# Activatability for tractable simulations of NP problems

## -- Application to Ecology --

Patrick Coquillard*, Alexandre Muzy**, Eric Wajnberg*

* Behavioral & Molecular Ecology, CNRS-INRA-Nice Univ., Sophia-Antipolis (Fr)

** UMR CNRS LISA University of Corsica, Corte (Fr).

- Powerful simulators involve space and time (biology/ecology, physics, economics, etc.)
- Limitations and variability of experiments & models require stochasticity
- Reduce state spaces? How? Why?

# Prisoner's dilemma

**1. The non spatial prisoner's dilemma**

- **If the other prisoner testifies:**
  If I remain silent, I will receive the full 10-years sentence;
  But if I testify, I will only receive a 5-years sentence

- **If he does not :**
  If I remain silent, I will receive a 6-monthes sentence;
  But if I testify, I will be free.

**«*Whatever his choice, I have interest to testify*»**

| | Silent | Testify |
|---|---|---|
| Silent | (-0.5,-0.5) | (-10,0) |
| Testify | (0,-10) | (-5,-5) |

**=> Non stable strategy, which converges to the disappearance of cooperators**

Temptation > Cooperation > Punishment > trickery: (0 > -0,5 > -5 > -10)

**2. The spatialized dilemma** b > 1

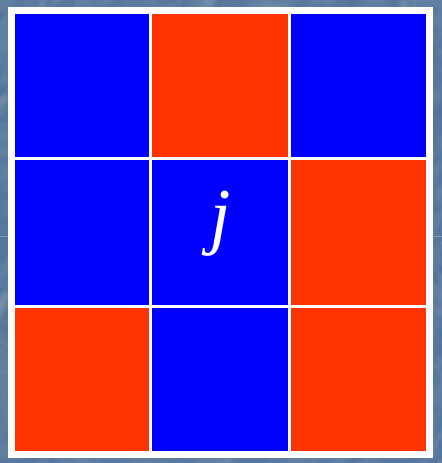
| | Silent | Testify |
|---|---|---|
| Silent | (1,1) | (b,0) |
| Testify | (0,b) | (0,0) |

## Example of spatialization necessity (2)

Evolution rules of spatialized dilemma (M. Nowak et R. May, 92-95)

$$P(j) = \sum_{i \in N_j} A_i^m s_i \Big/ \sum_{i \in N_j} A_i^m$$ = probability that *j* switches to cooperator state

- With $\begin{cases} s_i = 0 \text{ (selfish ■)} \\ s_i = 1 \text{ (cooperator ■)} \end{cases}$
- $N_j$ = Set of neighbors of j
- $A_i$ = Gains of cell i at time *t*

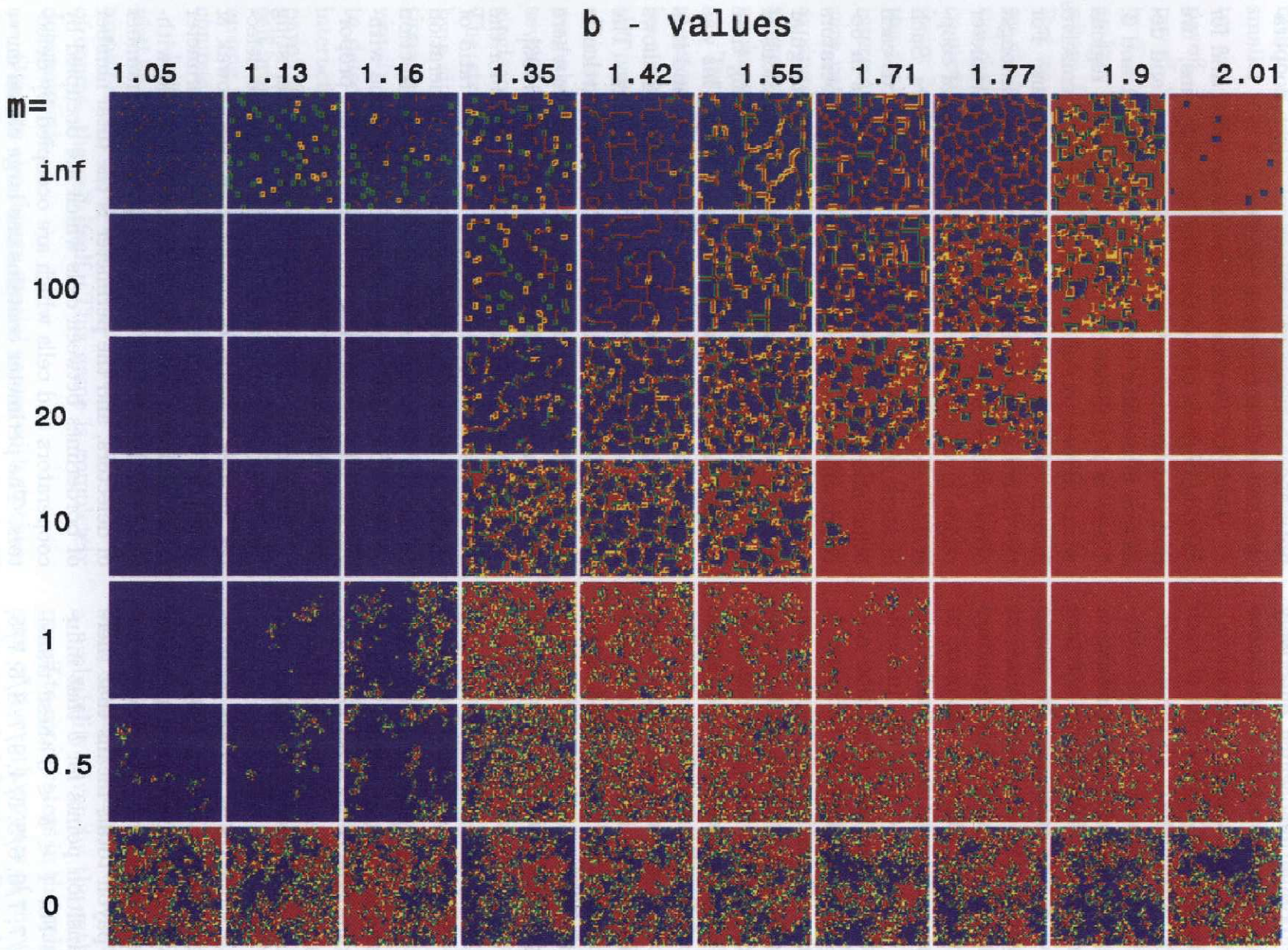



$m = 1$, original stochastic game

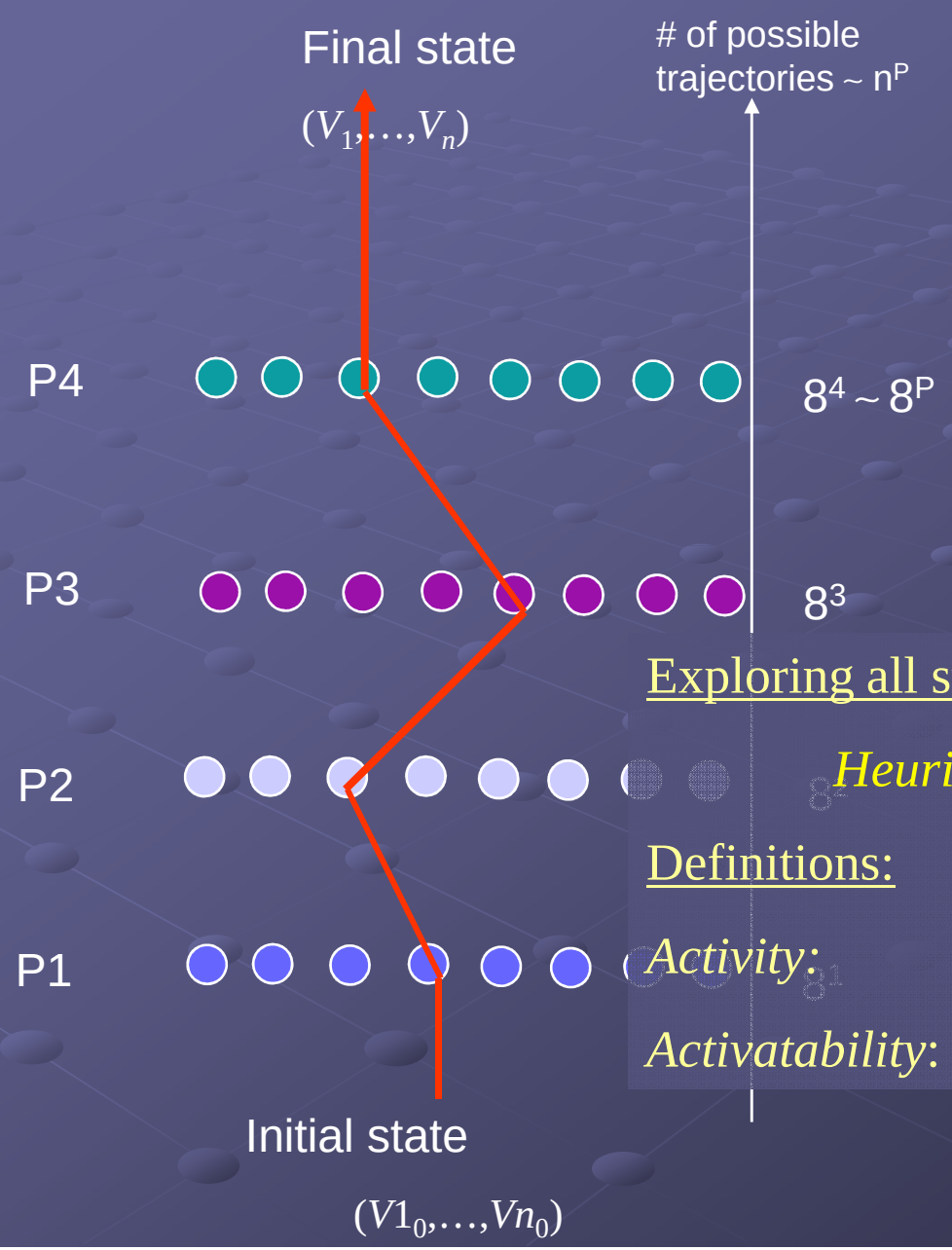
$m \to \infty$, the more the cell *j* will get a chance to adopt the more profitable strategy of its neighbors (deterministic game)

$m = 0$, the gains are worthless, neutral derivate between both strategies

# The 2 strategies can then coexist for subsets of values of m and b



## Finding a particular solution belongs to the NP class problems




The time needed to compute all possible trajectories is $O(n^p)$

<u>Exploring all states of a base model is intractable</u> :

*Heuristics need to be experimented*

<u>Definitions:</u>

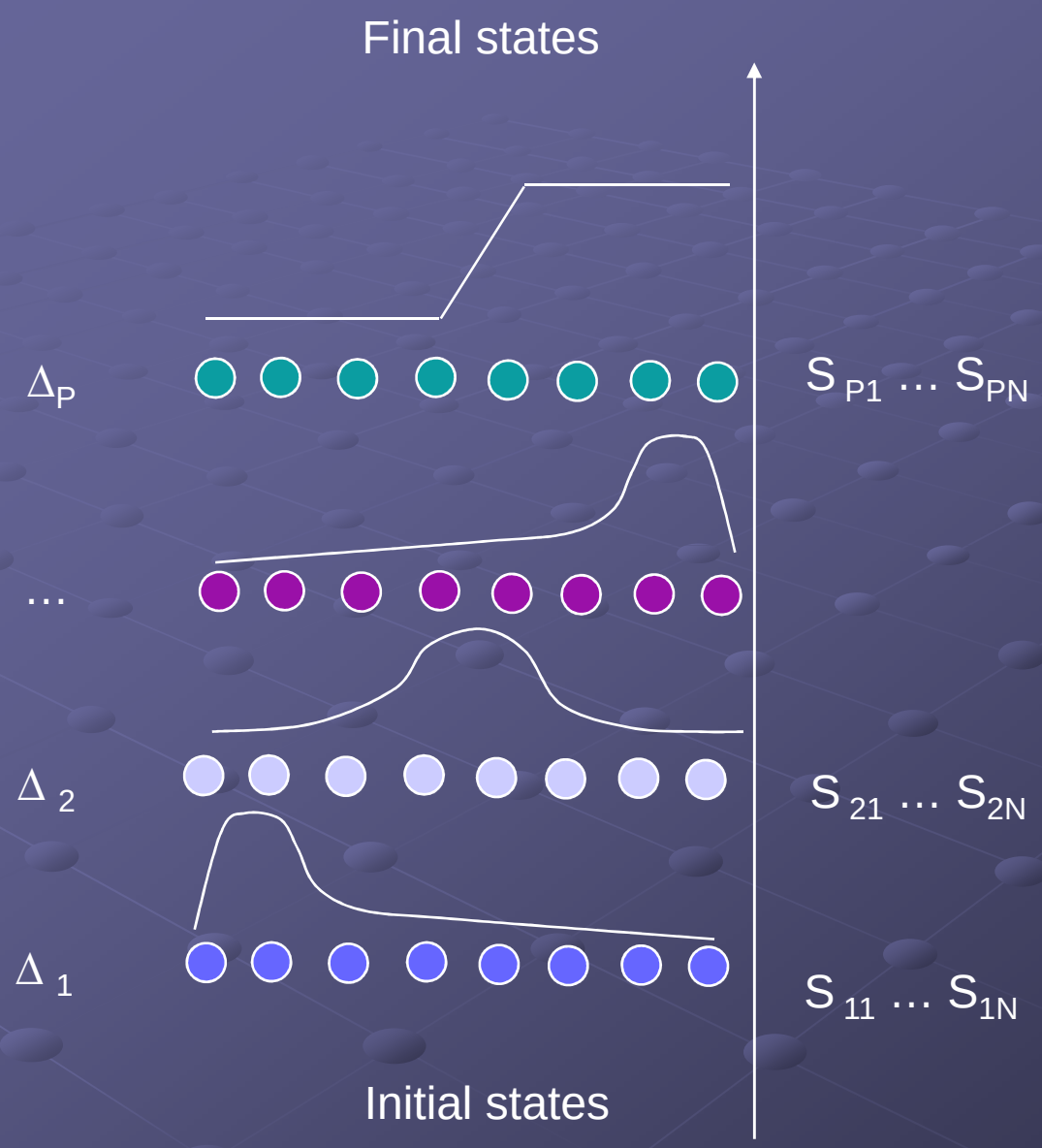
*Activity*: Number of transitions

*Activatability*: Probability of transition activation

# Pseudorandom generator and random variables [TMS76]

- Ideal generator hypothesis:
  - Uniformly distributed over [0,1]: $\Gamma : [0,1] \rightarrow [0,1]$
  - Deterministic: $r_{i+1} = \Gamma(r_i)$
- Random sampled variable:
  - Autonomous sequential machine: $\delta(s) = \pi^{-1}(r_{i+1})$
  - With *π(.)* a distribution law.
- Simulation: Monte Carlo, Student…

## Activity and processes




We assume a computational sequential machine: $< \Delta, S >$,

With : $\Delta = x\Delta_i$, $S = xS_i$

$\Delta_1(s_{11}, \ldots, s_{1N})$

…

$\Delta_P(s_{P1}, \ldots, s_{PN})$

*Activatability: $\pi(.)$*

$\Delta_1[\pi_1(s_{11}, \ldots, s_{1N})]$

…

$\Delta_P[\pi_N(s_{P1}, \ldots, s_{PN})]$

# Activity measures

- Find a particular trajectory (*i.e. maximal activity* $O(n^P)$) belongs to the NP complexity class problem

  ⇨ *Reformulate the question : find the most probable trajectory*

- Using activatability: *Activity* $\cong f(R, \%, P)$
  - Where R is the number of replications, % the confidence interval, and P the number of processes
  - The number of replications depends on $\sigma$ (standard error of the response) and %.

# Impact of activatability

- Kleijnen test (1992 : activating processes or not?) Experimental protocol design (type $2(_{II,IV,V})^{n-p}$) can be used.

- Test the effect (using GLM because of possible non Gaussian responses) of activating the $\Delta_i$ process on the values of the final states.

$$y_i \sim \Delta i + .. + \Delta j + \ldots + \Delta j\ \Delta i + \ldots + \varepsilon$$

# Proposal

## ACTIVATABILITY CYCLE SEARCH SPACE

Select activatable potentially pertinent processes

- Build an experimental design
  $2(_{II,IV,V})^{n-p}$, $p \in \{0,\ldots,k\}$
- Activate selected processes
  Using $\Delta_i\,[\pi N(s\,i_1\,,\ldots,\,si_N)]$

Run simulator

Test the effects (using GLM) and reject inefficient processes

$y_i \sim \Delta i + .. + \Delta j + \ldots + \Delta i\,\Delta j + \ldots + \varepsilon$

Run simulator

# (small) Application

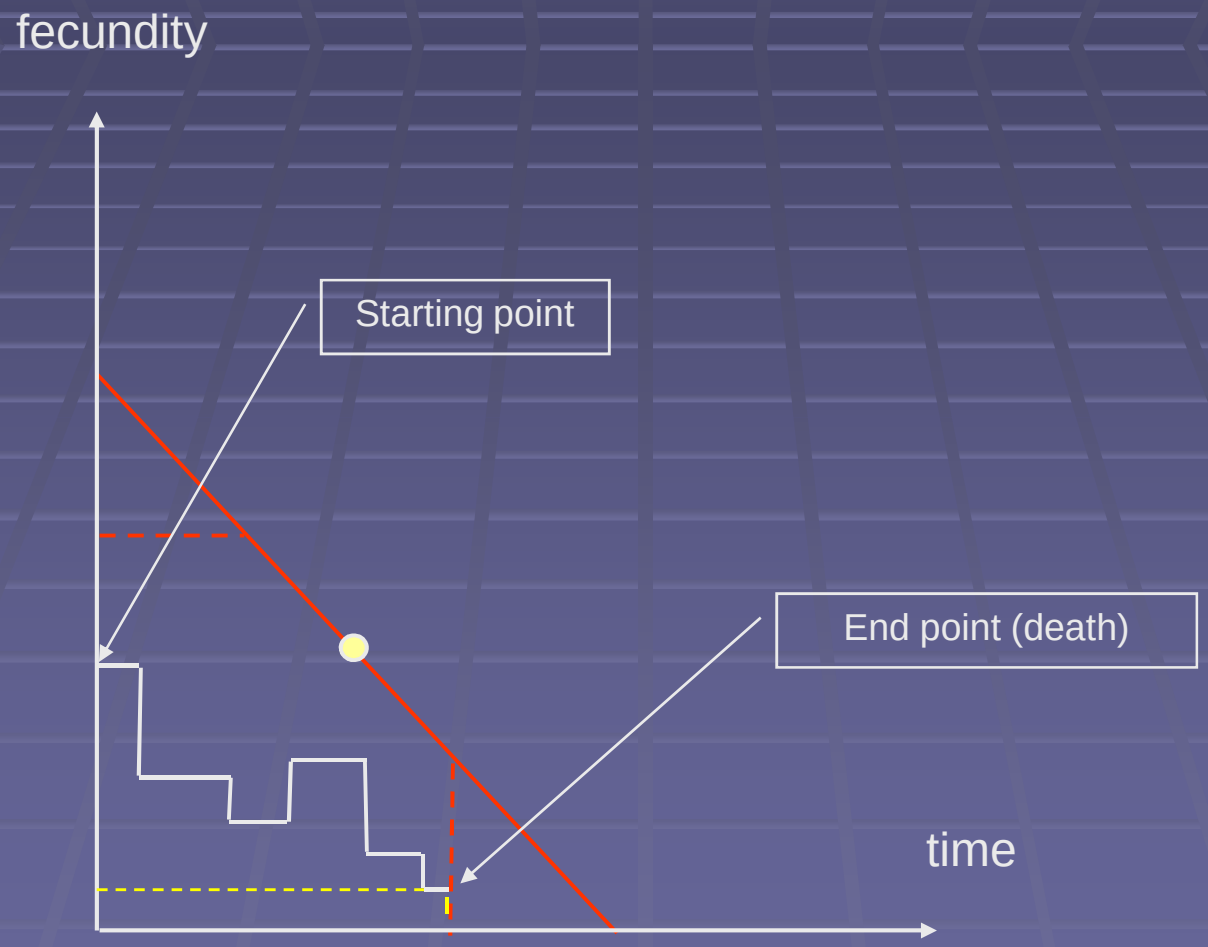
Search of an optimal foraging strategy  of a parasitoid wasp

Energ

cost

The wasp can move along the trade-off
within the range using (G3)
And starts from one of the two ends of the
range (G4)

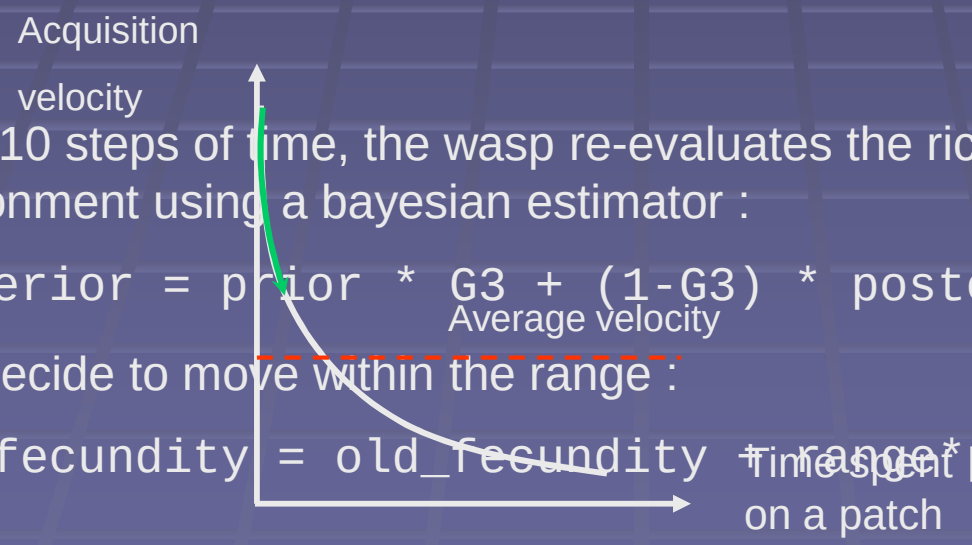
During its life, the wasp moves along its trade-off as a function of her estimation of the resource richness of the environment it is exploring



- When the wasp encounters a patch of eggs, the number *N*(*t*) of eggs laid follows :

$$N(t) = N_0 \exp(1 - \exp(\frac{-kt}{N_0}))$$

- The wasp leaves the patch according to the *marginal value theorem*, i.e. when its laying velocity reach the average velocity obtained in the environment.




- Each 10 steps of time, the wasp re-evaluates the richness of the environment using a bayesian estimator :

=> `Posterior = prior * G3 + (1-G3) * posterior;`

And decide to move within the range :

=> `New_fecundity = old_fecundity + range* posterior;`

- Finally, we have 4 stochastic processes embedding G1,..,G4 coded as genes hold by the C++ Galib (*Matthew Wall* (M.I.T.), 2007) and encapsulated in one single chromosome. Each wasp hold one chromosome.

```
GABin2DecPhenotype map;

map.add(16, 0, MAX_LIFE_TIME);  //G1 : abscissa of the point
map.add(16, 0, RANGE_MAX);      //G2 :  range around G1
map.add(16, 0, 1);              //G3 :  Bayesian parameter
map.add(16, 0, 1);              //G4 :  moving direction

GABin2DecGenome genome(map, objective); //genome structure and
                                        // function to optimize
GAIncrementalGA ga(genome);             // genetic algorithm
```

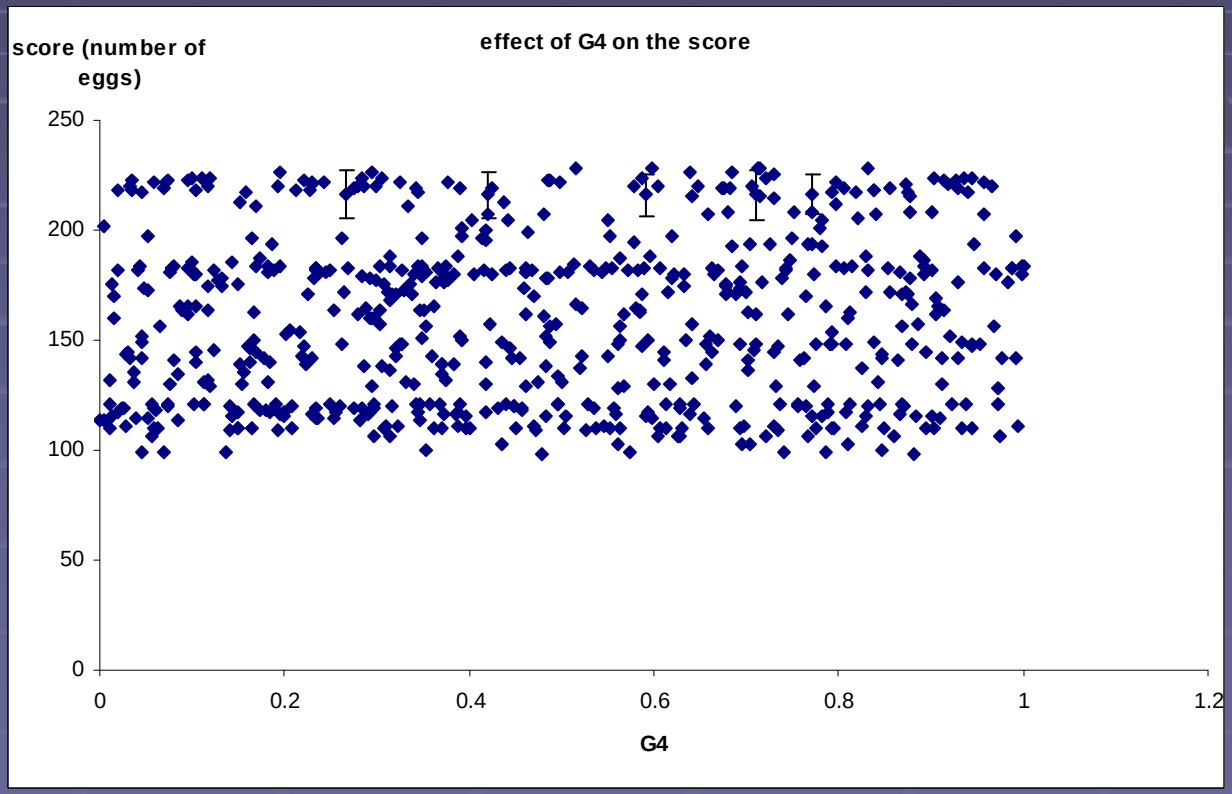
- The objective consists in find the vector {G1,..,G4} which maximises the score of the wasp, *i.e.* which maximises the number of eggs laid throughout its life.

- In this application we used a complete design protocol (all combinations of {G1,..,G4} explored); 2625 combinations of 5 parameters tested.

# Result of the GLM analysis (1)

No effect on G4 was detected whatever the variables and interactions between variables tested (p-value required : $p < 0.001$).

There was no effect of G4 on scores



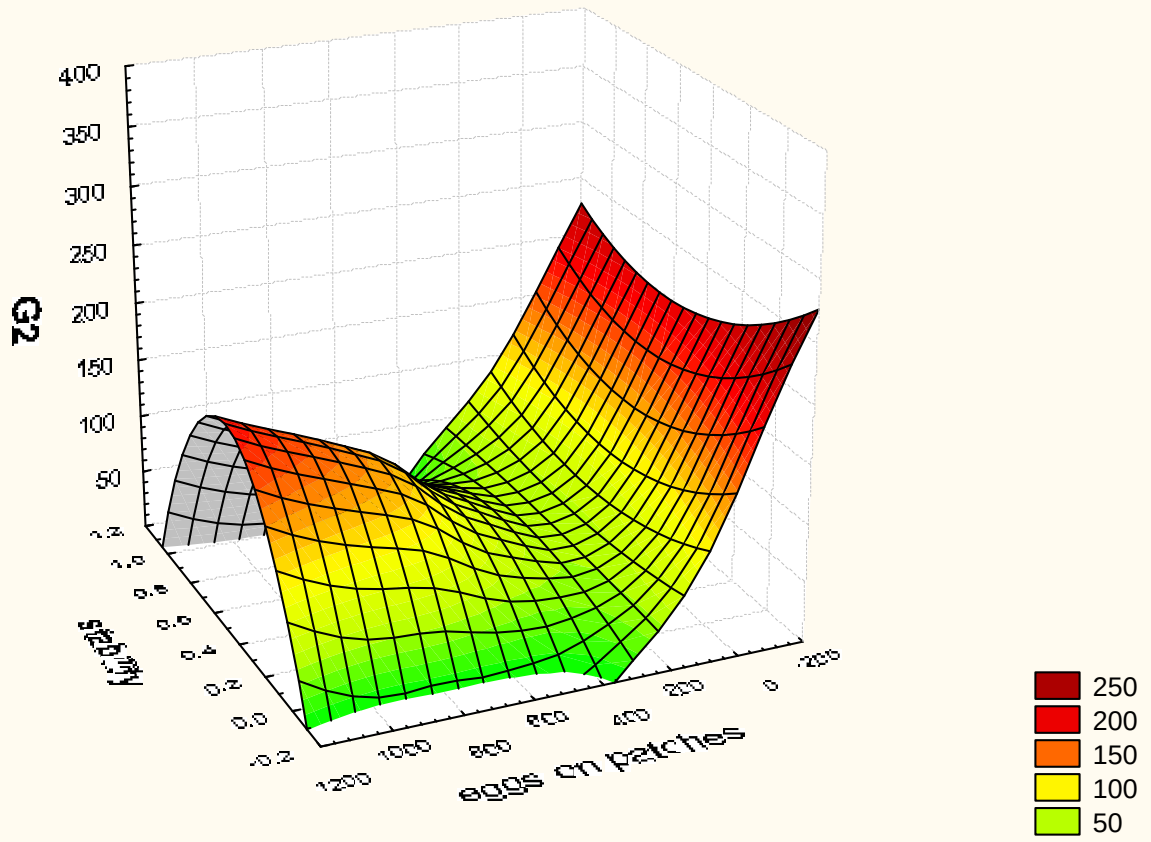
# Result of the GLM analysis (2)

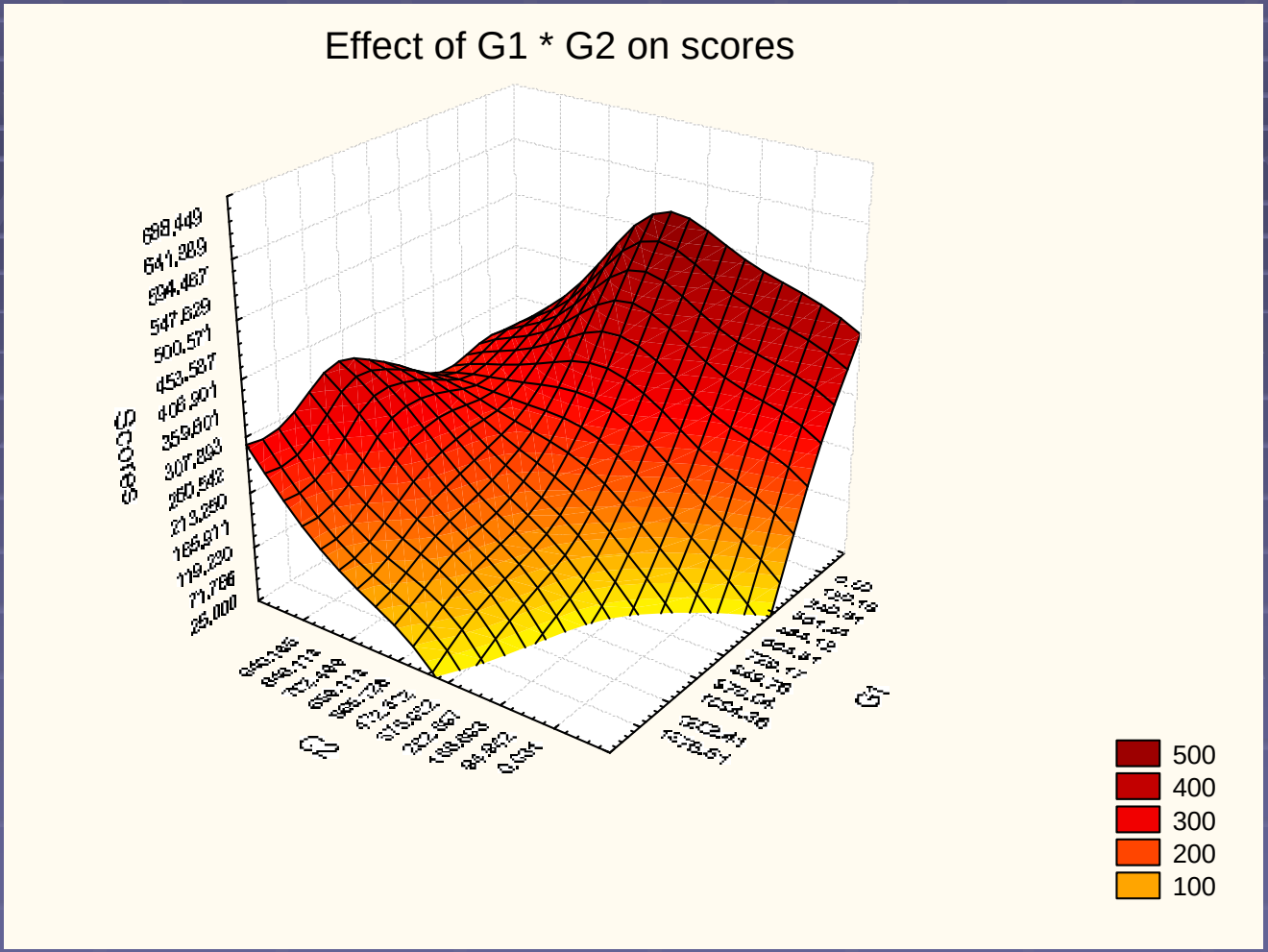
The G4 process was inactivated and the simulator restarted.

Huge and medium effects were found on each of the resting {G1,..,G3} processes

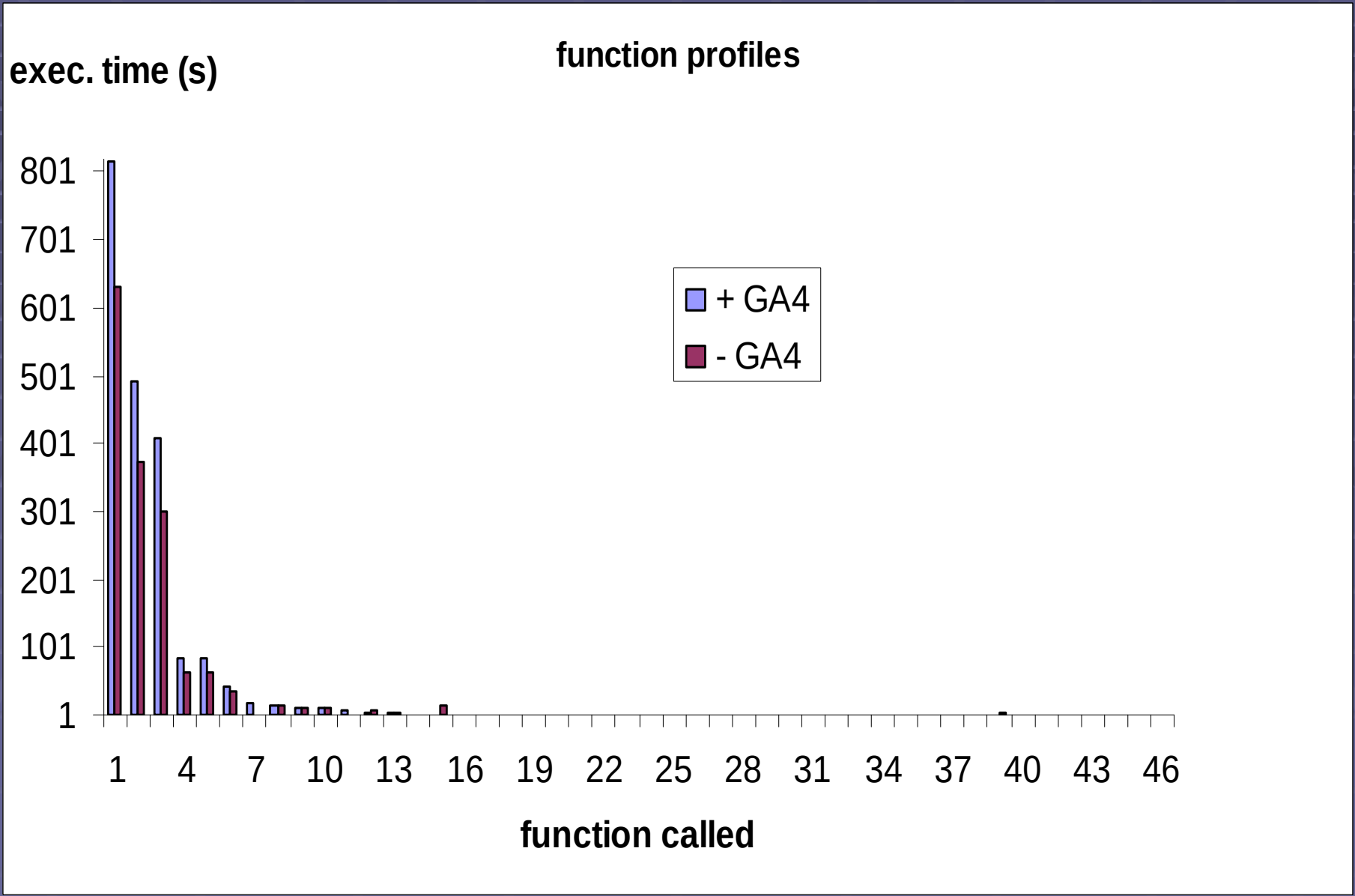
| Source of variation | F e | G1 |
|---|---|---|
| Stability (1) | 5 | 0.0852 |
| travel time ( | 2 | 0.1427 |
| 1 * 2 | 5 | 0.3962 |
| Cost (3) | 2 | <.0001 |
| 1 * 3 | 5 | 0.16 |
| 2 * 3 | 3 | 0.5389 |
| eggs (4) | 6 | 0.0005 |
| 1 * 4 | 3 | 0.6958 |
| 2 * 4 | 8 | 0.0209 |
| 3 * 4 | 1 | 0.7266 |
| stoch.(5) | 6 | 0.6331 |
| 1 * 5 | 3 | 0.9653 |
| 2 * 5 | 4 | 0.4054 |
| 3 *5 | 3 | 0.966 |
| 4*5 | 7 | 0.3829 |

effect of the interaction eggs * stability on G2




Effect of G1 * G2 on scores
SCORES
G2
G1
500
400
300
200
100

# Running execution time



# Conclusion

- The proposal is automatable
- Allows embedding into the simulator all the processes with no *a priori* exclusion
- Processes are activated according to their statistical effects on results

Drawbacks and troubles

- Reformulating the question, one loses the prediction abilities (statistical results) but recognize the dimension of complexity in the scientific explanation.
- The selection must be a conservative operation (the internal coherence must be preserved)
- The necessity of replicates strongly diminishes the information about spatial results. Thus, trajectories of particular interest cannot be identified.