ADVERSARIAL INTELLIGENCE IN NATURAL AND ARTIFICIAL SYSTEMS

IPAM WORKSHOP

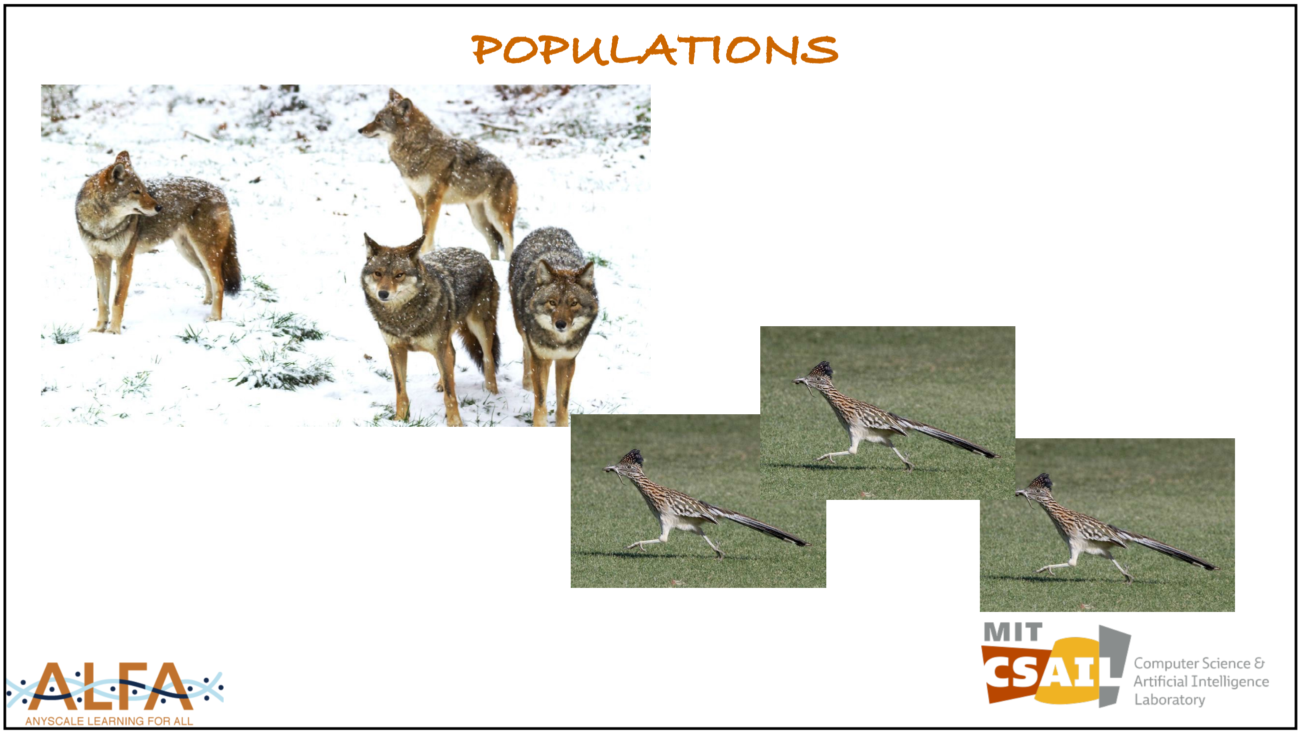
Fri, Nov 8, 2024

Una-May O'Reilly, unamay@csail.mit.edu

1

Adversarial Behavior

2



POPULATIONS

ALFA
ANYSCALE LEARNING FOR ALL
MIT CSAIL
Computer Science & Artificial Intelligence Laboratory

3

EVOLVED TO DEFEND

ALFA
ANYSCALE LEARNING FOR ALL
MIT CSAIL
Computer Science & Artificial Intelligence Laboratory

4

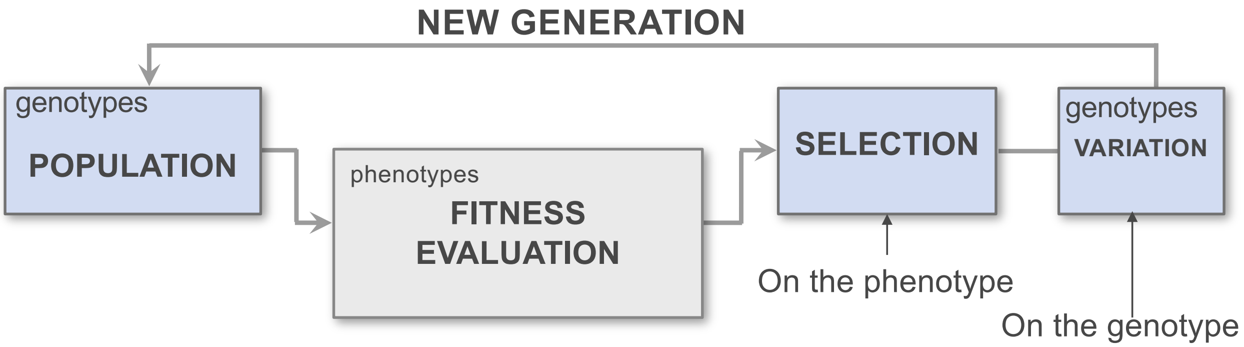
EVOLVED AS PREDATORS

ALFA
ANYSCALE LEARNING FOR ALL

MIT CSAIL
Computer Science & Artificial Intelligence Laboratory

5

EVOLUTIONARY ALGORITHM



NEW GENERATION

genotypes
POPULATION

phenotypes
FITNESS EVALUATION

SELECTION

genotypes
VARIATION

On the phenotype

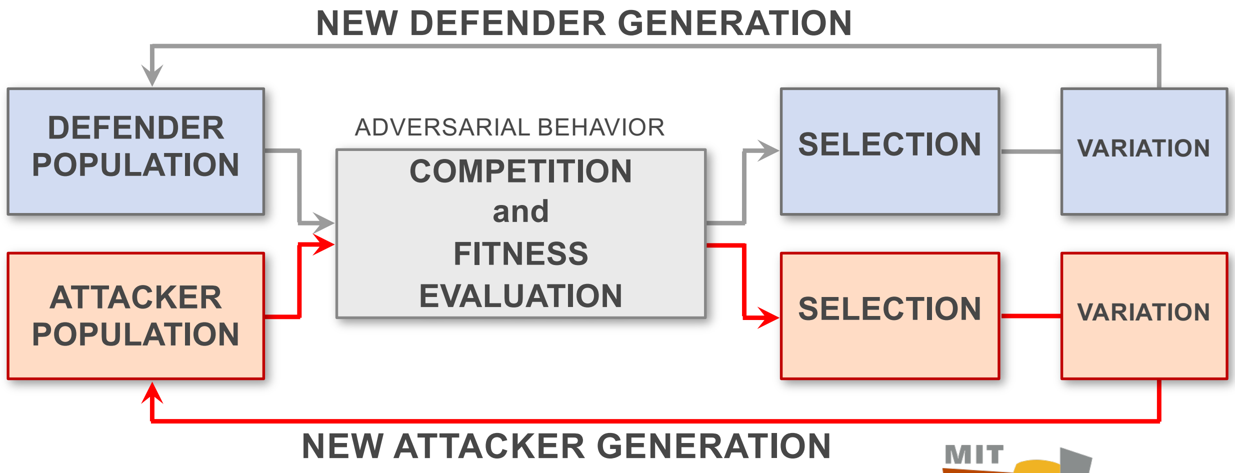
On the genotype

ALFA
ANYSCALE LEARNING FOR ALL

MIT CSAIL
Computer Science & Artificial Intelligence Laboratory

6

COMPETITIVE COEVOLUTIONARY ALGORITHM



ALFA ANYSCALE LEARNING FOR ALL

MIT CSAIL Computer Science & Artificial Intelligence Laboratory

7

HERE'S WHERE IT STARTS TO GET INTERESTING!

- 3 case studies
  - Each features
    - » an adversarial, attacker-defender (predator-prey) relationship
    - » an evolutionary arms race

  #1: Taxation

  #2: Generative adversarial networks

  #3: Cyber security agents*

ALFA ANYSCALE LEARNING FOR ALL

MIT CSAIL Computer Science & Artificial Intelligence Laboratory

8

CASE STUDY #1

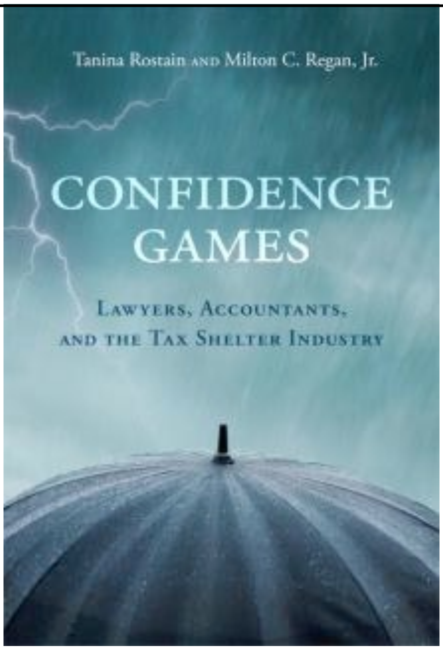
THE ARMS RACE BETWEEN TAX AVOIDANCE AND TAX AUDITING

SYSTEM: STEALTH
TEAM: ERIK HEMBERG, JACOB ROSEN, OSAMA BADAR, JEFF WARNER, SANITH WIJESINGHE
https://stealth.csail.mit.edu/publications.html

ANYSCALE LEARNING FOR ALL

Computer Science & Artificial Intelligence Laboratory

9

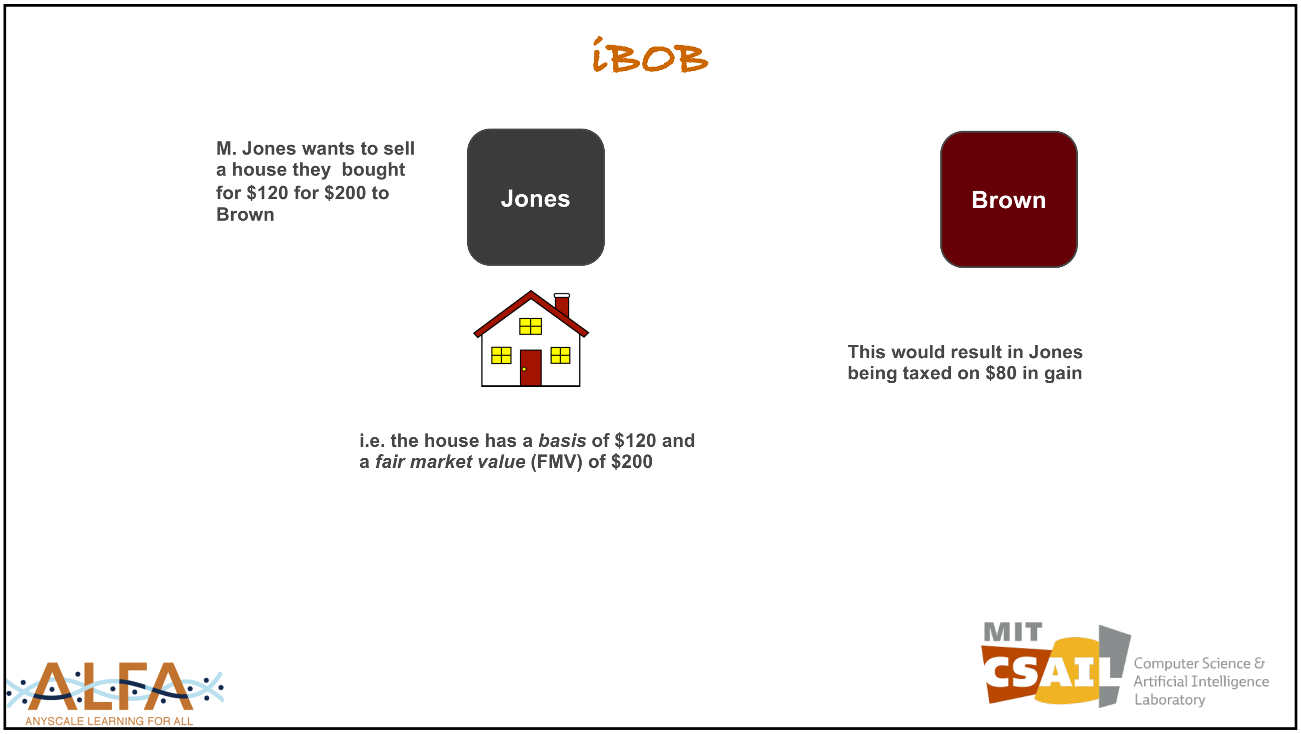

Tanina Rostain and Milton C. Regan, Jr.
CONFIDENCE GAMES
Lawyers, Accountants, and the Tax Shelter Industry

Video of iBOB and DAD explained available upon request

ANYSCALE LEARNING FOR ALL

Computer Science & Artificial Intelligence Laboratory

10



iBOB

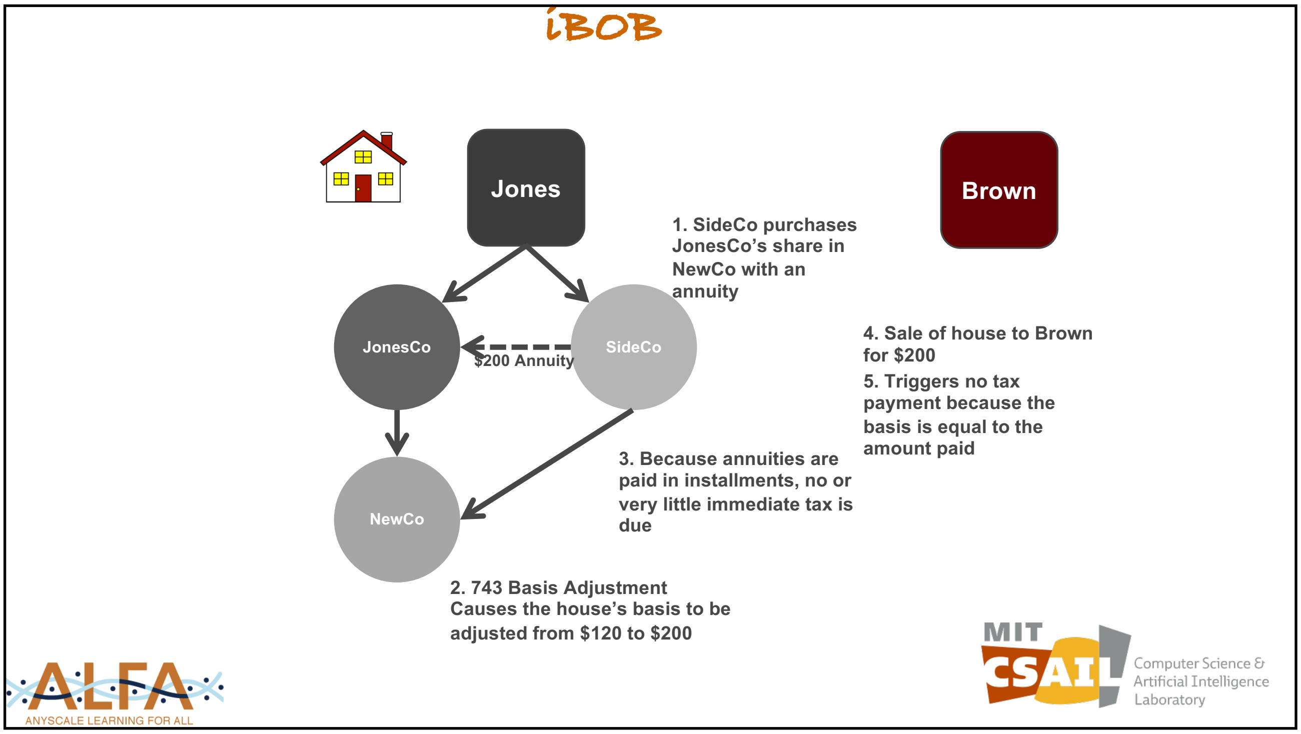
M. Jones wants to sell a house they bought for $120 for $200 to Brown

This would result in Jones being taxed on $80 in gain

i.e. the house has a *basis* of $120 and a *fair market value* (FMV) of $200

ALFA
ANYSCALE LEARNING FOR ALL
MIT CSAIL
Computer Science & Artificial Intelligence Laboratory

11



iBOB

Jones
Brown
JonesCo
SideCo
NewCo
$200 Annuity

1. SideCo purchases JonesCo's share in NewCo with an annuity

4. Sale of house to Brown for $200

5. Triggers no tax payment because the basis is equal to the amount paid

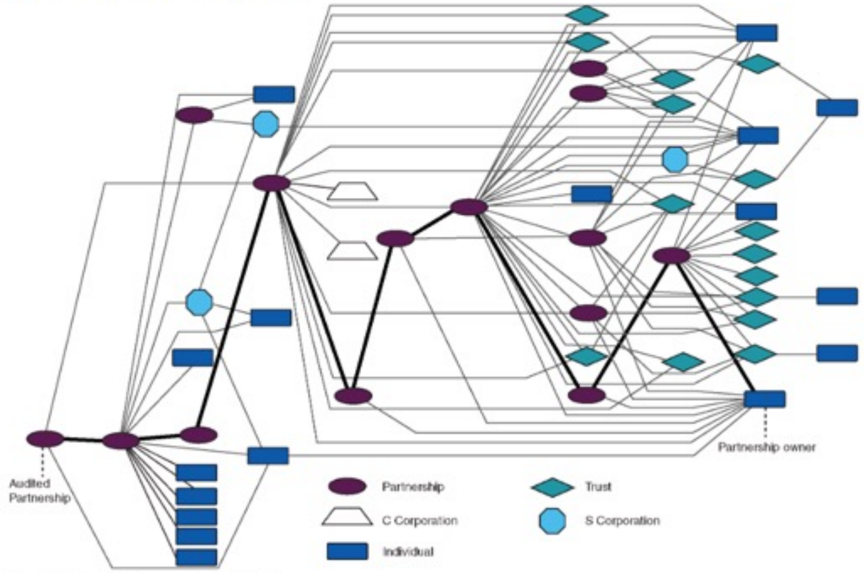
3. Because annuities are paid in installments, no or very little immediate tax is due

2. 743 Basis Adjustment Causes the house's basis to be adjusted from $120 to $200

ALFA
ANYSCALE LEARNING FOR ALL
MIT CSAIL
Computer Science & Artificial Intelligence Laboratory

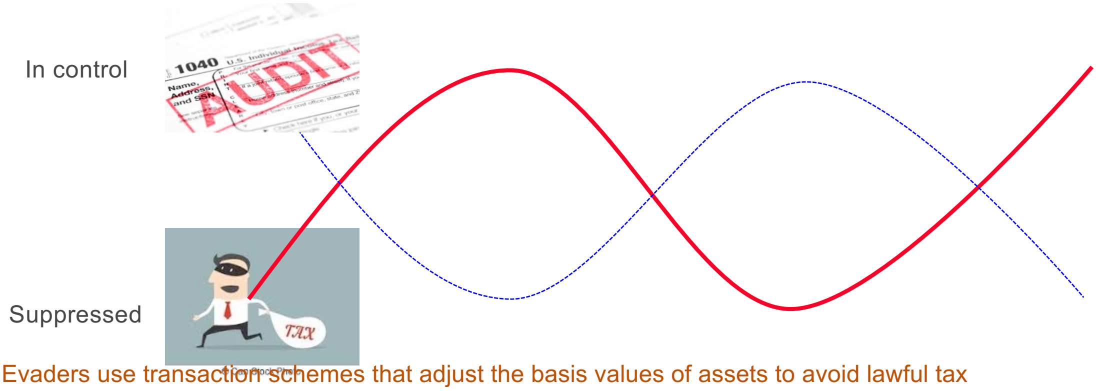
12

Figure 7: Example of Partnership Structure



Source: GAO analysis of IRS documentation. I GAO-14-732

COEVOLUTION OF TAX NON-COMPLIANCE AND AUDITING



- Evaders use transaction schemes that adjust the basis values of assets to avoid lawful tax payments
- Auditors shifts audit resources to some subset of transaction observables to flag them
- Evaders shift schemes to sneak by where there's no attention
- Adversarial co-evolution oscillation of successful auditing or non-compliance
- $91B tax gap from PARTNERSHIP activities, 2010s

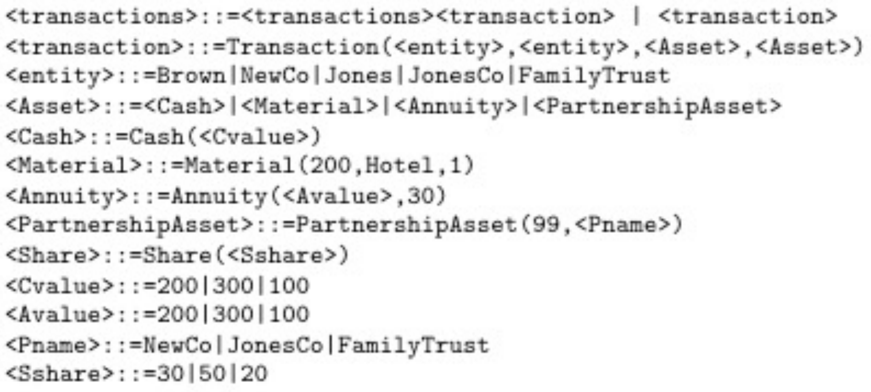
HOW DO WE REPLICATE THE COEVOLUTION of TAX NON-COMPLIANCE and AUDITING?

https://www.redbubble.com/i/poster/Tax-fraud-is-cool-by-ValDIFF

How do we represent both compliant and non-compliant transactions?

16

GRAMMAR OF EVADER'S PARTNERSHIP TRANSACTIONS

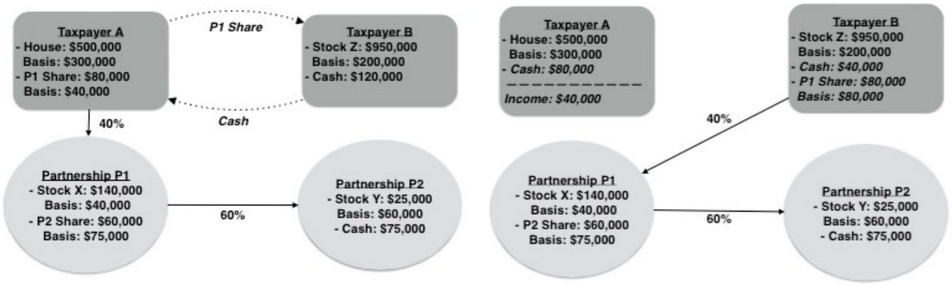


```
<transactions>::=<transactions><transaction> | <transaction>
<transaction>::=Transaction(<entity>,<entity>,<Asset>,<Asset>)
<entity>::=Brown|NewCo|Jones|JonesCo|FamilyTrust
<Asset>::=<Cash>|<Material>|<Annuity>|<PartnershipAsset>
<Cash>::=Cash(<Cvalue>)
<Material>::=Material(200,Hotel,1)
<Annuity>::=Annuity(<Avalue>,30)
<PartnershipAsset>::=PartnershipAsset(99,<Pname>)
<Share>::=Share(<Sshare>)
<Cvalue>::=200|300|100
<Avalue>::=200|300|100
<Pname>::=NewCo|JonesCo|FamilyTrust
<Sshare>::=30|50|20
```

GENOTYPE: Vector of integers
Translation: integers and start symbol –
PHENOTYPE: Sentence = transaction sequence
GRAMMATICAL EVOLUTION

18

PHENOTYPE: TRANSACTION SEQUENCE THAT TRANSFERS OWNERSHIP AND RESULTS IN ASSET BASIS ADJUSTMENT



(a) Initial state and transaction

(b) Network state after transaction

Ownership graph and record of basis adjustment after asset sale/transfer

HOW DO WE REPLICATE THE COEVOLUTION of TAX NON-COMPLIANCE and AUDITING?

https://www.redbubble.com/i/poster/Tax-fraud-is-cool-by-ValDIFF

How do we represent both compliant and non-compliant transactions?

How do we represent tax regulations?

TAXATION: REPRESENTING THE TAX CODE OF BASIS ADJUSTMENT

U.S. Code § 754 - Manner of electing optional adjustment to basis of partnership property

If a partnership files an election, in accordance with regulations prescribed by the Secretary, the basis of partnership property shall be adjusted, in the case of a distribution of property, in the manner provided in section 734 and, in the case of a transfer of a partnership interest, in the manner provided in section 743. Such an election shall apply with respect to all distributions of property by the partnership and to all transfers of interests in the partnership during the taxable year with respect to which such election was filed and all subsequent taxable years. Such election may be revoked by the partnership, subject to such limitations as may be provided by regulations prescribed by the Secretary.

22

BASIS ADJUSTMENT AND TAXATION

TAX AUTHORITY: from transaction sequence, calculates basis adjustments and tax liability

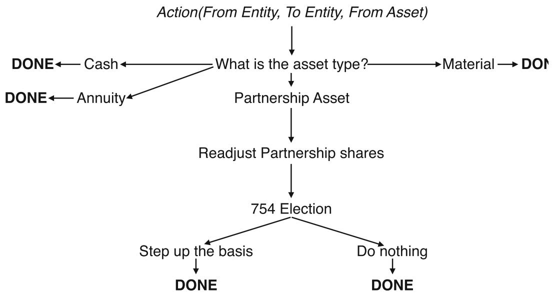


Fig. 4 A decision tree rule to evaluate asset basis changes

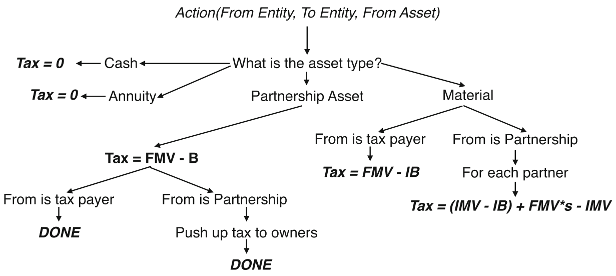
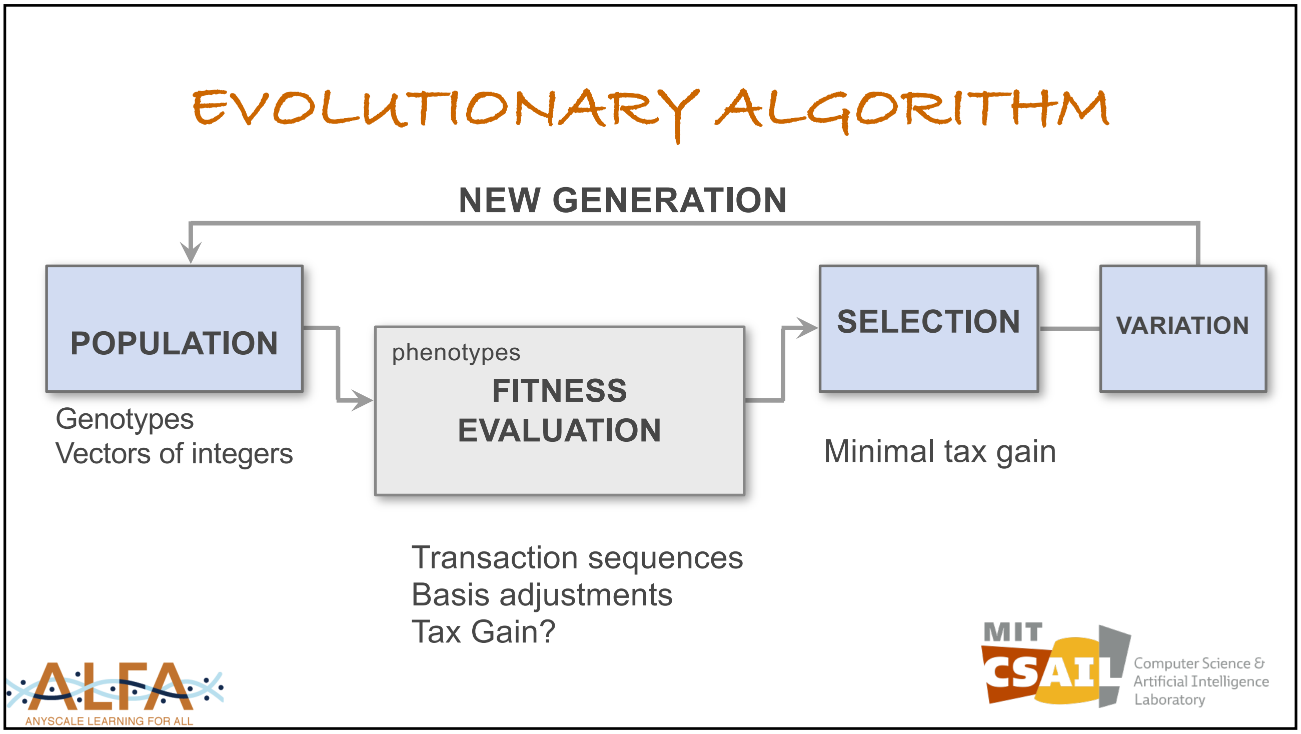


Fig. 5 A decision tree rule that shows the tax calculation on an asset transfer

DECISION TREE RULES

23



EVOLUTIONARY ALGORITHM

NEW GENERATION

POPULATION

Genotypes
Vectors of integers

phenotypes

FITNESS EVALUATION

SELECTION

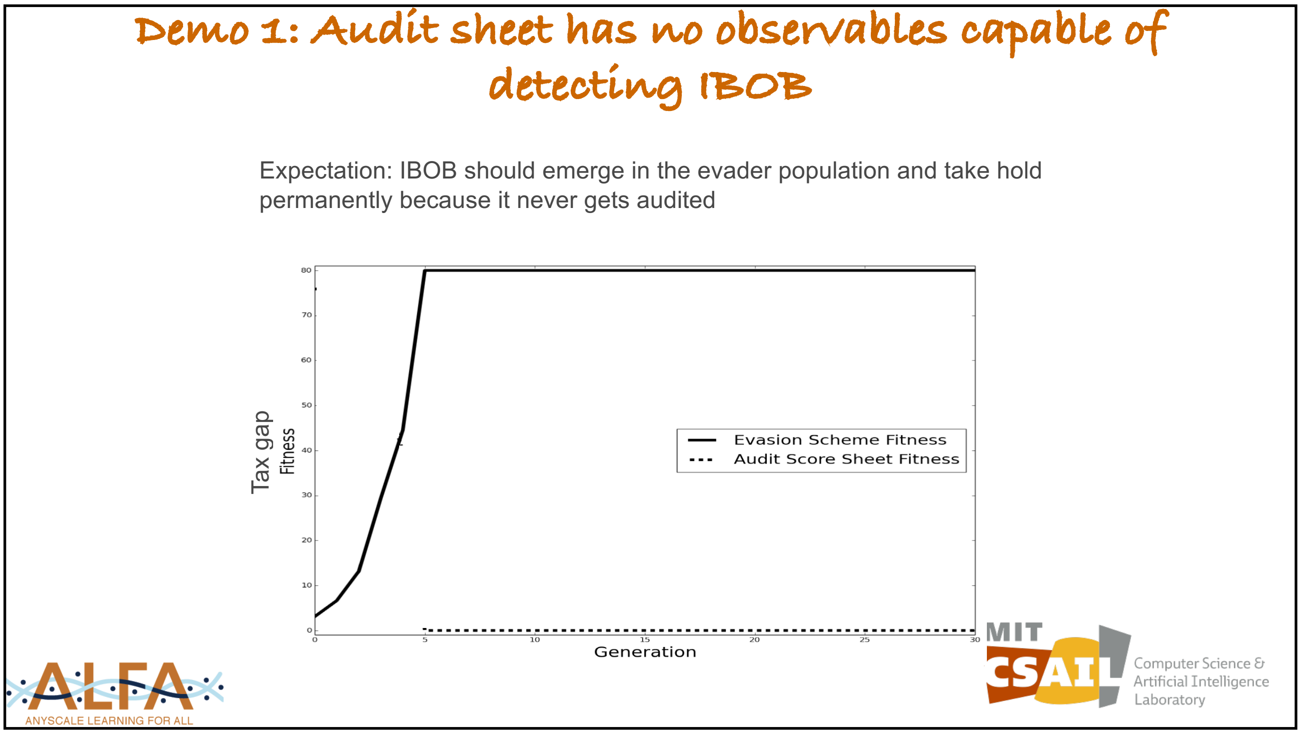
VARIATION

Minimal tax gain

Transaction sequences
Basis adjustments
Tax Gain?

ALFA
ANYSCALE LEARNING FOR ALL
MIT CSAIL
Computer Science & Artificial Intelligence Laboratory

24



Demo 1: Audit sheet has no observables capable of detecting IBOB

Expectation: IBOB should emerge in the evader population and take hold permanently because it never gets audited

26

HOW DO WE REPLICATE THE COEVOLUTION of TAX NON-COMPLIANCE and AUDITING?

https://www.redbubble.com/i/poster/Tax-fraud-is-cool-by-ValDIFF

How do we represent both compliant and non-compliant transactions?

How do we represent auditing?

ALFA
ANYSCALE LEARNING FOR ALL

MIT CSAIL
Computer Science & Artificial Intelligence Laboratory

27

AUDITOR'S OBSERVABLES

The basis of partnership property shall not be adjusted as the result of (1) a transfer of an interest in a partnership by sale or exchange or on the death of a partner unless (2) the election provided by § 754 (relating to optional adjustment to basis of partnership property) is in effect with respect to such partnership or (3) unless the partnership has a substantial built-in loss immediately after such transfer.

743 Alteration (2004)

Observables

1. The sale of a partnership interest in exchange for a taxable asset.
2. The partnership whose shares are being transferred has not made a § 754 election.
3. The seller's basis in respect to the non-cash assets owned by the partnership exceeds their FMV by more than $250.000

REPRESENTATION IS WEIGHTS ON OBSERVABLES

ALFA
ANYSCALE LEARNING FOR ALL

MIT CSAIL
Computer Science & Artificial Intelligence Laboratory

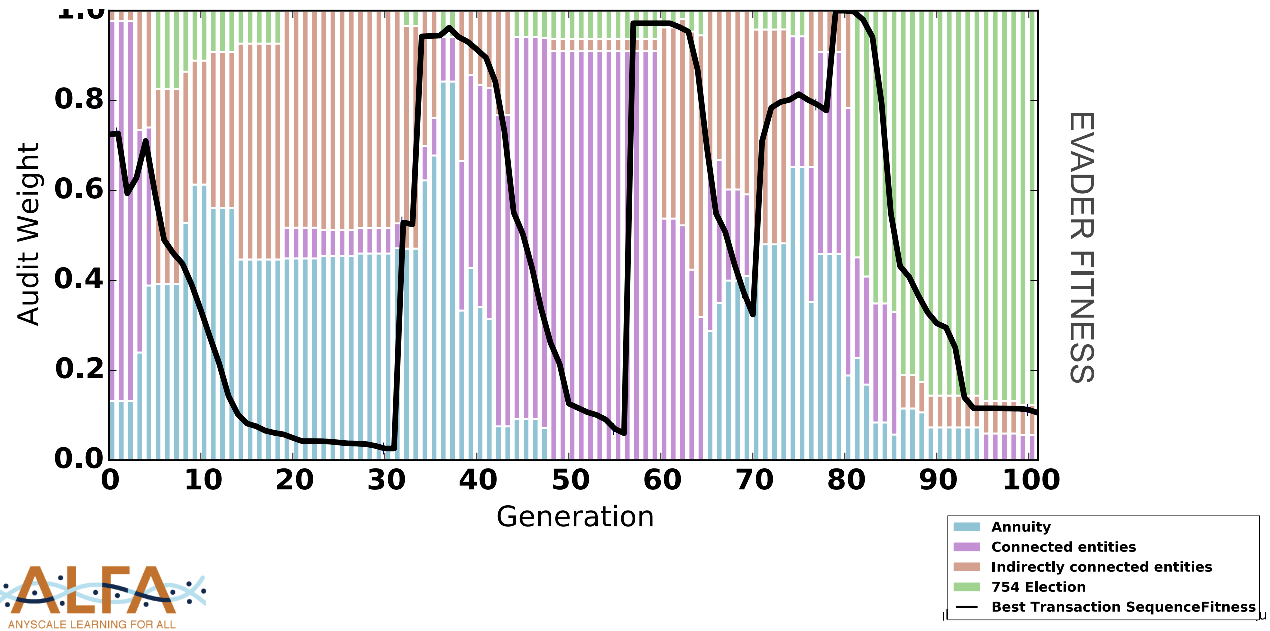
28

FITNESS FUNCTION

- An auditor is fit if they assign a high audit score to a highly non-compliant evader.
- An evader is fit if they receive a low audit score but are highly non-compliant.
- Corollaries
- An auditor is fit if they assign a low audit score to a compliant (non) evader
- An evader is not very fit if they are compliant.

30

BEST AUDITOR VS BEST EVADER



31

CASE STUDY #2

SPATIAL COEVOLUTION TO IMPROVE THE TRAINING ROBUSTNESS
of
GENERATIVE ADVERSARIAL NETWORKS
(GANs)

SYSTEM: LIPIZZANER
TEAM: Many, including Jamal Toutouh, Erik Hemberg, Adbullah Al Dujaili, and many, many students

ALFA ANYSCALE LEARNING FOR ALL

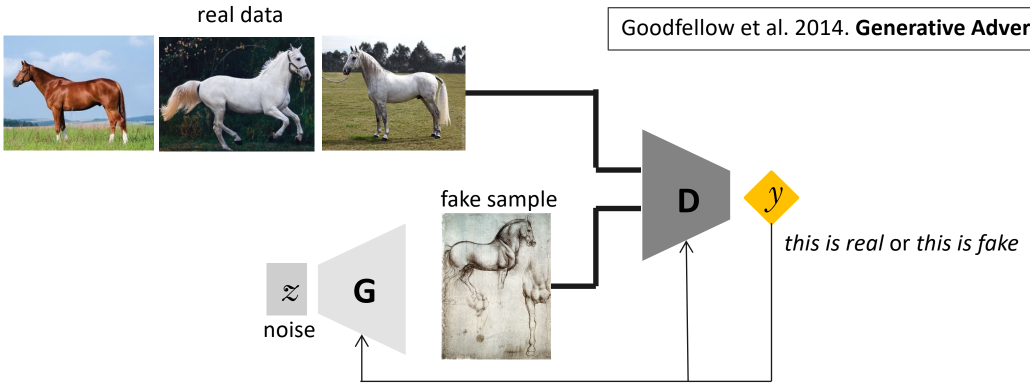
MIT CSAIL Computer Science & Artificial Intelligence Laboratory

32

AND... THERE'S A CS/ML ADVERSARIAL EXAMPLE!

Generative Adversarial Networks: Construct a generative model by exploiting an arms race between two neural networks, a generator and a discriminator

Goodfellow et al. 2014. **Generative Adversarial Nets**



$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

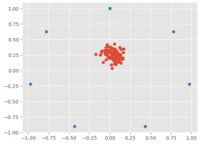
ALFA ANYSCALE LEARNING FOR ALL

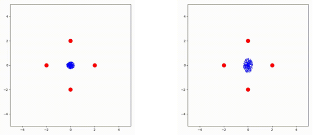
CSAIL Computer Science & Artificial Intelligence Laboratory

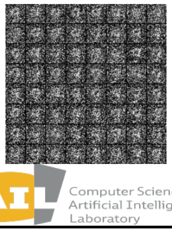
33

GAN TRAINING PATHOLOGIES

- Non-convergence: the model parameters oscillate, destabilize and never converge
- Mode collapse: the generator collapses which produces limited varieties of samples
- Diminished gradient: the discriminator gets too successful that the generator gradient vanishes and learns nothing





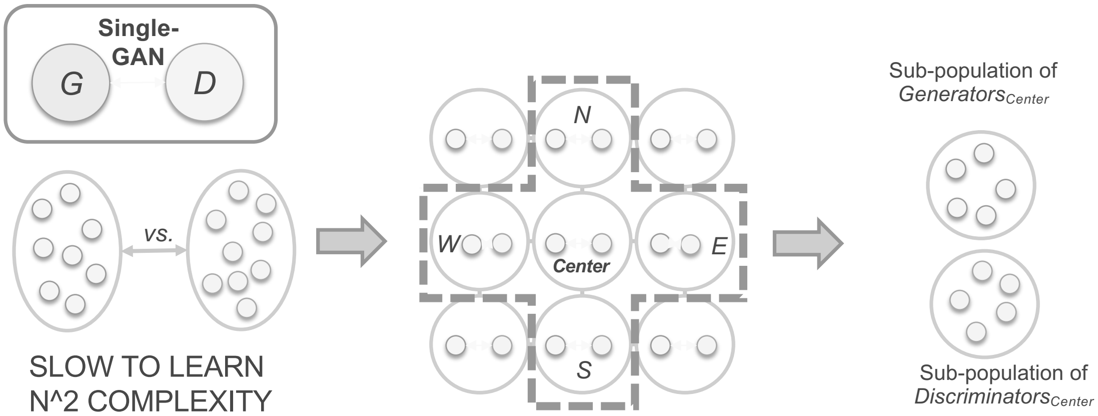


ALFA ANYSCALE LEARNING FOR ALL

MIT CSAIL Computer Science & Artificial Intelligence Laboratory

34

LIPIZZANER
SPATIAL COEVOLUTIONARY GAN TRAINING



Single-GAN

G

D

vs.

SLOW TO LEARN
N^2 COMPLEXITY

N

W

Center

E

S

Sub-population of *Generators*$_{Center}$

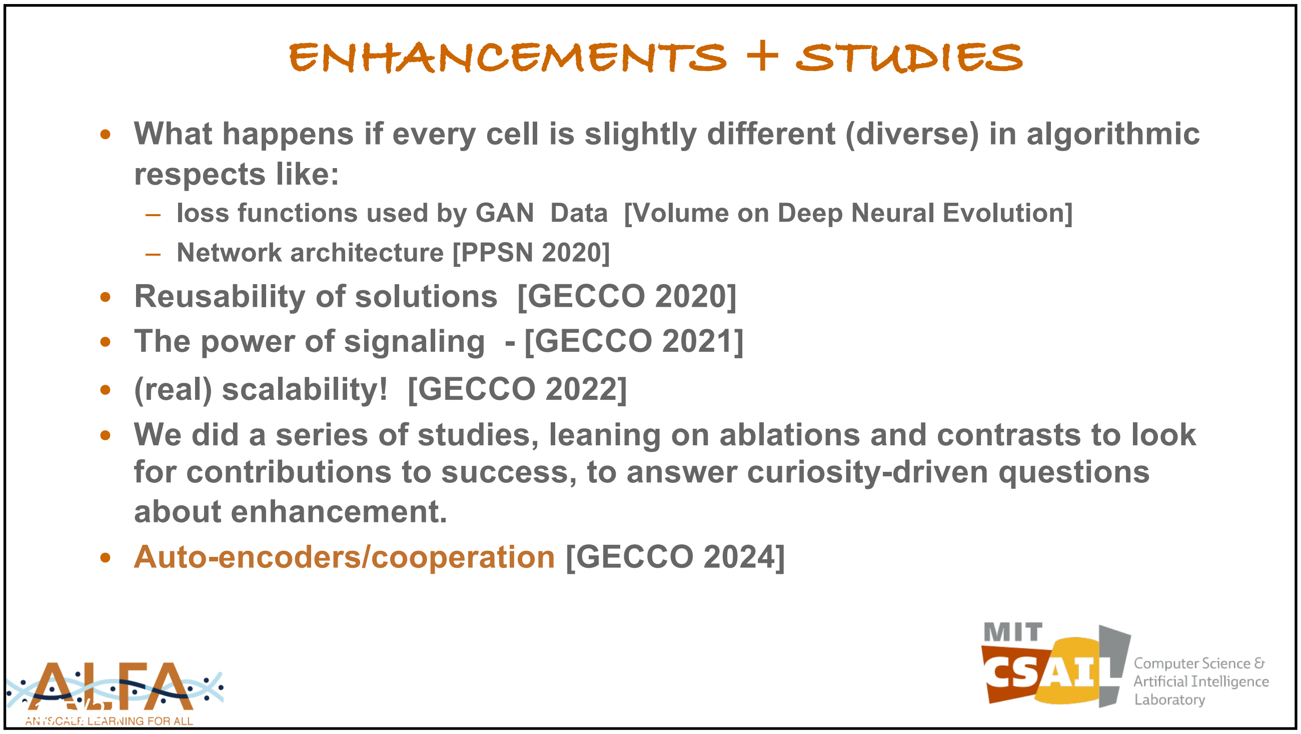
Sub-population of *Discriminators*$_{Center}$

COEVOLUTIONARY GAN TRAININGs

ALFA ANYSCALE LEARNING FOR ALL

MIT CSAIL Computer Science & Artificial Intelligence Laboratory

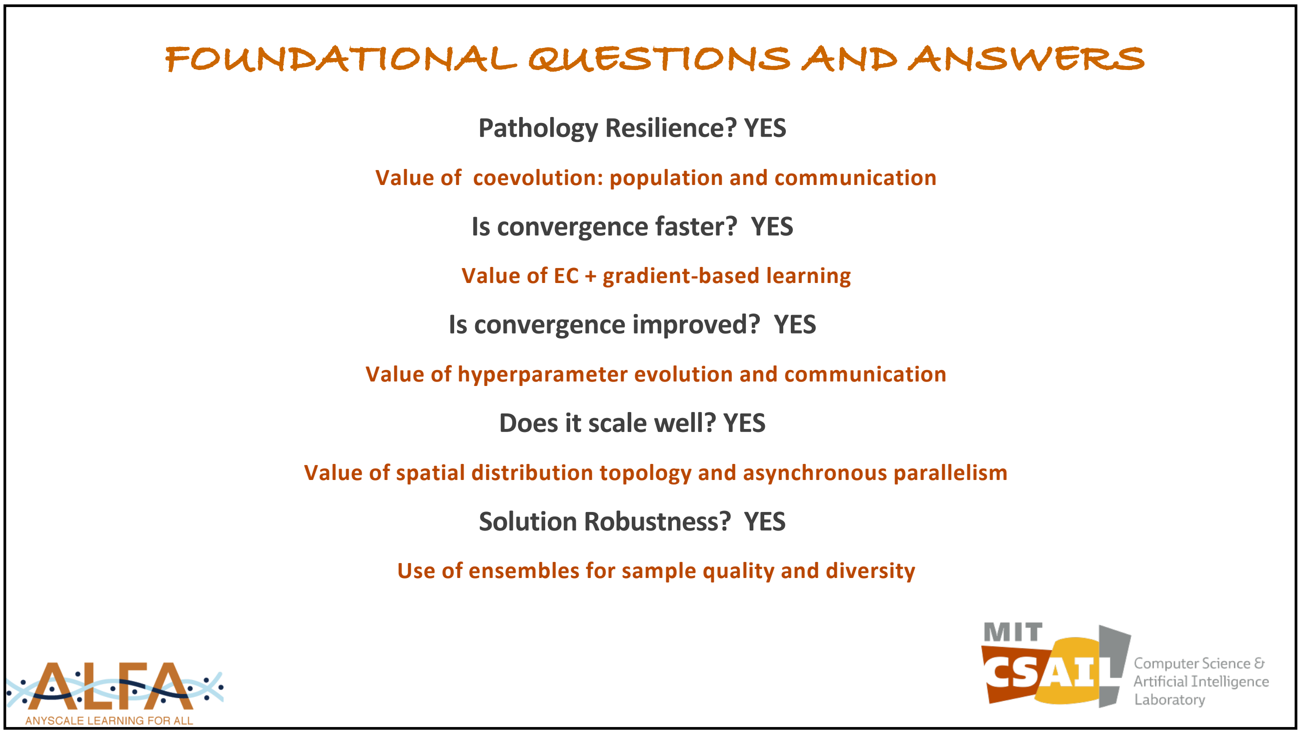
35



ENHANCEMENTS + STUDIES

- What happens if every cell is slightly different (diverse) in algorithmic respects like:
  - loss functions used by GAN Data [Volume on Deep Neural Evolution]
  - Network architecture [PPSN 2020]
- Reusability of solutions [GECCO 2020]
- The power of signaling - [GECCO 2021]
- (real) scalability! [GECCO 2022]
- We did a series of studies, leaning on ablations and contrasts to look for contributions to success, to answer curiosity-driven questions about enhancement.
- Auto-encoders/cooperation [GECCO 2024]

36



FOUNDATIONAL QUESTIONS AND ANSWERS

Pathology Resilience? YES

Value of coevolution: population and communication

Is convergence faster? YES

Value of EC + gradient-based learning

Is convergence improved? YES

Value of hyperparameter evolution and communication

Does it scale well? YES

Value of spatial distribution topology and asynchronous parallelism

Solution Robustness? YES

Use of ensembles for sample quality and diversity

37

CASE STUDY #3

THE COEVOLUTIONARY ARMS RACE
BETWEEN
CYBER NETWORK ATTACKS AND DEFENSES

SYSTEMS: RIVALS, RED-BLUE-AGENTS
ENVIRONMENT: CYBER NETWORKS

PREDATOR/ATTACKER: DDOS, RECONNAISSANCE, FULL CAMPAIGNS
PREY/DEFENDER: SELF-REPAIR, DECEPTION, Multi—agent Defenses

TEAM: Erik Hemberg and many, many students

ALFA
ANYSCALE LEARNING FOR ALL

MIT CSAIL
Computer Science & Artificial Intelligence Laboratory

38

CYBER THREATS AND DEFENSES

ALFA
ANYSCALE LEARNING FOR ALL

MIT CSAIL
Computer Science & Artificial Intelligence Laboratory

39

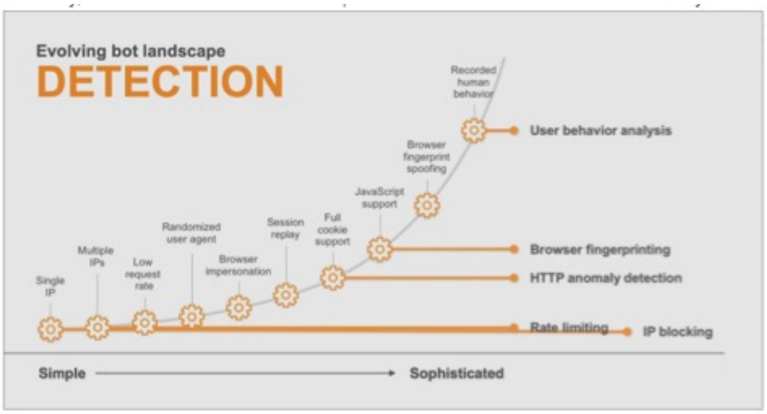


Figure 5: Common evasion tactics mapped to their logical defense mechanism, scaled by level of difficulty for the adversarial bot

Akamai SOTI Report, Jan 2019.

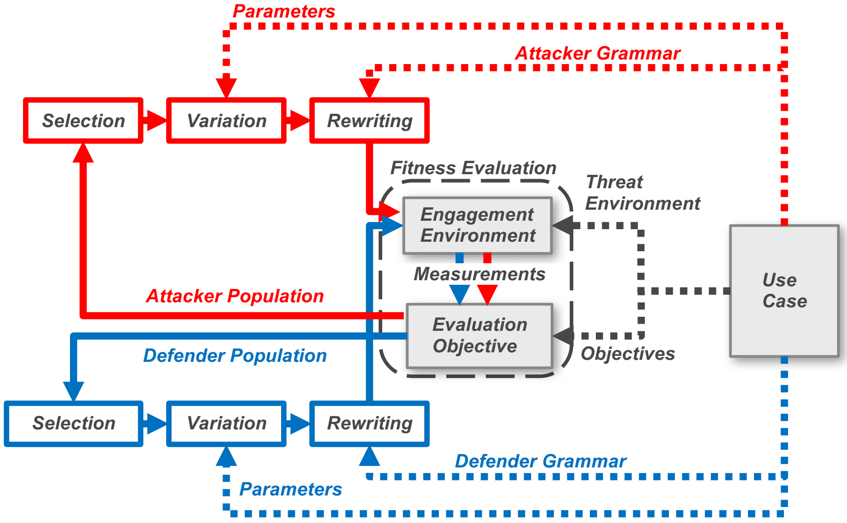
ALFA ANYSCALE LEARNING FOR ALL

Cyber Network Security Arms Race

MIT CSAIL Computer Science & Artificial Intelligence Laboratory

40

RIVALS



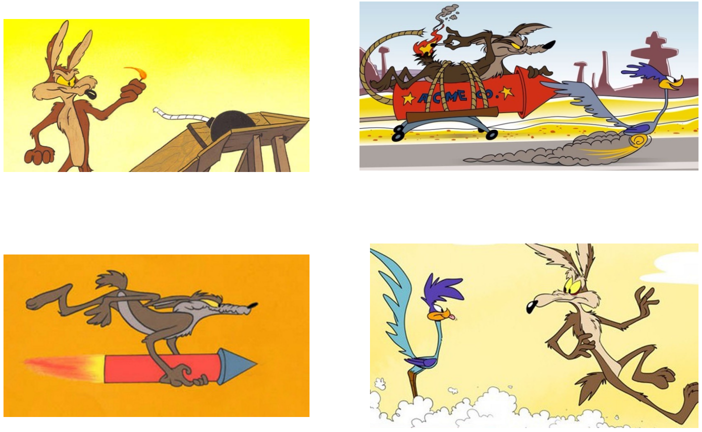
Predator/Attacker	Prey/Defender
D-Denial of Service	P2P Configs
Reconnaissance	Deception

ALFA ANYSCALE LEARNING FOR ALL

MIT CSAIL Computer Science & Artificial Intelligence Laboratory

41

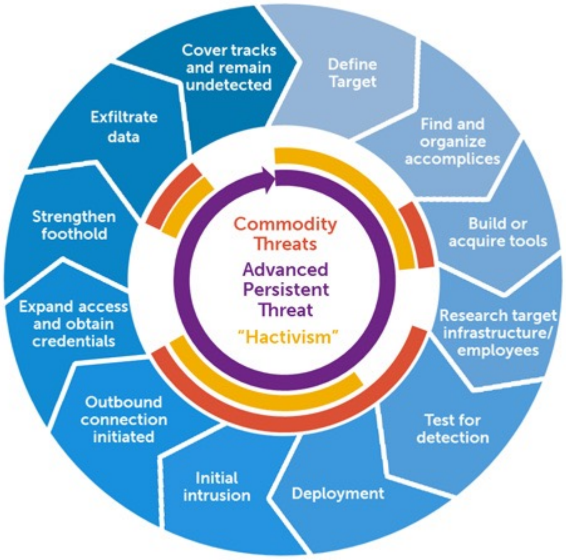
AND, WHAT'S ELSE IS REQUIRED FOR ARTIFICIAL ADVERSARIAL INTELLIGENCE?



Adversarial Intelligence and Learning

42

ADVANCED PERSISTENT THREATS

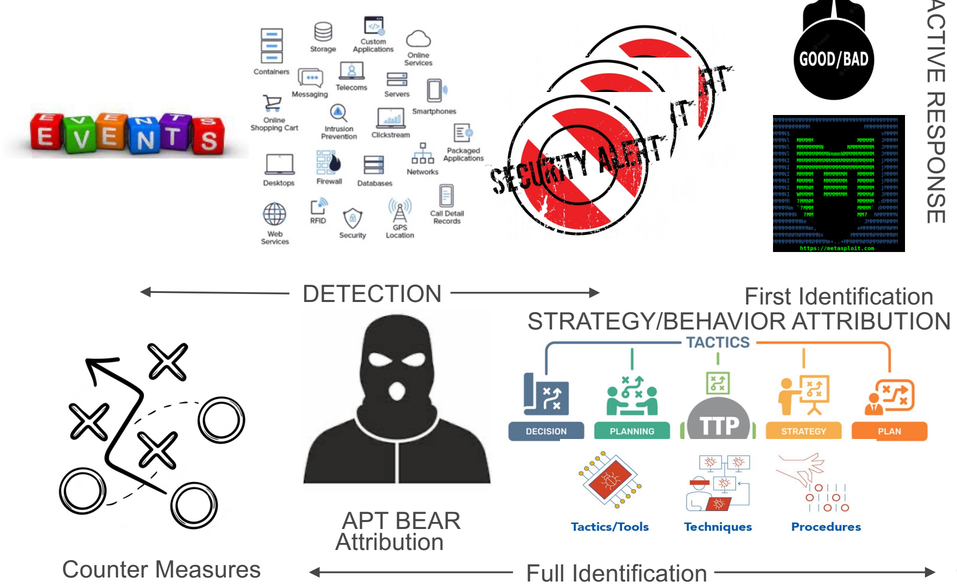

Define Target
Find and organize accomplices
Build or acquire tools
Research target infrastructure/ employees
Test for detection
Deployment
Initial intrusion
Outbound connection initiated
Expand access and obtain credentials
Strengthen foothold
Exfiltrate data
Cover tracks and remain undetected
Commodity Threats
Advanced Persistent Threat
"Hactivism"

- Planning
- Consulting text-based campaign information
- Choosing/using scripts and exploits
- Hiding, deception
- Sequential Decision making

SOURCE: Wikimedia Commons

43

CYBER DEFENCE



DETECTION

First Identification

STRATEGY/BEHAVIOR ATTRIBUTION

APT BEAR Attribution

Counter Measures

Full Identification

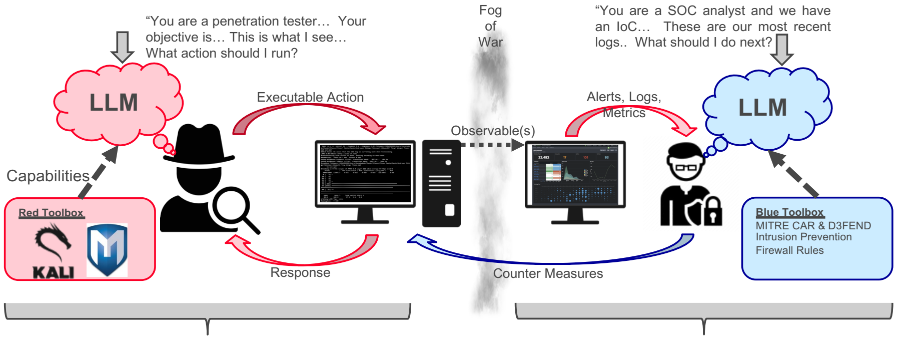
- Reactive
- Consulting text-based defensive information
- Choosing/using commands, scripts, tools
- Sequential decision making

ALFA ANYSCALE LEARNING FOR ALL

MIT CSAIL Computer Science & Artificial Intelligence Laboratory

44

LLM-SUPPORTED CYBER AGENTS



"You are a penetration tester… Your objective is… This is what I see… What action should I run?

"You are a SOC analyst and we have an IoC… These are our most recent logs.. What should I do next?

PEN-TESTER/RED TEAMER

DEFENDER

LLMs Killed the Script Kiddie: How Agents Supported by Large Language Models Change the Landscape of Network Threat Testing. https://arxiv.org/abs/2310.06936

Demo Video: https://youtu.be/OU0tiYmAQHo

ALFA ANYSCALE LEARNING FOR ALL

MIT CSAIL Computer Science & Artificial Intelligence Laboratory

45

AND, WHAT'S ELSE IS REQUIRED FOR ARTIFICIAL ADVERSARIAL INTELLIGENCE?

- symbolic reasoning
  - planning, knowledge consultation
    - Classical AI meets LLMs
  - Sequential decision making:
    - Plan-Act-Report, State machines
- agent learning:
  - hybridizing LLM and evolutionary algorithms and reinforcement learning

Home > Genetic Programming and Evolvable Machines > Article

g code with a large language model

hed: 12 September 2024

mber 21, (2024) Cite this article

Using Large Language Models for Evolutionary Search

~2022

ANYSCALE LEARNING FOR ALL

Artificial Intelligence Laboratory

46