Towards Metrics and Visualizations Sensitive to Coevolutionary Failures

Ari Bader-Natal Jordan B. Pollack
ari@cs.brandeis.edu

Brandeis University
DEMO Lab, Computer Science Department

AAAI 2005 Fall Symposium:
Coevolutionary and Coadaptive Systems

Outline

Motivation

Coevolutionary Fitness

- No objective fitness function required

Motivation

Coevolutionary Fitness

- No objective fitness function required
- No objective fitness function available

Motivation

Coevolutionary Fitness

- No objective fitness function required
- No objective fitness function available
- How to best monitor performance?

Notation

From (Bucci & Pollack)

- Candidate population C, Test population T

Notation

From (Bucci & Pollack)

- Candidate population C, Test population T
- Any candidate can be evaluated against any test.

Notation

From (Bucci & Pollack)

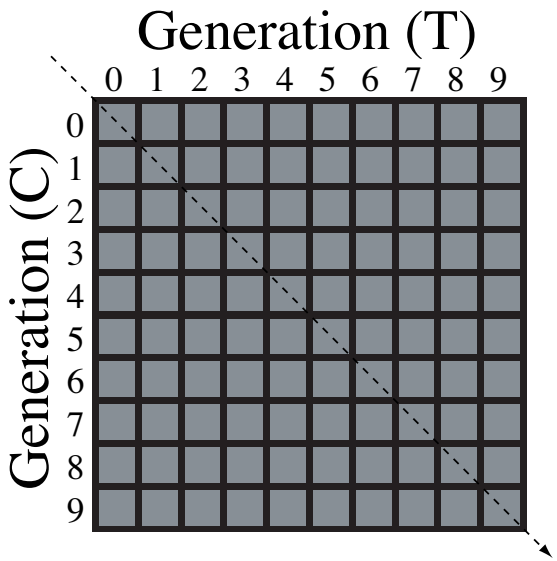
- Candidate population C, Test population T
- Any candidate can be evaluated against any test.
- Outcome is an element of some ordered set R

Notation

From (Bucci & Pollack)

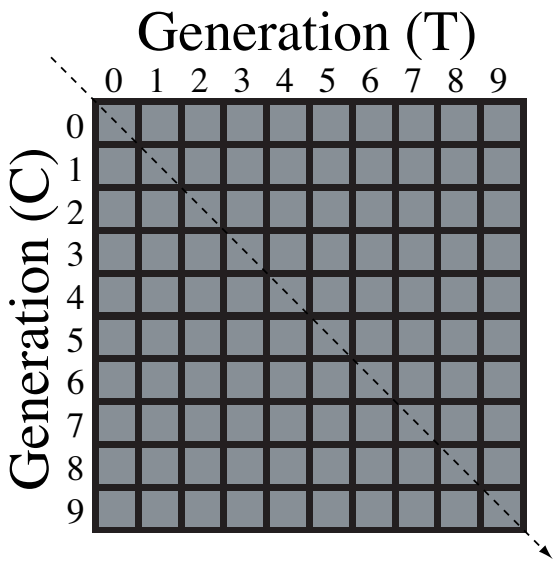
- Candidate population C, Test population T
- Any candidate can be evaluated against any test.
- Outcome is an element of some ordered set R
  $R = \{CandidateFailedTest < CandidateTiedTest < CandidatePassedTest\}$

Generation-Table Analysis



- CIAO data (Cliff and Miller)

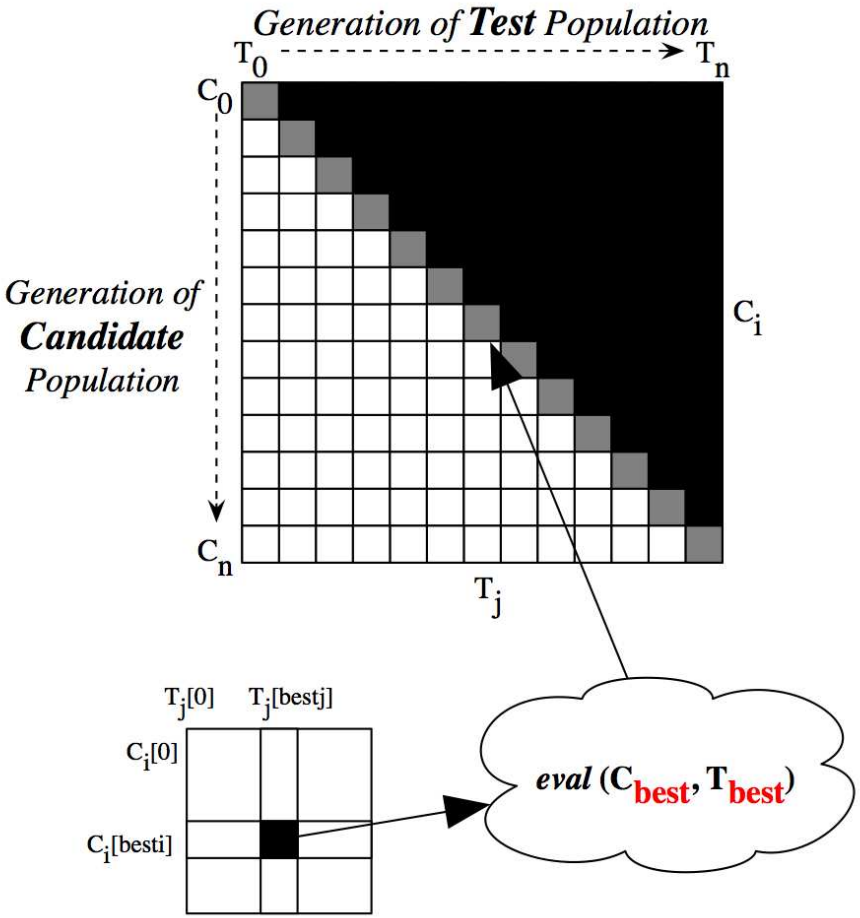
Generation-Table Analysis



- CIAO data (Cliff and Miller)
- Dominance Tournament (Stanley and Miikkulainen)
- Master Tournament (Nolfi and Floreano)
- Hall-of-Fame (Rosin and Belew)

Best-of-Generation (BOG) Approach

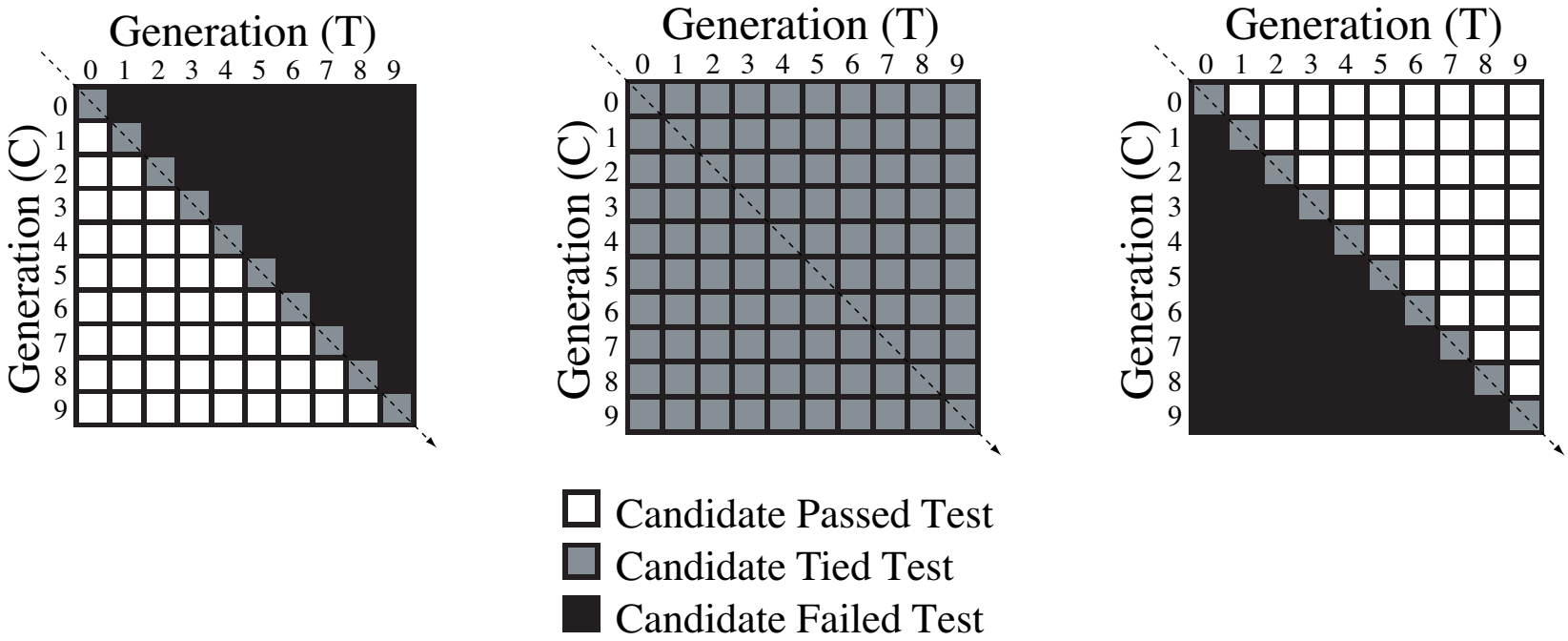
Evaluation



Candidate Passed Test
Candidate Tied Test
Candidate Failed Test

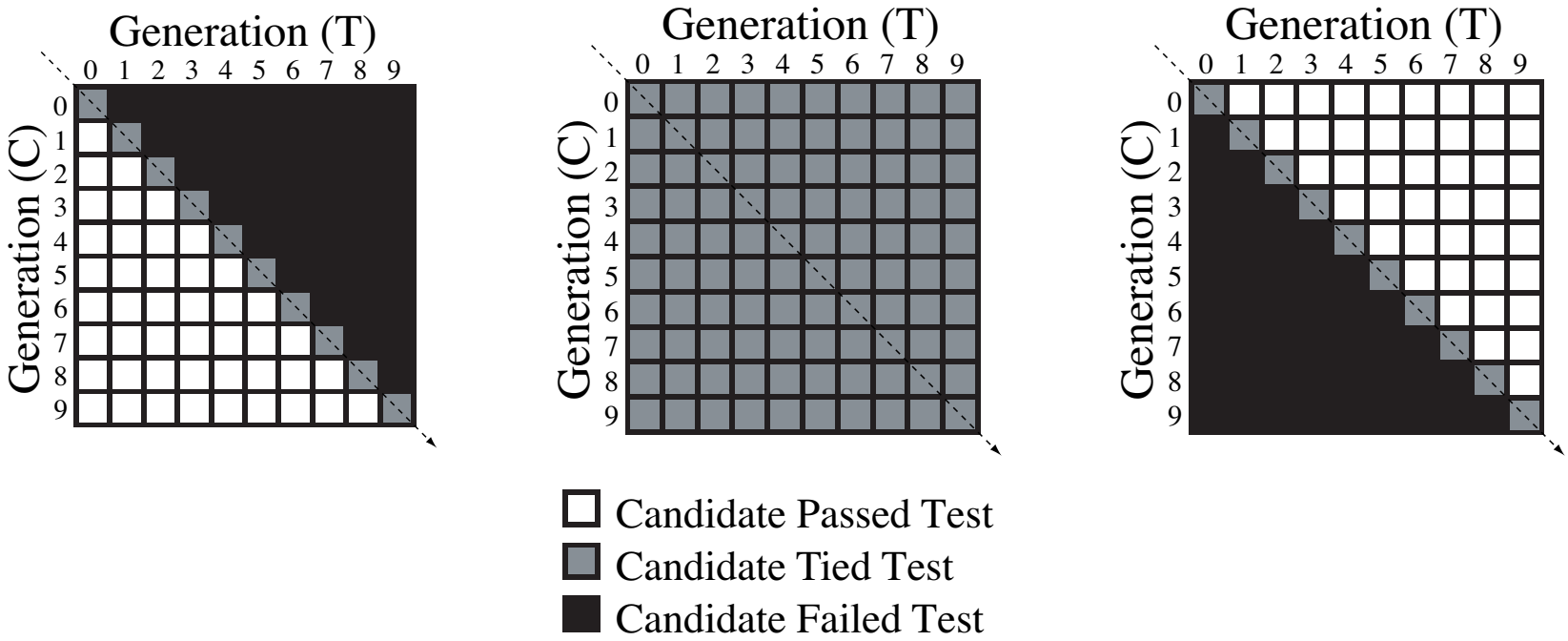
”Instantaneous Fitness”

”Red Queen” Dynamics



”Instantaneous Fitness”

”Red Queen” Dynamics



Simulations may drastically differ, yet may generate identical ”instantaneous fitness” profiles

Best-of-Generation (BOG) Approach

- Low computational cost

Best-of-Generation (BOG) Approach

- Low computational cost
- Definition of ”best” is pre-defined

Best-of-Generation (BOG) Approach

- Low computational cost
- Definition of ”best” is pre-defined
- Only reflects behavior of ”best” individuals

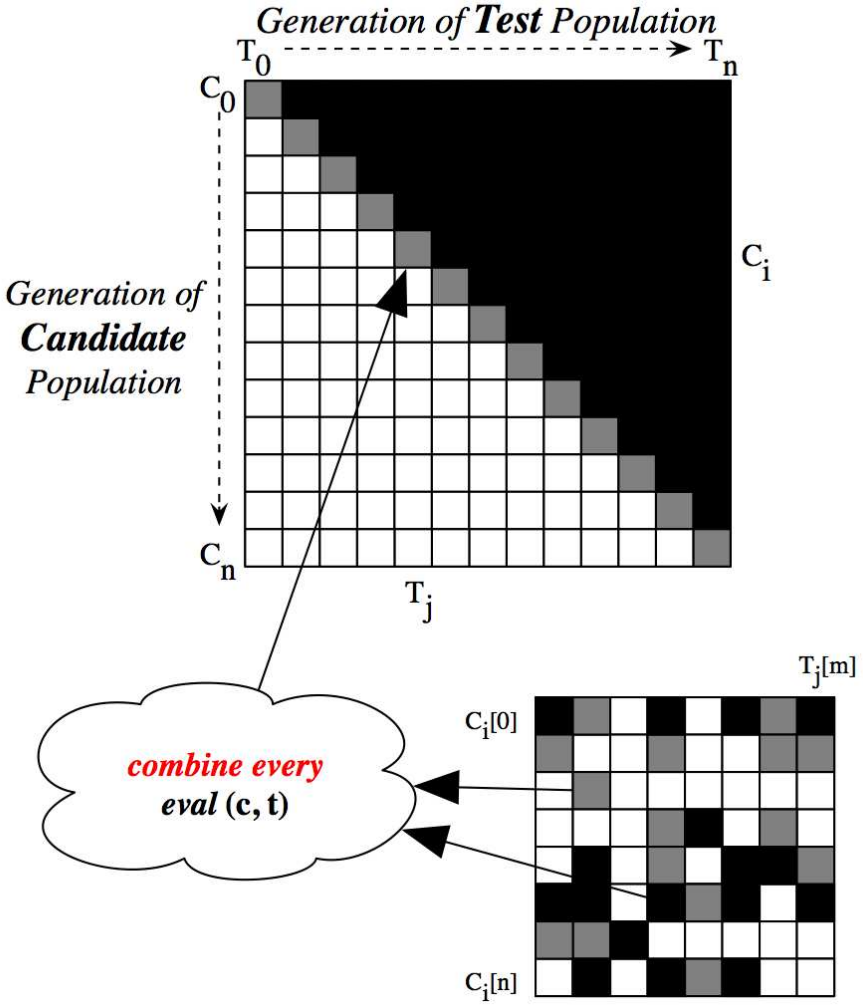
Best-of-Generation (BOG) Approach

- Low computational cost
- Definition of ”best” is pre-defined
- Only reflects behavior of ”best” individuals
- High sensitivity to coevolutionary success
  Low sensitivity to coevolutionary failure

Best-of-Generation (BOG) Approach

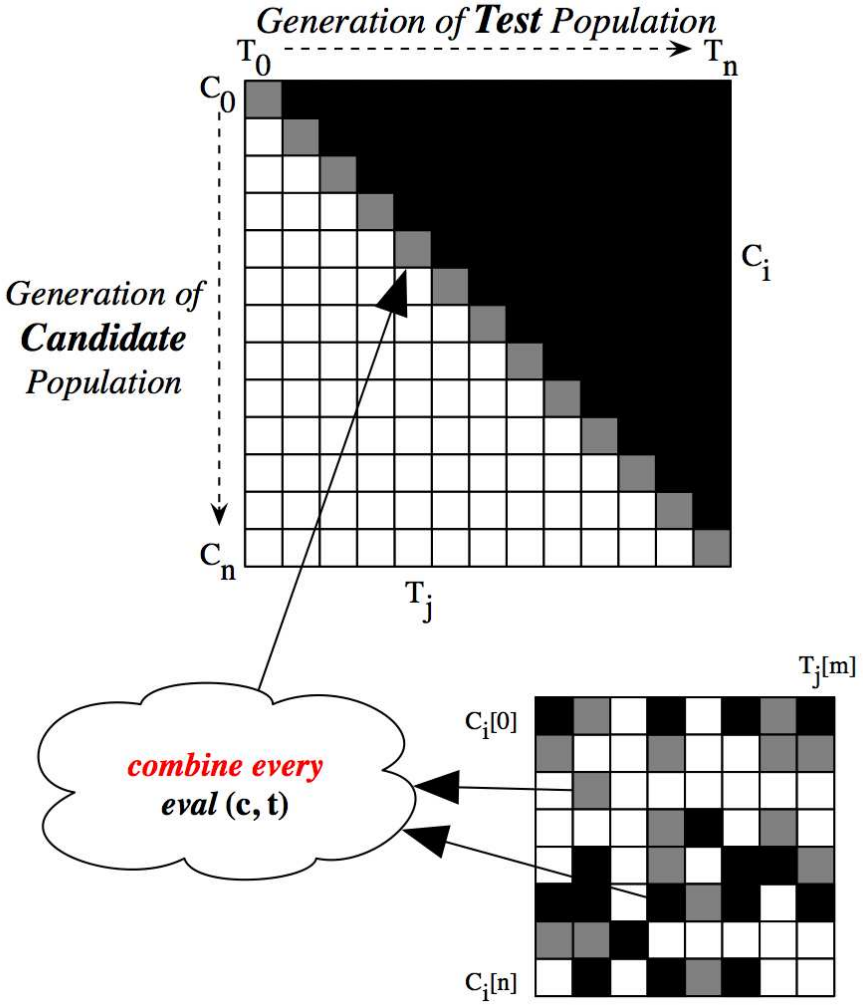
- Low computational cost
- Definition of ”best” is pre-defined
- Only reflects behavior of ”best” individuals
- High sensitivity to coevolutionary success
  Low sensitivity to coevolutionary failure
- Not appropriate for examining failures in coevolution.
  All-of-Generation (AOG) based analysis is a response...

All-of-Generation: Population-Grained Evaluation



All Candidates Passed All Tests
All Candidates Failed All Tests

All-of-Generation: Population-Grained Evaluation



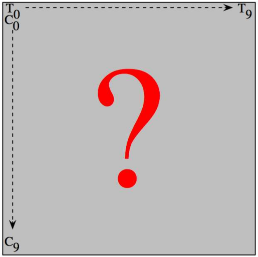
All Candidates Passed All Tests
All Candidates Failed All Tests

Here: simple average of outcome values

BOG vs. AOG Differentiating Example

0	1	2	3	4	5	6	7	8	9
0	-1	-2	-3	-4	-5	-6	-7	-8	-9
0	-1	-2	-3	-4	-5	-6	-7	-8	-9
0	-1	-2	-3	-4	-5	-6	-7	-8	-9
0	-1	-2	-3	-4	-5	-6	-7	-8	-9
0	-1	-2	-3	-4	-5	-6	-7	-8	-9
0	-1	-2	-3	-4	-5	-6	-7	-8	-9
0	-1	-2	-3	-4	-5	-6	-7	-8	-9
0	-1	-2	-3	-4	-5	-6	-7	-8	-9
0	-1	-2	-3	-4	-5	-6	-7	-8	-9

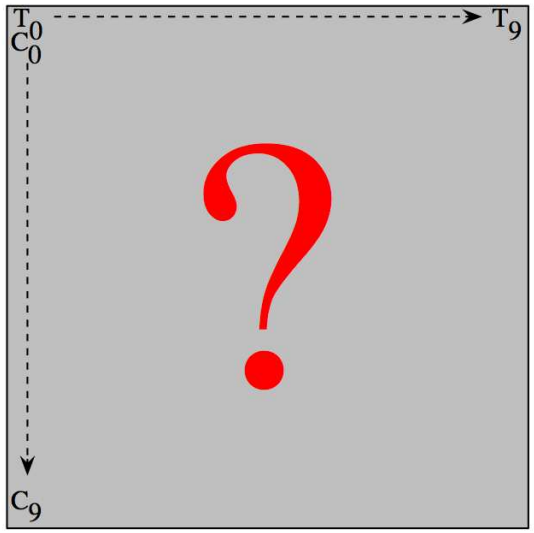
0	0	0	0	0	0	0	0	0	0
1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	-2	-2	-2	-2	-2	-2	-2	-2	-2
3	-3	-3	-3	-3	-3	-3	-3	-3	-3
4	-4	-4	-4	-4	-4	-4	-4	-4	-4
5	-5	-5	-5	-5	-5	-5	-5	-5	-5
6	-6	-6	-6	-6	-6	-6	-6	-6	-6
7	-7	-7	-7	-7	-7	-7	-7	-7	-7
8	-8	-8	-8	-8	-8	-8	-8	-8	-8
9	-9	-9	-9	-9	-9	-9	-9	-9	-9

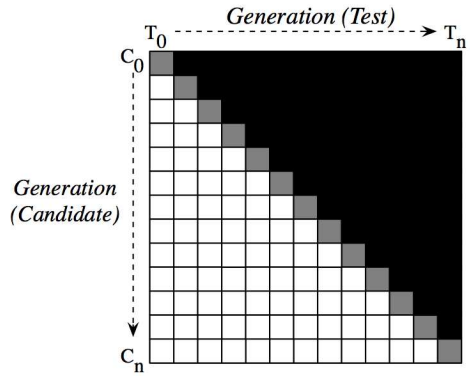


BOG vs. AOG Differentiating Example

	0	1	2	3	4	5	6	7	8	9
0	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
2	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
3	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
4	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
5	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
6	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
7	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
8	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
9	0	-1	-2	-3	-4	-5	-6	-7	-8	-9

0	0	0	0	0	0	0	0	0	0	0
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4
5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6
7	-7	-7	-7	-7	-7	-7	-7	-7	-7	-7
8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9





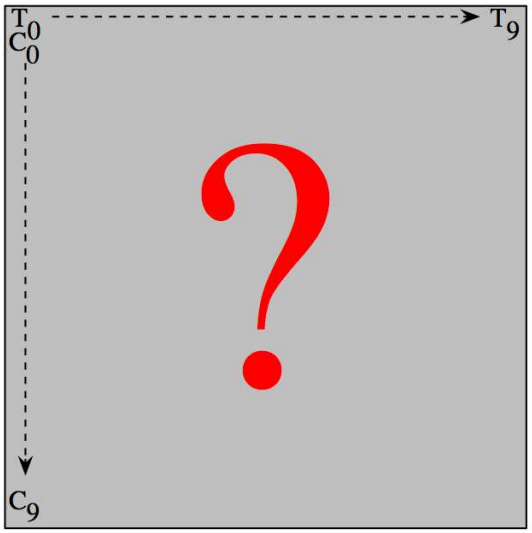
BOG table:

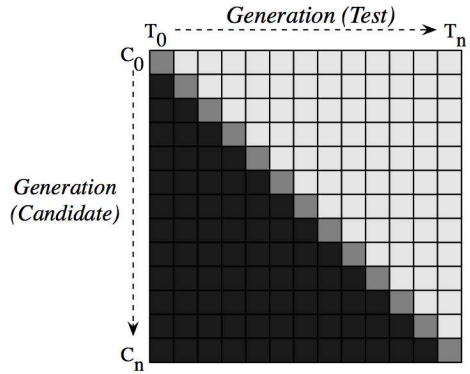
a success!

BOG vs. AOG Differentiating Example

	0	1	2	3	4	5	6	7	8	9
0	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
2	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
3	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
4	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
5	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
6	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
7	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
8	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
9	0	-1	-2	-3	-4	-5	-6	-7	-8	-9

	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4
5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6
7	-7	-7	-7	-7	-7	-7	-7	-7	-7	-7
8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9

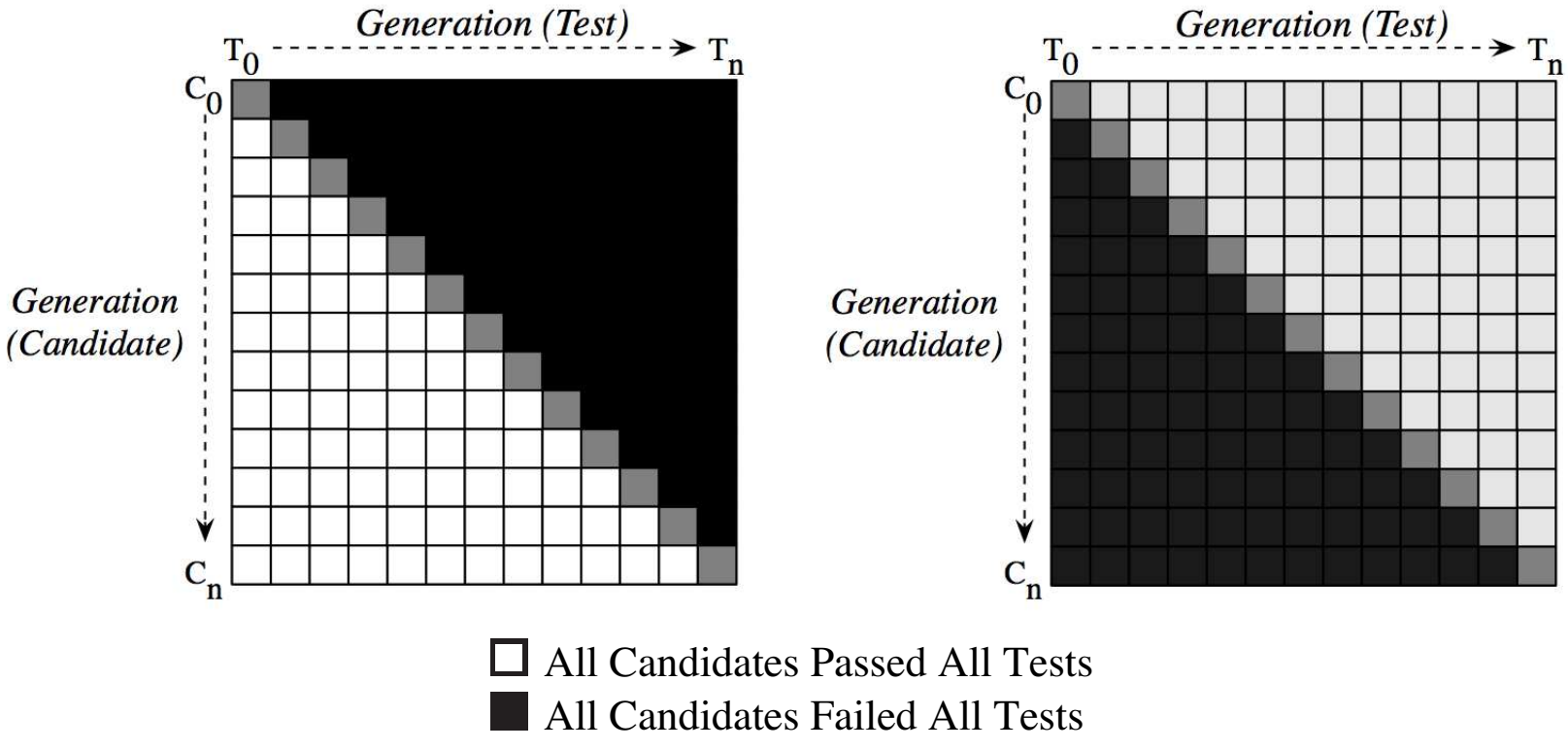




AOG table:
a failure

BOG vs. AOG Differentiating Example

Which is preferable?



The Population-Differential Metric

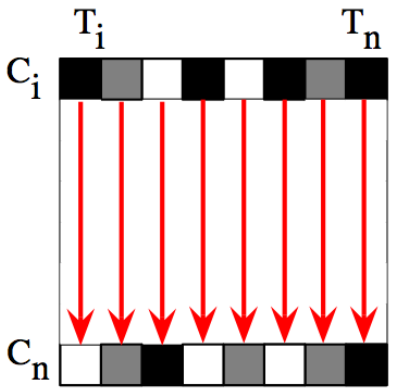
- ”How does the most recent generation compare to the least recent generation, with respect to each generation of the other population? And vice versa?

The Population-Differential Metric

- ”How does the most recent generation compare to the least recent generation, with respect to each generation of the other population? And vice versa?
- The resulting set of pairwise comparisons (of `PopEval` values) are binned according to the directionality of change.

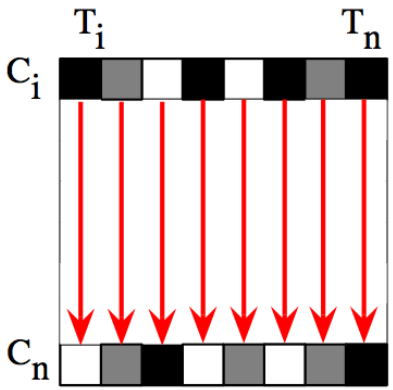
The Population-Differential Metric

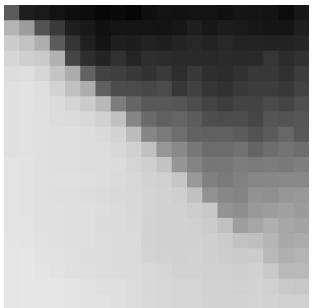
- ”How does the most recent generation compare to the least recent generation, with respect to each generation of the other population? And vice versa?
- The resulting set of pairwise comparisons (of `PopEval` values) are binned according to the directionality of change.
- Bin-counts determine value (at time *t*) ranging from +1 to -1.



The Population-Differential Metric

- ”How does the most recent generation compare to the least recent generation, with respect to each generation of the other population? And vice versa?
- The resulting set of pairwise comparisons (of `PopEval` values) are binned according to the directionality of change.
- Bin-counts determine value (at time *t*) ranging from +1 to -1.





Experimental Domain: ”Numbers Game” Variants

from (Watson & Pollack)

- Simple domains that produce interesting behavior.

Experimental Domain: ”Numbers Game” Variants

from (Watson & Pollack)

- Simple domains that produce interesting behavior.
- Objective metric exists, allowing for verification of subjective (population-differential) metric.

Experimental Domain: ”Numbers Game” Variants

from (Watson & Pollack)

- Simple domains that produce interesting behavior.
- Objective metric exists, allowing for verification of subjective (population-differential) metric.
- 1D variant: Individuals are points on a line. The higher value wins.

Experimental Domain: ”Numbers Game” Variants

from (Watson & Pollack)

- Simple domains that produce interesting behavior.
- Objective metric exists, allowing for verification of subjective (population-differential) metric.
- 1D variant: Individuals are points on a line. The higher value wins.
- ”Intransitive” variant: Individuals are points on 2D grid. The ”winner” has the higher value in dimension for which the two are closer.

Experimental Domain: ”Numbers Game” Variants

from (Watson & Pollack)

- Simple domains that produce interesting behavior.
- Objective metric exists, allowing for verification of subjective (population-differential) metric.
- 1D variant: Individuals are points on a line. The higher value wins.
- ”Intransitive” variant: Individuals are points on 2D grid. The ”winner” has the higher value in dimension for which the two are closer.
- ”Compare-on-one” variant: Individuals are points on 2D grid. The ”winner” has the higher value in test’s greater dimension.

Profiles of Coevolutionary Failures

Sought to reproduce common coevolutionary behaviors:

Profiles of Coevolutionary Failures

Sought to reproduce common coevolutionary behaviors:

- ”arms-race dynamic”

Profiles of Coevolutionary Failures

Sought to reproduce common coevolutionary behaviors:

- ”arms-race dynamic”
- ”lock-in failure”

Profiles of Coevolutionary Failures

Sought to reproduce common coevolutionary behaviors:

- ”arms-race dynamic”
- ”lock-in failure”
- ”variation”

Profiles of Coevolutionary Failures

Sought to reproduce common coevolutionary behaviors:

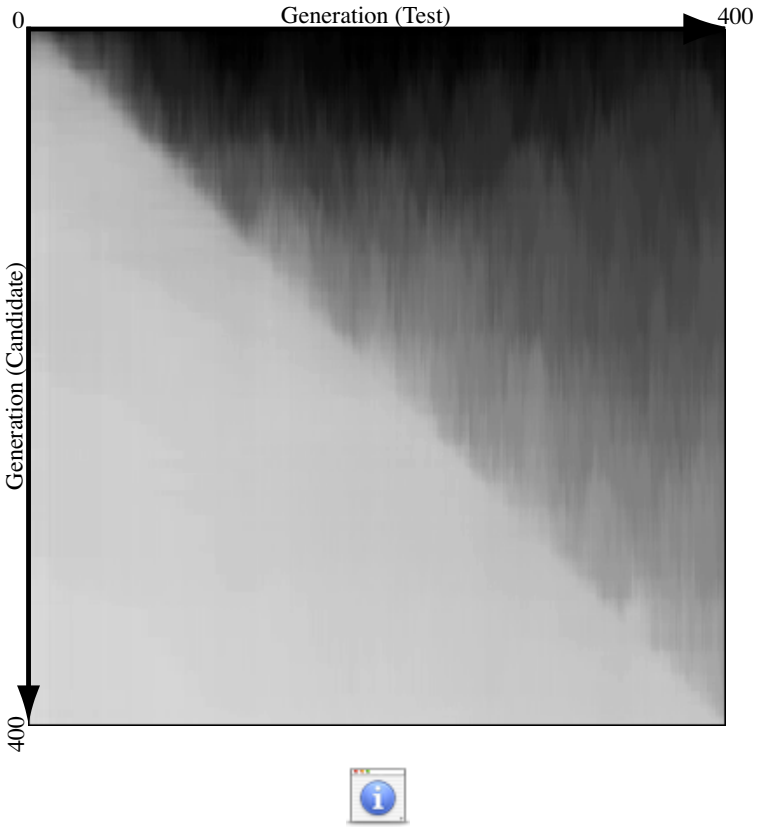
- ”arms-race dynamic”
- ”lock-in failure”
- ”variation”
- ”disengagement” (due to lack of gradient)

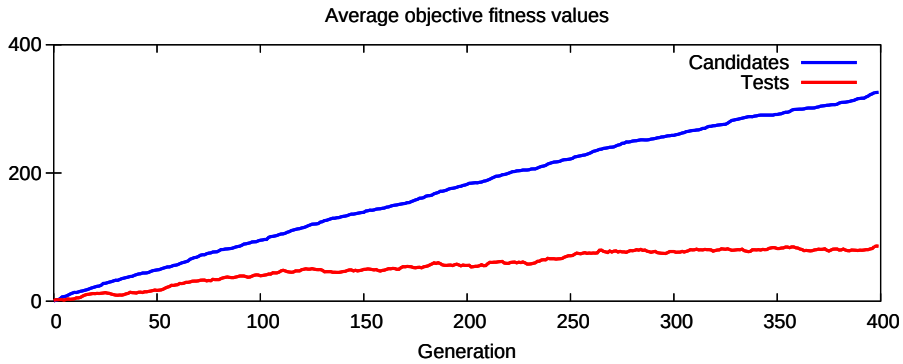
Profiles of Coevolutionary Failures

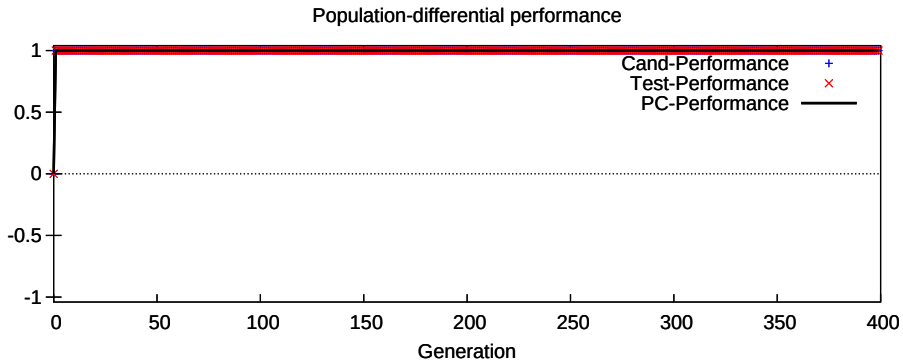
Sought to reproduce common coevolutionary behaviors:

- ”arms-race dynamic”
- ”lock-in failure”
- ”variation”
- ”disengagement” (due to lack of gradient)
- ”cycling”

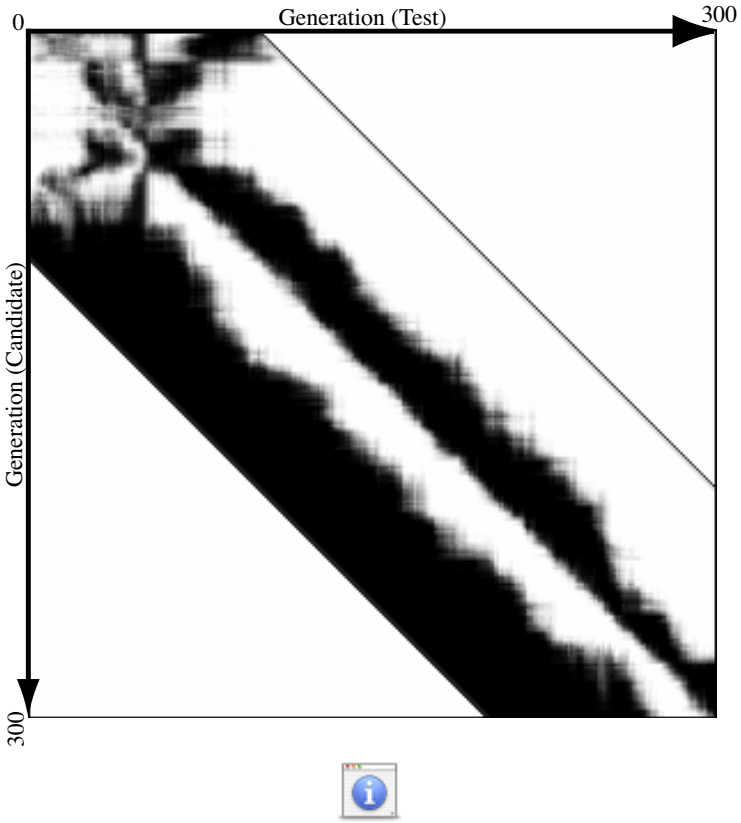
Behavior Example: Arms-Race Dynamics

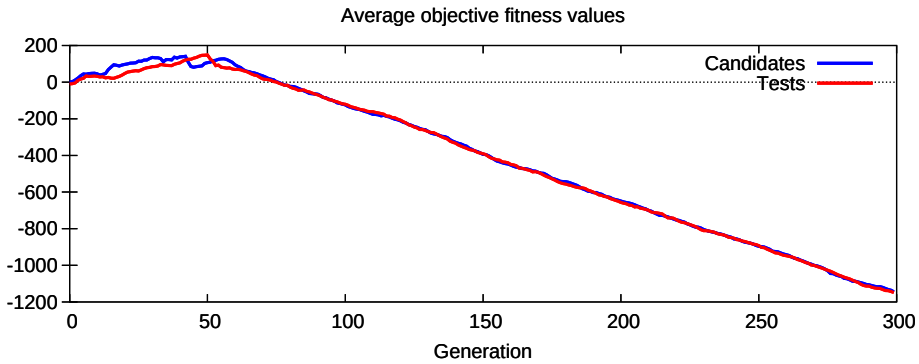


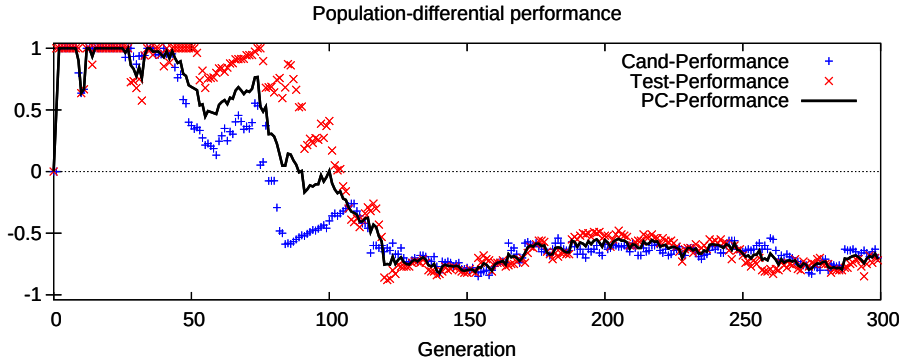




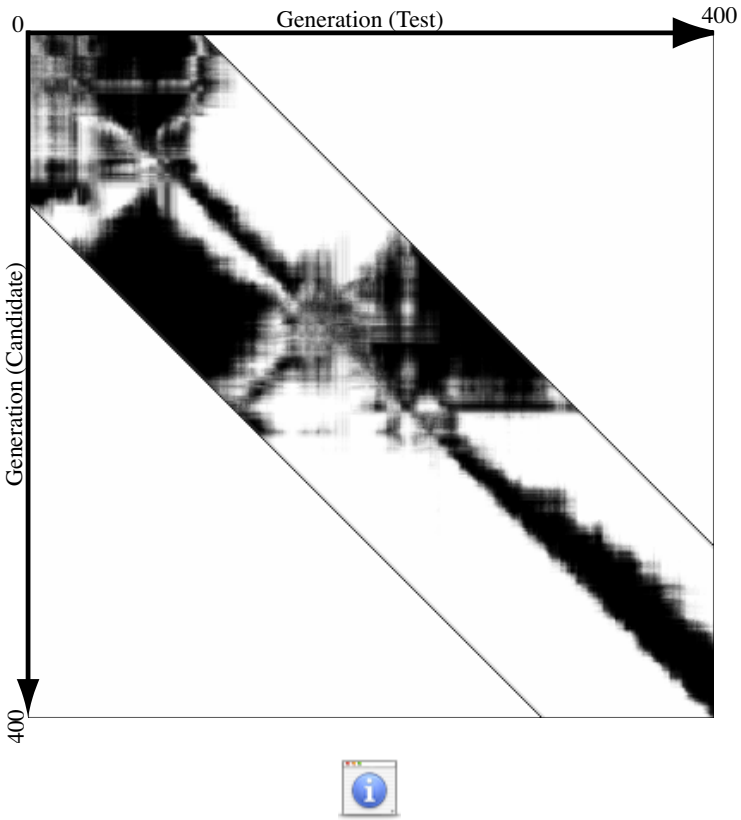
Behavior Example: Lock-In Failure

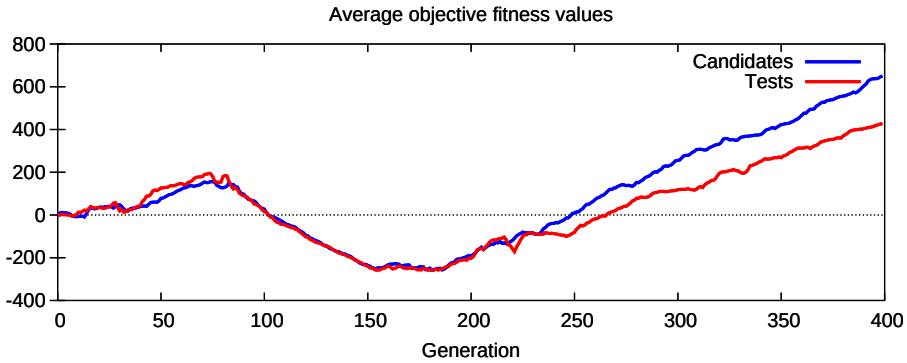


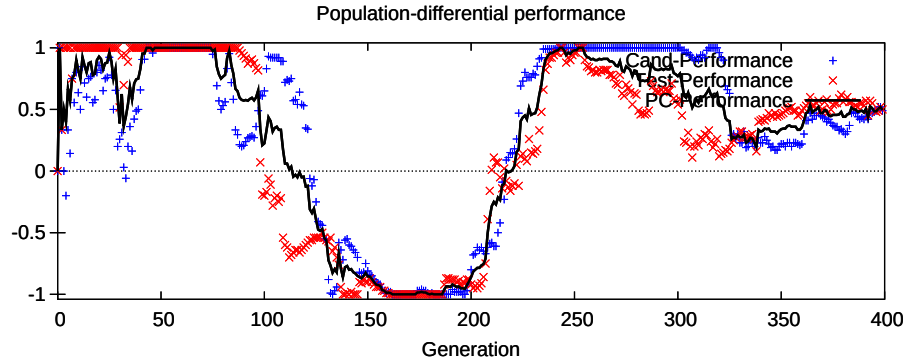




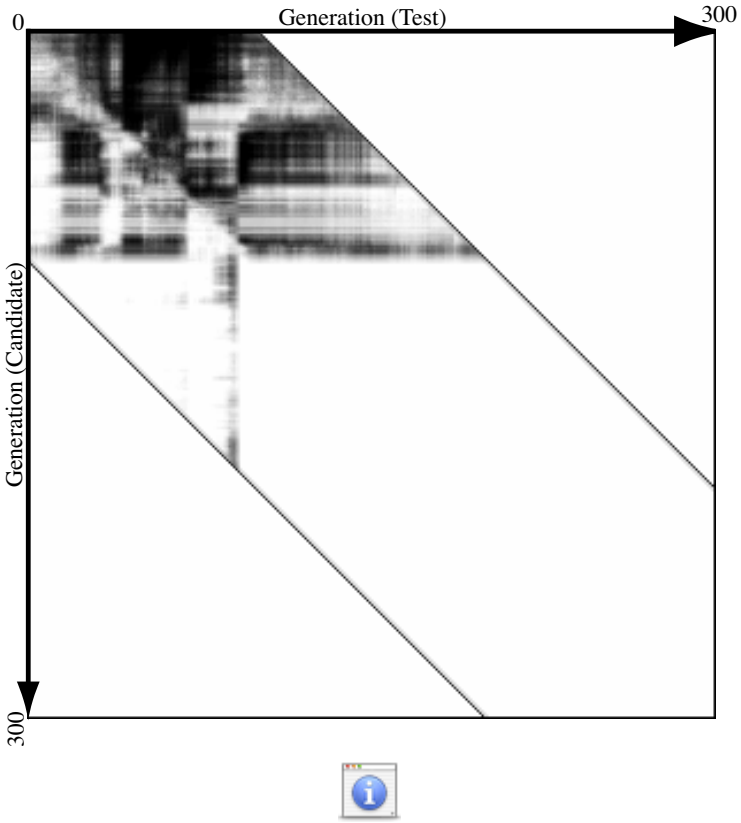
Behavior Example: Variation

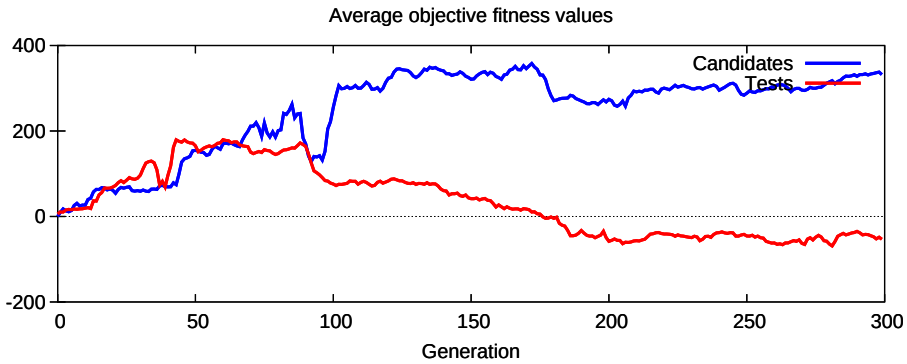


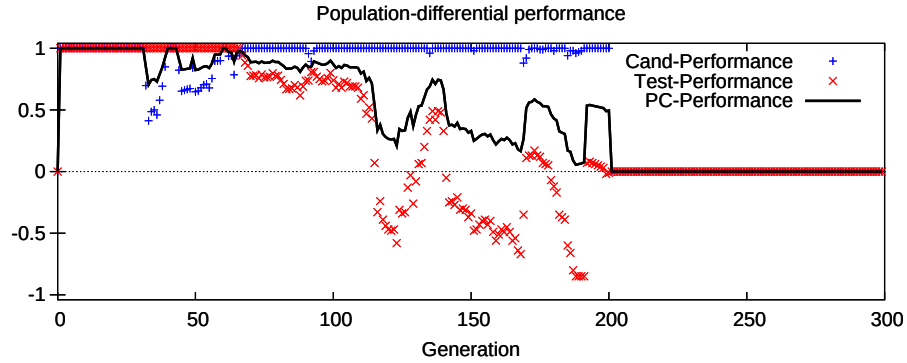




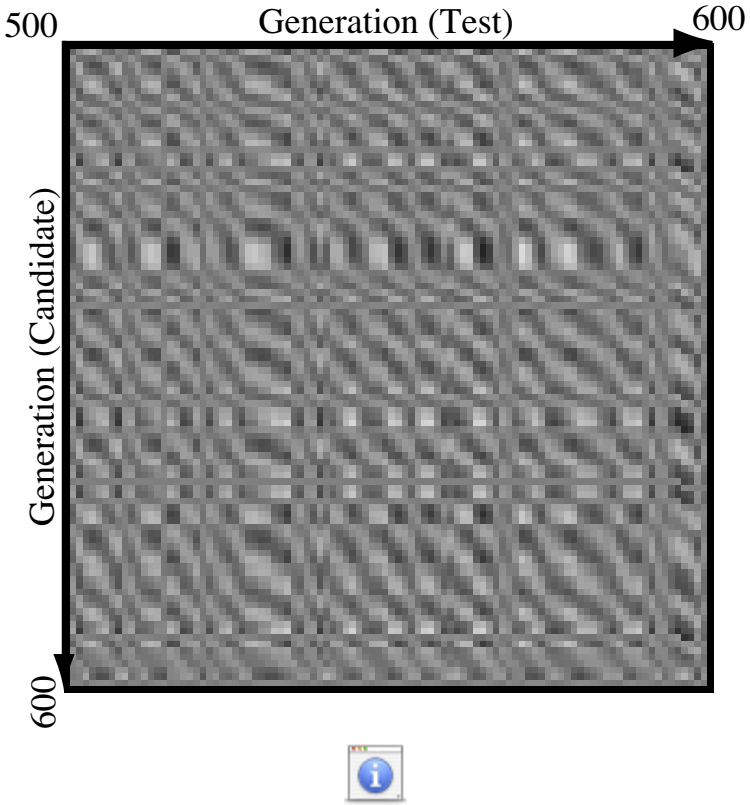
Behavior Example: Disengagement

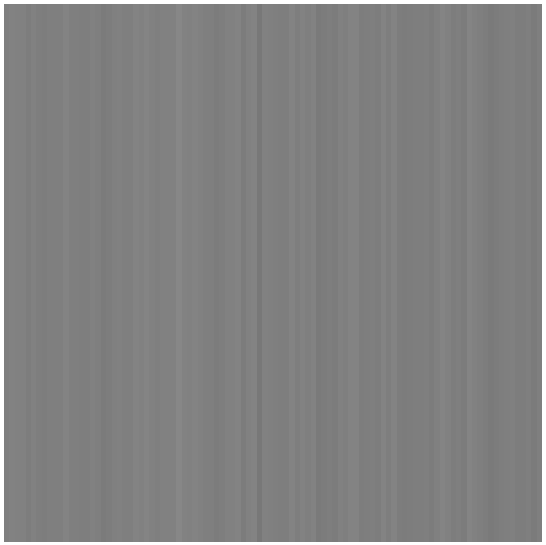


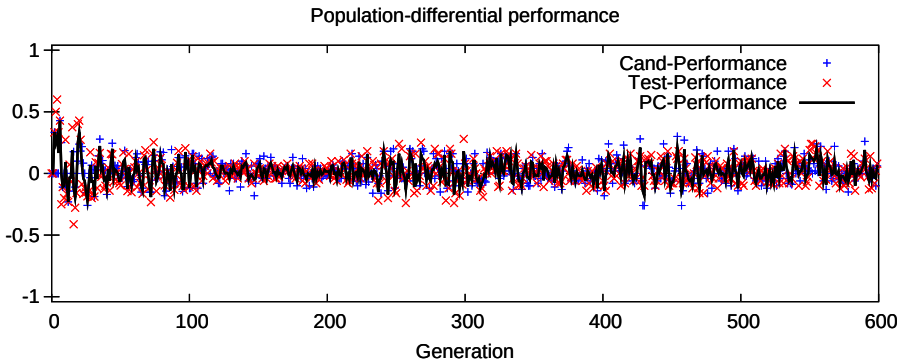




Behavior Example: Cycling







Summary

- BOG-based analysis is good for monitoring successes but not so good for monitoring failures
- AOG-based analysis uses information on the entire population (not just the ”best”)
- The ”Generation-Table approach” to coevolutionary analysis continues to yield useful techniques.

Thank you

Thank you!

Towards Metrics and Visualizations Sensitive to Coevolutionary Failures

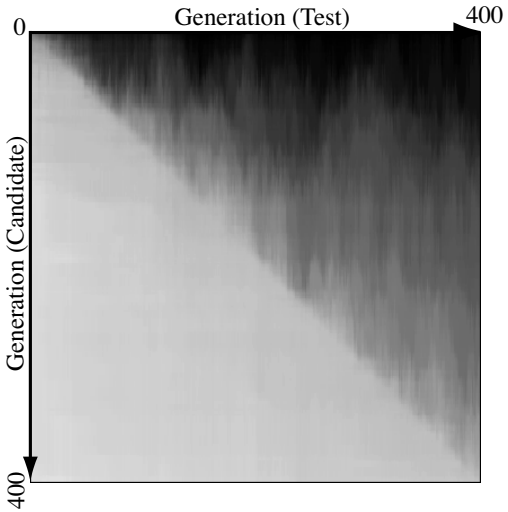
Ari Bader-Natal Jordan B. Pollack
ari@cs.brandeis.edu

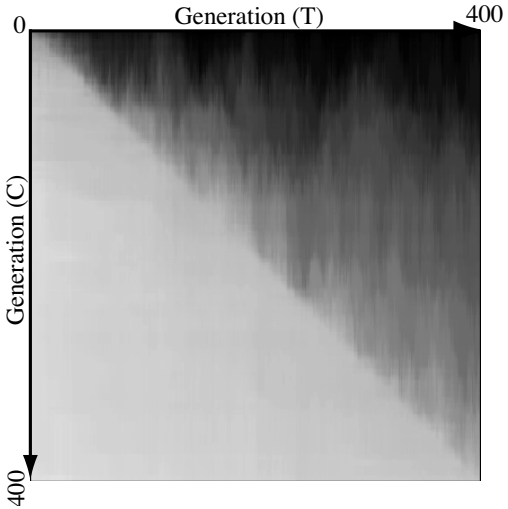
Brandeis University
DEMO Lab, Computer Science Department

AAAI 2005 Fall Symposium:
Coevolutionary and Coadaptive Systems

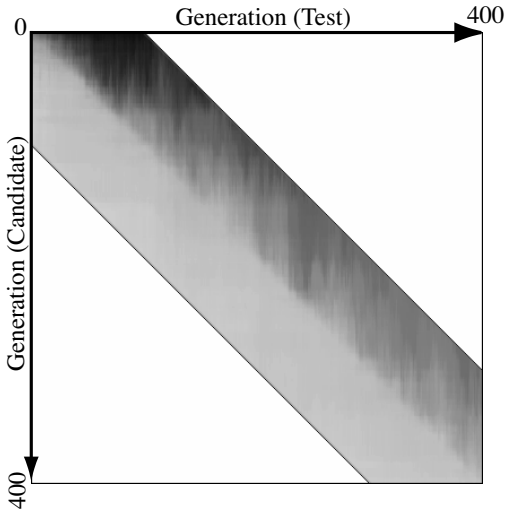
AOG Memory Policies to Reduce Computation

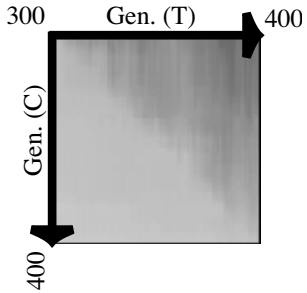
Lossless



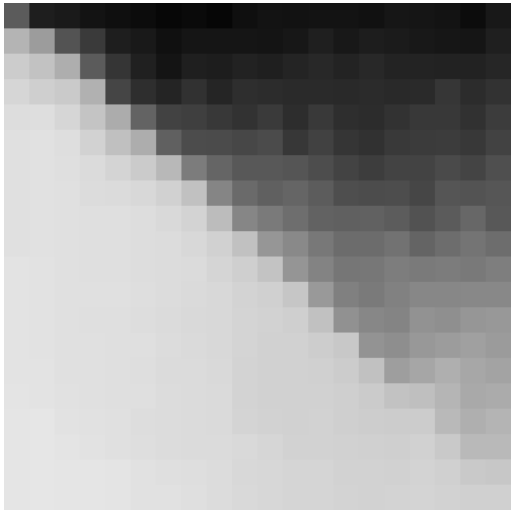


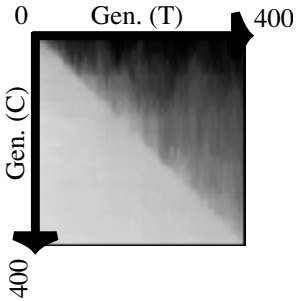
Bounded





Sampled





Memory Size Can Affect Interpretation