# Analysis of Adaptive Operator Selection Techniques on the Royal Road and Long k-Path Problems

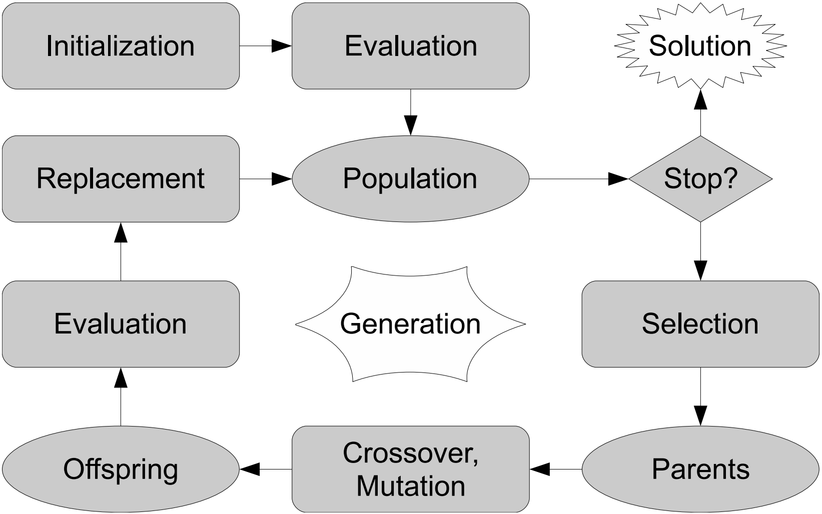
Álvaro Fialho[1], Marc Schoenauer[1,2], Michèle Sebag[1,2]

[1]Microsoft Research – INRIA Joint Centre
Orsay, France

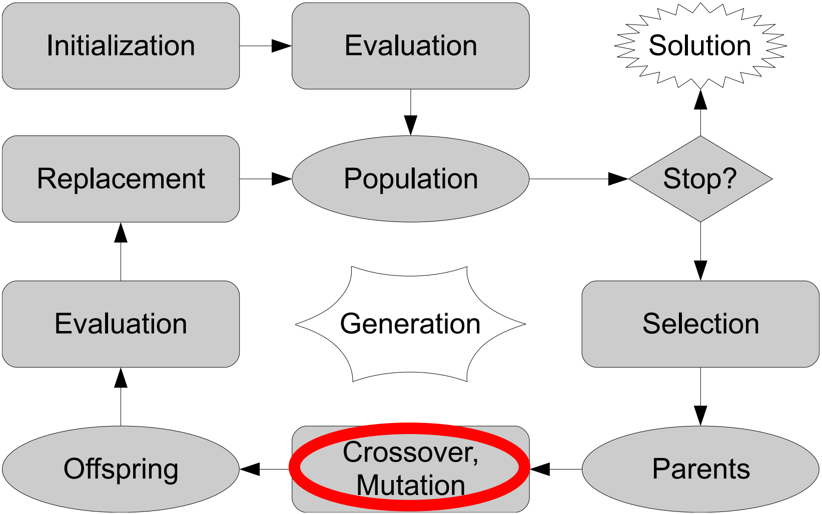
[2]Project-Team TAO, LRI / INRIA Saclay - Île-de-France
Orsay, France

GECCO 2009, Montreal, Canada
July 10th, 2009

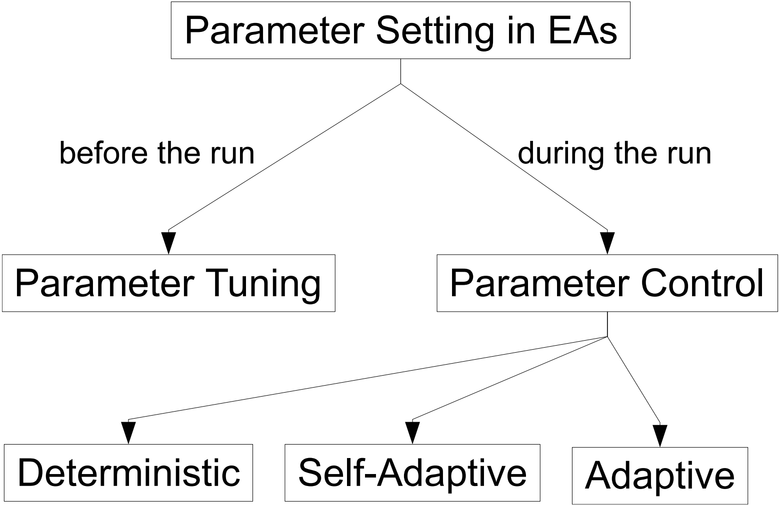
# Evolutionary Algorithms



# Evolutionary Algorithms

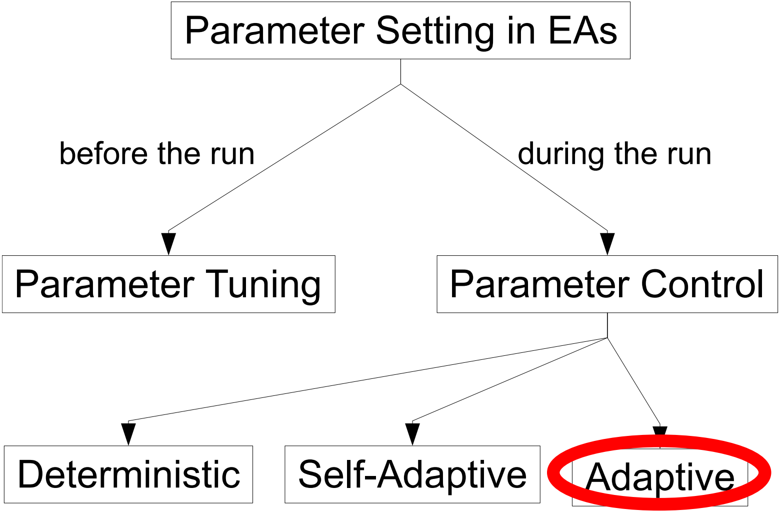


# Parameter Setting in Evolutionary Algorithms




(from [Eiben et al., 2007])

# Parameter Setting in Evolutionary Algorithms

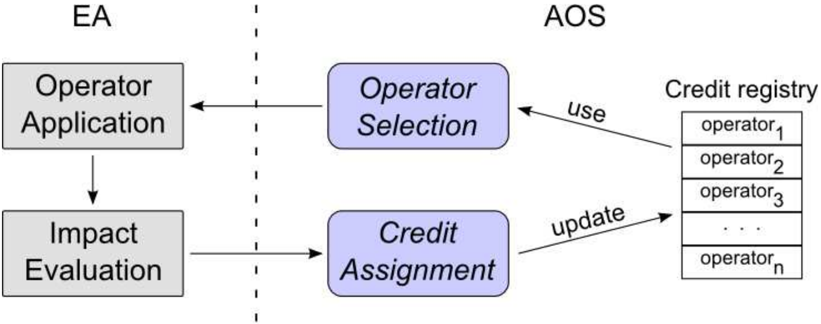



(from [Eiben et al., 2007])

# Adaptive Operator Selection

## Objective

Autonomously select the operator to be applied amongst available ones, based on its impact in the past.



# Previous Empirical Analyses

- Proposed Approach: Extreme - DMAB [Fialho et al., 2008]
  - Operator Selection: Dynamic Multi-Armed Bandits
  - Credit Assignment: Extreme Value Based
- Unimodal problems: OneMax and Long k-Path
- (1+50)-EA, selecting between mutation operators
- Extreme - DMAB achieved similar performance w.r.t. Optimal

# Previous Empirical Analyses

- Proposed Approach: Extreme - DMAB [Fialho et al., 2008]
  - Operator Selection: Dynamic Multi-Armed Bandits
  - Credit Assignment: Extreme Value Based
- Unimodal problems: OneMax and Long k-Path
- (1+50)-EA, selecting between mutation operators
- Extreme - DMAB achieved similar performance w.r.t. Optimal

**Q:** Is "Extreme - DMAB" really the best option?

- Different Operator Selection techniques
- Different Credit Assignment mechanisms
- Long K-Path and Royal Road problems

# Op. Selection: A (kind of) Multi-Armed Bandit problem

## Multi-Armed Bandits (MAB)

At time $t$, *gambler* plays arm $j$

$$\textit{reward at } t \; : r_t = \begin{cases} 1 & \text{with prob } = p_j \\ 0 & \text{with prob } = 1 - p_j \end{cases}$$

Goal: maximize cumulated reward

# Op. Selection: A (kind of) Multi-Armed Bandit problem

## Multi-Armed Bandits (MAB)

At time $t$, *gambler* plays arm $j$

$$\textit{reward at } t \; : r_t = \begin{cases} 1 & \text{with prob } = p_j \\ 0 & \text{with prob } = 1 - p_j \end{cases}$$

Goal: maximize cumulated reward

## MAB state-of-the-art: UCB1 [Auer et al., 2002]

- *Be optimistic in face of the unknown*
- At time $t$, choose arm $j$ maximizing:

$$\hat{r}_{j,t} + \sqrt{\frac{2 \log \sum_k n_{k,t}}{n_{j,t}}}, \text{where} \begin{cases} \hat{r}_{j,t} & \text{,estimated reward for arm } j \\ n_{j,t} & \text{,chosen times for arm } j \end{cases}$$

# Op. Selection with Multi-Armed Bandits: the true story

## *Exploration vs. Exploitation* (EvE) balance

- MAB framework: $\hat{r}_{j,t} \in [0, 1]$;
- AOS framework: $\hat{r}_{j,t} \in [a, b]$ (e.g. fitness improv., diversity)
- UCB1's EvE balance is broken, **Scaling** is needed:

$$\hat{q}_{i,t} = \hat{r}_{j,t} + \mathcal{C}\sqrt{\frac{2 \log \sum_k n_{k,t}}{n_{j,t}}}$$

# Op. Selection with Multi-Armed Bandits: the true story

## *Exploration vs. Exploitation* (EvE) balance

- MAB framework: $\hat{r}_{j,t} \in [0,1]$;
- AOS framework: $\hat{r}_{j,t} \in [a,b]$ (e.g. fitness improv., diversity)
- UCB1's EvE balance is broken, **Scaling** is needed:

$$\hat{q}_{i,t} = \hat{r}_{j,t} + C\sqrt{\frac{2\log \sum_k n_{k,t}}{n_{j,t}}}$$

## Operator Selection is a dynamic context

- UCB1: too long to recover.
- Page-Hinkley test [Page, 1954]
    - Detect change in the rewards distribution (threshold $\gamma$);
    - Restart the MAB from scratch.

# Op. Selection with Multi-Armed Bandits: the true story

## *Exploration vs. Exploitation* (EvE) balance

- MAB framework: $\hat{r}_{j,t} \in [0,1]$;
- AOS framework: $\hat{r}_{j,t} \in [a,b]$ (e.g. fitness improv., diversity)
- UCB1's EvE balance is broken, **Scaling** is needed:

$$\hat{q}_{i,t} = \hat{r}_{j,t} + \mathcal{C}\sqrt{\frac{2\log\sum_k n_{k,t}}{n_{j,t}}}$$

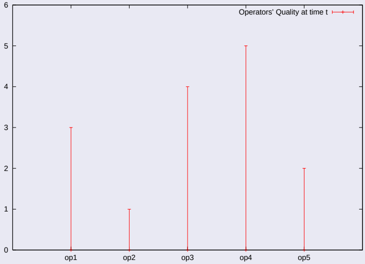
## Operator Selection is a dynamic context

- UCB1: too long to recover.
- Page-Hinkley test [Page, 1954]
  - Detect change in the rewards distribution (threshold $\gamma$);
  - Restart the MAB from scratch.

## Dynamic MAB: UCB1 + Scaling + Page-Hinkley

# Op. Selection: Probability Matching and Adaptive Pursuit

- $\hat{Q}_{j,t}$ = estimate of reward of operator $j$, for time $t$

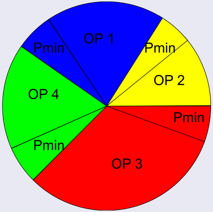



- At time $t$, op. $j$ is selected according to $s_j$; gets reward $r_{j,t}$;
- Its "empirical quality" is updated by relaxation:

$$\hat{Q}_{j,t+1} = (1-\alpha)\hat{Q}_{j,t} + \alpha\, r_{j,t}$$

# Op. Selection: Probability Matching and Adaptive Pursuit

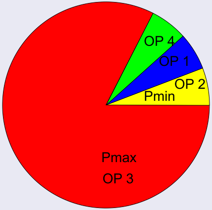
## Probability Matching




$s_i$ proportional to $Q_i$

$$s_{i,t+1} = p_{min} + (1 - K * p_{min}) \frac{\hat{Q}_{i,t+1}}{\sum_{j=1}^{K} \hat{Q}_{j,t+1}}$$

## Adaptive Pursuit




$s_{i^*}$ tends to $p_{max}$; others tend to $p_{min}$

$$\begin{aligned} i^* &= argmax\{\hat{Q}_{i,t}, i = 1 \ldots K\} \\ s_{i^*,t+1} &= s_{i^*,t} + \beta\left(p_{max} - s_{i^*,t}\right), \\ s_{i,t+1} &= s_{i,t} + \beta\left(p_{min} - s_{i,t}\right), \text{ for } i \neq i^* \end{aligned}$$

# Operator Selection techniques

## "Own" approaches

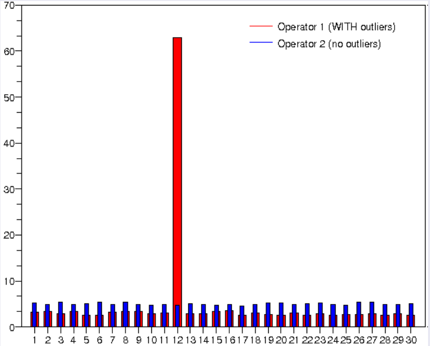
- Multi-Armed Bandits (MAB)
- Dynamic Multi-Armed Bandits (DMAB) [Fialho et al., 2008]

## Baseline approaches

- Probability Matching (PM)
- Adaptive Pursuit (AP) [Thierens, 2007]

# Credit Assignment: Extreme Value-Based (EVB)

- Outlier operators are rarely considered - smaller expectation.




- EC: Focus on extreme, rather than average events
  - Complex systems, e.g. epidemic propagation, financial markets
- Extreme Value-Based (EVB) Credit Assignment.
  - $\mathcal{R}$ = Extreme value over a Window

# Credit Assignment approaches

- Extreme versus usual Average rewards over a window

# Credit Assignment approaches

- Extreme versus usual Average rewards over a window

- Unimodal problems: fitness improvement
  - Variance of the credits reduces as the search goes on
- Alternative: normalize w.r.t. to the current higher one
  - $\mathcal{R} \in [0, 1]$

# Credit Assignment approaches

- Extreme versus usual Average rewards over a window

- Unimodal problems: fitness improvement
  - Variance of the credits reduces as the search goes on
- Alternative: normalize w.r.t. to the current higher one
  - $\mathcal{R} \in [0, 1]$

**To be coupled with the Operator Selection techniques**

- Extreme / Absolute [Fialho et al., 2008]
- Extreme / Normalized
- Average / Absolute
- Average / Normalized

# Meta-Parameters

- Operator selection:
  - MAB and DMAB: $\mathcal{C}$, for scaling ;
  - DMAB: $\gamma$ for the PH test'
  - AP and PM: learning rate $\alpha$
  - AP: adaptation rate $\beta$
- Credit assignment (EVB):
  - $\mathcal{W}$, the size of the sliding window

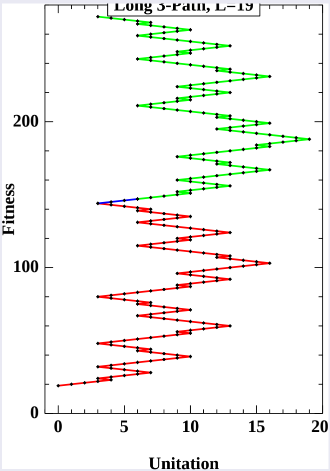
# Meta-Parameters

- Operator selection:
  - MAB and DMAB: $\mathcal{C}$, for scaling ;
  - DMAB: $\gamma$ for the PH test'
  - AP and PM: learning rate $\alpha$
  - AP: adaptation rate $\beta$
- Credit assignment (EVB):
  - $\mathcal{W}$, the size of the sliding window

- Tuned off-line by F-RACE [Birattari et al., 2002]
  - 11 initial runs, up to 50 runs or 1 configuration left
  - Friedman's test at 95% confidence level
- Comparison: each AOS combination with its best meta-parameters configuration found

# The Long k-Path Problem

- Unimodal, very long path to the optimum
  - Path length grows exponentially with size of the bitstring ($n$)
  - Hamming distance between two consecutive points is 1
  - Any other point at distance 1 is off the path
- "Shortcuts" require at least $k$ bit-flips
- (1+50)-EA: 1-bit, 3-bit, 5-bit, 1/n bit-flip, 3/n bit-flip mutations



# Comparative Results on the Long K-Path Problem I

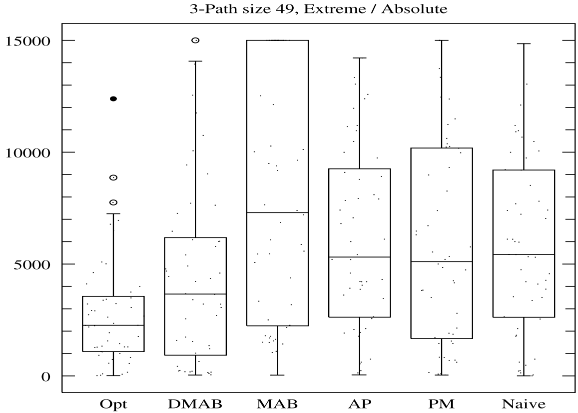



Figure: AOS performances on *Long k-Path*, using the EVB Credit Assignment (fitness improv.). From left to right: optimal strategy, DMAB, MAB, Adapt.Pursuit, Prob.Match. and the naive uniform sel.

# Comparative Results on the Long K-Path Problem II

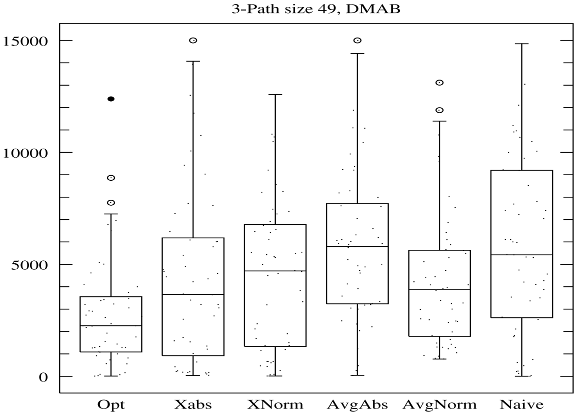



Figure: AOS performances on *Long k-Path*, using the DMAB Op. Selection. From left to right: optimal sel., ExtremeAbs, ExtremeNorm, AverageAbs and AverageNorm credit assign., and the naive uniform sel.

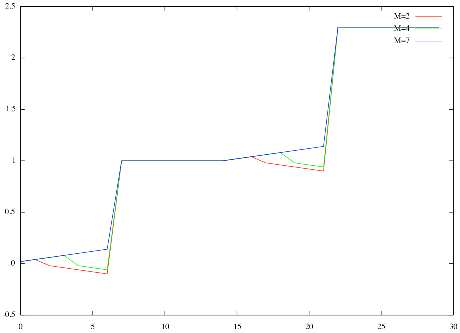
# The Royal Road Problem

- Easy for a GA, difficult for other optimization methods
- By exploring the building block hypothesis

| | |
|---|---|
| Level 4: | $\{B_0, B_1, ..., B_{15}\}$ |
| Level 3: | $\{B_0, B_1, ..., B_7\}, \{B_8, B_9, ..., B_{15}\}$ |
| Level 2: | $\{B_0, B_1, B_2, B_3\}, \{B_4, B_5, B_6, B_7\}, ..., \{B_{12}, B_{13}, B_{14}, B_{15}\}$ |
| Level 1: | $\{B_0, B_1\}, \{B_2, B_3\}, \{B_4, B_5\}, ..., \{B_{10}, B_{11}\}, \{B_{12}, B_{13}\}, \{B_{14}, B_{15}\}$ |
| Level 0: | $\{B_0\}, \{B_1\}, \{B_2\}, \{B_3\}, \{B_4\}, \{B_5\}, \{B_6\}, ..., \{B_{13}\}, \{B_{14}\}, \{B_{15}\}$ |

# The Royal Road Problem

- $2^k$ non-overlapping sections with length $b + g$
- Just the *b*lock is considered by the fitness, *g*ap doesn't matter
- Default settings [Holland, 1993]
  - Structure: $k = 4, b = 8, g = 7$ (240 bits)
  - Fitness: $m = 4, v = .02, u^* = 1, u = .3$




- (100, 100)-EA with weak elitism
- Operators
  - 1-pt Crossover
  - 2-pt Crossover
  - 4-pt Crossover
  - Unif. Crossover
  - 1/30 Mutation

# Comparative Results on the Royal Road I

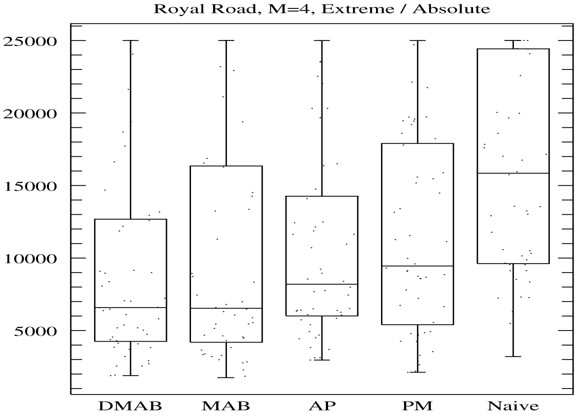



Figure: AOS performances on *Royal Road*, using the EVB Credit Assignment (fitness improv.). From left to right: optimal strategy, DMAB, MAB, Adapt.Pursuit, Prob.Match. and the naive uniform sel.

# Comparative Results on the Royal Road II

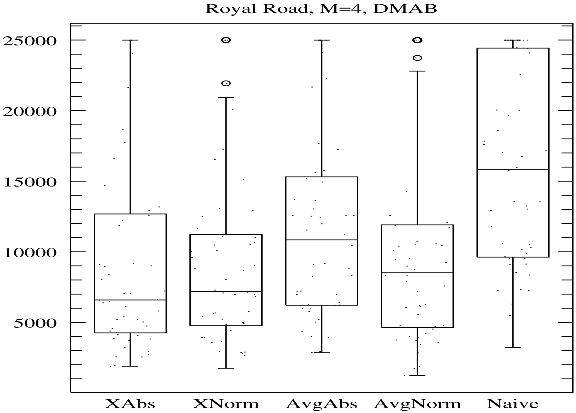



Figure: AOS performances on *Royal Road*, using the DMAB Op. Selection. From left to right: optimal sel., ExtremeAbs, ExtremeNorm, AverageAbs and AverageNorm credit assign., and the naive uniform sel.

# Conclusions and Perspectives

- AOS = Credit Assignment (CA) + Operator Selection (OS)
  - CA: Extreme Value Based rewards
  - OS: Dynamic Multi-Armed Bandits
- Showed to be efficient on 2 benchmark unimodal problems
- Off-line tuning of meta-parameters is still expensive.
- Real problems, optimal behavior is not know.
- Better than fixed, naive and known adaptive approaches.

# Conclusions and Perspectives

- AOS = Credit Assignment (CA) + Operator Selection (OS)
  - CA: Extreme Value Based rewards
  - OS: Dynamic Multi-Armed Bandits
- Showed to be efficient on 2 benchmark unimodal problems
- Off-line tuning of meta-parameters is still expensive.
- Real problems, optimal behavior is not know.
- Better than fixed, naive and known adaptive approaches.

## Near future

- High sensitivity: DMAB's $\mathcal{C}$ and $\lambda$ affect the EvE balance.
- Generalization: Rank-based rewarding
- Different multimodal benchmark problems
- Selection between heuristics instead of operators (HH)

# References I


Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002).
Finite-time analysis of the multiarmed bandit problem.
*Machine Learning*, 47(2/3):235–256.

Birattari, M., Stützle, T., Paquete, L., and Varrentrapp, K. (2002).
A racing algorithm for configuring metaheuristics.
In *Proc. GECCO'02*.

Eiben, A. E., Michalewicz, Z., Schoenauer, M., and Smith, J. E. (2007).
Parameter Control in Evolutionary Algorithms.
In *Parameter Setting in Evolutionary Algorithms*. Springer.

Fialho, A., Da Costa, L., Schoenauer, M., and Sebag, M. (2008).
Extreme value based adaptive operator selection.
In *Proc. PPSN'08*.

Holland, J. H. (1993).
Royal road functions.
In *Internet Genetic Algorithms Digest 7:22*. Massachusetts Institute of Technology.

# References II

Page, E. (1954).
Continuous inspection schemes.
*Biometrika*, 41:100–115.

Thierens, D. (2007).
Adaptive Strategies for Operator Allocation.
In Lobo, F., Lima, C., and Michalewicz, Z., editors, *Parameter Setting in Evolutionary Algorithms*, pages 77–90. Springer Verlag.

# Analysis of Adaptive Operator Selection Techniques on the Royal Road and Long k-Path Problems

Álvaro Fialho[1], Marc Schoenauer[1,2], Michèle Sebag[1,2]

[1]Microsoft Research – INRIA Joint Centre
Orsay, France

[2]Project-Team TAO, LRI / INRIA Saclay - Île-de-France
Orsay, France

GECCO 2009, Montreal, Canada
July 10th, 2009