From Reinforcement Learning to Sequential Decision Analytics, with Applications in Transportation and Logistics

MIT Mobility Forum
March 24, 2023

Warren B Powell
Chief Innovation Officer, Optimal Dynamics
Professor Emeritus, Princeton University


NASDAQ
COVID 19
Coronavirus Vaccine
AIR PRODUCTS
tell me more
www.airproducts.com
9533
cargolux
E-COMMERCE

CHALLENGES

Virtually every problem in the domain of human processes combines decisions and uncertainty

Engineering

Physical science

Economics

Public health

Health systems

Laboratory sciences

Energy systems

Finance

Manufacturing

Personal transportation

Supply chains

E-commerce

Freight transportation

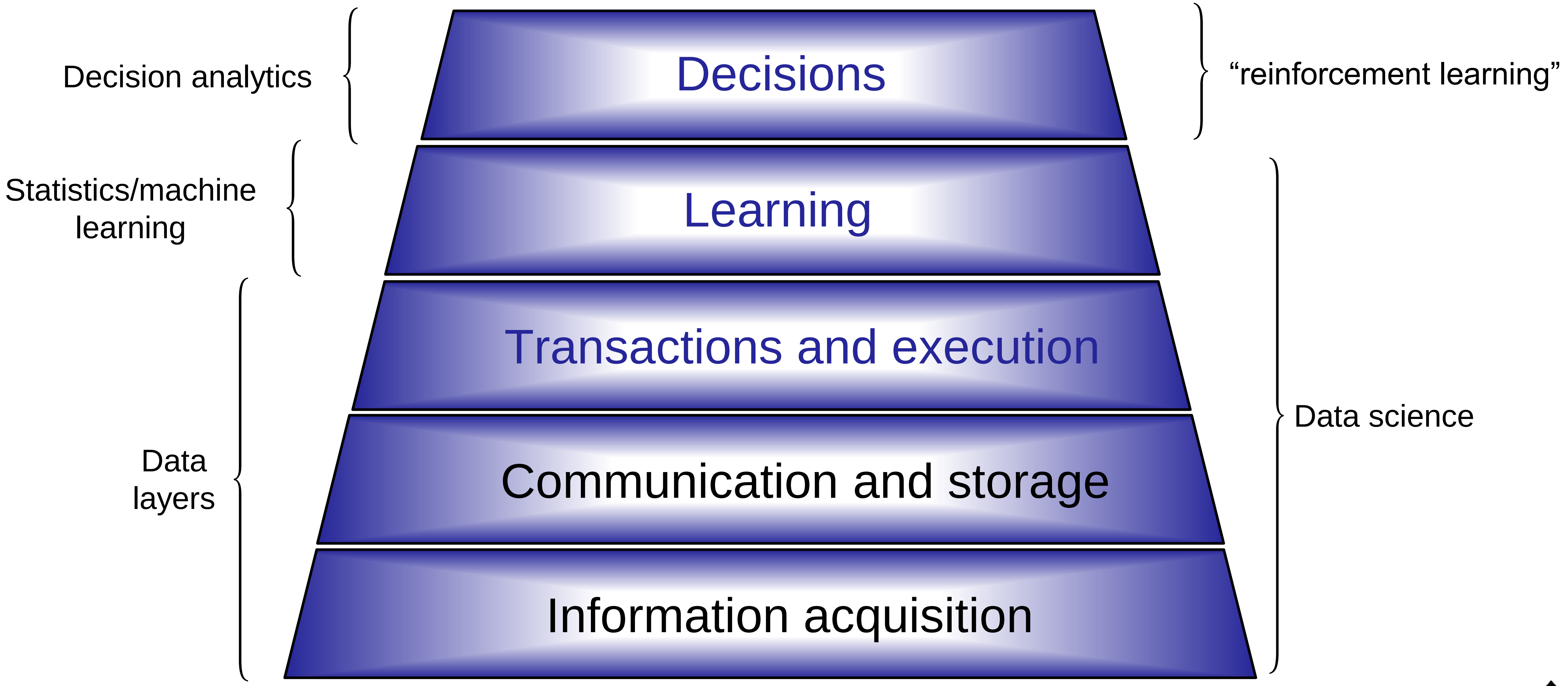
OUTLINE

- → The five layers of intelligence
- → Modeling sequential decision problems
- → Modeling uncertainty
- → Designing policies
- → A new educational field: sequential decision analytics

OUTLINE

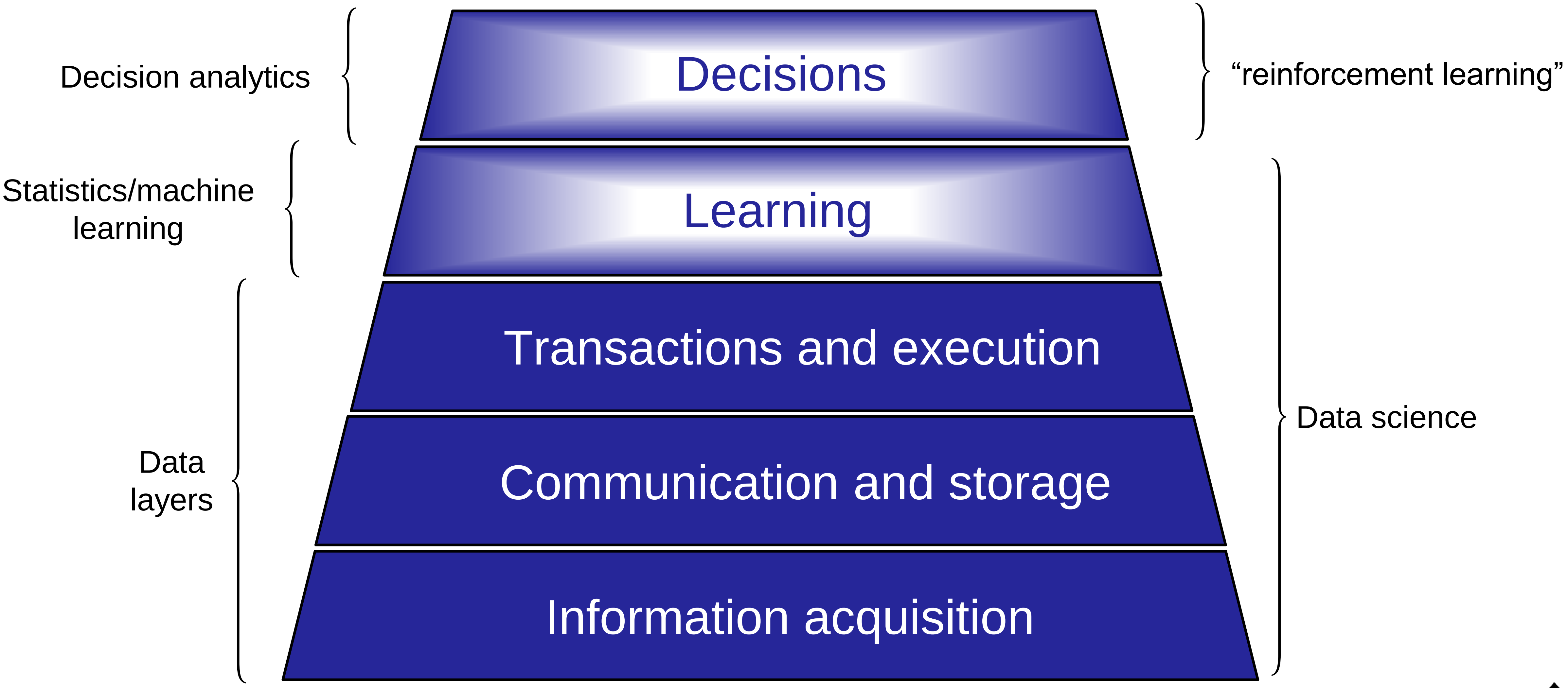
- → The five layers of intelligence
- → Modeling sequential decision problems
- → Modeling uncertainty
- → Designing policies
- → A new educational field: sequential decision analytics

THE 5 LAYERS OF INTELLIGENCE



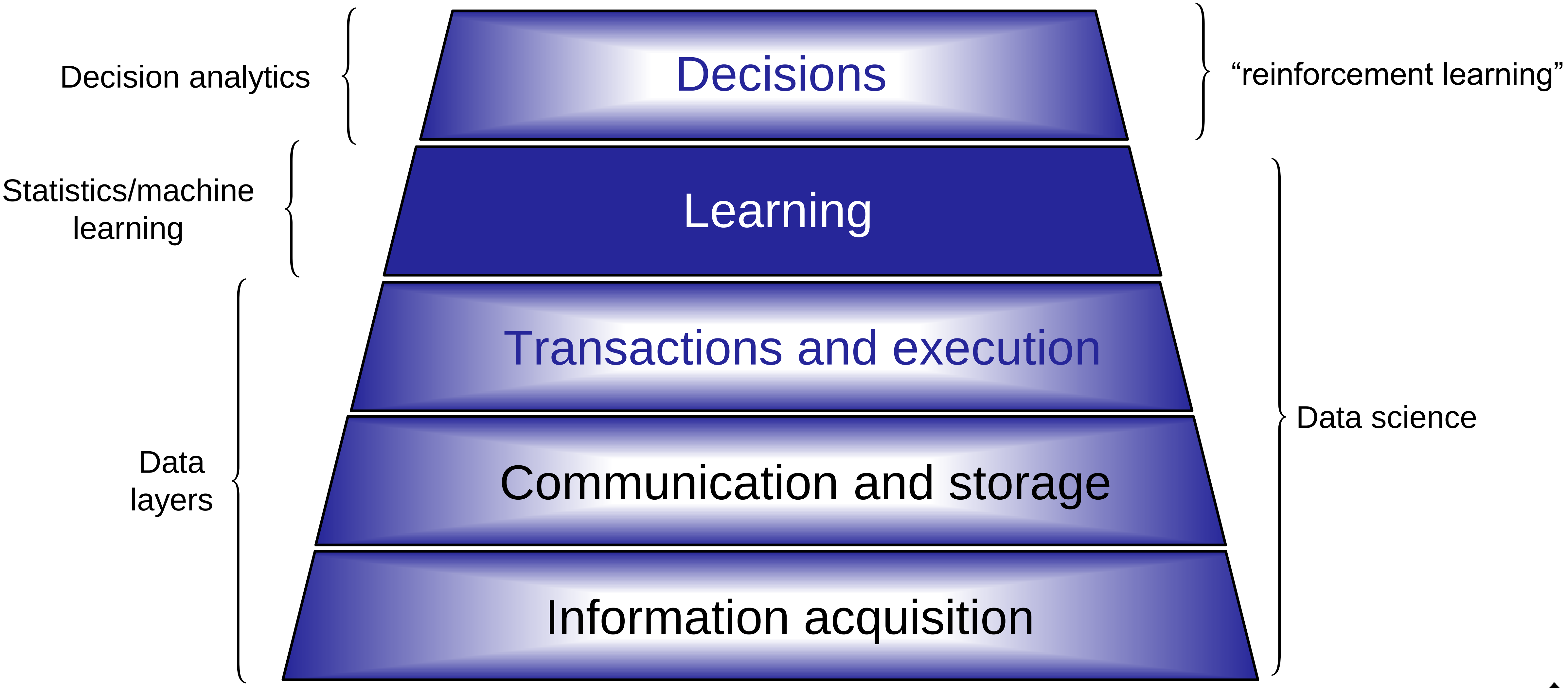
Optimal Dynamics

THE 5 LAYERS OF INTELLIGENCE



INFORMATION & DECISION PROCESSES

THE 5 LAYERS OF INTELLIGENCE



Optimal Dynamics

MACHINE LEARNING

Types of Learning

Pattern Matching	Classification	Inference	Prediction

» What is the voice saying?

» What is in the picture?

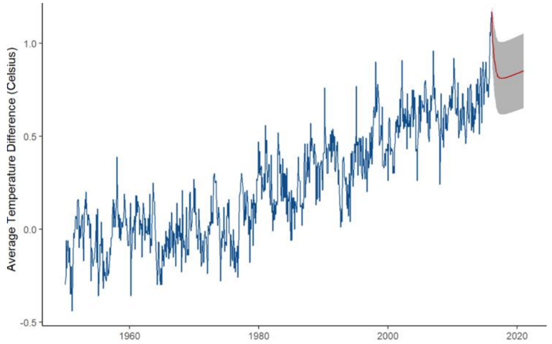
» What is the email asking for?

» What product should I recommend for this customer?

» What treatment should I recommend for this patient?

» How will an increase in price affect market demand?

» What is the condition of a piece of equipment?

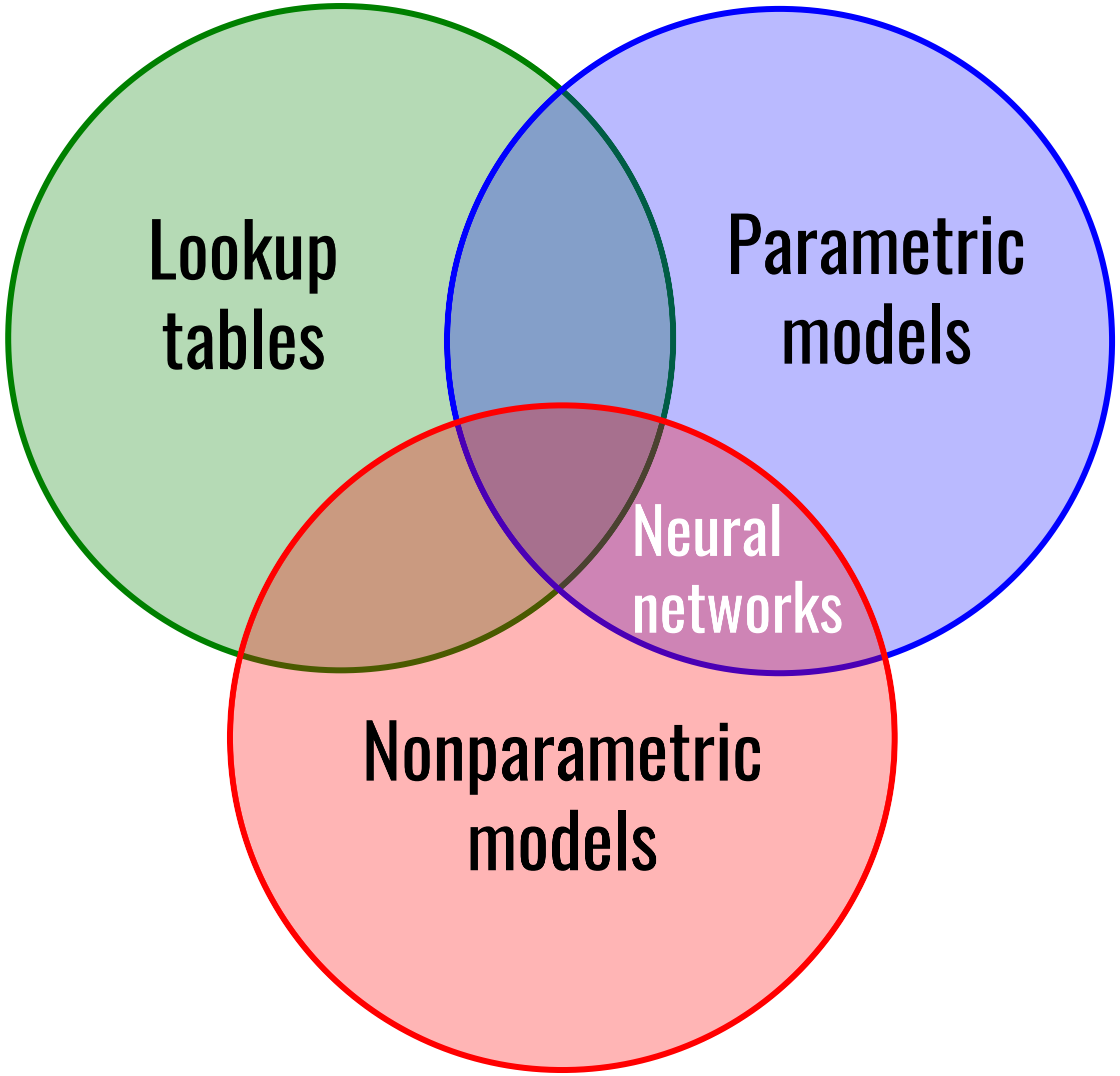


» What will the market demand be in three days?

» How many loads will the shipper need to move in a week?

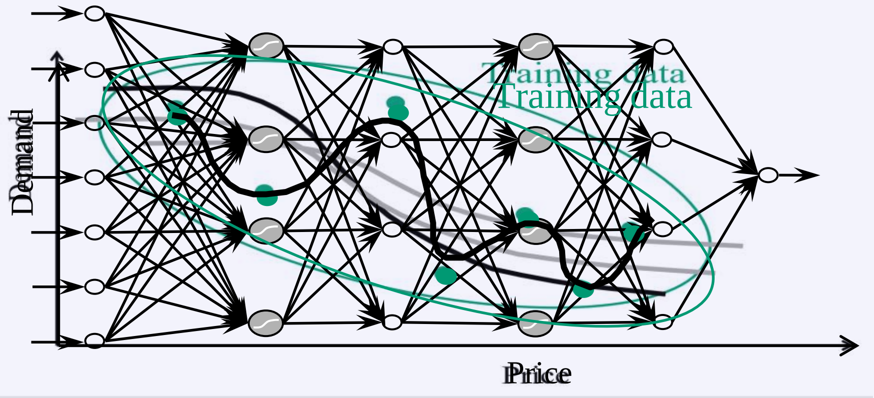
MACHINE LEARNING

Every single machine learning method falls in one of these three circles.



Machine learning as an optimization problem

$$\min_{f\in F,\theta\in\Theta^f} \frac{1}{N}\sum_{n=1}^{N}\left(y^n - f(x^n|\theta)\right)^2$$

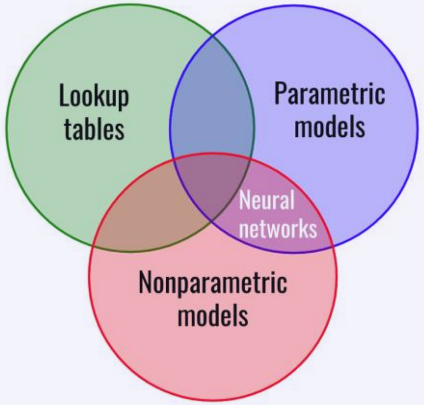


The first step is choosing a mathematical model that will do the best job of fitting the data (but be careful of overfitting with neural networks).

BRIDGING MACHINE LEARNING & SEQUENTIAL DECISIONS

Machine learning as an optimization problem

$$\min_{f\in F,\theta\in\Theta^f} \frac{1}{N}\sum_{n=1}^{N}\left(y^n - f\left(x^n|\theta\right)\right)^2$$

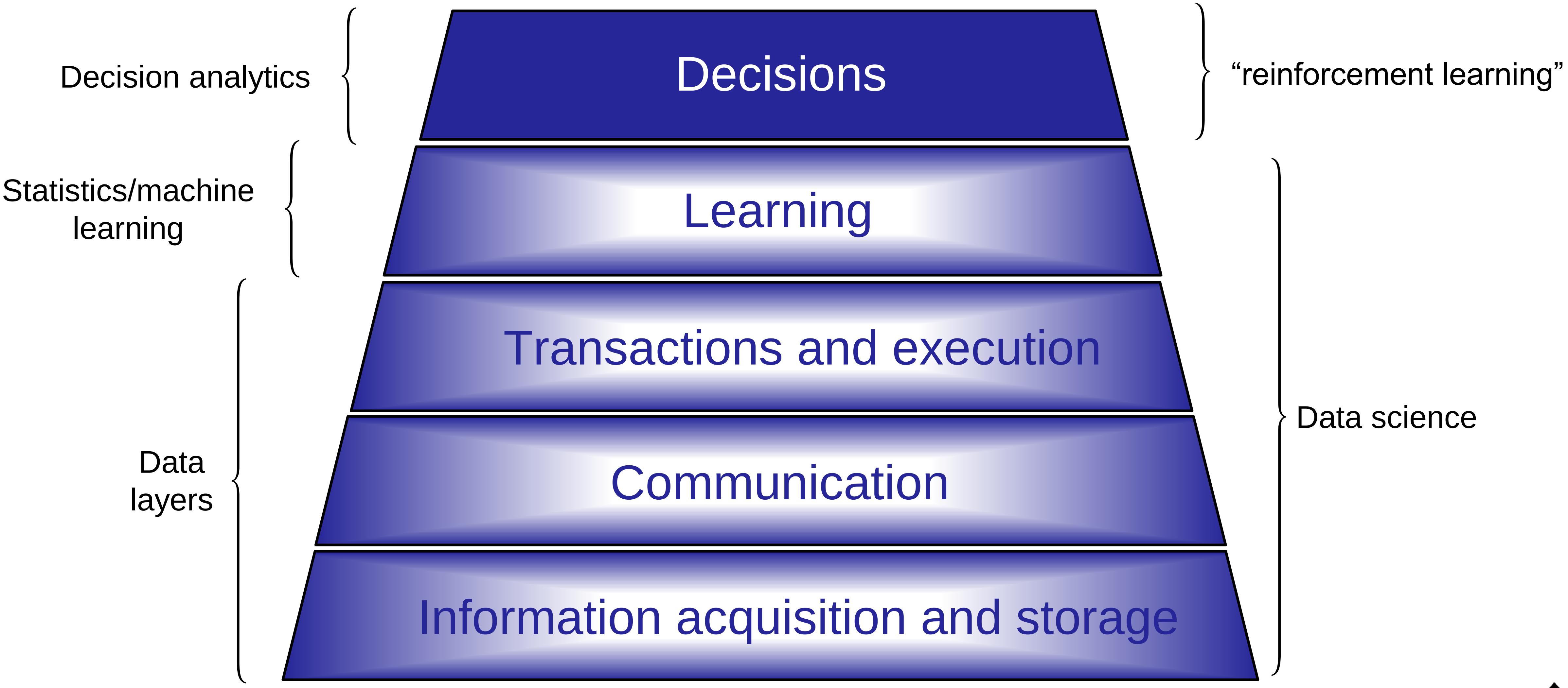


"Big dataset"

Searching over statistical models

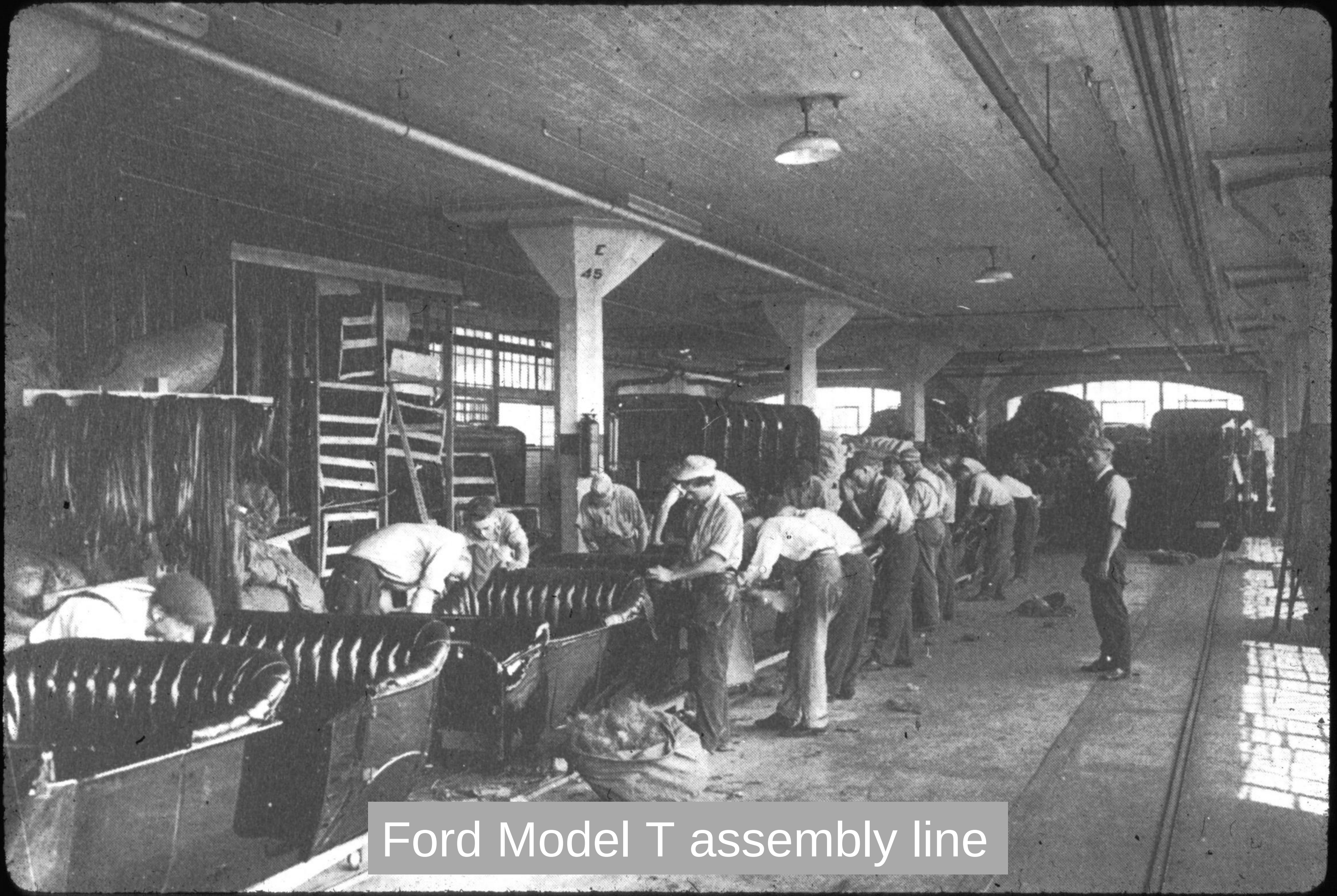
These consist of functions $f \in \mathcal{F}$
and tunable parameters $\theta \in \Theta^f$

THE 5 LAYERS OF INTELLIGENCE



Optimal Dynamics

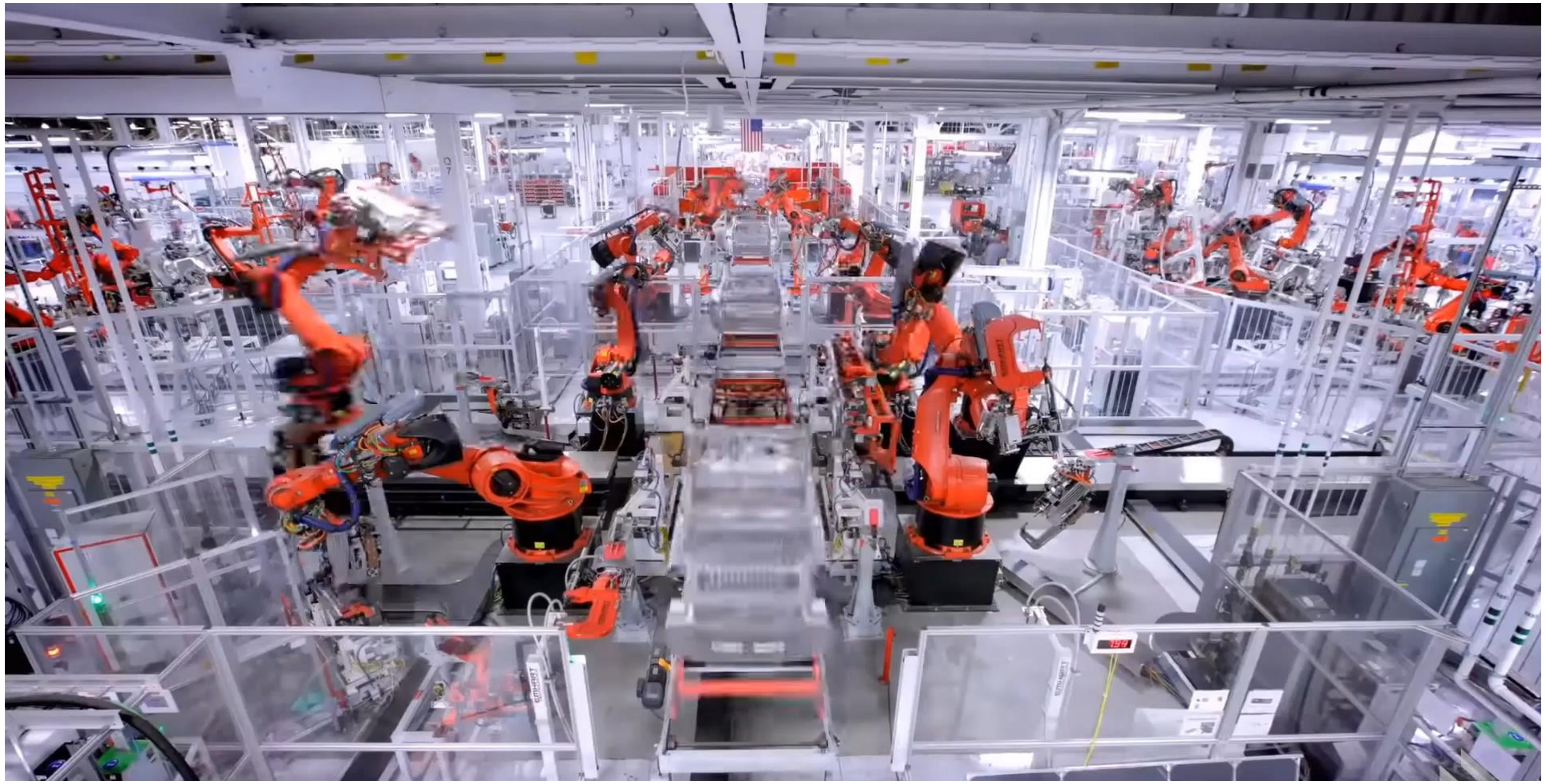
INFORMATION & DECISION PROCESSES



Ford Model T assembly line

Information and decision processes

- There are parallels between the process of making "decisions" and a manufacturing line making "products"



We have to approach information processing and decisions like a manufacturing process.


NASDAQ
NASDAQ COMPOSITE
S&P 500 Index
COVID 19
Coronavirus Vaccine
AIR PRODUCTS
tell me more
www.airproducts.com
9533
9533
cargolux
cargolux
E-COMMERCE

DECISIONS

What price to accept for a spot load?

Which load to accept now to move next week

How many dedicated drivers should we have

Where should drivers be domiciled?

Which driver should move a load?

What is the best policy for high-frequency trading?

What contracts to sign for raw materials?

Which vendor should supply each part?

What is the value of financial option?

How many syringes should be sent to each vaccination site, and when?

When should inventory be ordered?

How much battery sto needed to handle the variability of wind?

What price should be charged

Which physician should handle a procedure?

How many nurses should we have to serve local hospitals

When should I refill the customer's tank with liquid nitrogen

Which nurse should visit this doctor's office today?

When should gas turbines be scheduled to handle drops in wind?

Which customer tanks should we fill when we are in the area?

Where should a patient be assigned for specific treatment?

How many suppliers should you have for a particular part, and where?

Which material handling jobs should be done by robots, and which robot?

What bid should we place on Google for a set of ad-words?

How much energy should I purchase from the wind farm?

Which supplier should manufacture turbine blades?

How many jet engines should be made each day?

When should inventory be refilled at a fulfillment center?

Which fulfillment center should handle an order?

Optimal Dynamics

DECISIONS

Types of decisions.

Physical Decisions

» Managing inventories

» Assigning drivers and moving trucks

» Scheduling nurses and energy generators

Financial Decisions

» Pricing decisions

» Insurance decisions

» Managing investments

» Hedging contracts

Informational Decisions

» Sending/receiving information

» Marketing and advertising

» Running experiments (lab or field)

» Testing drugs

THE TIME FRAMES FOR DECISIONS

Strategic planning and design – We simulate operational decisions so we understand how a system would respond to decisions far in the future:

- » Where to source parts
- » How much production capacity to have
- » What markets to serve

Tactical planning decisions – We simulate operational decisions to help make decisions that impact the system in the near future,

- » What orders to place now for delivery in the future
- » Pricing decisions
- » Personnel scheduling

Real-time decisions – These are decisions that impact the system now:

- » Which driver should move a load of freight right now
- » Which production lines should be running today
- » Spot-pricing decisions

Who is making the decisions

C-suite decisions – Strategic decisions covering:

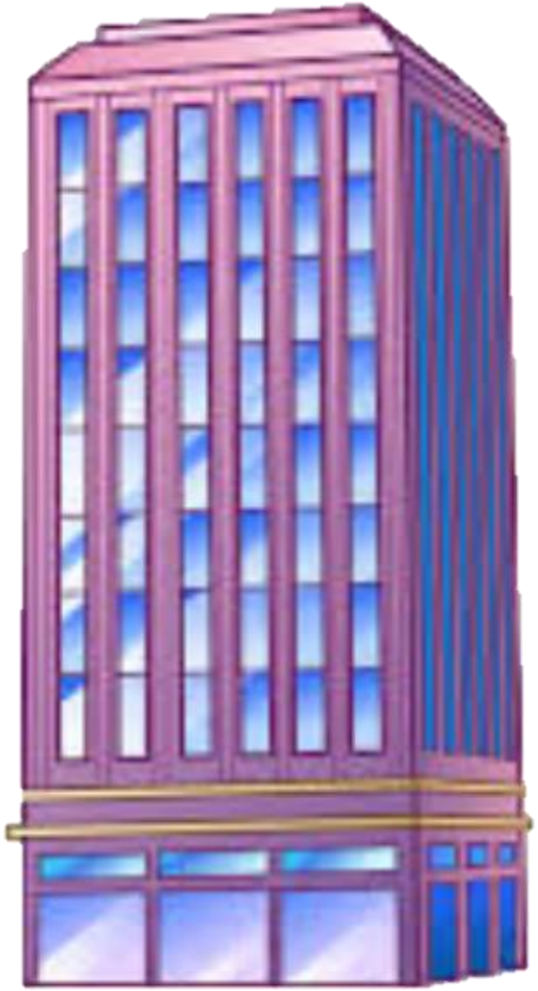
- Which products are being made, and where.
- How much production capacity.
- Which markets to enter?
- Top-line budgets for people, equipment, marketing, …

Middle management – Tactical planning decisions:

- Inventory planning
- Pricing, marketing and advertising
- Staffing, equipment distribution
- Setting performance metrics

Field operations – Day-to-day decisions such as:

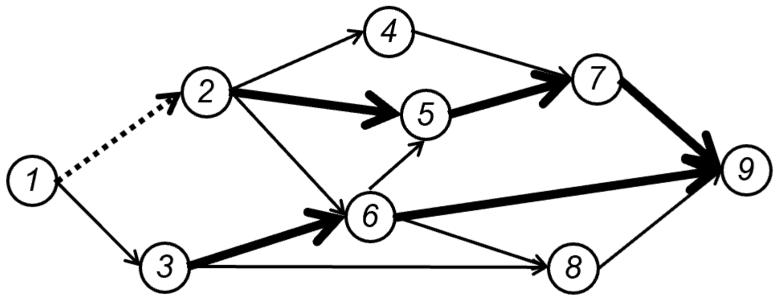
- Scheduling people and equipment
- Assigning jobs to people
- Dispatching trucks



DETERMINISTIC OPTIMIZATION

Low dimensional decisions

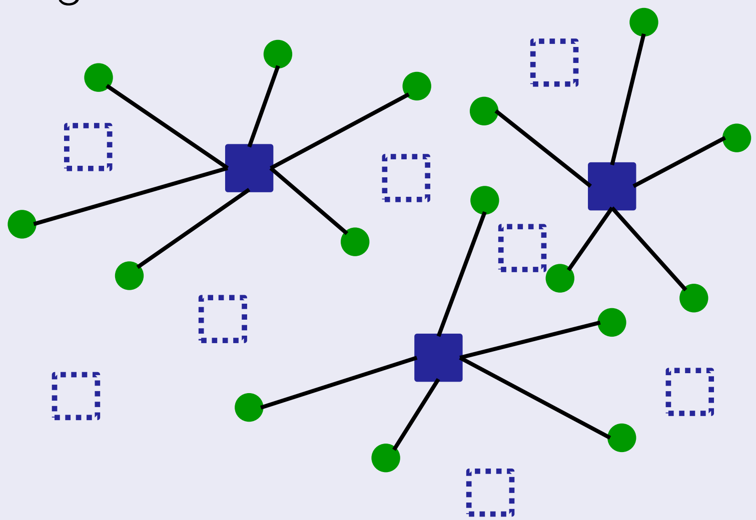
Planning a path to your destination



$$x_{ij} = \begin{cases} 1 & \text{If we move from node } i \text{ to node } j \\ 0 & \text{Otherwise} \end{cases}$$

High dimensional decisions

Optimizing facility locations



$$x_{i} = \begin{cases} 1 & \text{If we locate a facility at location } i \\ 0 & \text{Otherwise} \end{cases}$$

Optimal Dynamics

DETERMINISTIC OPTIMIZATION

Airline scheduling

Airlines | **Optimization Model** | **Airline Schedule**

$$\min_x cx$$
$$Ax = b$$
$$x \geq 0$$

Airlines around the world use tools that depend on this mathematical model to perform strategic and operational planning.

DETERMINISTIC OPTIMIZATION

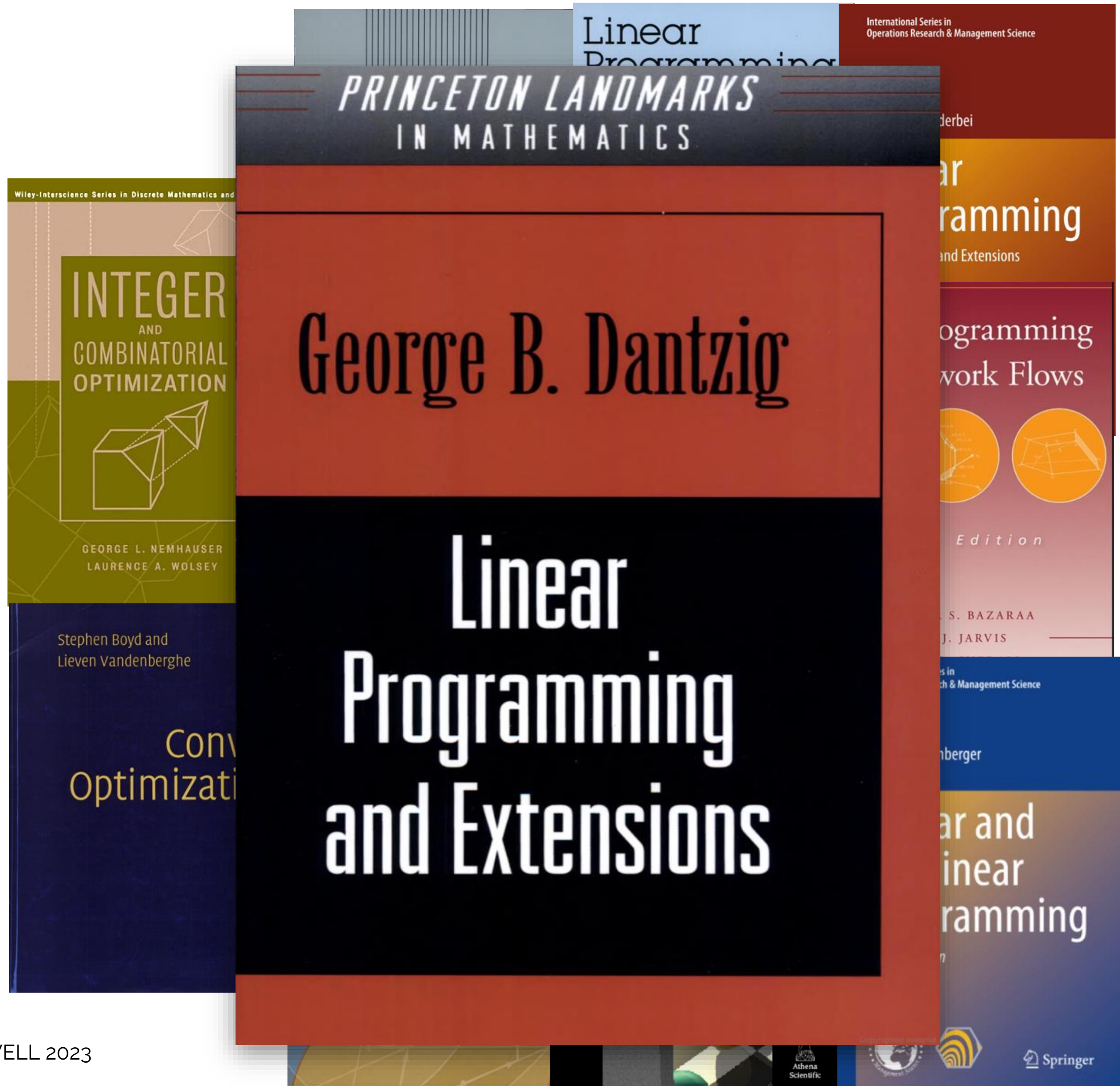
The language of deterministic optimization

$$\min_x cx$$

$$Ax = b$$

$$x \geq 0$$

- Spoken around the world.
- Many books communicate the same core theory
- Computer packages are available to solve realistic problems
- Many graduate programs producing thousands of students each year.




NASDAQ
NASDAQ COMPOSITE
S&P 500 Index
AIR PRODUCTS
tell me more
www.airproducts.com
COVID 19
Coronavirus Vaccine
9533
9533
cargolux
cargolux
E-COMMERCE

DECISIONS

What price to accept for a spot load?

Which load to accept now to move next week

Which driver should move a load?

What is the best policy for high-frequency trading?

What contracts to sign for raw materials?

Which vendor should supply each part?

What is the value of a financial option

Where should drivers be domiciled?

How many dedicated drivers should we have?

How many syringes should be sent to each vaccination site, and when?

When should inventory be ordered?

How much battery storage is needed to handle the variabilit of wind?

What price should be charged

Which physician should handle a procedure?

How many nurses should we have to se local hospitals and doctor's offices?

When should I refill the customer's tank with liquid nitrogen

When should gas turbines be scheduled to handle drops in wind?

Which nurse should visit this doctor's office today?

Which customer tanks should we fill when we are in the area?

Where should a patient be assigned for specific treatment?

How many suppliers should you have for a particular part, and where?

Which material handling jobs should be done by robots, and which robot?

What bid should we place on Google for a set of ad-words?

How much energy should I purchase from the wind farm?

Which supplier should manufacture turbine blades?

How many jet engines should be made each day?

When should inventory be refilled at a fulfillment center?

Which fulfillment center should handle an order?

INFORMATION

Market prices for spot freight

Driver requests for loads; time-at-home requests

Changes in asset prices

Prices of raw materials by region

Quality of orders provided by a vendor

Offered loads by shipper, by lane

Production delays in order fulfillme

New COVID-19 cases by county

Employment rate; unemployment filings

Transit delays

Driver application for jobs by region

Patient arrivals and symptoms

Competitor prices

Wind generation from a wind farm

Requests for nurses from doctor's offices

Customer usage rate of liquid nitrogen

Electricity prices on the grid

Number of nurses calling in sick

Equipment failures at customer nitrogen tanks

Availability of specialists to treat a condition

Capacity shutdowns at suppliers due to labor or political problems

Flow of different parts to each machining station

Whether a bid wins an ad-click auction

Lead times required by each manufacturer

The amount of energy that is generated from wind.

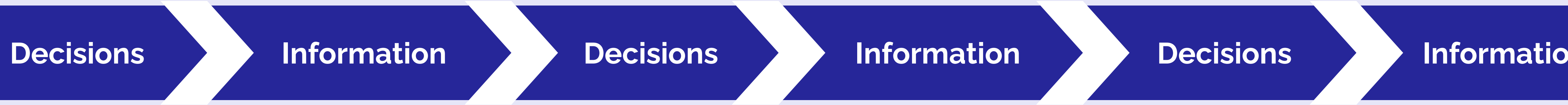
Daily production of new jet engines

Flow of orders for a product by region around the country

Orders for a product from different regions

SEQUENTIAL DECISIONS

In most settings, decisions are made over time…

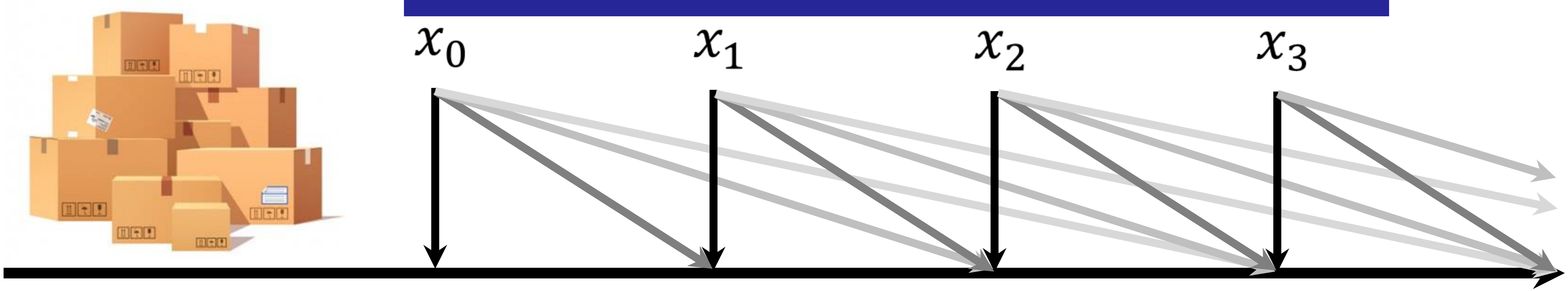


Information that arrives after a decision is made is not known when we made the decision.

SEQUENTIAL DECISIONS

Inventory management

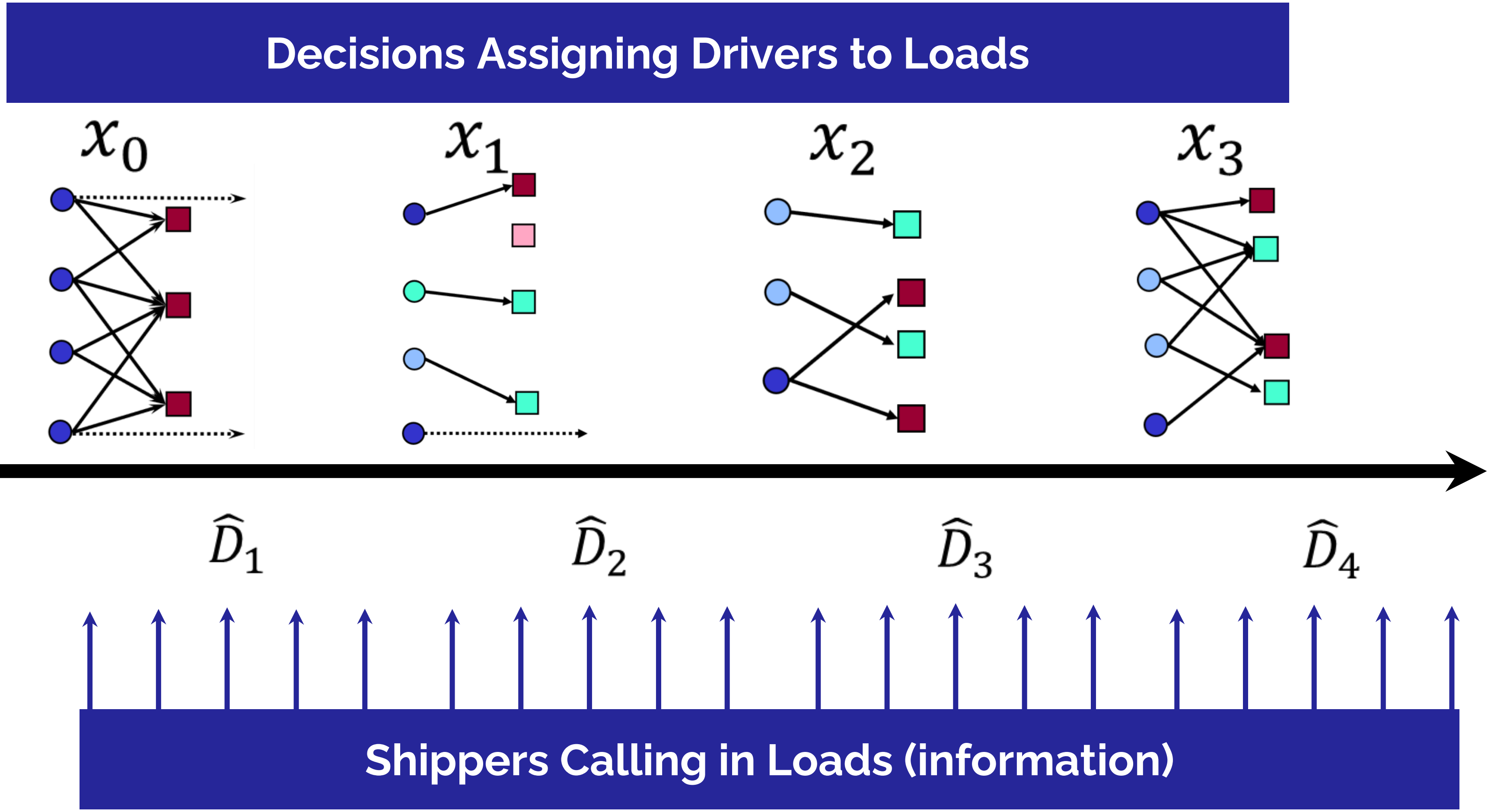
Inventory Ordering Decisions



Customer Demands (information)

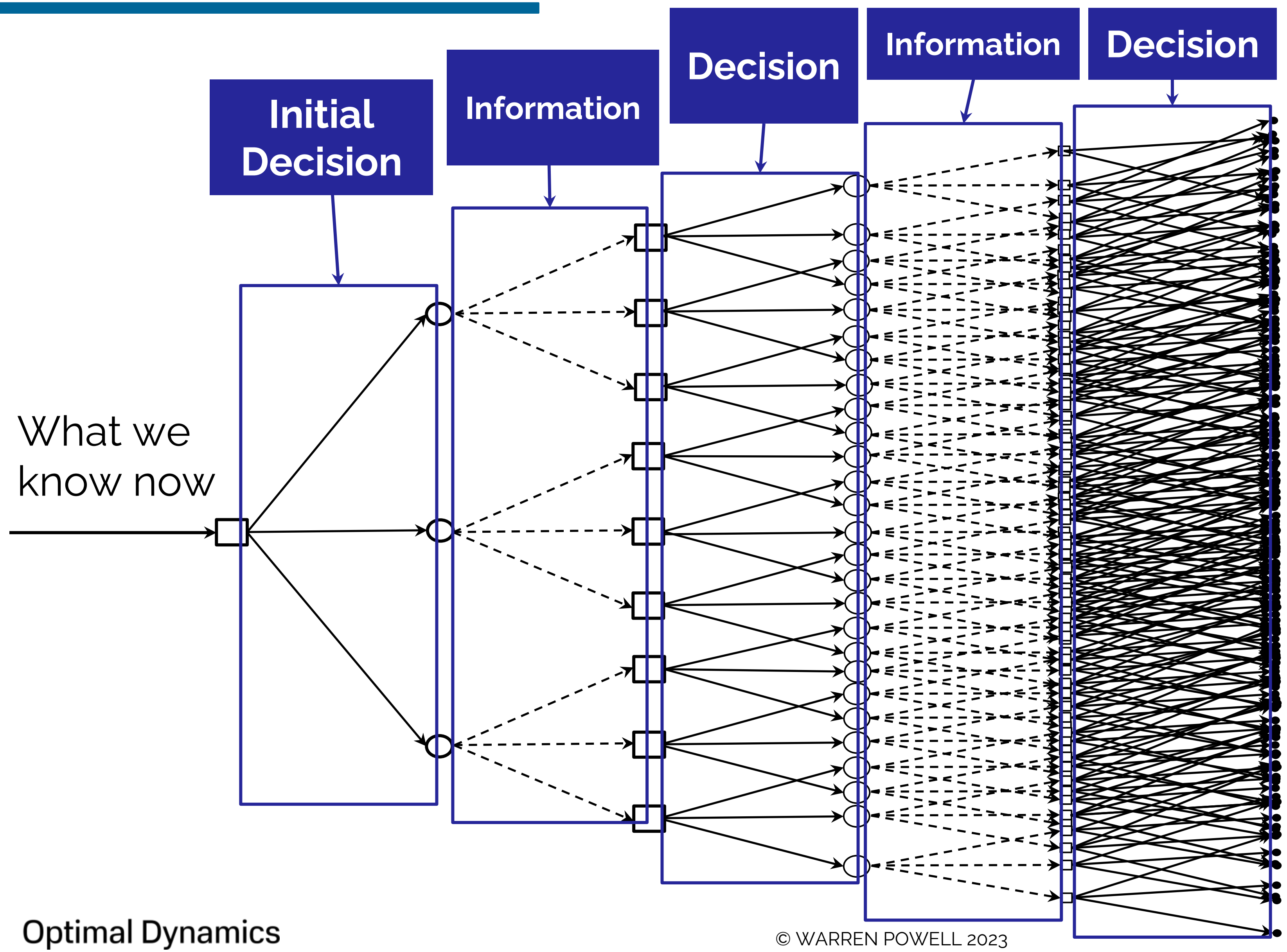
SEQUENTIAL DECISIONS

Driver dispatch for truckload trucking



Optimal Dynamics

SEQUENTIAL DECISIONS



Even small sequential decision problems explode dramatically as we plan into the future

OUTLINE

- → The five layers of intelligence
- → Modeling sequential decision problems
- → Modeling uncertainty
- → Designing policies
- → A new educational field: sequential decision analytics

MODELING SEQUENTIAL DECISION PROBLEMS

The biggest challenge when making decisions under uncertainty is ***modeling***.

Everyone writing out a deterministic optimization model, or machine learning model, knows how to write out their problem mathematically…

$$\text{Min } E\left\{\sum cx\right\}$$
$$Ax = b$$
$$x \geq 0$$

Organize class libraries, and set up communications and databases

Mathematical model

…we lack a standard modeling framework for sequential decisions.


Stochastic programming
Robust optimization
Decision analysis
Approximate dynamic programming
Simulation optimization
Dynamic Programming and control
Optimal learning
Model predictive control
Optimal control
Stochastic search
Bandit problems
Stochastic optimization
Reinforcement learning
Markov decision processes
Stochastic control
Online computation


John R. Birge
François Louveaux
Introduction to Stochastic Programming
Second Edition
Michael C. Fu Editor
Handbook of Simulation Optimization
Princeton Series in APPLIED MATH
Robust Optimization
Introduction to Decision Analysis
A Practitioner's Guide to
SECOND EDITION
Approximate Dynamic Programming
Solving the Curses of Dimensionality
Warren B. Powell
VOLUME 2 • 4th EDITION
Dynamic Programming and Optimal Control
APPROXIMATE DYNAMIC PROGRAMMING
Dimitri P. Bertsekas
WILEY
INTRODUCTION TO STOCHASTIC SEARCH AND OPTIMIZATION
Estimation, Simulation, and Control
JAMES C. SPALL
Wiley Series in Probability and Statistics
Optimal Learning
Warren B. Powell
Ilya O. Ryzhov
MULTI-ARMED BANDIT ALLOCATION INDICES
SECOND EDITION
John Gittins, Kevin Glazebrook and Richard Weber
SECOND EDITION
Model Predictive Control
nd C. Bordons
OPTIMAL CONTROL
Frank L. Lewis
Draguna L. Vrabie
Vassilis L. Syrmos
43
Jiongmin Yong
Xun Yu Zhou
Stochastic Controls
Hamiltonian Systems and HJB Equations
ations of Mathematics
Modelling and Applied Probability
Reinforcement Learning
An Introduction
second edition
Richard S. Sutton and Andrew G. Barto
Markov Decision Processes
Discrete Stochastic Dynamic Programming
MARTIN L. PUTERMAN
LECTURES ON STOCHASTIC PROGRAMMING
Modeling and Theory
Alexander Shapiro
Darinka Dentcheva
Andrzej Ruszczyński
MOS-SIAM Series on Optimization
Online Computation and Competitive Analysis
Allan Borodin
Ran El-Yaniv

Modeling sequential decision problems

- Any sequential decision problem can be written:

$$(S_0, x_0, W_1, S_1, x_1, W_2, \ldots, S_t, x_t, W_{t+1}, S_{t+1} \ldots, S_T)$$

What we know (or believe) — **The decision** — **What we observe (or learn)**

- Each time we make a decision, we receive a contribution $C(S_t, x_t)$.
- Decisions are made with a method or *policy* $X^\pi(S_t)$ which we design later.
- State variables evolve using a transition function: $S_{t+1} = S^M(S_t, x_t, W_{t+1})$.
- The goal is to find the policy that maximizes expected contributions:

$$\max_\pi \mathbb{E}\{\sum_{t=0}^T C(S_t, X^\pi(S_t))|S_0\}$$

Modeling sequential decision problems

Every sequential decision problem can be modeled using 5 core components

- State variables $S_t = (R_t, I_t, B_t)$
  - Physical state R_t, other information I_t, beliefs B_t.
- Decision variables x_t (or action a_t, or control u_t)
  - Decisions x_t are determined by a policy $X^\pi(S_t)$.
- Exogenous variables W_{t+1}
  - This is new information that arrives between t and $t+1$.
- Transition function $S_{t+1} = S^M(S_t, x_t, W_{t+1})$
  - This is how our state variable evolves given x_t and W_{t+1}.
- Objective function for finding the best policy
  - $\max_\pi E\{\sum_{t=0}^{T} C(S_t, X^\pi(S_t)) | S_0\}$

These five elements describe any sequential decision problem.

Optimal Dynamics

Modeling sequential decision problems

The complete model:

- Objective function
  - Cumulative reward ("online learning")

$$\max_\pi \mathbb{E}\left\{\sum_{t=0}^{T} C_t(S_t, X_t^\pi(S_t))\,|S_0\right\}$$

  - Final reward ("offline learning")

$$\max_\pi \mathbb{E}\{F(x^{\pi,N}, \widehat{W})|S_0\}$$

  - Risk:

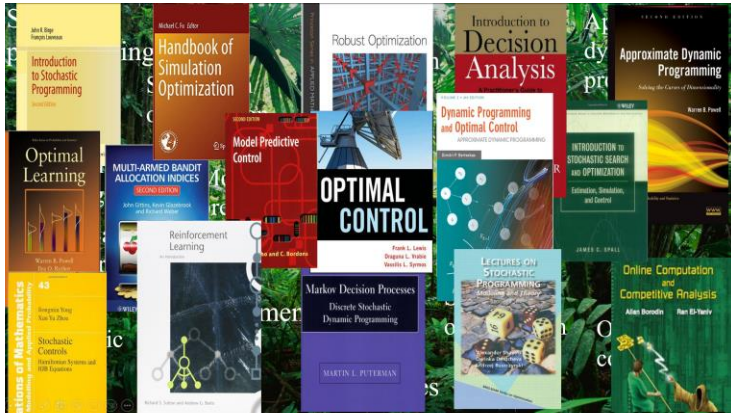
$$\max_\pi \rho\{C(S_0, X_0^\pi(S_0)), C(S_1, X_1^\pi(S_1)), \ldots, C(S_T, X_T^\pi(S_T))|S_0\}$$

- Transition function:

$$S_{t+1} = S^M(S_t, x_t, W_{t+1})$$

- Exogenous information:

$$(S_0, W_1, W_2, \ldots, W_T)$$



Optimal Dynamics

Modeling sequential decision problems

Optimizing over policies

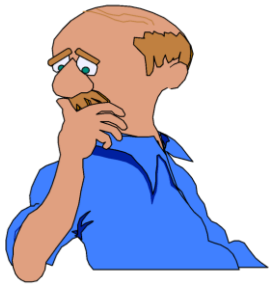
$$max_{\pi=(f\in F,\theta\in\Theta^f)} \sum_{t=0}^{T} C_t(S_t(\omega), X_t^\pi(S_t(\omega)))$$

$$S_t = S^M(S_t, x_t, W_{t+1})$$

Type of policy (structure of function)

Algorithmic tuning

Computer simulation



$$\theta^{n+1} = \theta^n + \alpha_n \nabla_\theta F(\theta^n, W^{n+1})$$

Optimal Dynamics

Evaluating policies

1) Theoretically

- Optimality proofs
- Regret bounds
- Asymptotic convergence

2) Through numerical simulations

3) In the field

Optimal Dynamics

Modeling sequential decision problems

Application

Step 1
Identify:

- Performance metrics
- Types of decisions
- Sources of uncertainty

Step 2:
Mathematical model:

» State variables $S_t = (R_t, I_t, B_t)$
- Physical state R_t, other information I_t, belief state B_t.

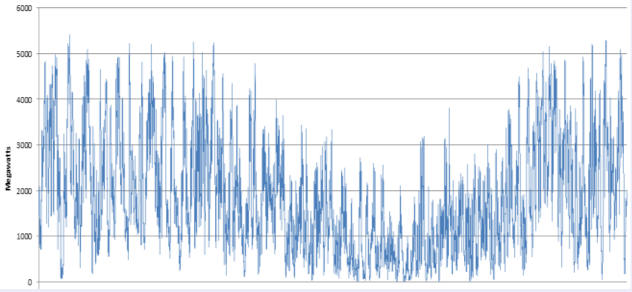
» Decision variables (x_t, a_t, u_t)
- Made with *policy* $X^\pi(S_t|\theta)$ (or $A^\pi(S_t)$ or $U^\pi(S_t)$)

» Exogenous information W_{t+1}
- What do we learn for the first time between t and $t+1$?

» Transition function $S_{t+1} = S^M(S_t, x_t, W_{t+1})$
- How do the state variables evolve over time?

» Objective function
- $\max_\pi \mathbb{E}_{S_0} \mathbb{E}_{W_1,\ldots,W_T|S_0} \sum_{t=0}^{T} C(S_t, X^\pi(S_t))$

Step 3:
Uncertainty modeling



Step 4:
Designing policies

$$\max_\pi \mathbb{E}\left\{\sum_{t=0}^{T} C(S_t, X^\pi(S_t))|S_0\right\}$$

Step 5:
Computer model

Step 6:
Implementation/analysis

OUTLINE

- → The five layers of intelligence
- → Modeling sequential decision problems
- → Modeling uncertainty
- → Designing policies
- → A new educational field: sequential decision analytics

Modeling uncertainty

Language of models

12 Classes of uncertainty

- » Observational uncertainty
- » Prognostic uncertainty (forecasting)
- » Experimental noise/variability
- » Transitional uncertainty
- » Inferential uncertainty
- » Model uncertainty
- » Systematic exogenous uncertainty
- » Control/implementation uncertainty
- » Communication errors/biases
- » Algorithmic noise
- » Goal uncertainty
- » Environmental uncertainty

Language of the problem domain

- » Suppliers:
  - Daily production, yield
  - Future commitments
  - Delivery times
  - Costs
- » Market/customers
  - Orders, returns
  - Price paid
  - Service requirements
- » Personnel
  - Availability
  - Departures, hiring
  - Performance
- » Equipment
  - Up-time, failures
  - Productivity
- » Network
  - Transit times
  - Weather, earthquakes

OUTLINE

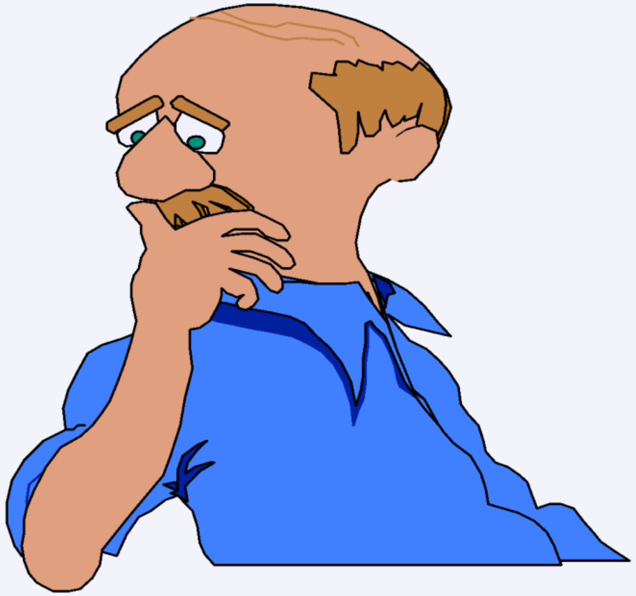
- → The five layers of intelligence
- → Modeling sequential decision problems
- → Modeling uncertainty
- → Designing policies
- → A new educational field: sequential decision analytics

Designing policies

What is a policy?

A policy is method that makes a decision using the information in the state variable.

… *any method.*



Designing policies

Policies and the English language

Algorithm	Formula	Prejudice
Behavior	Grammar	Principle
Belief	Habit	Procedure
Bias	Heuristics	Process
Canon	Laws/bylaws	Protocols
Code	Manner	Recipe
Commandment	Method	Ritual
Conduct	Mode	Rule
Control law	Mores	Strategy
Convention	Norm	Style
Culture	Orthodox	Syntax
Customs	Patterns	Technique
Duty	Plans	Template
Etiquette	Policies	Tenet
Fashion	Practice	Tradition
Format	Precedent	Way of life

http://tinyurl.com/policiesanddecisions

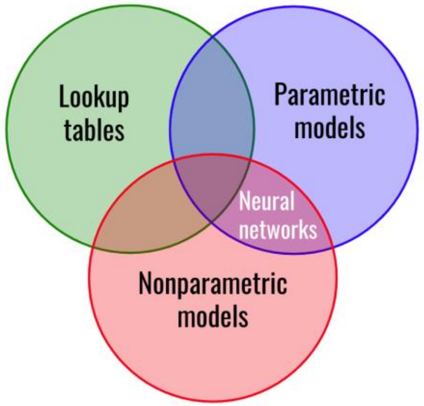
BRIDGING MACHINE LEARNING & SEQUENTIAL DECISIONS

Machine learning

$$\min_{f\in F,\theta\in\Theta^f} \frac{1}{N}\sum_{n=1}^{N}\left(y^n - f(x^n|\theta)\right)^2$$

Searching over functions

"Big dataset"



Sequential decisions

$$\max_{\pi=(f\in F,\theta\in\Theta^f)} \frac{1}{N}\sum_{n=1}^{N}\sum_{t=0}^{T} C(S_t^n, X^\pi(S_t^n|\theta))$$

$$S_{t+1} = S^M(S_t, x_t, W_{t+1})$$

Searching over policies

System model

Designing policies

There are two fundamental strategies for designing policies

Policy search – Search over a class of methods for making decisions to optimize some metric over time.

» Finding the best class of policy.

» Finding the best policy within the class.

Lookahead approximations – Approximate the impact of a decision now on the future.

» The contribution from the first period, plus

» An approximation of the sum of contributions in future time periods resulting from the first decision.

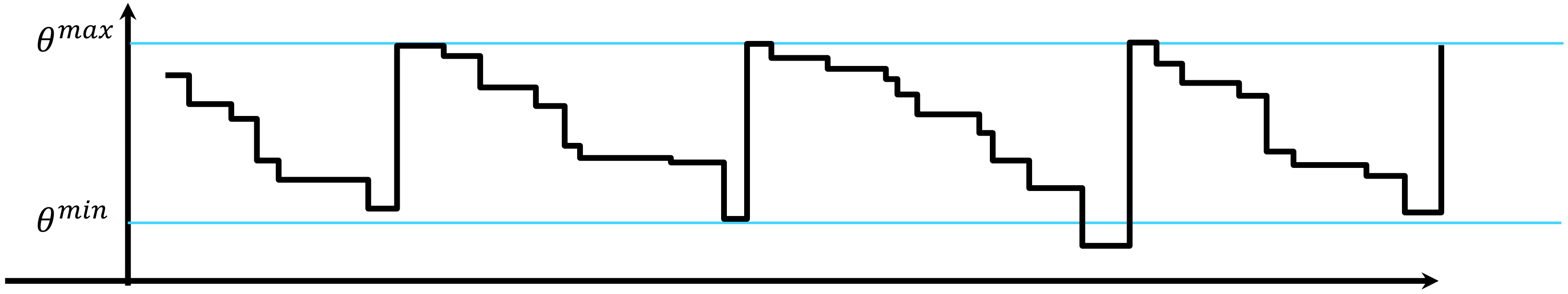
Policy search

1) Policy function approximation (PFA)

These are analytical functions that specify what to do given what we know.

Examples:

a) Order-up-to inventory policy $\theta = (\theta^{min}, \theta^{max})$



b) Buy when the price goes **below $\boldsymbol{\theta^{min}}$** and sell when it goes **above $\boldsymbol{\theta^{max}}$**

c) Lookup tables, linear/nonlinear models, neural networks, nonparametric models, … any function we might use in machine learning.

Policy search

2) Cost function approximations (CFAs)

These are parameterized optimization problems:

a) Find the shortest path to a destination, but add a buffer θ (e.g. 15 minutes) to make sure you arrive on time.

b) Schedule drivers for $\theta = 32$ hours per week, which allows for unforeseen delays.

c) Advertise the product x which solves:

$$X^{UCB}(S^n|\theta) = \arg\max_x (Estimated\ revenue_x^n + \theta \cdot Standard\ deviation\ of\ estimated\ revenue_x^n)$$

Parametric CFAs are widely used in industry yet dismissed by the academic research community. This is actually quite a powerful strategy.

Cost function approximations

- Inventory management
  - » How much product should I order to anticipate future demands?
  - » Need to accommodate different sources of uncertainty.
    - Market behavior
    - Transit times
    - Supplier uncertainty
    - Product quality

Cost function approximations

- Imagine that we want to purchase parts from different suppliers. Let x_{tp} be the amount of product we purchase at time t from supplier p to meet forecasted demand D_t. We would solve

$$X_t^\pi(S_t) = \text{argmax}_{x_t \in X_t} \sum_{p \in P} c_p x_{tp}$$

subject to

$$\left.\begin{aligned} \sum_{p \in P} x_{tp} &\geq D_t \\ x_{tp} &\leq u_p \\ x_{tp} &\geq 0 \end{aligned}\right\} X_t$$

  » This assumes our demand forecast D_t is accurate.

Cost function approximations

- Imagine that we want to purchase parts from different suppliers. Let x_{tp} be the amount of product we purchase at time t from supplier p to meet forecasted demand D_t. We would solve

$$X_t^{\pi}(S_t|\theta) = \mathrm{argmax}_{x_t \in X_t} \sum_{p \in P} c_p x_{tp}$$

subject to

$$\left.\begin{aligned} \sum_{p \in P} x_{tp} &\geq \theta^{reserve} D_t \\ x_{tp} &\leq u_p \\ x_{tp} &\geq \theta^{buffer} \end{aligned}\right\} X_t$$

Reserve

Buffer stock

  - » This is a *parametric cost function approximation*.

Cost function approximations

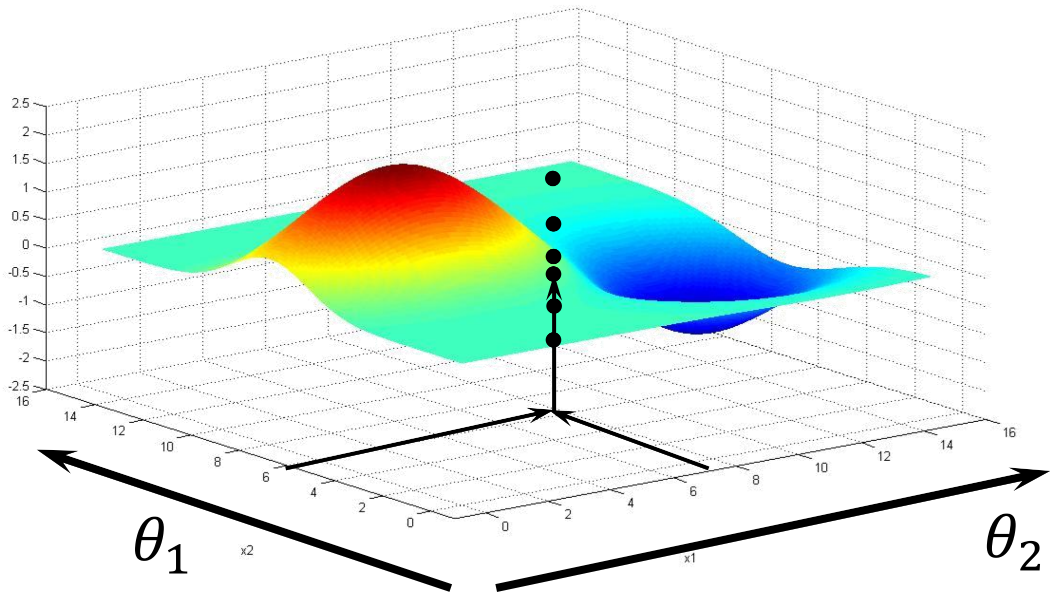
- Other applications
  - » Airlines optimizing schedules with schedule slack to handle weather uncertainty.
  - » Manufacturers using buffer stocks to hedge against production delays and quality problems.
  - » Grid operators scheduling extra generation capacity in case of outages.
  - » Adding time to a trip planned by Google maps to account for uncertain congestion.

  - » See: https://tinyurl.com/cfapolicy for an introduction to CFAs.

Policy search

- Both PFAs and CFAs have tunable parameters θ which have to be tuned. We write this mathematically as

$$\max_{\theta} \mathbb{E}\left\{\sum_{n=1}^{N} C(S^n, X^\pi(S^n|\theta))\,|S_0\right\}$$



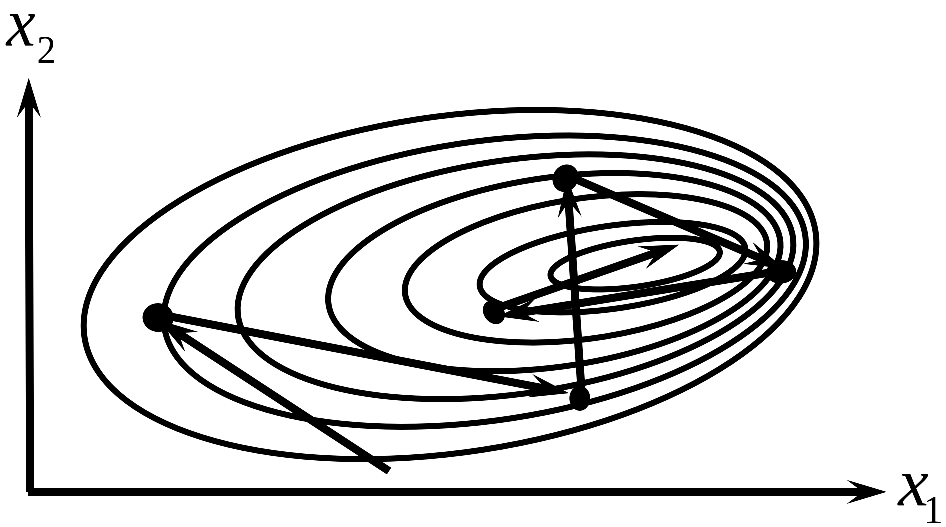
- There are two ways to evaluate a policy:
  - » In a simulator – This allows us to perform extensive testing in a controlled environment.
  - » In the field – This is "learning by doing"

Policy function approximations

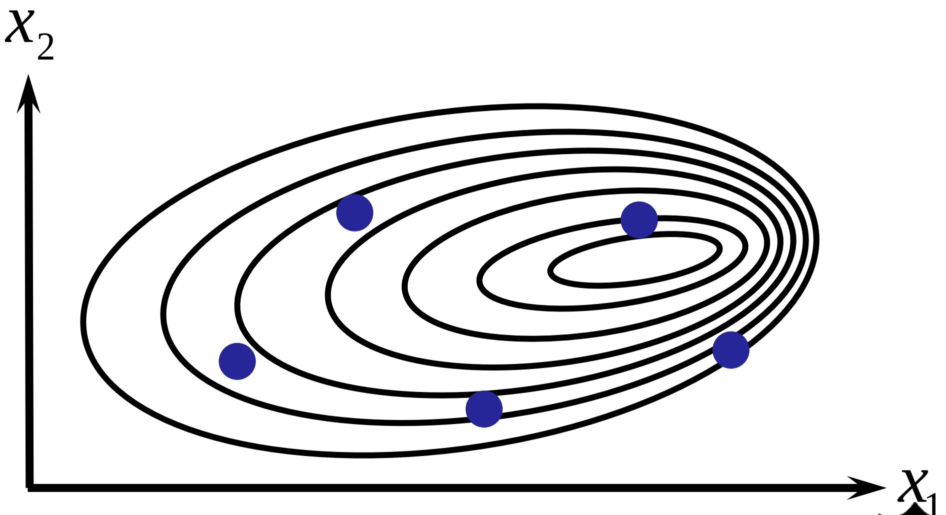
- How do we search for the best θ?
  - » Derivative-based
    - Stochastic gradient methods:

$$\theta^{n+1} = \theta^n + \alpha_n \nabla_\theta F(\theta^n, W^{n+1})$$

Decision



  - » Derivative-free
    - Build a belief model $\bar{F}(\theta) \approx \mathbb{E}F(\theta, W)$ that approximates our function.



  - » Both of these approaches are sequential decision problems!

Designing policies

There are two fundamental strategies for designing policies

Policy search – Search over a class of methods for making decisions to optimize some metric over time.

» Finding the best class of policy.

» Finding the best policy within the class.

Lookahead approximations – Approximate the impact of a decision now on the future.

» The contribution from the first period, plus

» An approximation of the sum of contributions in future time periods resulting from the first decision.

Lookahead approximations

- Lookahead approximations combine:
  - » The immediate contribution (or cost) of a decision made now…
  - » … and an approximation of future contributions (or costs)

Lookahead approximations

Lookahead policies are based on solving

$$X_t^*(S_t) = argmax_x \left(C(S_t, x_t) + \mathbb{E}\left\{ \max_\pi \left\{ \mathbb{E} \sum_{t'=t+1}^{T} C\left(S_t, X_{t'}^\pi(S_{t'})\right) | S_{t+1} \right\} | S_t, x_t \right\} \right)$$

Contribution we receive now

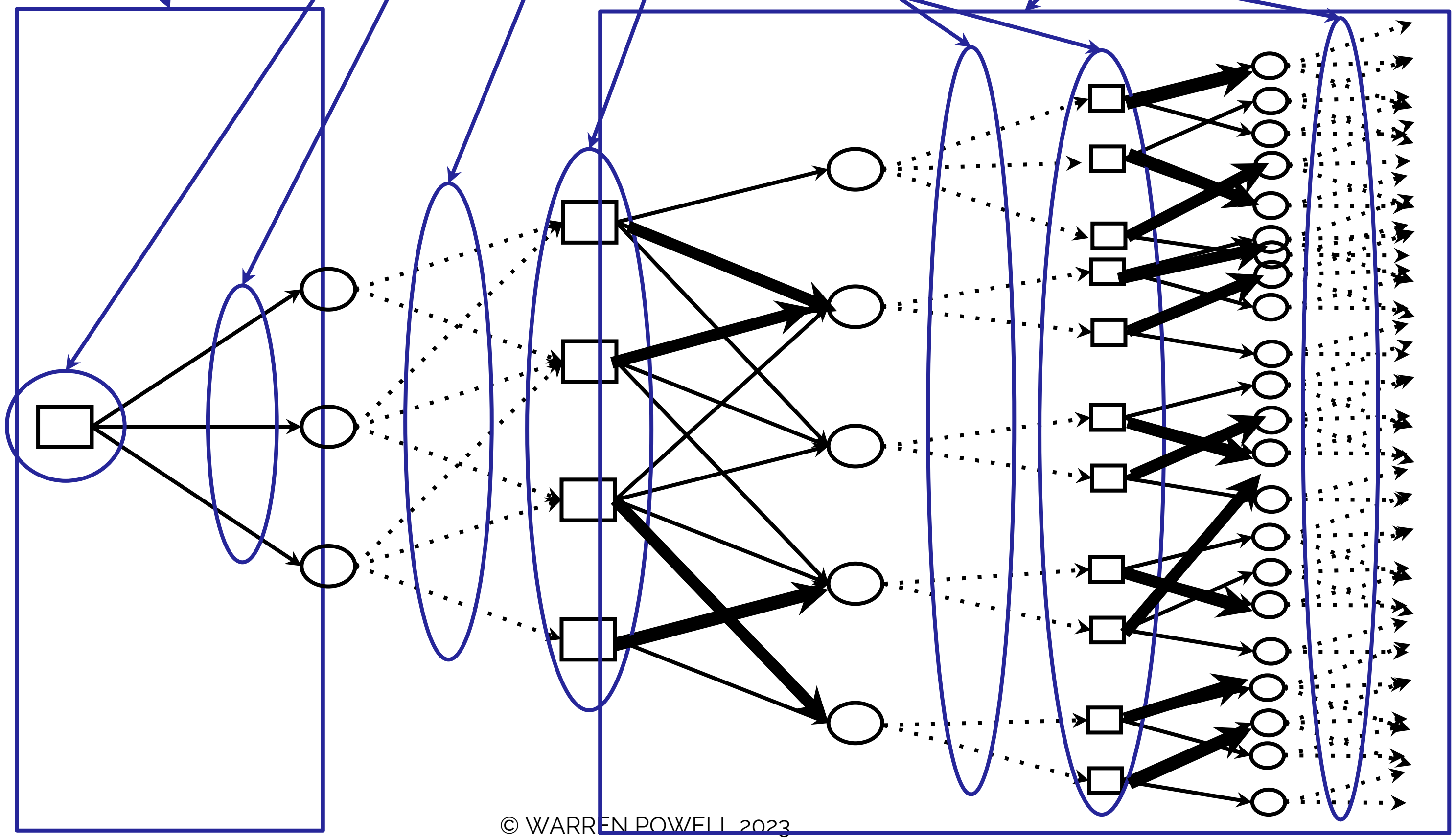
Future contributions

- This looks like scary mathematics, but it is what all of us are doing when we make decisions now that consider what might happen in the future.
- The challenge is … *how to compute it!!!*

Lookahead approximations

Lookahead policies are based on solving

$$X_t^*(S_t) = argmax_x \left(C(S_t, x_t) + \mathbb{E}\left\{ \max_\pi \left\{ \mathbb{E} \sum_{t'=t+1}^{T} C\left(S_t, X_{t'}^\pi(S_{t'})\right) | S_{t+1} \right\} | S_t, x_t \right\} \right)$$



Lookahead approximations

Lookahead approximations

Approximate the impact of a decision now on the future

$$X_t^*(S_t) = argmax_x \left(C(S_t, x_t) + \mathbb{E} \left\{ \max_\pi \left\{ \mathbb{E} \sum_{t'=t+1}^{T} C\left(S_t, X_{t'}^\pi(S_{t'})\right) |S_{t+1} \right\} |S_t, x_t \right\} \right)$$

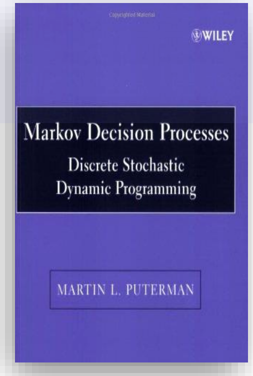
3) Value function approximations (VFAs)

$$X_t^*(S_t) = \arg\max_{x_t} \left(C(S_t, x_t) + \mathsf{E} \left\{ V_{t+1}(S_{t+1}) \,|\, S_t, x_t \right\} \right)$$

$$X_t^{VFA}(S_t) = \arg\max_{x_t} \left(C(S_t, x_t) + \mathsf{E} \left\{ \bar{V}_{t+1}(S_{t+1}) \,|\, S_t, x_t \right\} \right)$$

$$= \arg\max_{x_t} \left(C(S_t, x_t) + \bar{V}_t^x(S_t^x) \right)$$

$$= \arg\max_{x_t} \bar{Q}_t(S_t, x_t) \quad \text{("Q-learning")}$$



Lookahead approximations

Lookahead approximations

Approximate the impact of a decision now on the future

$$X_t^*(S_t) = argmax_x \left(C(S_t, x_t) + \mathbb{E}\left\{ \max_\pi \left\{ \mathbb{E} \sum_{t'=t+1}^{T} C\left(S_t, X_{t'}^{\pi}(S_{t'})\right) | S_{t+1} \right\} | S_t, x_t \right\} \right)$$

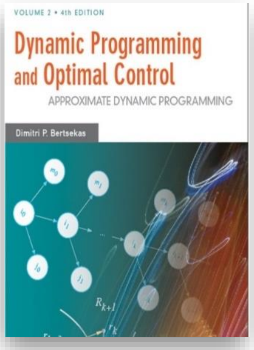
3) Value function approximations (VFAs)

$$X_t^*(S_t) = \arg\max_{x_t} \left(C(S_t, x_t) + E\left\{ V_{t+1}(S_{t+1}) \mid S_t, x_t \right\} \right)$$

$$X_t^{VFA}(S_t) = \arg\max_{x_t} \left(C(S_t, x_t) + E\left\{ \overline{V}_{t+1}(S_{t+1}) \mid S_t, x_t \right\} \right)$$

$$= \arg\max_{x_t} \left(C(S_t, x_t) + \overline{V}_t^x(S_t^x) \right)$$

$$= \arg\max_{x_t} \overline{Q}_t(S_t, x_t) \quad (\text{"Q-learning"})$$



Lookahead approximations

Lookahead approximations

Approximate the impact of a decision now on the future

$$X_t^*(S_t) = argmax_x \left(C(S_t, x_t) + \mathbb{E}\left\{ \max_\pi \left\{ \mathbb{E} \sum_{t'=t+1}^{T} C\left(S_t, X_{t'}^\pi(S_{t'})\right) | S_{t+1} \right\} | S_t, x_t \right\} \right)$$

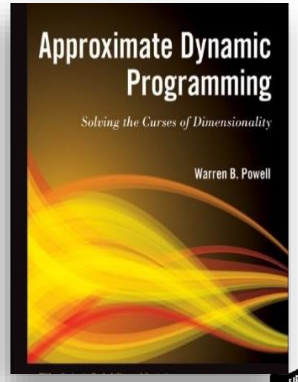
3) Value function approximations (VFAs)

$$X_t^*(S_t) = \arg\max_{x_t} \left(C(S_t, x_t) + \mathrm{E}\left\{ V_{t+1}(S_{t+1}) \mid S_t, x_t \right\} \right)$$

$$X_t^{VFA}(S_t) = \arg\max_{x_t} \left(C(S_t, x_t) + \mathrm{E}\left\{ \overline{V}_{t+1}(S_{t+1}) \mid S_t, x_t \right\} \right)$$

$$= \arg\max_{x_t} \left(C(S_t, x_t) + \overline{V}_t^x(S_t^x) \right)$$

$$= \arg\max_{x_t} \overline{Q}_t(S_t, x_t) \quad (\text{"Q-learning"})$$



Lookahead approximations

Lookahead approximations

Approximate the impact of a decision now on the future

$$X_t^*(S_t) = argmax_x \left(C(S_t, x_t) + \mathbb{E}\left\{ \max_\pi \left\{ \mathbb{E} \sum_{t'=t+1}^{T} C\left(S_t, X_{t'}^\pi(S_{t'})\right) | S_{t+1} \right\} | S_t, x_t \right\} \right)$$

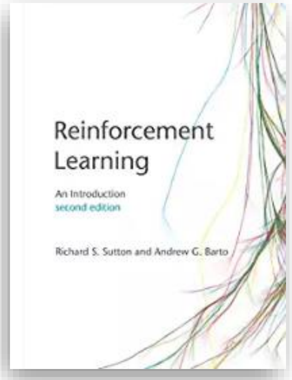
3) Value function approximations (VFAs)

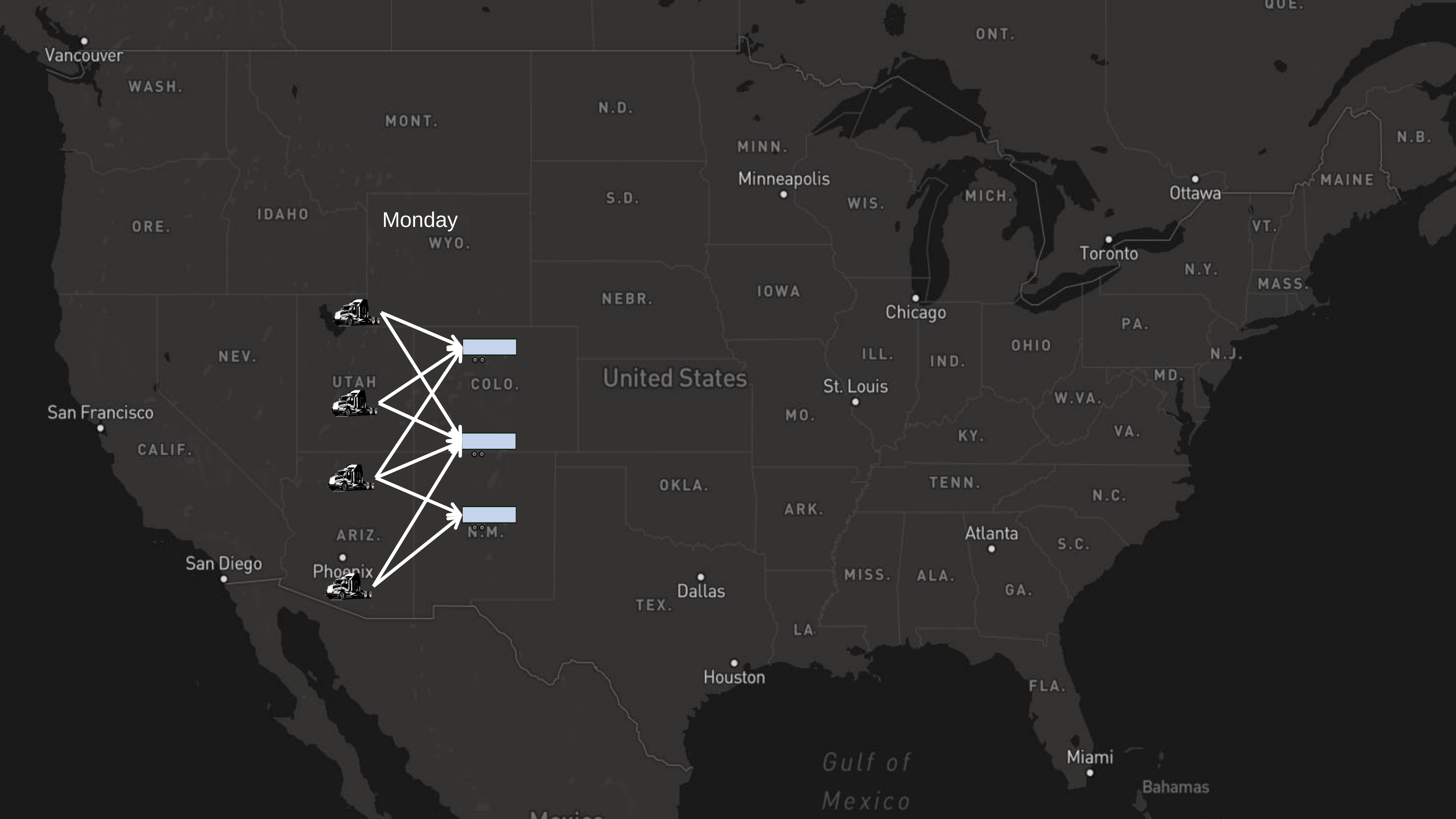
$$X_t^*(S_t) = \arg\max_{x_t} \left(C(S_t, x_t) + \mathrm{E}\left\{ V_{t+1}(S_{t+1}) \mid S_t, x_t \right\} \right)$$

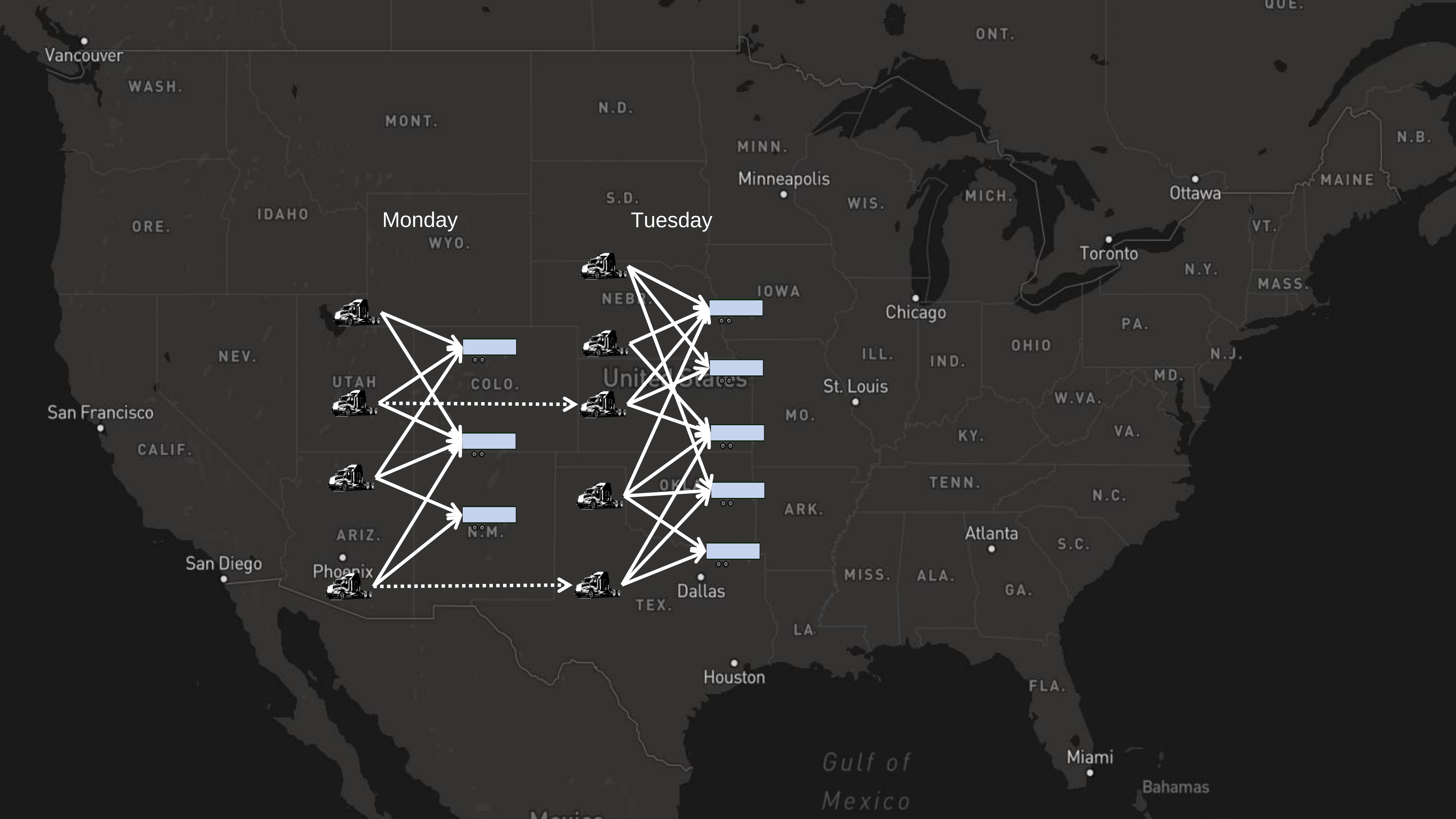
$$X_t^{VFA}(S_t) = \arg\max_{x_t} \left(C(S_t, x_t) + \mathrm{E}\left\{ \overline{V}_{t+1}(S_{t+1}) \mid S_t, x_t \right\} \right)$$

$$= \arg\max_{x_t} \left(C(S_t, x_t) + \overline{V}_t^x(S_t^x) \right)$$

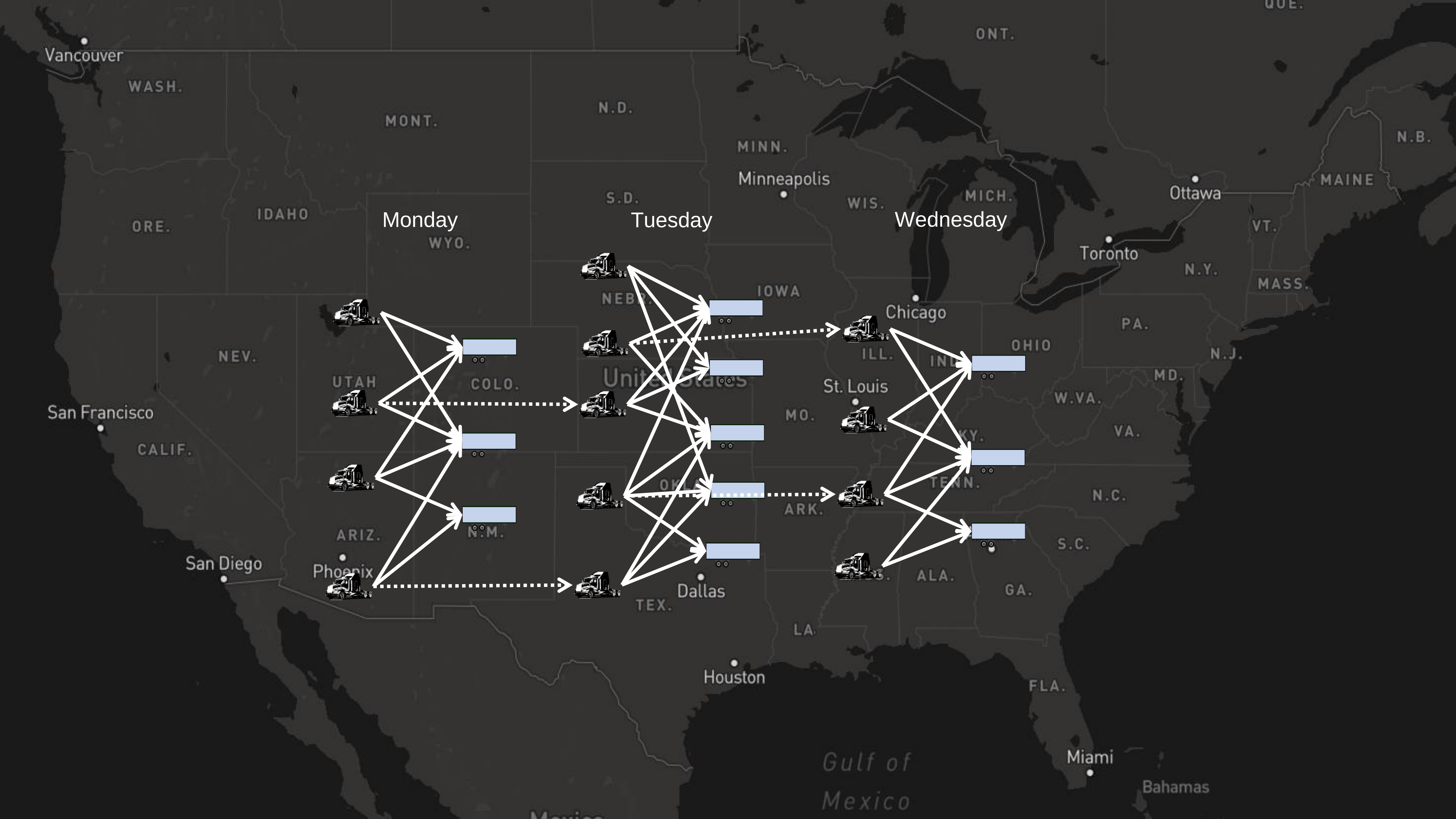
$$= \arg\max_{x_t} \overline{Q}_t(S_t, x_t) \quad (\text{"Q-learning"})$$



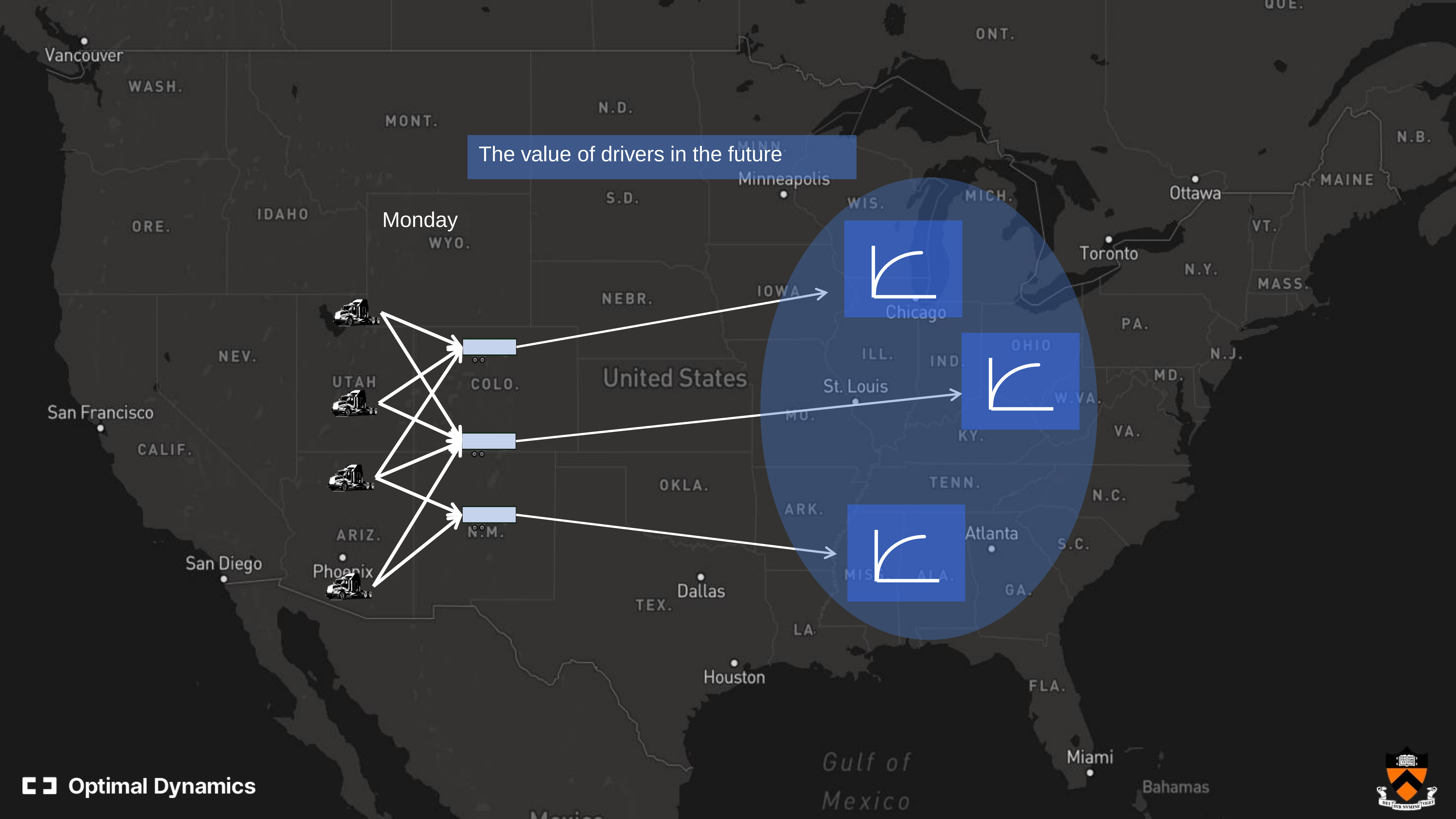

Monday
Vancouver
WASH.
MONT.
N.D.
MINN.
ONT.
QUE.
N.B.
MAINE
Minneapolis
S.D.
WIS.
MICH.
Ottawa
ORE.
IDAHO
WYO.
VT.
Toronto
N.Y.
MASS.
IOWA
NEBR.
Chicago
PA.
NEV.
UTAH
COLO.
United States
ILL.
IND.
OHIO
N.J.
MD.
San Francisco
St. Louis
W.VA.
CALIF.
MO.
KY.
VA.
TENN.
OKLA.
N.C.
ARK.
ARIZ.
N.M.
Atlanta
S.C.
San Diego
Phoenix
MISS.
ALA.
Dallas
GA.
TEX.
LA.
Houston
FLA.
Gulf of Mexico
Miami
Bahamas

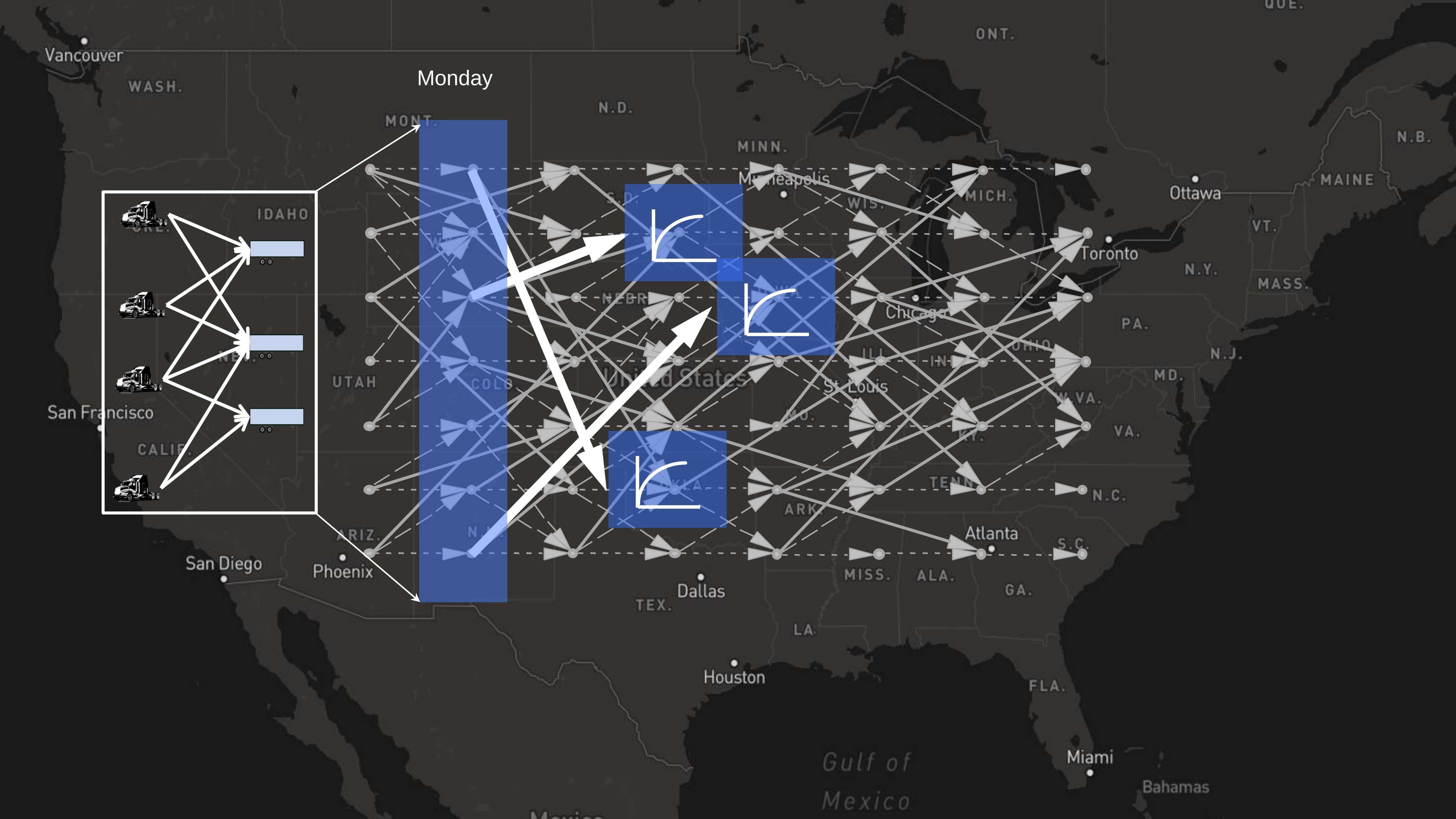


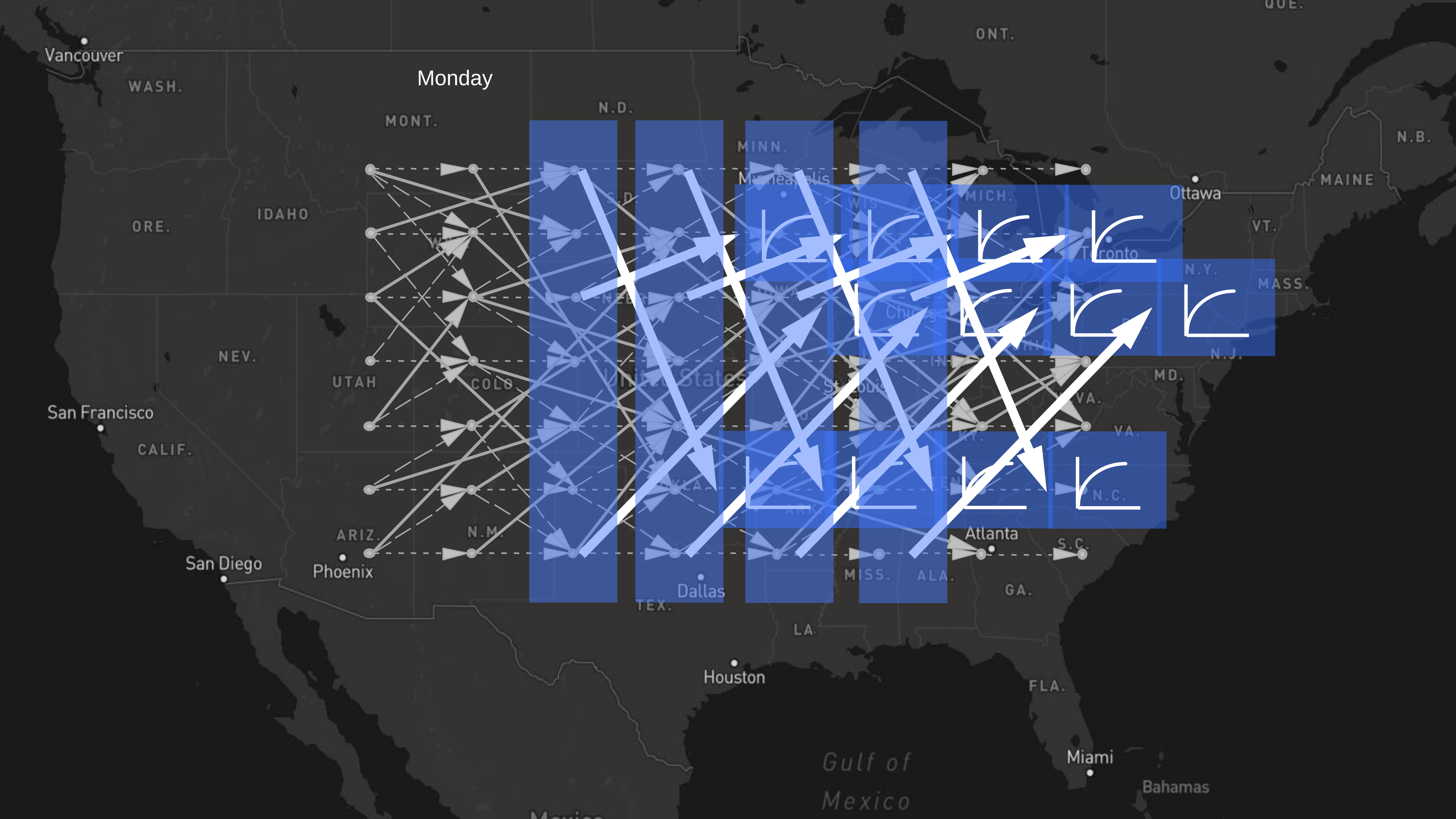
Monday
Tuesday

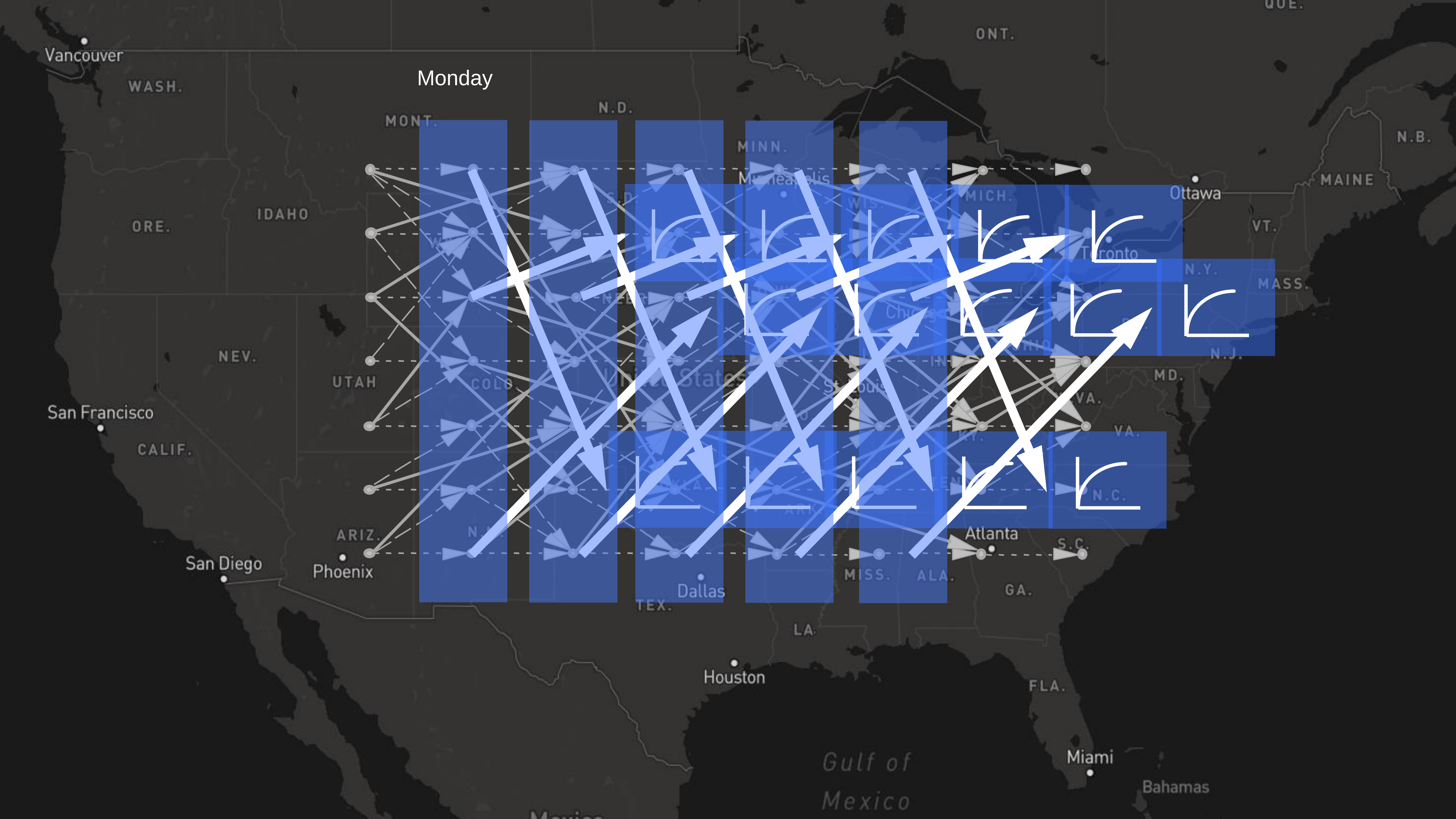

Monday
Tuesday
Wednesday
Vancouver
WASH.
MONT.
N.D.
MINN.
ONT.
QUE.
N.B.
Minneapolis
S.D.
WIS.
MICH.
Ottawa
MAINE
ORE.
IDAHO
WYO.
Toronto
VT.
N.Y.
MASS.
NEBR.
IOWA
Chicago
PA.
NEV.
UTAH
COLO.
ILL.
IND.
OHIO
N.J.
MD.
San Francisco
St. Louis
W.VA.
MO.
VA.
CALIF.
KY.
TENN.
N.C.
ARIZ.
N.M.
ARK.
S.C.
San Diego
Phoenix
ALA.
GA.
Dallas
TEX.
LA.
Houston
FLA.
Gulf of Mexico
Miami
Bahamas

The value of drivers in the future




Monday

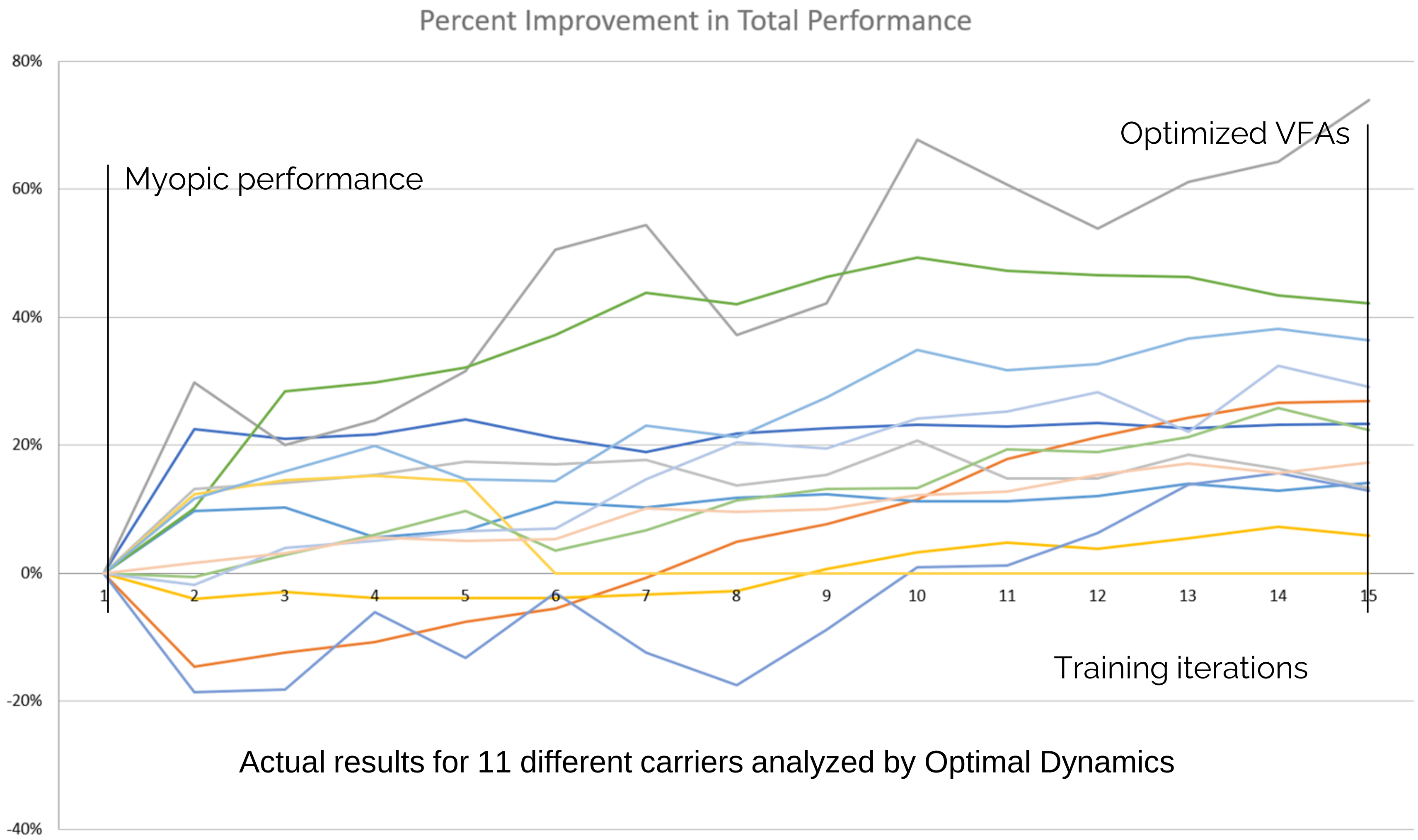

Monday
Vancouver
WASH.
MONT.
N.D.
MINN.
ONT.
QUE.
N.B.
MAINE
Ottawa
VT.
MASS.
N.Y.
MICH.
IDAHO
ORE.
NEV.
UTAH
COLO.
San Francisco
CALIF.
ARIZ.
N.M.
San Diego
Phoenix
Dallas
TEX.
LA.
Houston
MISS.
ALA.
GA.
Atlanta
S.C.
N.C.
VA.
MD.
N.J.
FLA.
Miami
Bahamas
Gulf of
Mexico



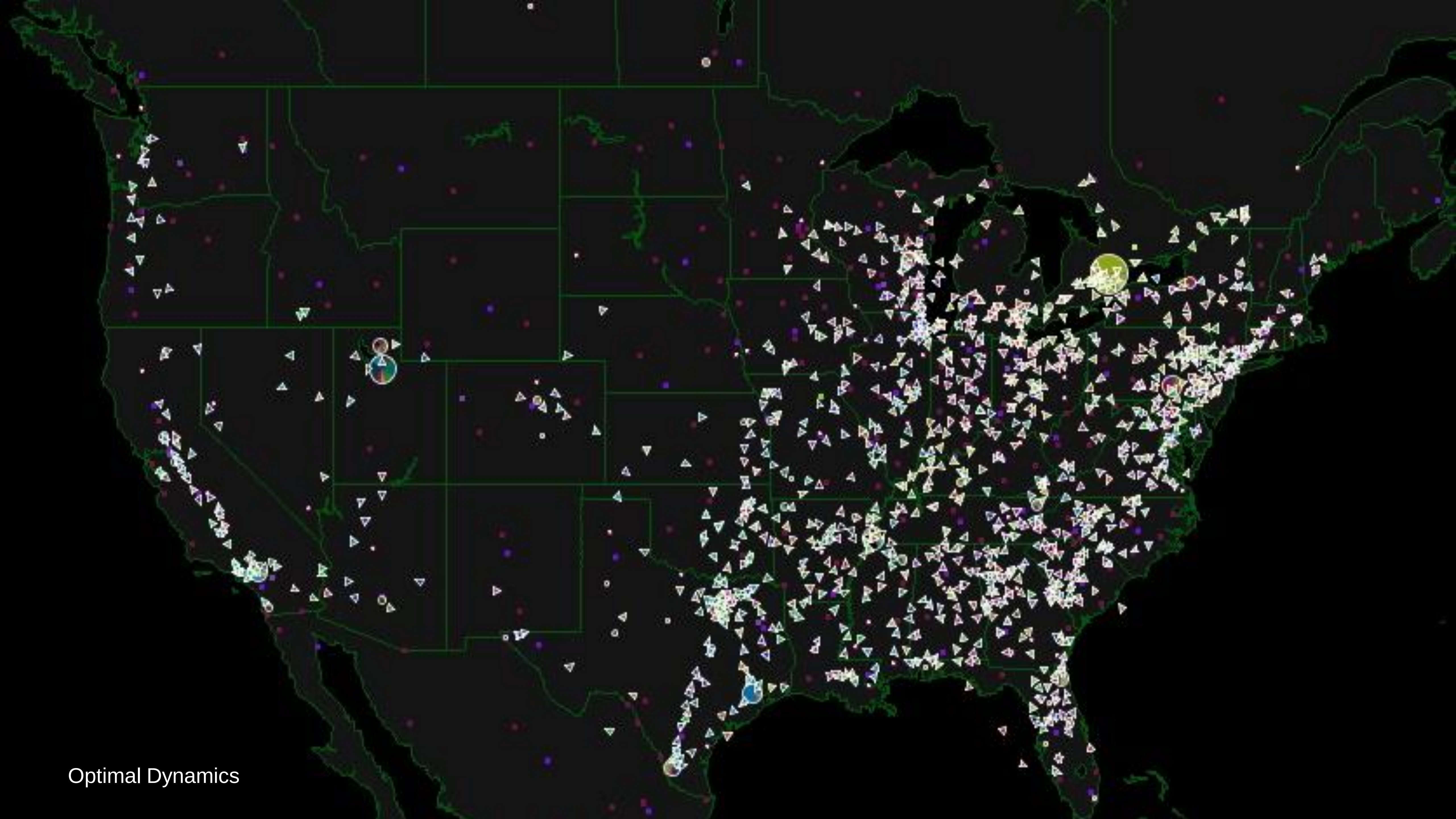
Monday
Vancouver
WASH.
MONT.
N.D.
ONT.
QUE.
MINN.
N.B.
MAINE
Ottawa
IDAHO
ORE.
MICH.
VT.
N.Y.
MASS.
Toronto
NEV.
UTAH
N.J.
MD.
San Francisco
CALIF.
VA.
ARIZ.
Atlanta
S.C.
San Diego
Phoenix
Dallas
MISS.
ALA.
GA.
TEX.
LA.
Houston
FLA.
Miami
Bahamas
Gulf of Mexico

Approximate Dynamic Programming for Fleet Optimization

Percent improvement due to value function training



Optimal Dynamics



Optimal Dynamics

Direct lookahead policies

4) Direct lookahead policies (DLAs) – Here we create an approximation called the *approximate lookahead model:*

$$(\tilde{S}_{tt}, \tilde{x}_{tt}, \widetilde{W}_{t,t+1}, \tilde{S}_{t,t+1}, \tilde{x}_{t,t+1}, \widetilde{W}_{t,t+2}, \dots, \tilde{S}_{tt'}, \tilde{x}_{tt'}, \widetilde{W}_{t,t'+1}, \dots)$$

There are seven classes of approximations we can introduce.

Our direct lookahead policy now requires solving:

$$X_t^{DLA}(S_t|\theta) = argmax_x \left(C(S_t, x_t) + \tilde{E} \left\{ \max_{\tilde{\pi}} \tilde{E} \left\{ \sum_{t'=t+1}^{t+H} C\left(\tilde{S}_{t'}, \tilde{X}_{t'}^{\pi}(\tilde{S}_{t'})\right) |\tilde{S}_{t+1} \right\} |S_t, x_t \right\} \right)$$

Sampled information process

Restricted horizon

Limited decisions

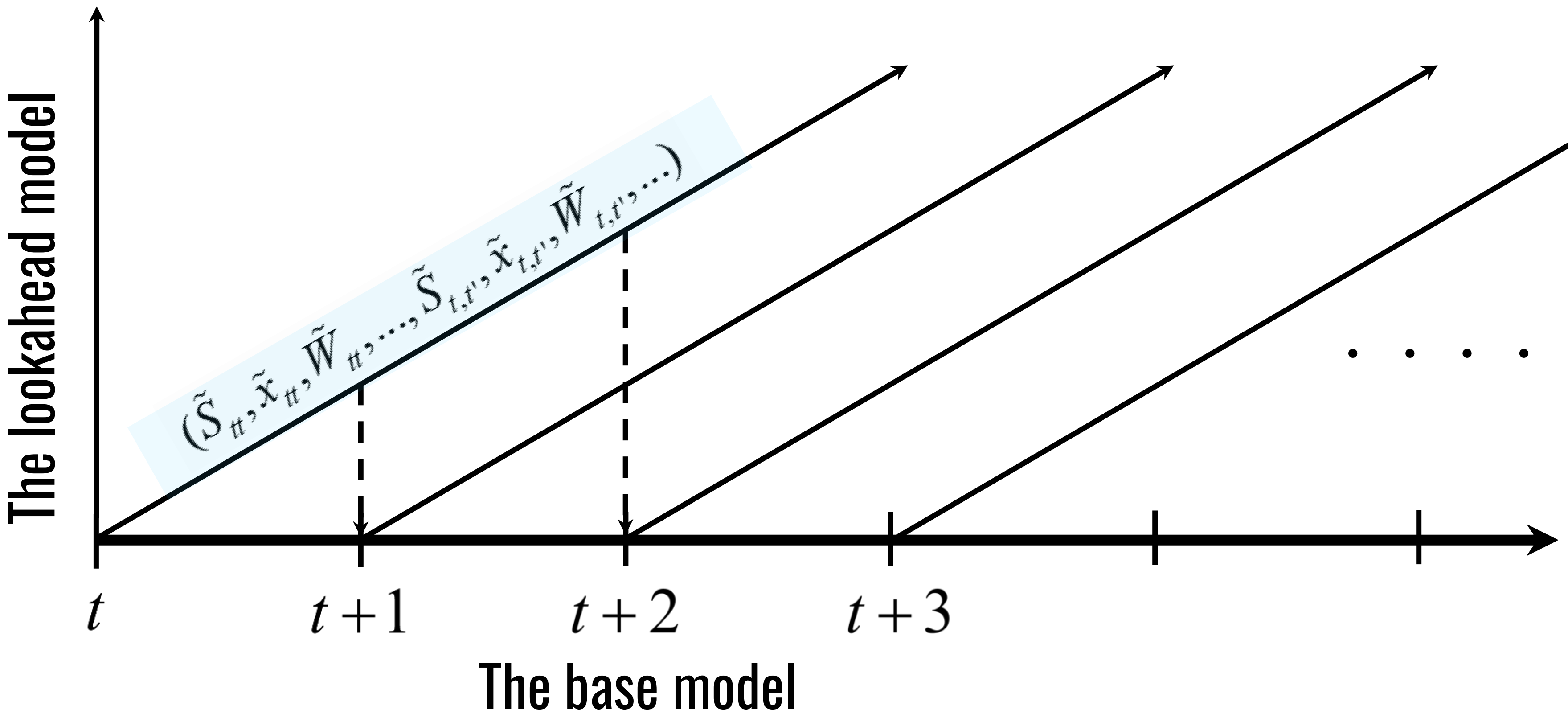
Simplified policies

Reduced state space

Direct lookahead policies

Direct Lookahead Policies (DLAs)

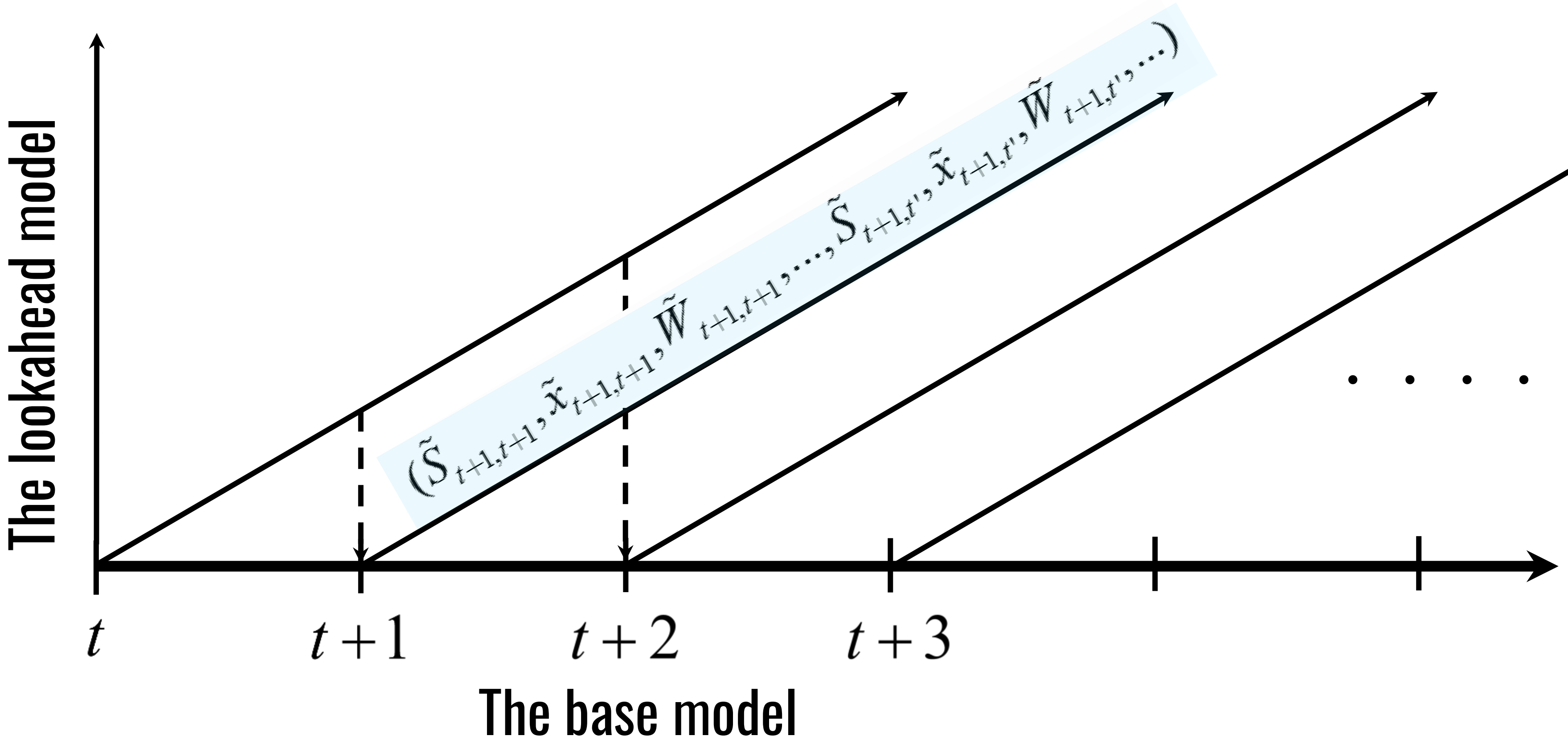
» Tilde variables are used to model approximate lookahead



Direct lookahead policies

Direct Lookahead Policies (DLAs)

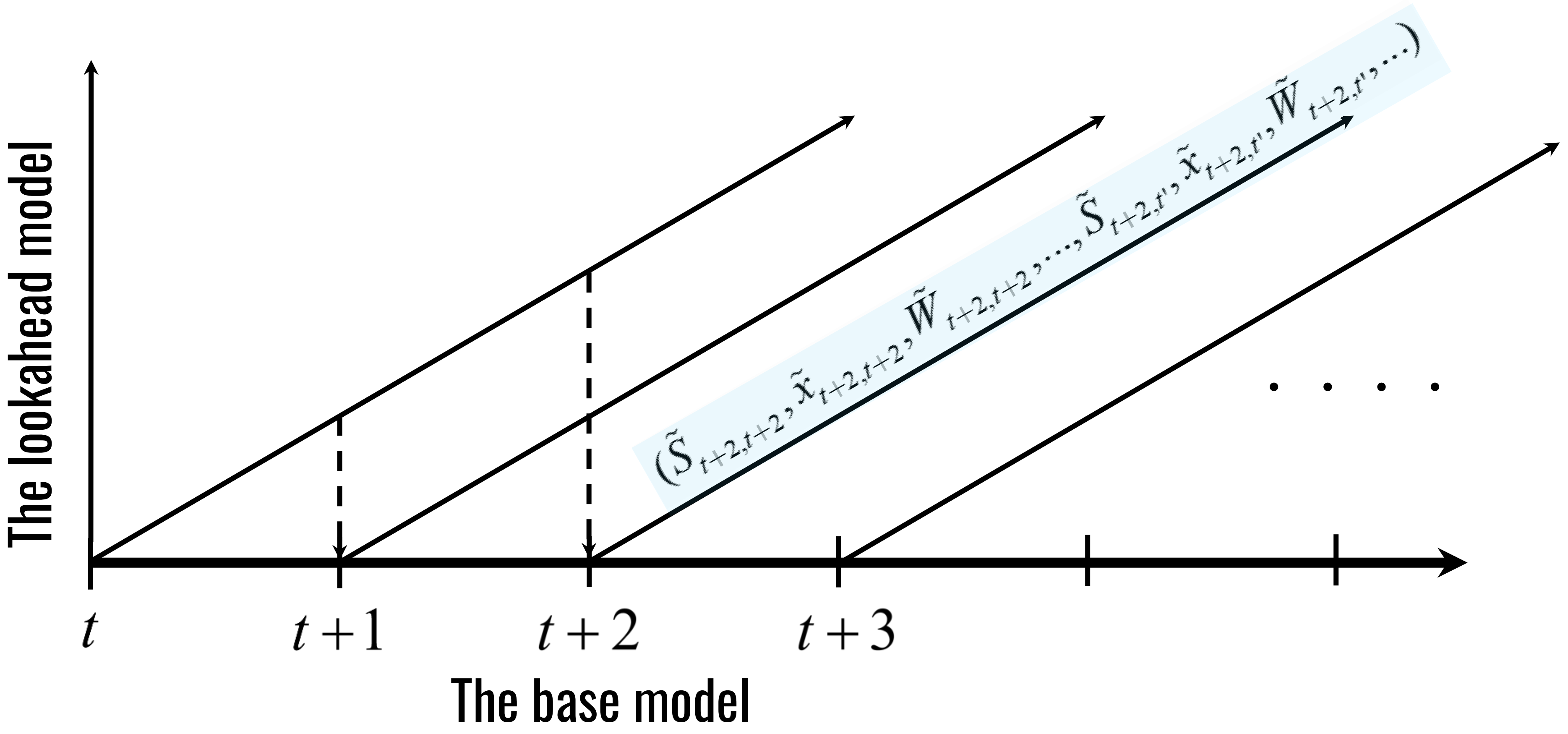
» Tilde variables are used to model approximate lookahead



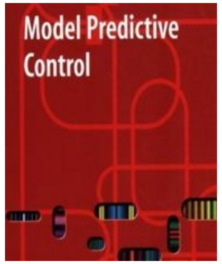
Direct lookahead policies

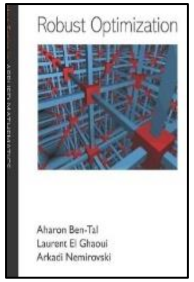
Direct Lookahead Policies (DLAs)

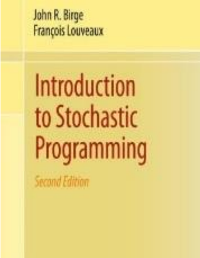
» Tilde variables are used to model approximate lookahead

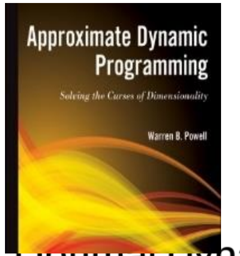


Direct lookahead policies



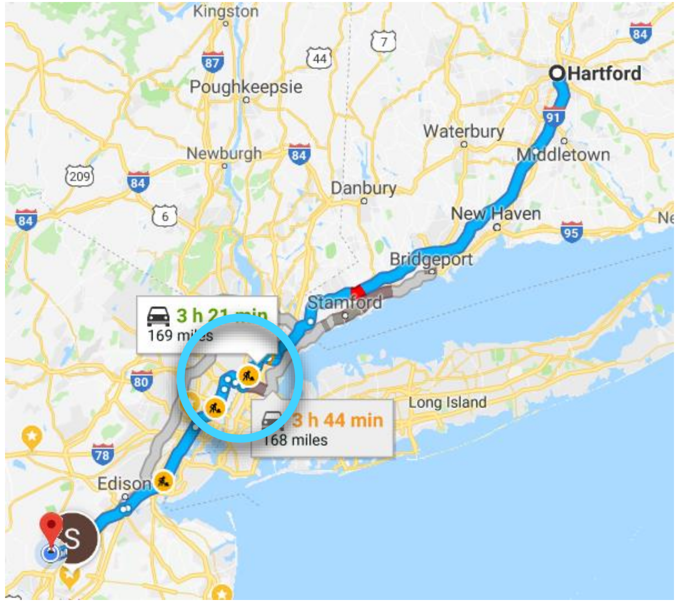


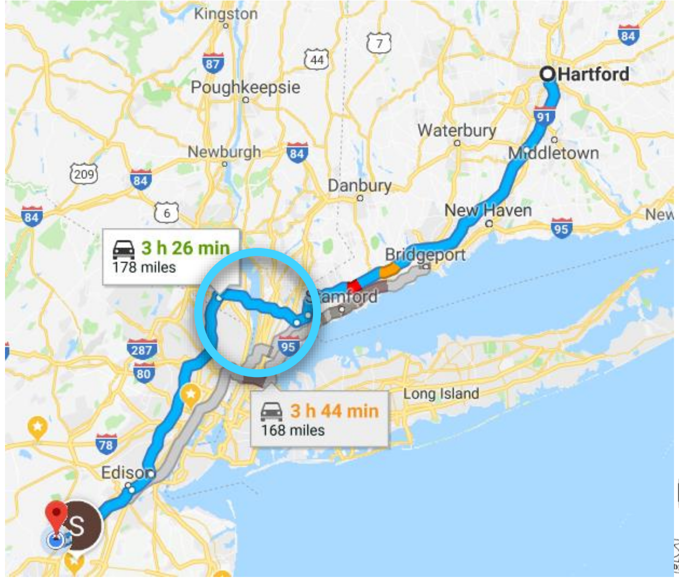




Examples of Lookahead Models

- **The deterministic lookahead model**
  - This is what is most widely used in practice.
  - Standard approach is to use a "best estimate" (which means deterministic) of travel times in the future.
  - This is often referred to as "model predictive control"
- **Robust optimization** - We could use the 90^{th} percentile of travel times.
- **Stochastic programming** – We represent the future using, say, 20 samples.
- **Approximate dynamic programming applied to approximate lookahead model**
- **Chance constrained programming** – Impose constraint on the probability of being late.





Designing policies

Policy search policies

Policy function approximations (PFAs)

- Simple rules, functions
- Examples:
  - Order up to
  - Buy low, sell high

Cost function approximations (CFAs)

- Parameterized cost models
- Examples
  - Schedule slack for trips
  - Buffer stocks for inventory

Lookahead policies

Value function approximations (VFAs)

- Making a decision now using the value of being in a future state
- Examples:
  - The value of a truck driver
  - The value of holding an asset

Direct lookaheads (DLAs)

- Models that optimize over a planning horizon (deterministically/stochastically)
- Examples:
  - Google maps
  - Energy planning models

Designing policies

The four classes of policies are *universal* – they cover every method for making decisions described in the research literature or used in practice.

This means you are already using one of the four classes of policies (or a hybrid) in your own decisions.

Direct lookahead policies

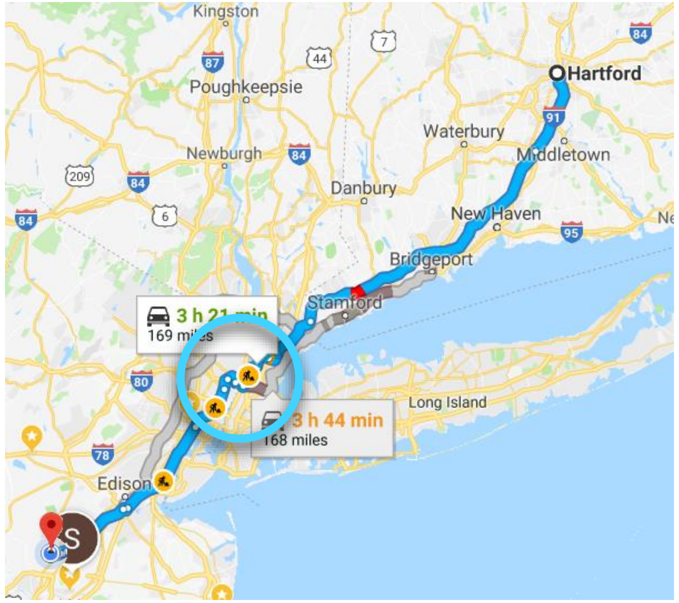
Parameterized deterministic lookahead

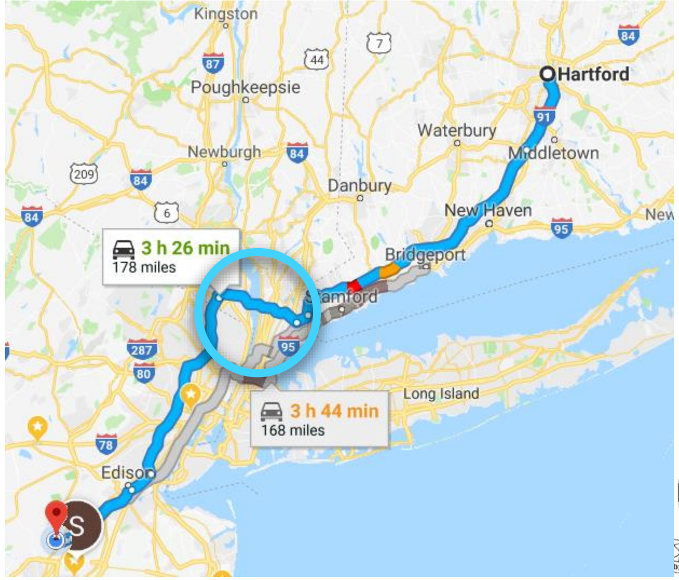
Instead of using a complex stochastic lookahead:

- Use the θ −percentile of the travel time distribution for each link:

  $\tilde{c}^p_{ij}(\theta)$ =The θ −percentile of $\hat{c}_{ij}$

- …which means $Prob\left[\hat{c}_{ij} \leq \tilde{c}^p_{ij}(\theta)\right] = \theta$.
- Now solve deterministic shortest path problems using costs $\tilde{c}^p_{ij}(\theta)$.
- This is no more complicated than our original deterministic shortest path problem, but …
- … we have to tune θ.





Parameterized deterministic lookahead

- The θ –percentile policy.
  - » Solve the linear program (shortest path problem):

$$X_t^\pi(S_t^n|\theta) = \arg\min \sum_{i\in N} \sum_{j\in N_i^+} \tilde{c}_{tij}^p(\theta)\tilde{x}_{tij} \quad \text{(Vector with } x_{tij} = 1 \text{ if decision is to take } (i,j))$$

  - » subject to

$$\sum_j \tilde{x}_{t,i_t^n,j} = 1 \quad \text{Flow out of current node where we are located}$$

$$\sum_i \tilde{x}_{tir} = 1 \quad \text{Flow into destination node } r$$

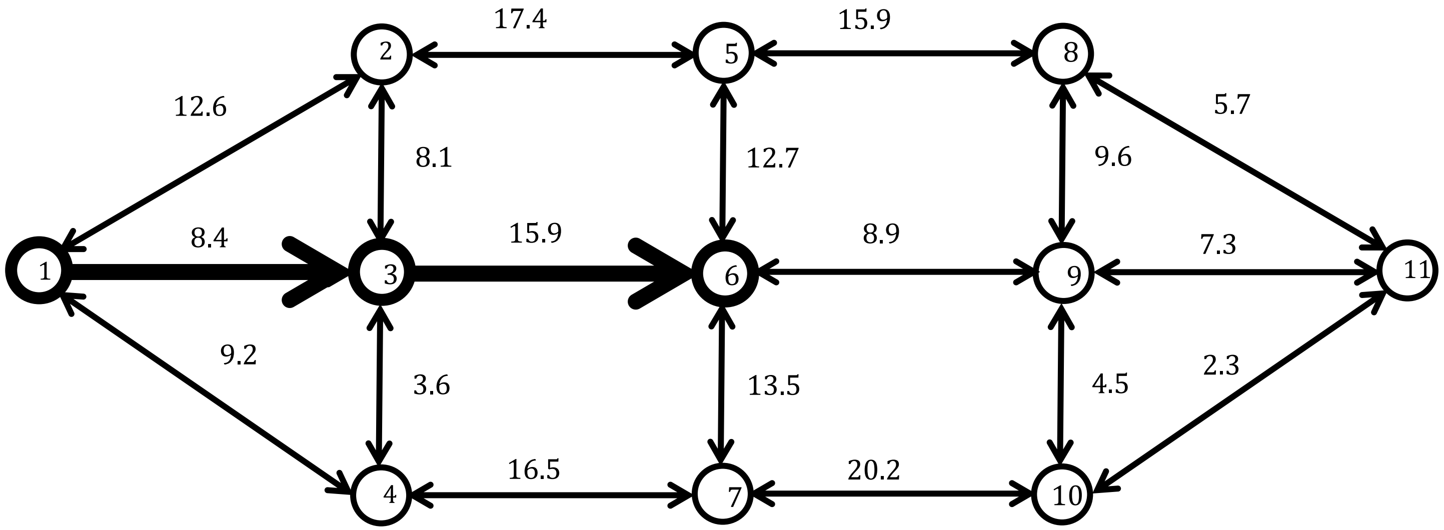
$$\sum_i \tilde{x}_{tij} - \sum_k \tilde{x}_{tjk} = 0 \quad \text{for all other nodes.}$$

  - » This is a deterministic shortest path problem.

Optimal Dynamics

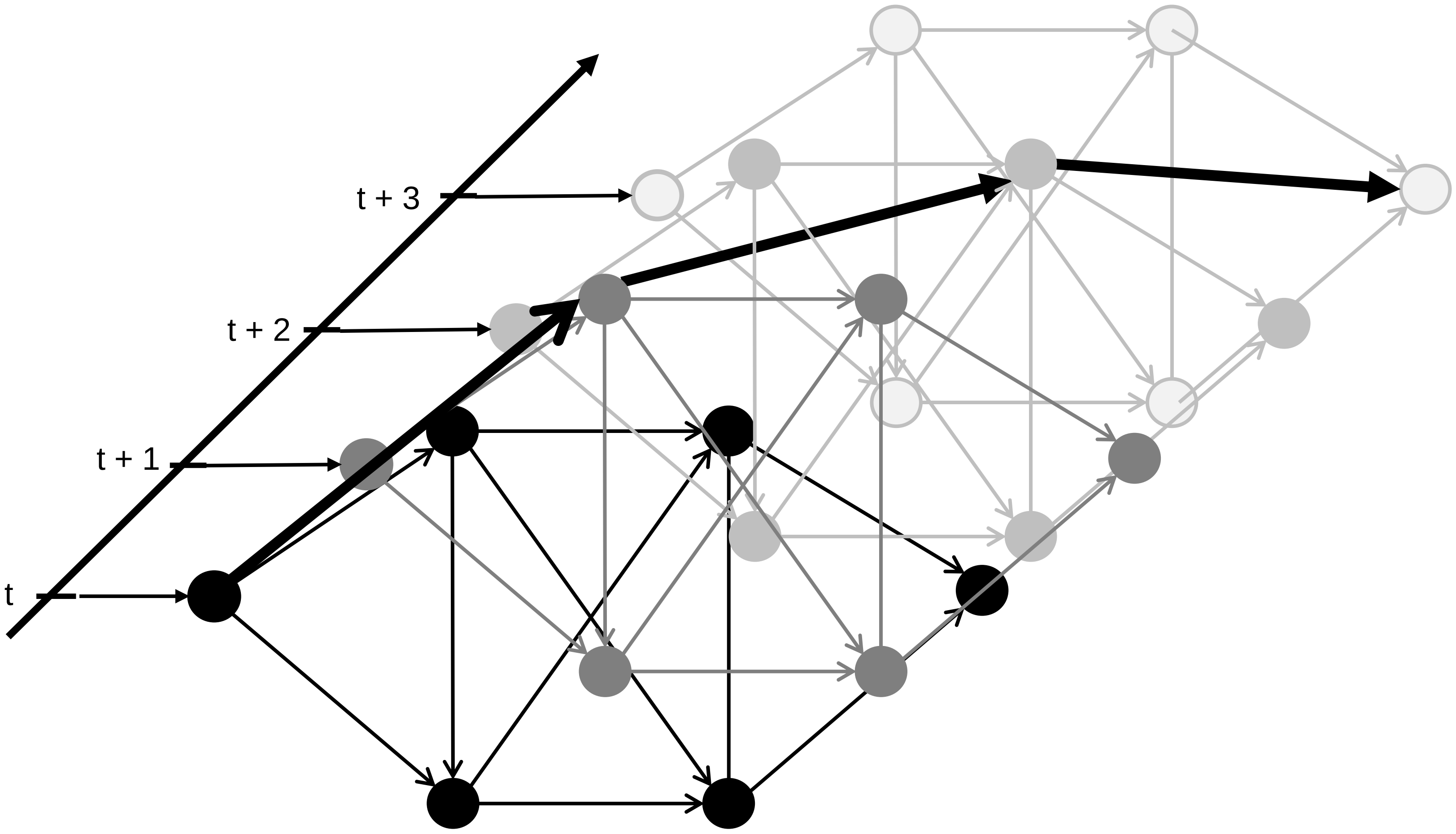
Dynamic shortest paths

- A static, deterministic network



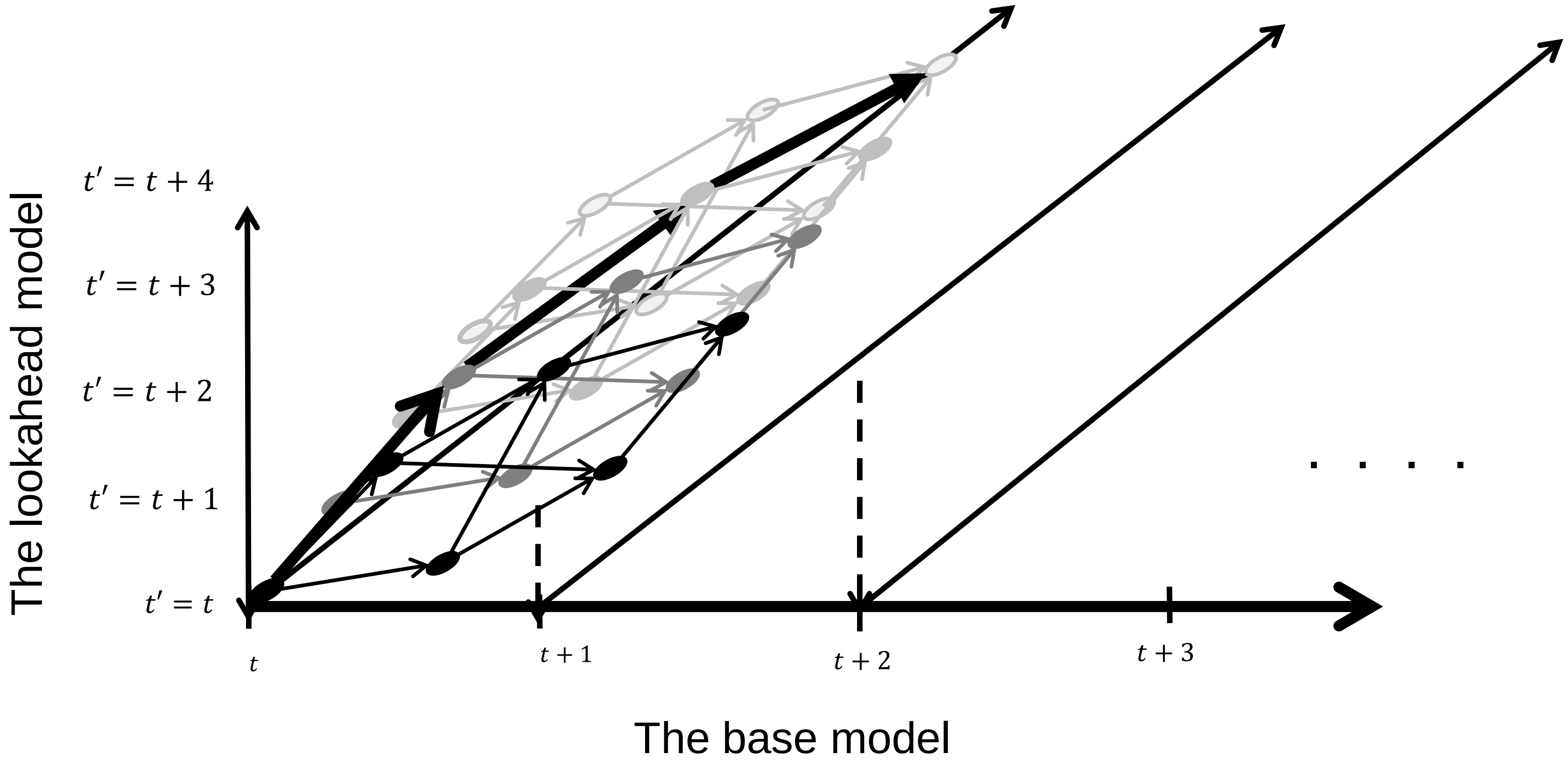
Dynamic shortest paths

- A time-dependent, deterministic network



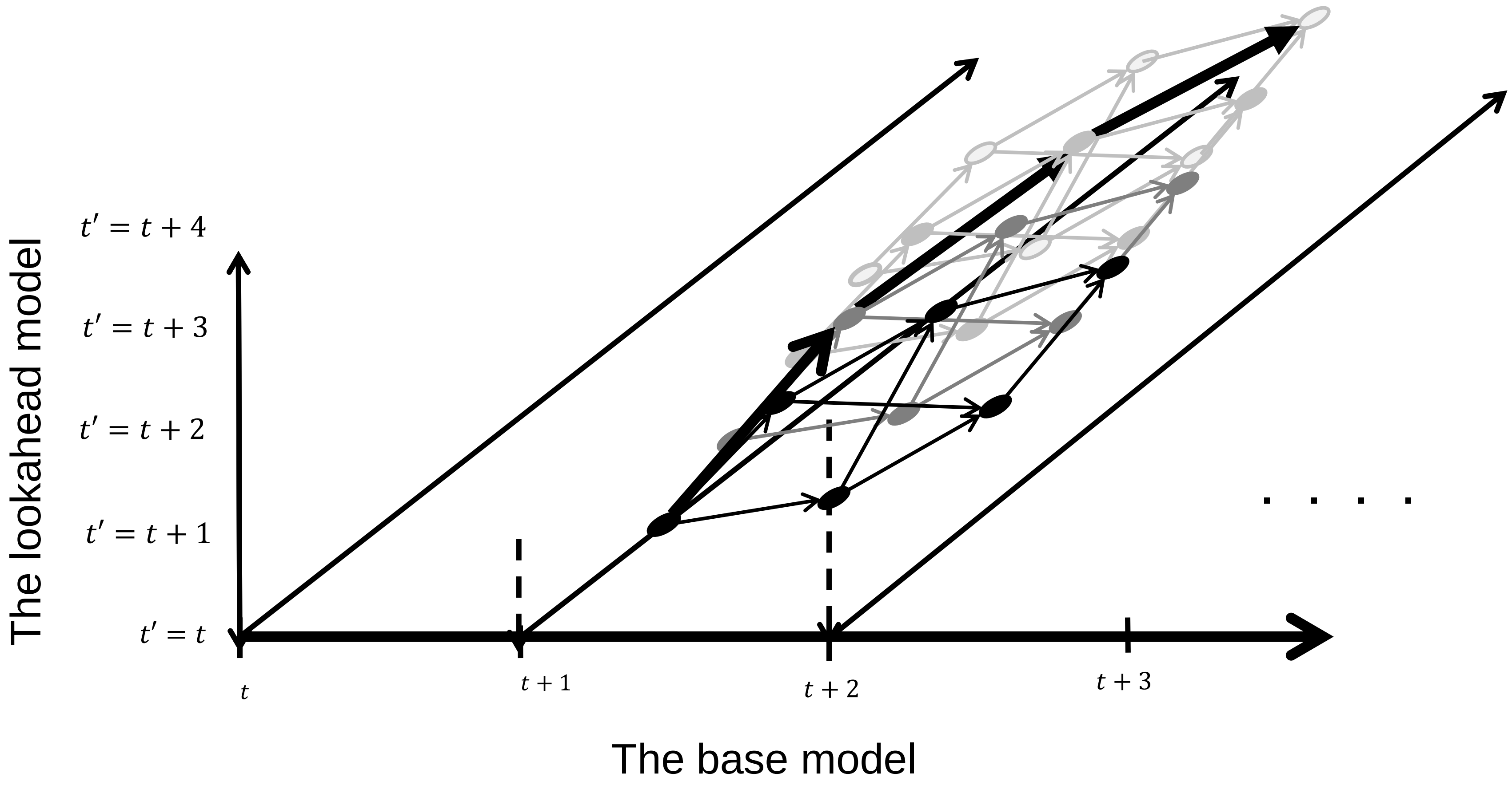
Dynamic shortest paths

- A time-dependent, deterministic lookahead network



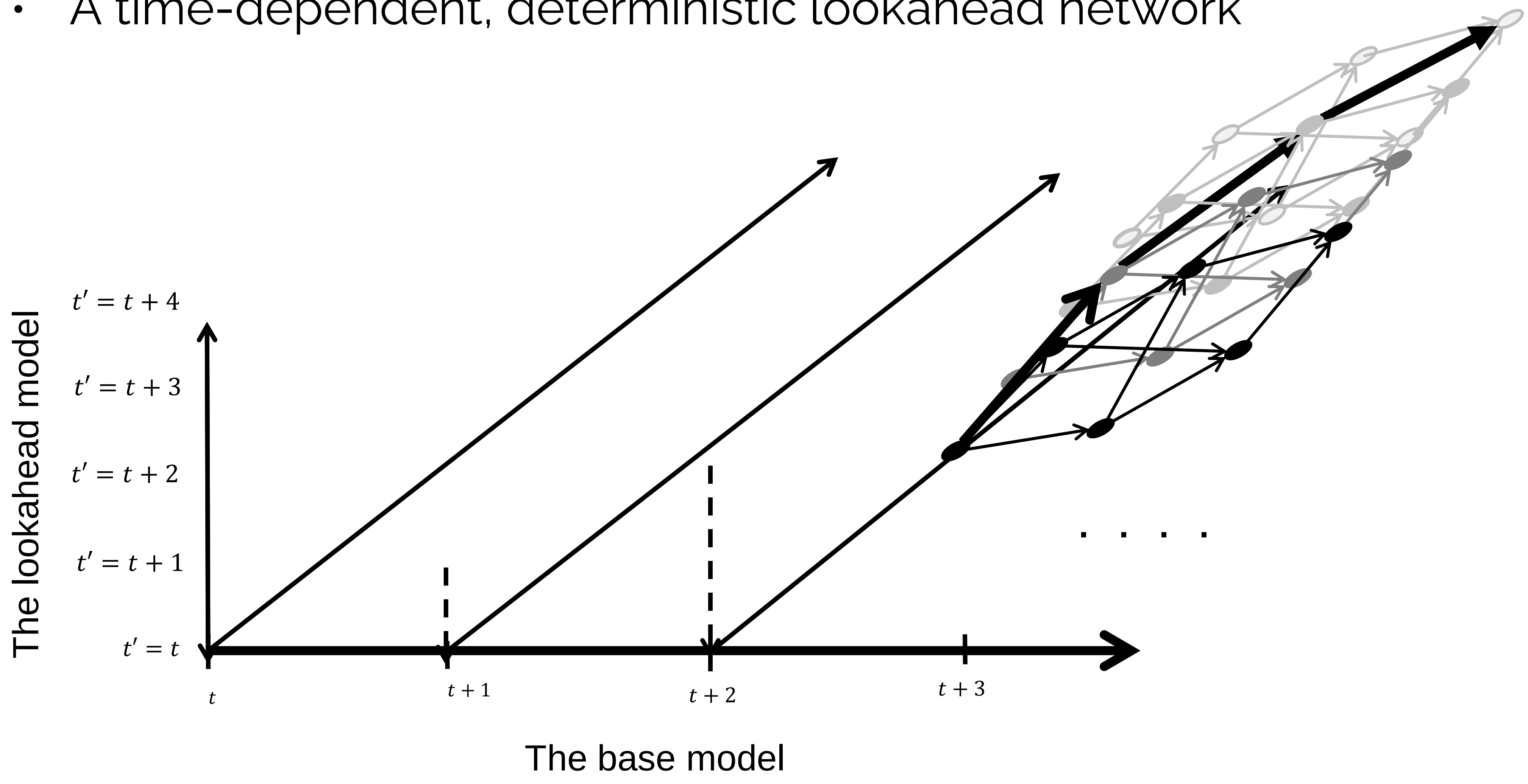
Dynamic shortest paths

- A time-dependent, deterministic lookahead network



Dynamic shortest paths

- A time-dependent, deterministic lookahead network



Parameterized deterministic lookahead

- Simulating a lookahead policy

  Let ω be a sample realization of costs

  $\hat{c}_{t,t',ij}(\omega), \hat{c}_{t+1,t',ij}(\omega), \hat{c}_{t+2,t',ij}(\omega), \ldots$

  Now simulate the policy

  $$\hat{F}^{\pi}(\theta|\omega^n) = \sum_{t=0}^{T} \sum_{i,j} \hat{c}_{t,t',ij}(\omega) X_t^{\pi}(S_t(\omega^n)|\theta)$$

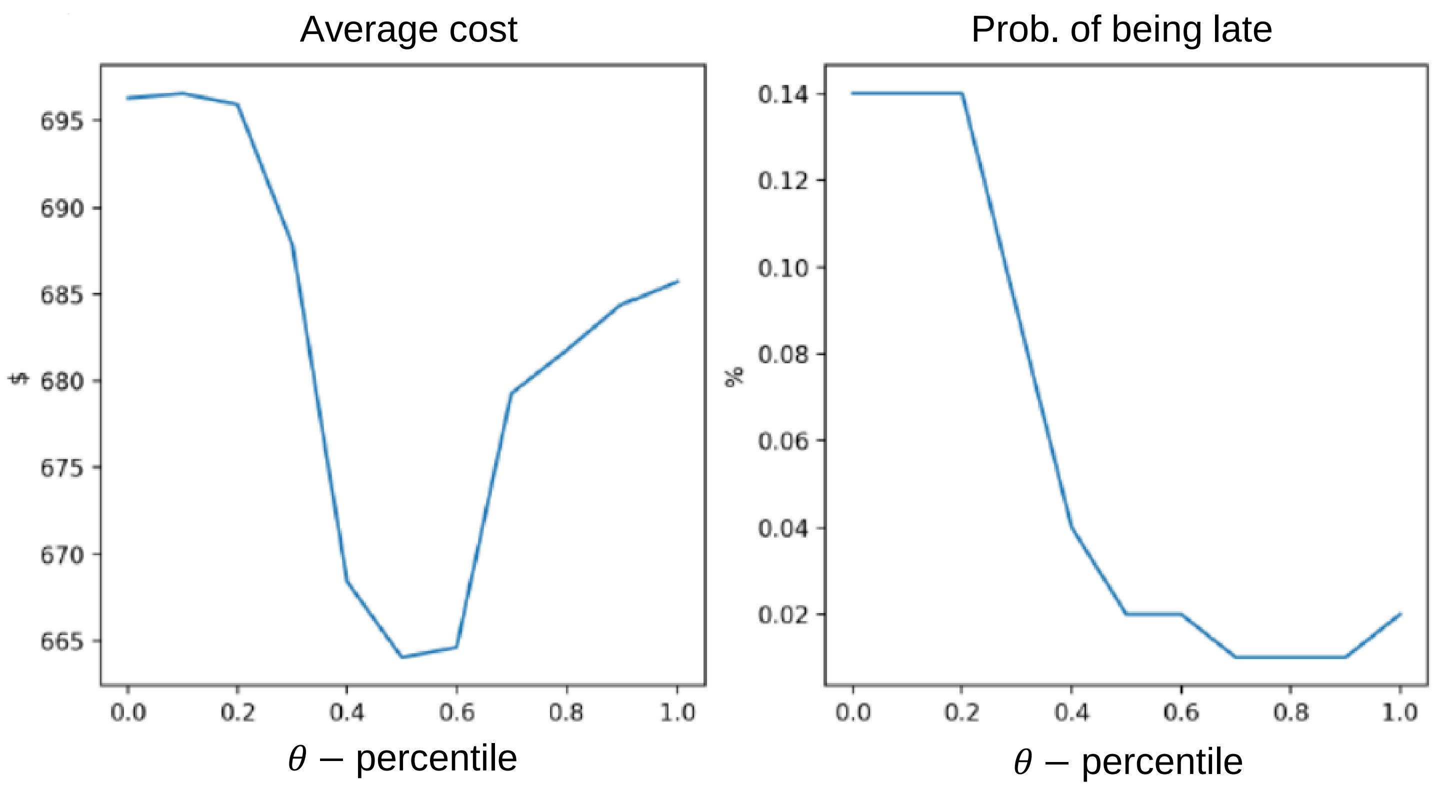
  Finally, get the average performance

  $$\bar{F}^{\pi}(\theta) = \frac{1}{N} \sum_{n=1}^{N} \hat{F}^{\pi}(\omega^n)$$

Optimal Dynamics

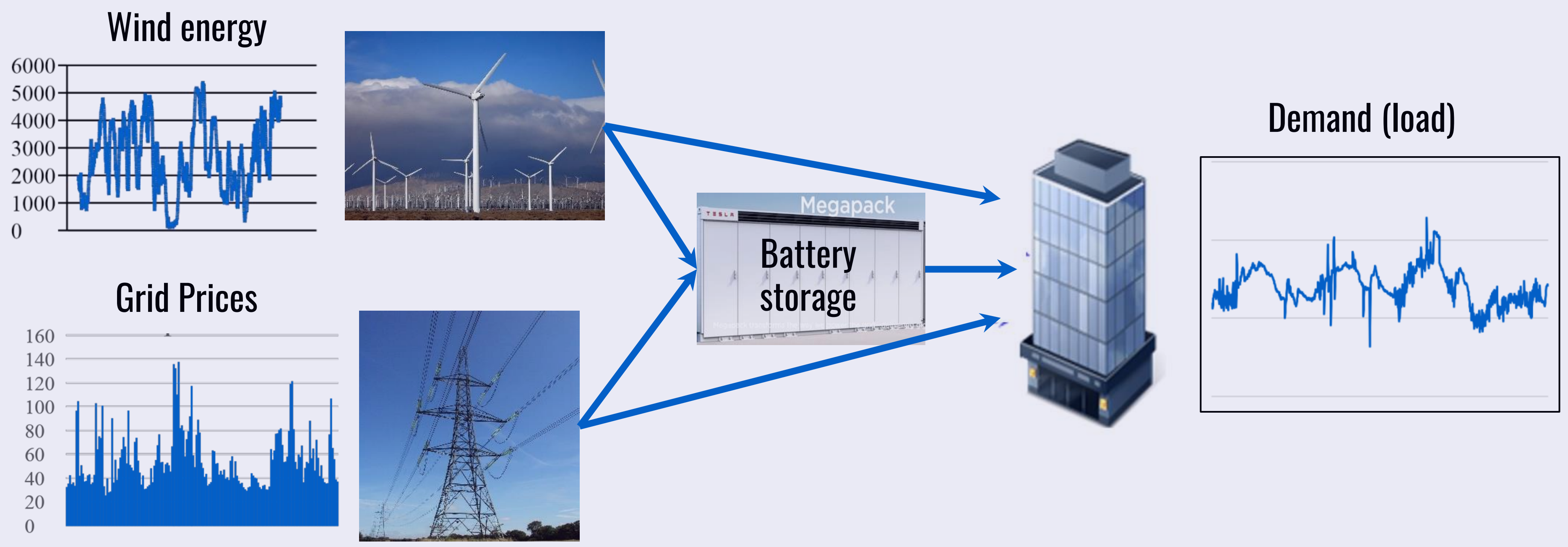
Parameterized deterministic lookahead

- Policy tuning
  - » Cost vs. lateness (risk)



An energy storage application

Consider a basic energy storage problem



We are going to show that with minor variations in the characteristics of this problem, we can make *each* class of policy work best.

An energy storage application

Each policy is best on certain problems

Problem:	Problem description	PFA	CFA	VFA	DLA	DLA/CFA
A	A stationary problem with heavy-tailed prices, relatively low noise, moderately accurate forecasts.	**0.959**	**0.839**	**0.936**	**0.887**	**0.887**
B	A time-dependent problem with daily load patterns, no seasonalities in energy and price, relatively low noise, less accurate forecasts.	**0.714**	**0.752**	**0.712**	**0.746**	**0.746**
C	A time-dependent problem with daily load, energy and price patterns, relatively high noise, forecast errors increase over horizon.	**0.865**	**0.590**	**0.914**	**0.886**	**0.886**
D	A time-dependent problem, relatively low noise, very accurate forecasts.	**0.962**	**0.749**	**0.971**	**0.997**	**0.997**
E	Same as (C), but the forecast errors are stationary over the planning horizon.	**0.865**	**0.590**	**0.914**	**0.922**	**0.934**

Joint research with Prof. Stephan Meisel, University of Muenster, Germany.

BRIDGING MACHINE LEARNING & SEQUENTIAL DECISIONS

Machine learning

$$\min_{f \in F, \theta \in \Theta^f} \frac{1}{N} \sum_{n=1}^{N} \left(y^n - f(x^n | \theta) \right)^2$$

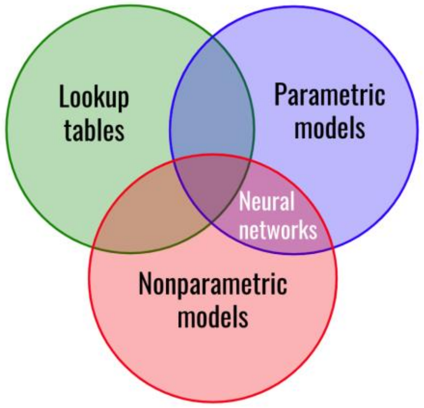
$$S_{t+1} = S^M(S_t, x_t, W_{t+1})$$

Searching over functions | **"Big dataset"**

Sequential decisions

$$\max_{\pi = (f \in F, \theta \in \Theta^f)} \frac{1}{N} \sum_{n=1}^{N} \sum_{t=0}^{T} C(S_t^n, X^\pi(S_t^n | \theta))$$

Searching over policies | **System model**

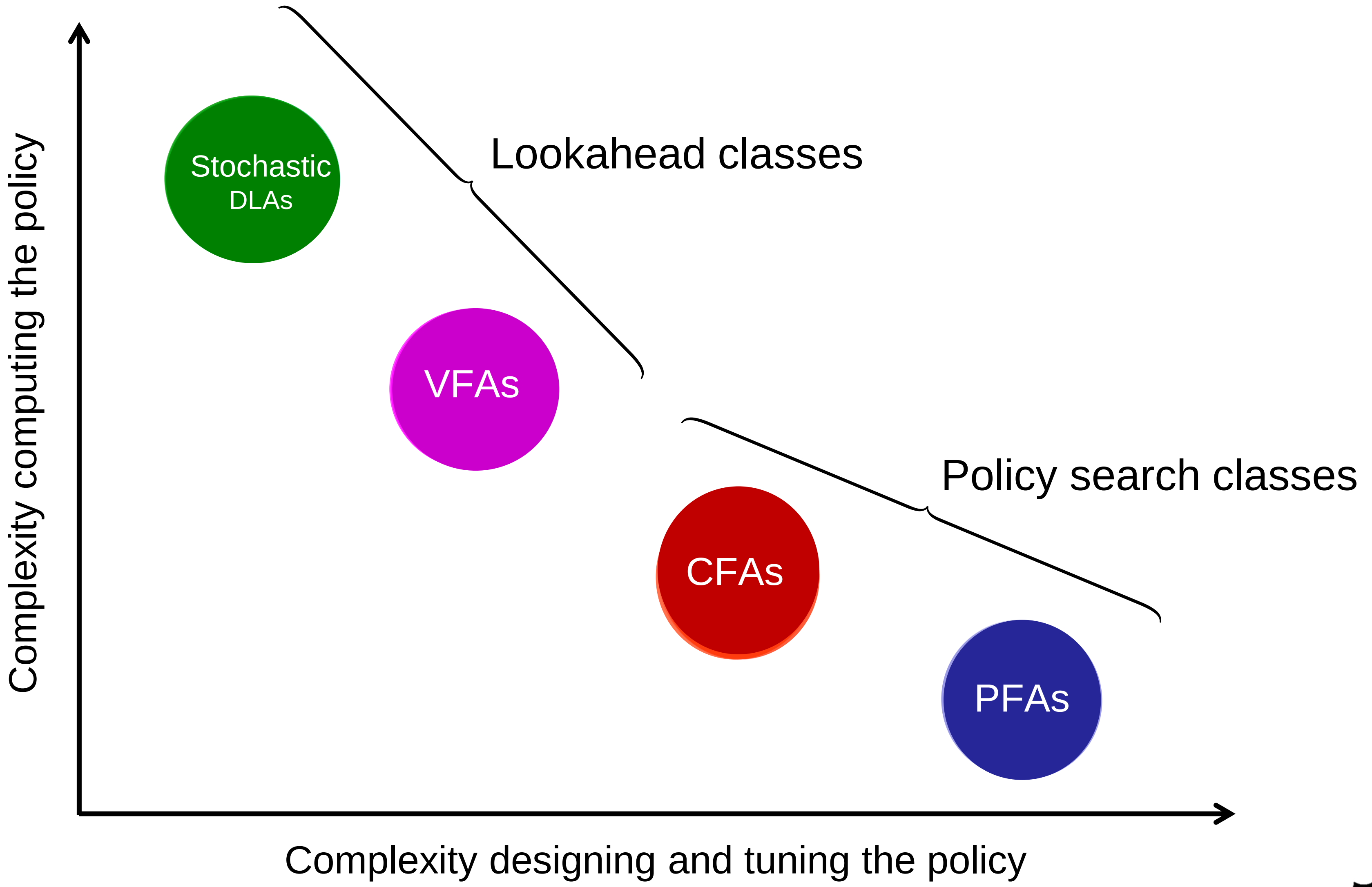


Searching over policies	System model
Policy function approximations	Analytical functions
Cost function approximations	Optimization problem
Value function approximations	Optimization problem
Direct lookahead approximations	Optimization problem

Choosing a policy class

There is a natural tradeoff between how well we approximate the impact of a decision on the future…

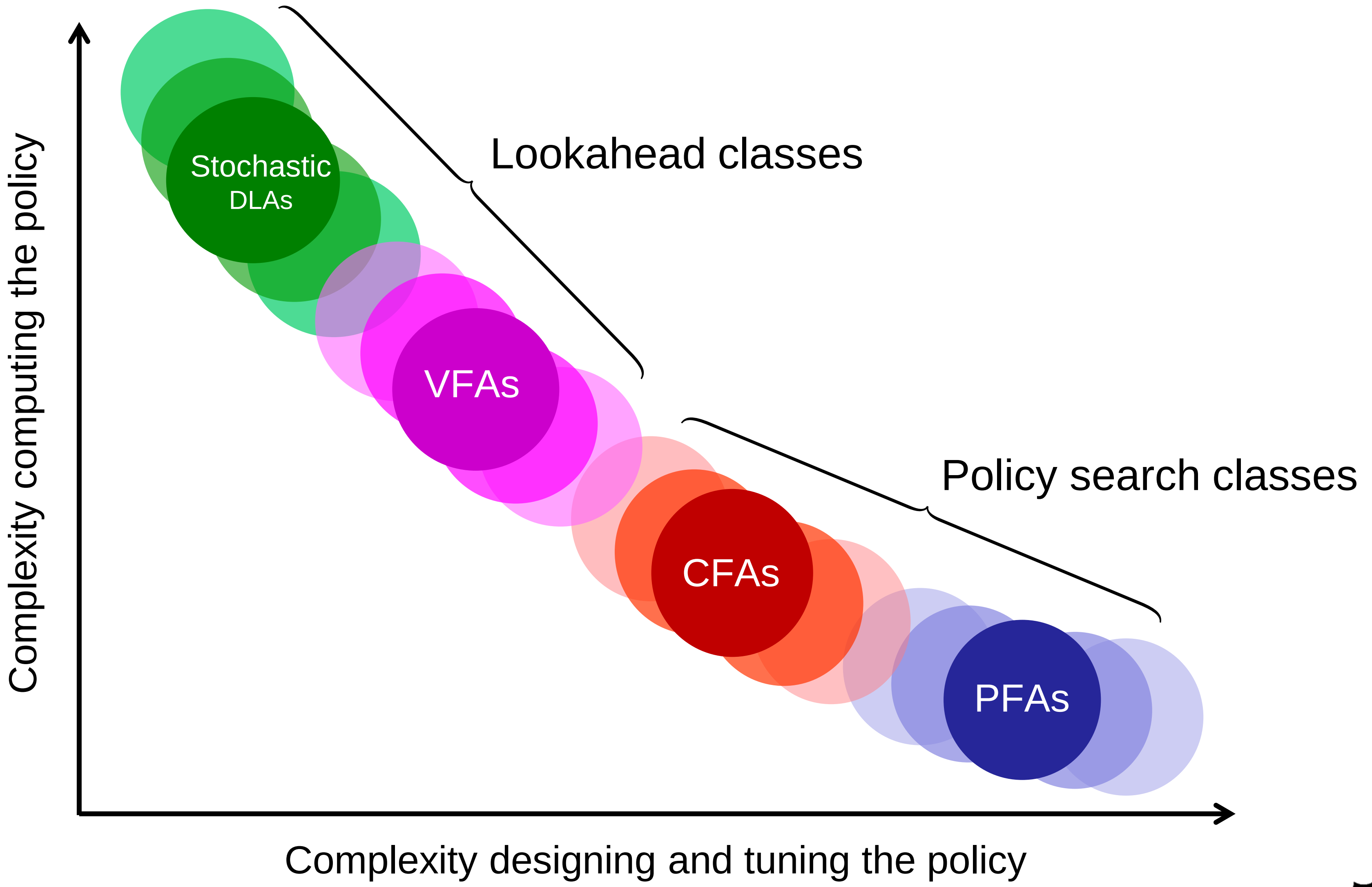
… and the complexity of tuning a policy (any policy) to work well over time.



Choosing a policy class

There is a natural tradeoff between how well we approximate the impact of a decision on the future…

… and the complexity of tuning a policy (any policy) to work well over time.



Choosing a policy class

- It helps to identify five types of policies:
  - 1) Policy function approximations (PFAs) – Simple rules, analytical functions.
  - 2) Cost function approximations (CFAs) – Parameterized deterministic optimization models (typically static)
  - 3) Policies based on value function approximations (VFAs) – Policies that use an approximation of the value of landing in a downstream state
  - 4) Policies based on direct lookahead approximations (DLAs) – These should be divided into two subclasses:
    - 4a) DLAs using deterministic lookaheads (Det-DLA) – These may be parameterized.
    - 4b) DLAs using stochastic lookaheads (Stoch-DLA)

- So, which are the most useful?

Choosing a policy class

- We can divide the five types of policies into three categories:

  » Category 1 – This category consists of:
    - 1) PFAs – Rules/analytical functions
    - 2) CFAs – Parameterized det. optimization
    - 4a) Det-DLAs – Deterministic lookaheads

  By far the most widely used in practice. The choice among the three tends to be obvious from the structure of the problem.

  » Category 2 – This category consists of:
    - 4b) Stochastic direct lookaheads

  Useful for more complex problems where planning into an uncertain future is required, and risk is important.

  » Category 3 – This category consists of:
    - 3) Policies based on VFAs.

  A very powerful strategy for a very small number of specialized problems.

BRIDGING MACHINE LEARNING & SEQUENTIAL DECISIONS

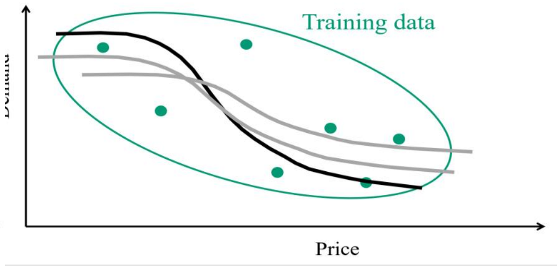
Machine learning

$$\min_{f\in F}\min_{\theta\in\Theta^f}\frac{1}{N}\sum_{n=1}^{N}\left(y^n - f(x^n|\theta)\right)^2$$

Searching over functions

Parameter tuning

(Deterministic optimization)



Parameterized policies

$$\min_{f\in F}\min_{\theta\in\Theta^f}\frac{1}{N}\sum_{n=1}^{N}\sum_{t=0}^{T} C\left(S_t^n, X^\pi(S_t|\theta)\right)$$

$$S_{t+1}=S^M(S_t,x_t,W_{t+1})$$

Searching over functions

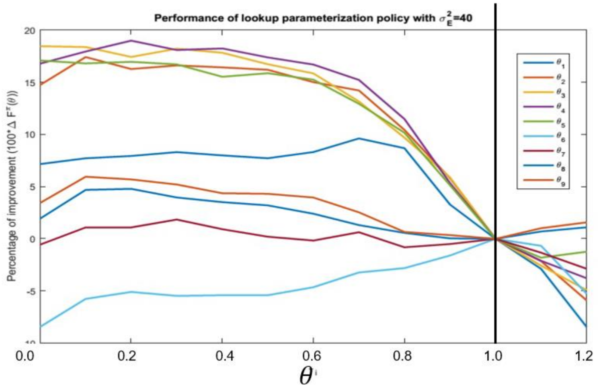
Parameter tuning

Stochastic search

PFAs

CFAs

Determ.
DLAs

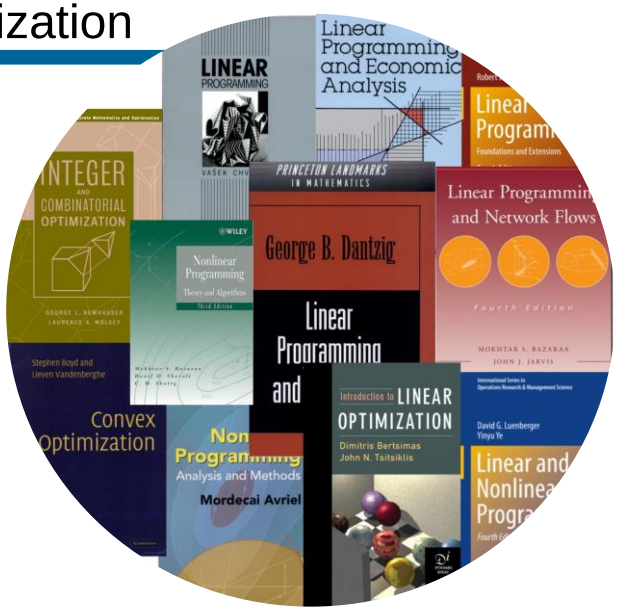


OUTLINE

- → The five layers of intelligence
- → Modeling sequential decision problems
- → Modeling uncertainty
- → Designing policies
- → A new educational field: sequential decision analytics

Decision analytics

Optimization



Machine learning

Each of these fields have well-defined communities, using common notation and established tools.

There are widely used textbooks that cover common material, with standard notational frameworks..

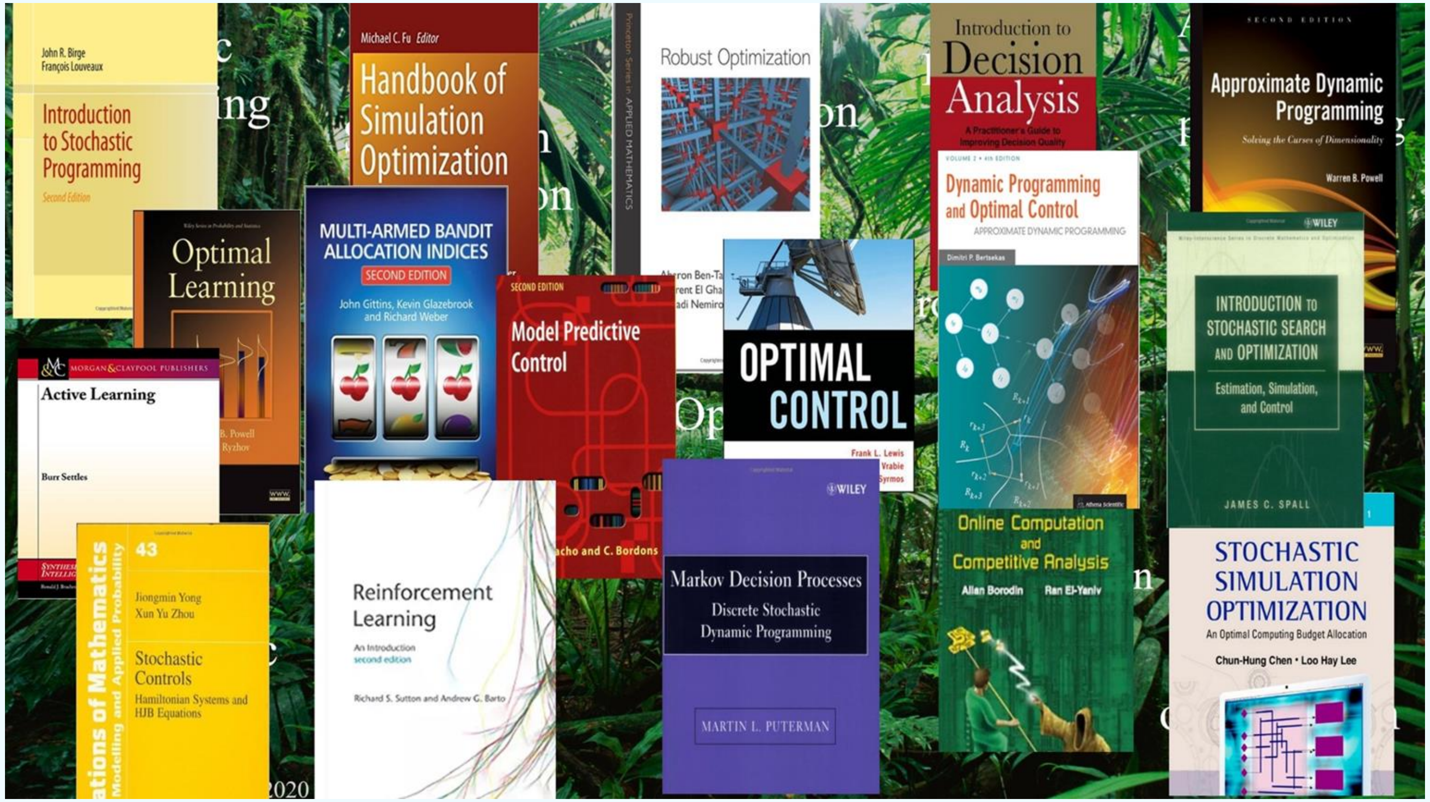
The concepts are taught in hundreds of academic programs, producing thousands of graduates each year which can be hired by industry.

Simulation

Decision analytics

The fields that deal with decisions and uncertainty are completely fragmented.

- Sequential decision analytics is not a recognized field.
- There are 15 distinct communities that deal with decisions under uncertainty
- Each community offers tools that work only for specific problems
- Real applications require skills that span a wide range of problem settings.

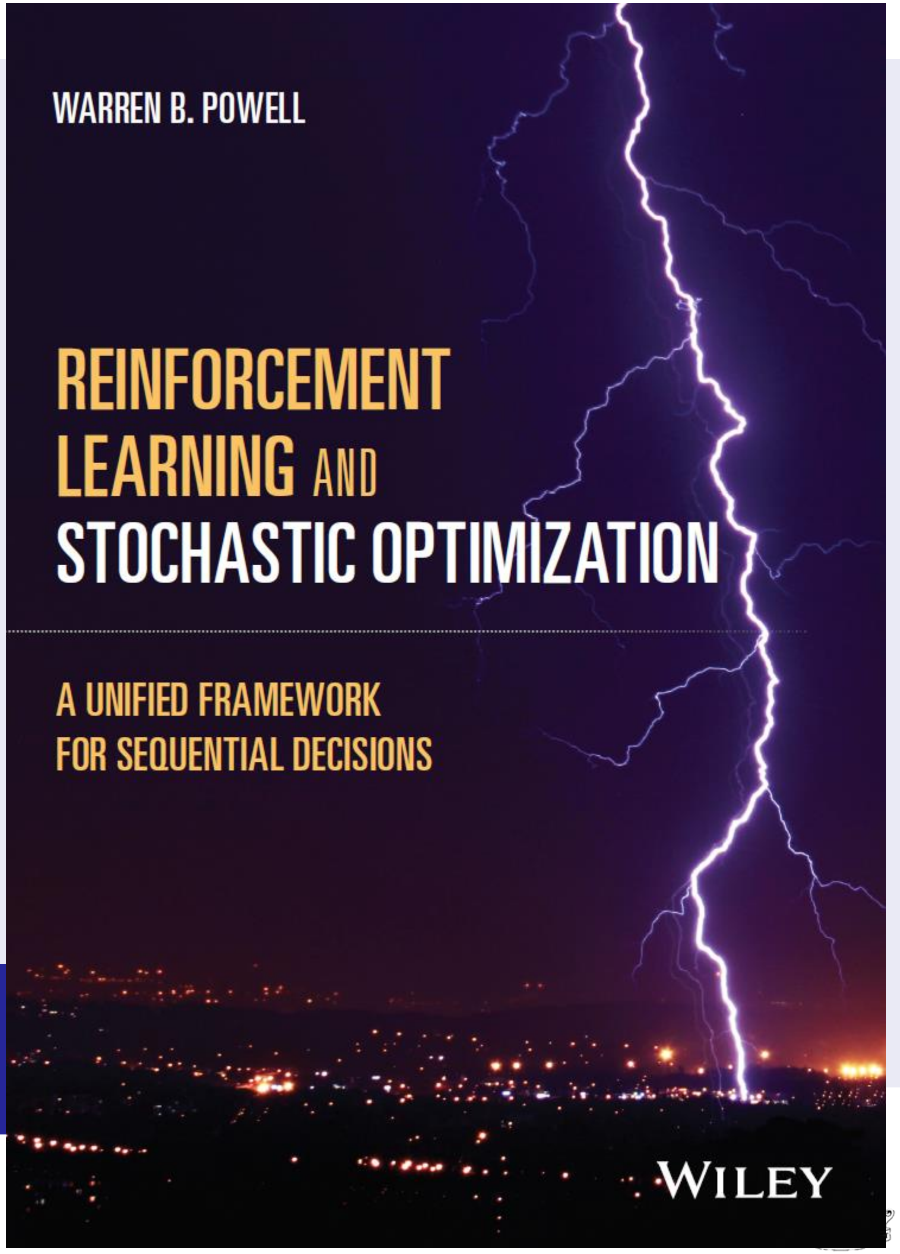


Decision analytics

A new book:

- First book to introduce a universal modeling framework, covering all four classes of policies.
- Describes the tools for modeling and solving *any* sequential decision problem, from simple learning problems to truckload fleets to complex supply chains.
- Aimed at a technical audience interested in writing software to develop models such as those described in this presentation.
- Provides the foundation for a new field we are calling *sequential decision analytics.*

http://tinyurl.com/RLandSO/



Decision analytics

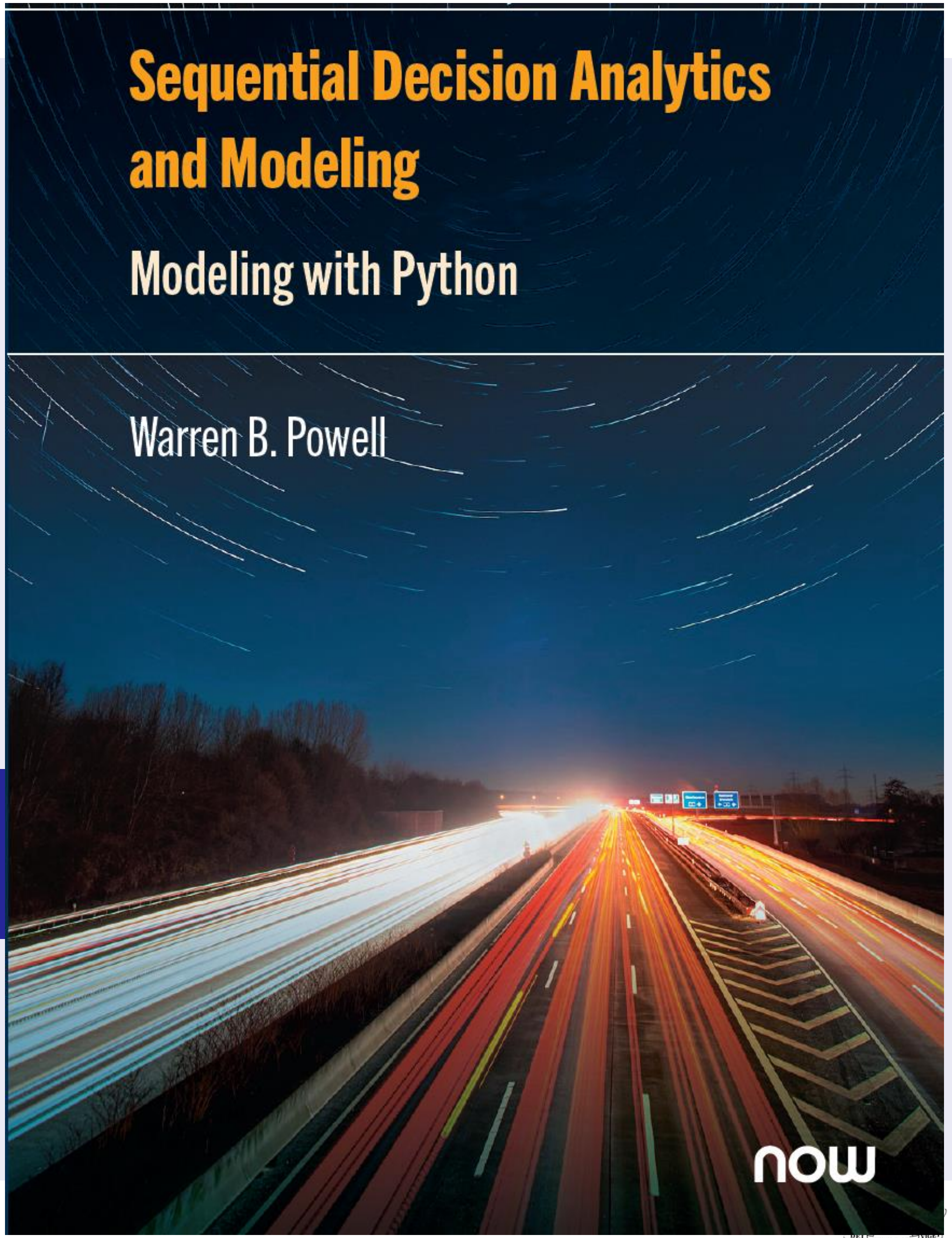
An introductory book:

- Uses a teach-by-example style
- Illustrates how to model sequential decision problems using a rich set of examples
- Illustrates all four classes of policies
- Highlights uncertainty modeling

http://tinyurl.com/sdamodeling

- Free download of the book:

  https://tinyurl.com/PowellSDAMbook



Teaching decision analytics

It is time to start teaching *sequential decision analytics*.

- » Can be taught to a broad audience spanning science and engineering.
- » Teaches students how to *think* about sequential decision problems.
- » Emphasizes identifying metrics, types of decisions and sources of uncertainties.
- » Start with the simplest policies that are most widely used.

These will be the first books to present sequential decision problems and solution methods in a unified way.

Thank you!

Some additional references:

A webpage on sequential decision analytics:
http://tinyurl.com/sdafield/

My new book:
http://tinyurl.com/RLandSO/

An information resource page for sequential decisions:
http://tinyurl.com/SDAlinks

Part 1. Literature

1. Sutton and Barto Book
2. Prof. Powell's Books
3. Papers from NeurIPS, ICLR, ICML and more.

Part 2. Recent News

1. GPT-4 and RLHF
2. AlphaTensor

Part 3a. Alex Jacquillat and Daniel Freund Questions / Comments

Sequential decision analytics - from theory to methods to practice. Alex framed his journey in transportation, logistics and optimization over the past 10 to 15 years in the context of Powell's book, "Reinforcement Learning and Stochastic Optimization". These are new tools that Alex wishes he knew back then when he started out. He was trying to model complex problems and ran into a whole bunch of scalability issues and he encourages everyone to learn from that today.

Alex commented on the definition of a state variable in this book and that we need to carefully define what we need to do or know at each stage. He also mentioned that we should look into the unifying concepts that this book teaches.

Powell worked closely with industry back in the day - in high profile projects as well, and he was also the CTO of the company. It's rare to see academics do this and it's great to see this coming into play.

Dynamic Pricing and Routing for Same-Day Delivery - Martin Ulmer. He really liked this paper as Ulmer took all the concepts in the talk from Powell today, and integrated them with modern day important problems in last mile delivery, logistics, etc. It also won the best paper prize.

These things are hard to implement and Alex looks forward to using Powell's book to make the course more accessible. In Alex's chart, he discusses three circles of large-scale optimization, sequential decision analytics, and stochastic models. Powell made a remark that optimization and stochastic models are in sequential decision analytics. It's interesting that he takes this "ultimate view" of "everything is a sequential decision analytic" problem. That's a very interesting perspective. Powell asked for an example of a purely "large-scale optimization", and Alex gave an example but Powell argued that the example is within sequential decision analytics! Facility location is thought of as a static problem, but it's not! There's always downstream implications of that.

Daniel appreciates the unified approach. Daniel was curious about partial observability - one player observes a part of the state, and another player observes another. Powell talks about multi-agent in his book as well. There are many problems that have belief states, and a lot of people ignore them. Powell mentions that the multi-agent field is massive and there are very poor models there. He also comments that it's very hard to publish in multi-agent because people are still thinking about this in the old framework.

Finally, Powell claims that any algorithm written by someone that solves some sequential decision making problem, can be mapped into his framework. I think it's probably true!

Part 3b. Audience Questions

Not Applicable.

Part 4. Reflection

I've briefly heard of Warren Powell and his lecture was very insightful for me. Coming from an RL background, I've only heard of popular books like Sutton and Barton (which he cited), and of course the papers from OpenAI, DeepMind, NeurIPS and ICML and many more. It was a refreshing take.

I enjoyed his perspective on how everything is a "sequential decision making" problem, and also, his view on machine learning as optimization, which is very similar to Bertsimas' work. It was ambitious of him to make a claim that he has come up with a unified view of this approach in his book, "Reinforcement Learning and Stochastic Optimization". He also briefly talked about it towards the end of his presentation and claimed that all sequential decision making problems can fit in his framework.

I'm curious as to why he called it a "cost function approximator", where general literature refers to this as a value function approximation. Interestingly, I recently came across his latest textbook in my lab titled "Reinforcement Learning and Stochastic Optimization". I took a quick browse, and he seems to come from the optimization perspective and not RL.

For the purposes of modern RL (in the world of GPT and all), his methods and textbook may not be as relevant. However, his textbook could be good for a graduate level course in RL and Optimization. He effectively explores the synergies between reinforcement learning and stochastic optimization.

Part 5. Other Information

Not Applicable.