $L = Pr(data \mid b_0, b_1, model) = \exp(Tk - \Sigma ce(t)^2)$
- But $Pr(model \mid data) = L*Pr(model)/Pr(data)$!
  - Bayes Law: Pr(model) can be specific OR uninformative
- Translate human insight $\Leftrightarrow$ stochastic model
  - Example of Econometric Methods, PURHAPS, crucial when only thousands of data points
  - A Key human ability in need of cultivation

Where Did ANNs Come From?

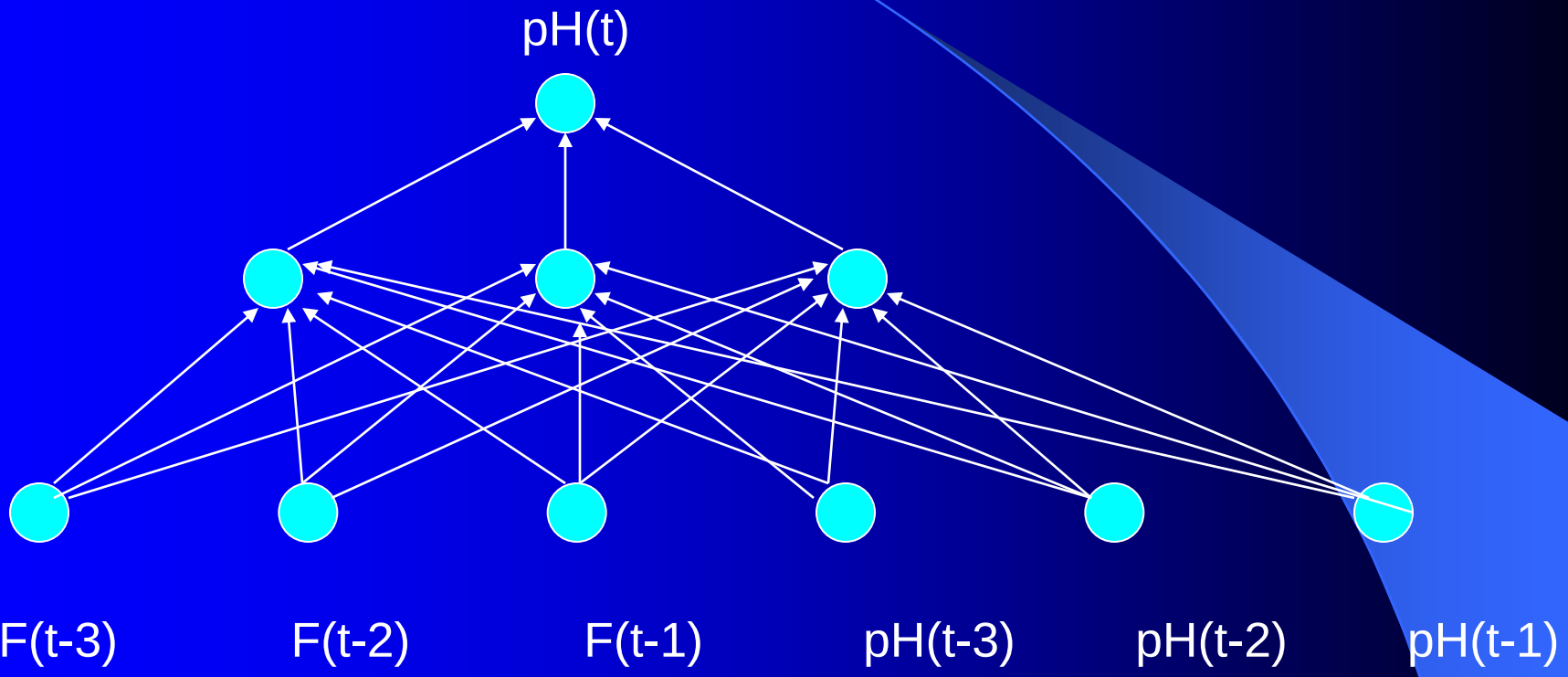


Offer to Minsky to Coauthor BP/TLU (see Talking Nets)



- Real neurons are not 1/0 asynchronous binary digits! Every 100 ms or so, a “volley” of continuous intensity. Clocks, Richmond, Llinas

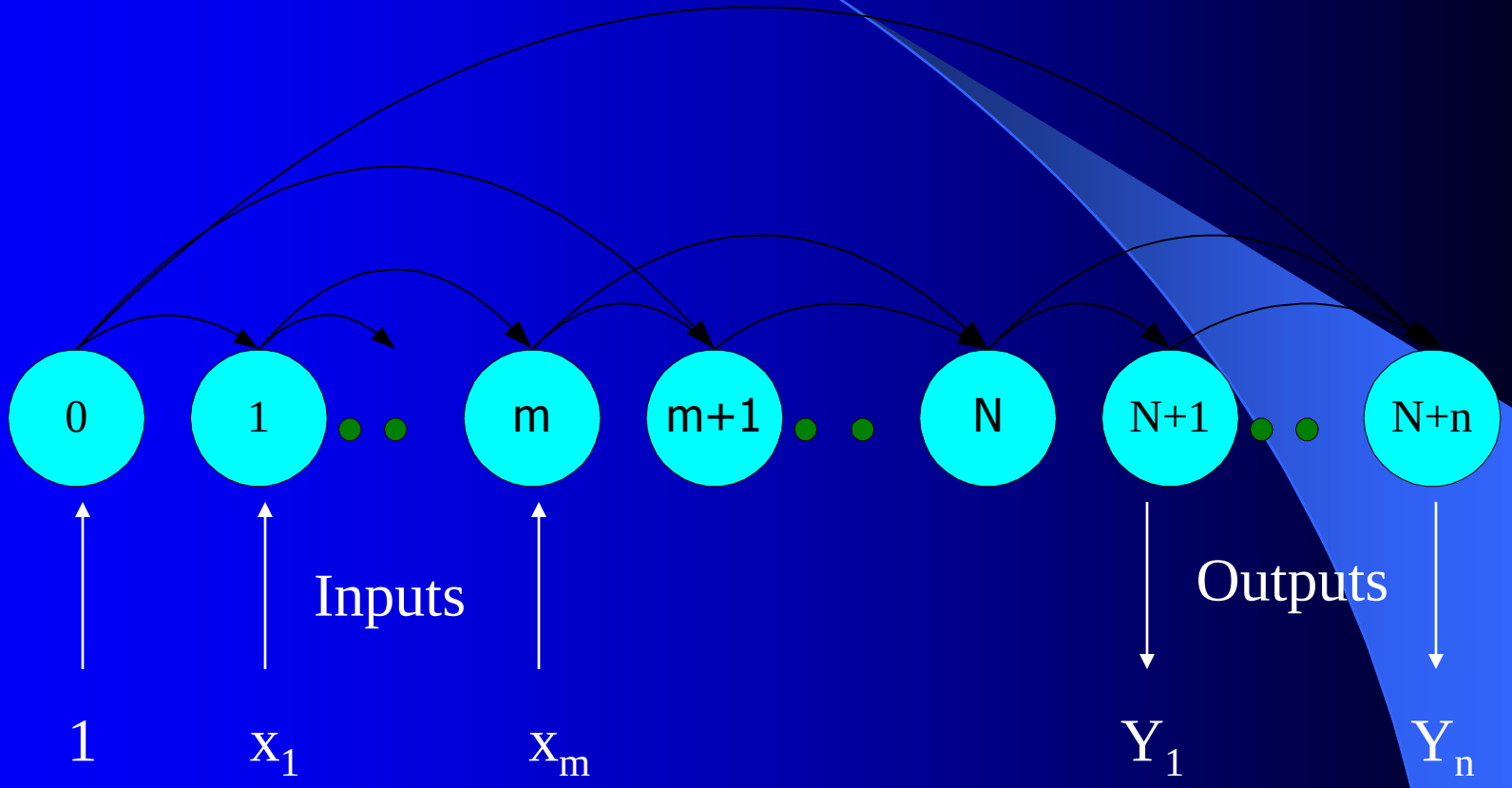
Myth 1: Training Multilayer Perceptrons (MLP) is not black magic, is not an alternative to statistics



- Any MLP represents a function **Y**=**f**(**X**,**W**), **X** the inputs, **W** the weights.
- Minimizing the mean square value of (**actual Y** – **f**(**X**,**W**)) over **W** is nonlinear regression. All the usual error and significance and standard error statistics apply. It's just a more general choice of **f** than usual (able to approximate any nonlinear smooth function efficiently) and it comes with faster more reliable convergence.

Standard errors are less with more data and fewer weights.

Generalized MLP



EQUATIONS OF GENERALIZED MLP

$x_i = X_i$ i=1 to m, read-in

do for i=m+1 to N+n

$$v_i = \sum_{j=0}^{m} W_{ij} x_j$$

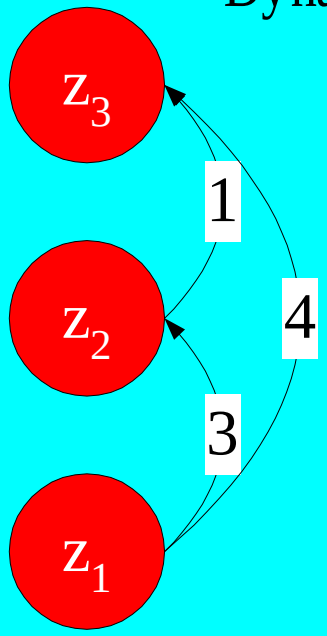
$$x_i = s(v_i)$$

$Y_i = x_{i-N}$ i=1 to N, read-out

$$E = 1/2 \sum_{i=1}^{n} (Y_i - Y_i^*)^2$$

How calculate the derivatives?

Dynamics: $z_3 = z_2 + 4z_1 ; \quad z_2 = 3z_1$



$$\frac{\partial z_3}{\partial z_2} = 1 \qquad \frac{\partial z_3}{\partial z_1} = 4$$

$$\frac{\partial^+ z_3}{\partial z_1} = \frac{\partial z_3}{\partial z_1} + \frac{\partial^+ z_3}{\partial z_2}\frac{\partial z_2}{\partial z_1} = 7$$

$$\frac{\partial^+ z_n}{\partial z_i} = \frac{\partial z_n}{\partial z_i} + \sum_{j=i+1}^{n-1} \frac{\partial^+ z_n}{\partial z_j}\frac{\partial z_j}{\partial z_i}$$

A Chain Rule For Ordered Derivatives

Historical Note: In IFIP 1981, backpropagation for deep neural networks (one for prediction F)

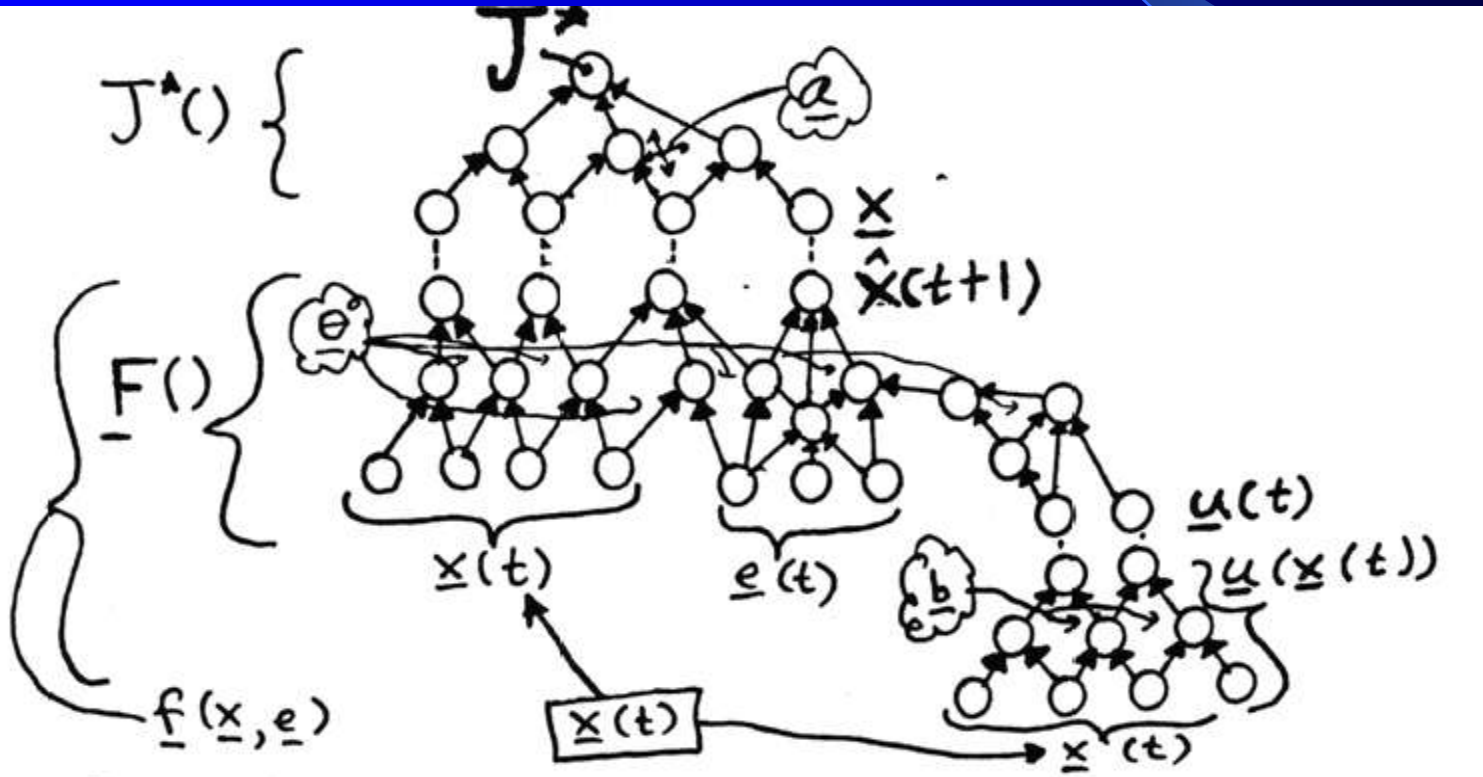


Figure 5: Realization of GDHP As a Triple Network to Make Decisions

The Economic Crunch of 2008

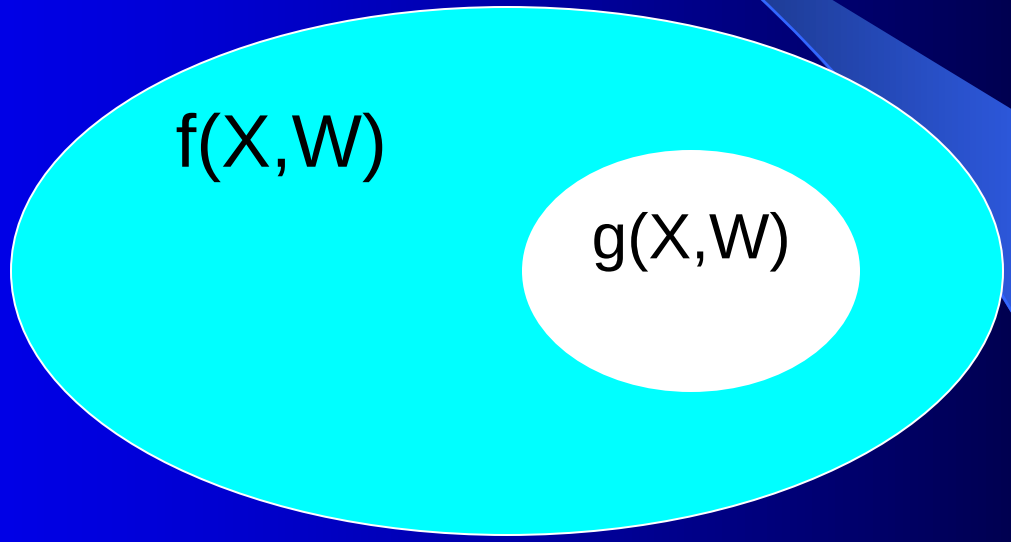
- Finance spends a lot on prediction and optimal decision. But they had many failures in 2008. Today I have time for just one.
- The trigger of the collapse:
  - Big financial firms predicted low probability of big loss in packages of mortgages
  - Given M mortgages, i=1,…,M, estimate Pr(default-i) from "FICO scores"
  - Assume independent probabilities such that

    Pr(total default) = Pr(default-1) * Pr(default-2) *…*Pr(default-M)

But FICO does not give a probability! It shifted from neural nets to SVM, from scores based on a probability method to scores based on Vapnik thinking, when this became popular. Also, no cross-time analysis or external variable conditions reported.

“No Free Lunch” Is Not a Theorem

- Some approximating functions g(X,W) can approximate only a subset of what others f(X,W) can approximate! May require a little bounded extra time to learn it, but can learn anything the other can…..



Simple ARMA is a SUBSET of vector ARMA; with time to learn, vector ARMA can learn anything simple ARMA can, and also learn what simple ARMA cannot.

In the same way, vector ARMA is a SUBSET of Time-Lagged Recurrent Networks (TLRN). TLRN is more universal (NARMAX), and it won the time-series competitions in IJCNN07 & IJCNN11. TLRN is todayt’s best vector predictor BUT CAN BE IMPROVED.

Time-Series Prediction: A Challenge to the Neural Network Field IJCNN07, IJCNN11

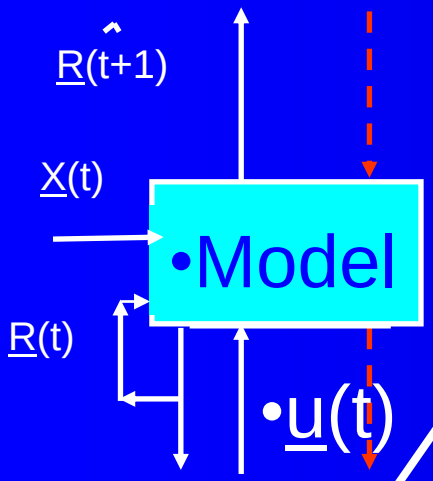
- NSF funding support via Guyon, interest
- Neural network people need to respond, but only in the right way
- Need to develop, teach and use the fundamental statistical principles which make brain-like “cognitive” prediction possible.
- How to win: lessons from past competitions, formal and informal

Best Existing Universal Learning Systems for Linear Systems

- Box and Jenkins (1971): minimize square error e:
  - $e(t) = x(t) - x\hat{}(t)$
  - $x\hat{}(t) = a_1x(t\text{-}1) + \ldots + a_px(t\text{-}p) + b_1e(t\text{-}1) + \ldots + b_qe(t\text{-}q) + c_0y(t) + \ldots c_my(t\text{-}m)$: "ARMAX model"
  - Error comparisons used to pick p, q, m; identifiable
- Werbos (1974, 1994): backpropagation allows rapid estimation of vector case (models causality):
  - $\underline{e}(t) = \underline{x}(t) - \underline{x}\hat{}(t)$
  - $\underline{x}\hat{}(t) = A_1\underline{x}(t\text{-}1) + \ldots + A_p\underline{x}(t\text{-}p) + B_1\underline{e}(t\text{-}1) + \ldots + B_q\underline{e}(t\text{-}q) + C_0\underline{y}(t) + \ldots C_m\underline{y}(t\text{-}m)$: "vector ARMAX model"

Roadmap for Cognitive Prediction

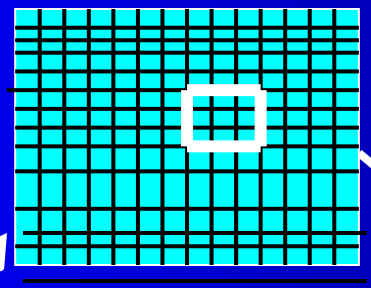
Reward direct simplicity



0. Vector Prediction (robustified SRN/TLRN)
HIC Chapter 10 on web.

Reward symmetry

1. AT&T winning ZIP code recognizer and new COPN work



Networks for inputs with more spatial complexity using symmetry – CSRN, ObjectNets, ….

Predicts What Will Happen Over Multiple Time Intervals Harmonized

2. reptile

Space-like cognitive map of the space of Possibilities, to support higher creativity

3. Mouse

To see how you could do better than even them, and break the world records again… or to see the research needs to fulfill gthis roadmap… see

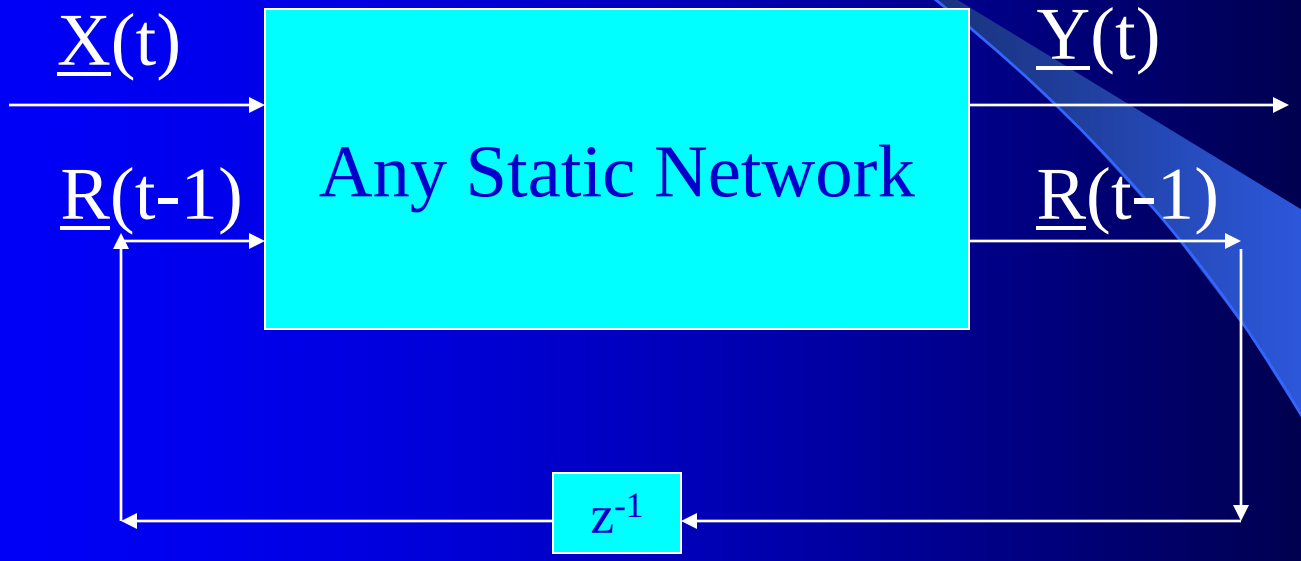
www.werbos.com/Erdos.pdf

Universal Vector Prediction System: Principles To be Explained

- For smooth functions $\underline{\mathbf{Y}}=\underline{\mathbf{f}}(\underline{\mathbf{X}})$, Multilayer Perceptron (MLP) minimizes complexity and hence estimation error. Barron.
- For general functions $\underline{\mathbf{Y}}=\underline{\mathbf{f}}(\underline{\mathbf{X}})$, add simultaneous recurrence ($\underline{\mathbf{y}}[n+1]=F(\underline{\mathbf{y}}[n],\underline{\mathbf{X}})$ for Turing-like universality. SRN.
- For dynamic or time-series prediction, add time-lagged recurrence $\underline{\mathbf{Y}}(t)=\underline{\mathbf{f}}(\underline{\mathbf{Y}}(t-1),\underline{\mathbf{X}}(t))$ for universal "NARMAX" capability (TLRN)
- Unify maximum likelihood (least squares training) with precedent-based forecasting, "uninformative priors" (penalty functions), & weights for multiperiod prediction and salience – especially for real-time "incremental" learning.

⇒ Learning speed also an issue, harder with better prediction. Many useful tricks known. Kozma/Ilin/Werbos patent just a useful start.

Time-Lagged Recurrent Network (TLRN): 50% of coal generators, Neuco Siemens….



$\underline{Y}(t)=\underline{f}(\underline{X}(t), \underline{R}(t-1)); \underline{R}(t)=\underline{g}(\underline{X}(t), \underline{R}(t-1))$

$\underline{f}$ and $\underline{g}$ represent 2 outputs of one network

All-encompassing, NARMAX($1 \equiv n$)

Felkamp/Prokhorov Yale03: >>EKF,≈ hairy

Why It Was Seen As Impossible: 4 Schools of Thought in Statistics

- Probabilism ("We don't do inference. We just prove stuff.")
- Maximum Likelihood (Simplified from Jeffreys and Carnap)

  $\Pr(f, h \mid \text{Data}) \approx \Pr(\text{Data} \mid f, h)$

- Bayesian (e.g. Raiffa)
  - Most popular: $\Pr(f, h \mid \text{Data})$

    $= \Pr(\text{Data} \mid f, h) * \Pr(f, h) / \Pr(\text{Data})$
  - Sometimes minimize utility-based loss function
- Robust statistics (Tukey, Mosteller): try to get useful results without assuming model must be true for some value of weights W. (Also used by Raiffa, Werbos, and Vapnik.)

Uninformative Priors: The Big Picture

- Philosophers studying human learning have known for centuries that we cannot explain human learning without "uninformative priors"
  - Reverend Occam: assume higher Pr(f,h) for "simpler models" f and h
  - Emmanuel Kant: the "apriori synthetic"
- Solomonoff/Werbos (60's): if f and h are instructions to a Turing machine, assume Pr(f,h)=a exp(-kC), where C, the complexity, is the number of symbols needed to express the Turing machine. This is universal to all Turing machines, to within some finite "learning time" to adapt from one Turing machine to another. In a way, this is the perfect best possible general foundation for a universal learning machine, but….
- How can we approximate its implications for f and h implemented, for example, as networks of neurons?
  - "reward" (higher Pr assumed) for networks of greater direct simplicity
  - also "reward" the greater simplicity implied by symmetry (reflecting how Turing machines can reuse subroutines)
- How do we handle x and y being continuous? And our inability to integrate over all possible models? -- direct priors, nonlinear version of "empirical Bayes" (Efron)? Is brain limited to 3+1-D (Kant)?

"Bayes" versus "Vapnik": today's debate

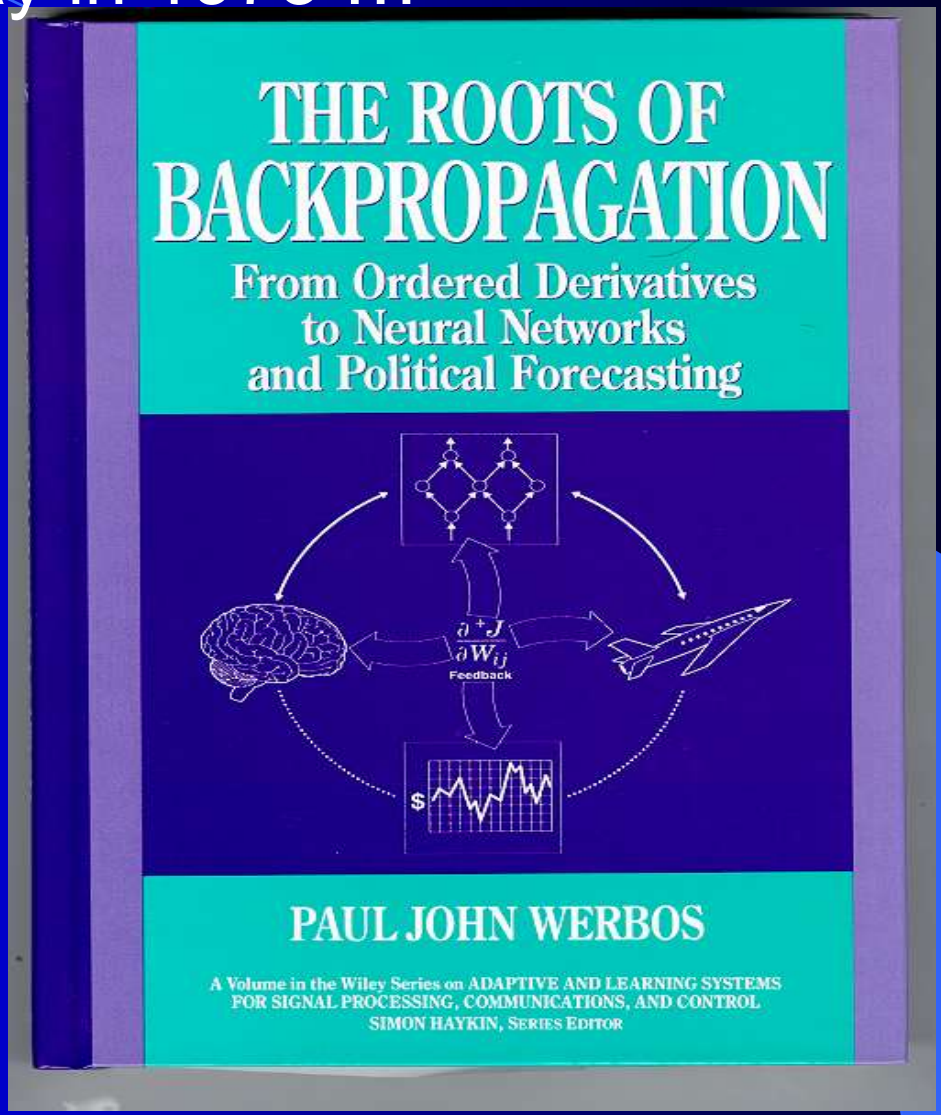
- Theorem: Pr(A|B) = Pr(B|A)*Pr(A)/Pr(B)
- Platonic Bayes:
  - Predict by using stochastic model Pr($\underline{\mathbf{x}}$(t)|past)
  - Find model with highest probability of being true: Pr(Model_W|database) = Pr(database|Model_W)* Pr(Model_W)/Pr(database)
  - Neural $\underline{\mathbf{x}}(t+1)=\underline{\mathbf{f}}(\underline{\mathbf{x}}(t),\ldots,W)+\underline{\mathbf{e}}(t)$ is just another stochastic model, with full NL regression statistics
  - Many variations; e.g. "Box-Jenkins" ARMA methods
  - "anything else is Las Vegas numerology"
- Vapnik says NO. "New" philosophy: if you want $, not truth, pick Model_W which would have maximized $ in the past (database)

But Platonic Bayes fails very badly in some ways, as I learned the hard way in 1973 …

Vector ARMA (f) had twice the prediction error of simple extrapolator (g), on 100-year political data and simulated dirty datasets

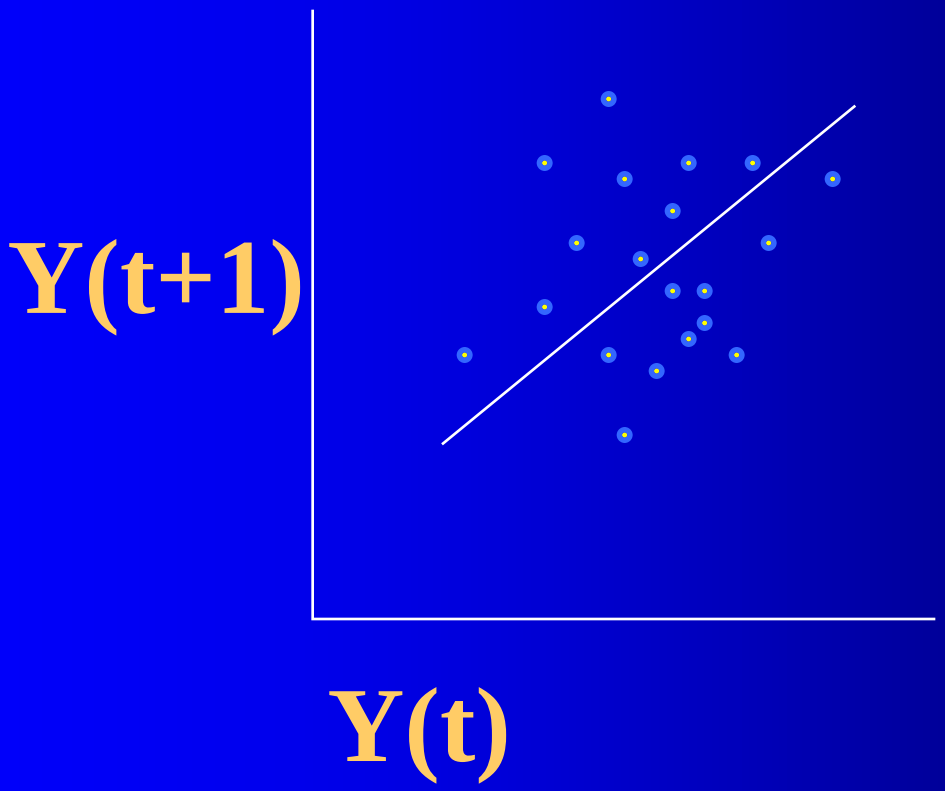
"Vapnik" style "pure robust method"

BRAINS absolutely require multiperiod robustness beyond what Platonic Bayes offers

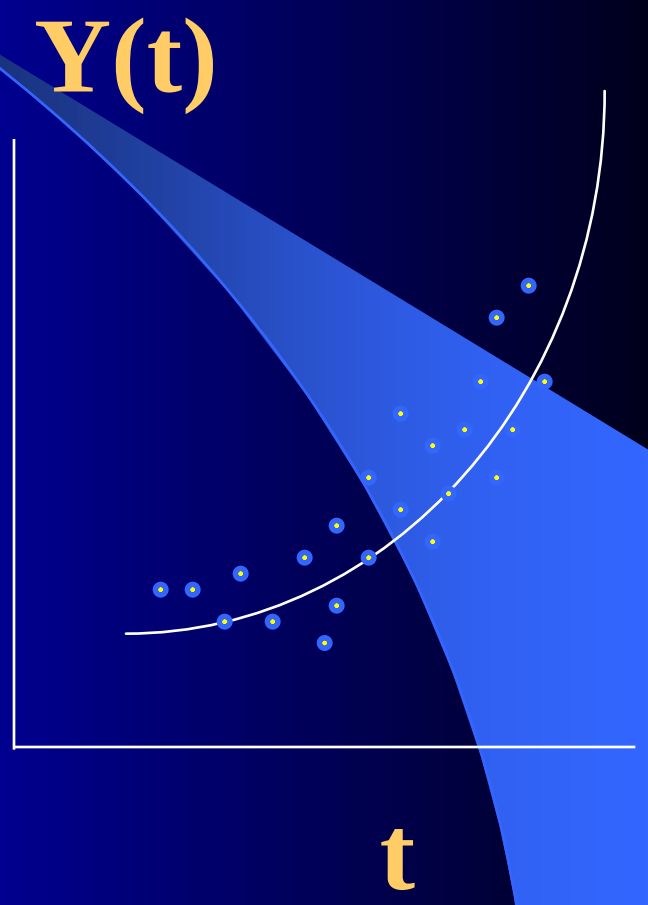


1974 Harvard PhD in subject of statistics, Mosteller on committee (Dempster help)

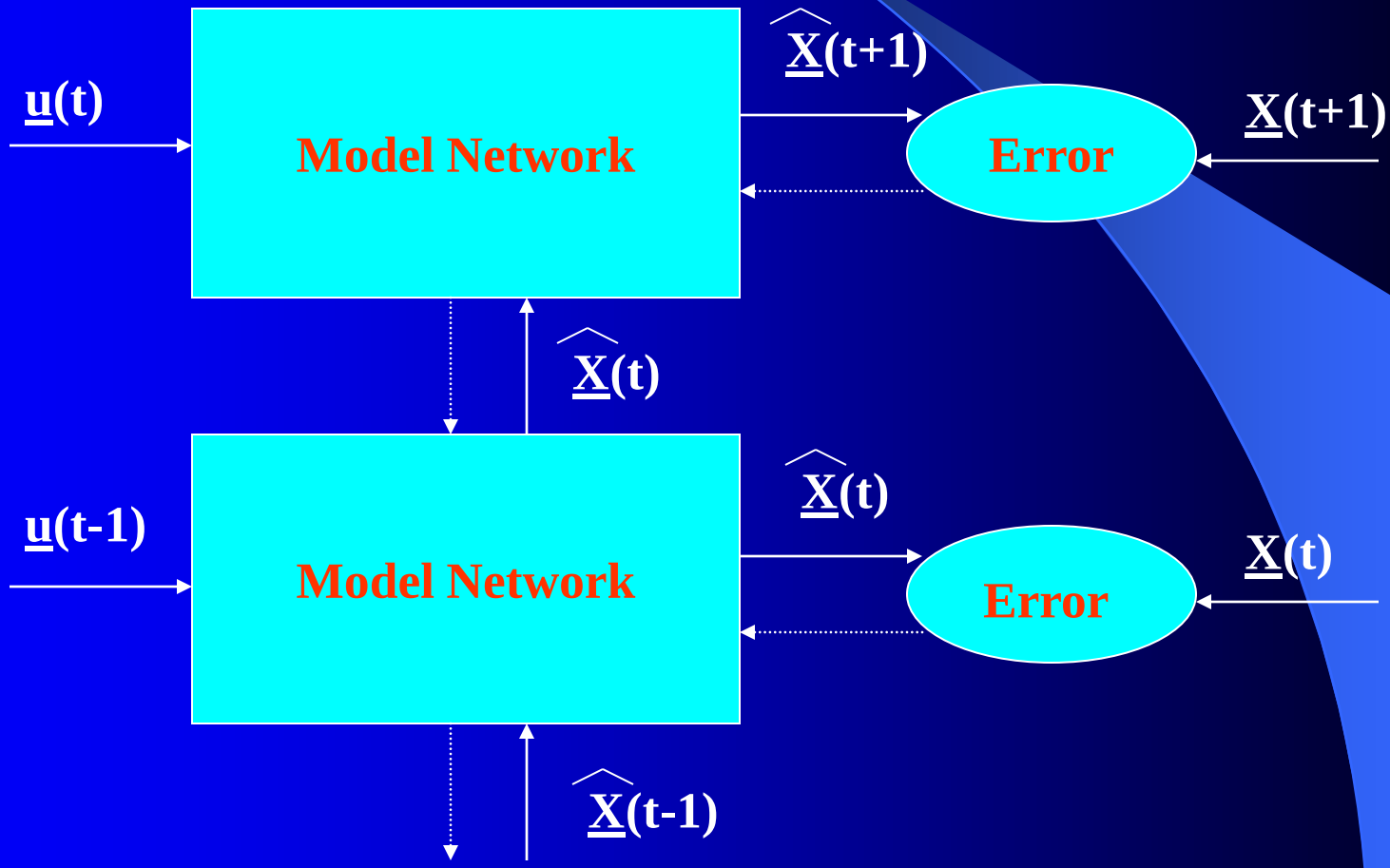
Conventional Least Squares

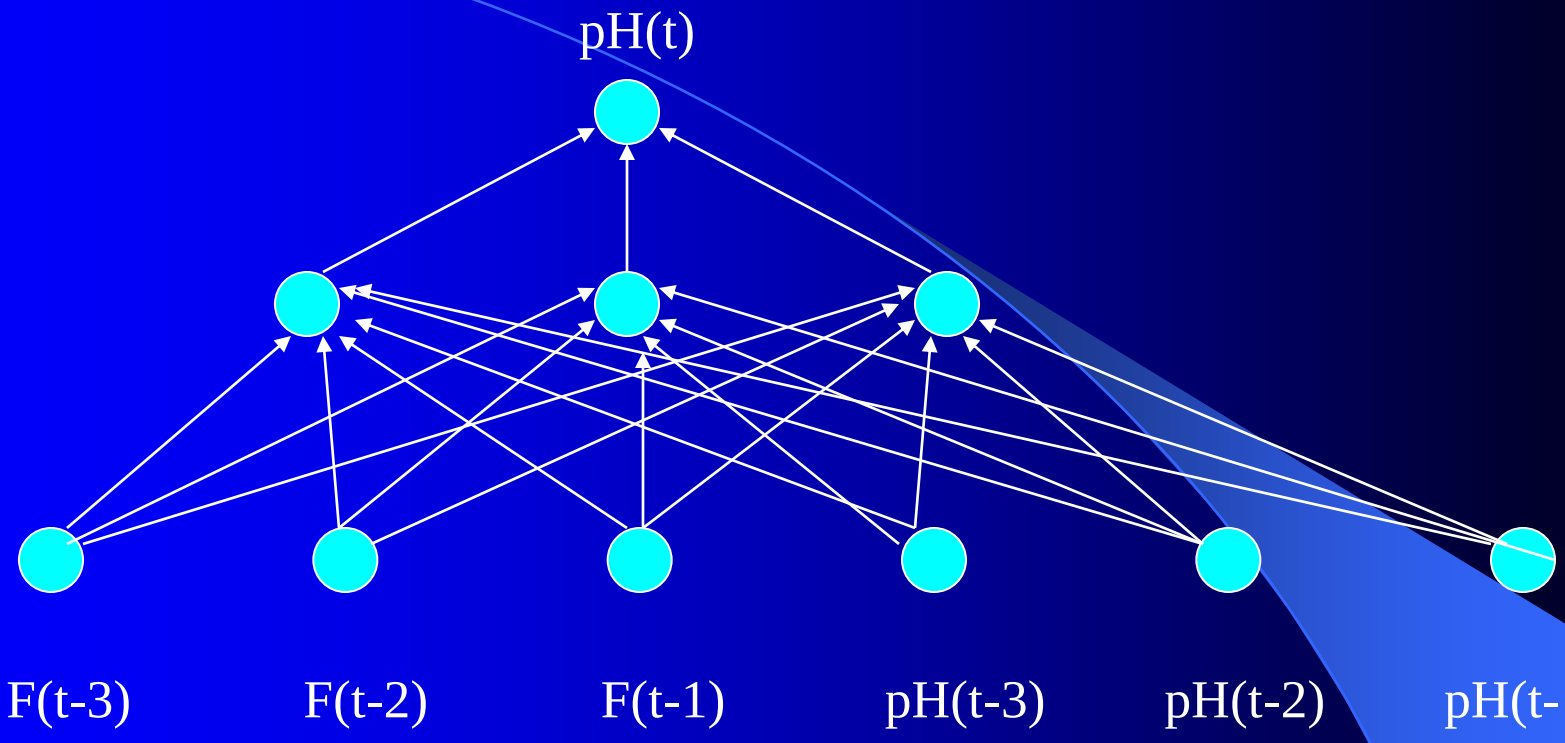


Pure Robust



PURE ROBUST METHOD

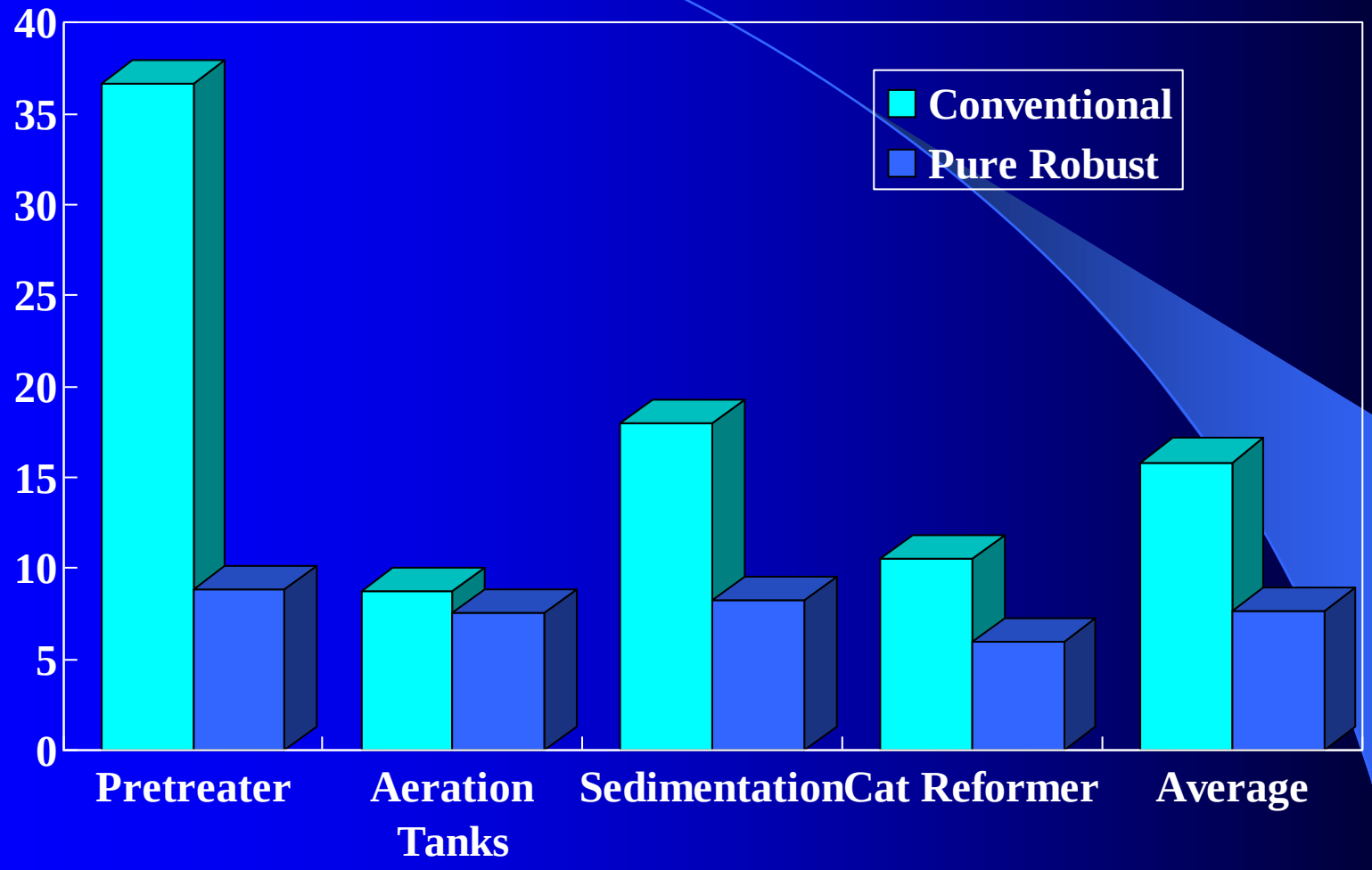




Example of TDNN used in HIC, Chapter 10

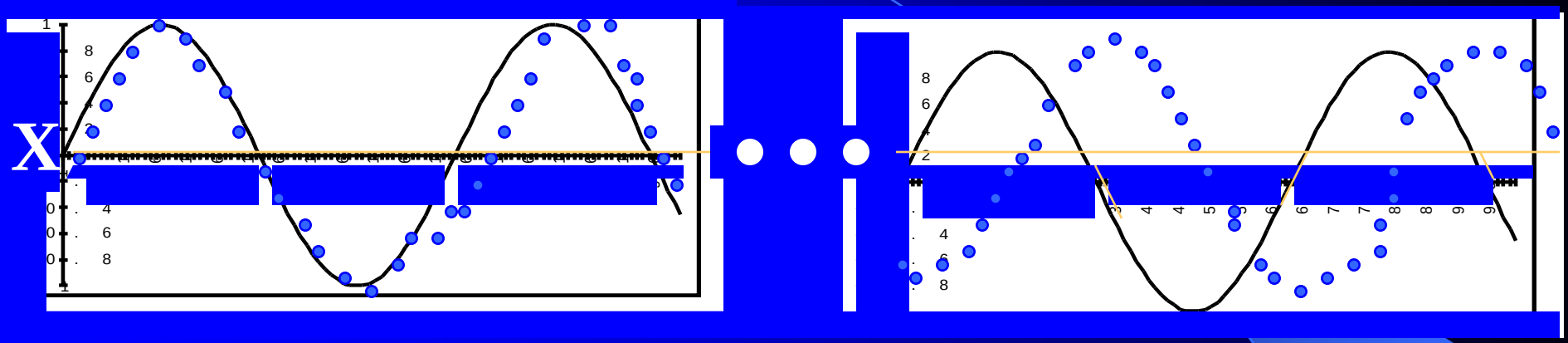
TDNNs learn NARX or FIR Models, not NARMAX or IIR

Prediction Errors (HIC p.319)



- Greatest advantage on real-world data (versus simulated)
- Full details in chapter 10 of HIC, posted at www.werbos.com.
- Statistical theory (and how to do better) in second half of that chapter.

But Pure Robust ("Vapnik") Can Fail Badly Too: Phase Drift



$$R(t+1) = R(t) + w + e_p(t)$$

$$X(t) = \sin R(t) + e_m(t)$$

TINY

A unified method cut GNP errors in half on Latin American data, versus maximum likelihood and pure robust both (SMC 78, econometric).

Best Hybrid Known So Far

- Cut error 50% in predicting GNP in Latin America versus the best of ML and Pure Robust
- (see page 327, Chapter 10, HIC)

$$general\ example: \quad e(t) = \tfrac{1}{2}\sum_i (y_i(t) - \hat{y}_i(t))^2$$

$$ML\ version: \quad \hat{y}_i = \tilde{f}_i(y(t-1), x(t))$$

$$Pure\ Robust\ version: \quad \hat{y}_i = \tilde{f}_i(\hat{y}(t-1), x(t))$$

$$Compromise\ Method\ Version\ (1977\ version):$$

$$\hat{y}_i = \tilde{f}_i(\tilde{y}(t-1), x(t))$$

$$\tilde{y}_i(t) = (1 - w_i)\hat{y}_i(t) + w_i y_i(t)$$

$$Minimize \sum_t \frac{(y_i(t) - \hat{y}_i(t))^2}{\sigma_{y_i}^2 (1 - |w|}$$

"Vapnik" approach is not new even in the static case

- Utilitarian Bayes: google "Raiffa Bayesian": pick model and weights W so as to minimize a loss function L.
- Example of the issue: to weight or not weight your regression (in actual DOE/EIA model and conflict model):

  Energy(state,year)=a*income(state,year)+e(year) (1)

  (energy(state,year)/income(state,year)=a+e(year) (2)

  If big states different, equation (1) is more consistent

  If big states few, (2) has more information, less random error

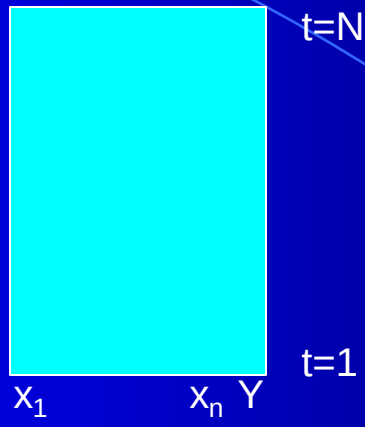
  Platonic approach: use F tests to see which is more true, but..

NonBayesian methods in econometrics for consistency under more general conditions

Lowest level of the ladder: static x-to-y, not yet conquered

- Just assume y=f(x,e), for a database of x and y, where e is random and y is exogenous, where f is one sample from Pr(f)=c exp(-kC), and C is a Sobolev measure or the max of the length of the gradient of f, or such. Can we combine both theorems and simulations to move towards a universal learning system – a system which approaches the best possible performance in this case? And outperforms the many ad hoc methods now being used in “data mining” for this purpose?

Model-Based Versus Precedent-Based: Which Is Better?



- Model-based: Pick W to fit Y=g(x,W) across examples t. Given a new x(T), predict Y(T) as g(x(T),W). Exploit Barron's Theorem that smooth (low C) functions f are well approximated by simple MLP neural nets – though not by Taylor series. Also add penalty function to error measure, ala empirical Bayes, Phatak – min e+f(W).
- Precedent-Based: Find t whose x(t) is closest to x(T). Predict Y(T) as Y(t). Kernel is similar, weighted sum of near values.
- Best is optimal hybrid, needed by brain. "Syncretism" – chapter 3 of HIC…. Next 2 slides

“Syncretism” Design

Basic Idea:

$$\hat{Y}(t) = \tilde{f}(x(t)) + \sum_{\tau} K(x(t) - x(\tau))(Y(\tau) - \tilde{f}(x(\tau))$$

Practical Implementation/Approximation:

-- Associative Memory of Prototype $x(\tau), Y(\tau), Y(\tau)\text{-}f^*(x(\tau))$

-- Update $Y(\tau)\text{-}f^*(x(\tau))$ on occasion as f^* is changed

In other words: Keep training f* to match examples or prototypes in memory, especially high-error examples.

Predict Y(t) by f* plus adjustment for errors of f in nearby memory.

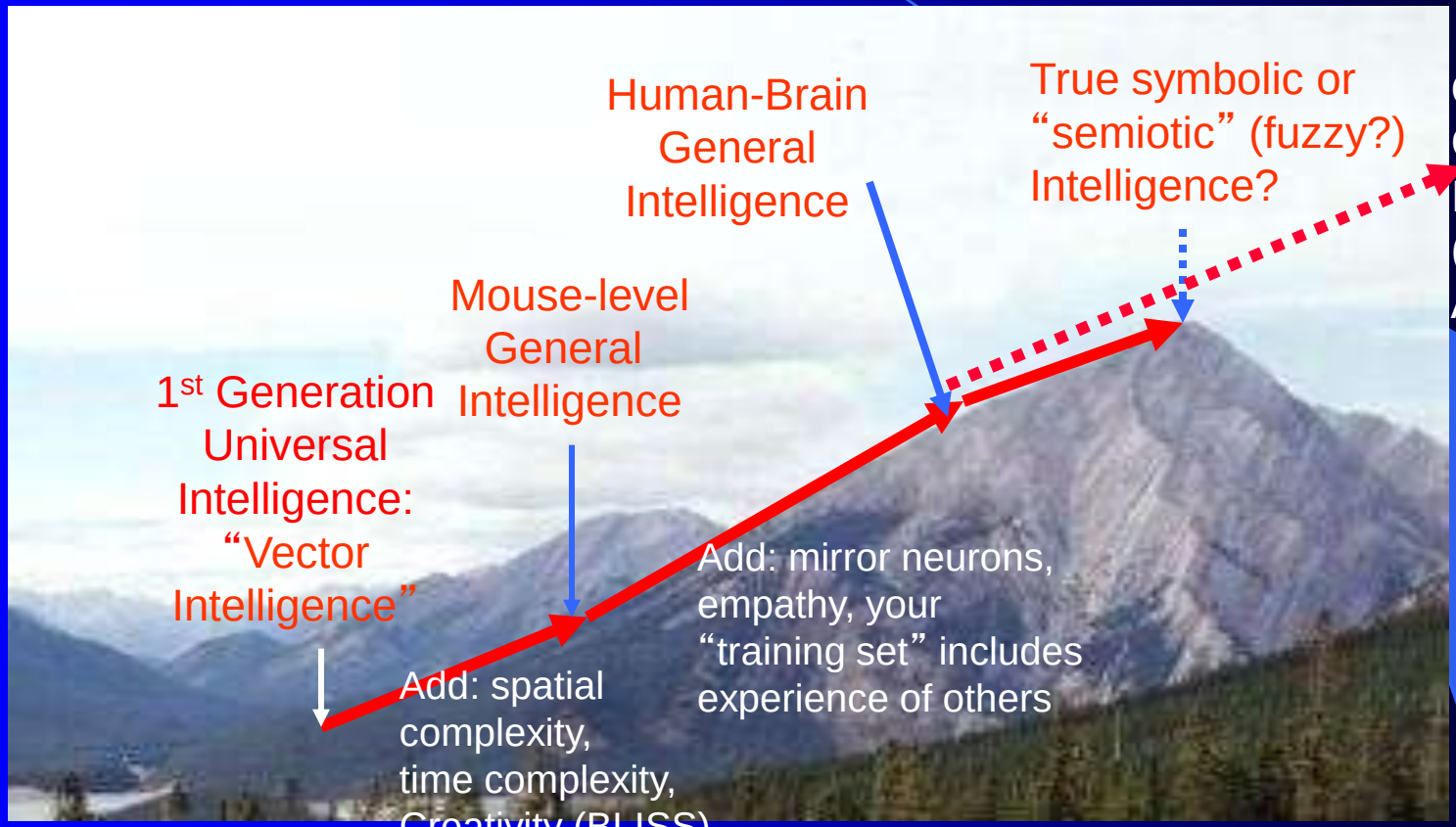
Closest so far: Principe kernel applied to model residuals; Atkeson’s memory-based learning.

Exactly fits Freud’s description of ego versus id in neurodynamics.

Example of Freud and Syncretism

- A Freudian story:
  - Nazi hurts child, a traumatic memory
  - For years, he is terrified when anyone in black shirt appears (precedent based prediction/expectation) – the kernel-based "id" is at work!
  - Later he learns about Nazis in subjective model of world (f), "ego"
  - After that learning, if he relives that memory (trains on memory), f error on the memory is low; memory loses power to cause irrational bias
- Key corollaries:
  - False hope from memory is as dangerous as false fear
  - We still need id when exploring new realms we can't yet reliably predict

From Brain to Mind: What Can We Learn Of Use Beyond the Level of the Mouse Brain?



www.werbos.com/pi/Confucius_talk.pdf

And Neural Networks 2012

Time-Symmetric Physics: A Radical New Approach to Analog Quantum Computing and Reduced Decoherence

keynote talk posted at arxiv.org

given at PIRCAI (Australia) Dec. 4, 2014

With links to audio, slides

See also 2016 paper by Paul and Ludmilla in Quantum Information Processing, and new book Freeman, Kozma eds