NEW DIRECTIONS IN PLANNING

Sridhar Mahadevan

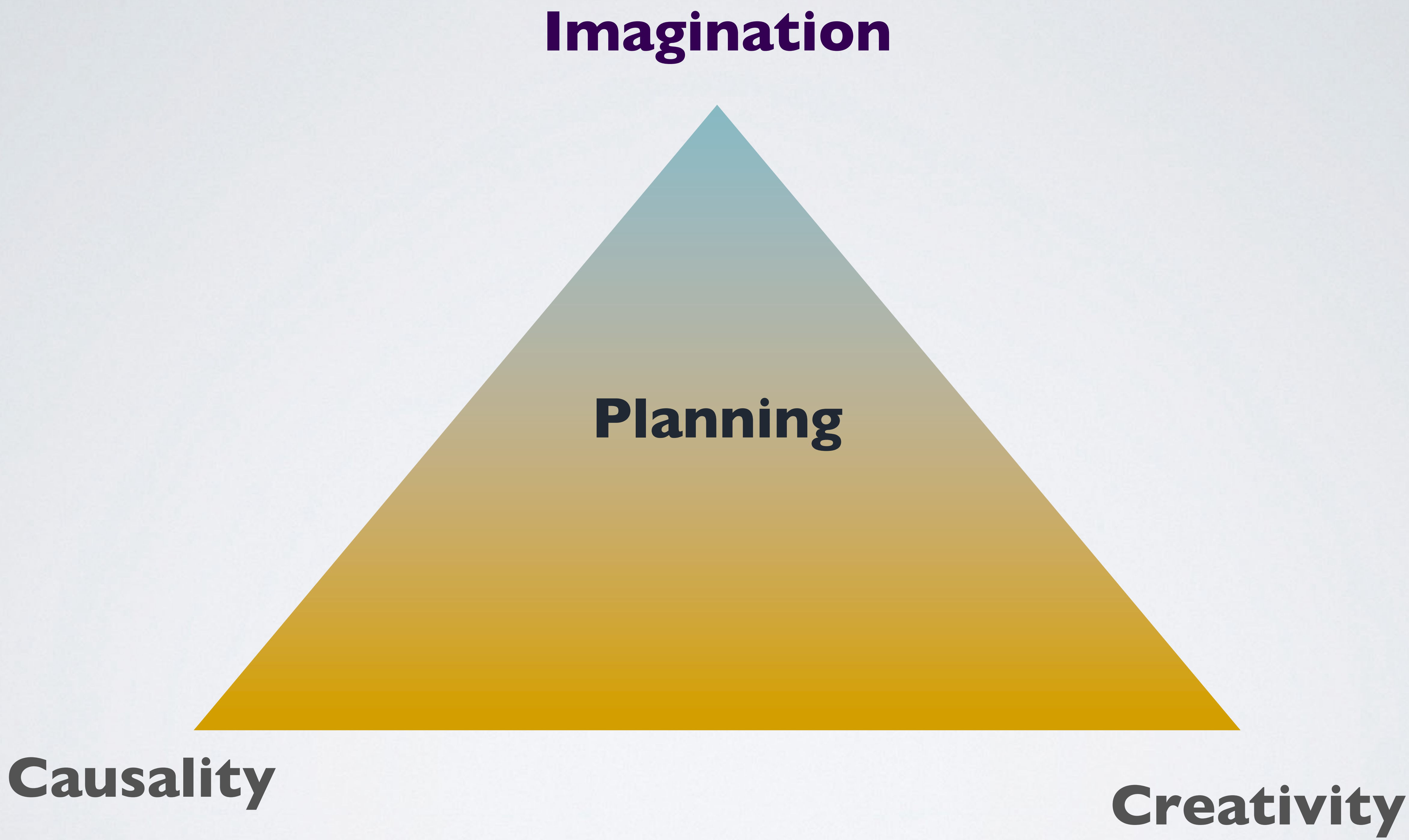
Adobe Research

UMass Amherst
College of Information and Computer Sciences

ICAPS Tutorial (October 20th 11:00-2:00 p.m. PST)

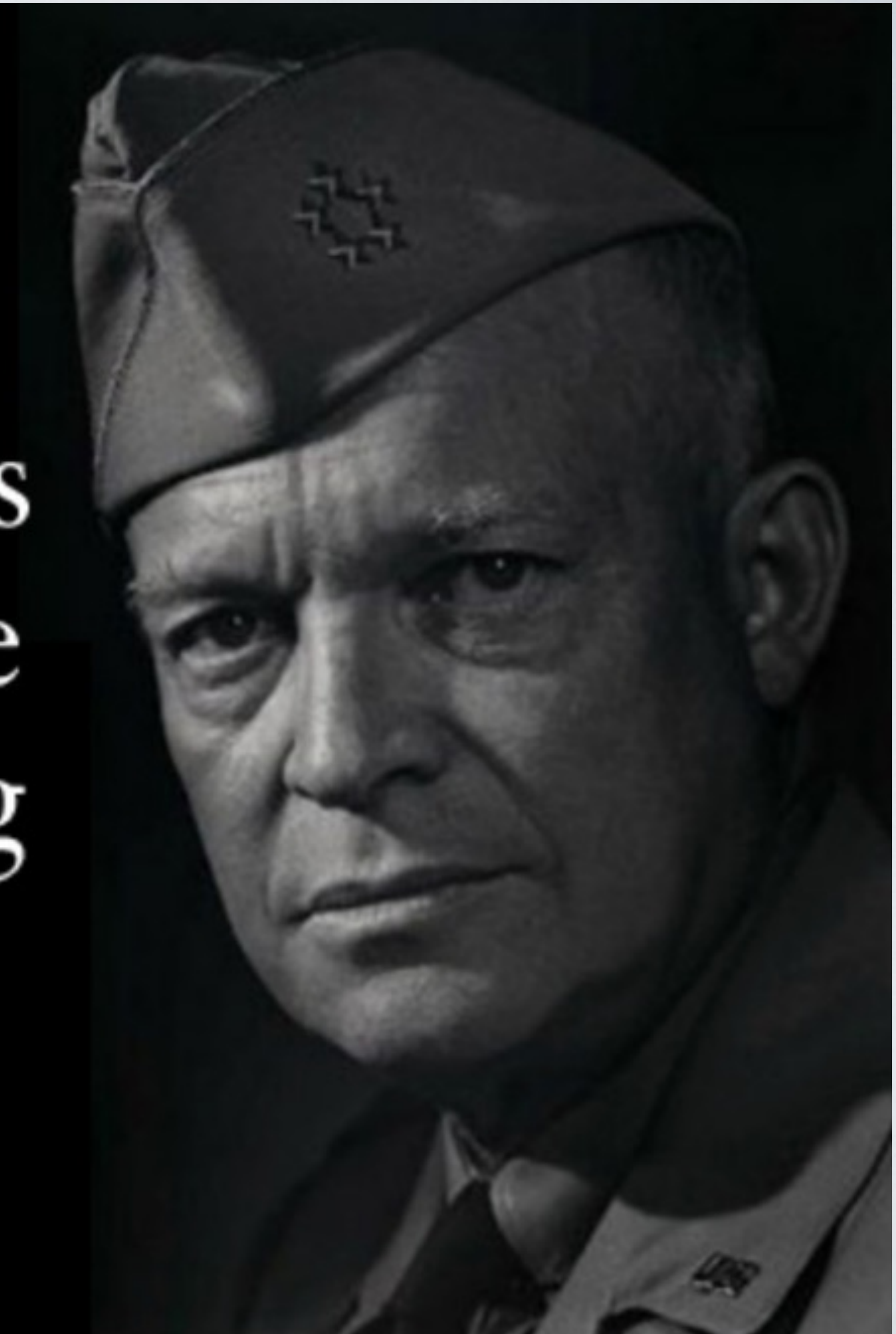
SCHEDULE OF TUTORIAL
(PACIFIC STANDARD TIME)

- Session 1: 11:00 a.m. – 12:00 noon
- Discussion: 12:00-12:15 p.m.
- Break: 12:15-12:45 p.m. (lunch, tea, dinner etc.)
- Session 2: 12:45-1:45 p.m.
- Discussion: 1:45-2:00 p.m.


Imagination
Planning
Causality
Creativity

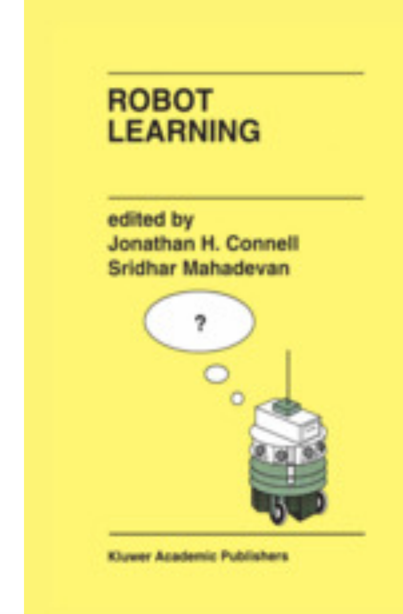
In preparing for battle, I have always found that plans are useless but planning is indispensable.

– Dwight D. Eisenhower



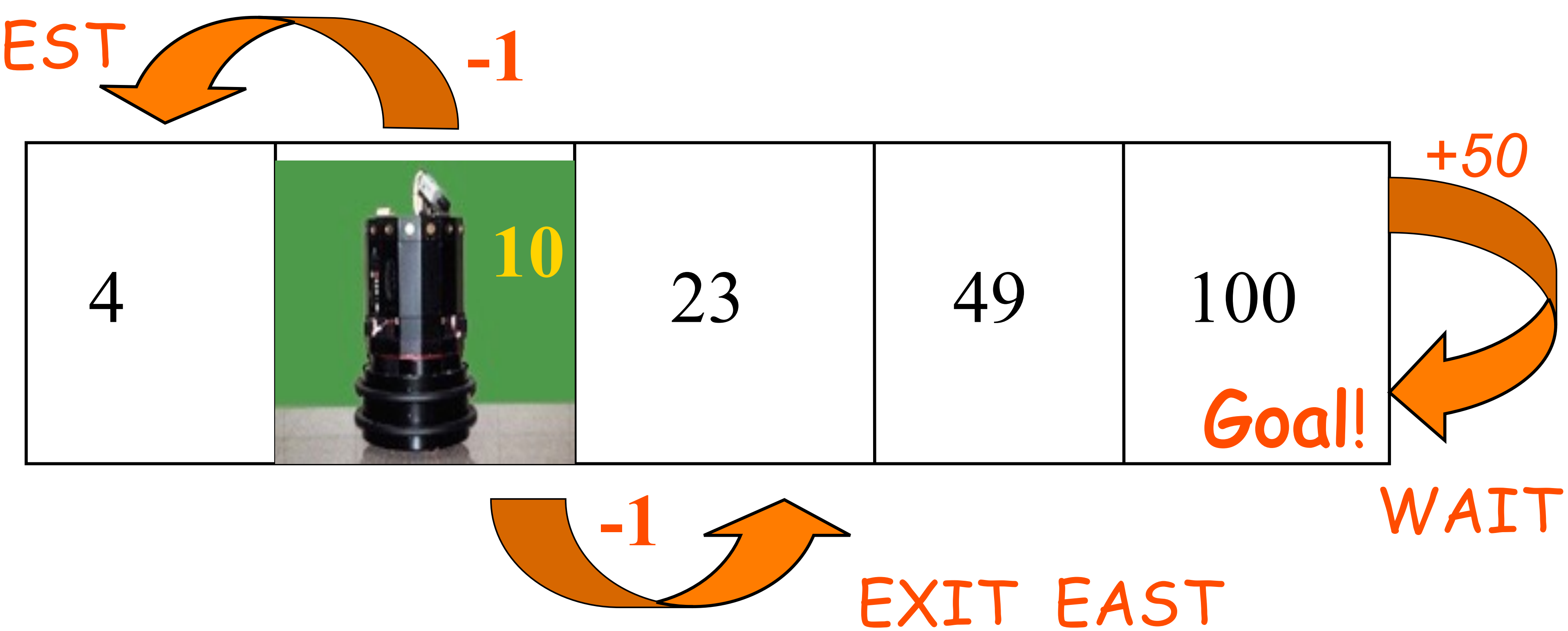
Markov Decision Processes (Bellman, Howard)

Kluwer
1992



$$V^*(x) = \max_{a \in A(x)} \left(r(x,a) + \gamma \sum_{y} P^a_{xy} V^*(y) \right)$$

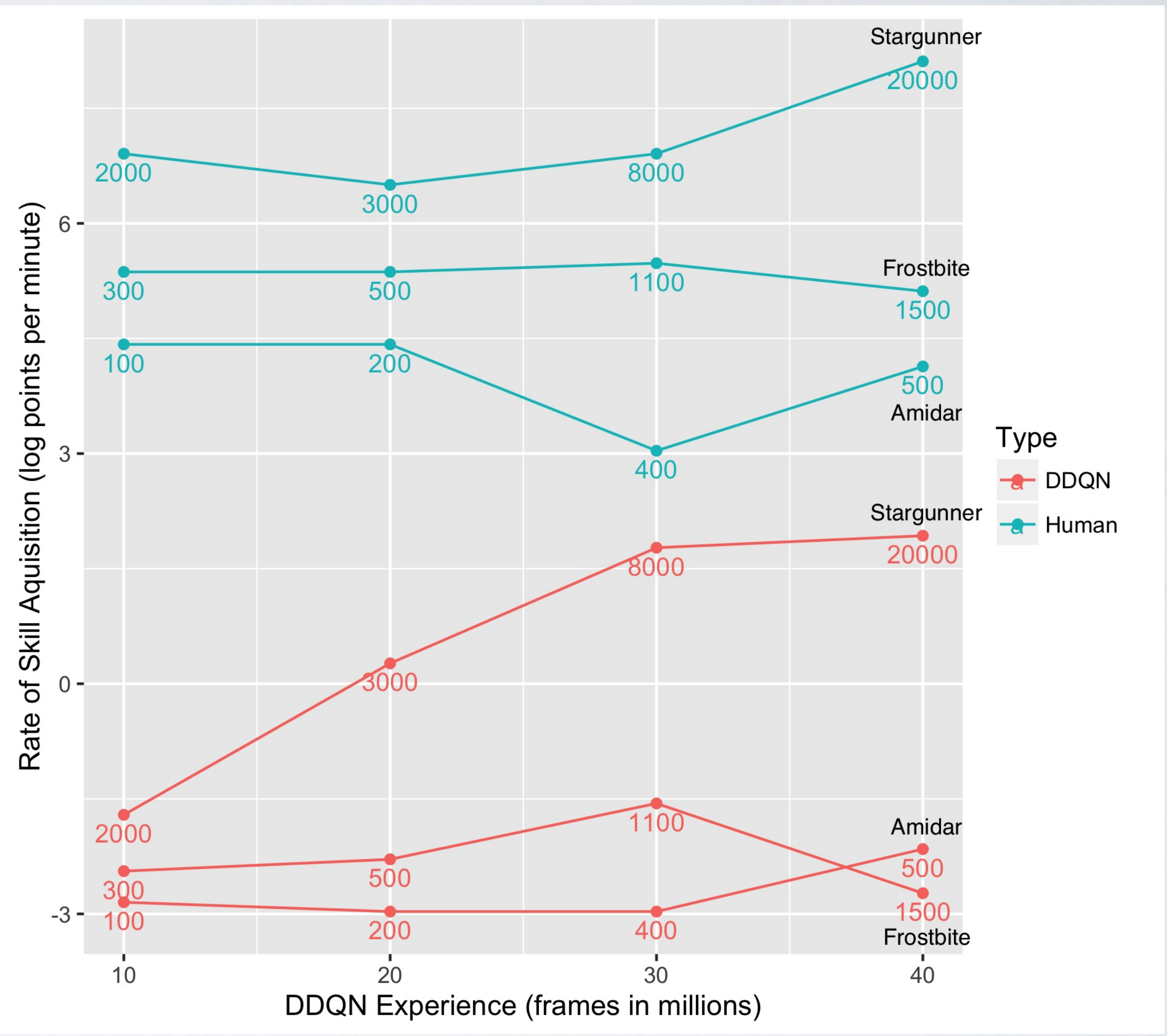
Choose
greedy action
given V*



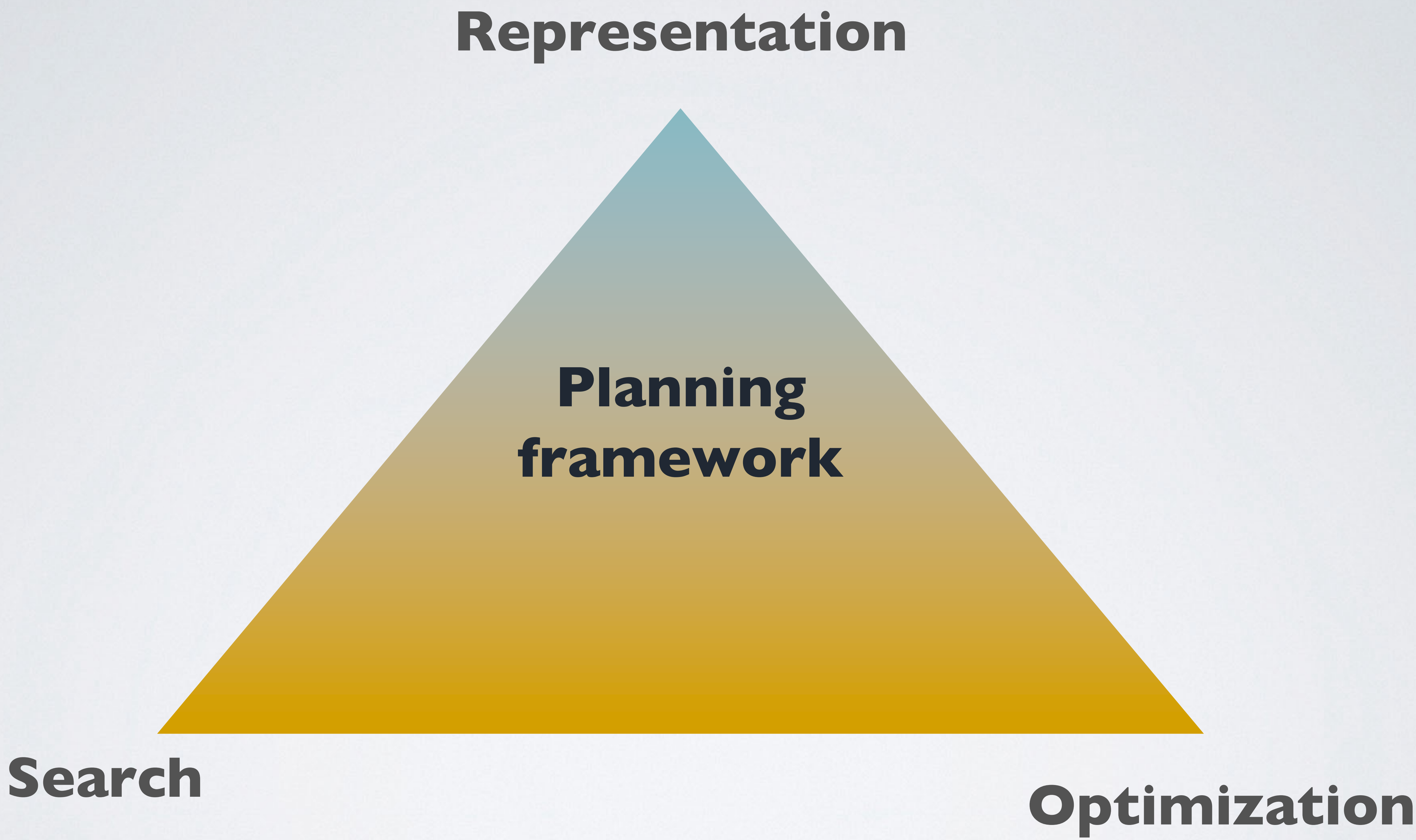
HUMANS VS DEEP RL

Frostbite

Humans learn Atari 1000x faster than any deep RL framework

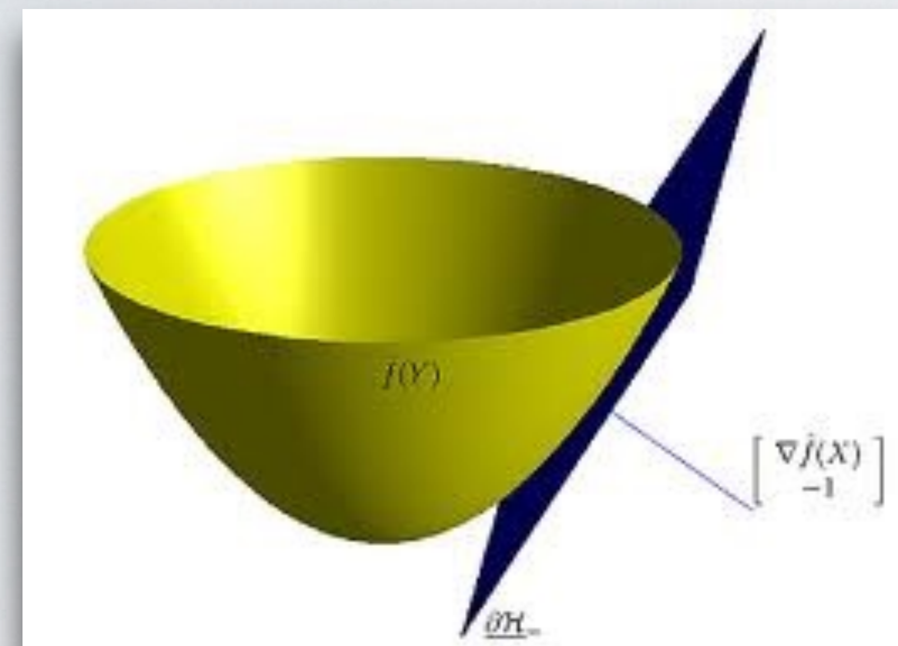


Tsividis et al.
AAAI 2017


Representation
Planning
framework
Search
Optimization

Optimization

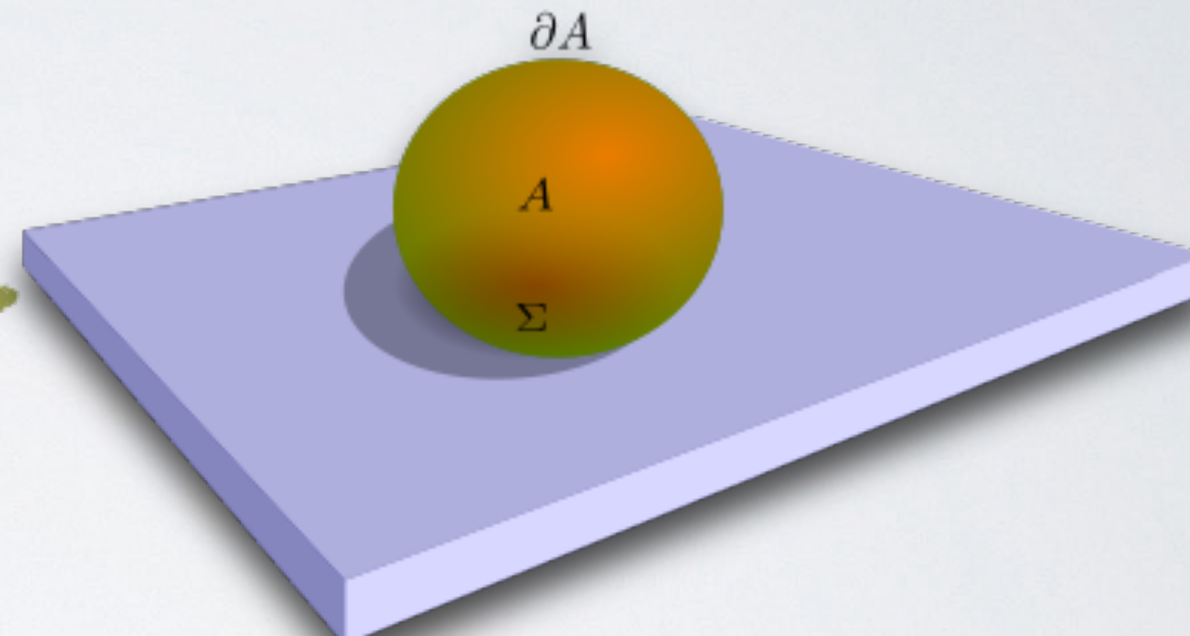
Equilibration



min f(x)

x in feasible set K

(Gauss, Newton, Shor, Hestenes 1800s-1950s)



$$\langle F(x^*), x - x^* \rangle \geq 0, \ \forall x \in K$$

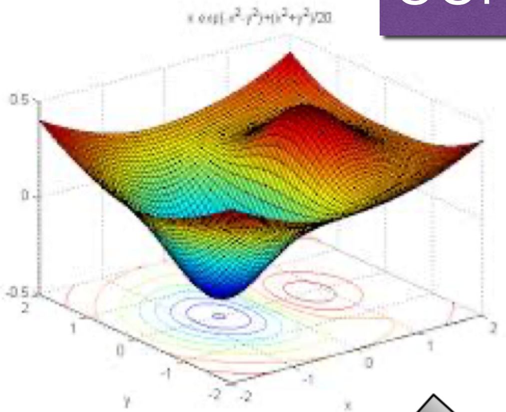
(Stampacchia, 1960s)

Optimization to Equilibration

AAAI 2015, AAMAS 2020 Tutorials

20th century AI

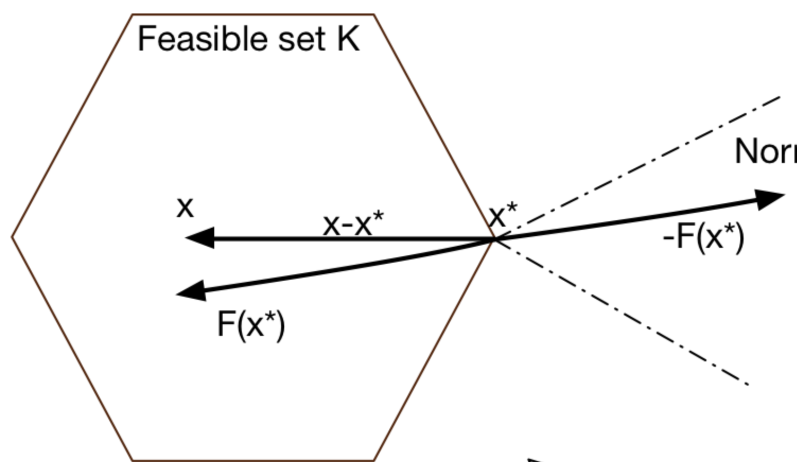
Optimization



		Player A	
		Cooperate	Defect
Player B	Cooperate	3, 3	1, 4
	Defect	4, 1	2, 2

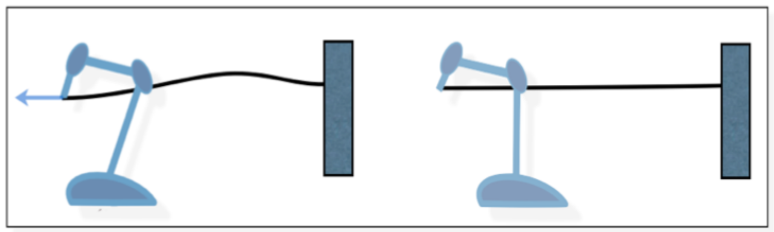
Game theory

21st century AI



Variational Inequalities

Complementarity problems



$$\frac{\partial u}{\partial x_1} + \frac{\partial u}{\partial x_2} = 0 \quad \text{is linear .}$$

$$\frac{\partial u}{\partial x_1} + \left(\frac{\partial u}{\partial x_2}\right)^2 = 0 \quad \text{is nonlinear .}$$

$$\frac{\partial u}{\partial x_1} + \frac{\partial u}{\partial x_2} + u^2 = 0 \quad \text{is nonlinear .}$$

$$\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} = x_1 \quad \text{is linear .}$$

$$\frac{\partial^2 u}{\partial x_1^2} + u\frac{\partial^2 u}{\partial x_2^2} = 0 \quad \text{is quasilinear .}$$

Nonlinear equation solving

Traffic equilibrium problem

GENERALIZING GRADIENT DESCENT

$$w_{t+1} \leftarrow w_t - \alpha_t \nabla f(w_t)$$

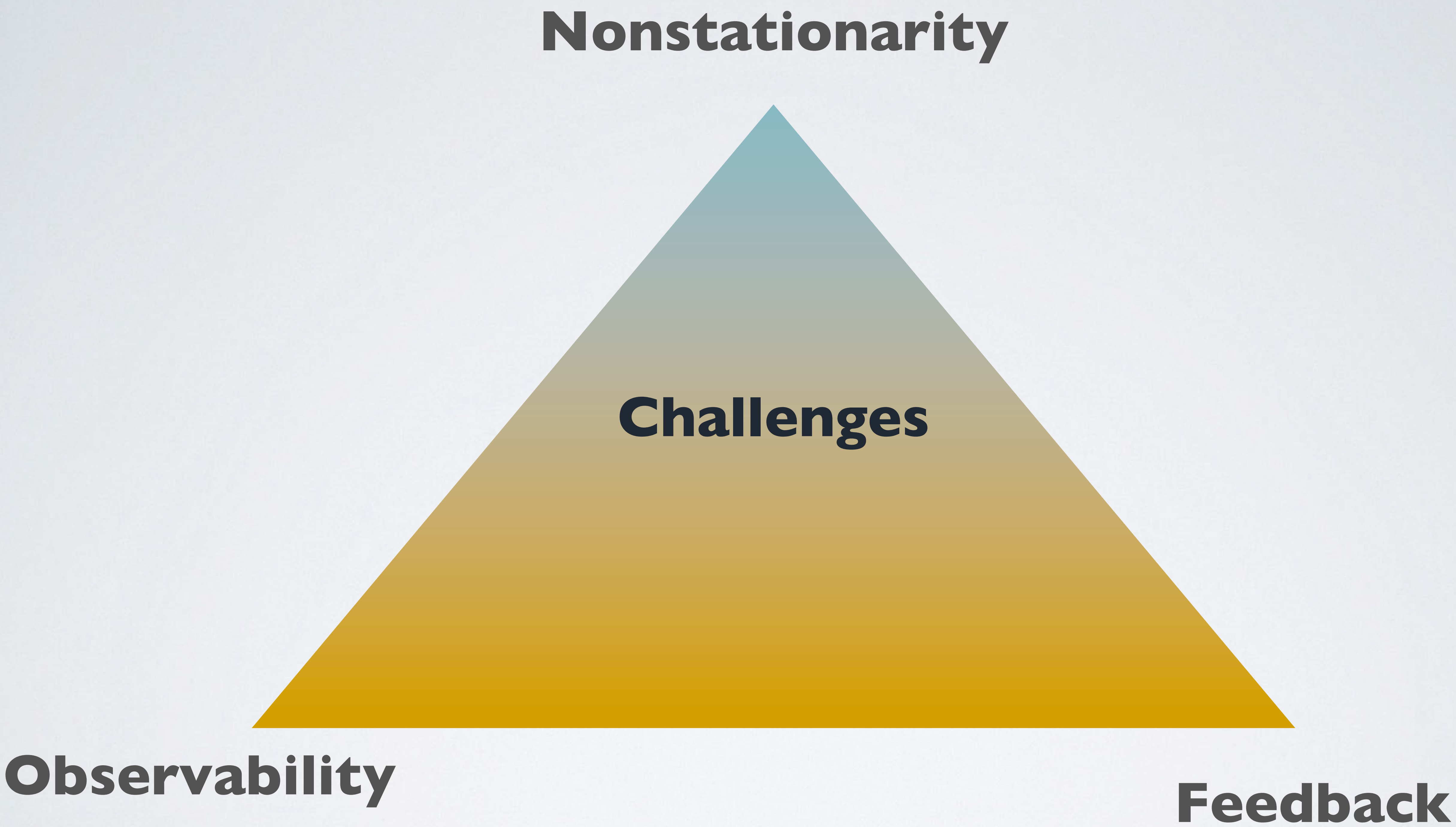
$$\downarrow$$

$$w_{t+1} \leftarrow w_t - \alpha_t F(w_t)$$

$$\downarrow$$

Extragradient method for VIs

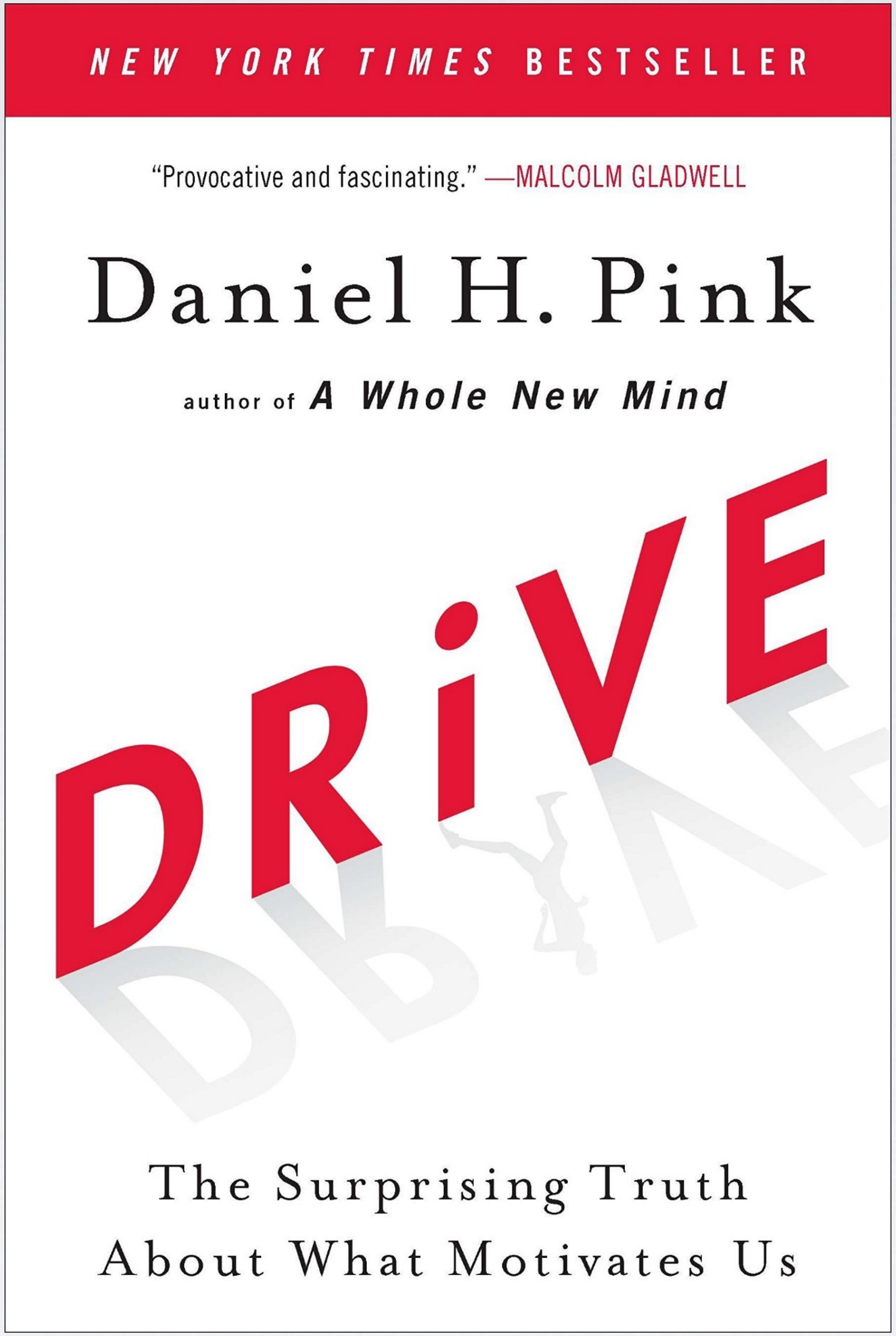
$$w_{t+1} \leftarrow \Pi_K(w_t - \alpha_t \Pi_K(w_t - \alpha_t F(w_t)))$$


Nonstationarity
Challenges
Observability
Feedback

SPARSITY OF REWARDS

- Often, rewards are very delayed and sparse
- What can substitute for task-specific rewards?
- Intrinsic rewards: encourage exploration to discover representations of long-term utility

INTRINSIC MOTIVATION

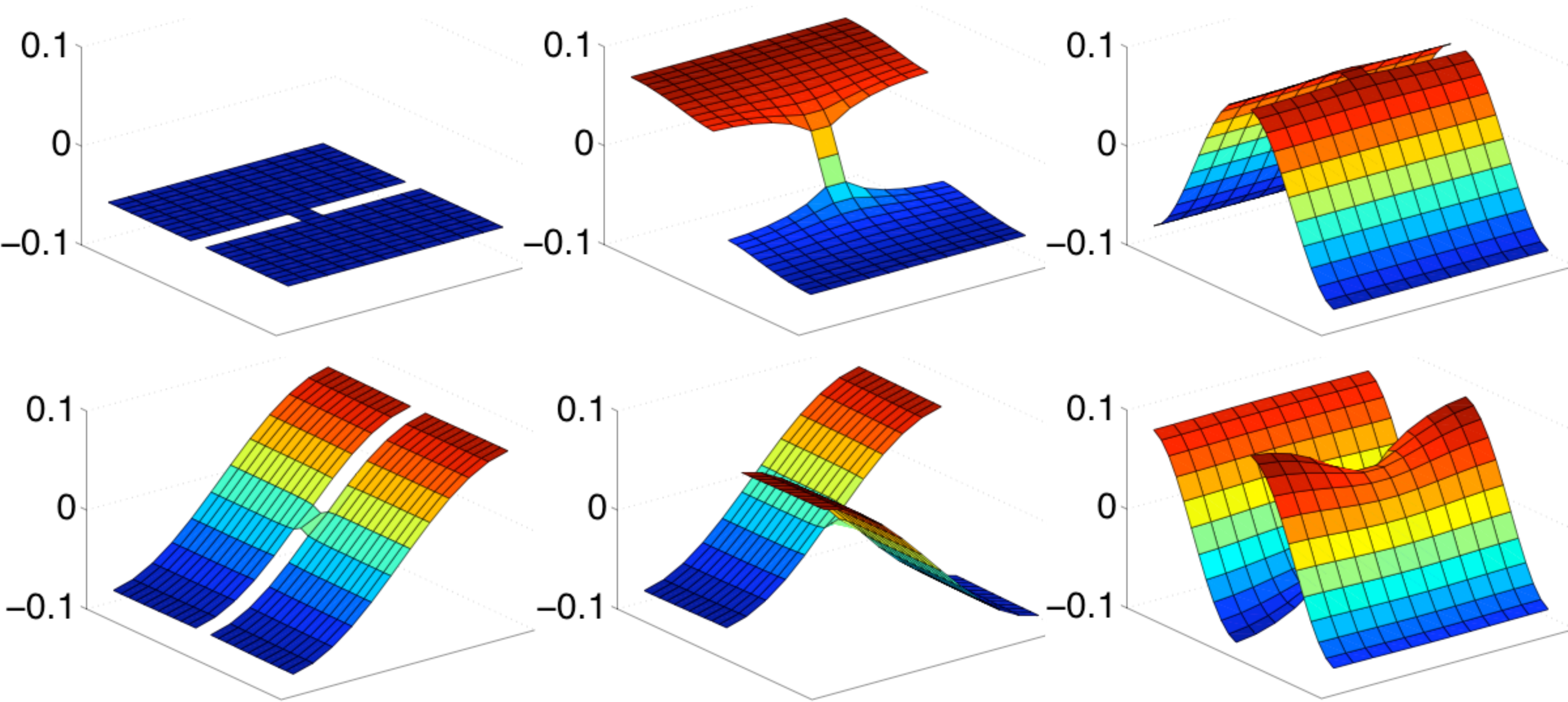


Proto-Value Functions

(Mahadevan, ICML 2005)

Eigenvectors of the graph Laplacian = D-W

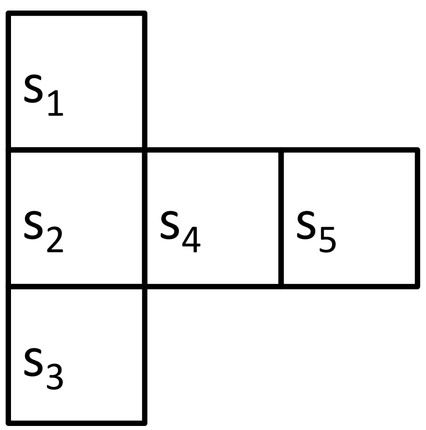
Φ



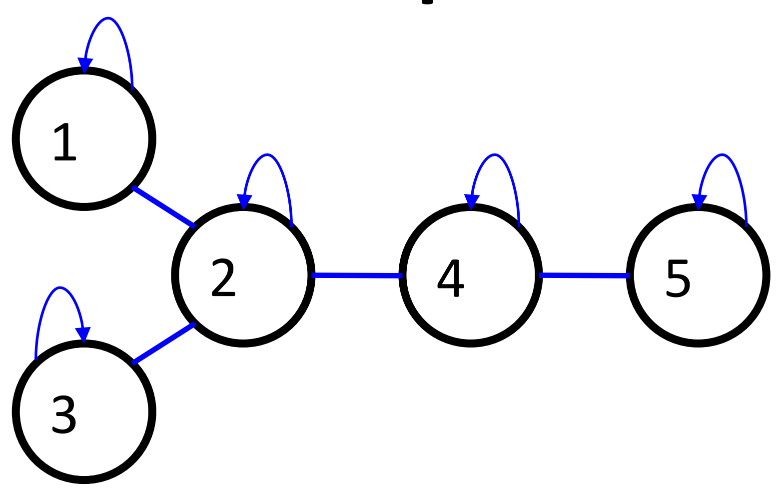
Reward-invariant, orthogonal, diagonalized, flat

Topological Representations for Planning

MDP



Graph



Weight Matrix, W

1	1	0	0	0
1	1	1	1	0
0	1	1	0	0
0	1	0	1	1
0	0	0	1	1

Random walk on graph = $D^{-1} W$

[Mahadevan, ICML 2005; Johns and Mahadevan, ICML 2007; Osentoski and Mahadevan, ICML 2007]

The Successor Representation: Its Computational Logic and Neural Substrates

Journal of Neuroscience 15 August 2018, 38 (33) 7193-7200;

Assuming rewards are linear functions of features: $R(s_t) = \phi_{s_t} \cdot \mathbf{w}$,

$$Q^\pi(s,a) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t) | s_0 = s, a_0 = a\right], \quad (4)$$

$$= \mathbb{E}\left[R(s_0) + \gamma^1 R(s_1) + \gamma^2 R(s_2) + \ldots | s_0 = s, a_0 = a\right] \quad (5)$$

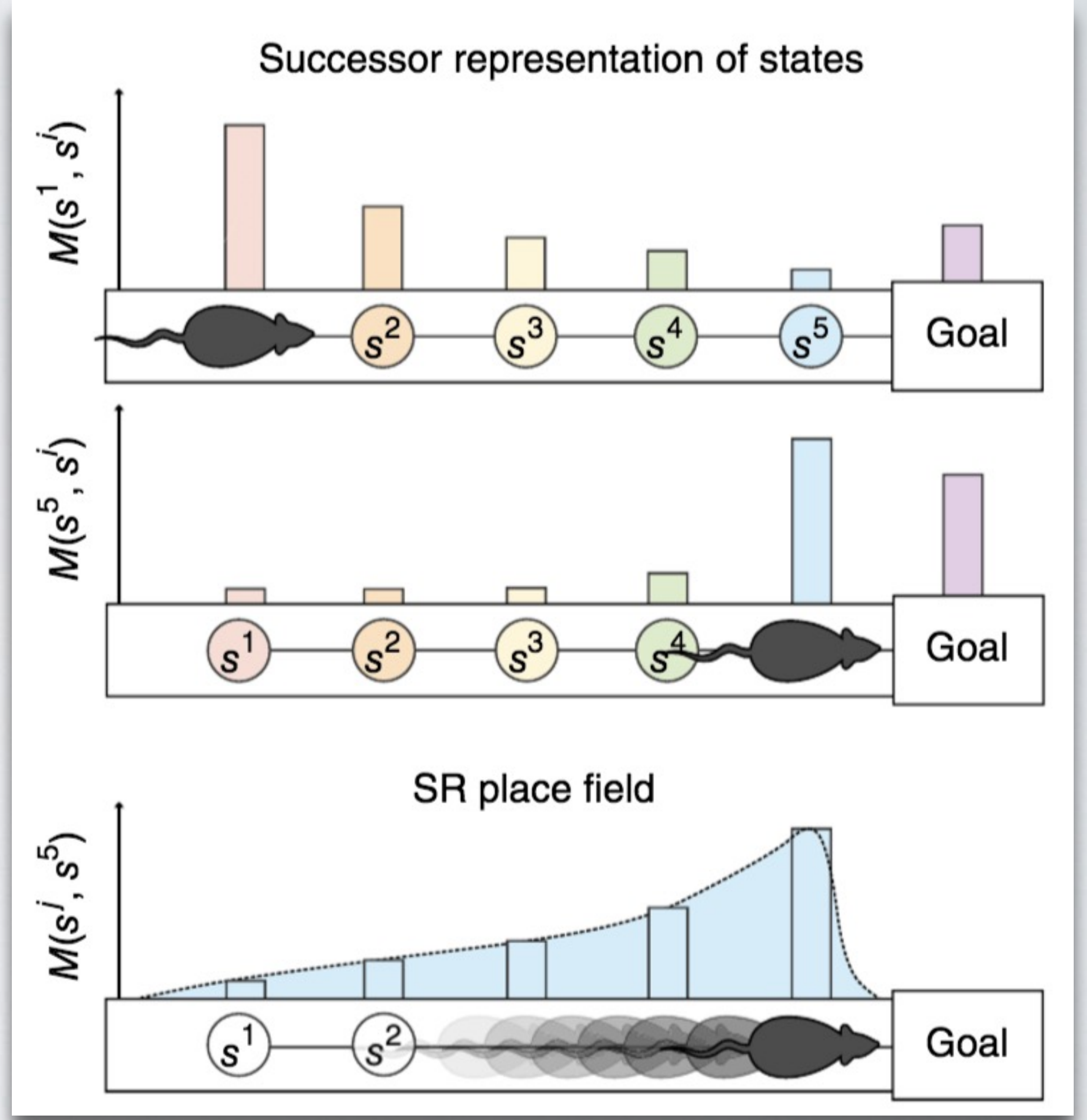
$$= \mathbb{E}\left[\phi_{s_0} \cdot \mathbf{w} + \gamma^1 \phi_{s_1} \cdot \mathbf{w} + \gamma^2 \phi_{s_2} \cdot \mathbf{w} + \ldots | s_0 = s, a_0 = a\right] \quad (6)$$

$$= \mathbb{E}\left[\phi_{s_0} + \gamma^1 \phi_{s_1} + \gamma^2 \phi_{s_2} + \ldots | s_0 = s, a_0 = a\right] \cdot w \quad (7)$$

$$= M^\pi(s,a) \cdot w \quad (8)$$

M represents the policy dependent expected features or a **partial model** and w represents the goal.

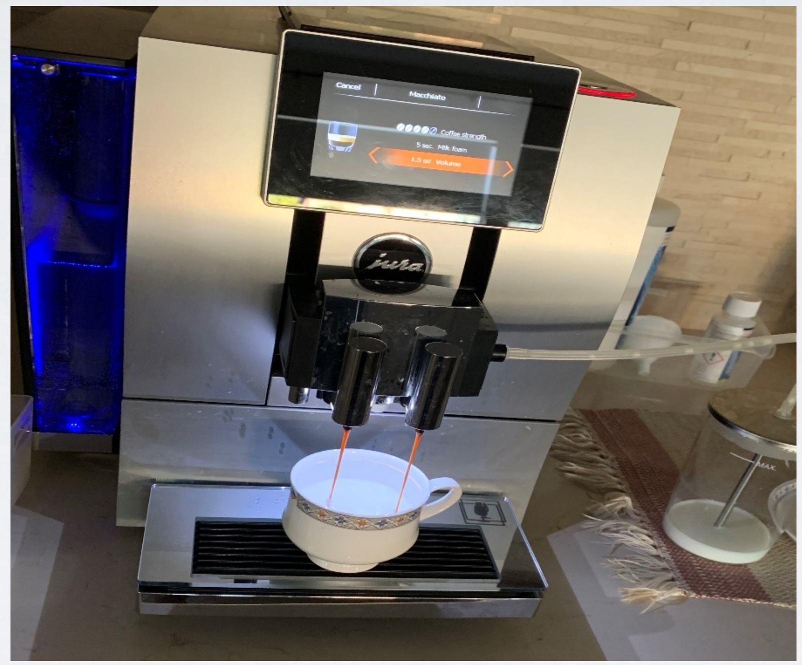
(Dayan, MLJ)



PVF & SUCCESSOR REPRESENATIONS

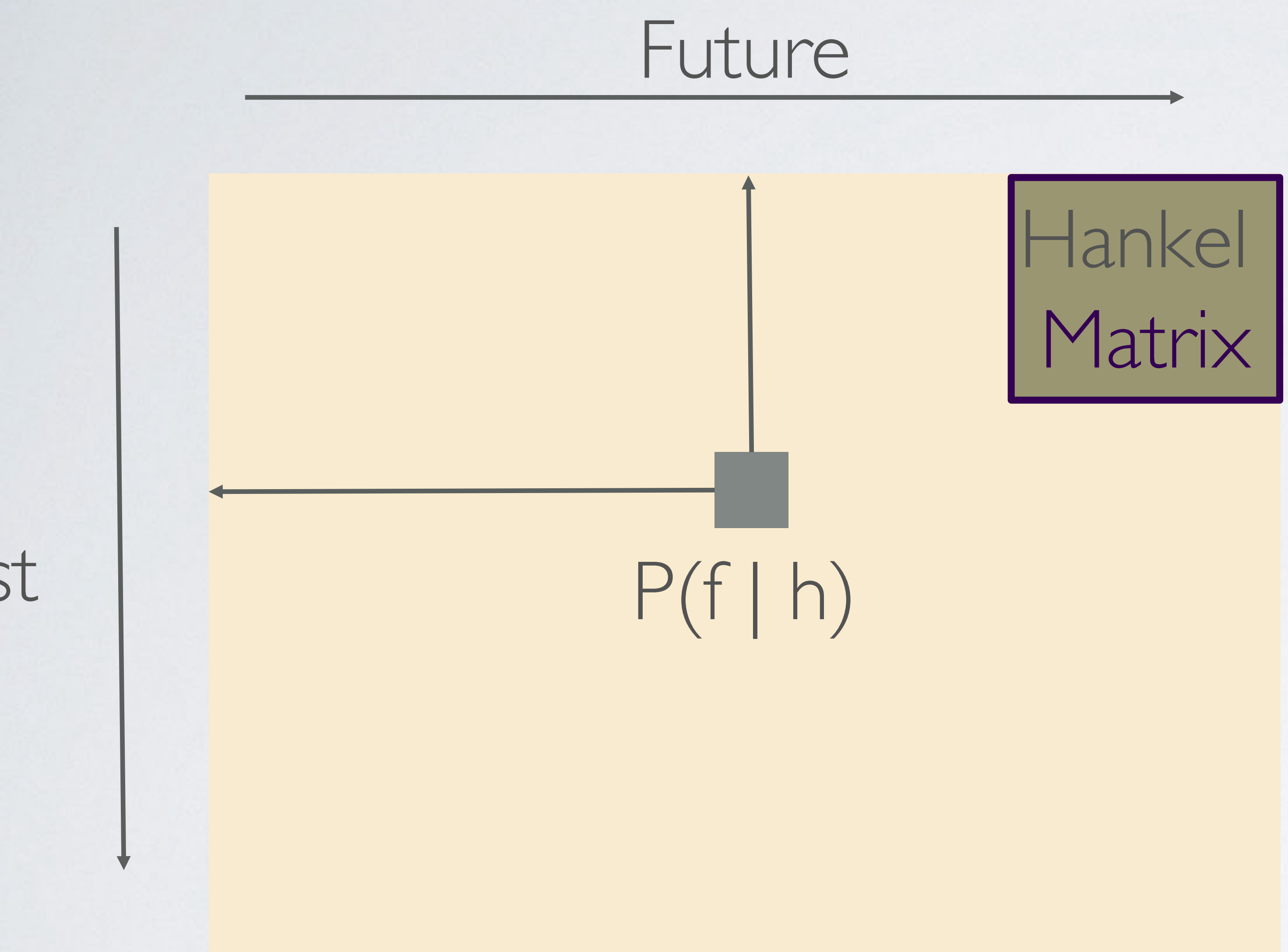
- These two representations are closely related (Stachenfield, 2014)
- Another related representation is slow feature analysis (SFA, Wiscott & Sejnowski, 2002)
- It is possible to unify and generalize these approaches using more sophisticated topological concepts

TOPOLOGY AND AFFORDANCE




We know the past but cannot control it. We control the future but cannot know it.
— Claude Shannon —
AZ QUOTES

SYSTEM IDENTIFICATION



Observable operators using

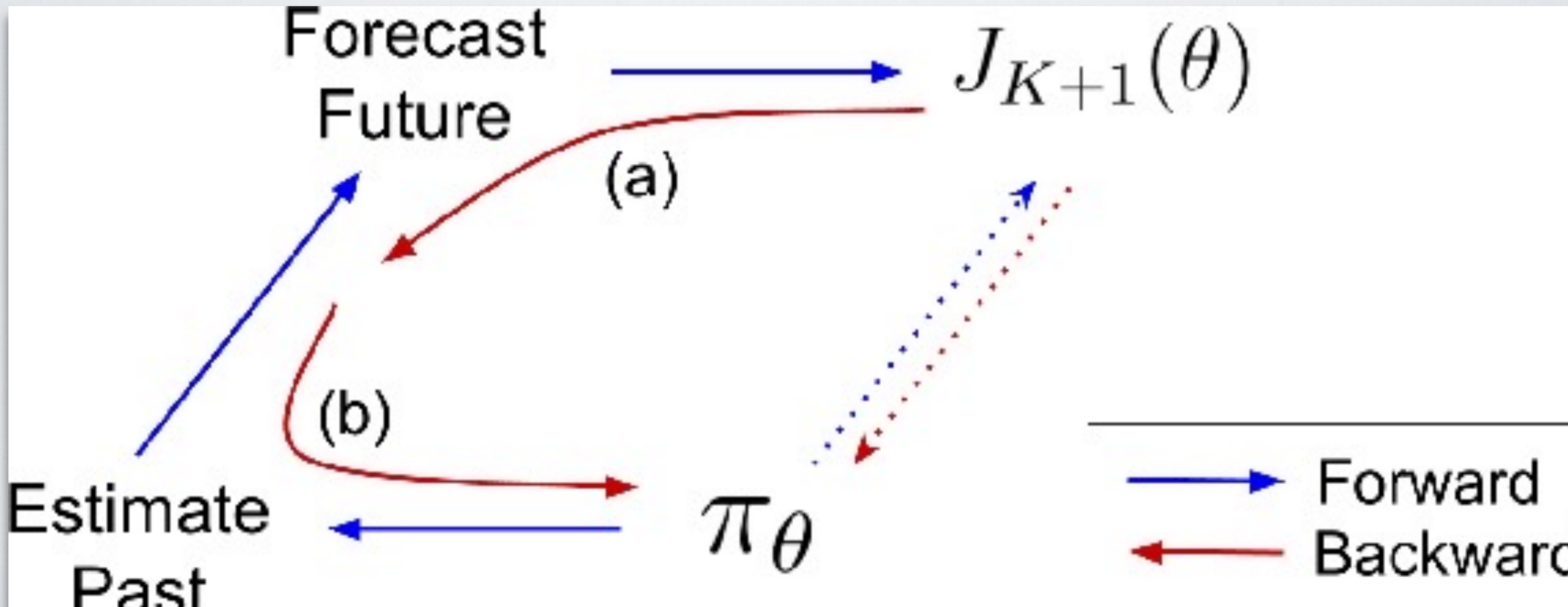
Canonical Variate Analysis (Overschee, et al.)

Originally pioneered in algebraic automata theory (Schutzenberger, 1960s)
Rediscovered in AI much later: OOMs, PSRs, TPSRs (2005-x)

RL IN NONSTATIONARY WORLDS

(Chandak et al., ICML 2020)

Learn a policy that optimizes for tomorrow's world, not today's!



Imagination

First published Mon Mar 14, 2011; substantive revision Mon Oct 8, 2018

To imagine is to form a mental representation that does not aim at things as they actually, presently, and subjectively are. One can use imagination to represent possibilities other than the actual, to represent times other than the present, and to represent perspectives other than one's own. Unlike perceiving and believing, imagining something does not require one to consider that something to be the case. Unlike desiring or anticipating, imagining something does not require one to wish or expect that something to be the case.

Imagination is involved in a wide variety of human activities, and has been explored from a wide range of philosophical perspectives. Philosophers of mind have examined imagination's role in mindreading and in pretense. Philosophical aestheticians have examined imagination's role in creating and in engaging with different types of artworks. Epistemologists have examined imagination's role in theoretical thought experiments and in practical decision-making. Philosophers of language have examined imagination's role in irony and metaphor.

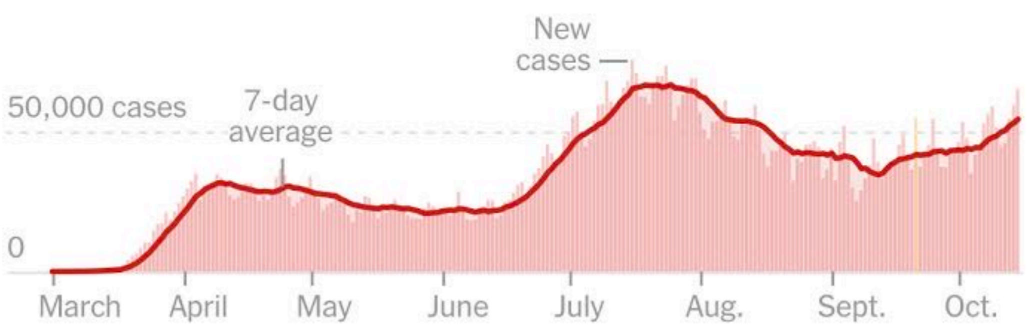
WORLD COUNTRIES ▾ | U.S.A. STATES ▾ COLLEGES

Covid in the U.S.: Latest Map and Case Count

By The New York Times

Updated October 16, 2020, 12:06 P.M. E.T.

Leer en español



	TOTAL REPORTED	ON OCT. 15	14-DAY CHANGE
Cases	8.0 million+	65,327	+25%
Deaths	217,867	793	–2%

Day with data reporting anomaly.

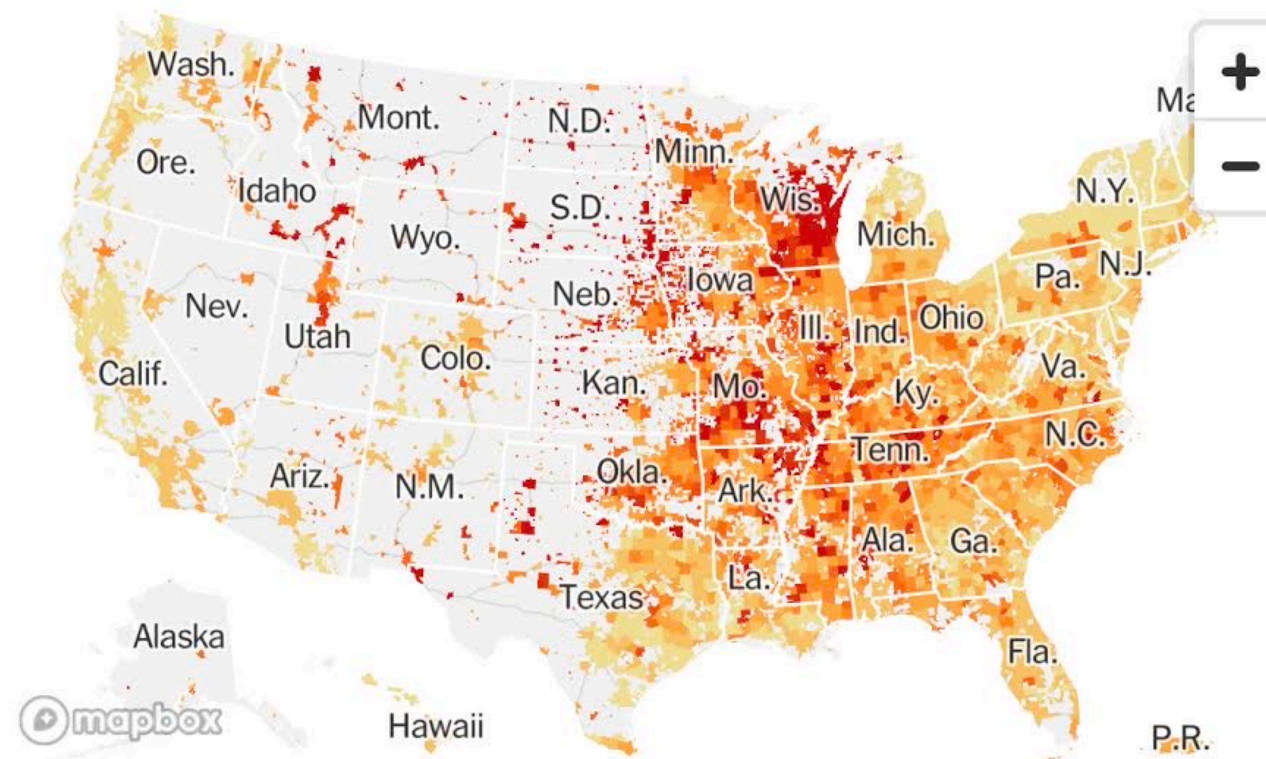
Includes confirmed and probable cases where available. 14-day change trends use 7-day averages.

Hot spots | Total cases | Deaths | Per capita

Average daily cases per 100,000 people in the past week

8 24 40 56 Few or no cases

Use two fingers to pan and zoom. Tap for details.



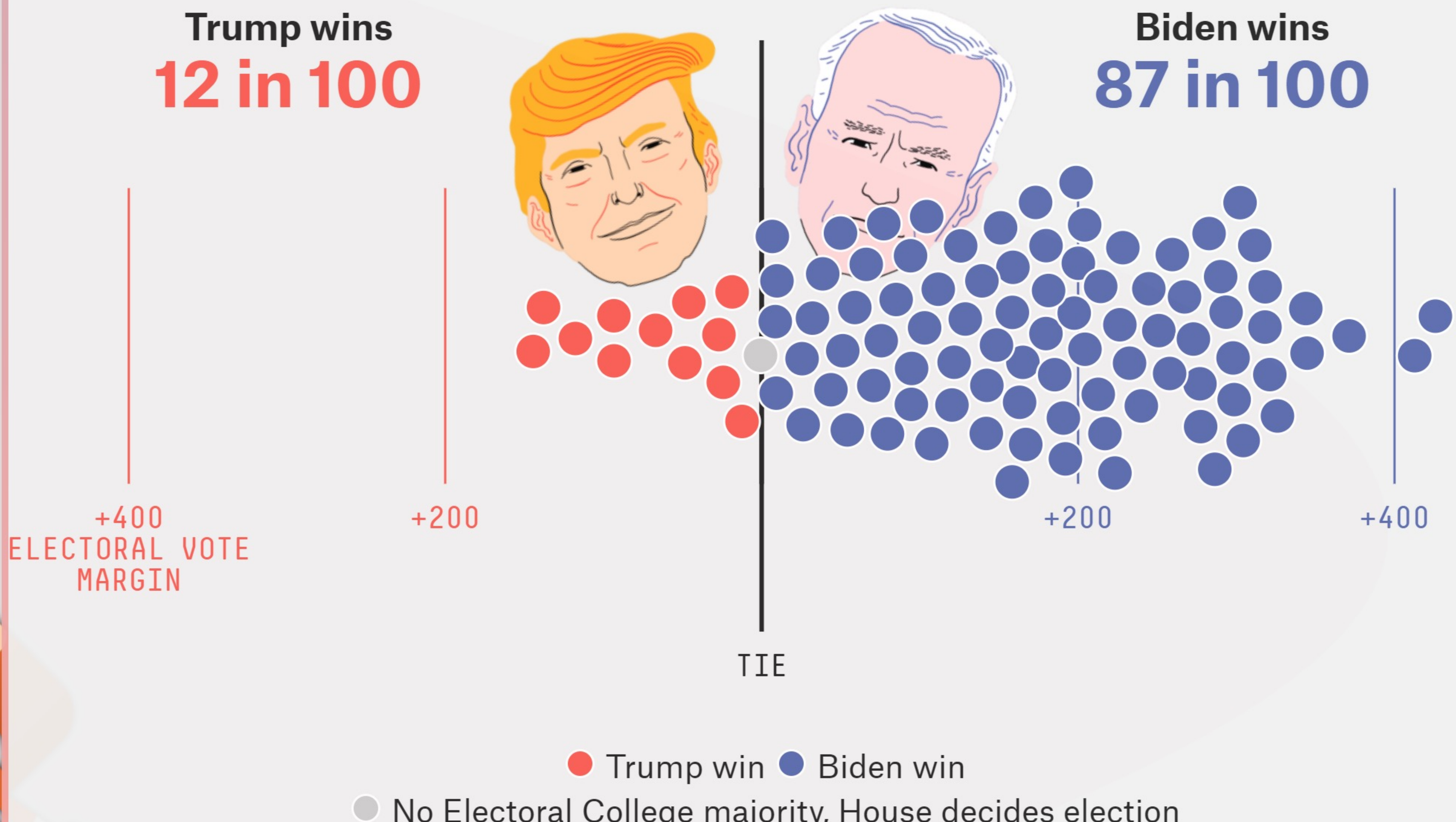
Sources: State and local health agencies. Population and demographic data from Census Bureau.

▶ About this data

fivethirtyeight.com

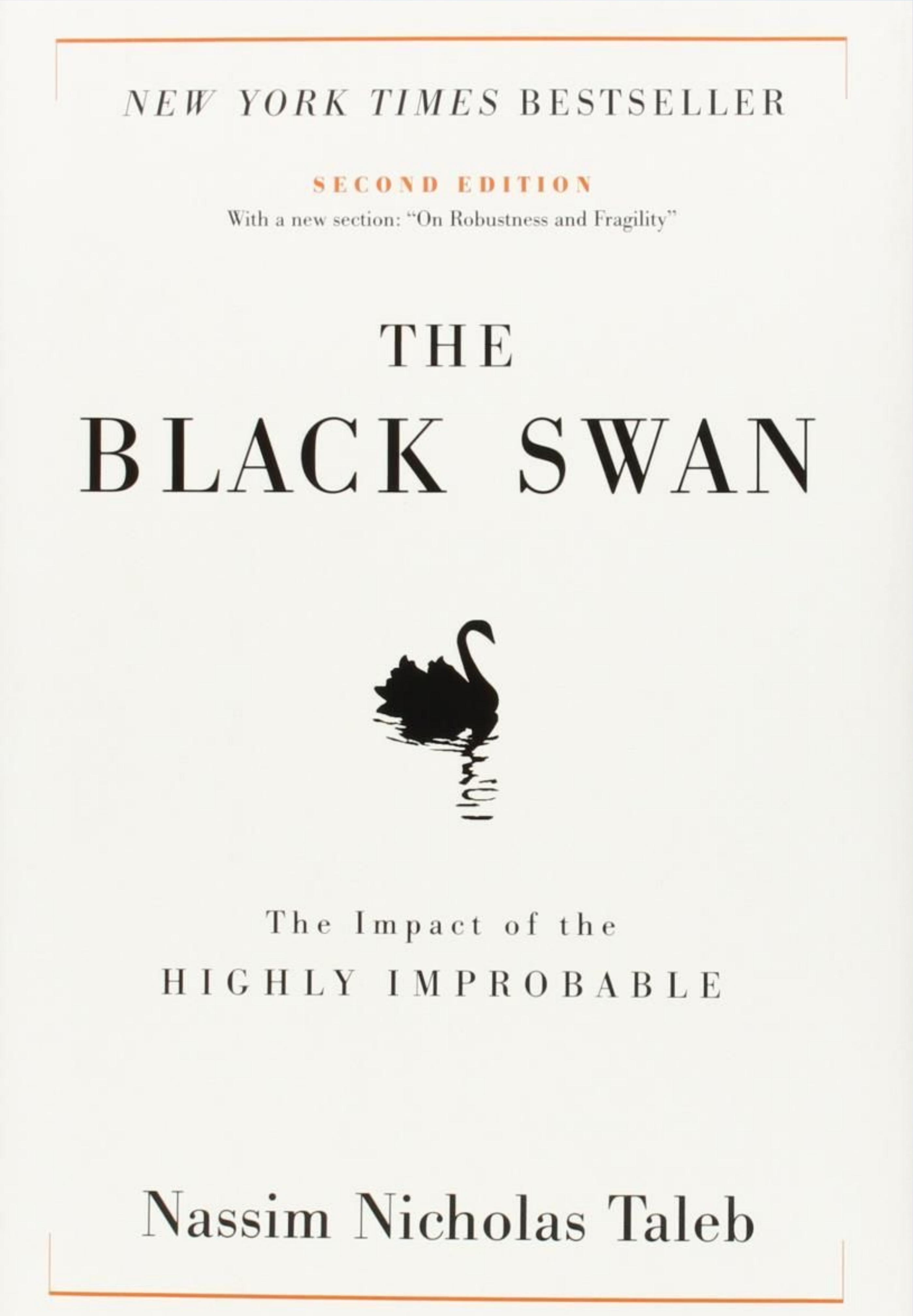
Biden is *favored* to win the election

We simulate the election 40,000 times to see who wins most often. The sample of 100 outcomes below gives you a good idea of the range of scenarios our model thinks is possible.



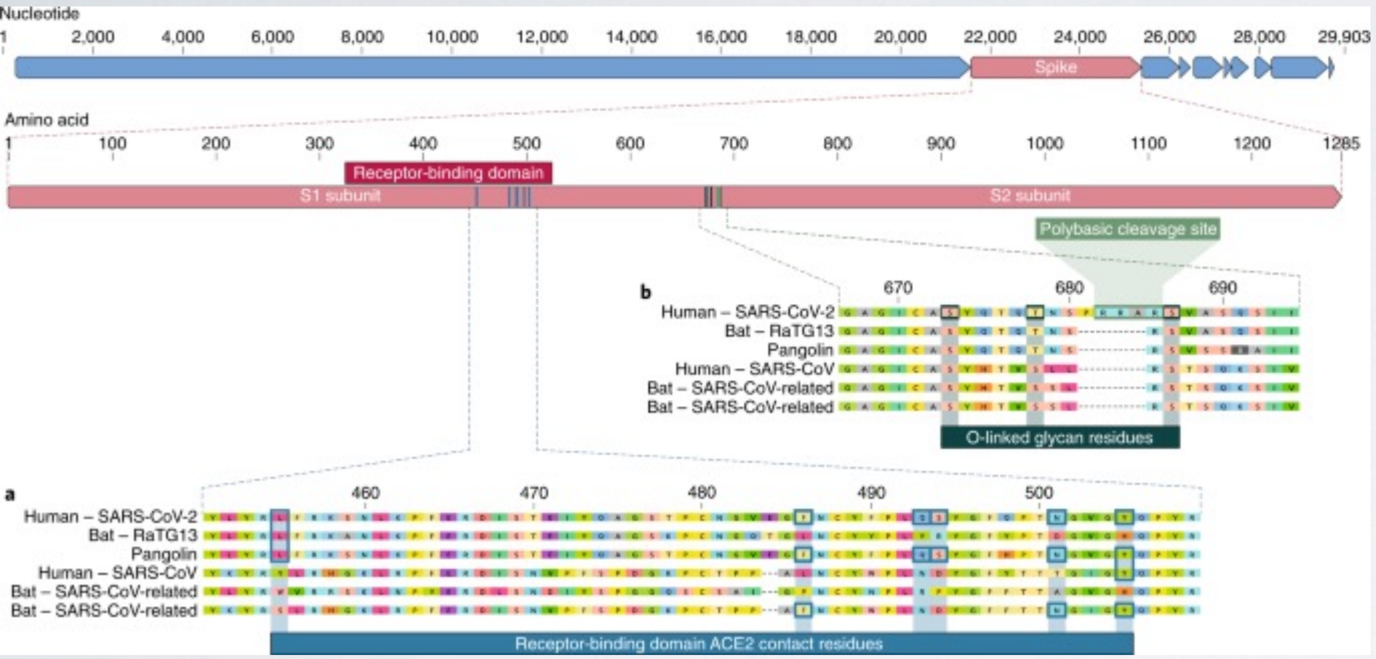
2020: Covid-19
2008: Financial crisis
2001: World Trade Center

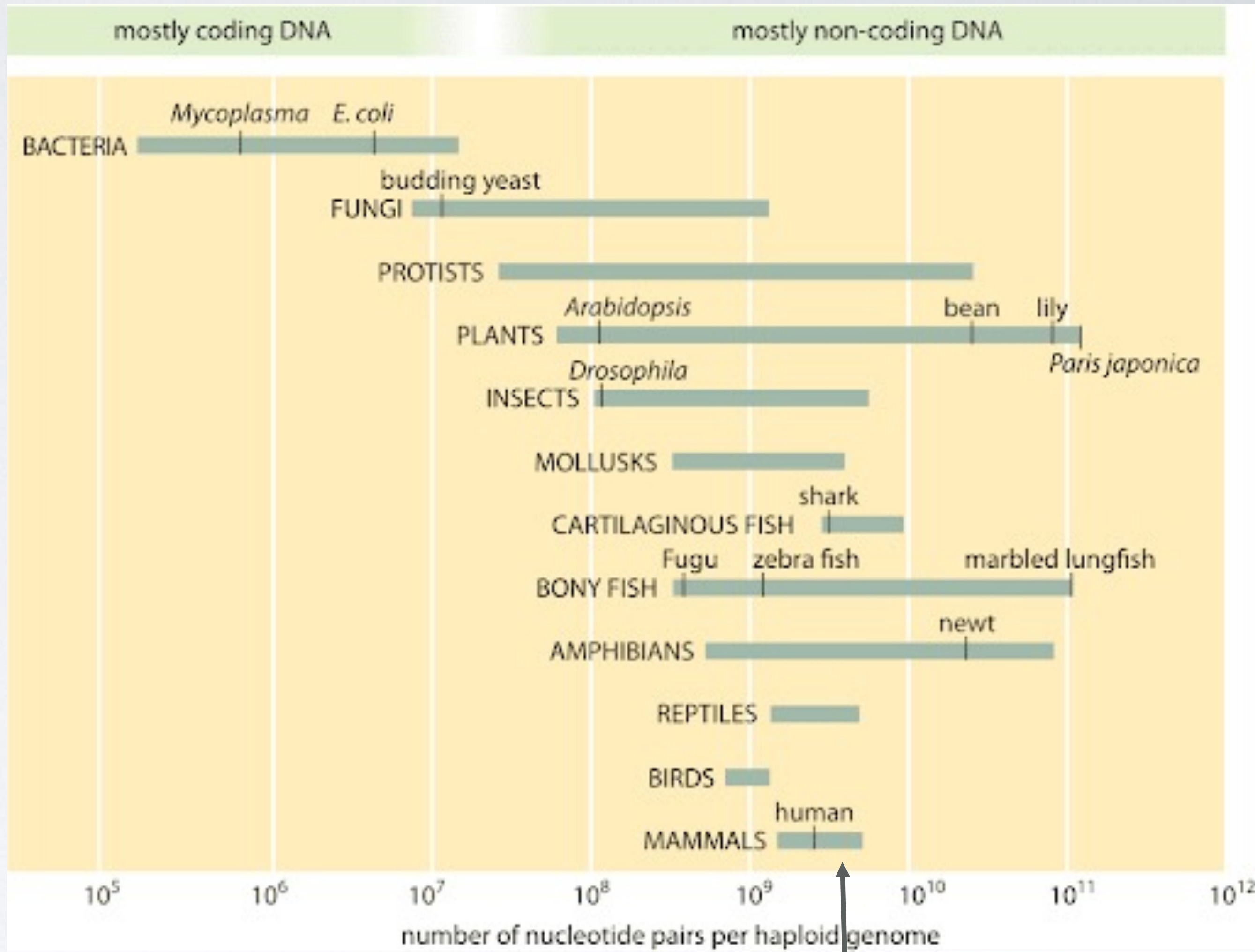
AI/ML models
fail in modeling
Black swan
outlier events



Virus genome is 100,000 times smaller than human DNA

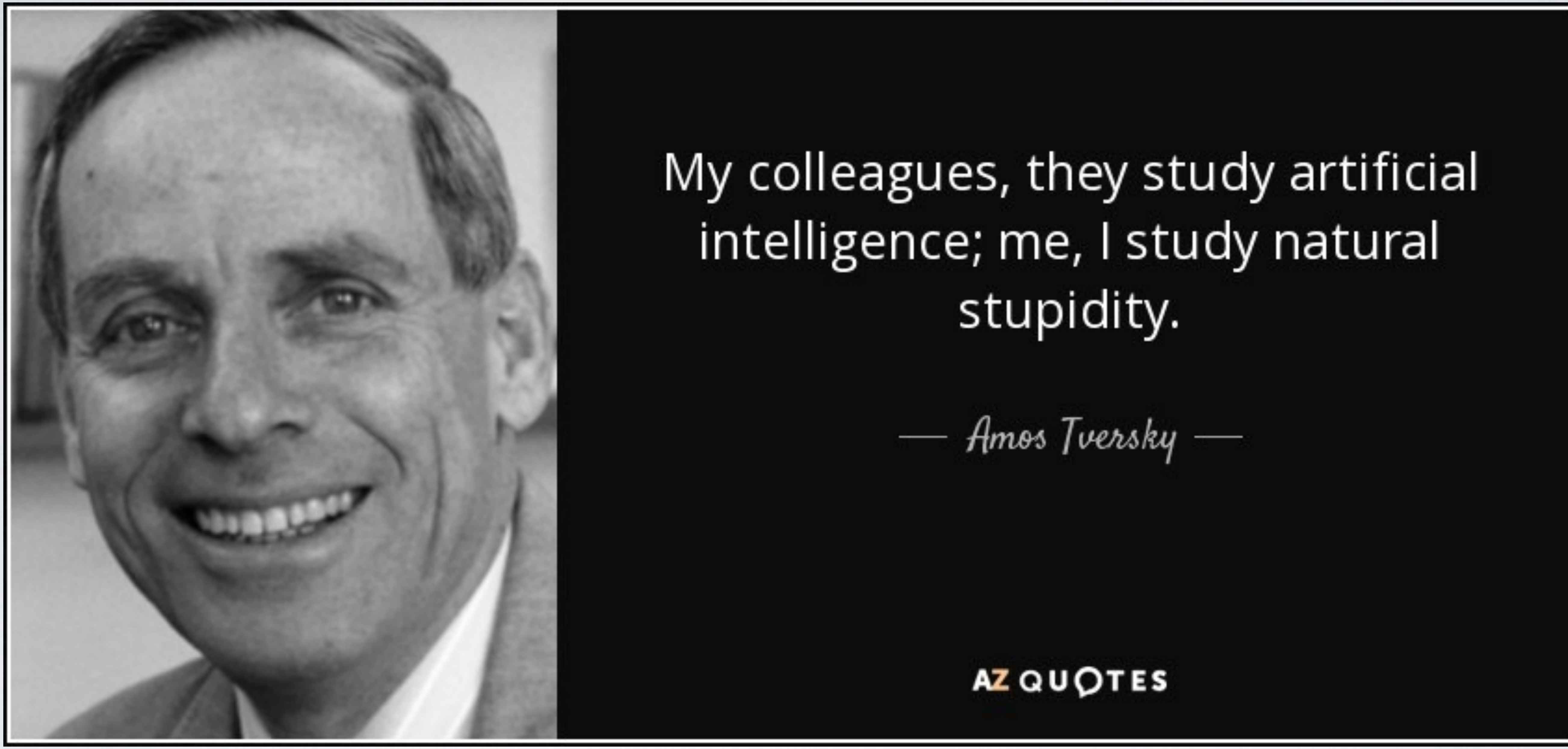
Corona virus: 30,000 base pairs





Human DNA

SUPER-SPREADER COVID-19 EVENTS


My colleagues, they study artificial intelligence; me, I study natural stupidity.
— Amos Tversky —
AZ QUOTES

Common Myths around
COVID-19

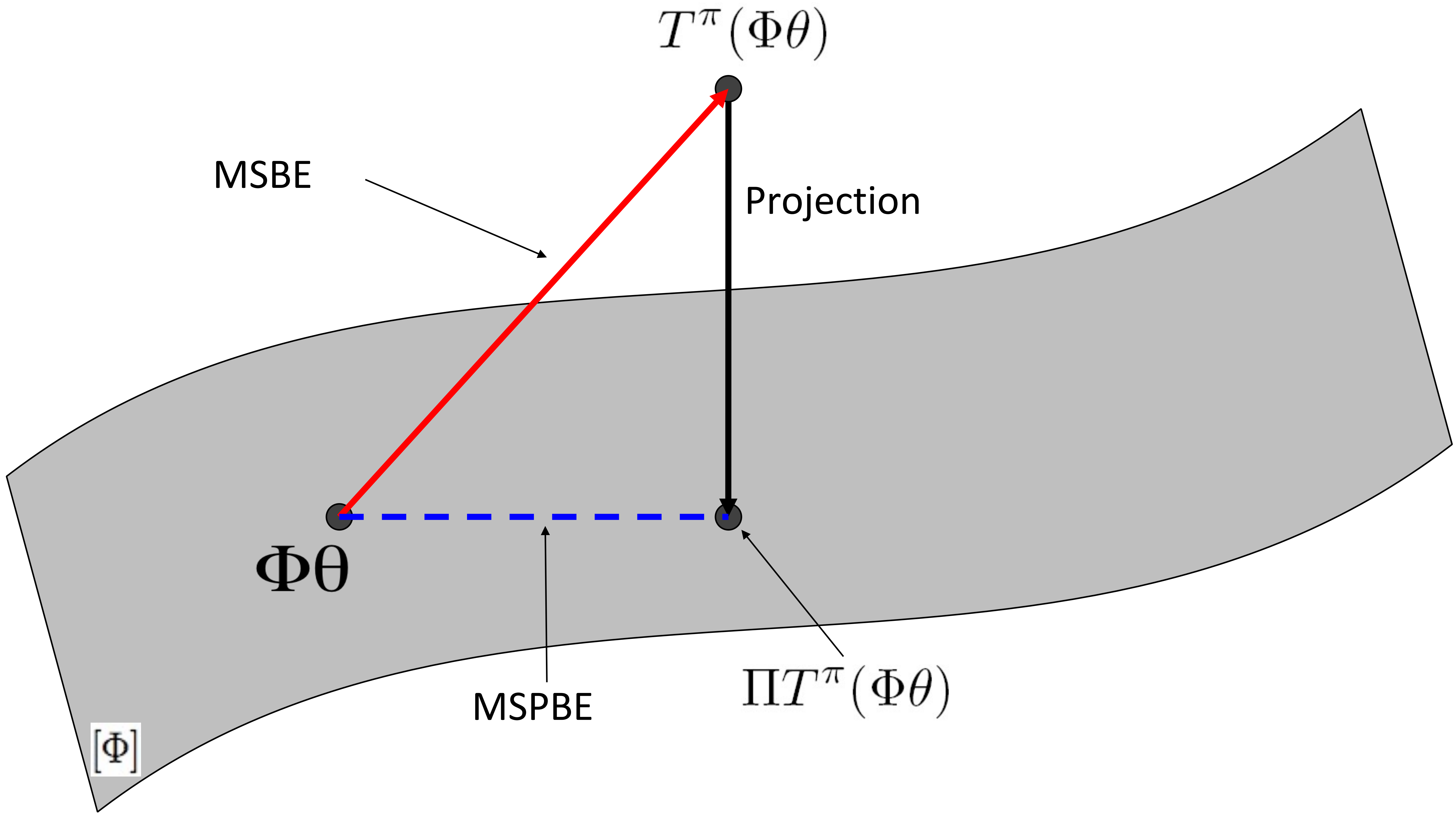
- Ayurveda and other herbal medicines can help cure coronavirus
- Only old people are affected by coronavirus
- Onset of summers mean a decline of the virus
- Eating meat can cause coronavirus
- Spraying alcohol can kill coronavirus
- Vitamin C can cure the coronavirus infection
- Eating garlic will help prevent the infection
- Taking a hot bath can prevent the virus from infecting you
- You don't have COVID if you can hold your breath for ten seconds without discomfort
- Dogs can spread coronavirus

INFERENCE IN SOCIAL NETWORKS

Examples:
Traffic
Migration
Pandemic
Conspiracy
Marketing
Healthcare

(Manski, 2013; Hudgins and Halloran, 2018)

PROJECTIONS IN REINFORCEMENT LEARNING



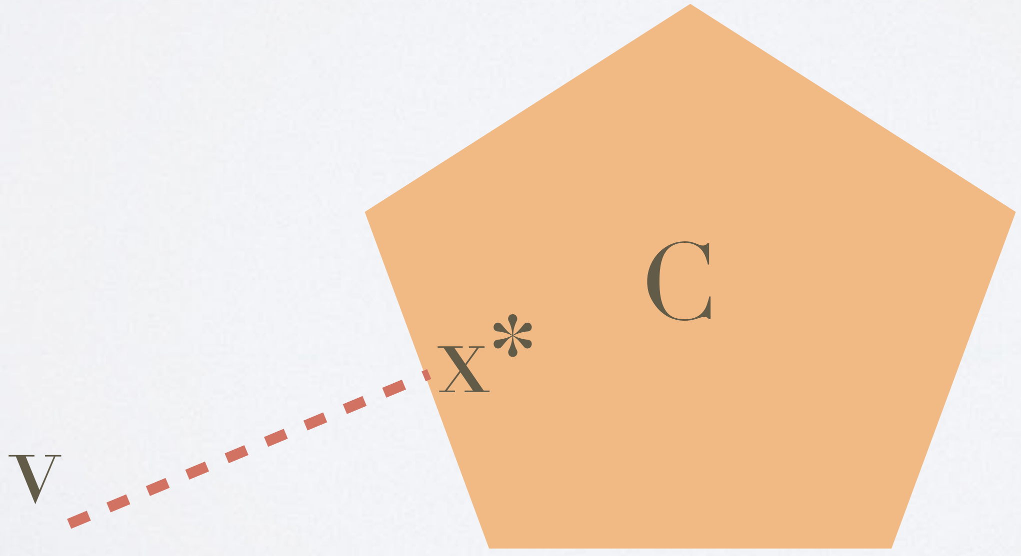
PROXIMAL MAPPING

Moreau, 1965:

$$\mathrm{prox}_f(v) = \mathrm{argmin}_x\left(f(x) + \frac{1}{2}\|x - v\|_2^2\right)$$

Proximal operator generalizes projection:

$$f(x) = I_{\mathcal{C}}(x) : \mathrm{prox}_{I_{\mathcal{C}}}(v) = \Pi_{\mathcal{C}}(v) = \mathrm{argmin}_{x \in \mathcal{C}}\|x - v\|_2$$



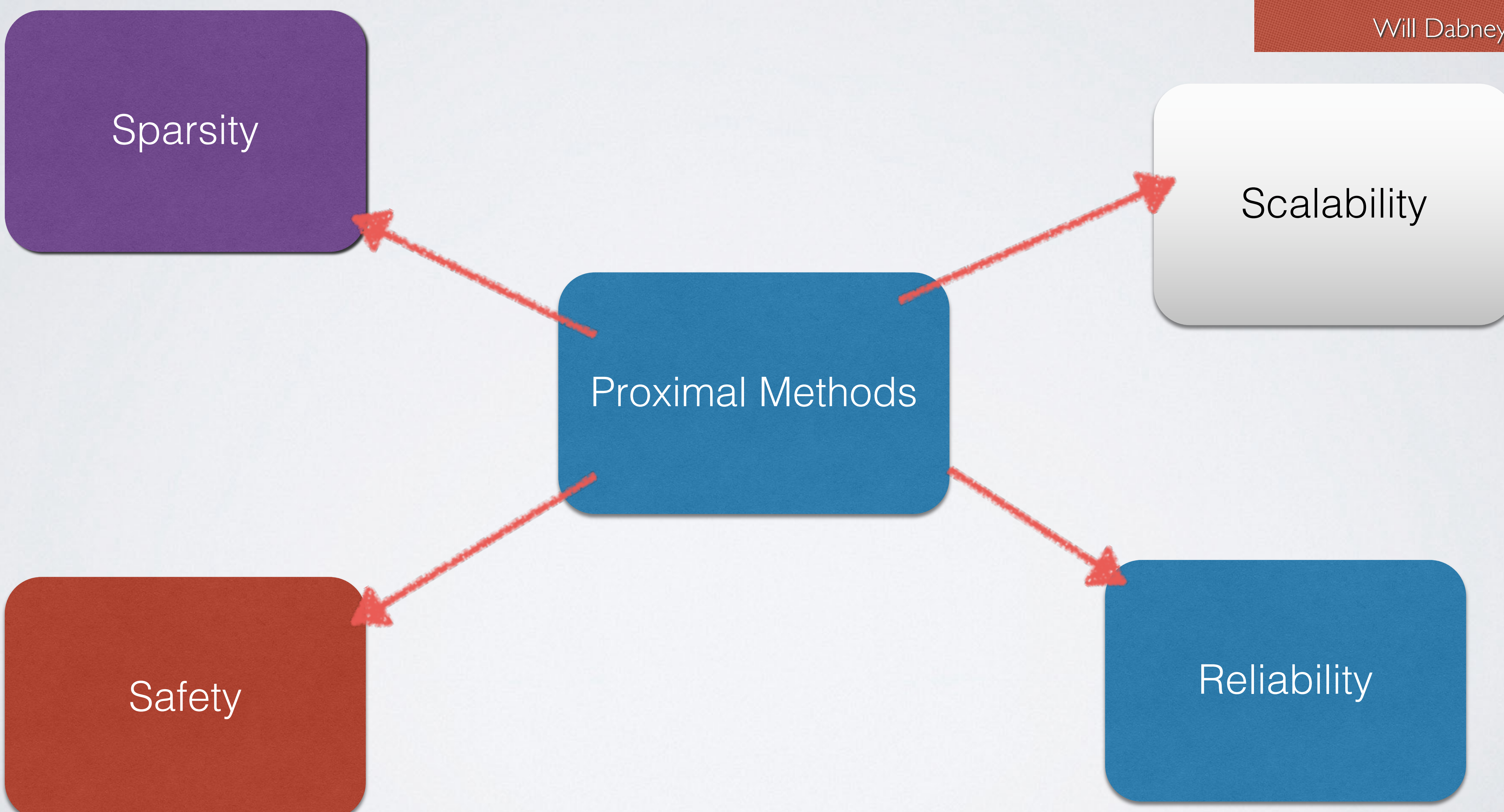
PROXIMAL ABSTRACTIONS

	Property	$f(x)$	$\operatorname{prox}_f x$
i	translation	$\varphi(x-z),\ z\in\mathbb{R}^N$	$z+\operatorname{prox}_\varphi(x-z)$
ii	scaling	$\varphi(x/\rho),\ \rho\in\mathbb{R}\smallsetminus\{0\}$	$\rho\operatorname{prox}_{\varphi/\rho^2}(x/\rho)$
iii	reflection	$\varphi(-x)$	$-\operatorname{prox}_\varphi(-x)$
iv	quadratic perturbation	$\varphi(x)+\alpha\|x\|^2/2+u^\top x+\gamma$ $u\in\mathbb{R}^N,\ \alpha\geq 0,\ \gamma\in\mathbb{R}$	$\operatorname{prox}_{\varphi/(\alpha+1)}\big((x-u)/(\alpha+1)\big)$
v	conjugation	$\varphi^*(x)$	$x-\operatorname{prox}_\varphi x$
vi	squared distance	$\frac{1}{2}d_C^2(x)$	$\frac{1}{2}(x+P_Cx)$
vii	Moreau envelope	$\widetilde{\varphi}(x)=\inf_{y\in\mathbb{R}^N}\varphi(y)+\frac{1}{2}\|x-y\|^2$	$\frac{1}{2}\big(x+\operatorname{prox}_{2\varphi}x\big)$
viii	Moreau complement	$\frac{1}{2}\|\cdot\|^2-\widetilde{\varphi}(x)$	$x-\operatorname{prox}_{\varphi/2}(x/2)$
ix	decomposition in an orthonormal basis $(b_k)_{1\leq k\leq N}$	$\sum_{k=1}^N\phi_k(x^\top b_k)$ $\phi_k\in\Gamma_0(\mathbb{R})$	$\sum_{k=1}^N\operatorname{prox}_{\phi_k}(x^\top b_k)b_k$
x	semi-orthogonal linear transform	$\varphi(Lx)$ $L\in\mathbb{R}^{M\times N},\ LL^\top=\nu I,\ \nu>0$	$x+\nu^{-1}L^\top\big(\operatorname{prox}_{\nu\varphi}(Lx)-Lx\big)$
xi	quadratic function	$\gamma\|Lx-y\|^2/2$ $L\in\mathbb{R}^{M\times N},\ \gamma>0,\ y\in\mathbb{R}^M$	$(I+\gamma L^\top L)^{-1}(x+\gamma L^\top y)$
xii	indicator function	$\iota_C(x)=\begin{cases}0 & \text{if } x\in C\\ +\infty & \text{otherwise}\end{cases}$	P_Cx
xiii	distance function	$\gamma d_C(x),\ \gamma>0$	$\begin{cases}x+\gamma(P_Cx-x)/d_C(x) & \text{if } d_C(x)>\gamma\\ P_Cx & \text{otherwise}\end{cases}$
xv	function of distance	$\phi(d_C(x))$ $\phi\in\Gamma_0(\mathbb{R})$ even, differentiable at 0 with $\phi'(0)=0$	$\begin{cases}x+\left(1-\frac{\operatorname{prox}_\phi d_C(x)}{d_C(x)}\right)(P_Cx-x) & \text{if } x\notin C\\ x & \text{otherwise}\end{cases}$
xv	support function	$\sigma_C(x)$	$x-P_Cx$
xvii	thresholding	$\sigma_C(x)+\phi(\|x\|)$ $\phi\in\Gamma_0(\mathbb{R})$ even and not constant	$\begin{cases}\frac{\operatorname{prox}_\phi d_C(x)}{d_C(x)}(x-P_Cx) & \text{if } d_C(x)>\max\operatorname{Argmin}\phi\\ x-P_Cx & \text{otherwise}\end{cases}$

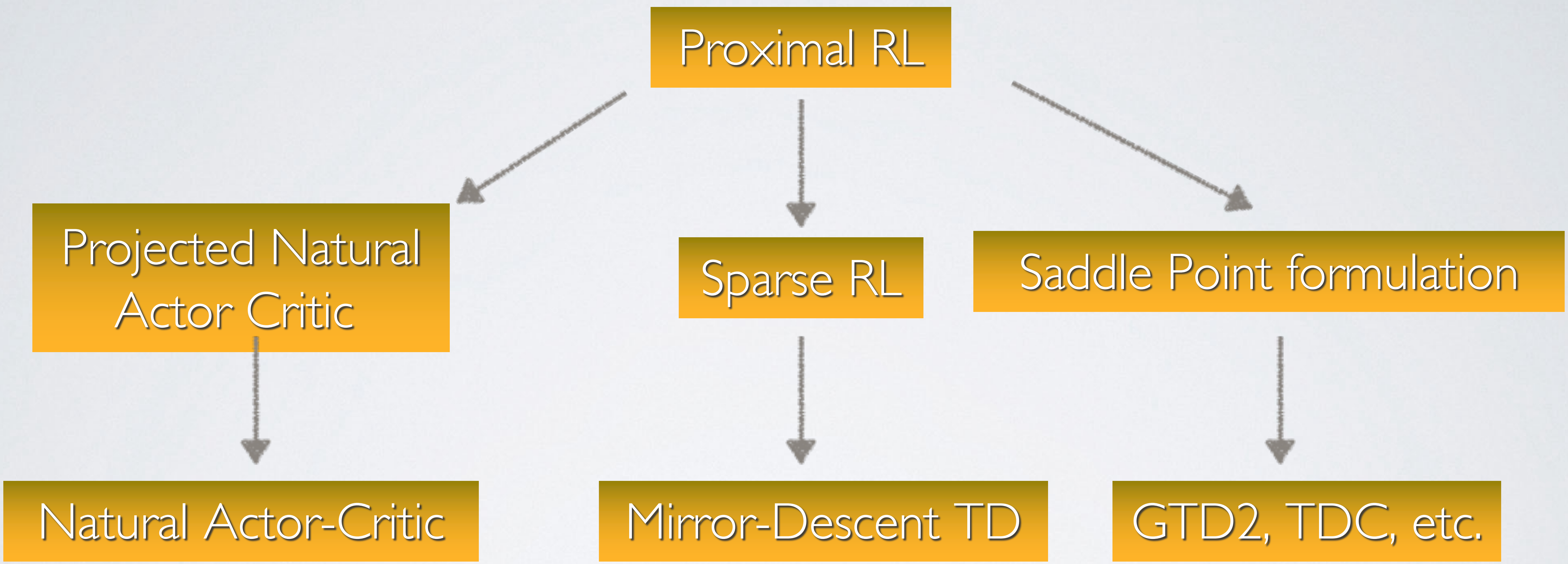
PROXIMAL PLANNING IN PRIMAL-DUAL SPACES

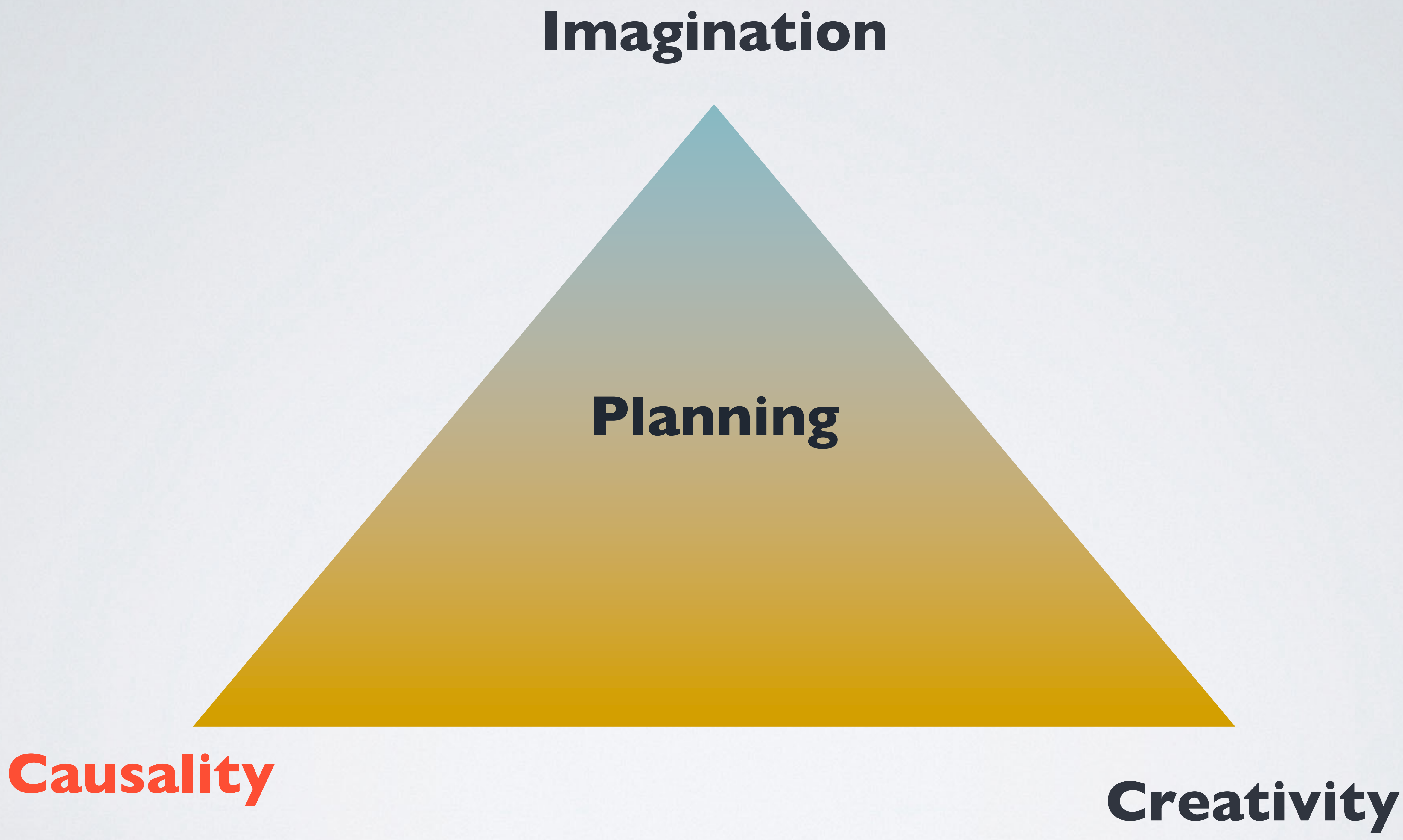
(Mahadevan et al., Arxiv 2015)

Bo Liu
Ian Gemp
Philip Thomas
Stephen Giguere
Ji Liu
Nicholas Jacek
Will Dabney



PROXIMAL FRAMEWORK COVERS MANY RL ALGORITHMS



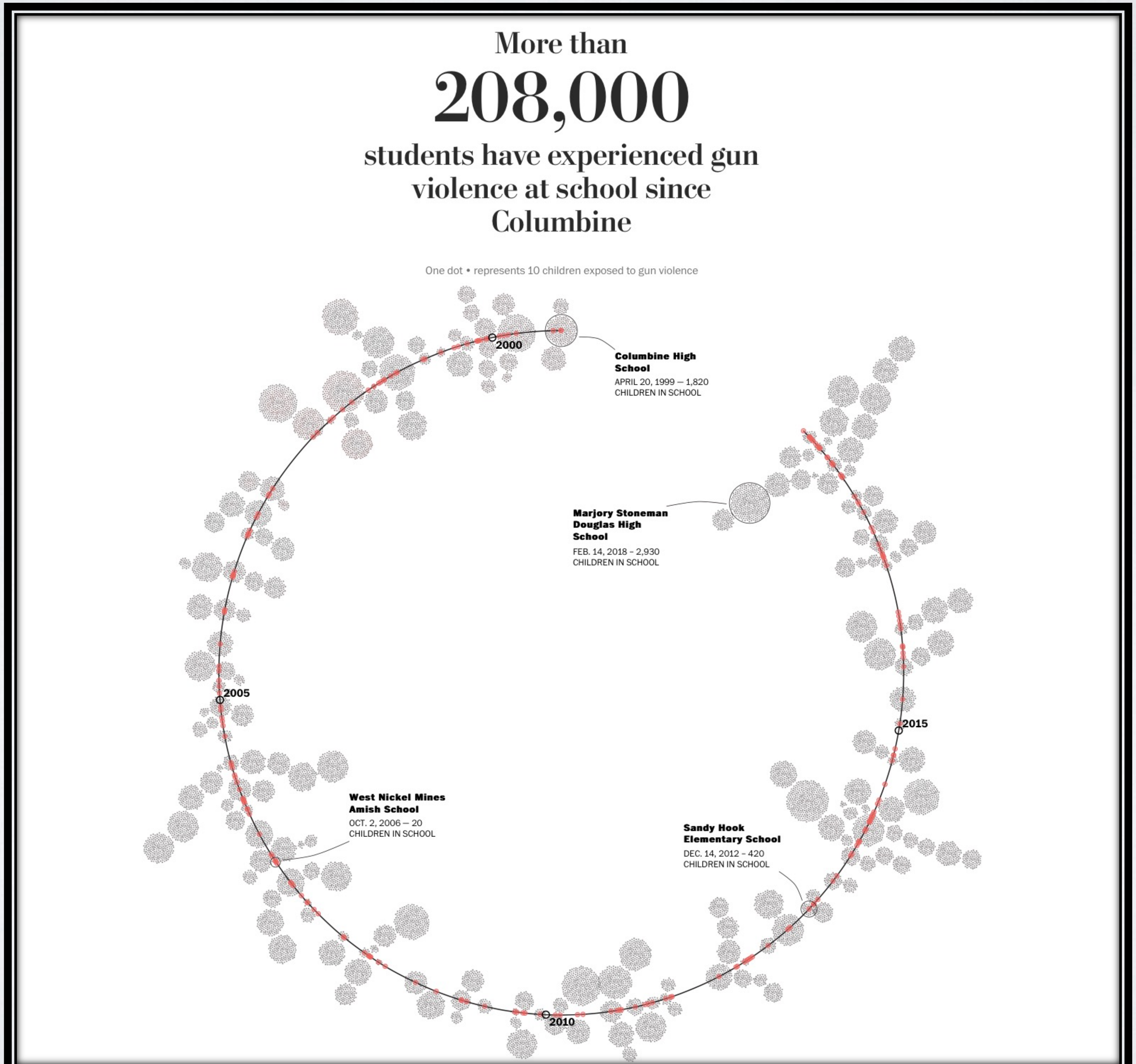

Imagination
Planning
Causality
Creativity

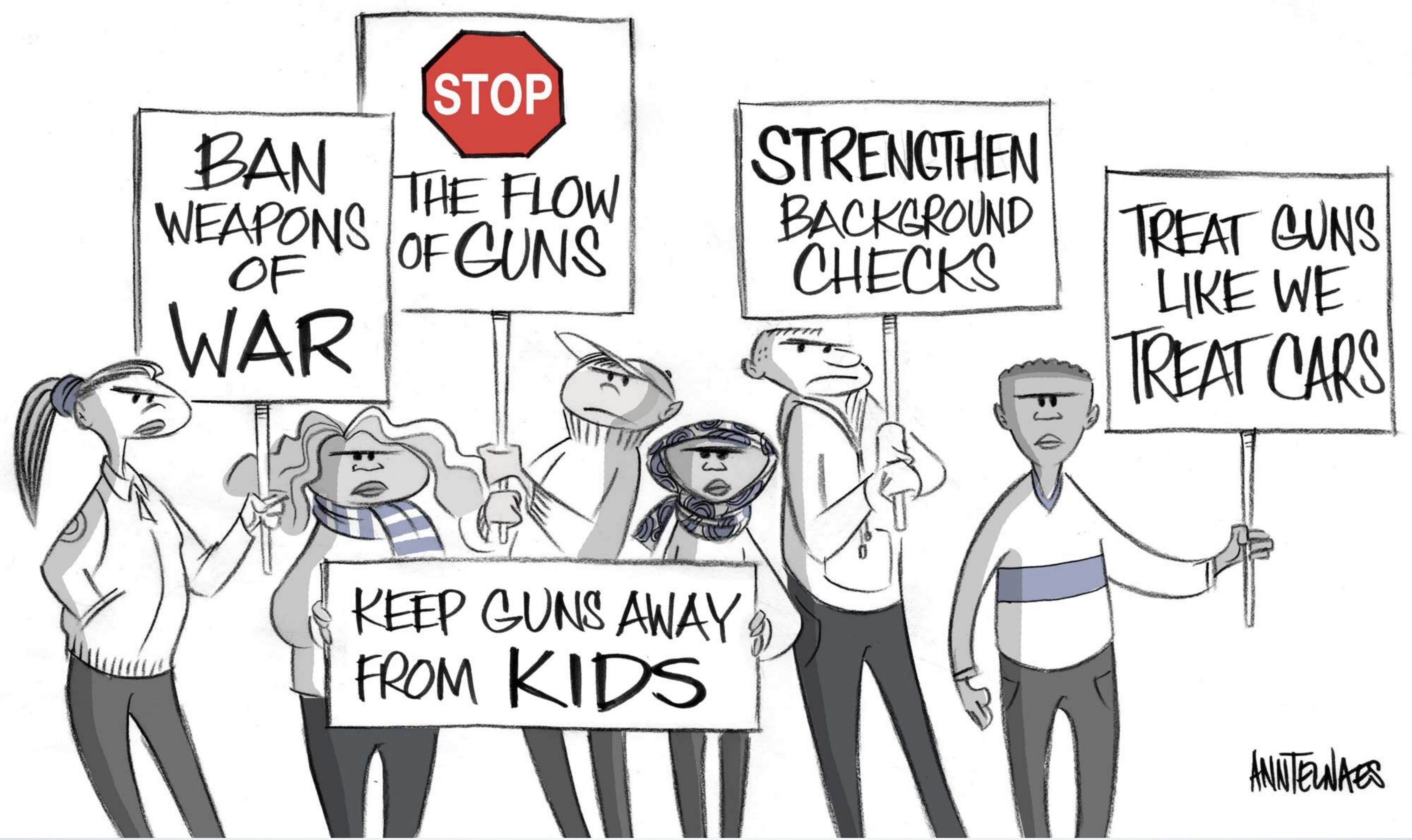
CAUSAL INTERVENTION

New Delhi, India

WASHINGTON POST DATABASE ON GUNS IN SCHOOLS

April 20 2018




BAN WEAPONS OF WAR
STOP
THE FLOW OF GUNS
STRENGTHEN BACKGROUND CHECKS
TREAT GUNS LIKE WE TREAT CARS
KEEP GUNS AWAY FROM KIDS
ANNTELNAES

CAUSALITY VS PLANNING

- The aim of causal inference is to answer the “Why?” question: to understand the world, not only for planning, but also for explanation
- Aggregate model: decisions are often irrevocable
  - Will remdesivir improve the chance of survival in a Covid-19 patient?
- Outcomes are only partially observed
  - Control units cannot be observed under treatment
  - Treated units cannot be observed under control

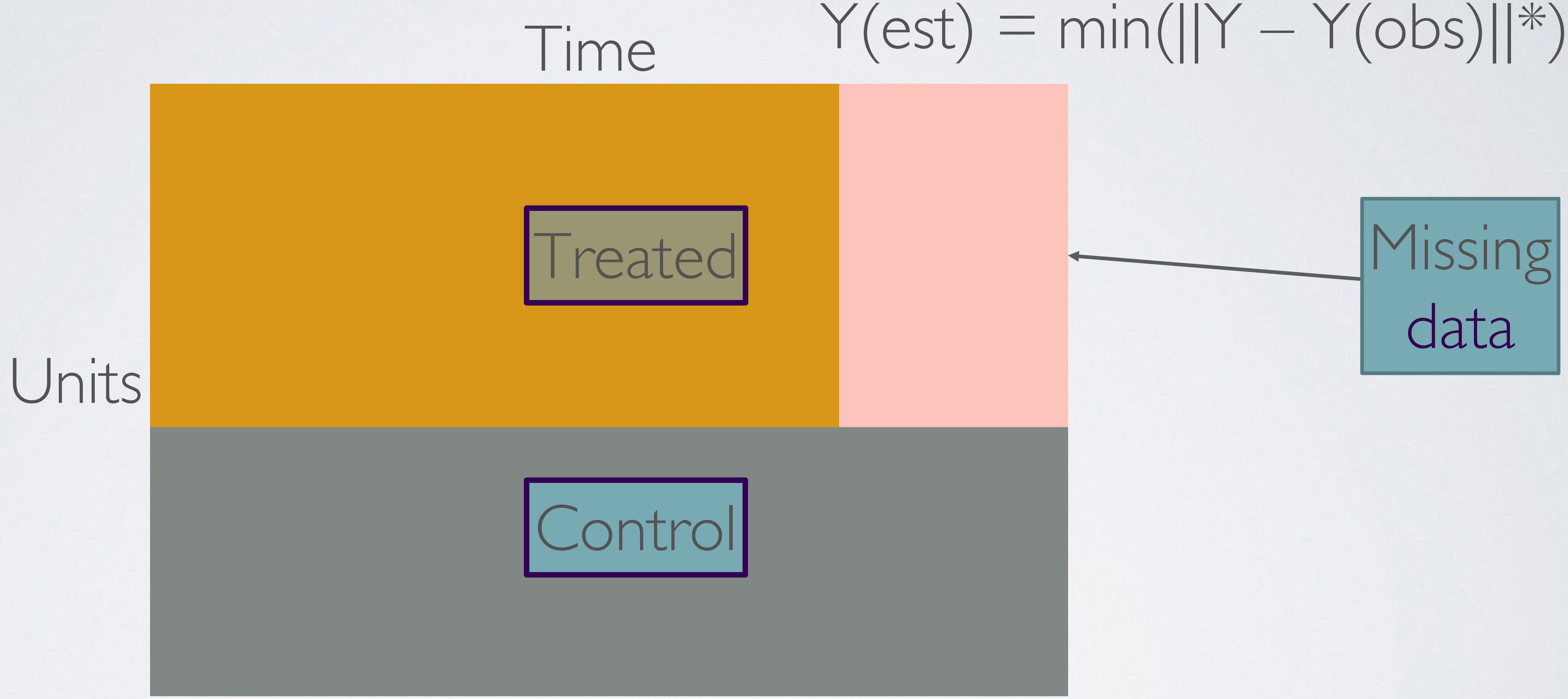
POTENTIAL OUTCOMES

- Rubin (1974) introduced the potential outcomes model of causal inference to reason about counterfactuals
- If we classify countries based on their response to Covid-19, we can do causal analysis of potential outcomes
- Treatment units: countries that mandated masks
- Control units: countries where masks were not mandated
- Stable unit treatment value assumption (SUTVA)

CAUSAL INFERENCE AS MATRIX COMPLETION

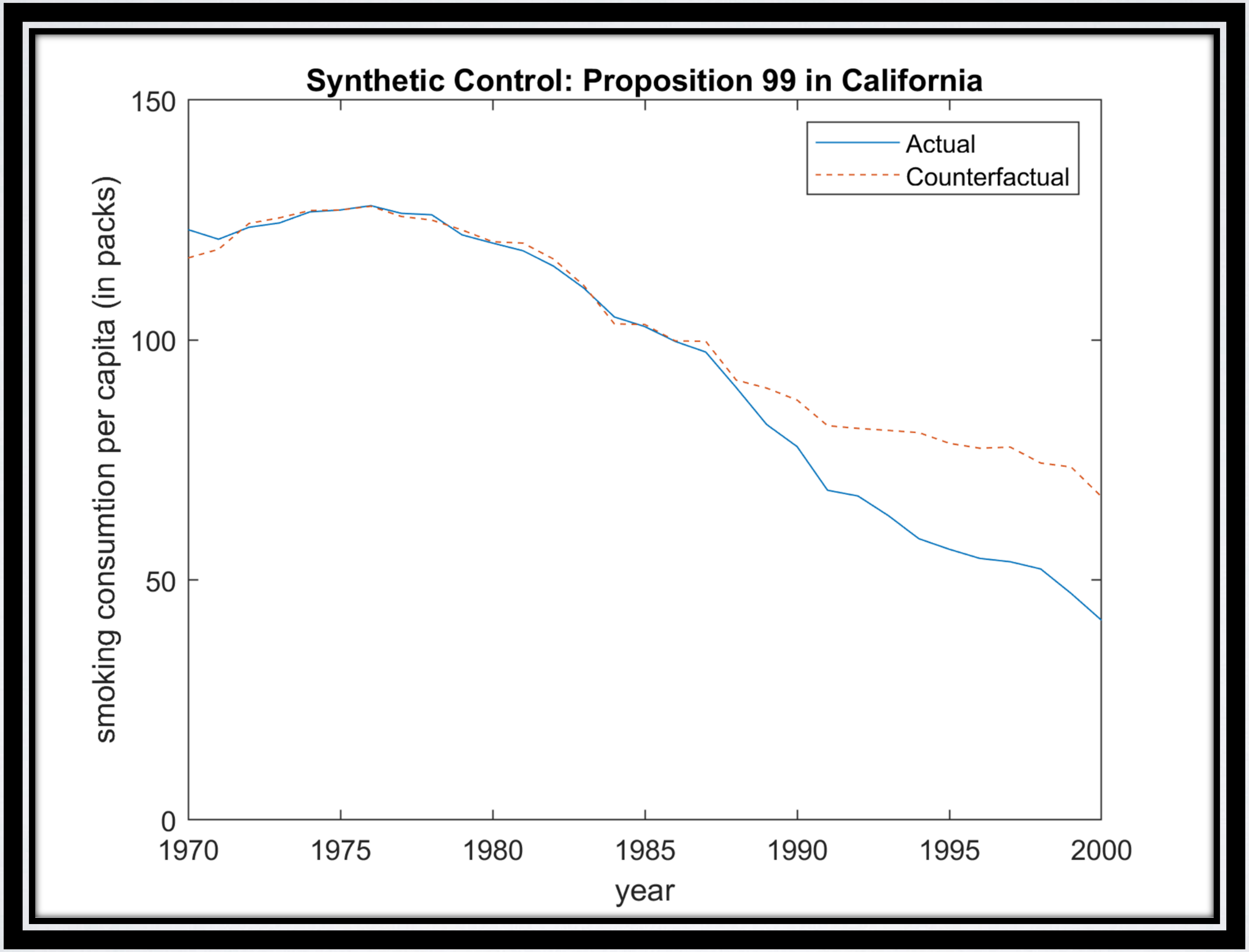
Nuclear norm regularization:

Y(est) = min(||Y – Y(obs)||*)

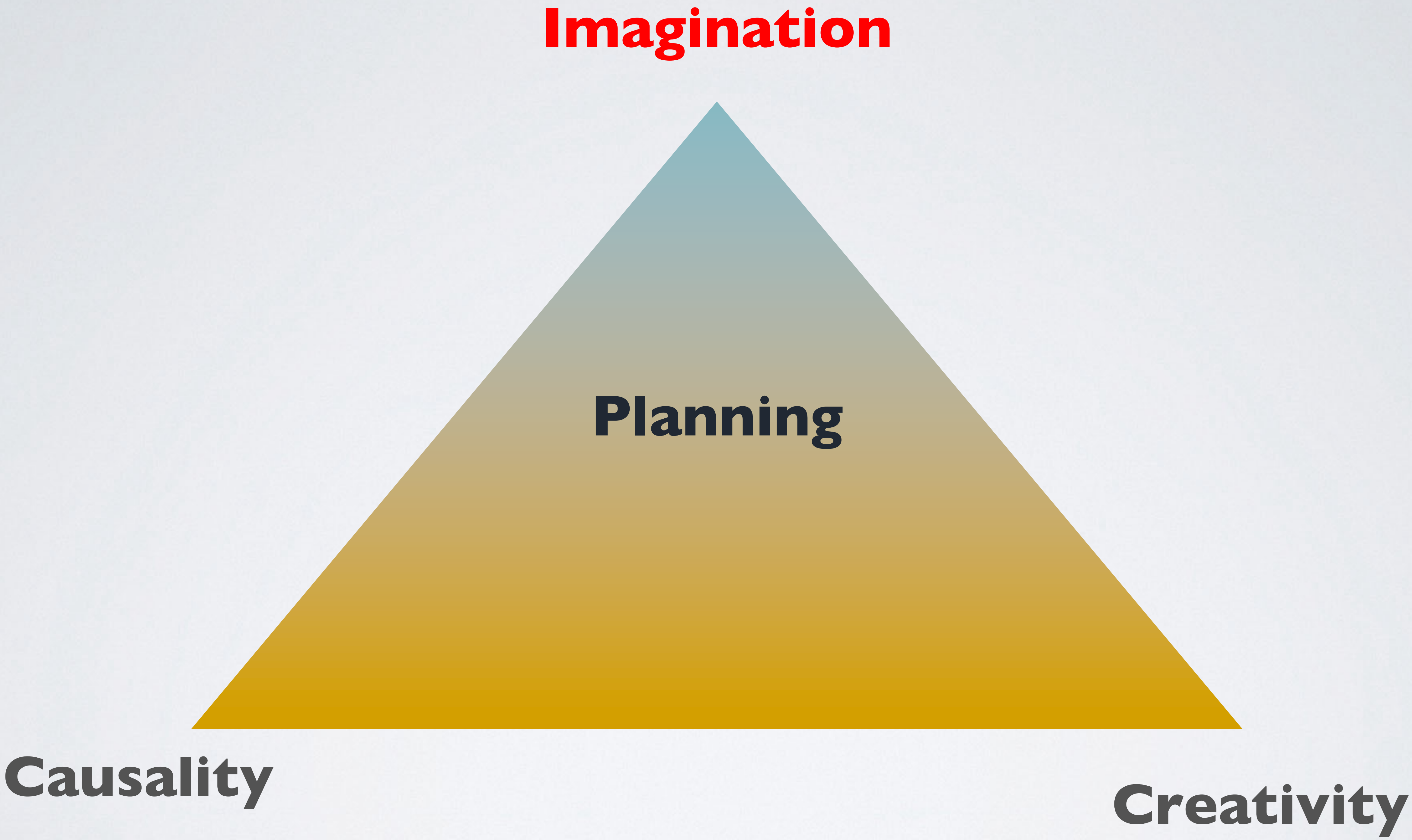


(Athey et al., 2017)

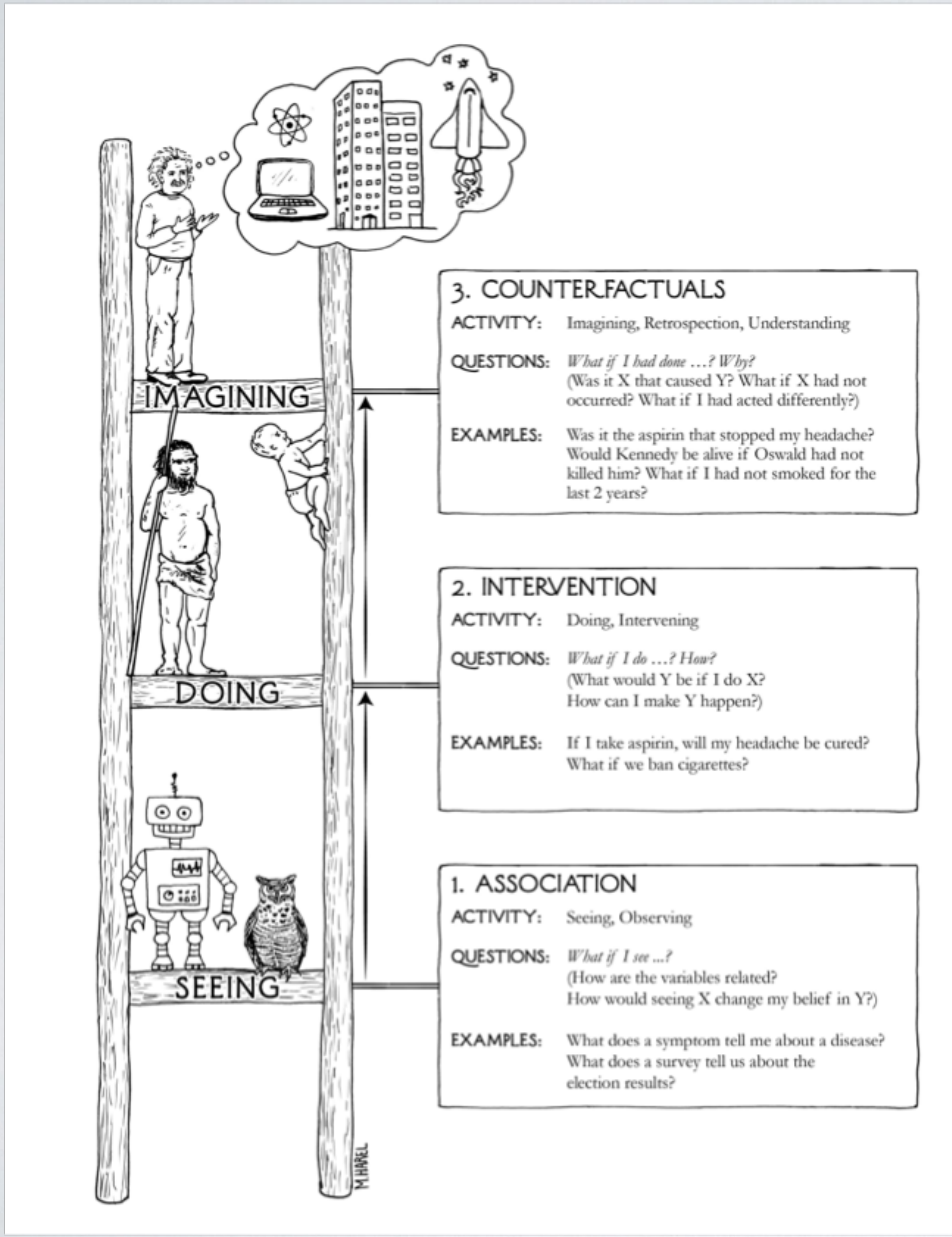
SYNTHETIC CONTROL

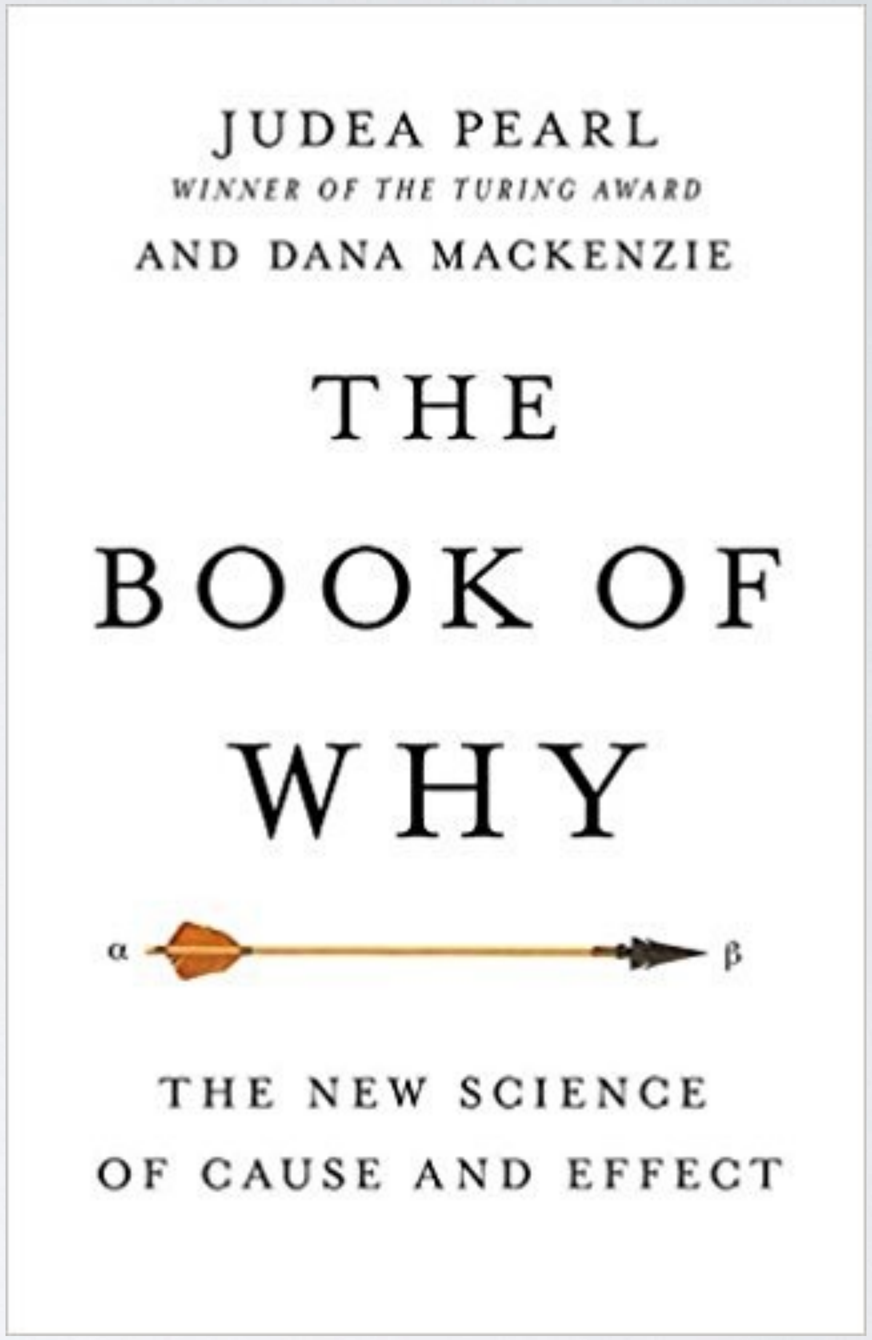


(Abadi)


Imagination
Planning
Causality
Creativity

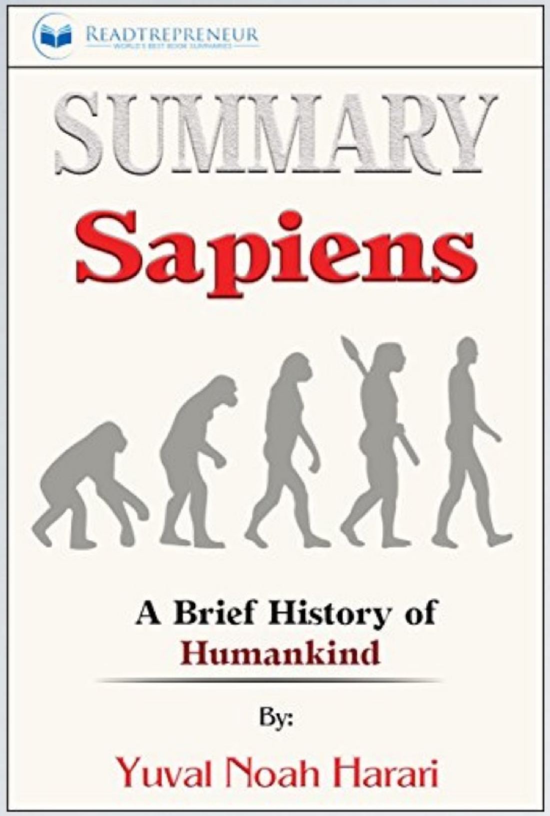
Ladder of Causation

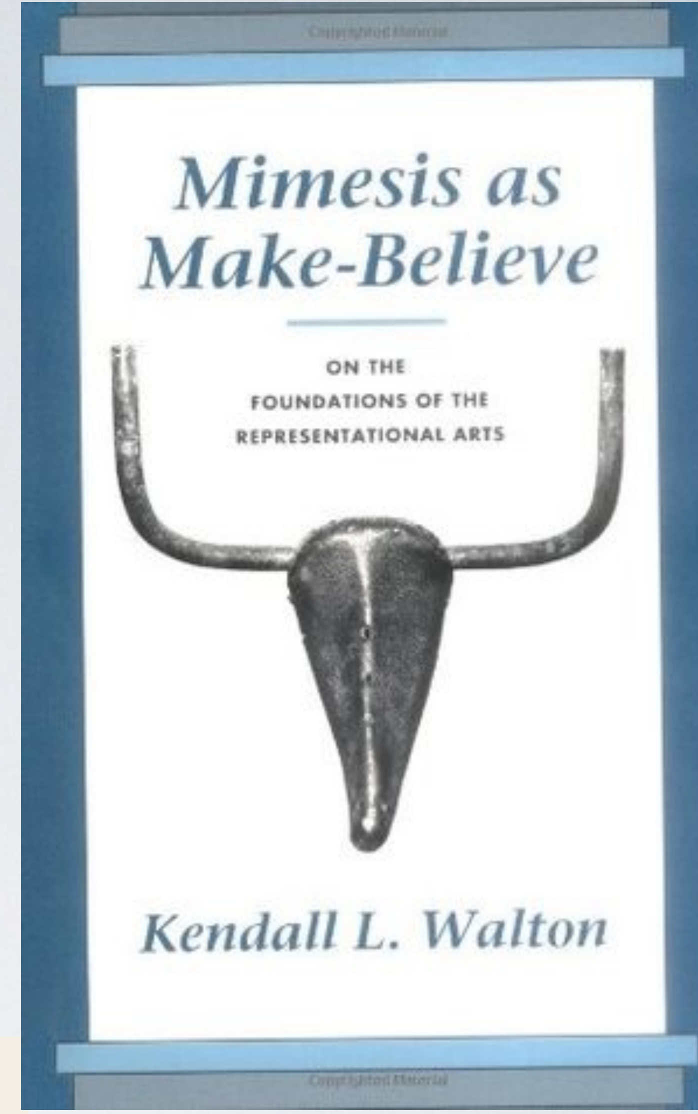


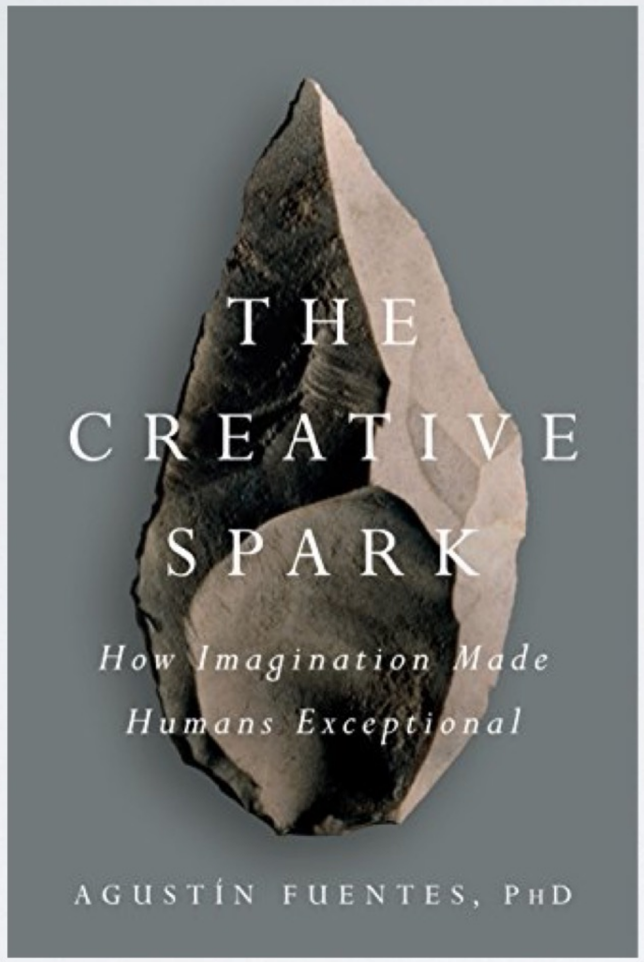


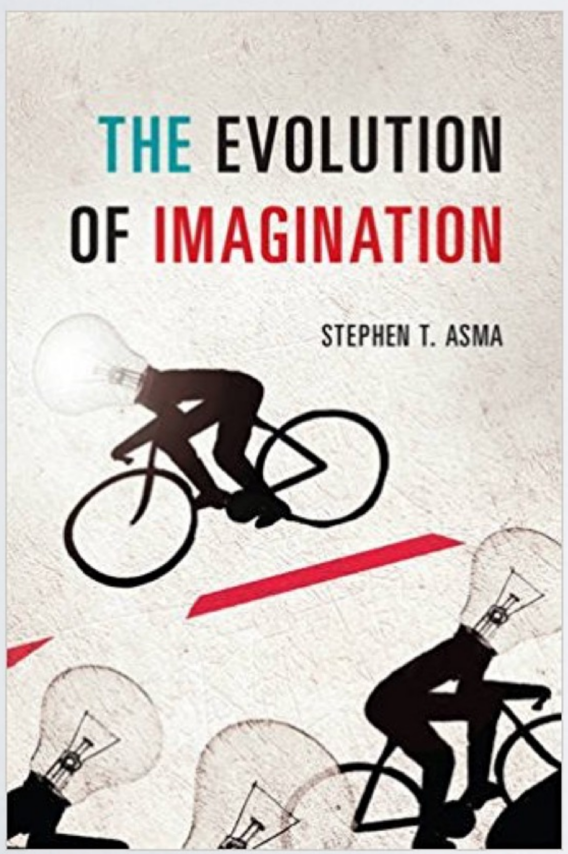
Lion Man at Stadel cave (40,000 years ago)

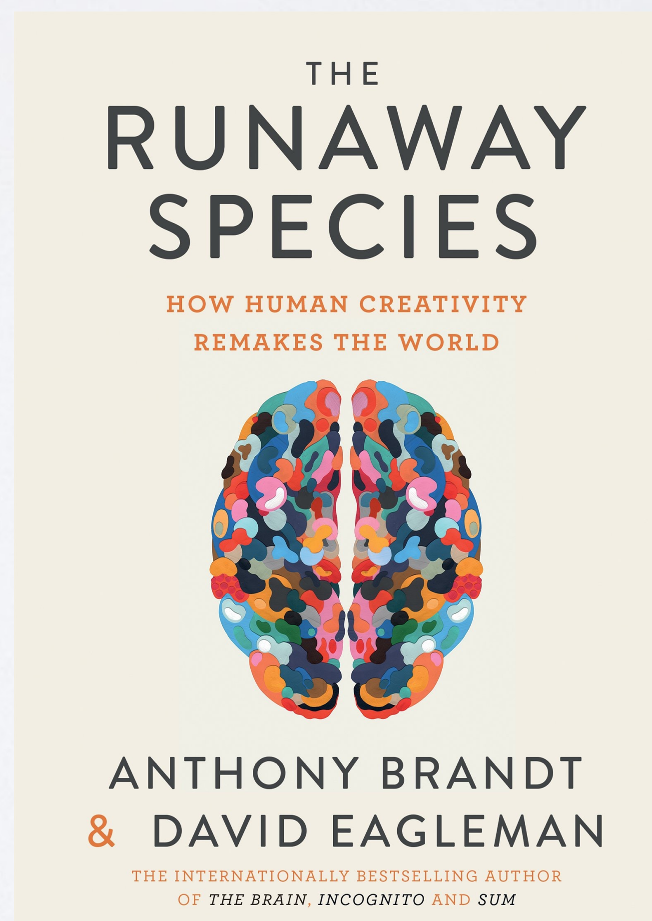
imagining "impossible" objects

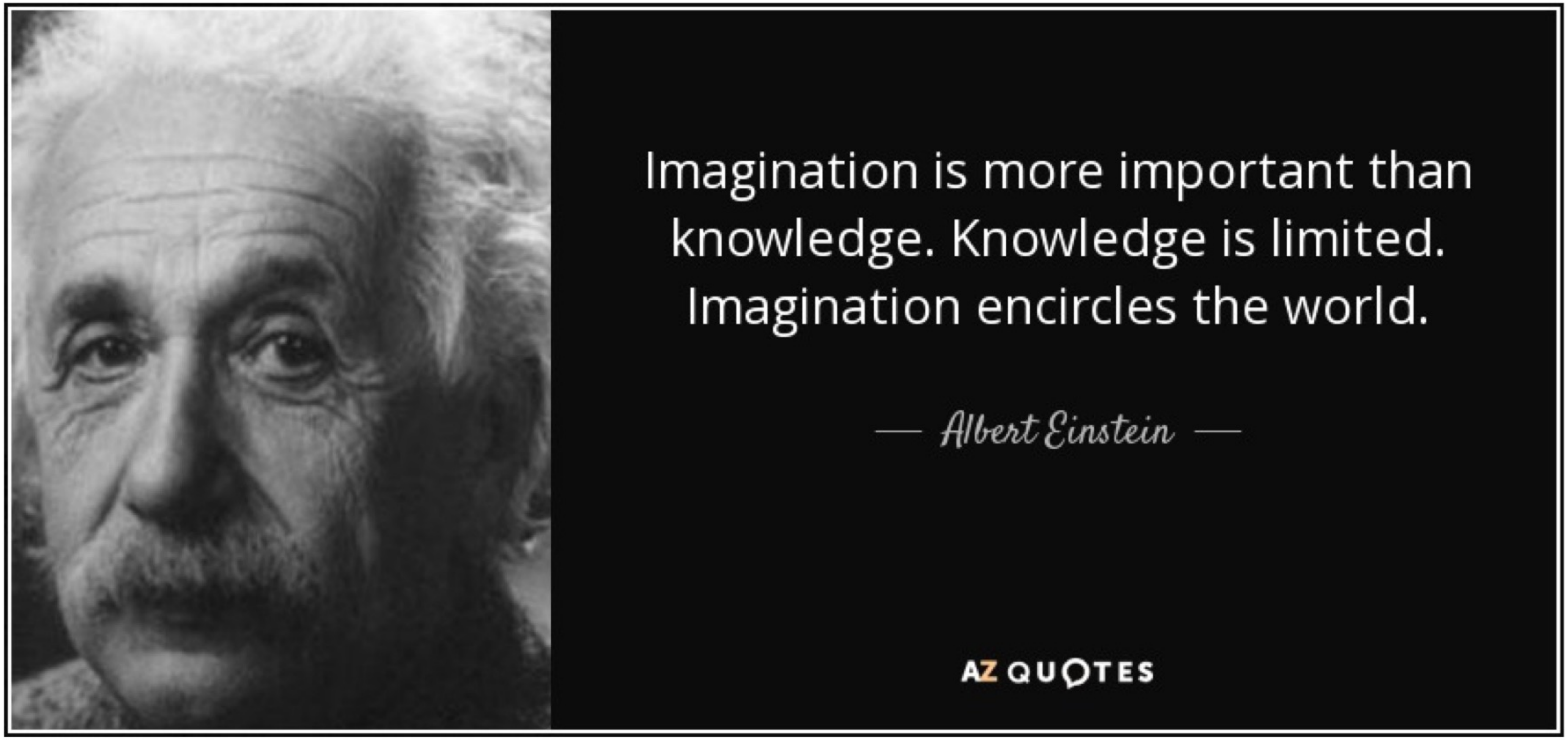

READTREPRENEUR
SUMMARY
Sapiens
A Brief History of Humankind
By:
Yuval Noah Harari


Mimesis as Make-Believe
ON THE FOUNDATIONS OF THE REPRESENTATIONAL ARTS
Kendall L. Walton


THE CREATIVE SPARK
How Imagination Made Humans Exceptional
AGUSTÍN FUENTES, PHD


THE EVOLUTION OF IMAGINATION
STEPHEN T. ASMA

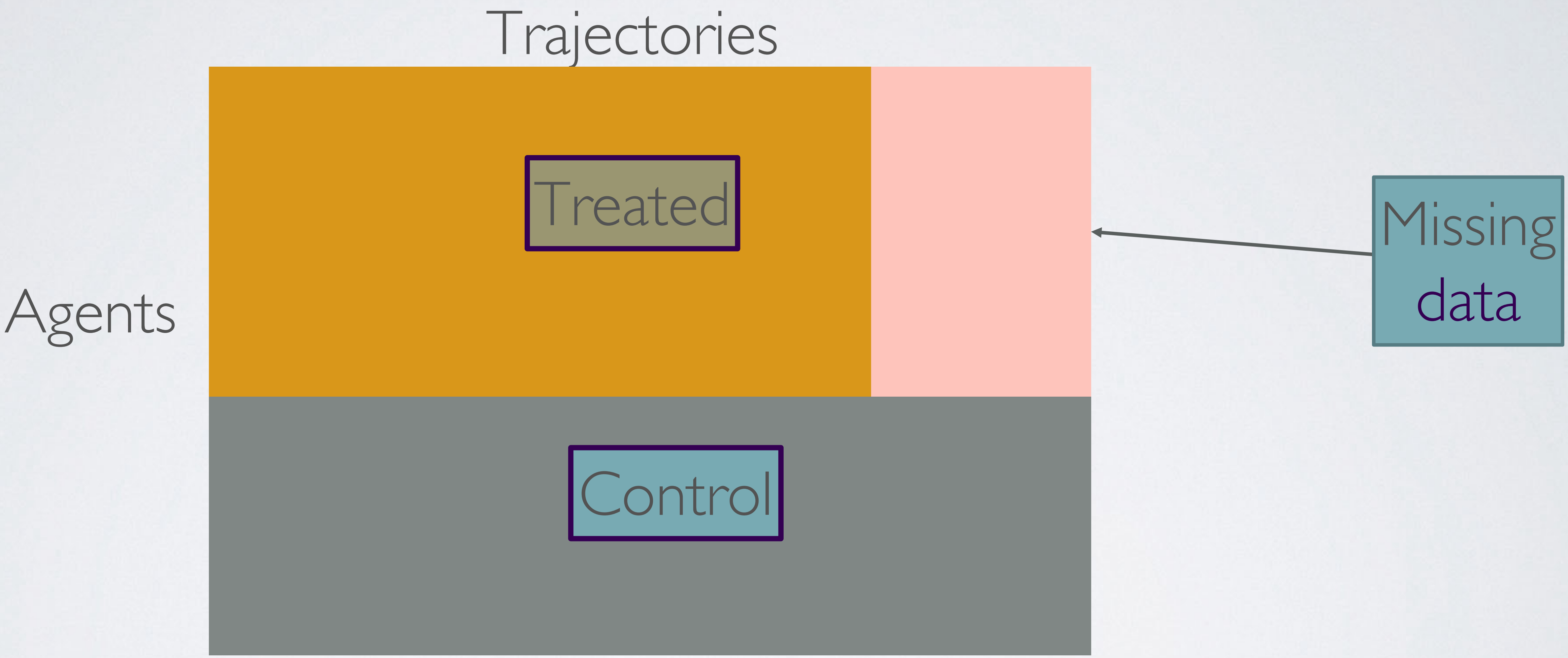

THE RUNAWAY SPECIES
HOW HUMAN CREATIVITY REMAKES THE WORLD
ANTHONY BRANDT & DAVID EAGLEMAN
THE INTERNATIONALLY BESTSELLING AUTHOR OF THE BRAIN, INCOGNITO AND SUM

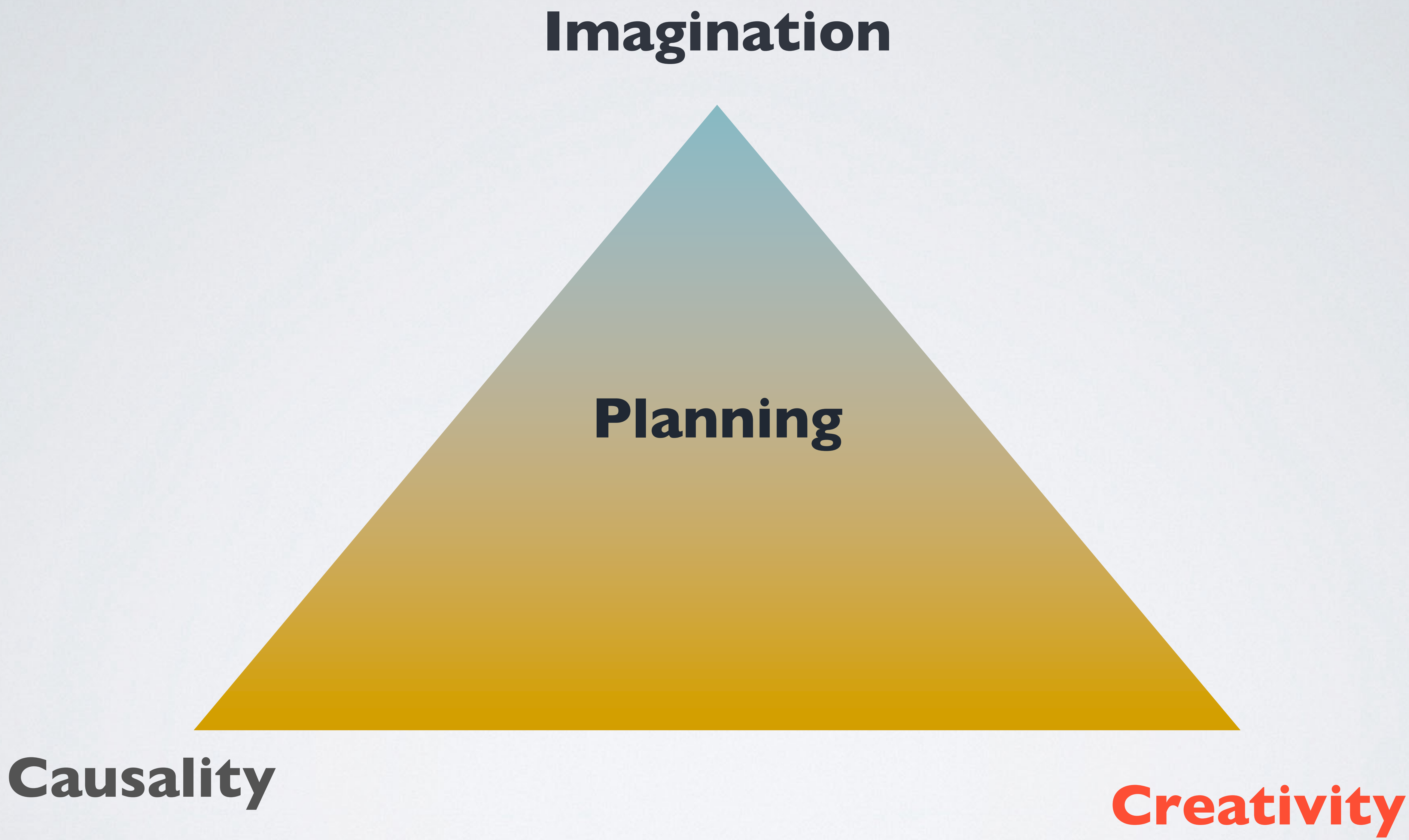

Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world.
— Albert Einstein —
AZ QUOTES

IMAGINATION VALUES

- Q-learning estimates value of actions by trial and error
- How to extend Q-learning to counterfactual learning?
  - How about values of novel actions in novel states?
- Imagination values: infer action values using synthetic control by imputing values from other agents
  - Similar to off-policy RL (Sutton, Liu et al.)

SYNTHETIC Q-LEARNING




Imagination
Planning
Causality
Creativity

CREATIVITY IN ART

Cy Twombly
NY Times, April 11 2018

Jean Michel Basquiat
Sold in NY for $110M

GENERATIVE ADVERSARIAL NETWORKS (Goodfellow et al., 2014)

MIT Technology Review
Feb 2018

Intelligent Machines

The GANfather: The man who's given machines the gift of imagination

By pitting neural networks against one another, Ian Goodfellow has created a powerful AI tool. Now he, and the rest of us, must face the consequences.

by Martin Giles February 21, 2018

11

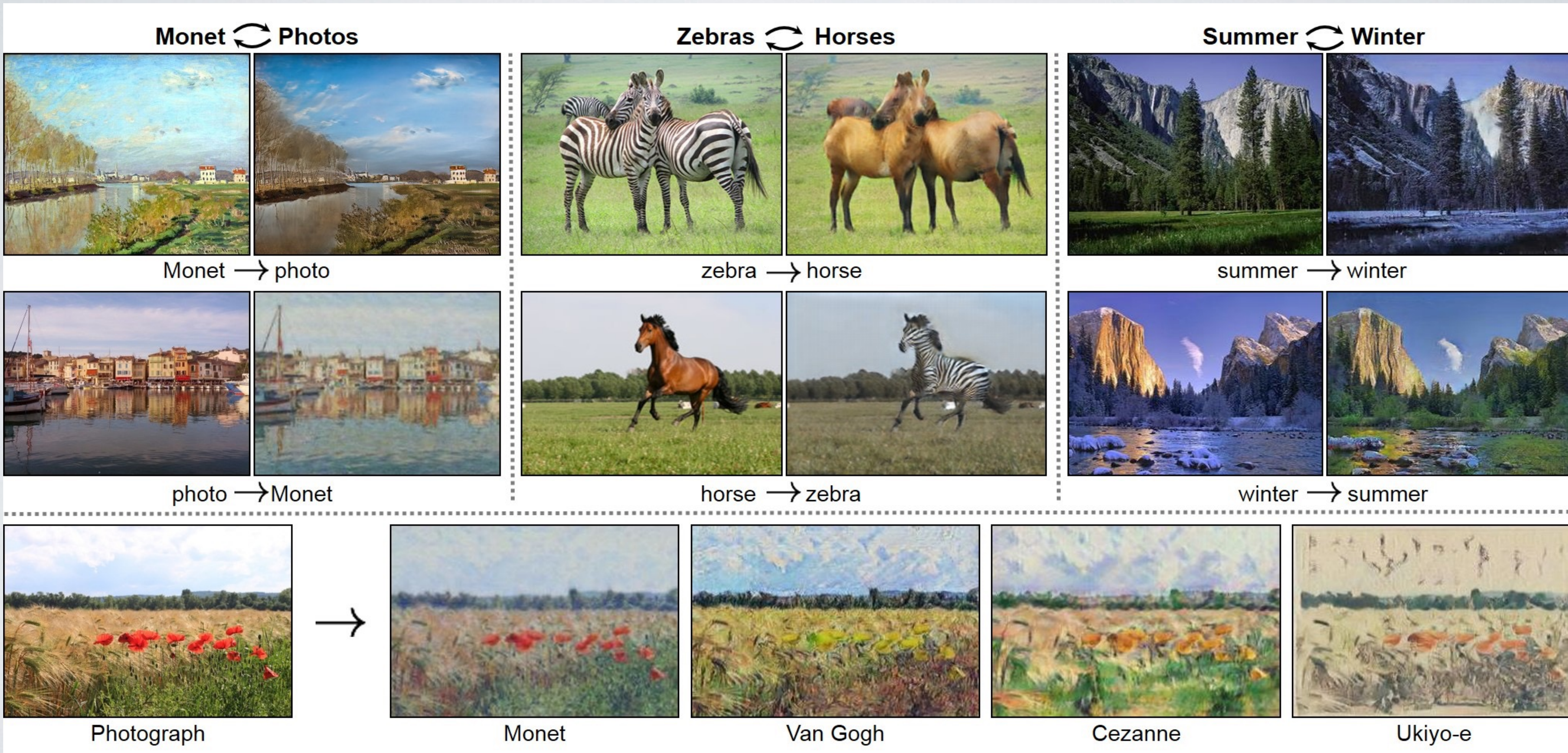
One night in 2014, Ian Goodfellow went drinking to celebrate with a fellow doctoral student who had just graduated. At Les 3 Brasseurs (The Three Brewers), a favorite Montreal watering hole, some friends asked for his help with a thorny project they were working on: a computer that could create photos by itself.

Researchers were already using neural networks, algorithms loosely modeled on the web of neurons in the human brain, as "generative" models to create plausible new data of their own. But the results were often not very good: images of a computer-generated face tended to be blurry or have errors like missing ears. The plan Goodfellow's friends were proposing was to use a complex statistical analysis of the elements that make up a photograph to help machines come up with images by themselves. This would have required a massive amount of number-crunching, and Goodfellow told them it simply wasn't going to work.

But as he pondered the problem over his beer, he hit on an idea. What if you pitted two neural networks against each other? His friends were skeptical, so once he got home, where his girlfriend was already fast asleep, he decided to give it a try. Goodfellow coded into the early hours and then tested his software. It worked the first time.

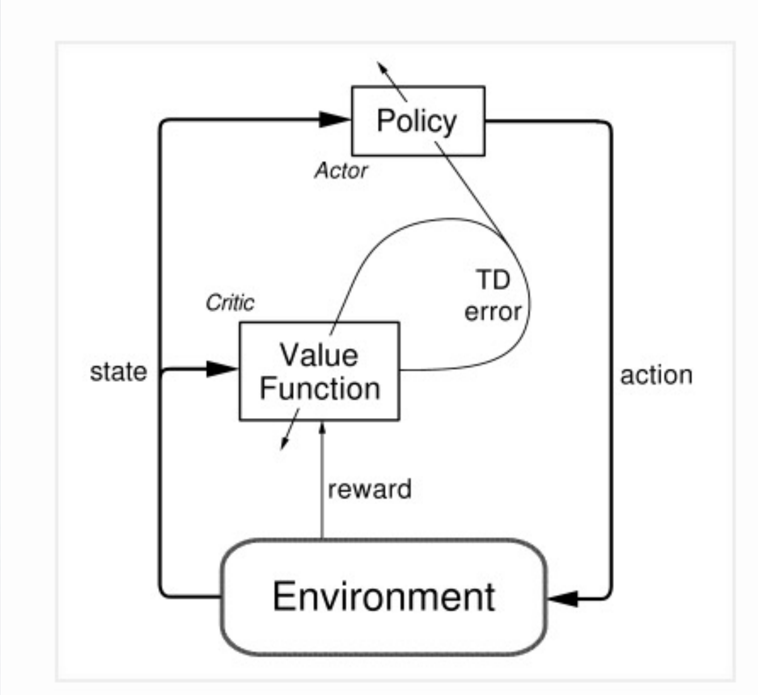
What he invented that night is now called a GAN, or "generative adversarial network." The technique has sparked huge excitement in the field of machine learning and turned its creator into an AI celebrity.

IMAGINATION WITH GANS

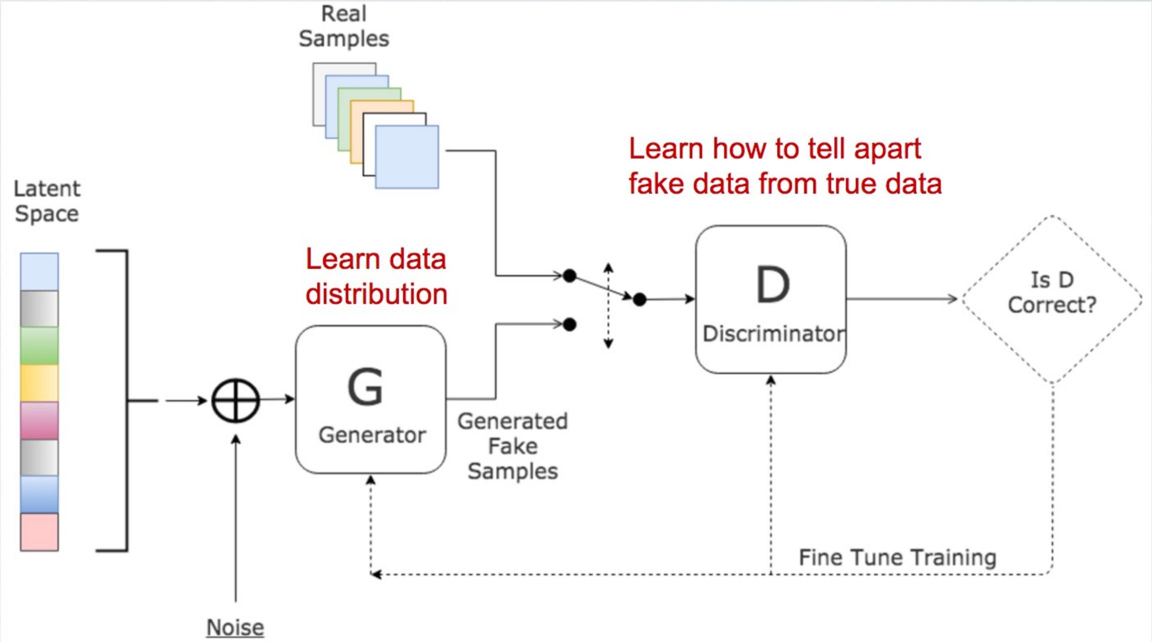


(Zhu et al., 2017)

GAN ~ ACTOR-CRITIC



Actor-Critic (Barto et al., 1983)



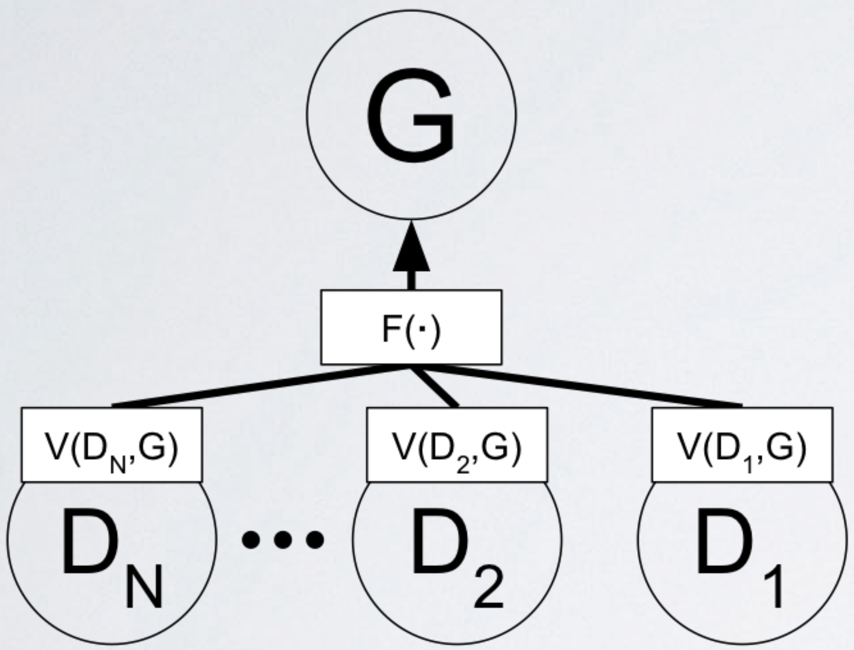
(Goodfellow et al., 2014)

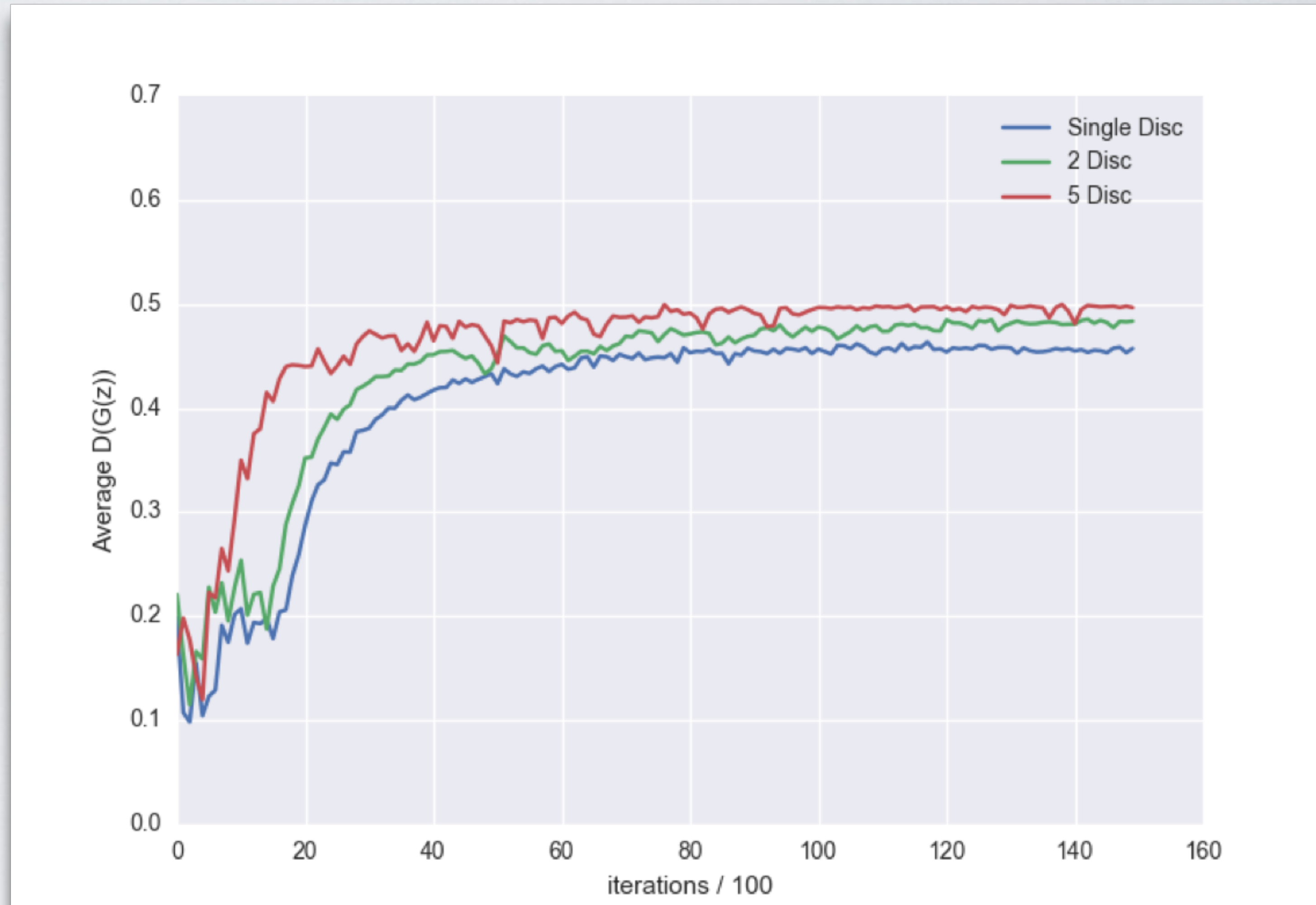
$$
\begin{aligned}
F(D,G) &= -\mathbb{E}_{w\sim p_{\text{data}}}[\log D(w)] - \mathbb{E}_{z\sim\mathcal{N}(0,I)}[\log(1-D(G(z)))] \\
f(D,G) &= -\mathbb{E}_{z\sim\mathcal{N}(0,I)}[\log D(G(z))]
\end{aligned}
$$

$$
\begin{aligned}
F(Q,\pi) &= \mathbb{E}_{s_t,a_t\sim\pi}[\mathcal{D}(\mathbb{E}_{s_{t+1},r_t,a_{t+1}}[r_t+\gamma Q(s_{t+1},a_{t+1})]||Q(s_t,a_t))] \\
f(Q,\pi) &= -\mathbb{E}_{s_0\sim p_0,a_0\sim\pi}[Q^\pi(s_0,a_0)]
\end{aligned}
$$

Pfau and Vinyals, NeurIPS 2017

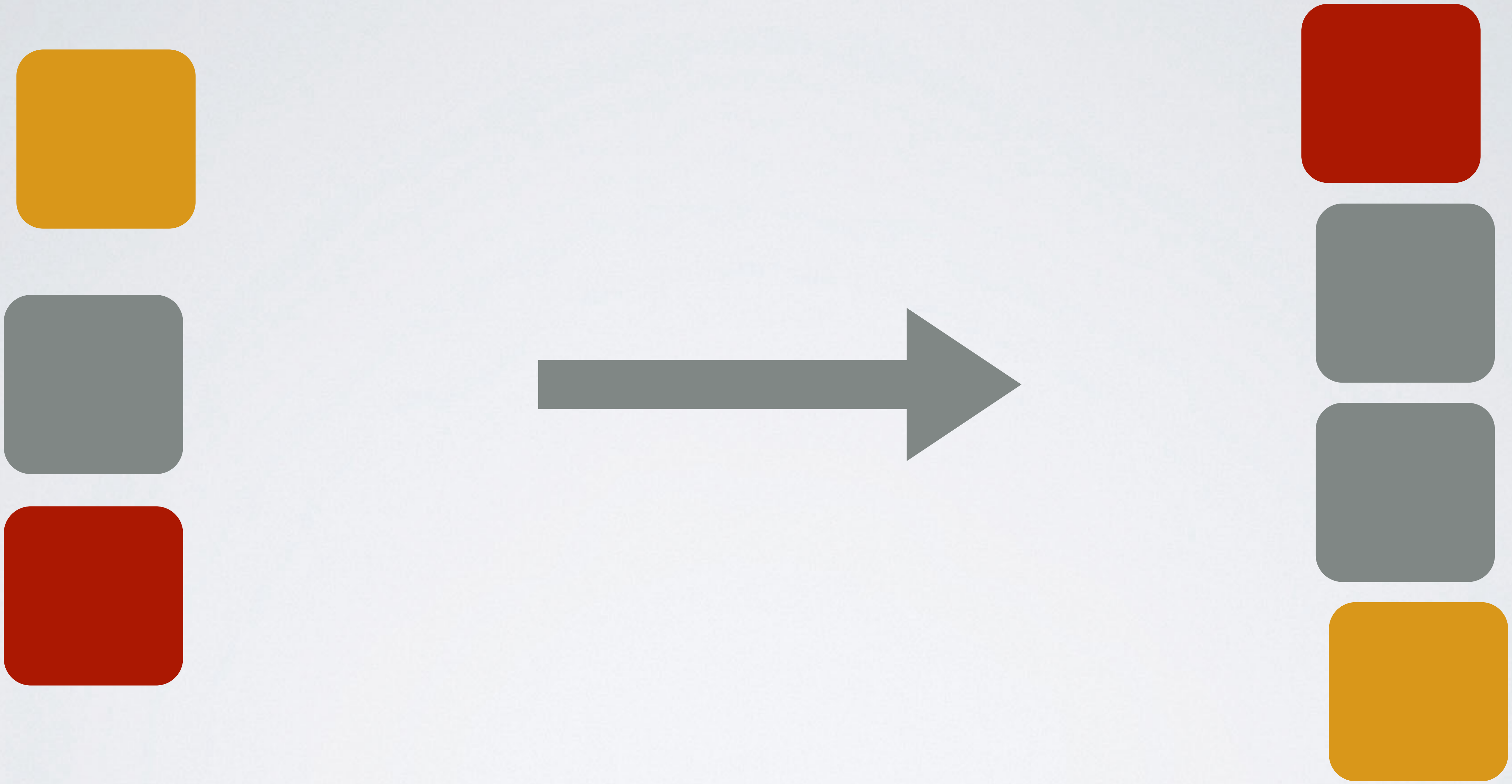
GENERATIVE MULTI-ADVERSARIAL NETWORKS (GMAN)





(Durugkar, Gemp, and Mahadevan, ICLR, 2017)

MIX GAN MODELS



Generators

Discriminators

(Arora et al., ICML 2017)

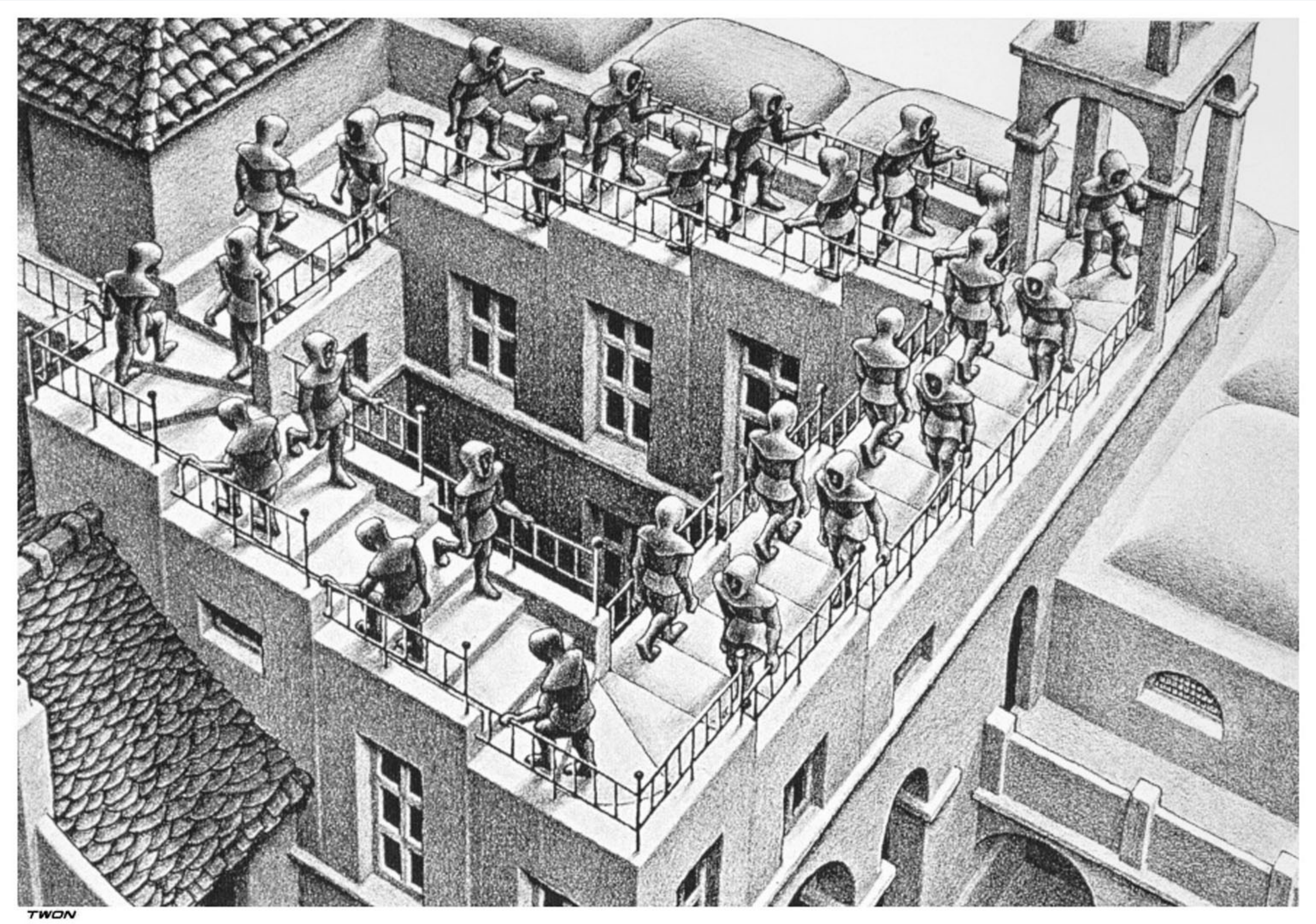
SIMULTANEOUS GRADIENT ASCENT

Algorithm 1 Simultaneous Gradient Ascent (SimGA)

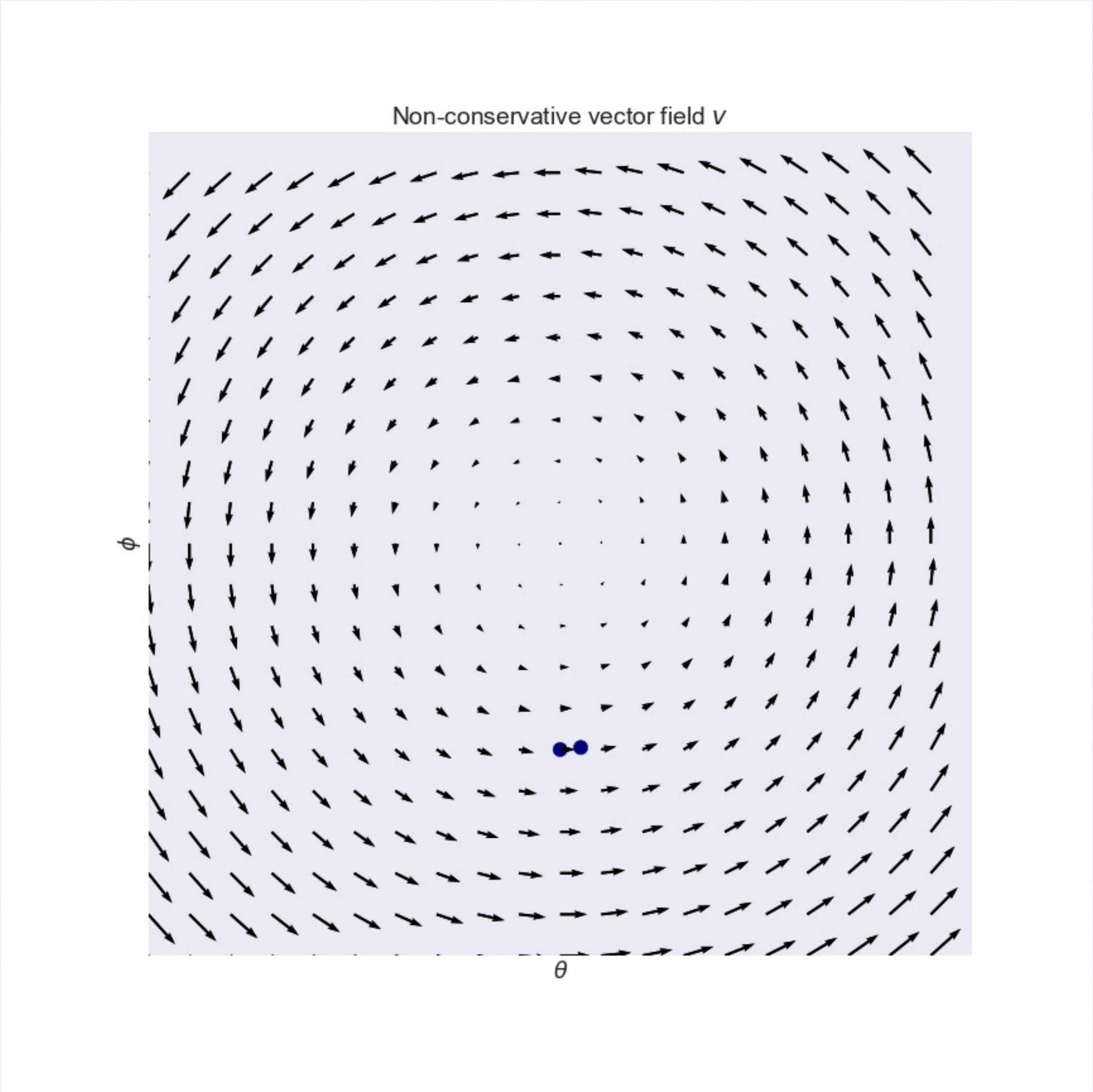
1: **while** not converged **do**
2: $\quad v_\phi \leftarrow \nabla_\phi f(\theta, \phi)$
3: $\quad v_\theta \leftarrow \nabla_\theta g(\theta, \phi)$
4: $\quad \phi \leftarrow \phi + h v_\phi$
5: $\quad \theta \leftarrow \theta + h v_\theta$
6: **end while**

(Mescheder et al., Numerics of GANs, Arxiv, 2017)

M. C. Escher

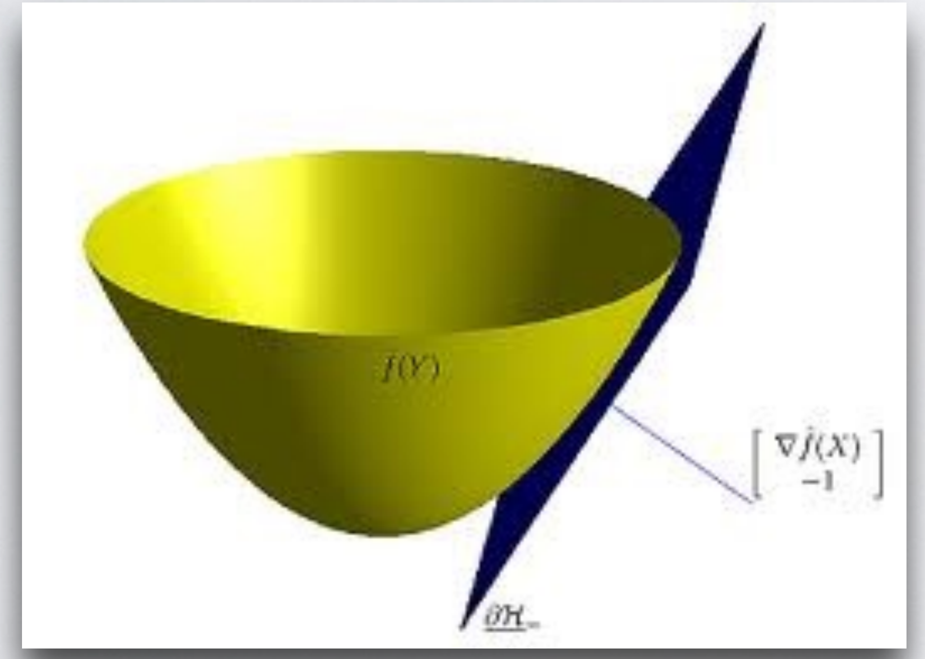


$$f(\theta, \phi) = -g(\theta, \phi) = (-\phi, \theta)$$

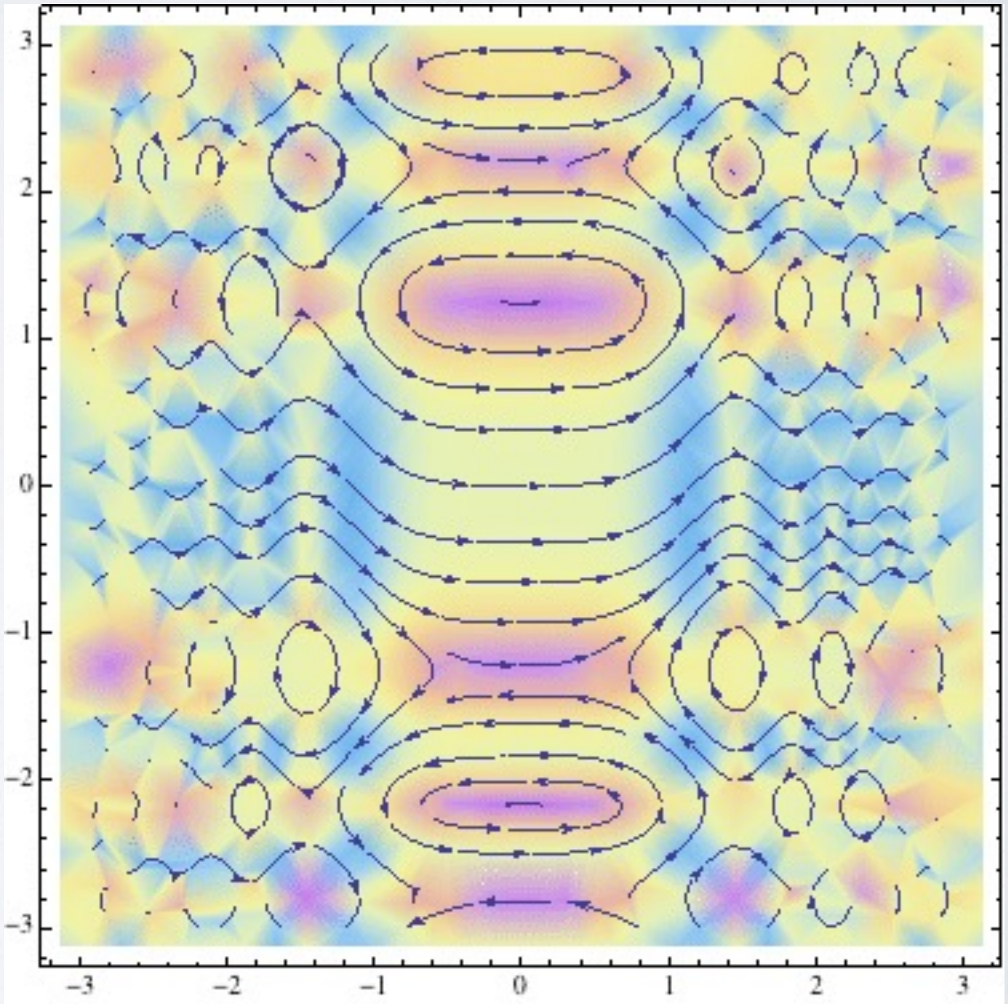


GRADIENTS TO VECTOR FIELDS

$$w_{t+1} \leftarrow w_t - \alpha_t \nabla f(w_t)$$



$$w_{t+1} \leftarrow w_t - \alpha_t F(w_t)$$



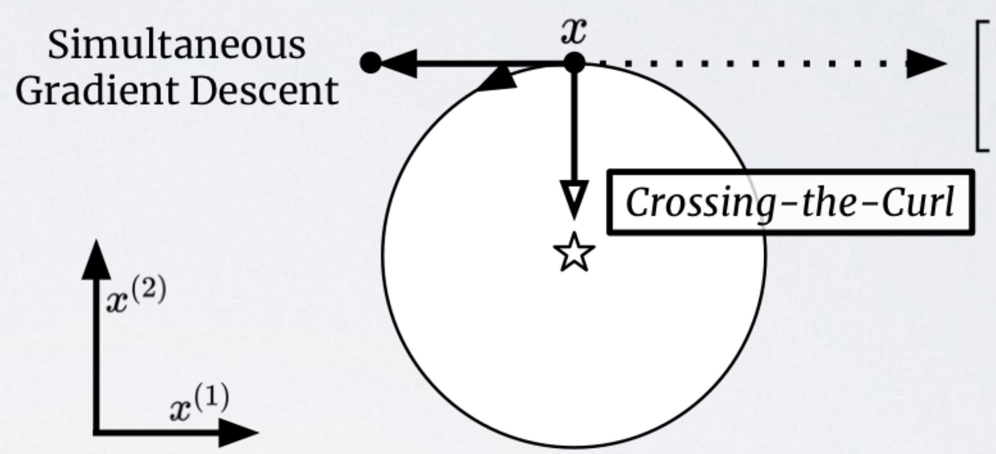
NOVEL GAN ALGORITHMS: VARIATIONAL INEQUALITIES

- Joint work with my PhD student Ian Gemp
- In equilibrium problems, traversing the steepest descent direction is not sufficient for convergence
- **Orthogonal** directions turn out to be crucial for convergence

VECTOR FIELD GAN ALGORITHM

Example:

$$\begin{bmatrix} x^{(1)} \\ x^{(2)} \end{bmatrix} \leftarrow \begin{bmatrix} x^{(1)} - \rho \nabla_{x^{(1)}} f^{(1)}(x) \\ x^{(2)} - \rho \nabla_{x^{(2)}} f^{(2)}(x) \end{bmatrix}$$



$$\begin{bmatrix} \nabla_{x^{(1)}} f^{(1)}(x) \\ \nabla_{x^{(2)}} f^{(2)}(x) \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \end{bmatrix}$$

Player **1**'s loss is: $f^{(1)}(x) = 10x^{(1)}x^{(2)}$.

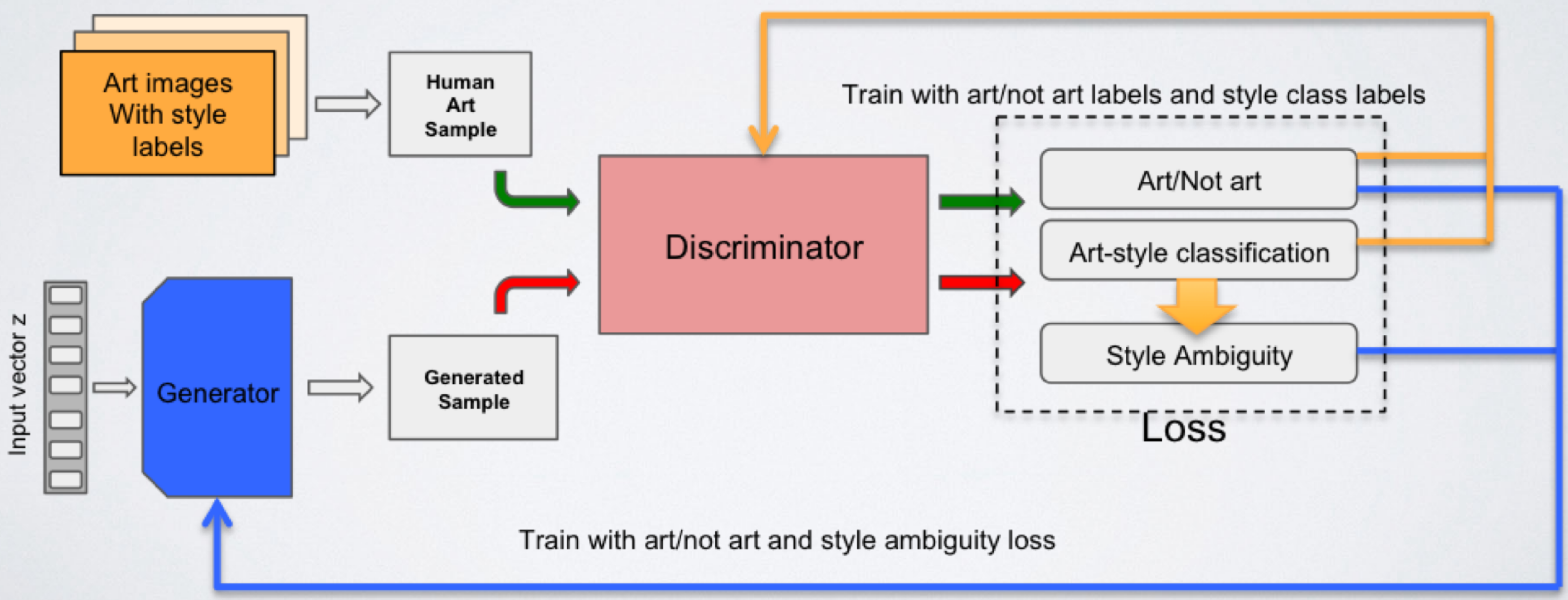
Player **2**'s loss is: $f^{(2)}(x) = -10x^{(1)}x^{(2)}$.

Let $x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\rho = 0.1$.


R. KEITH SAWYER
the
SCIENCE
of
HUMAN INNOVATION
Explaining Creativity
SECOND EDITION
OXFORD

CAN: Creative Adversarial Networks Generating “Art” by Learning About Styles and Deviating from Style Norms*

Ahmed Elgammal[1†] Bingchen Liu[1] Mohamed Elhoseiny[2] Marian Mazzone[3]
The Art & AI Laboratory - Rutgers University
[1] Department of Computer Science, Rutgers University, NJ, USA
[2] Facebook AI Research, CA, USA
[3] Department of Art History, College of Charleston, SC, USA



CAN loss function

$$\min_G \max_D V(D,G) =$$

$$\mathbb{E}_{x,\hat{c}\sim p_{data}}[\log D_r(x) + \log D_c(c=\hat{c}|x)] +$$

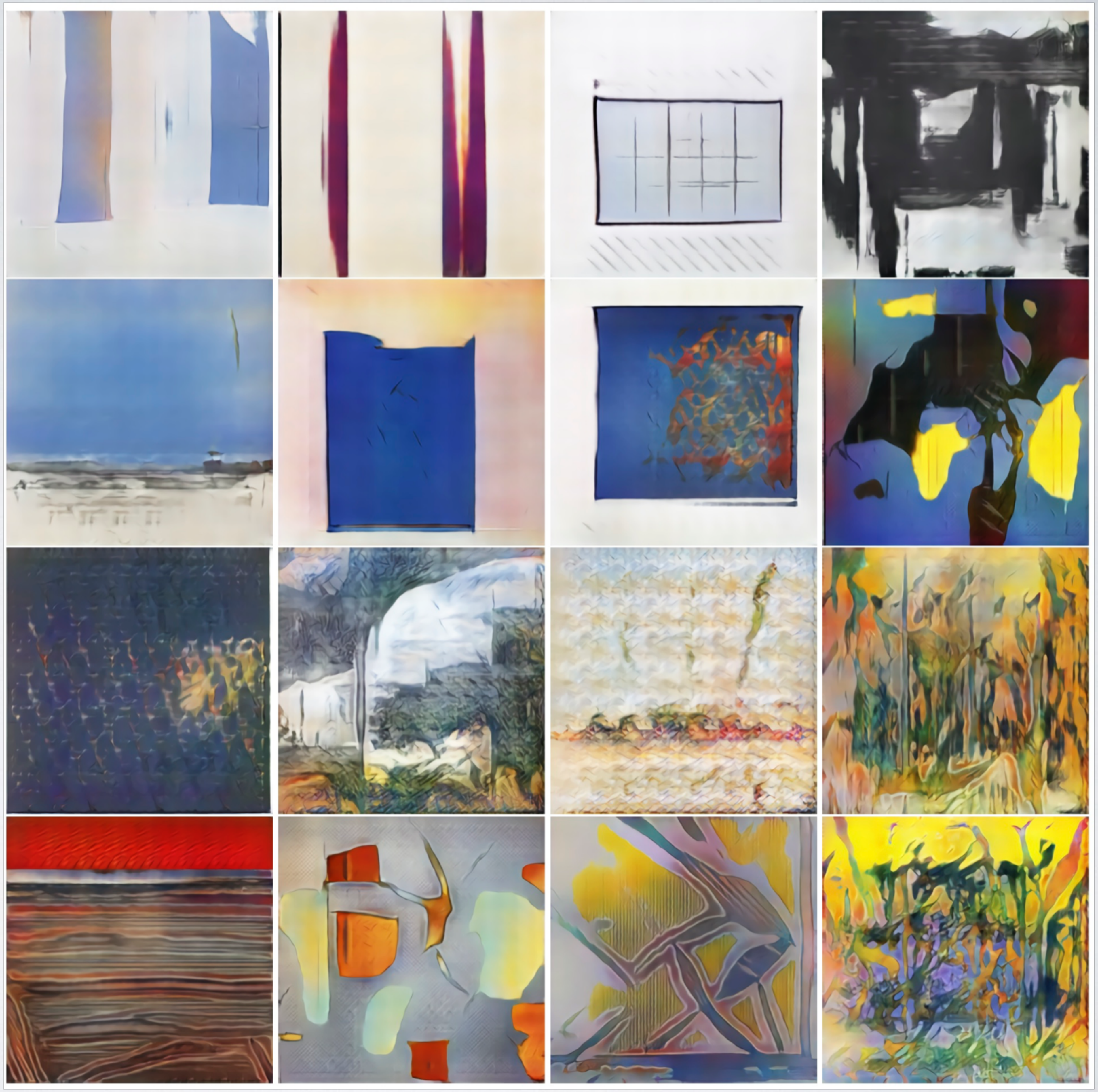
$$\mathbb{E}_{z\sim p_z}[\log(1-D_r(G(z))) - \sum_{k=1}^{K}\Big(\frac{1}{K}log(D_c(c_k|G(z))+$$

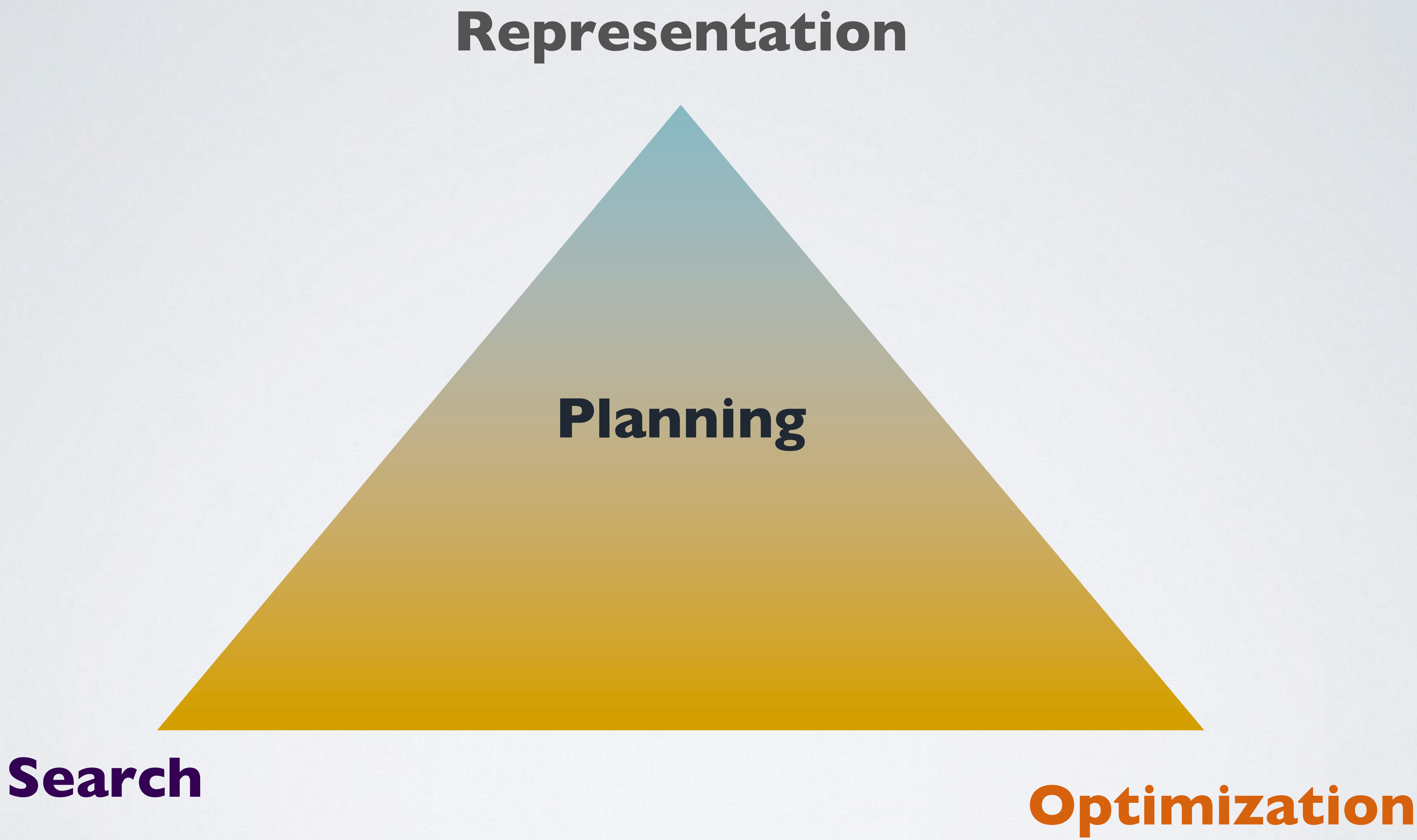
$$(1-\frac{1}{K})log(1-D_c(c_k|G(z))\Big)],$$

Table 1: Artistic Styles Used in Training

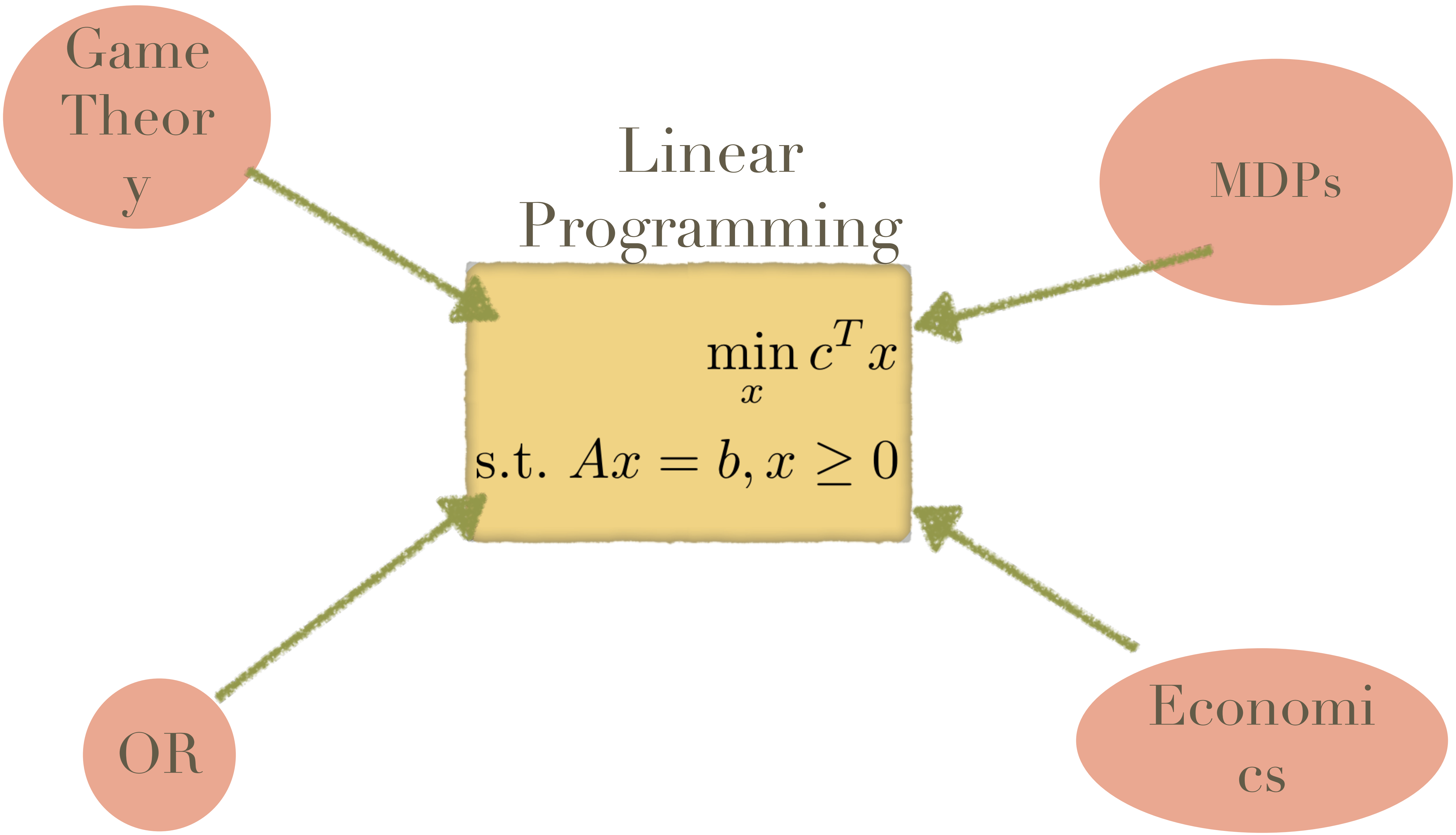
Style name	Image number	Style name	Image number
Abstract-Expressionism	2782	Mannerism-Late-Renaissance	1279
Action-Painting	98	Minimalism	1337
Analytical-Cubism	110	Naive Art-Primitivism	2405
Art-Nouveau-Modern	4334	New-Realism	314
Baroque	4241	Northern-Renaissance	2552
Color-Field-Painting	1615	Pointillism	513
Contemporary-Realism	481	Pop-Art	1483
Cubism	2236	Post-Impressionism	6452
Early-Renaissance	1391	Realism	10733
Expressionism	6736	Rococo	2089
Fauvism	934	Romanticism	7019
High-Renaissance	1343	Synthetic-Cubism	216
Impressionism	13060	Total	75753

Sample art
created by CAN




Representation
Planning
Search
Optimization

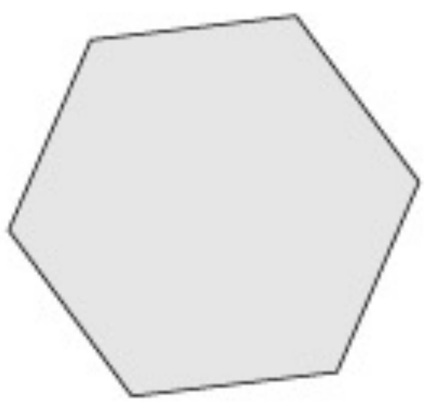
Convex Optimization



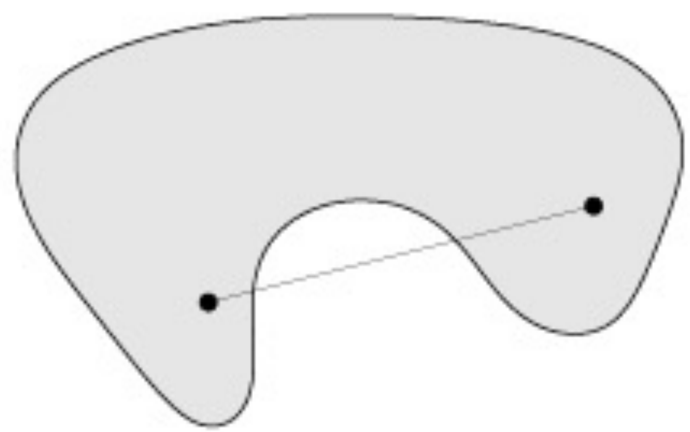
Convex set

contains the line segment between any two points in the set

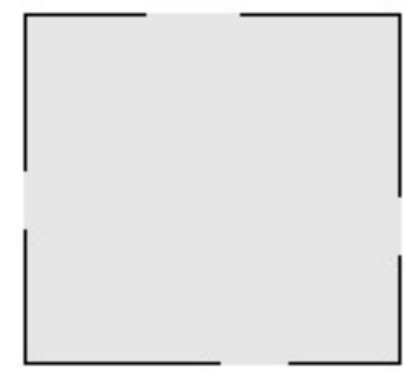
$$x_1, x_2 \in C, \quad 0 \leq \theta \leq 1 \quad \Longrightarrow \quad \theta x_1 + (1-\theta)x_2 \in C$$



convex

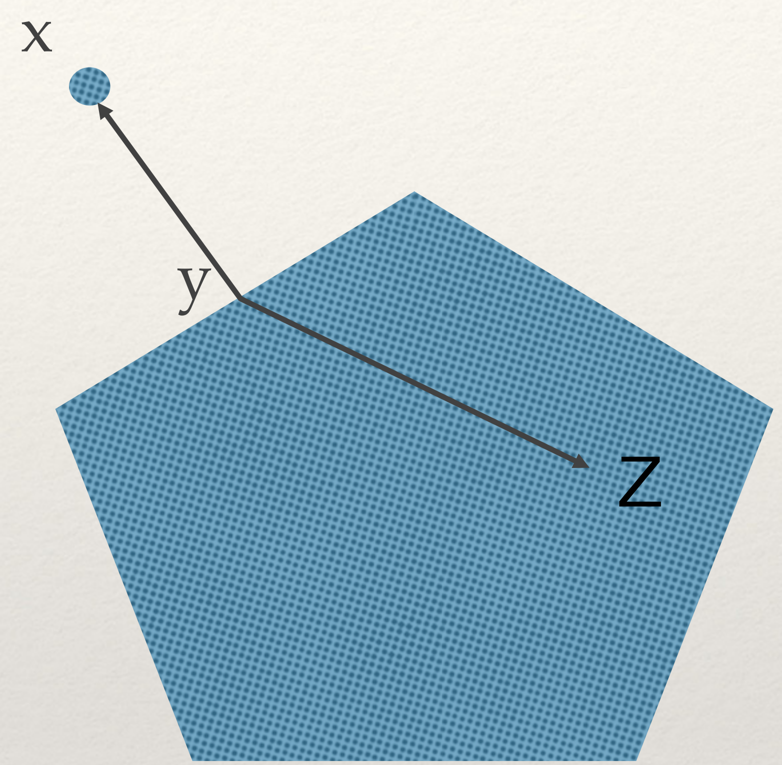


not convex



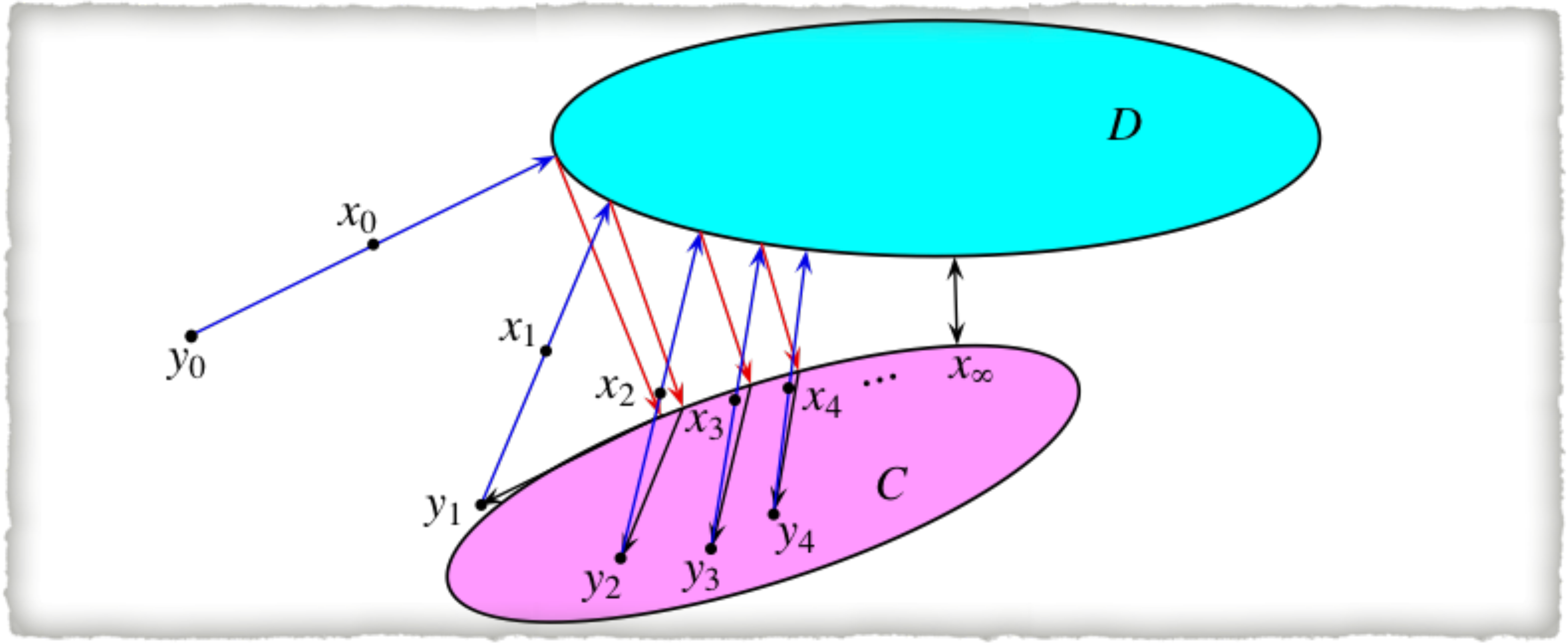
not convex

Projection onto Convex Sets



$$\text{if } y = \Pi_{\hat{S}}(x), \text{ then } (y-x)^T(z-y) \geq 0, z \in \hat{S}$$

Convex Feasibilty



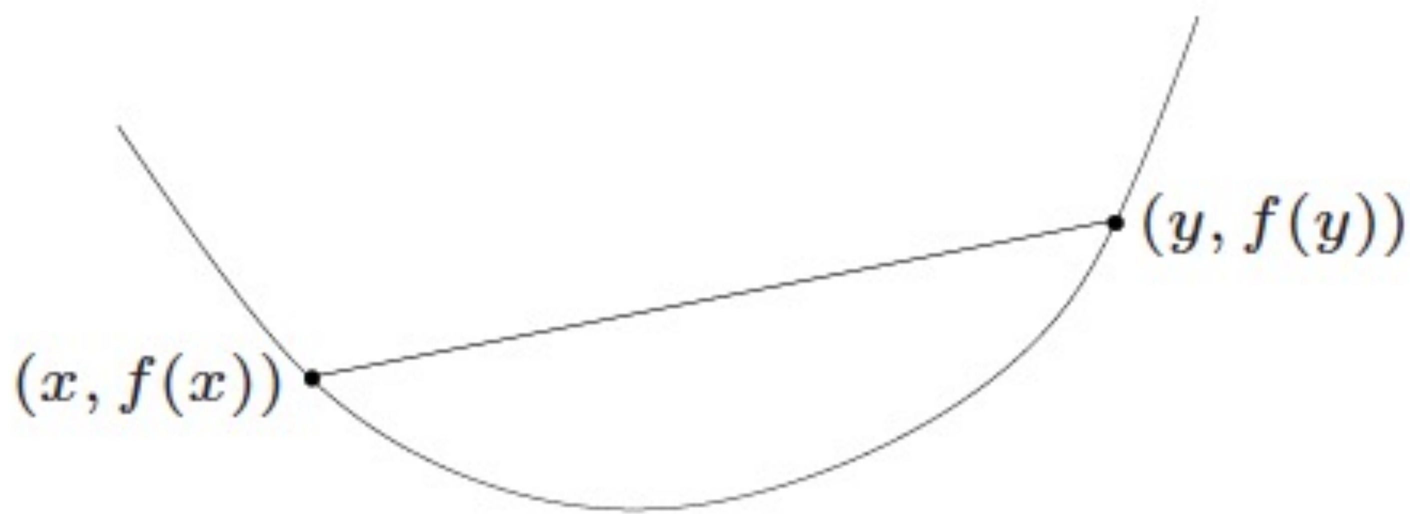
Proximal splitting methods in signal processing
Combetti and Pesquet

Convex function

convex function: domain is a convex set and Jensen's inequality holds:

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$

for all $x, y \in \textbf{dom}\, f$, $0 \leq \theta \leq 1$



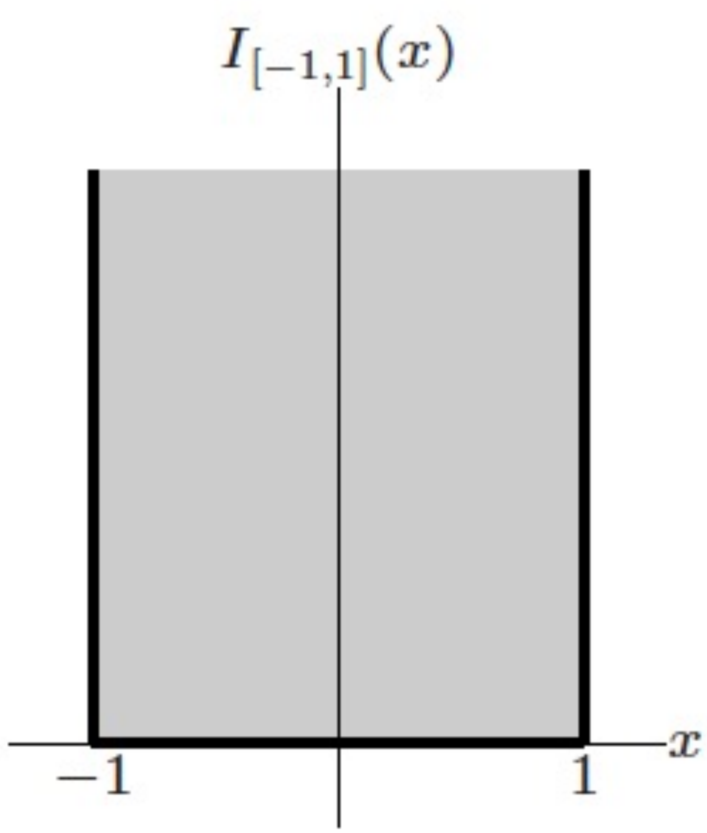
f is **strictly convex** if Jensen's inequality is strict for $0 < \theta < 1$, $x \neq y$

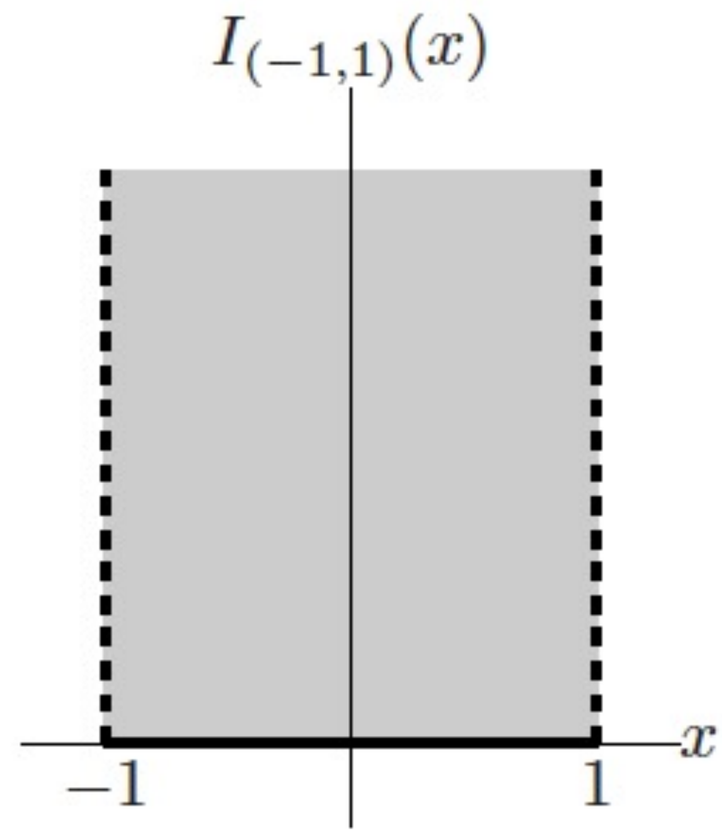
f is (strictly) **concave** if $-f$ is (strictly) convex

Indicator function

the indicator function of a set C is

$$I_C(x) = \begin{cases} 0 & x \in C \\ +\infty & \text{otherwise} \end{cases}$$

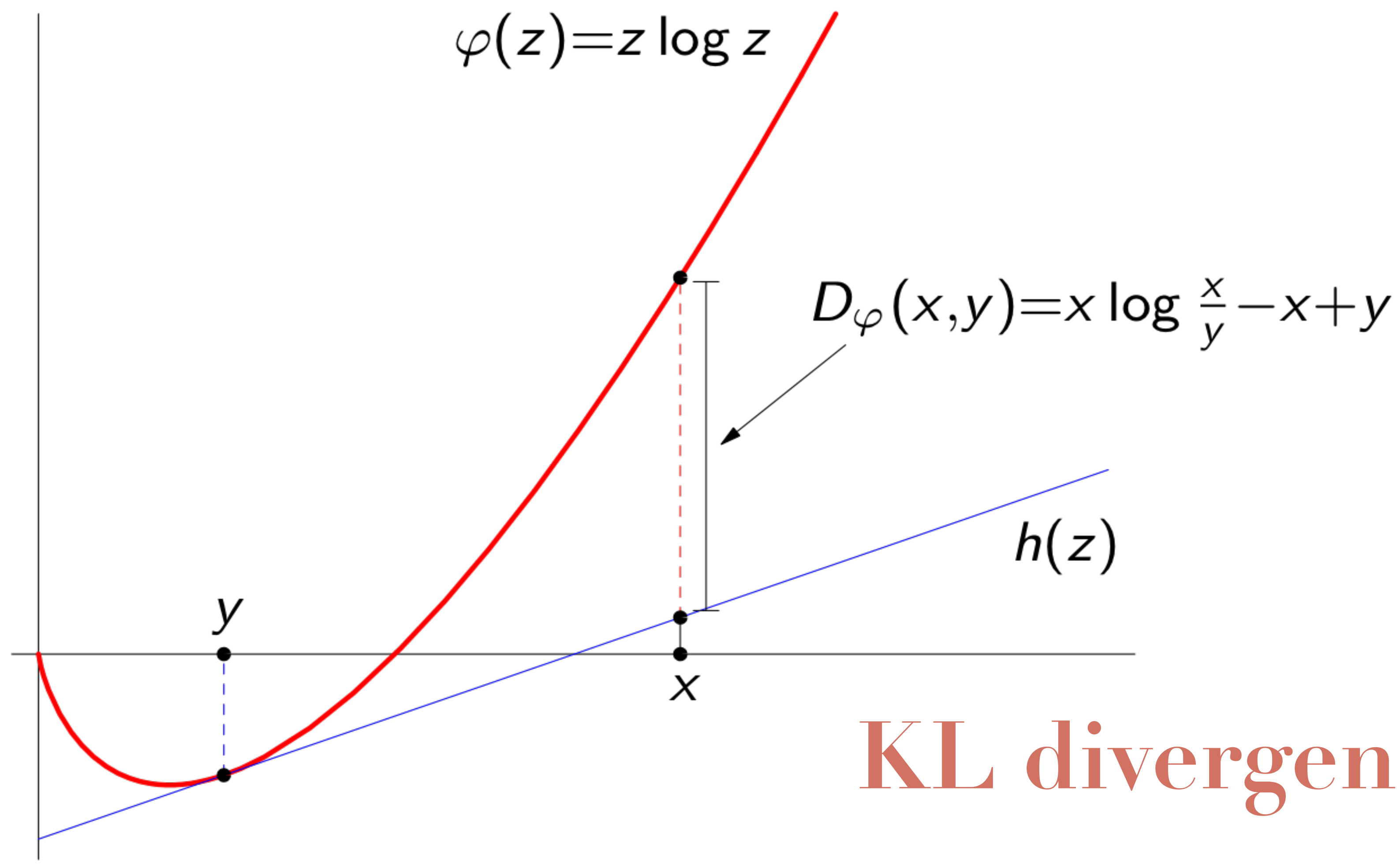




the indicator function of a convex set is a convex function

Bregman Divergence

$$D_\varphi(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{x}) - \varphi(\mathbf{y}) - (\mathbf{x} - \mathbf{y})^T \nabla \varphi(\mathbf{y})$$



KL divergence

Mirror Descent = Proximal Algm + Bregman Divergence

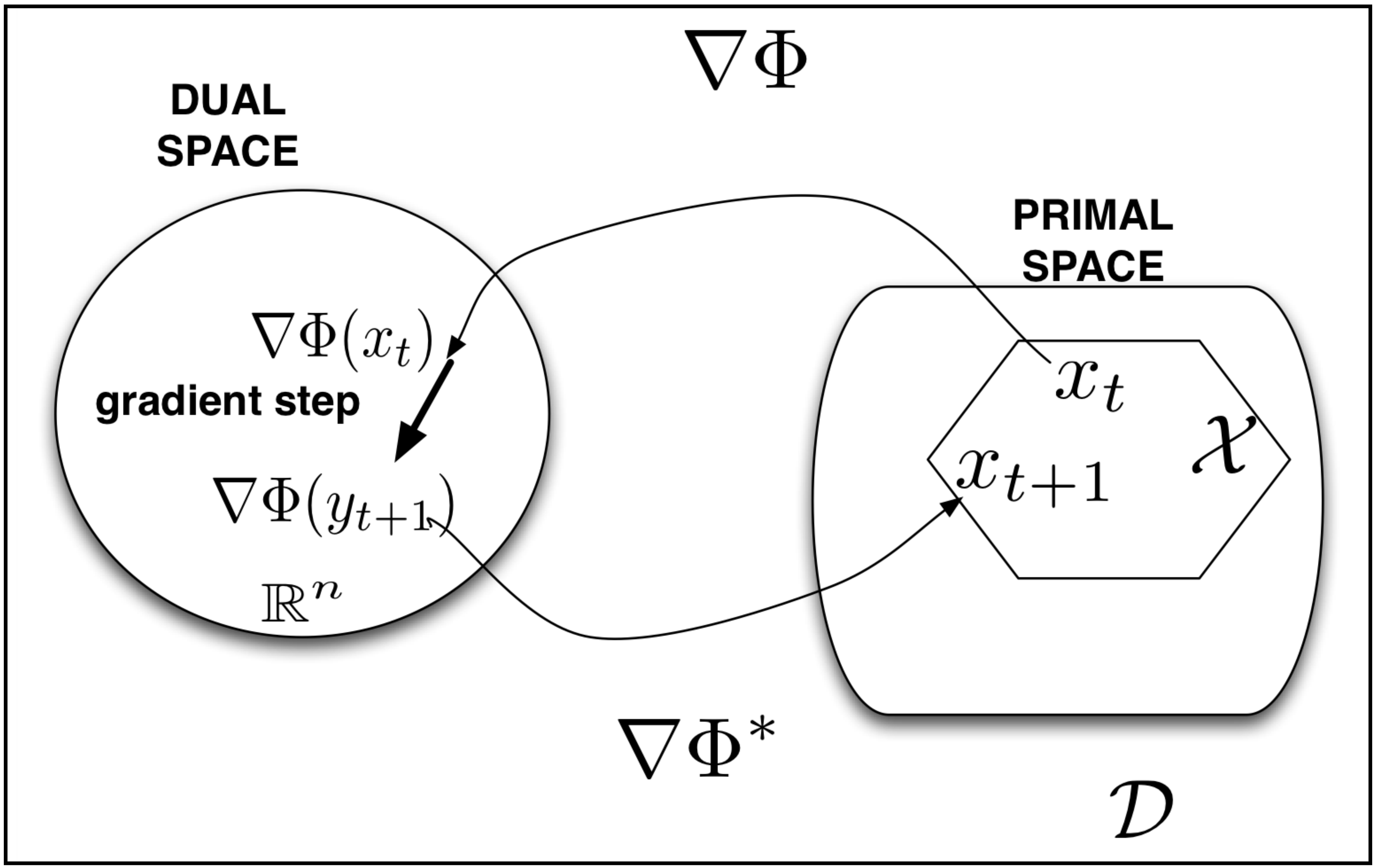
- Mirror descent can be viewed as a proximal method using a Bregman divergence

$$x_{k+1} = \operatorname{argmin}_{x \in X} \left(\langle x, \partial f(x_k) \rangle + \frac{1}{t_k} D_\phi(x, x_k) \right)$$

- Mirror descent can outperform regular subgradient method by O(n/log(n)) (Beck and Teboulle, 2003)

Mirror Maps

(Nemirovski and Yudin, 1980s; Bubeck, 2014)

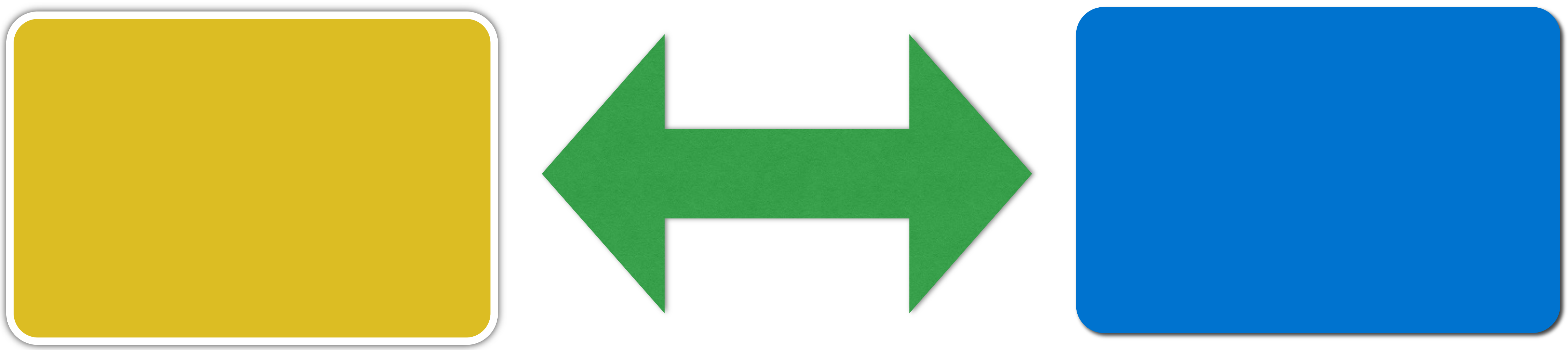


Mirror Descent

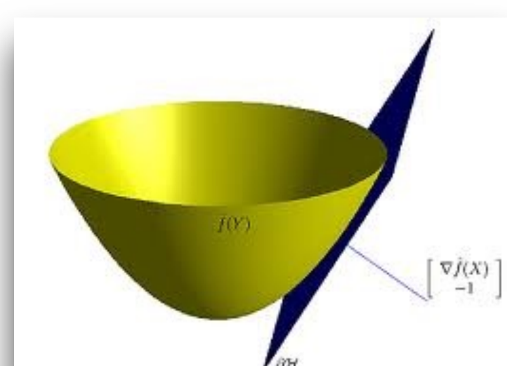
(Nemirovsky and Yudin)

“Imagination Space”

Dual   Primal



$$x_{k+1} = \nabla\psi^{*}\left(\nabla\psi(x_k) - t_k\partial f(x_k)\right)$$



Mirror Descent = "Natural" Gradient

(Nemirovsky and Yudin; Amari, 1980s)

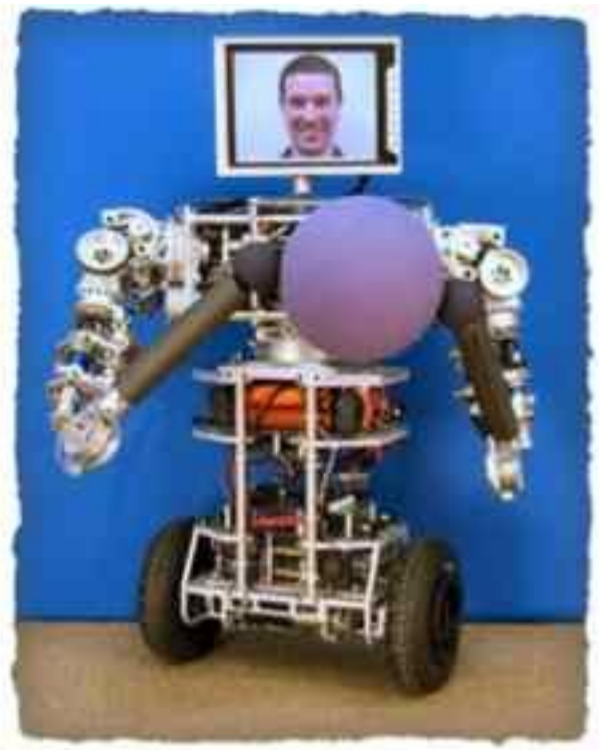
Mirror Descent

Mirror Map

Natural gradient

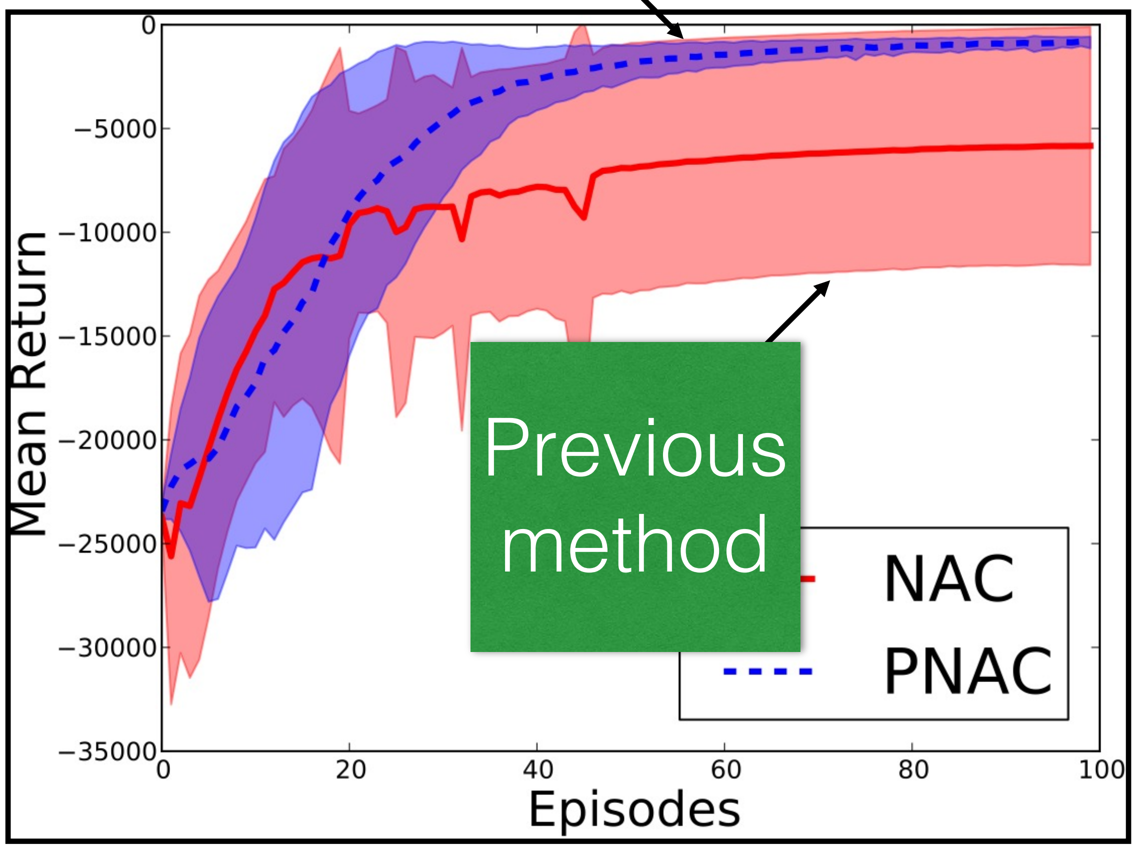
Thomas, Dabney, Mahadevan, Giguere, NIPS 2013

Safe Robot Learning

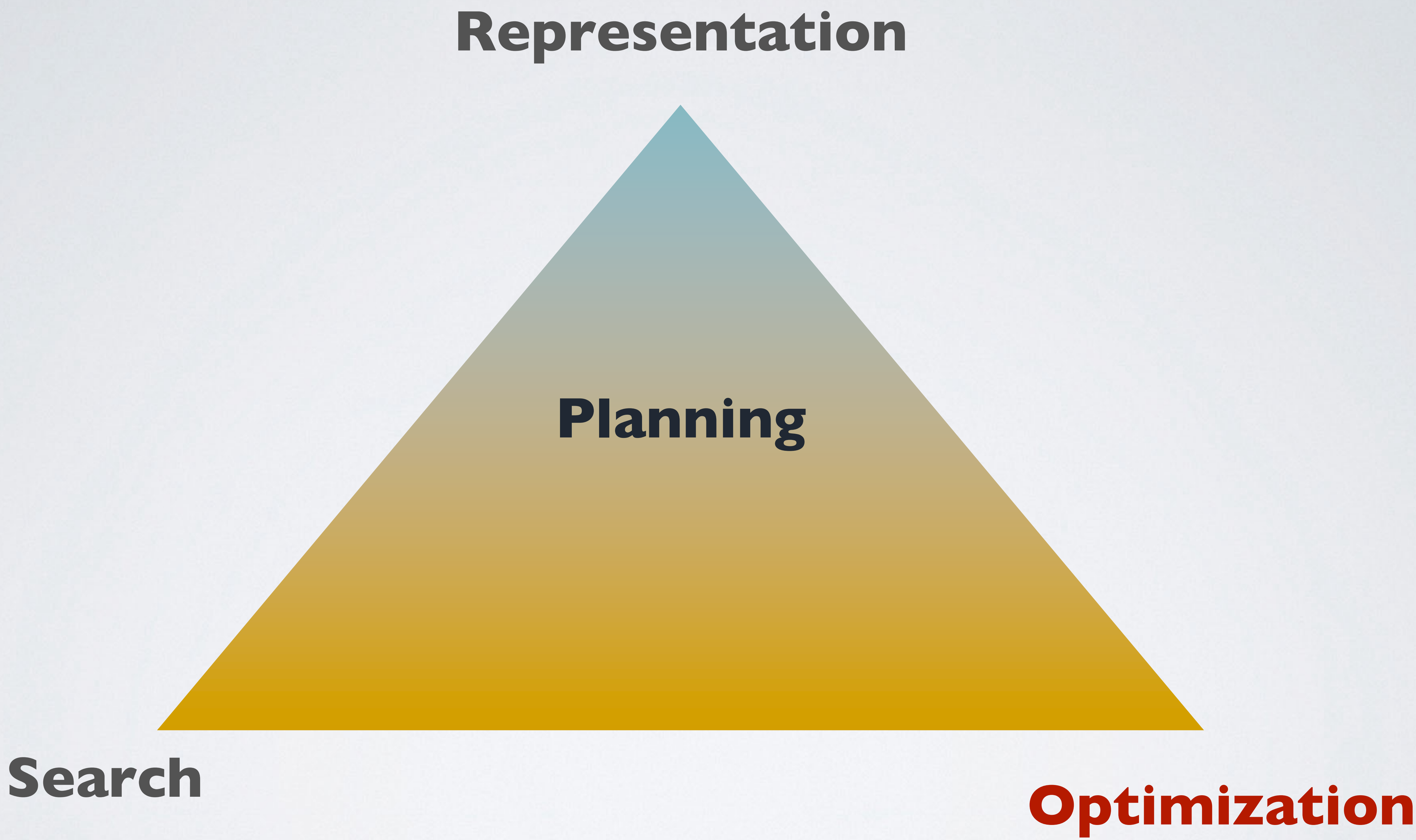


UBot, Laboratory of Perceptual Robotics

Our new method

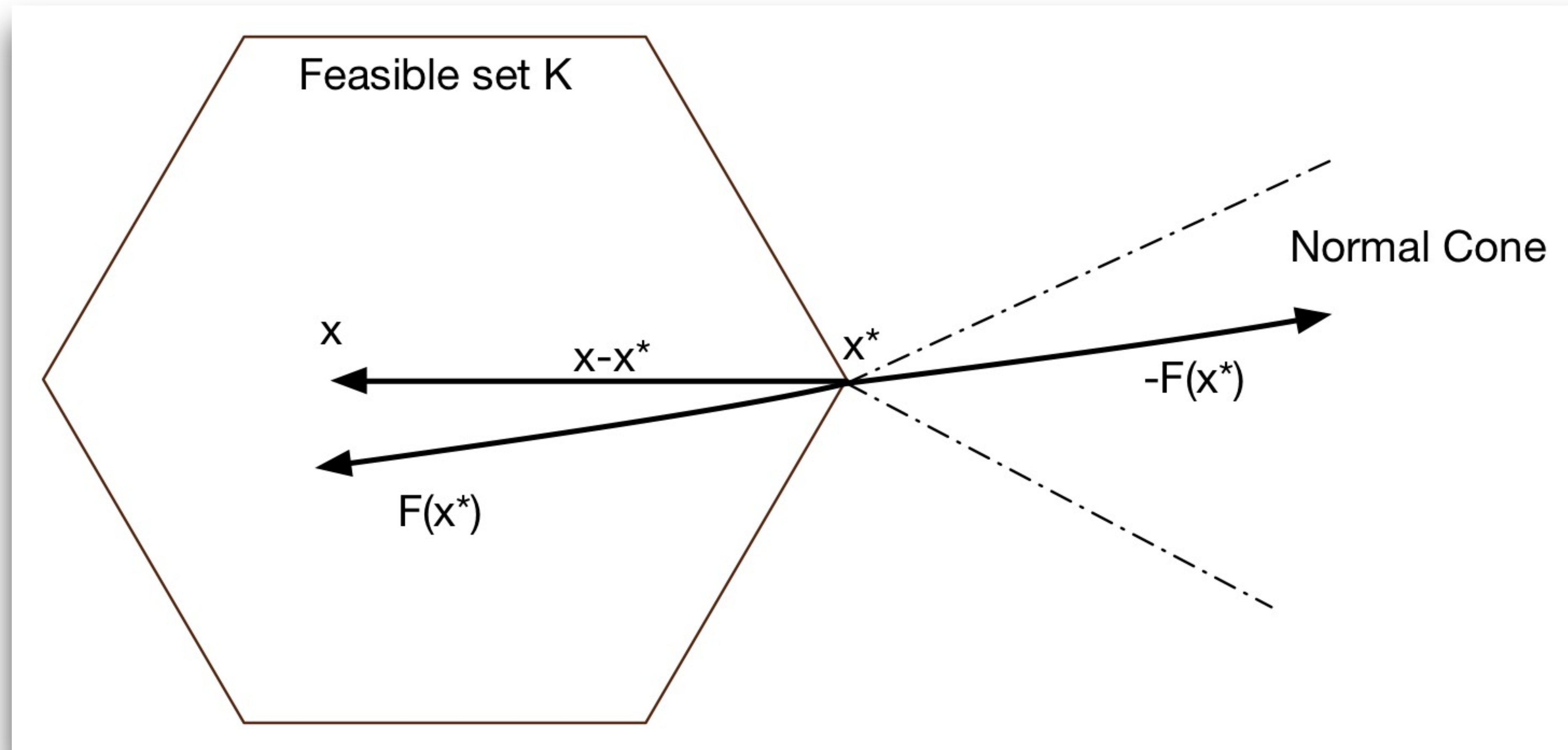


Thomas, Dabney, Mahadevan, Giguere, NIPS 2013


Representation
Planning
Search
Optimization

Variational Inequality

(Stampacchia, 1960s)



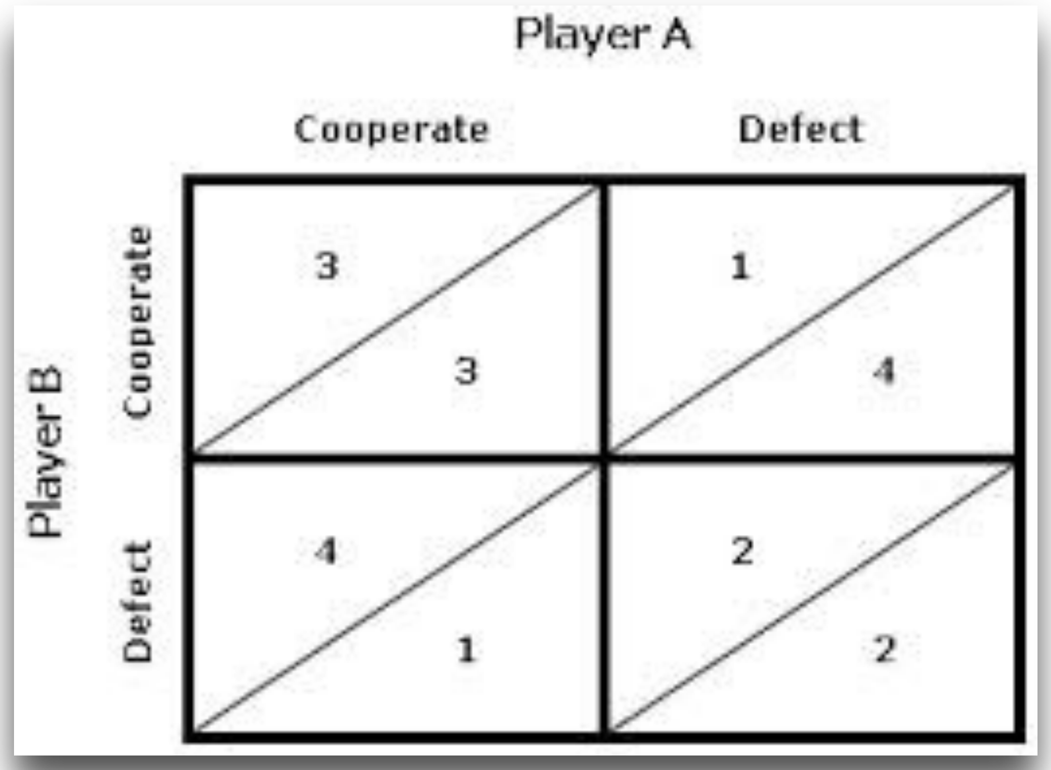
$$\langle F(x^*), x - x^* \rangle \geq 0, \ \forall x \in K$$

Vector field

Game theory => VI

- A CN game consists of m players, where player i chooses a strategy $x_i \in X_i$
- Let the joint payoffs for player i be $F_i(x_1,\ldots,x_m)$
- A set of strategies x* is in Nash equilibrium if

$$\langle (x_i - x_i^*), \nabla_i F_i(x_i^*) \rangle \geq 0$$



Optimization vs VIs

Property	Optimization	VI
Mapping	(Strong) Convexity	(Strong) Monotonicity
Jacobian	Positive definite and symmetric	Asymmetric
Objective function	Single fixed	Multiple or none

Monotone Operators

- relation F on $\mathbf{R}^n$ is *monotone* if

$$(u-v)^T(x-y) \geq 0 \quad \text{for all } (x,u),\ (y,v) \in F$$

- F is *maximal monotone* if there is no monotone operator that properly contains it

for f convex, $F(x) = \partial f(x)$ is monotone

- suppose $u \in \partial f(x)$ and $v \in \partial f(y)$
- then

$$f(y) \geq f(x) + u^T(y-x), \qquad f(x) \geq f(y) + v^T(x-y)$$

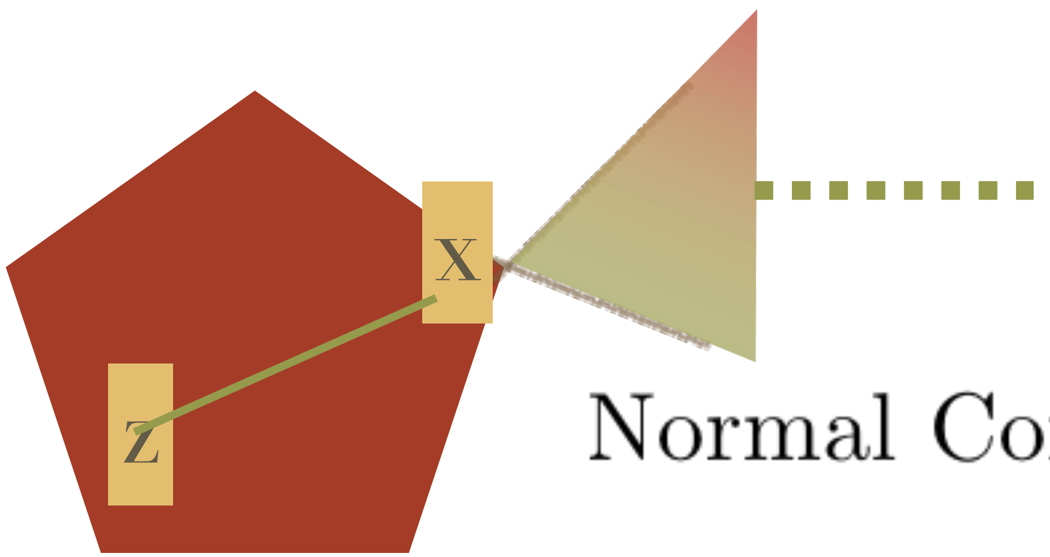
- add these and cancel $f(y) + f(x)$ to get

$$0 \leq (u-v)^T(x-y)$$

Subdifferentials

Subdifferential of a convex function:

$$\partial f(x) = \{v \in \mathbb{R}^n : f(z) \geq f(x) + v^T(z - x), \forall z \in \text{dom}(f)\}$$



Normal Cone: $\partial I_{\mathcal{C}}(x) = N_{\mathcal{C}}(x)$

$I_{\mathcal{C}}(x)$

Monotone Inclusion Problem

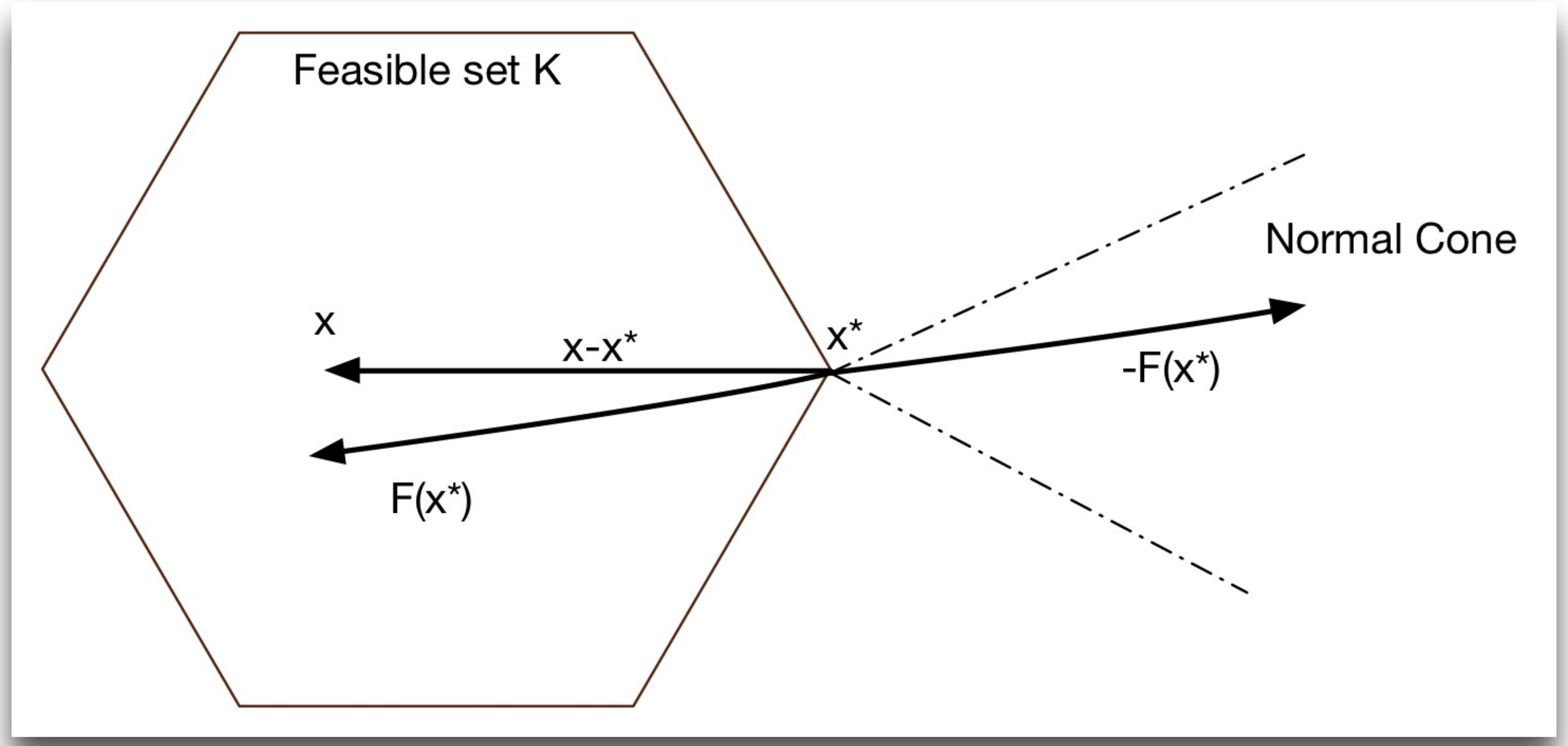
Given a monotone operator F, find x $s.t.$ $0 \in F(x)$

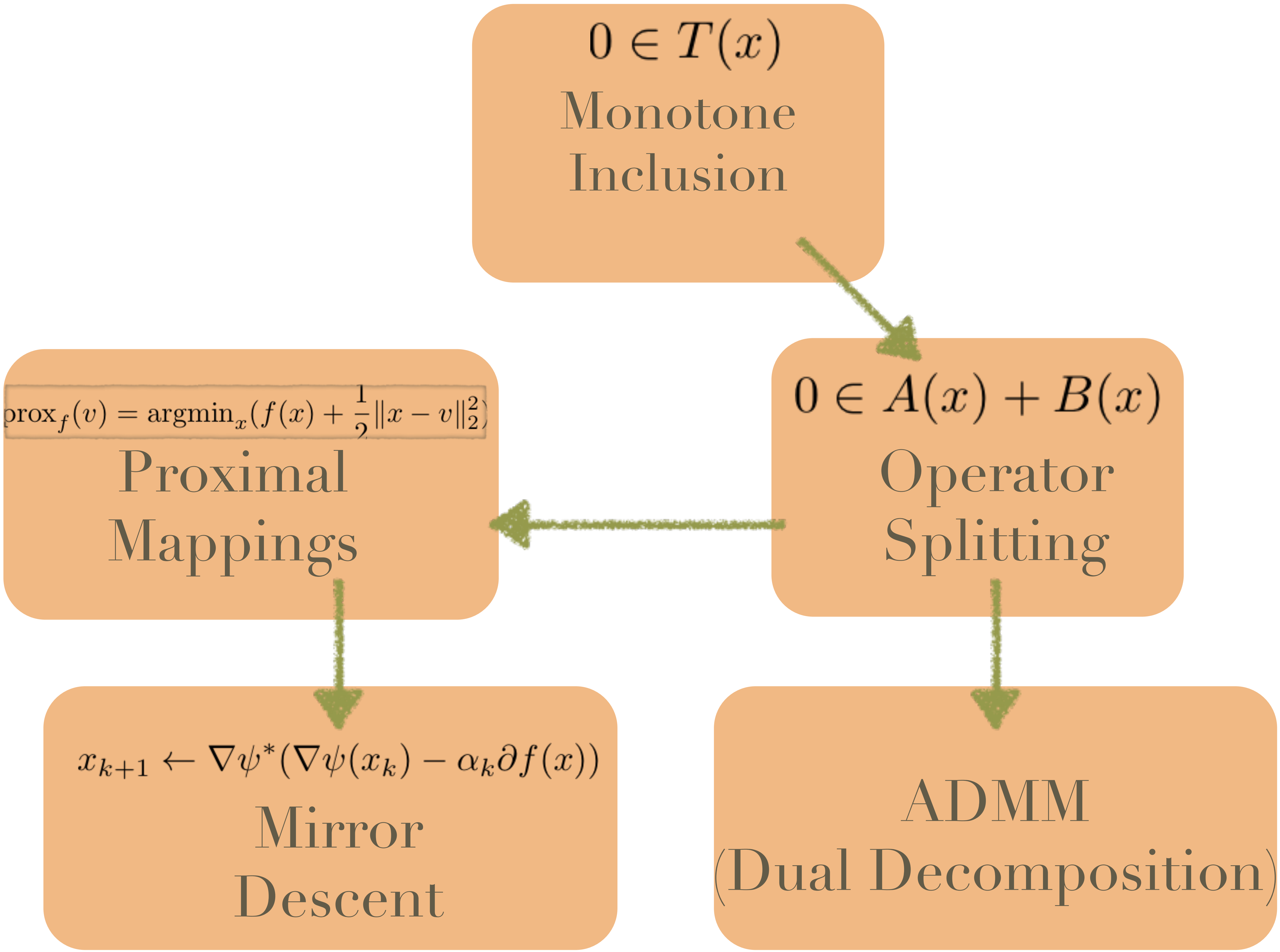
This means that $(x, 0) \in F$

For convex f, if x^* minimizes f, $0 \in \partial f(x^*)$

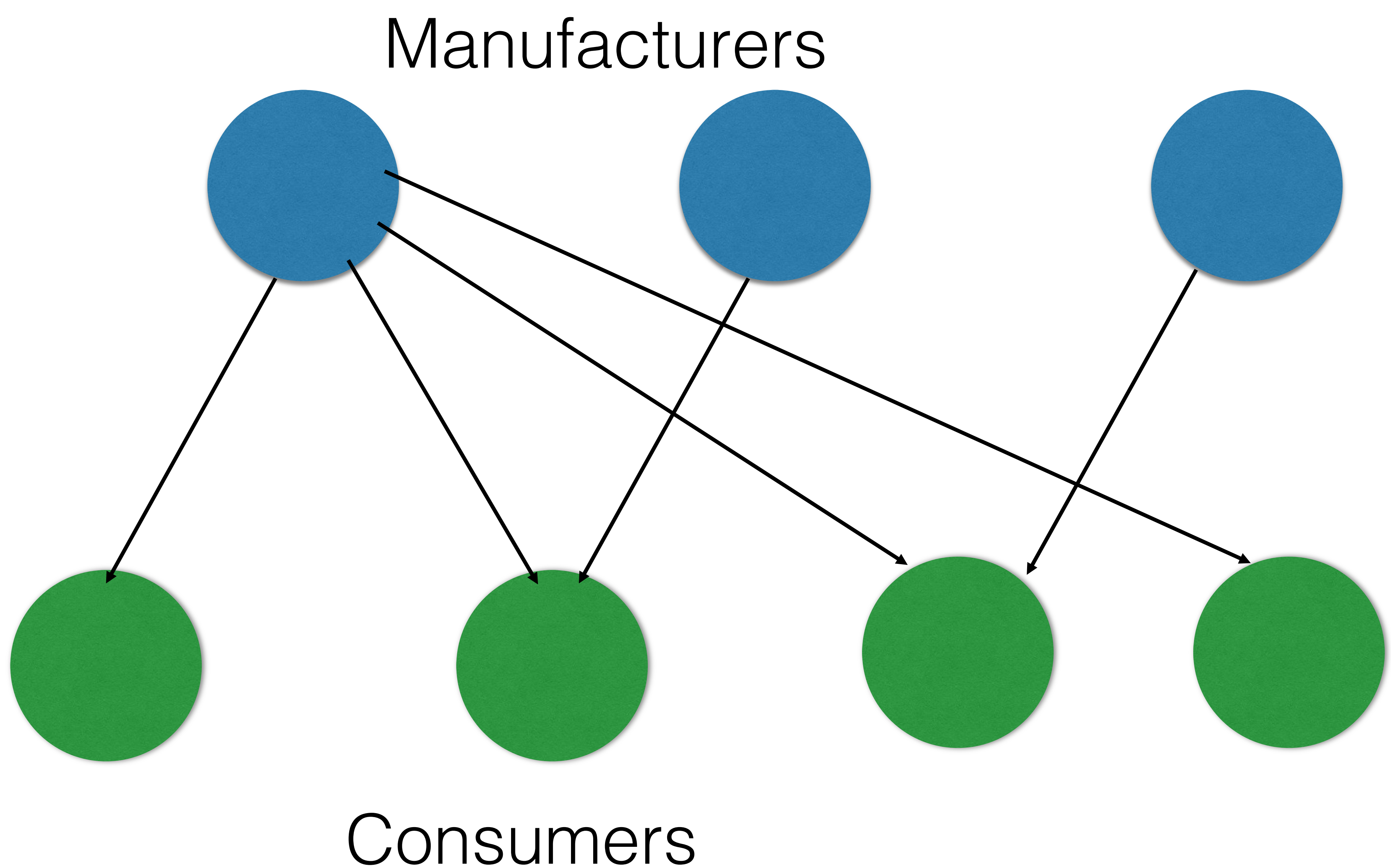
VI as monotone inclusion

$$0 \in F(x^*) + N_K(x^*)$$




$0 \in T(x)$
Monotone Inclusion
$0 \in A(x) + B(x)$
Operator Splitting
$\text{prox}_f(v) = \text{argmin}_x(f(x) + \frac{1}{2}\|x - v\|_2^2)$
Proximal Mappings
$x_{k+1} \leftarrow \nabla\psi^*(\nabla\psi(x_k) - \alpha_k \partial f(x))$
Mirror Descent
ADMM
(Dual Decomposition)

Spatial Price Equilibria



Spatial Price Equilibrium Model

$$\pi_i + c_{ij} = \begin{cases} \rho_j & \text{if } Q^*_{ij} > 0 \\ \geq \rho_j & \text{if } Q^*_{ij} = 0 \end{cases}$$

Supply level	**s_i**
Supply price	**π_j**
Quantity demanded	**d_i**
Demand price	**ρ_j**
Cost of transportation	**Q_{ij}**

Spatial Price Equilibrium Model as a VI

$$\langle \pi(s^*), s - s^* \rangle + \langle c(Q^*), Q - Q^* \rangle - \langle \rho(d^*), d - d* \rangle \geq 0 \quad \forall (s, Q, d) \in K$$

if and only if

$$\pi_i + c_{ij} = \begin{cases} \rho_j & \text{if } Q^*_{ij} > 0 \\ \geq \rho_j & \text{if } Q^*_{ij} = 0 \end{cases}$$

Can SPE be reduced to an Optimization Problem?

Only when

$$\frac{\partial \pi_i}{\partial s_k} = \frac{\partial \pi_k}{\partial s_i}$$

$$\frac{\partial \rho_j}{\partial d_l} = \frac{\partial \rho_l}{\partial d_j}$$

This is obviously not true in general

Optimization => VI

Suppose $x^* = \text{argmin}_{x \in \mathcal{K}} f(x)$

where f is differentiable

Then x^* solves the VI

$\langle \nabla f(x^*), x - x^* \rangle \geq 0. \ \forall x \in \mathcal{K}$

Proof: Define $\phi(t) = f(x^* + t(x - x^*))$

Since $\phi(0)$ achieves the minimum

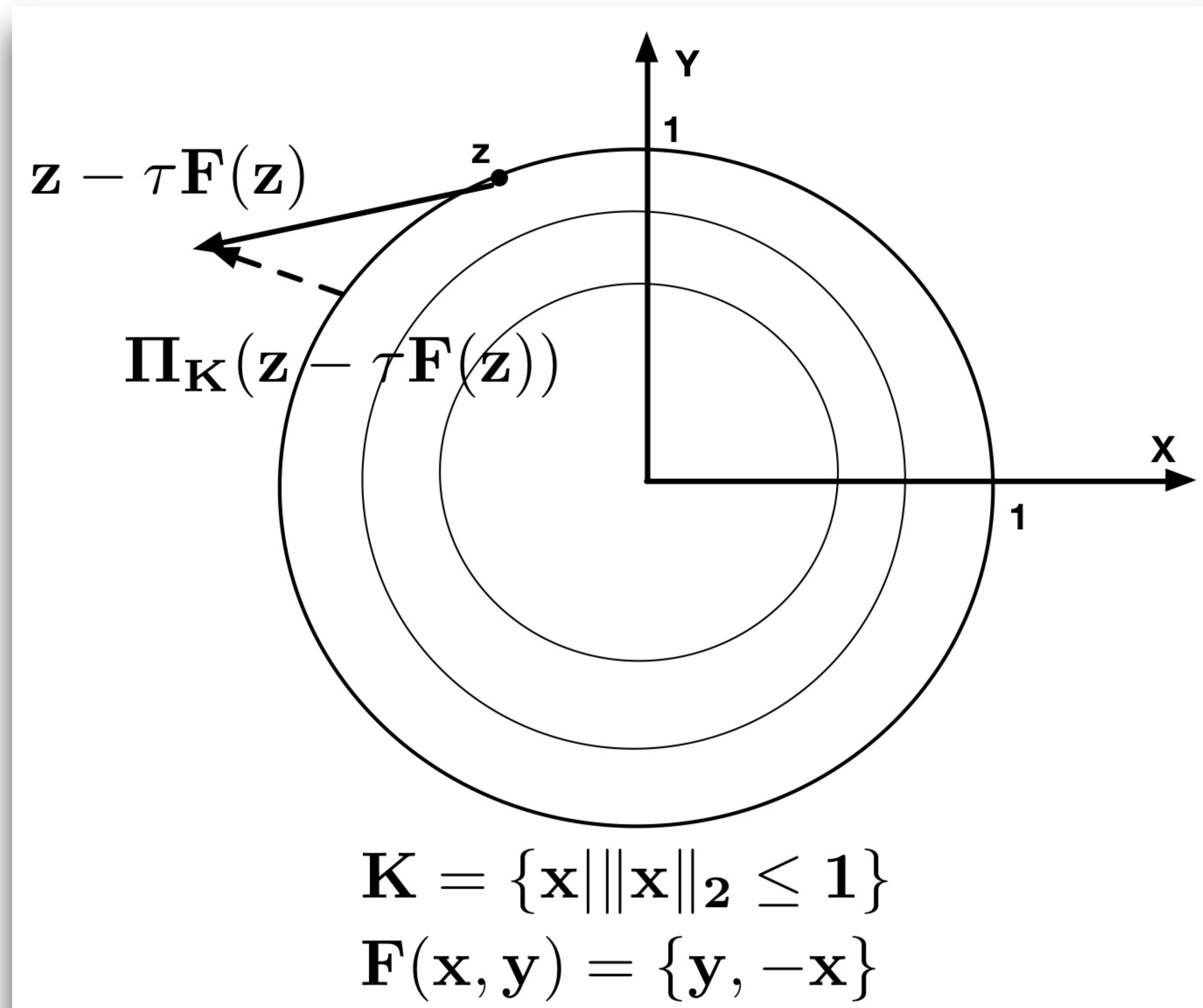
$\phi'(0) = \langle \nabla f(x^*), x - x^* \rangle \geq 0$

When VI => optimization?

$$\text{Given } VI(F,K), \text{ define } \nabla F(x) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial F_n}{\partial x_1} & \cdots & \frac{\partial F_n}{\partial x_n} \end{bmatrix}$$

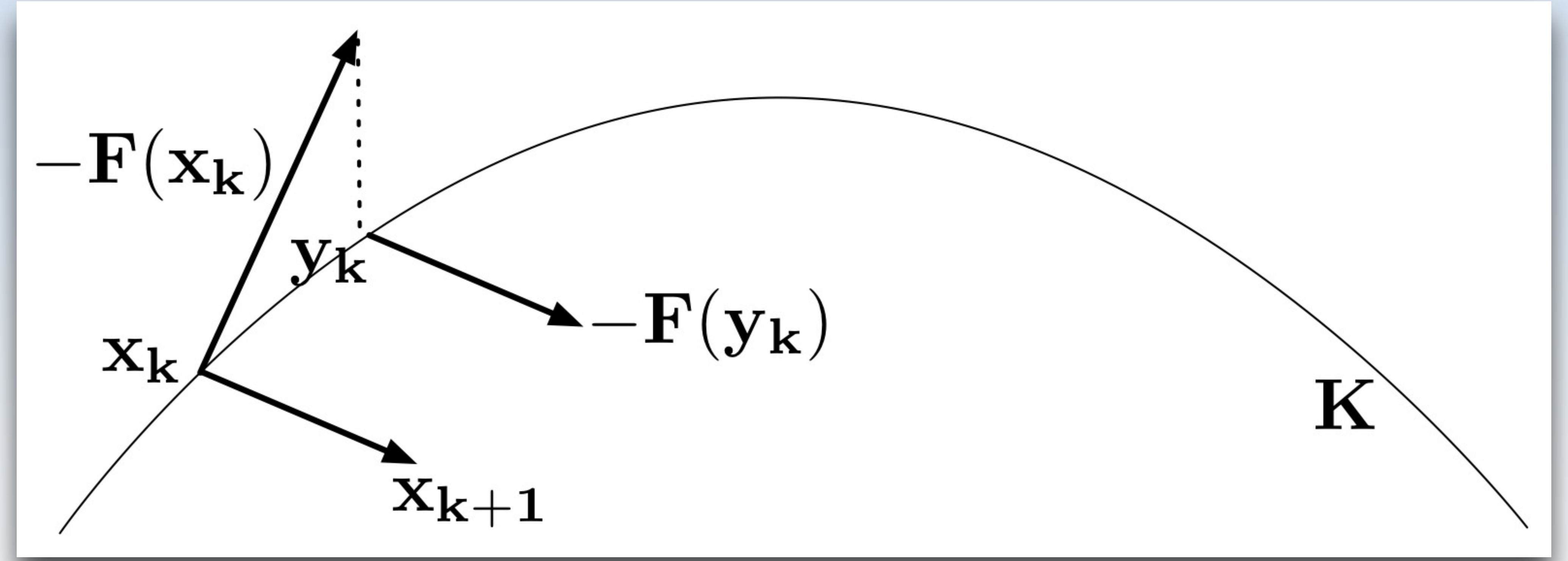
When ∇F is symmetric and positive semi-definite VI(F,K) can be reduced to an optimization problem,

Projection Method Fails



Bertsekas and Tsitsiklis, Parallel and Distributed Computation, Athena Scientific.

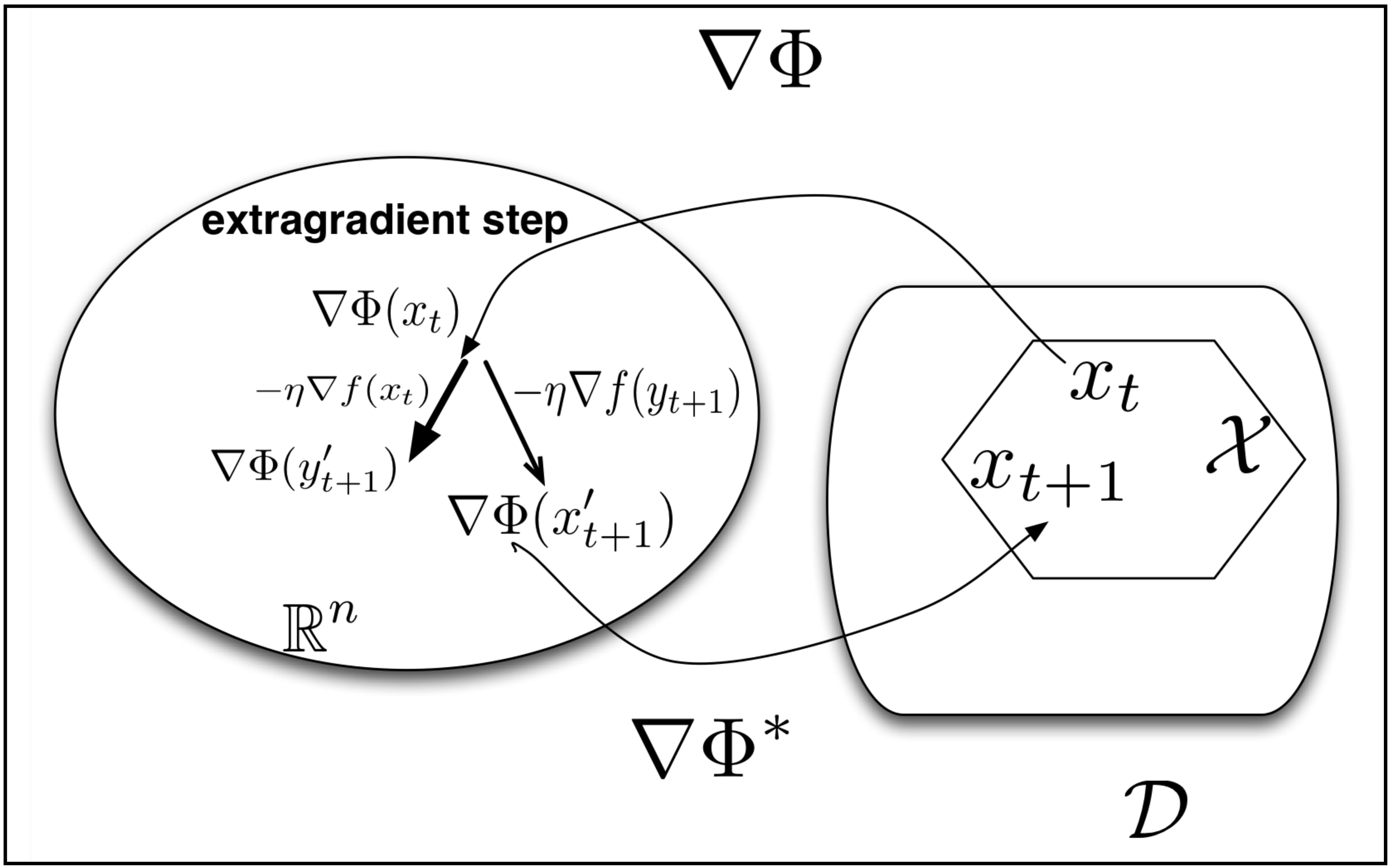
Extragradient Method



Korpolevich (1970s) developed the extragradient method for solving saddle point problems and variational inequalities

Mirror-Prox: Non-Euclidean Extragradient

(Nemirovski, 2005)



Runge-Kutta Method for VIs

(Gemp & Mahadevan, AAAI Workshop 2014, Spring Symposium 2015)

Runge Kutta (4) Gradient Descent

$$k_1 = \alpha \nabla F(x_k)$$
$$k_2 = \alpha \nabla F(x_k - \tfrac{1}{2} k_1)$$
$$k_3 = \alpha \nabla F(x_k - \tfrac{1}{2} k_2)$$
$$k_4 = \alpha \nabla F(x_k - k_3)$$

$$x_{k+1} = x_k - \tfrac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

General Runge Kutta Gradient Descent

$$k_1 = \alpha \nabla F(x_k)$$
$$k_2 = \alpha \nabla F(x_k - a_{21} k_1)$$
$$k_3 = \alpha \nabla F(x_k - a_{31} k_1 - a_{32} k_2)$$
$$\vdots$$
$$k_s = \alpha \nabla F(x_k - a_{s1} k_1 - a_{s2} k_2 - \ldots - a_{s,s-1} k_{s-1})$$

$$x_{k+1} = x_k - \Sigma_{i=1}^{s} b_i k_i$$

Runge Kutta (4) Non-Euclidean Extragradient

$$k_1 = \alpha F(x_k)$$
$$k_2 = \alpha F(\nabla \psi_k^*(\nabla \psi_k(x_k) - \tfrac{\alpha}{2} k_1))$$
$$k_3 = \alpha F(\nabla \psi_k^*(\nabla \psi_k(x_k) - \tfrac{\alpha}{2} k_2))$$
$$k_4 = \alpha F(\nabla \psi_k^*(\nabla \psi_k(x_k) - \alpha k_3))$$
$$x_{k+1} = \nabla \psi_k^*(\nabla \psi_k(x_k) - \tfrac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4))$$

General RK Non-Euclidean Extragradient

$$k_1 = \alpha F(x_k)$$
$$k_2 = \alpha F(\nabla \psi_k^*(\nabla \psi_k(x_k) - a_{21} k_1))$$
$$k_3 = \alpha F(\nabla \psi_k^*(\nabla \psi_k(x_k) - a_{31} k_1 - a_{32} k_2))$$
$$\vdots$$
$$k_s = \alpha F(\nabla \psi_k^*(\nabla \psi_k(x_k) - a_{s1} k_1 - a_{s2} k_2 - \ldots - a_{s,s-1} k_{s-1}))$$

$$x_{k+1} = \nabla \psi_k^*(\nabla \psi_k(x_k) - \Sigma_{i=1}^{s} b_i k_i)$$

Next Generation Internet Model [Nagurney et al., 2014]

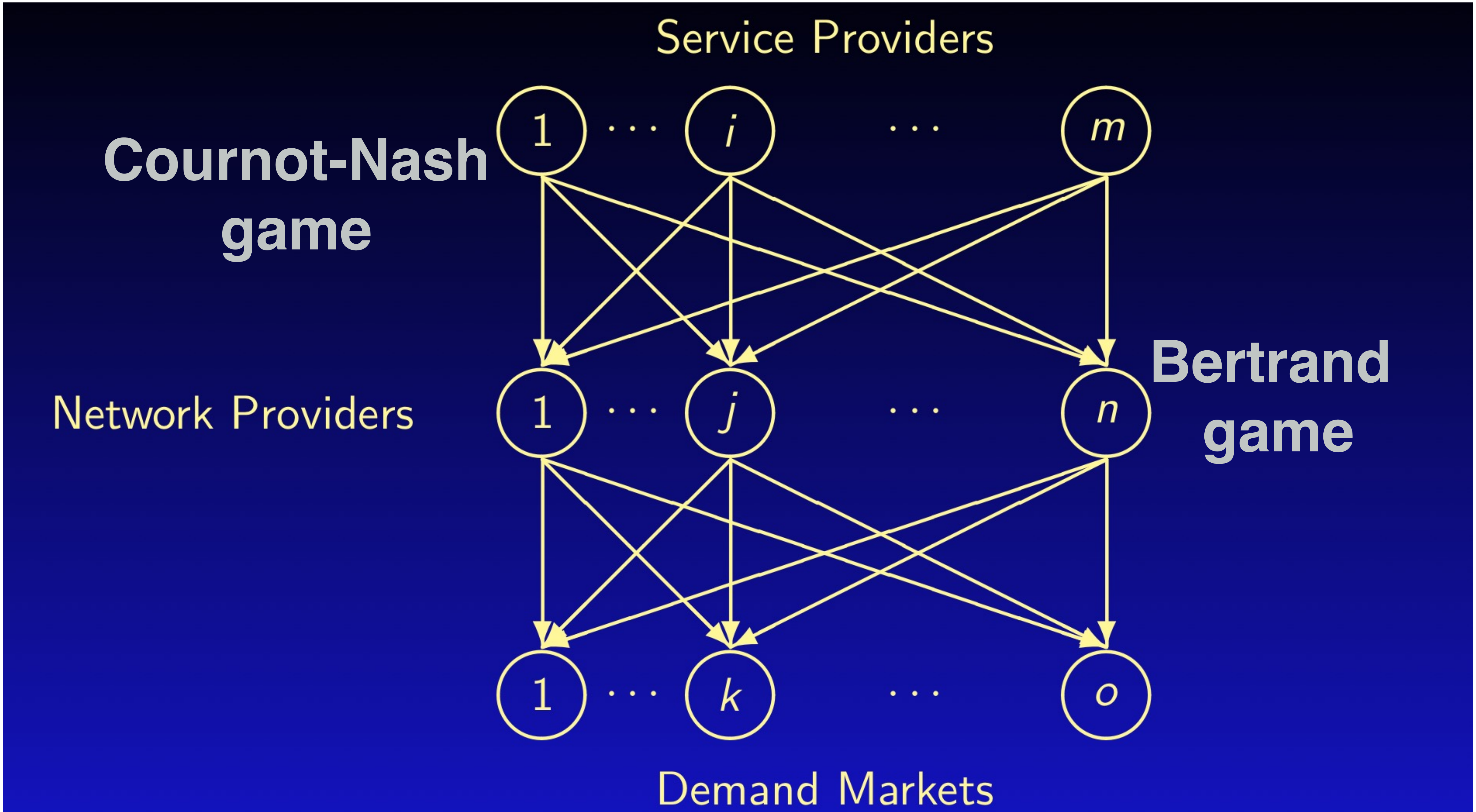


Figure 1: The Network Structure of the Cournot-Nash-Bertrand Model for a Service-Oriented Internet

Problem Formulation

Table 1: Notation for the Game Theoretic Cournot-Nash-Bertrand Model

Notation	Definition
Q_{ijk}	the nonnegative service volume from i to k via j. We group the $\{Q_{ijk}\}$ elements for all j and k into the vector $Q_i \in R_+^{no}$ and then we group all the vectors Q_i for all i into the vector $Q \in R_+^{mno}$.
s_i	the service volume (output) produced by service provider i. We group the $\{s_i\}$ elements into the vector $s \in R_+^m$.
q_{ijk}	the nonnegative quality level of network provider j transporting service i to k. We group the q_{ijk} for all i and k into the vector $q_j \in R_+^{mo}$ and all the vectors q_j for all j into the vector $q \in R_+^{mno}$.
π_{ijk}	the price charged by network provider j for transporting a unit of service provided by i via j to k. We group the π_{ijk} for all i and k into the vector $\pi_j \in R_+^m$ and then we group all the vectors π_j for all j into the vector $\pi \in R_+^{mno}$.
$f_i(s)$	the total production cost of service provider i.
$\hat{\rho}_{ijk}(Q,q)$	the demand price at k associated with service i transported via j.
$c_{ijk}(Q,q)$	the total transportation cost associated with delivering service i via j to k.
$oc_{ijk}(\pi_{ijk})$	the opportunity cost associated with pricing by network provider j services transported from i to k.

VI Fomulation

- Production cost function f(Q) - cost of providing a certain volume of content

$$\langle F(X^*), X - X^* \rangle \geq 0, \quad \forall X \in \mathcal{K}, \qquad X \equiv (Q, q, \pi)$$

Demand price function - user offer depends on content quality and market volume

- Transportation cost function c(Q,q) - cost of transporting content from service provider to user

- Opportunity cost function oc(π) - network providers lose business due to high prices

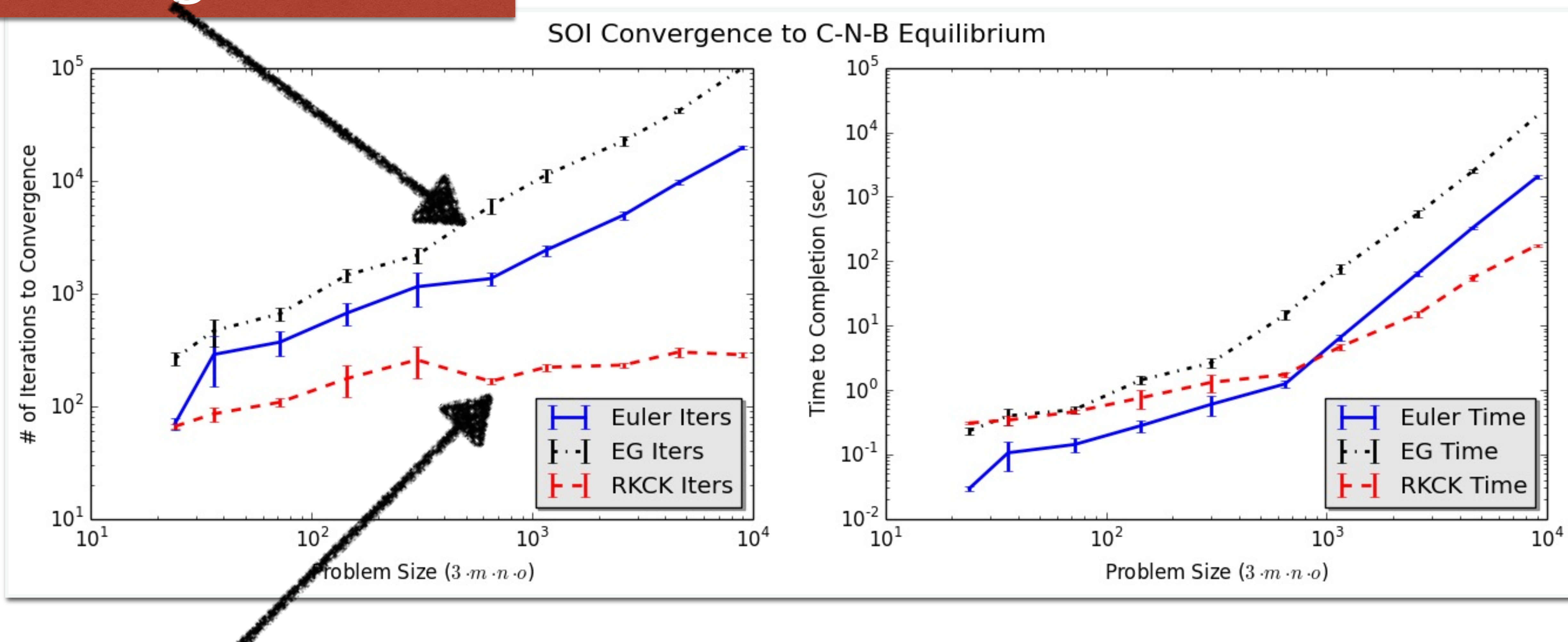
$$F^1_{ijk}(X) = \frac{\partial \hat{f}_i(Q)}{\partial Q_{ijk}} + \pi_{ijk} - \hat{\rho}_{ijk}(Q,q) - \sum_{h=1}^{n} \sum_{l=1}^{o} \frac{\partial \hat{\rho}_{ihl}(Q,q)}{\partial Q_{ijk}} \times Q_{ihl},$$

$$F^2_{ijk}(X) = \sum_{h=1}^{m} \sum_{l=1}^{o} \frac{\partial c_{hjl}(Q,q)}{\partial q_{ijk}},$$

$$F^3_{ijk}(X) = -Q_{ijk} + \frac{\partial oc_{ijk}(\pi_{ijk})}{\partial \pi_{ijk}}.$$

Results on Internet VI Problem

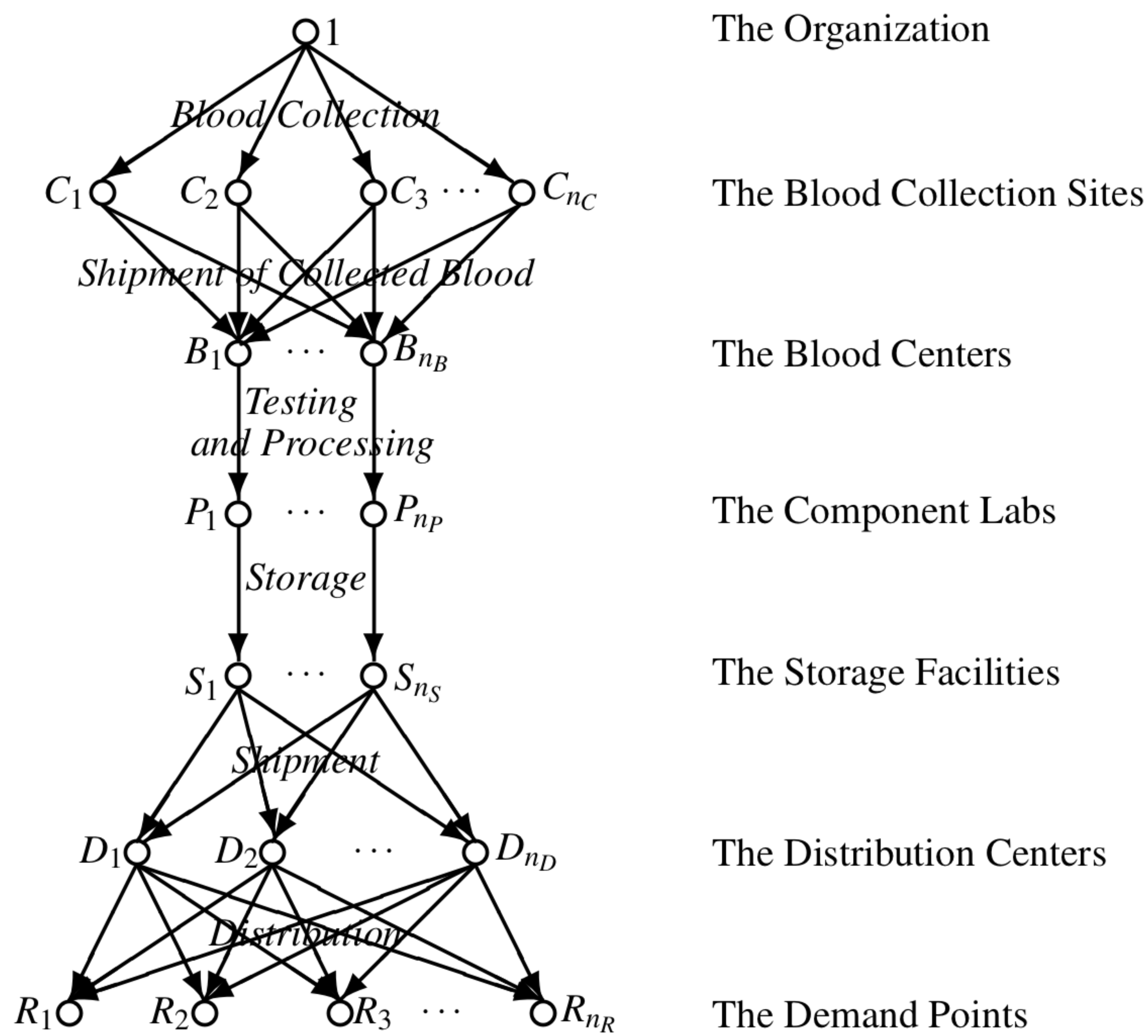
Extragradient



Our new algorithm

Substainable Blood Banking

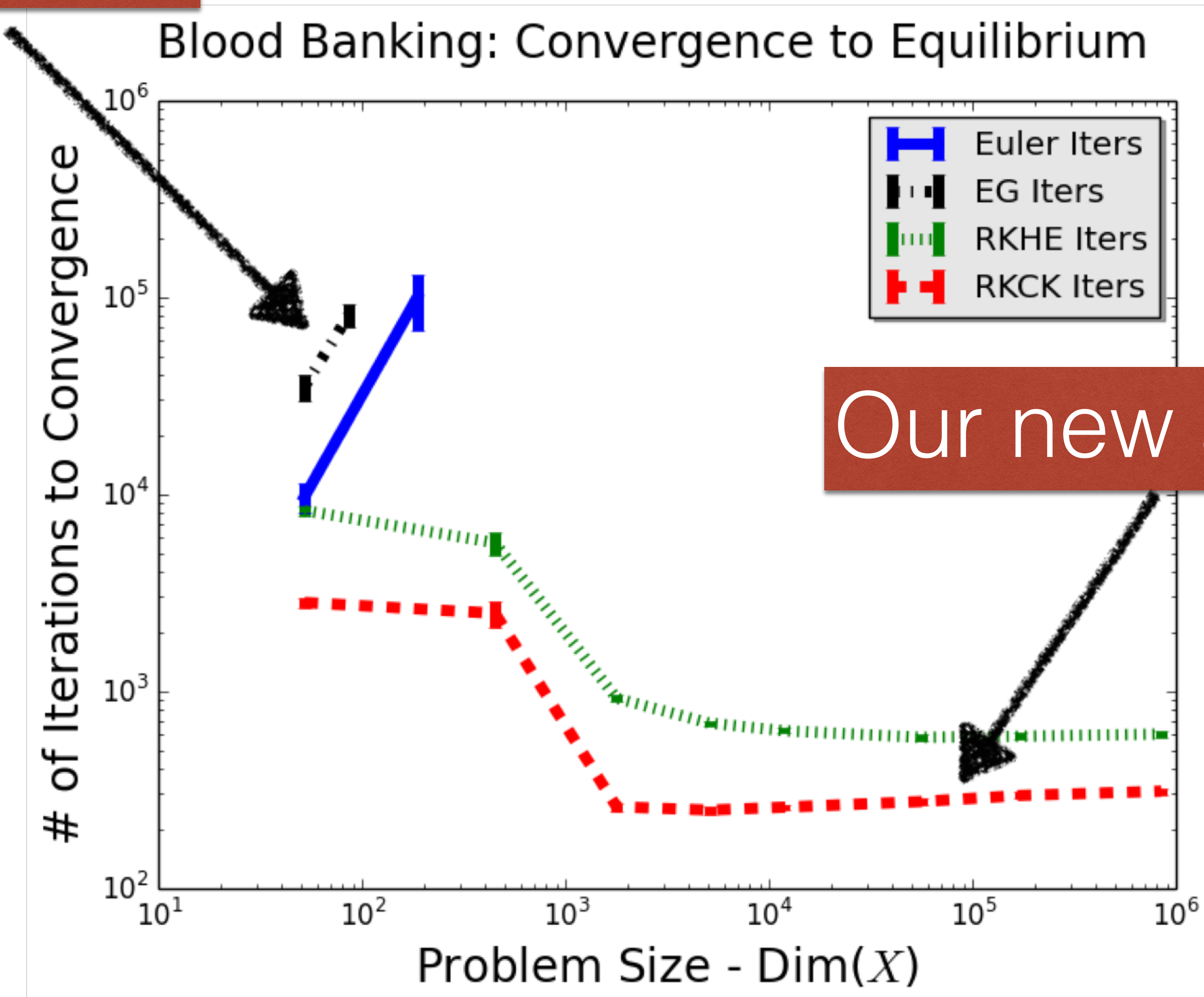
(Nagurney and Masoumi, 2012)



Experimental Results

(Gemp and Mahadevan, AAAI Workshop on Computational Sustainability, 2015)

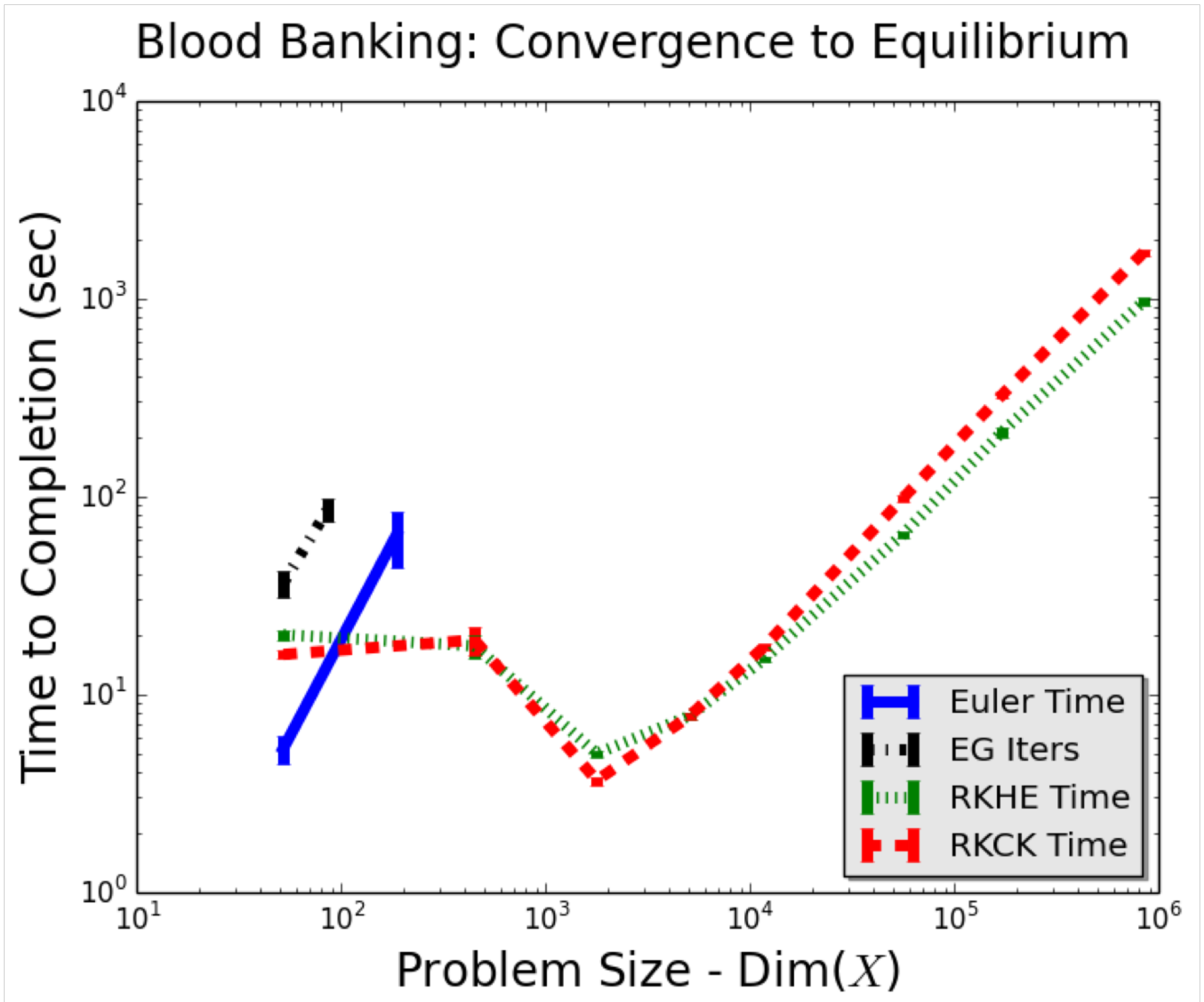
Extragradient



Our new algorithm

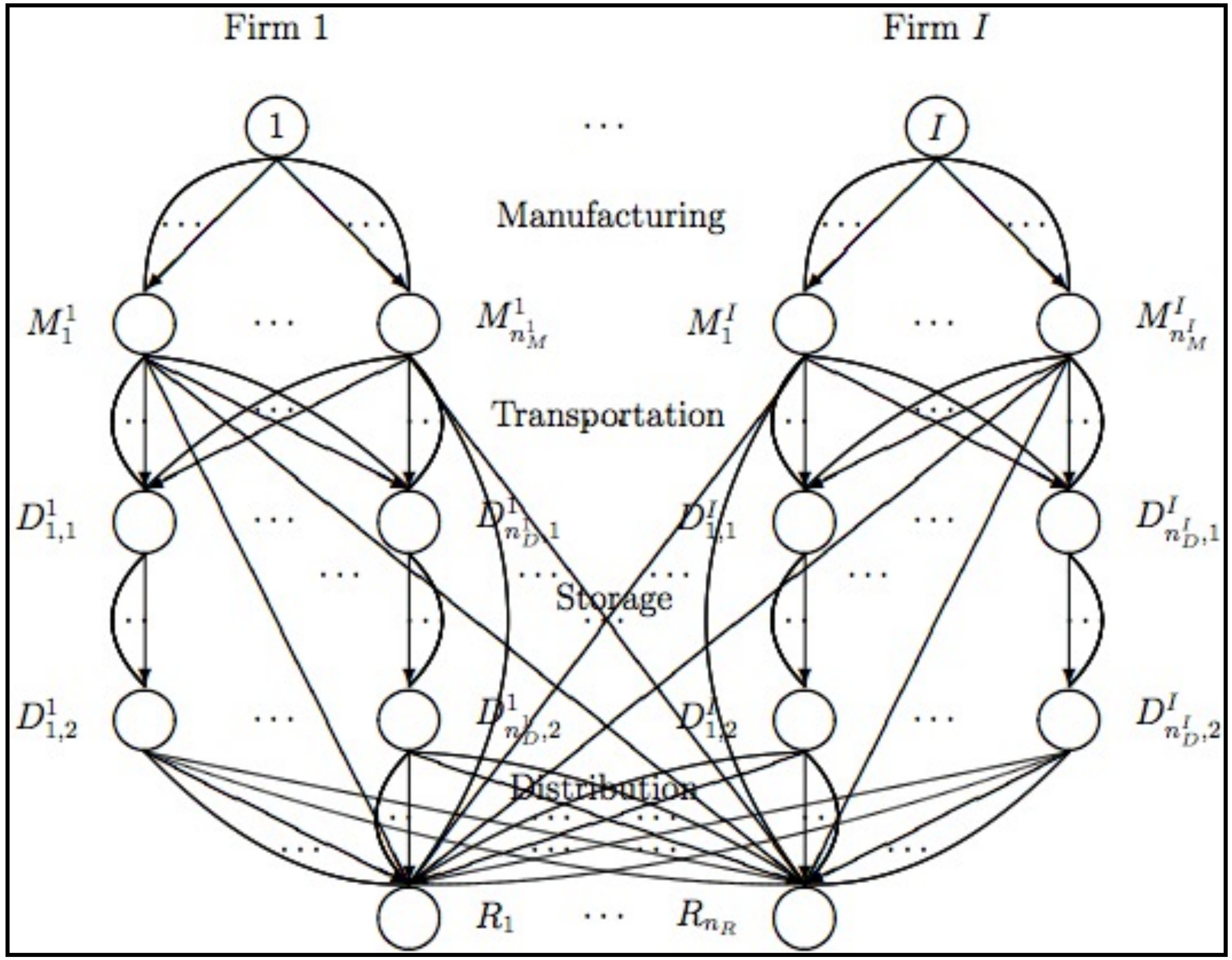
Experimental Results

(Gemp and Mahadevan, AAAI Workshop on Computational Substainability, 2015)



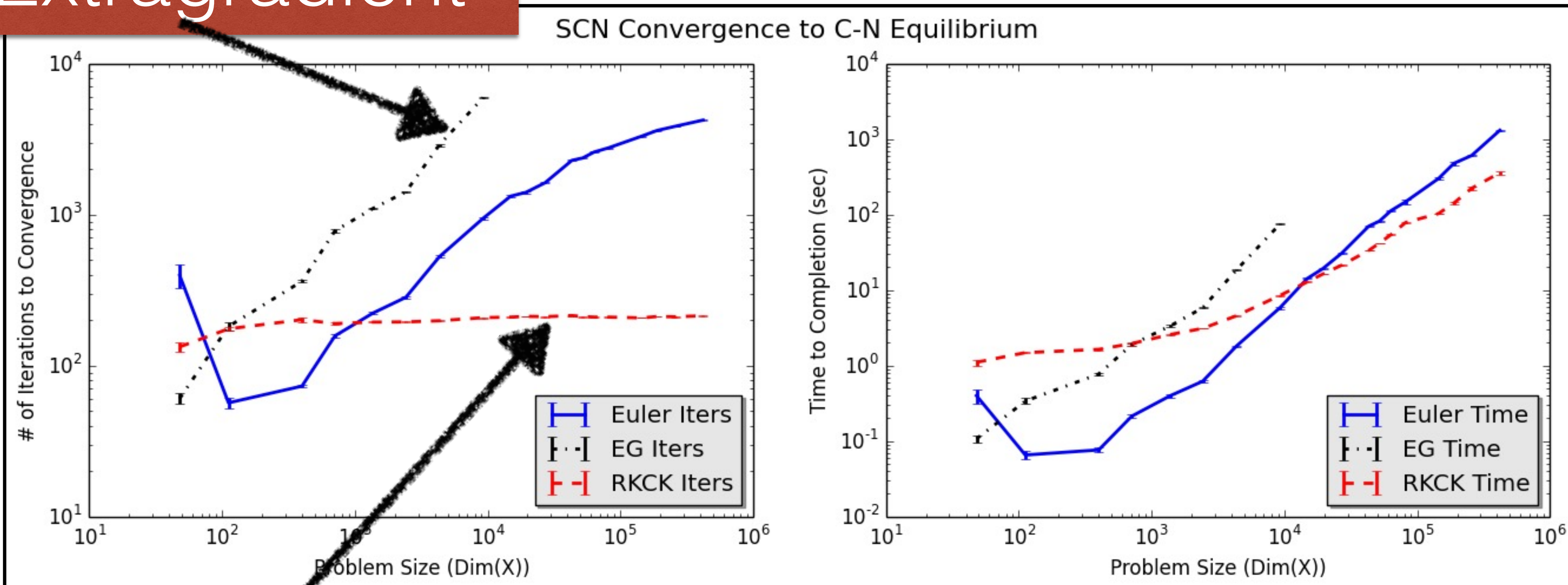
Sustainable Supply Chain

Nagurney et al.

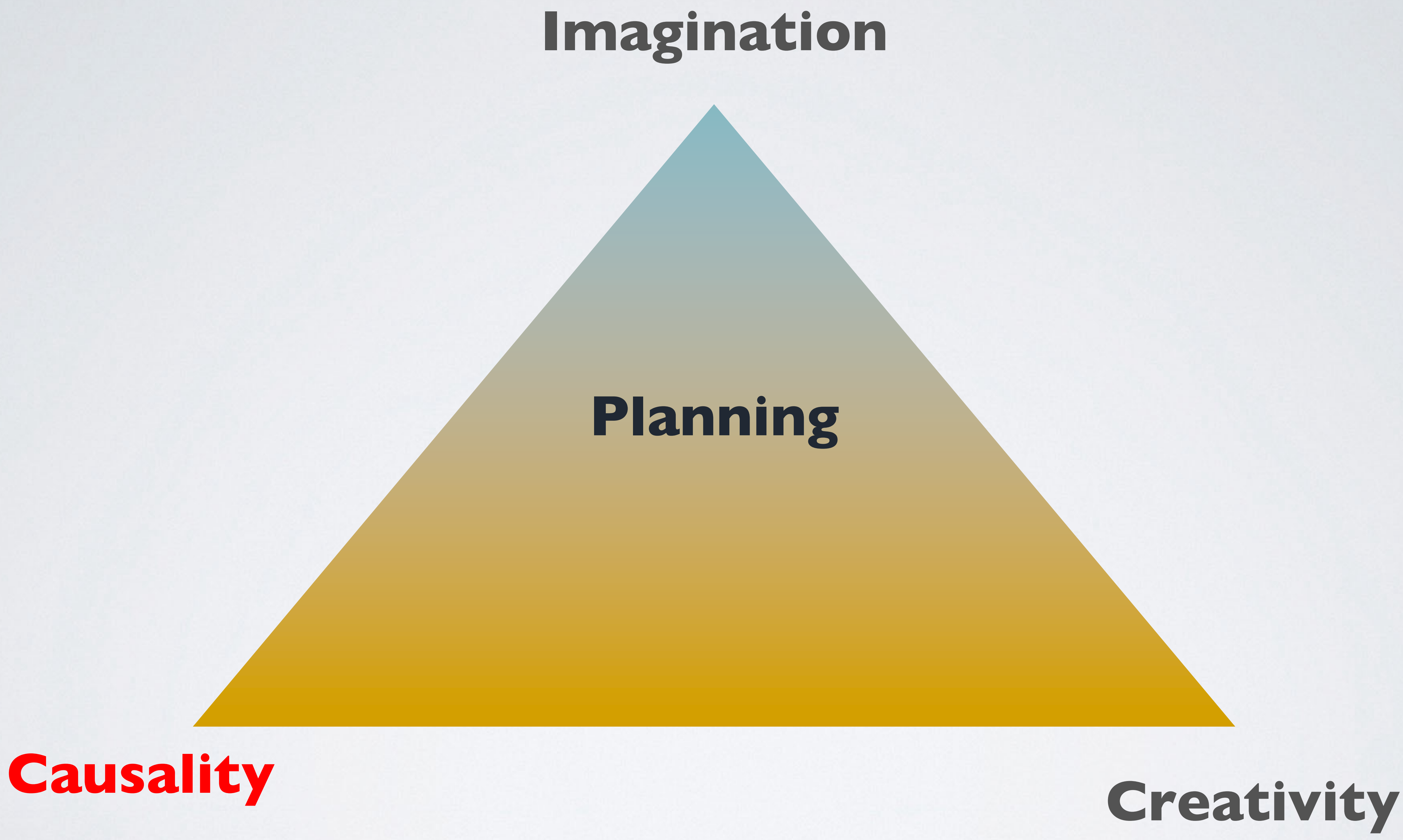


Results on Sustainable Supply Chain VI Problem

Extragradient



Our new algorithm


Imagination
Planning
Causality
Creativity

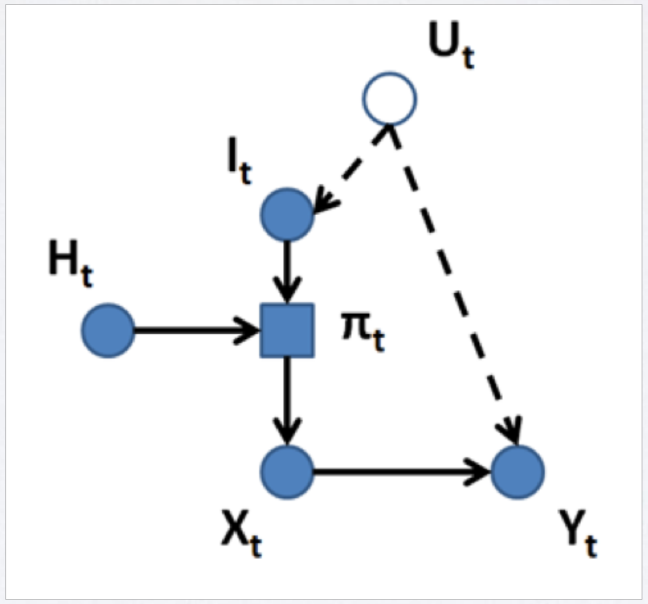
COUNTERFACTUALS

- $P(y_x \mid x', y')$ is the (counterfactual) probability that $Y=y$ if we artificially set $X = x$
- given that we actually observed Y to be y' when X naturally took on the value x'

IMAGINATION IN BANDITS

$$\underset{a}{\text{argmax}} \;\; E[Y_{X=a} = 1 | X = i]$$

Play the arm that maximizes the counterfactual estimate of the reward if action a is selected, when the agent's policy suggests choosing action i



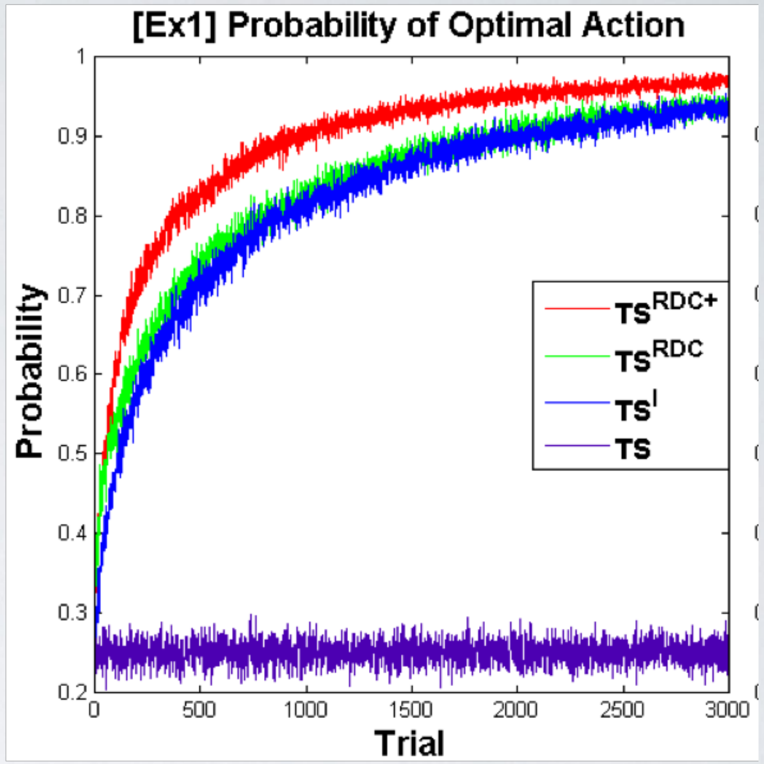
(Forney, Pearl and Bareinboim, 2017)

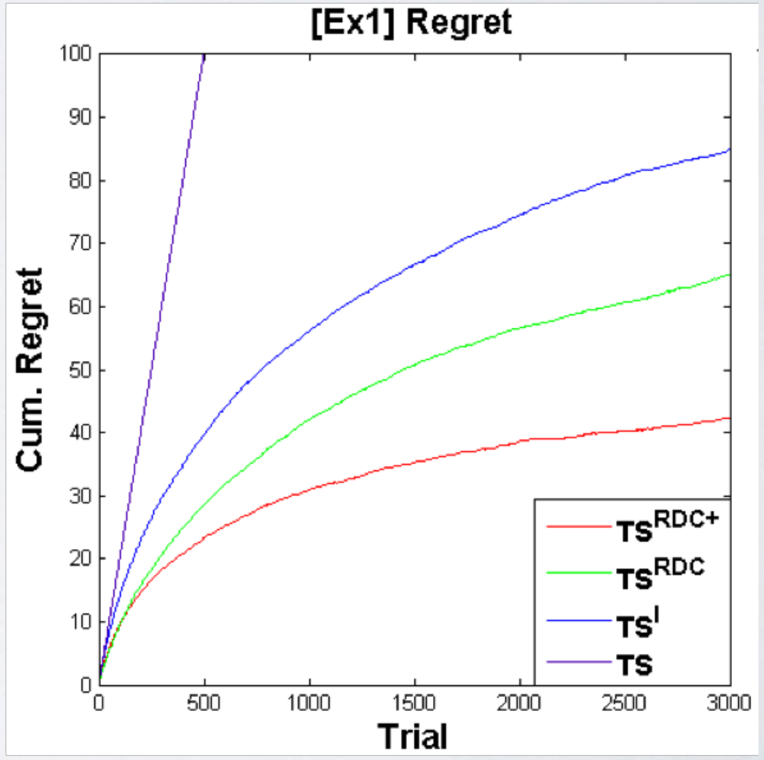
IMAGINATION WORKS BETTER

(a)	$D=0$		$D=1$	
$E[y_1 \mid X,B,D]$	$B=0$	$B=1$	$B=0$	$B=1$
$X=0$	*0.20	0.30	0.50	0.60
$X=1$	0.60	*0.20	0.30	0.50
$X=2$	0.50	0.60	*0.20	0.30
$X=3$	0.30	0.50	0.60	*0.20

(b)	$E[y_1 \mid X]$	$E[y_1 \mid do(X)]$
$X=0$	0.20	0.40
$X=1$	0.20	0.40
$X=2$	0.20	0.40
$X=3$	0.20	0.40

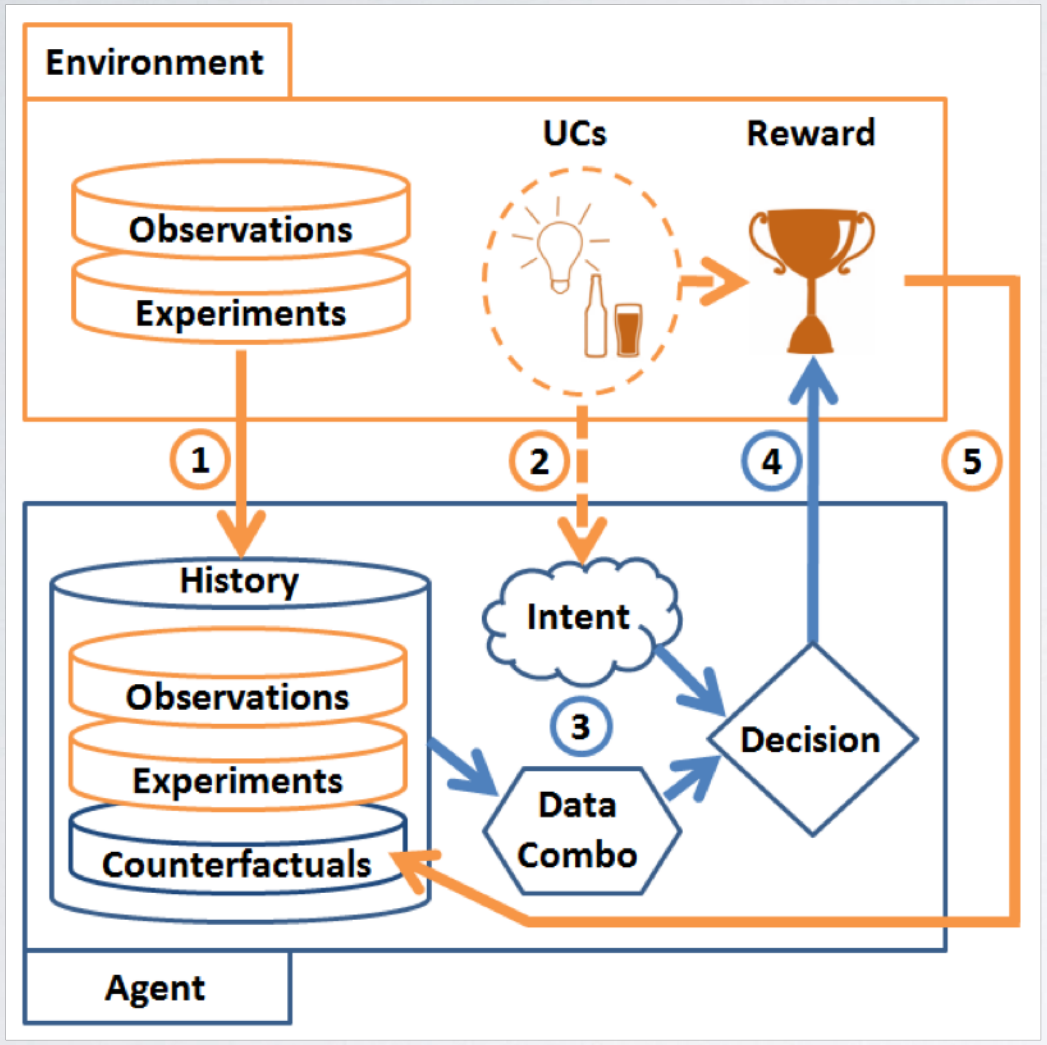
Table 1: (a) Payout rates decided by reactive slot machines as a function of arm choice X, sobriety D, and machine conspicuousness B. Players' natural arm choices ($f_x = B+2D$) under D, B are indicated by asterisks. (b) Payout rates according to the observational, $E[y_1|X]$, and experimental $E[y_1|do(X)]$, distributions, where $Y = y_1$ represents winning (shown in the table).





(Forney, Pearl and Bareinboim, 2017)

IMAGINATION AND IMITATION LEARNING



$$E[Y_x] = E[Y_x|x_1]P(x_1) + \ldots + E[Y_x|x_K]P(x_K)$$

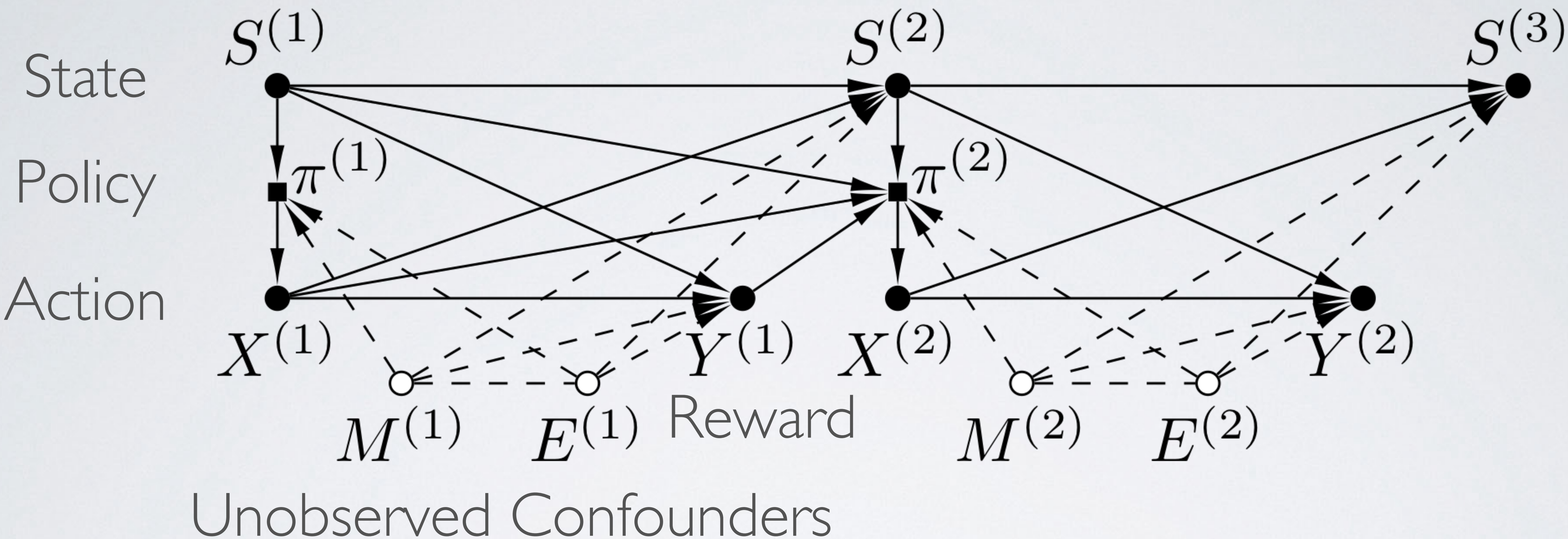
Counterfactual

Observational

Experimental

(Forney, Pearl and Bareinboim, 2017)

MDPS WITH UNOBSERVED CONFOUNDERS



(Zhang and Bareinboim, Arxiv, 2016)

IMAGINATION VALUES

Ternary function over states, hypothetical actions, and policy recommended actions

$$Q^{\pi}(s_t, a'_t, a_t) = E\left(R^t_{a_t} | s_t, a'_t\right) +$$

$$\gamma \sum_{s_{t+1} \in S} \sum_{a'_{t+1} \in A} P(s^{a_t}_{t+1}, (a')^{a_t}_{t+1} | s_t, a'_t) V^{\pi}(s_{t+1}, a'_{t+1})$$

SUMMARY

- **We explored several novel frontiers of planning**
  - **Causality, creativity and imagination**
- Planning is much more useful than plans, as it can be used to build long-term representations that transfer across tasks
- Planning in dual spaces using mirror descent leads to novel RL algorithms that have attractive properties
- Variational inequalities enable generalizing to problems without objective functions (traffic, blood distribution, Internet content distribution)

SUMMARY OF TUTORIAL

- GAN models are examples of equilibrium systems
- Gradient descent is not the optimal way to train equilibrium models
- We introduced variational inequalities as a framework for modeling equilibrium systems
- VI algorithms provide powerful tools for GANs, RL, and many other areas