



# A Distributed Framework for Semi-Automatically Developing Architectures of Brain and Mind

---

Fernand Gobet  
Brunel University

Peter Lane  
University of Hertfordshire



# Overview

---

- The problem of information overload in science
- Computational modelling
- Using the grid for pooling together information from
  - Empirical researchers
  - Modellers
  - Theorists
- Evolutionary scientific discovery
- Future work



# Information Overload

---

- Science aims to develop theories that summarise and unify a large body of empirical data
- About 1,500 journals (mostly experimental) devoted to scientific psychology
- How to cope with information overload?
- Databases (partly) solve this problem
- But accumulation of knowledge is not necessarily the same as scientific understanding



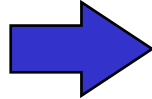
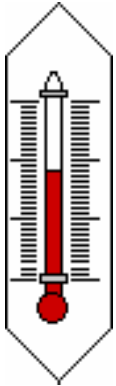
# Computational Modelling

---

- Given the amount and complexity of data, theories expressed as computer models are necessary
- Models can
  - Specify the processes carried out
  - Explain both quantitative and qualitative data
- Current difficulties
  - Access to empirical data may be difficult for the modeller
  - Lack of standard techniques for comparing models and optimising them

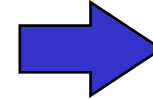
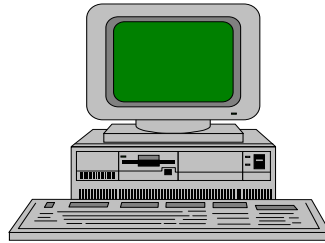
**Initial conditions:**

- temperature
- air pressure, etc.



**Laws of physics**

*Computer simulations*

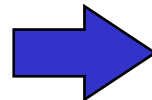


**Weather forecast**



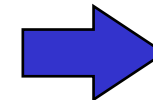
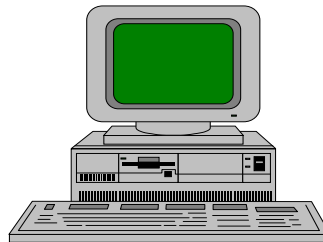
**Initial conditions:**

- emotional state
- environment, etc.

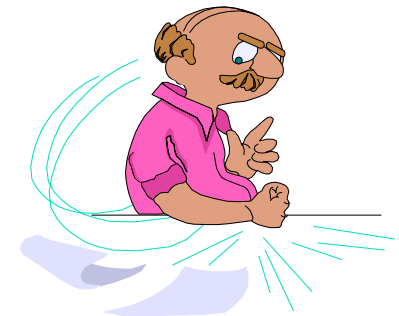


**Laws of psychology**

*Computer simulations*



**Prediction of  
behaviour**





# Computational Modelling

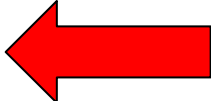
---

- Given the amount and complexity of data, theories expressed as computer models are necessary
- Models can
  - Specify the processes carried out
  - Explain both quantitative and qualitative data
- Current difficulties
  - Access to empirical data may be difficult for the modeller
  - Lack of standard techniques for comparing models and optimising them



# Evolutionary Scientific Discovery

---

- Three main roles
  - Empirical researcher → Empirical data
  - Computational modeller → Models 
  - Theorist → Generalisation from models
- Each of these roles can be supported by
  - On-line databases
  - Standardised data format
- Ideally, data, models and theories are made available on the Web
  - Can be inspected and used by all researchers
  - This is far from being the case in Psychology!



# Publishing Computational Models

Lane & Gobet (2003)

---

- Framework inspired from software engineering
- Models are published with
  - Detailed documentation
  - A suite of tests establishing that the model in fact does what is expected from it
  - A suite of experimental data and a description of how well (or how badly) the model explains them
- This allows other researchers to replicate the simulations, or re-implement the model



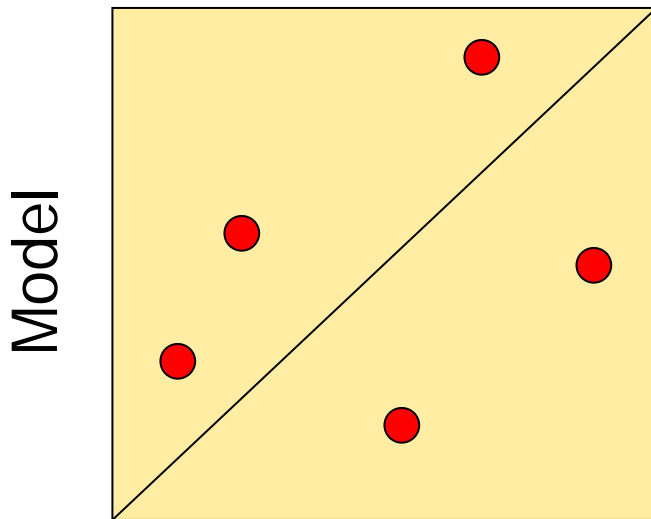
# Optimising Models

---

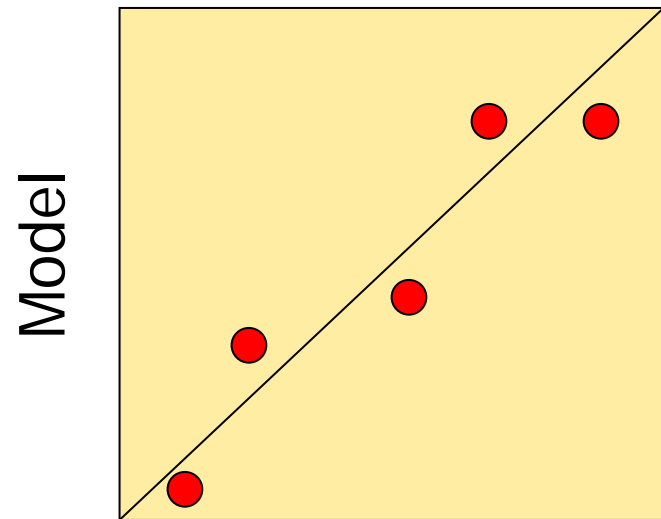
- No point in comparing models that have not been optimised
  - Differences may be due either to real features of the models or to lack of optimization
- Some techniques (e.g., genetic algorithms) have sometimes been used for this purpose (e.g., Ritter, 1991)
- However, they have been used only with one type of model fitted on one type of data
- Can we optimise different types of models with different sets of data simultaneously?

# The Basic Idea

- Models typically have free parameters
  - e.g., decay in short-term memory
  - e.g., weight of the links in a neural network
- Tuning these parameters can improve the fit



Data



Data



# Evolutionary Computation

---

- Evolutionary computation (e.g., genetic algorithms) can help the search for suitable parameters
- Large populations of genotypes (possible solutions) are evolved
  - The fittest solutions tend to survive and reproduce
  - Variability is obtained by crossover and mutation
  - Sets of parameters are encoded
  - Fitness function determines the quality of the solution
  - In our case, the fitness function is how well the model explains the data (e.g., variance explained)



# Multi-Objective Optimisation

---

- The goal is to evolve several classes of models using different types of data
- Some models can be better on some data, and others better on other data
- The approach we used: non-dominated sorting genetic algorithm
- The fitness function focuses on non-dominated models
- One model dominates another if
  - it is at least as good as the other in all constraints
  - better in at least one constraint

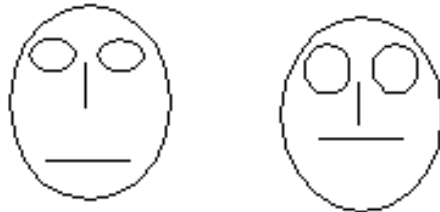


# A Case Study: Concept Formation

---



Category A



Category B



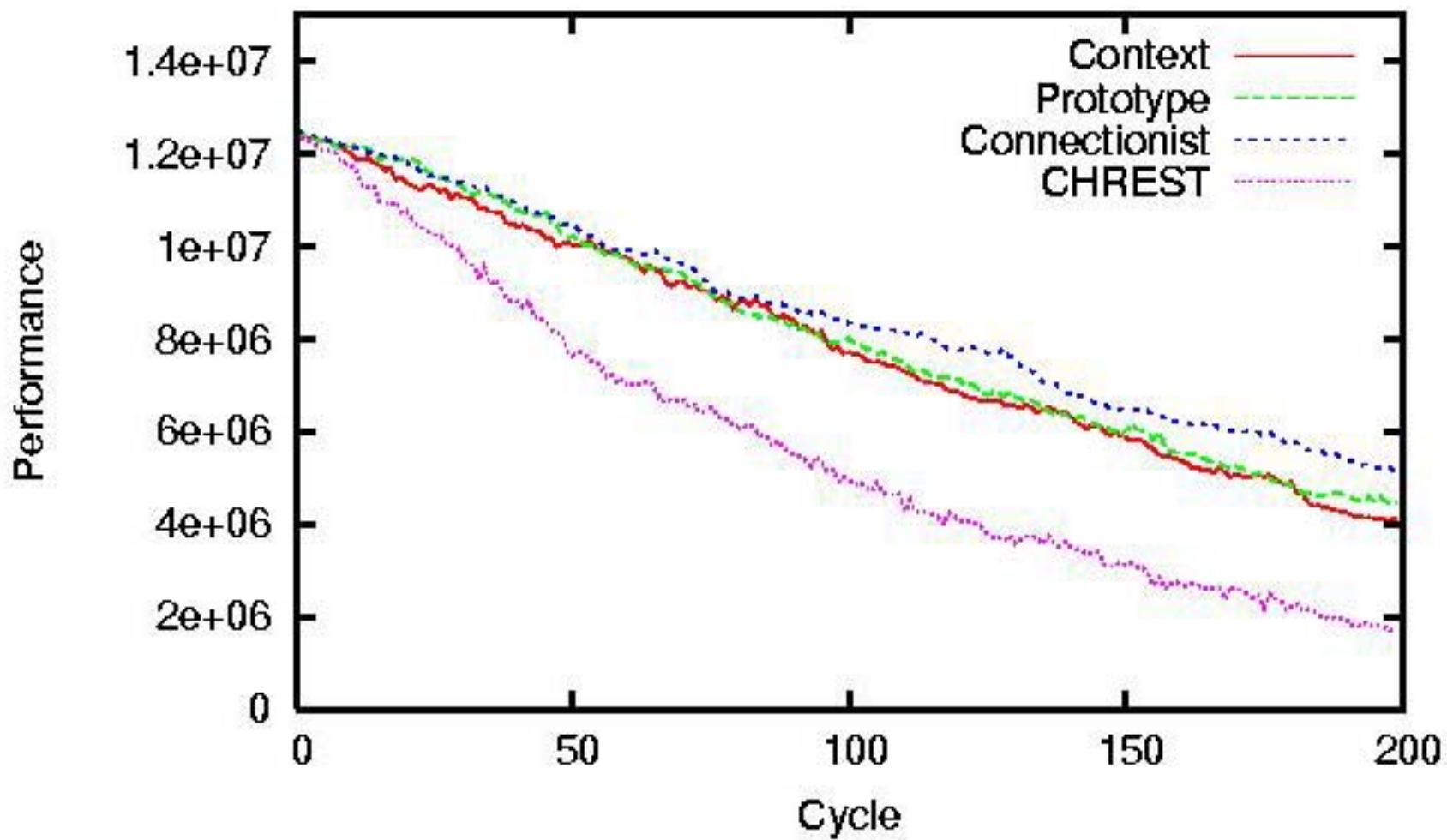
Which Category?

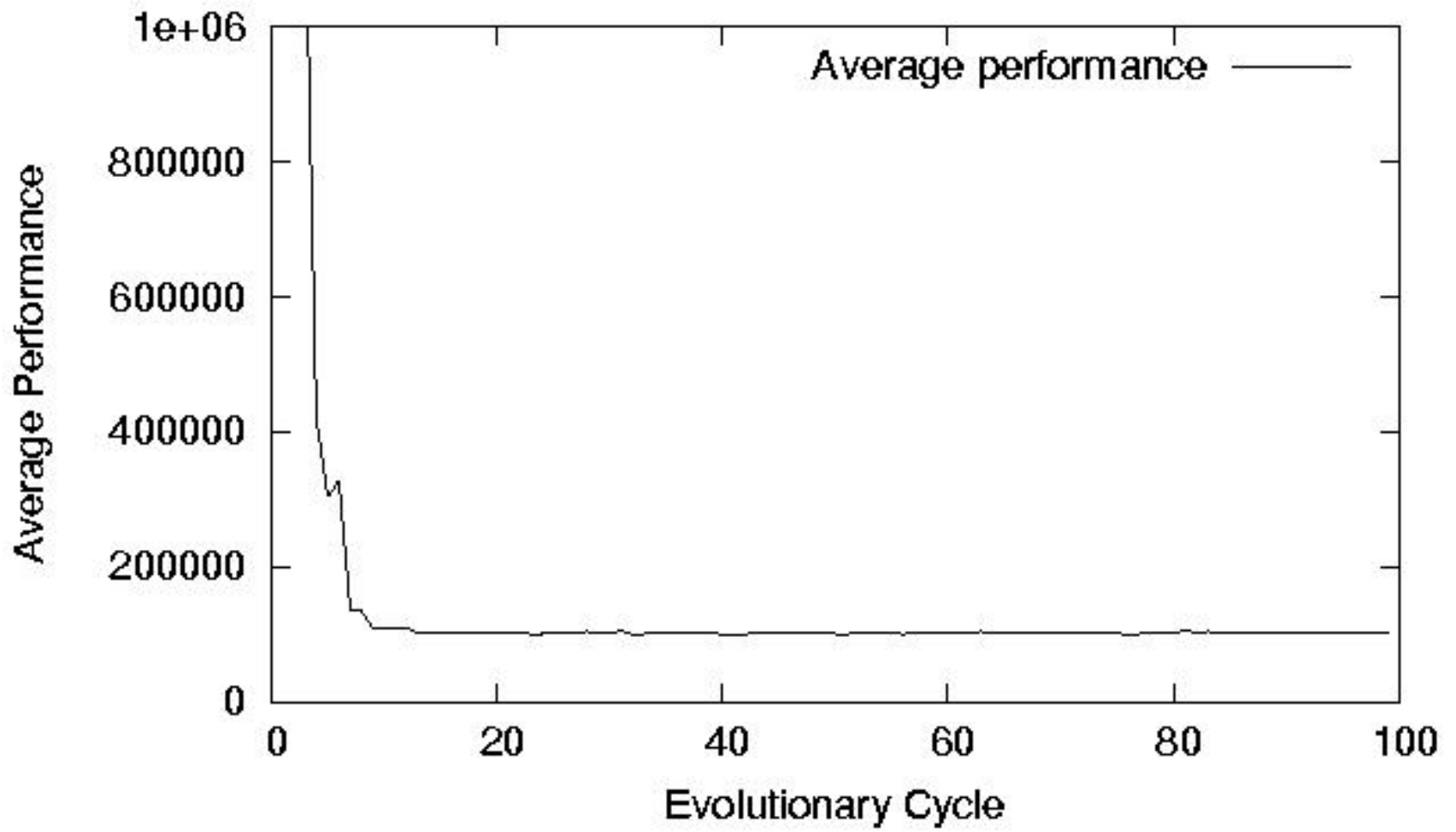


# A Case Study: Concept Formation

---

- The data to fit:
  - Different measures of percentage correct
  - Different measures of time needed
- Different classes of models are evolved simultaneously
  - Two mathematical models
    - Context model
    - Prototype model
  - One connectionist model
  - One discrimination net model (CHREST)







# Results

---

- Multiple classes of models can be evolved on multiple tasks
- Some of the models+parameters found are better than those published in the literature
- Evolving several classes of models enforces standardisation
  - In order to enable crossover between different 'species'
  - In order to simulate different types of task
- Takes about one hour to run on a PC
- More complex problems will require parallelism



# Models as Data

---

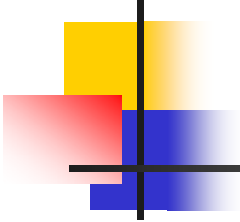
- Why did some models do better than others?
- Analytic and visualisation tools necessary for extracting patterns
- Some features enabling some models to perform well could be generalised to other models
  - E.g., if short-term memory appears to be capacity-limited in several models, this could be a feature of all models



# Generalisation for the Social Sciences

---

- Approach should apply both to quantitative and qualitative data
  - e.g., verbal protocols during problem solving in chess
  - Difficulty lies with current state of the art of theories, not with the kind of data
- Should apply both to controlled experiments and observational data
  - e.g., data on the national economy
  - Enables one to develop causal process theories, not only correlational theories

- 
- 
- Applies beyond psychology
  - The requirement is that theories should be expressed as computer programs
  - Highlights one of the disadvantages of informal theories
    - They cannot be optimised or otherwise manipulated formally



# Further Work

---

- Templates for representing experiments
  - Design
  - Instruction to subjects
  - Results
- Guidelines for writing papers (e.g., APA) could be used
- Standard format for representing models
- May help answer some long-standing questions in philosophy of science
  - Parsimony
  - Commensurability (are theories comparable?)



# Conclusion

---

- Information overload requires the development of computational models for summarising and explaining data
- Current lack of communication between empirical researchers, modellers, and theorists can be alleviated by making data, models, and theories available on the net
- This would enable the use of powerful computational techniques (e.g., optimisation) for developing better theories