The Prisoner's Dilemma

The Prisoner's Dilemma

- Idealized game-theoretic model of **cooperation**
- Formulated in 1950 by Melvin Dresher and Merrill Flood at RAND Corporation
- Implications for social science, political science, economics
- Genetic algorithms can be applied to the Prisoner's Dilemma to study the **evolution of cooperation**

The Prisoner's Dilemma

- **2 prisoners** (Alice and Bob) held in separate cells
- Neither one can communicate with the other
- Each prisoner has to make a **choice**:
  - Maintain innocence and don't talk (**cooperate**)
  - Agree to testify against the other one (**defect**)
- Each is being offered the **same deal**
- Each **knows** that the other is being offered the same deal

The Prisoner's Dilemma

- If Alice and Bob **both refuse** to testify against the other:

  2 years in prison for both

- If **Alice refuses** to talk, and **Bob testifies** against Alice:

  5 years for Alice, 0 years for Bob

- If **Alice testifies** against Bob, and **Bob refuses** to talk:

  0 years for Alice, 5 years for Bob

- If Alice and Bob **both testify** against the other:

  4 years for both

Payoff Matrix

	Bob cooperates	Bob defects
Alice cooperates	–2 , –2	–5 , 0
Alice defects	0 , –5	–4 , –4

Alice's payoff   Bob's payoff

Reward ***R*** = –2   Punishment ***P*** = –4   Temptation ***T*** = 0   Sucker payoff ***S*** = –5

Payoff Matrix (Standard Version)

	Bob cooperates	Bob defects
Alice cooperates	3 , 3	0 , 5
Alice defects	5 , 0	1 , 1

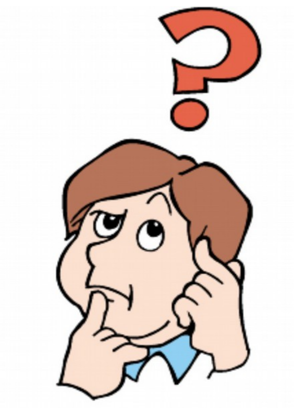
Alice's payoff   Bob's payoff

Reward ***R*** = 3   Punishment ***P*** = 1   Temptation ***T*** = 5   Sucker payoff ***S*** = 0

Alice's Point of View

- Bob will either cooperate or defect
- Suppose **Bob cooperates**:
  - If I cooperate, I will get 2 years in prison
  - If I defect, I will get 0 years ← better choice
- Suppose **Bob defects**:
  - If I cooperate, I will get 5 years
    (and Bob will get off scot-free!)
  - If I defect, I will get 4 years ← better choice
- Either way, the better choice for me is clearly to **defect**

Bob's Point of View



- Alice will either cooperate or defect
- Suppose **Alice cooperates**:
  - If I cooperate, I will get 2 years in prison
  - If I defect, I will get 0 years ← better choice
- Suppose **Alice defects**:
  - If I cooperate, I will get 5 years
    (and Alice will get off scot-free!)
  - If I defect, I will get 4 years ← better choice
- Either way, the better choice for me is clearly to **defect**

And So...

- Both Alice and Bob sleep on it
- The next day, both decide to **defect**
- Both end up in prison for **4 long years**
- If they had just cooperated, they'd be in for **half as long!**

Does logic prevent cooperation?

The Iterated Prisoner's Dilemma

- One possible formulation:
  - You exchange bags of **money** for **diamonds** with a dealer whom you've never met
  - Transactions occur **once a month** at separate drop and pickup locations in the forest
  - Each month, you must decide whether to **cooperate** (leave a full bag of money) or **defect** (leave an empty bag)
  - As far as you know, this arrangement will **continue indefinitely**, once a month

The Iterated Prisoner's Dilemma

- Suppose one day your dealer defects (leaves an empty bag)
- What should you do next time?
- Use a **strategy** to decide what to do (Cooperate or Defect) based on the recent **history** of the game

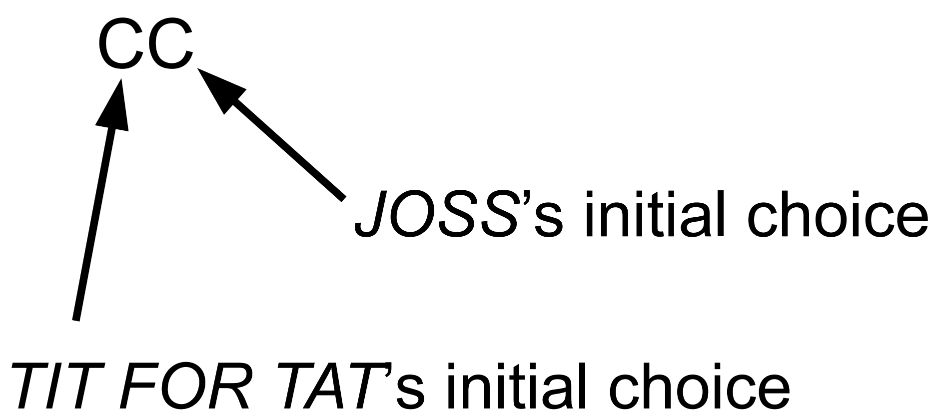
Some Example Strategies

- Always cooperate, no matter what
- Always defect, no matter what
- Cooperate for a while, then defect forever afterwards
- Cooperate until the other player defects, then defect forever afterwards (*MASSIVE RETALIATION*)

Some Example Strategies

- Cooperate or defect at random (*RANDOM*)
- Cooperate the first time, then do whatever the other player did on the previous step (*TIT FOR TAT*)
- Like *TIT FOR TAT*, but defect only when the other player defects twice in a row (*TIT FOR TWO TATS*)
- Like *TIT FOR TAT*, but with a 10% chance of defecting after the other player cooperates (*JOSS*)

TIT FOR TAT vs. *JOSS*



TIT FOR TAT vs. *JOSS*

CC CC CC … and so it goes

JOSS rewards *TIT FOR TAT*'s cooperation

TIT FOR TAT rewards *JOSS*'s cooperation

TIT FOR TAT vs. *JOSS*

CC CC CC C**D**

until *JOSS* attempts to exploit the situation

TIT FOR TAT vs. *JOSS*

CC CC CC C**D** DC

JOSS goes back to “playing nice”

TIT FOR TAT responds to *JOSS*’s defection

TIT FOR TAT vs. *JOSS*

CC CC CC C**D** DC CD

JOSS punishes *TFT*’s defection

TIT FOR TAT rewards *JOSS*’s cooperation

TIT FOR TAT vs. *JOSS*

CC CC CC C**D** DC CD DC

JOSS rewards TFT's cooperation

TIT FOR TAT punishes *JOSS*'s defection

TIT FOR TAT vs. *JOSS*

CC CC CC C**D** DC CD DC CD DC CD … and so it goes

TIT FOR TAT vs. *JOSS*

CC CC CC C**D** DC CD DC CD DC CD D**D**

until *JOSS* inevitably tries again

TIT FOR TAT vs. *JOSS*

CC CC CC C**D** DC CD DC CD DC CD D**D** DD

TIT FOR TAT punishes *JOSS*'s defection

JOSS punishes *TFT*'s defection

TIT FOR TAT vs. *JOSS*

CC CC CC C**D** DC CD DC CD DC CD D**D** DD DD DD …

TIT FOR TAT punishes *JOSS*'s defection

JOSS punishes *TFT*'s defection

… and so it goes, forever after

TIT FOR TAT vs. *JOSS*

CC CC CC C**D** DC CD DC CD DC CD D**D** DD DD DD …

- Result: a **complete breakdown** of trust and cooperation
- *JOSS*'s attempt at exploitation **backfires**
- The same thing likely happens when *JOSS* plays against other strategies, **limiting its overall gain** in the long run
- *TIT FOR TAT* likely **does better in the long run** when it plays against other "nicer" strategies

No Single Best Strategy Exists

- It all depends on the strategy the other player is using, and how long the game may last
- If Bob's strategy is *ALWAYS DEFECT*, the best strategy for Alice is also to always defect

  Payoff: *PPPPPP* …
- If Bob's strategy is *MASSIVE RETALIATION*, the best strategy for Alice is to cooperate for as long as possible, then defect on the very last move

  Payoff: *RRRRRRT*
- If Bob's strategy doesn't depend in any way on what Alice actually does, the best strategy for Alice is to always defect
- Cooperation is good for Alice only if Bob can be influenced by what Alice does

First IPD Computer Tournament

- Organized by **Robert Axelrod** at the University of Michigan
- Invited **game theory experts** in mathematics, economics, political science, and social science to submit strategies for an Iterated Prisoner's Dilemma computer tournament
- Strategies were encoded as **computer programs**
  Input: history of previous 3 games (e.g. CC CD DD)
  Output: move for this game (C or D)
- Some strategies incorporated randomness
- Some strategies were very sophisticated and complex (e.g., some used Markov models, or Bayesian inference)

First IPD Computer Tournament

- 15 strategies in all, including *RANDOM*
- All strategies played each other 200 times, round-robin style
- Strategy score: average number of points earned per game
- Tournament was run 5 times, and results averaged

- *And the winner was...*

First IPD Computer Tournament

- 15 strategies in all, including *RANDOM*
- All strategies played each other 200 times, round-robin style
- Strategy score: average number of points earned per game
- Tournament was run 5 times, and results averaged

- Winner: *TIT FOR TAT* (simplest strategy of all)

Cooperate the first time, then do whatever the other player did on the previous step

Strategy Properties

- ***Nice***: "Don't be the first to defect"
- ***Forgiving***: "Don't use massive retaliation" or "Don't hold a grudge after punishing a defection"
- ***Provocable***: "Do retaliate for a defection"
- ***Responsive***: "Base your behavior, at least in part, on what the other player does"
- ***Clarity***: "Be responsive in a way that is recognizable"
- ***Robust***: "Be effective against a variety of other strategies"

TIT FOR TAT has all of these properties

Second IPD Computer Tournament

- Strategies submitted by experts in the same fields as before, plus evolutionary biology, physics, and computer science
- 63 strategies in all, including *RANDOM*
- Many more sophisticated strategies than in Tournament 1
- Participants could submit any type of strategy, and were aware of *TIT FOR TAT* and the other strategies from Tournament 1

- *And the new winner was …*

TIT FOR TAT, once again!

Second IPD Computer Tournament

- Among the **top 15 strategies**, only one was not “nice” (#8)
- Among the **bottom 15 strategies**, only one was “nice”
- *TIT FOR TWO TATS* came in 24^{th}

- **Cooperation** was a key feature of the most successful strategies
- **Provocability** was important, but **restrained** retaliation was more successful than **massive** retaliation
- **Forgiveness** helped to restore cooperation in the face of occasional defections

- General Lesson: be **nice**, **provocable**, and **forgiving**

Evolving Strategies with a Genetic Algorithm

Evolving Strategies with a Genetic Algorithm

- Encoding a strategy as a genome:

History		*Move*
CC CC CC		**C**
CC CC CD		**D**
CC CC DC		**D**
CC CC DD		**C**
CC CD CC	→	**C**
CC CD CD		**D**
⋮		⋮
CD CC CC		**D**
⋮		⋮
DD DD DD		**C**

64 rows

+ 6 more symbols to encode initial "history"

CCCDCC CDDCCD . . . D . . . C

70 bits

2^{70} possible strategies in all

= 1,180,591,620,717,411,303,424

An Example Game

CCCDCC CDDCCD . . . D . . . C

History		*Move*
CC CC CC		**C**
CC CC CD		**D**
CC CC DC		**D**
CC CC DD		**C**
CC CD CC	→	**C**
CC CD CD		**D**
⋮		⋮
CD CC CC		**D**
⋮		⋮
DD DD DD		**C**

My first move: **CCCDCC** → **C**

Other player's first move: **C**

An Example Game

CCCDCC CDDCCD . . . D . . . C

History		*Move*
CC CC CC		**C**
CC CC CD		**D**
CC CC DC		**D**
CC CC DD		**C**
CC CD CC	→	**C**
CC CD CD		**D**
⋮		⋮
CD CC CC		**D**
⋮		⋮
DD DD DD		**C**

My first move: **CCCDCC** → **C**
Other player's first move: **C**

My next move: **CDCCCC** → **D**
Other player's next move: **C**

An Example Game

CCCDCC CDDCCD . . . D . . . C

History		*Move*
CC CC CC		**C**
CC CC CD		**D**
CC CC DC		**D**
CC CC DD		**C**
CC CD CC	→	**C**
CC CD CD		**D**
⋮		⋮
CD CC CC		**D**
⋮		⋮
DD DD DD		**C**

My first move: **CCCDCC** → **C**
Other player's first move: **C**

My next move: **CDCCCC** → **D**
Other player's next move: **C**

My next move: **CCCCDC** → **D**
Other player's next move: **D**

An Example Game

CCCDCC CDDCCD . . . D . . . C

History		*Move*
CC CC CC		**C**
CC CC CD		**D**
CC CC DC		**D**
CC CC DD		**C**
CC CD CC	→	**C**
CC CD CD		**D**
⋮		⋮
CD CC CC		**D**
⋮		⋮
DD DD DD		**C**

My first move: **CCCDCC** → **C**
Other player's first move: **C**

My next move: **CDCCCC** → **D**
Other player's next move: **C**

My next move: **CCCCDC** → **D**
Other player's next move: **D**

My next move: **CCDCDD** → *etc*.

The Genetic Algorithm: Details

- Population size: 20 strategies
- Fitness-proportionate selection
- Multi-point crossover (average of one crossover per genome)
- Mutation rate: 0.7% per position

- Each GA run evolved for 50 generations

- Performed 40 separate runs with random initial populations

- Fitness: average score of a strategy in an **environment**

Environment 1

- Fitness of a strategy
  - Strategy plays against **8 representative strategies** from IPD Tournament 2
  - 151 moves per game
  - Fitness is the average score over all games played
- Environment 1 is **fixed**

Results: Environment 1

- 29 of 40 runs evolved strategies **similar** to *TIT FOR TAT*
- 11 of 40 runs evolved strategies **better** than *TIT FOR TAT*
  - More exploitative than *TIT FOR TAT*
  - Always defect on first move (and sometimes second)
  - Not nice
  - Use player's responses to decide how to proceed
  - With unexploitable players, they "apologize" and then try to mutually cooperate
  - With exploitable players, they continue to exploit
  - Less robust, but highly adapted to Environment 1

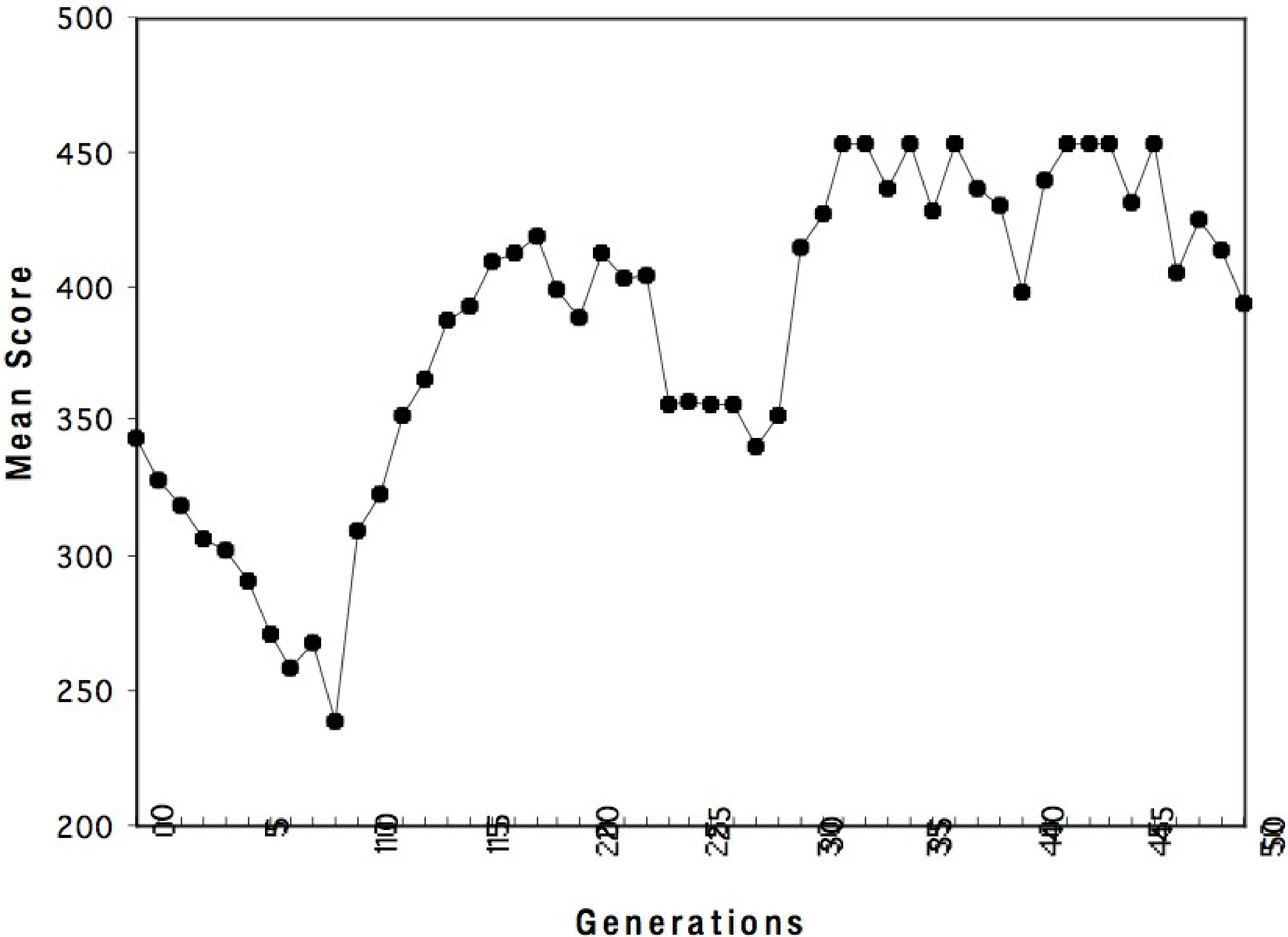
Turning Off Crossover

- 40 additional runs were conducted **without crossover**
- Each offspring strategy included information from only **one parent** instead of two
- Same fitness evaluation as before, using Environment 1
- **Results:**
  - Again, most runs found strategies that were **similar** to *TIT FOR TAT* in their effectiveness
  - Only about **half of the runs** (5 out of 40) found strategies that were substantially **better** than *TIT FOR TAT* (compared to 11 out of 40 when using crossover)

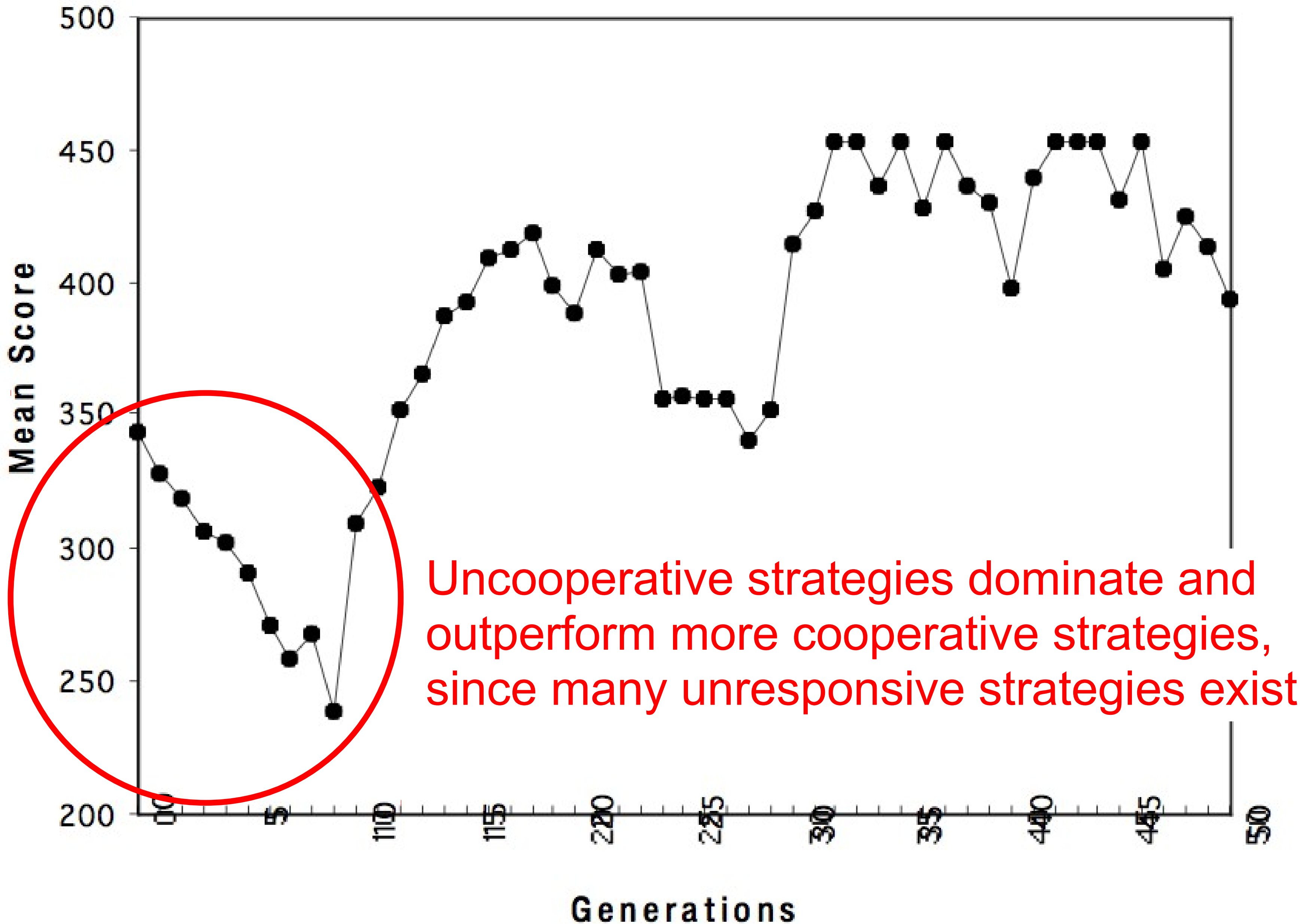
Environment 2

- Crossover was reinstated
- Fitness of a strategy
  - Strategy plays against **all 20 strategies in the current GA population**, including itself
  - 151 moves per game
  - Fitness is the average score over all games played
- Environment 2 **changes** from generation to generation
- 10 runs were performed

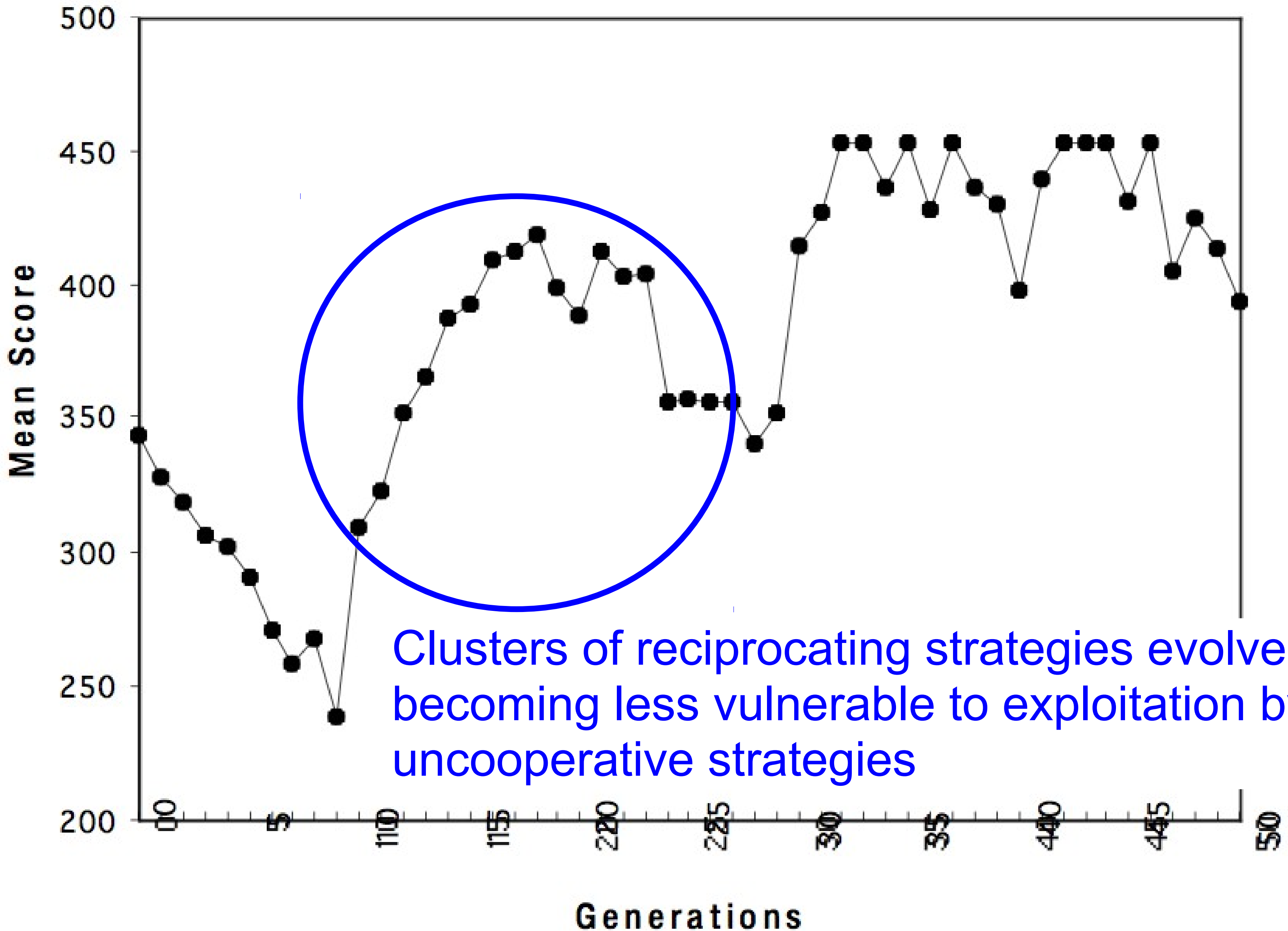
Results: Environment 2



Results: Environment 2

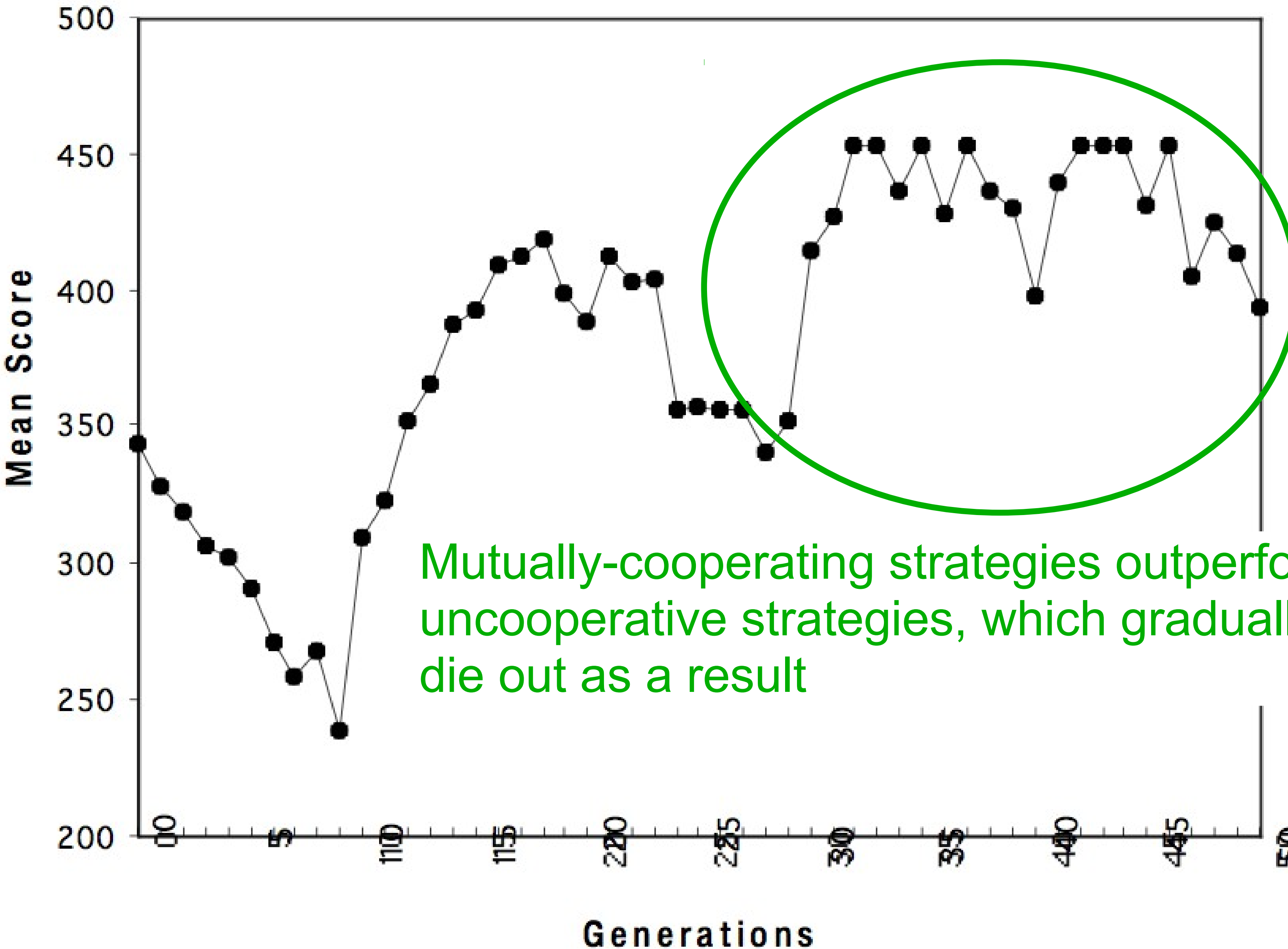


Results: Environment 2



Clusters of reciprocating strategies evolve, becoming less vulnerable to exploitation by uncooperative strategies

Results: Environment 2



Mutually-cooperating strategies outperform uncooperative strategies, which gradually die out as a result

Axelrod's Conclusions

- In the presence of many unresponsive strategies, uncooperative defectors have a strong advantage
- Cooperation can establish a foothold in a population of defectors through small clusters of reciprocating strategies
- The reciprocating strategies do well enough amongst themselves to offset being exploited by the defectors
- As cooperative strategies proliferate, the proportion of strategies vulnerable to exploitation by defectors shrinks, driving the defectors toward extinction

Modeling Social Norms

- Follow-up work by Axelrod investigated the effect of adding **social norms** to the GA model
- When a player defects, other players may **witness** the defection and punish the player, with some probability
- Each player that witnesses a defection may decide to **punish** the defector by subtracting points from its score
- New inherited traits, subject to mutation:
  - **Boldness**: a player's probability of defecting
  - **Vengefulness**: a player's probability of punishing an observed defection

Modeling Social Norms

- **Hypothesis**:
  Norms will facilitate the evolution of cooperation, with vengefulness evolving to counteract boldness

- **Simulation results**:
  - With no social norms (vengefulness values 0 in the initial population), defectors ended up dominating
  - Norms are not enough to reliably induce cooperation

Meta-Norms

- So Axelrod added **meta-norms** to the simulation:

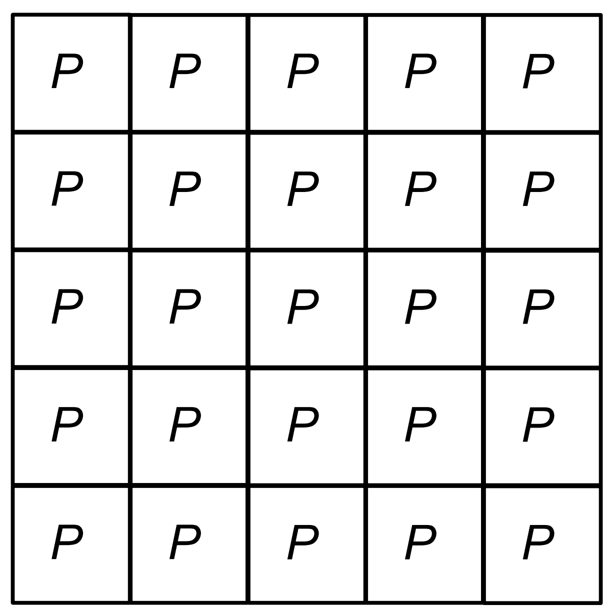
  Witnesses can be punished for not punishing the defectors!

- Example of a meta-norm:
  Bystanders' disapproving looks in a supermarket when a parent fails to discipline their child for being disruptive

- **Simulation results**:
  - Non-punishers tended to evolve into punishers
  - Defectors tended to evolve into cooperators
  - Meta-norms can indeed promote and sustain cooperation

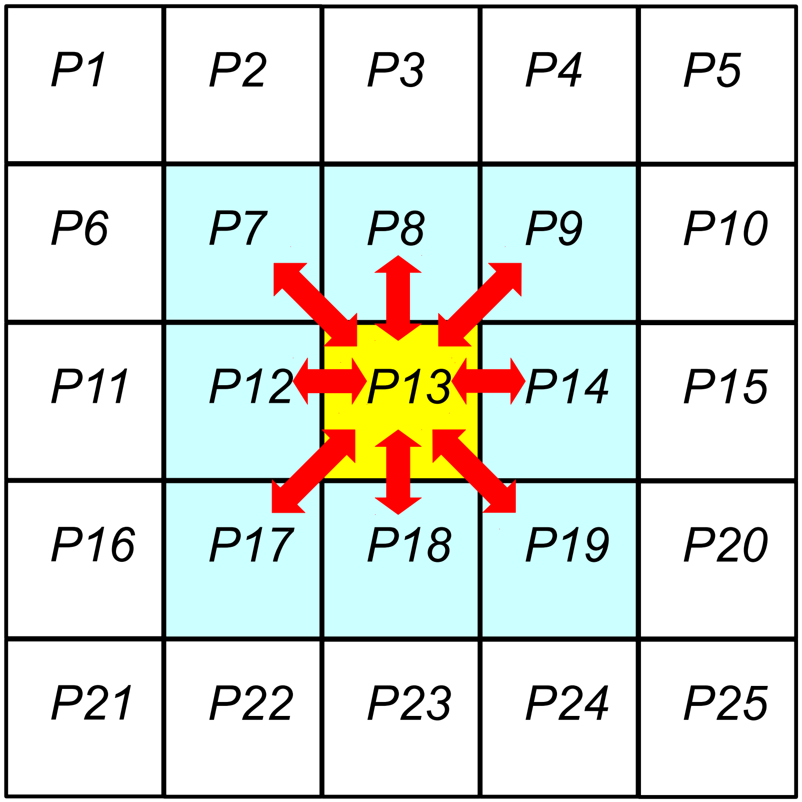
Modeling Spatial Structure

- Martin Nowak and Robert May added **spatial structure** to a simple version of the Prisoner's Dilemma
- Players either **always cooperate** or **always defect**
- Players are distributed across a **2-D lattice**



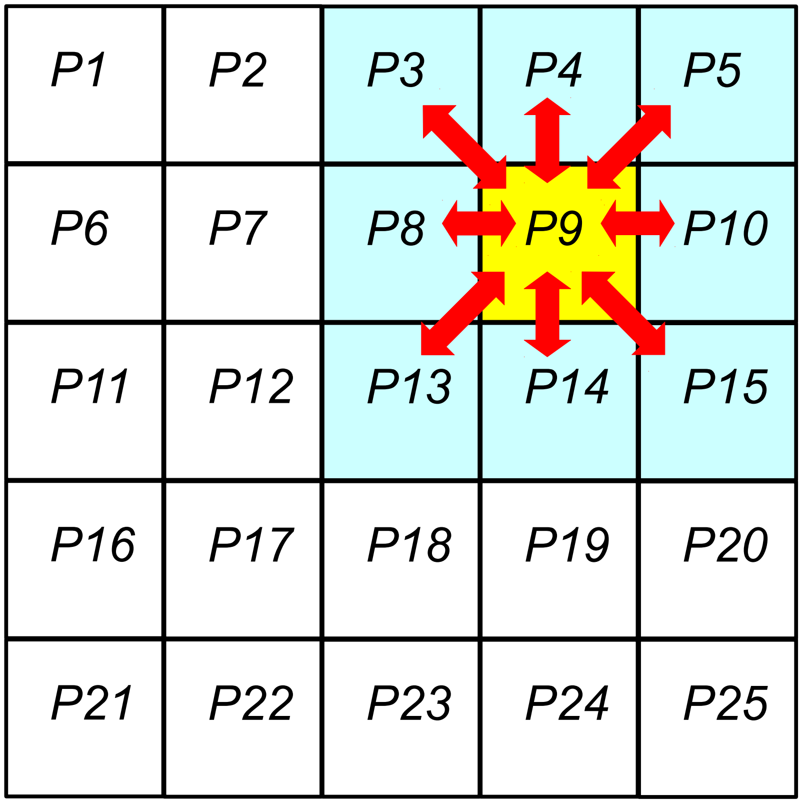
Modeling Spatial Structure

- Each player only plays against its **local neighbors**



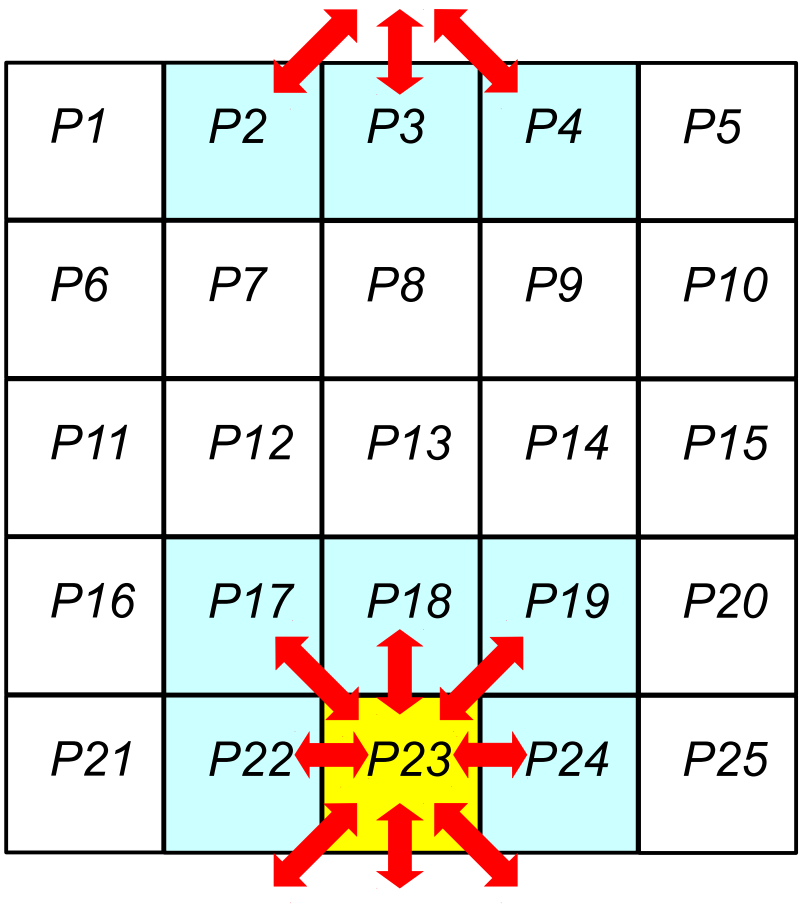
Modeling Spatial Structure

- Each player only plays against its **local neighbors**



Modeling Spatial Structure

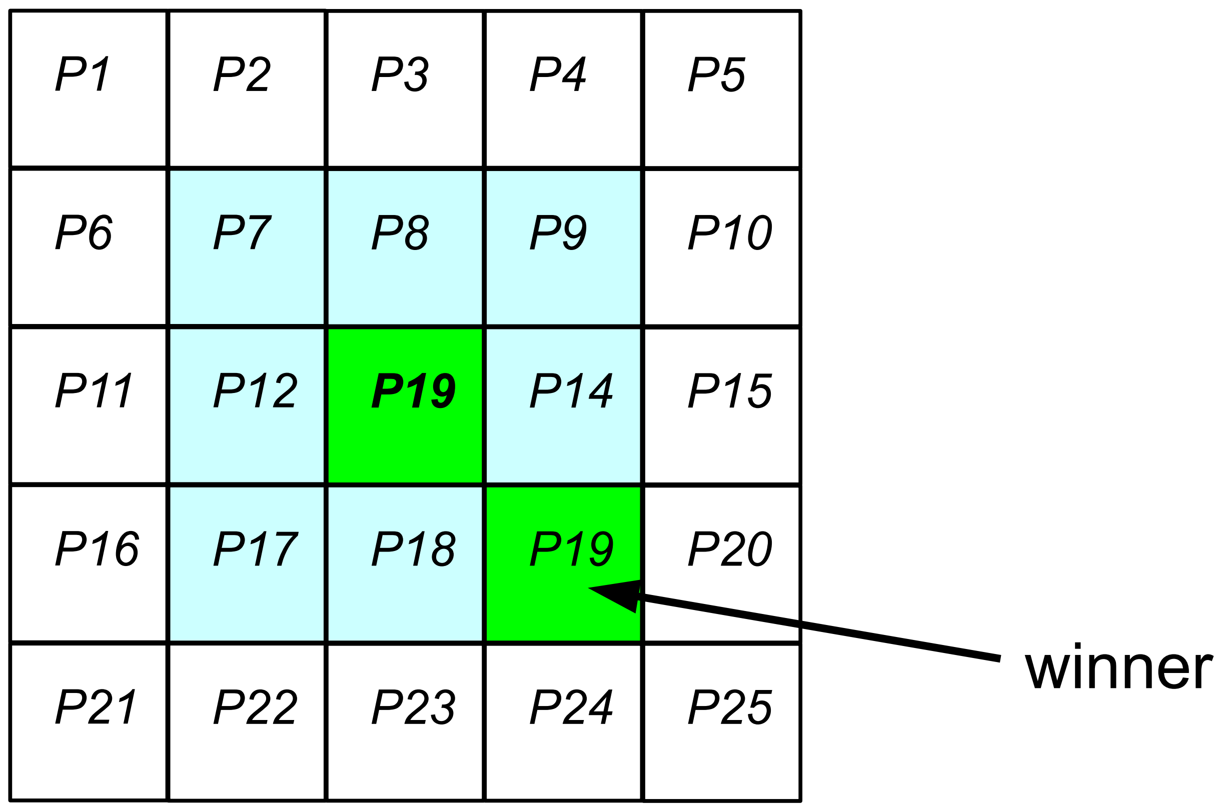
- Each player only plays against its **local neighbors**



And so on ...

Modeling Spatial Structure

- Each player only plays against its **local neighbors**
- Each player is **replaced** by the highest scoring player in its neighborhood, with no crossover or mutation



Modeling Spatial Structure

- Nowak and May experimented with different:
  - Mixtures of cooperators and defectors
  - Values of the payoff matrix

- **Results:**
  - Cooperation persisted indefinitely in the population
  - Distribution of cooperators and defectors either:
    - Oscillated indefinitely
    - Exhibited chaotic dynamics

Conclusion: Territoriality favors cooperation

Computer Modeling of the Real World

All models are wrong, but some are useful.

—George Box and Norman Draper

- Computer models of evolution (GAs) and social cooperation (Prisoner's Dilemma) are highly simplified and idealized
- Nevertheless, they can serve as a useful guide in thinking about the real phenomena being modeled
- They can provide new insights, suggest new questions, and enable controlled experiments to be performed that otherwise would be impossible

Also check out this fun website:

The Evolution of Trust

https://ncase.me/trust