Modelling Biochemical Pathways with Stochastic Process Algebra

Jane Hillston.
LFCS, University of Edinburgh

13th April 2007

The PEPA project

- The PEPA project started in Edinburgh in 1991.

The PEPA project

- The PEPA project started in Edinburgh in 1991.
- It was motivated by problems encountered when carrying out performance analysis of large computer and communication systems, based on numerical analysis of Markov processes.

The PEPA project

- The PEPA project started in Edinburgh in 1991.
- It was motivated by problems encountered when carrying out performance analysis of large computer and communication systems, based on numerical analysis of Markov processes.
- Process algebras offered a compositional description technique supported by apparatus for formal reasoning.

The PEPA project

- The PEPA project started in Edinburgh in 1991.
- It was motivated by problems encountered when carrying out performance analysis of large computer and communication systems, based on numerical analysis of Markov processes.
- Process algebras offered a compositional description technique supported by apparatus for formal reasoning.
- Performance Evaluation Process Algebra (PEPA) sought to address these problems by the introduction of a suitable process algebra.

The PEPA project

- The PEPA project started in Edinburgh in 1991.
- It was motivated by problems encountered when carrying out performance analysis of large computer and communication systems, based on numerical analysis of Markov processes.
- Process algebras offered a compositional description technique supported by apparatus for formal reasoning.
- Performance Evaluation Process Algebra (PEPA) sought to address these problems by the introduction of a suitable process algebra.
- The project has sought to investigate and exploit the interplay between the process algebra and the continuous time Markov chain (CTMC).

Outline

Outline

Systems Biology

- Biological advances mean that much more is now known about the components of cells and the interactions between them.

Systems Biology

- Biological advances mean that much more is now known about the components of cells and the interactions between them.
- Systems biology aims to develop a better understanding of the processes involved.

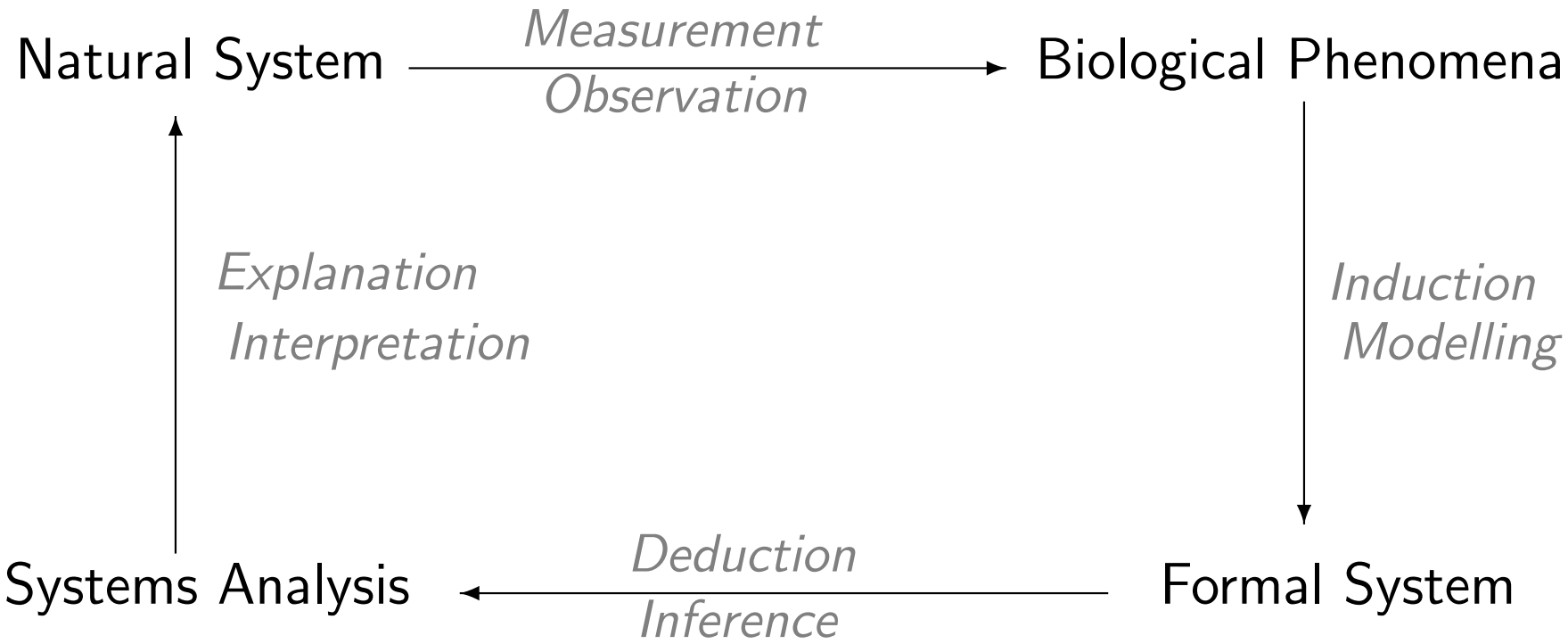
Systems Biology

- Biological advances mean that much more is now known about the components of cells and the interactions between them.
- Systems biology aims to develop a better understanding of the processes involved.
- It involves taking a systems theoretic view of biological processes — analysing inputs and outputs and the relationships between them.

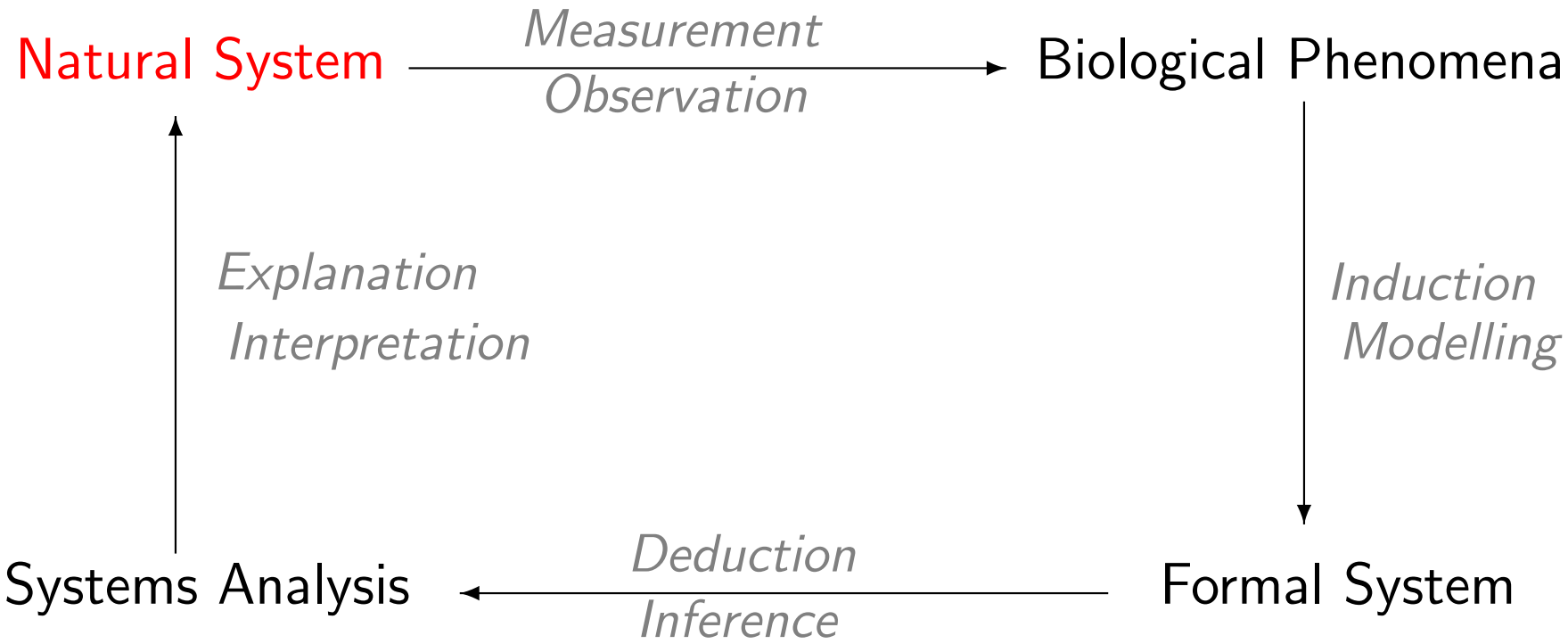
Systems Biology

- Biological advances mean that much more is now known about the components of cells and the interactions between them.
- Systems biology aims to develop a better understanding of the processes involved.
- It involves taking a systems theoretic view of biological processes — analysing inputs and outputs and the relationships between them.
- A radical shift from earlier reductionist approaches, systems biology aims to provide a conceptual basis and a methodology for reasoning about biological phenomena.

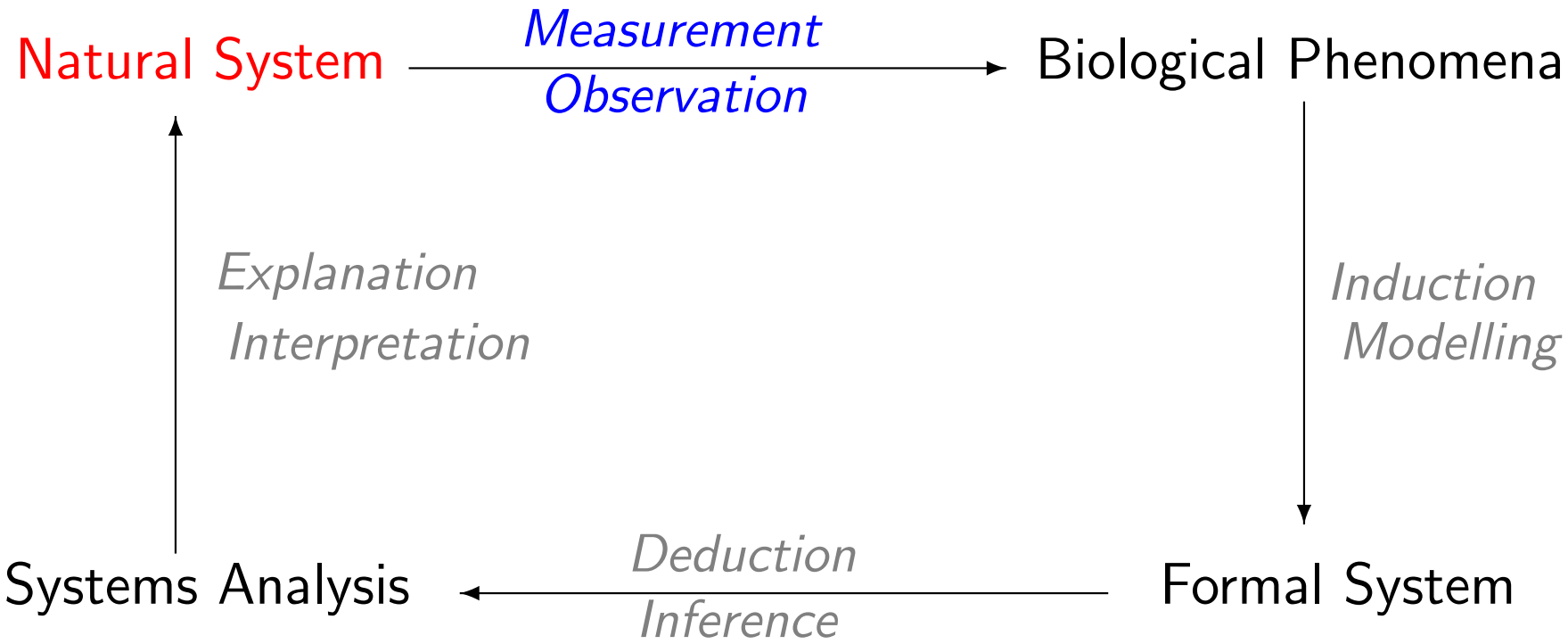
Systems Biology Methodology



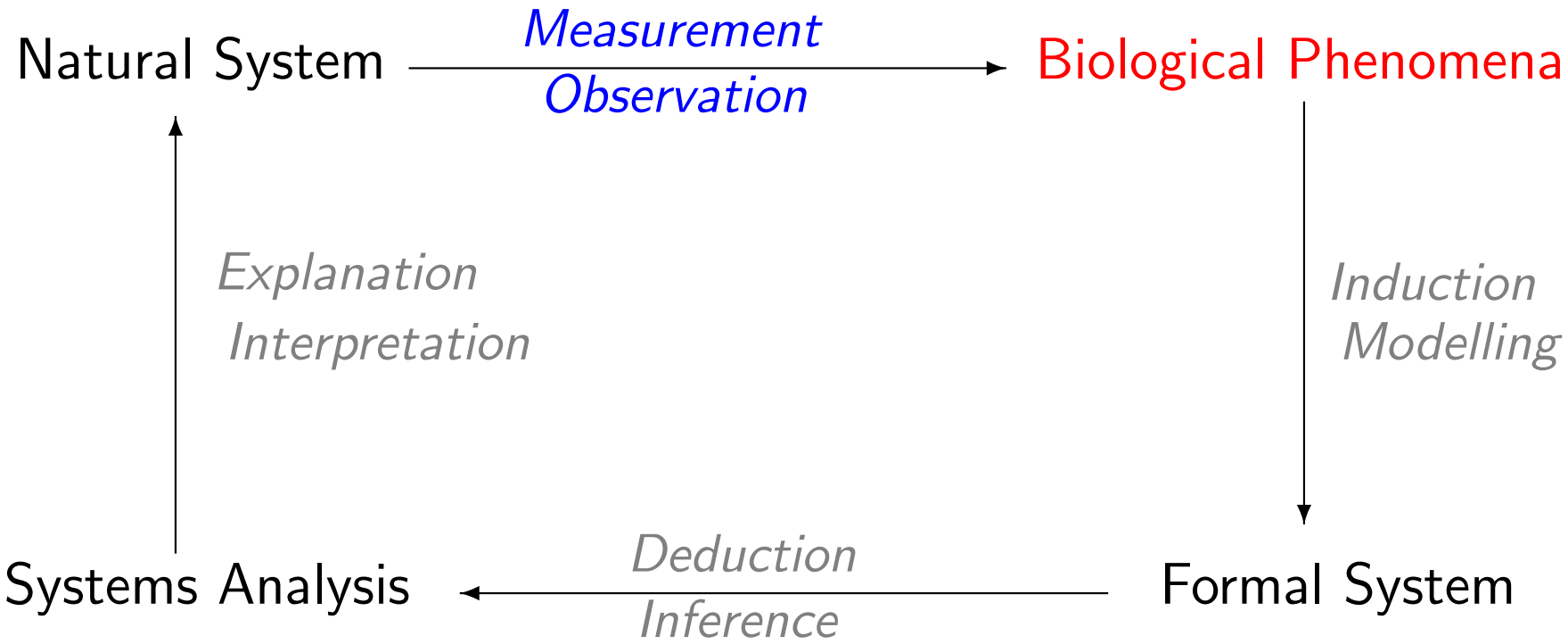
Systems Biology Methodology



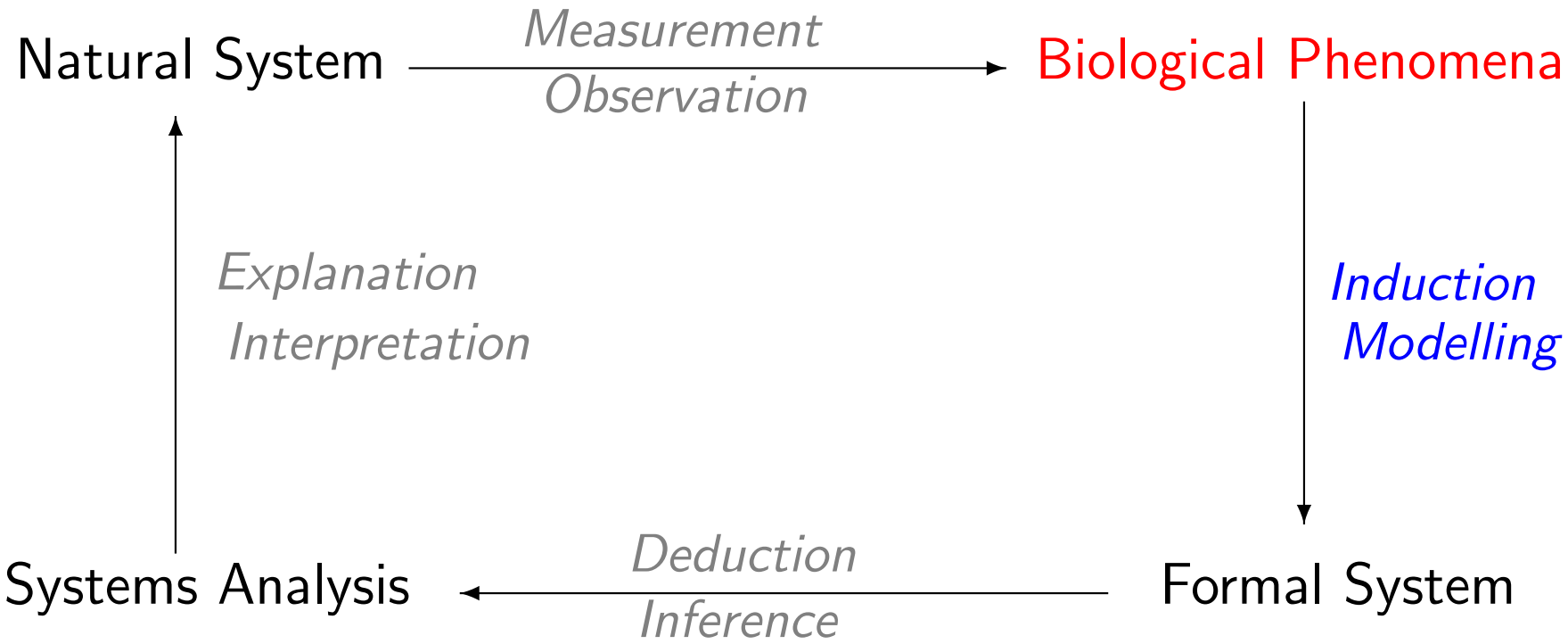
Systems Biology Methodology



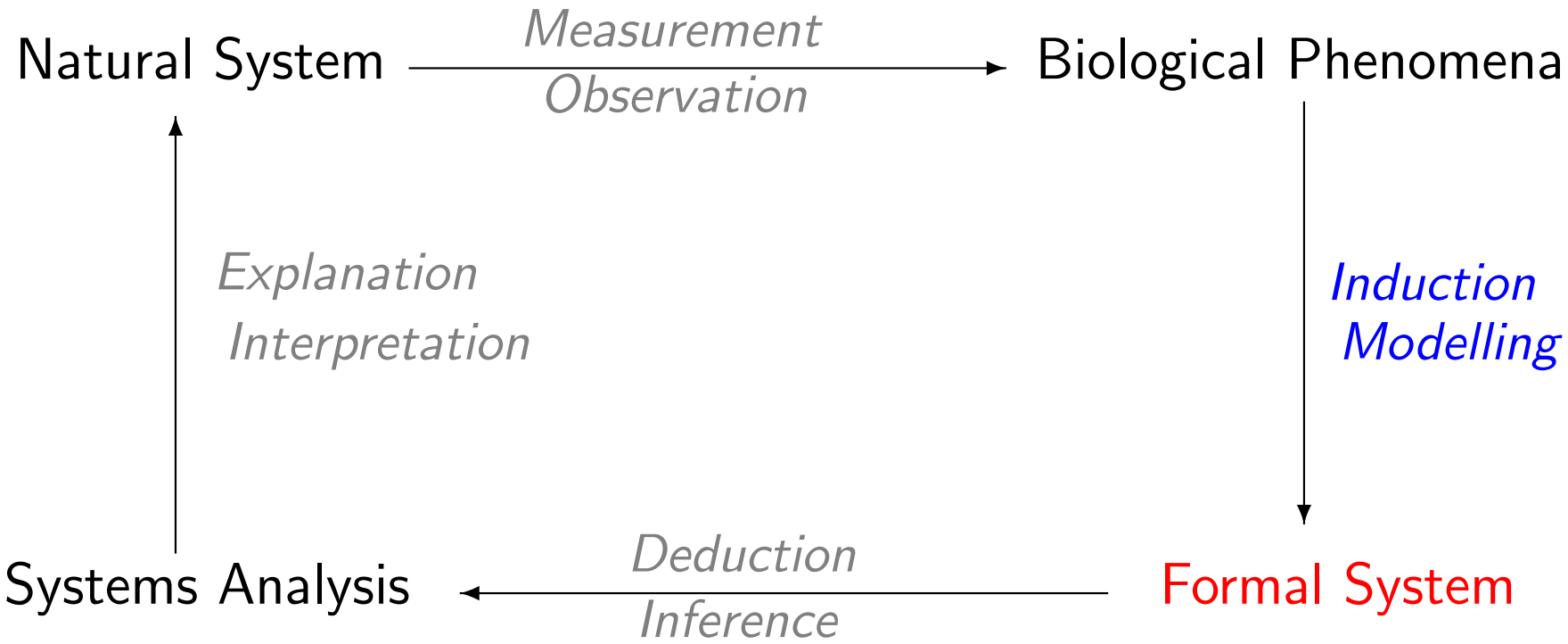
Systems Biology Methodology



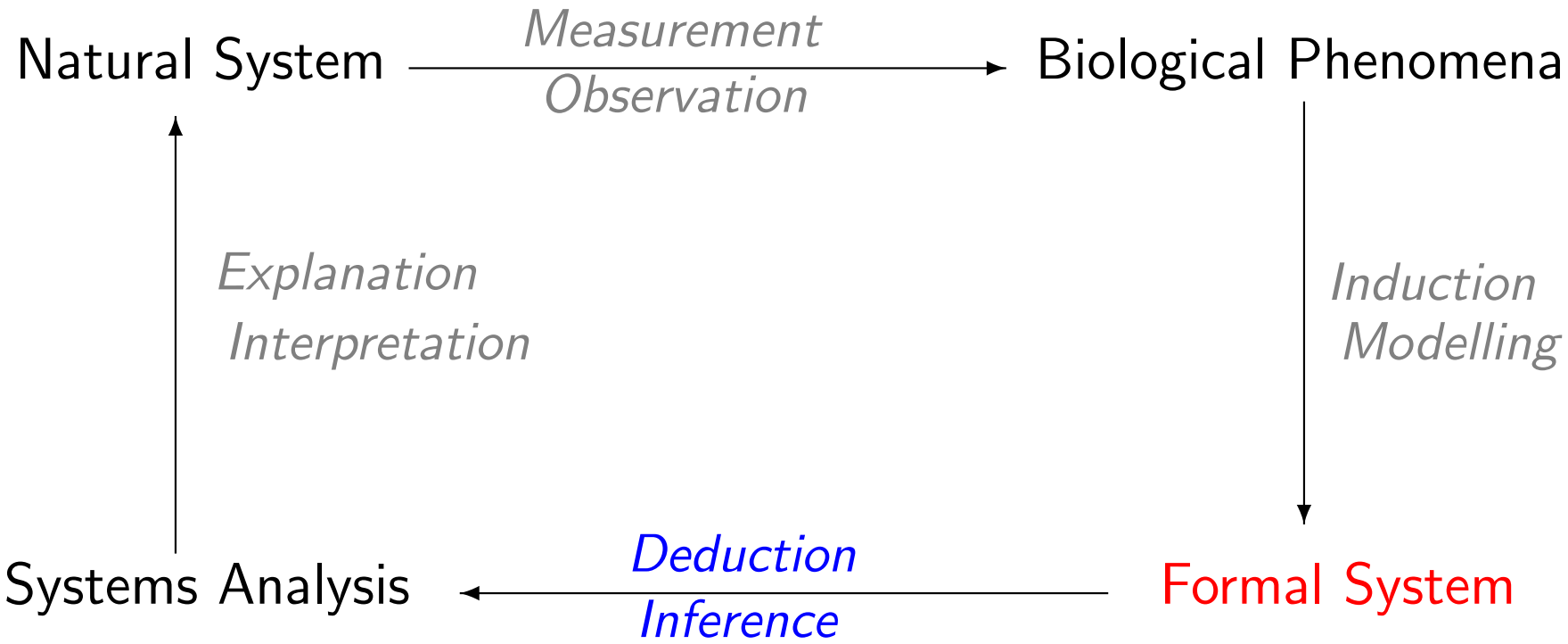
Systems Biology Methodology



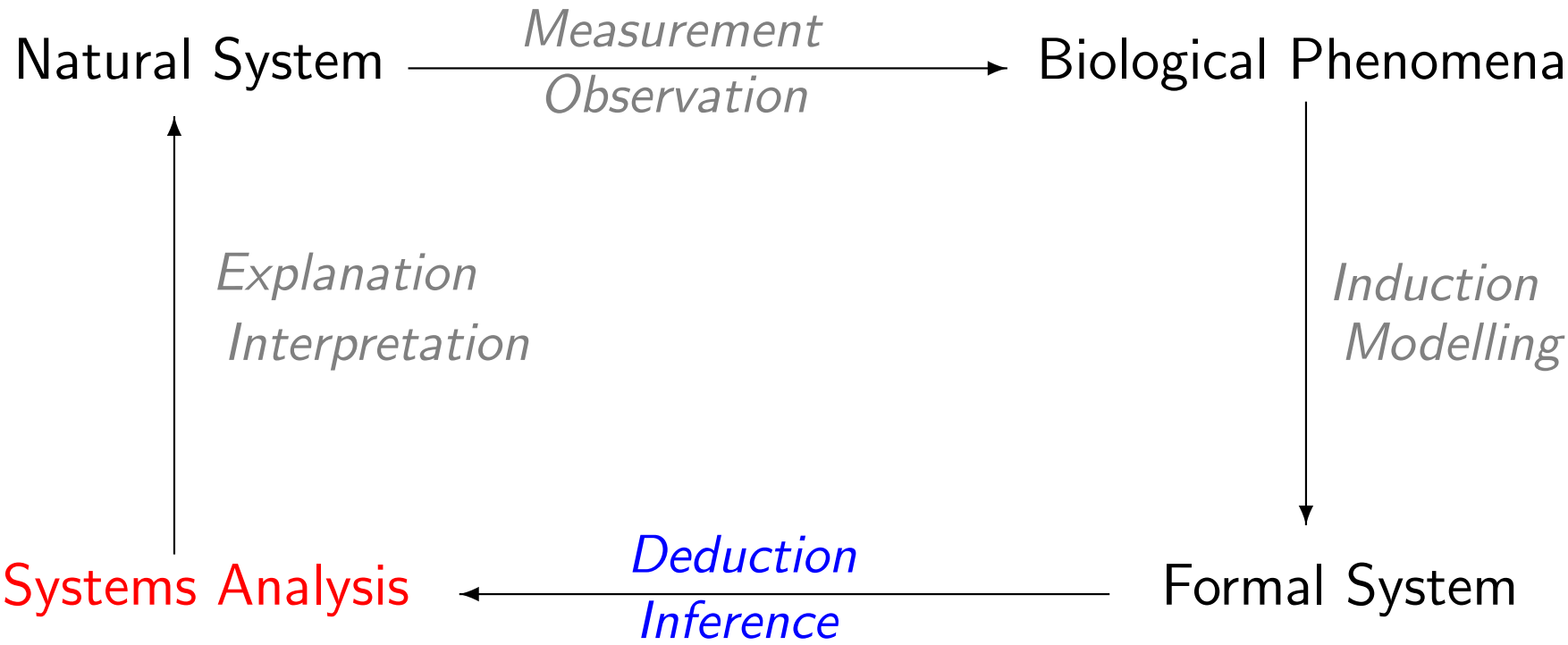
Systems Biology Methodology



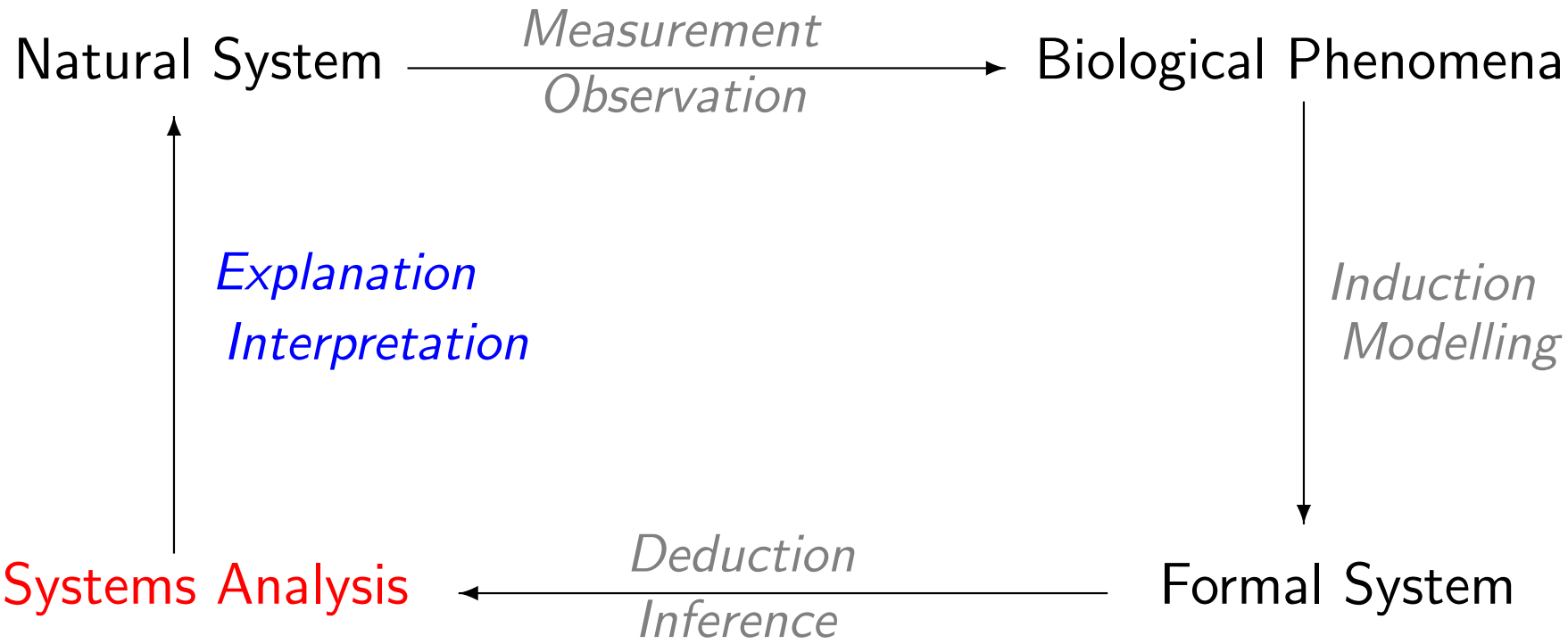
Systems Biology Methodology



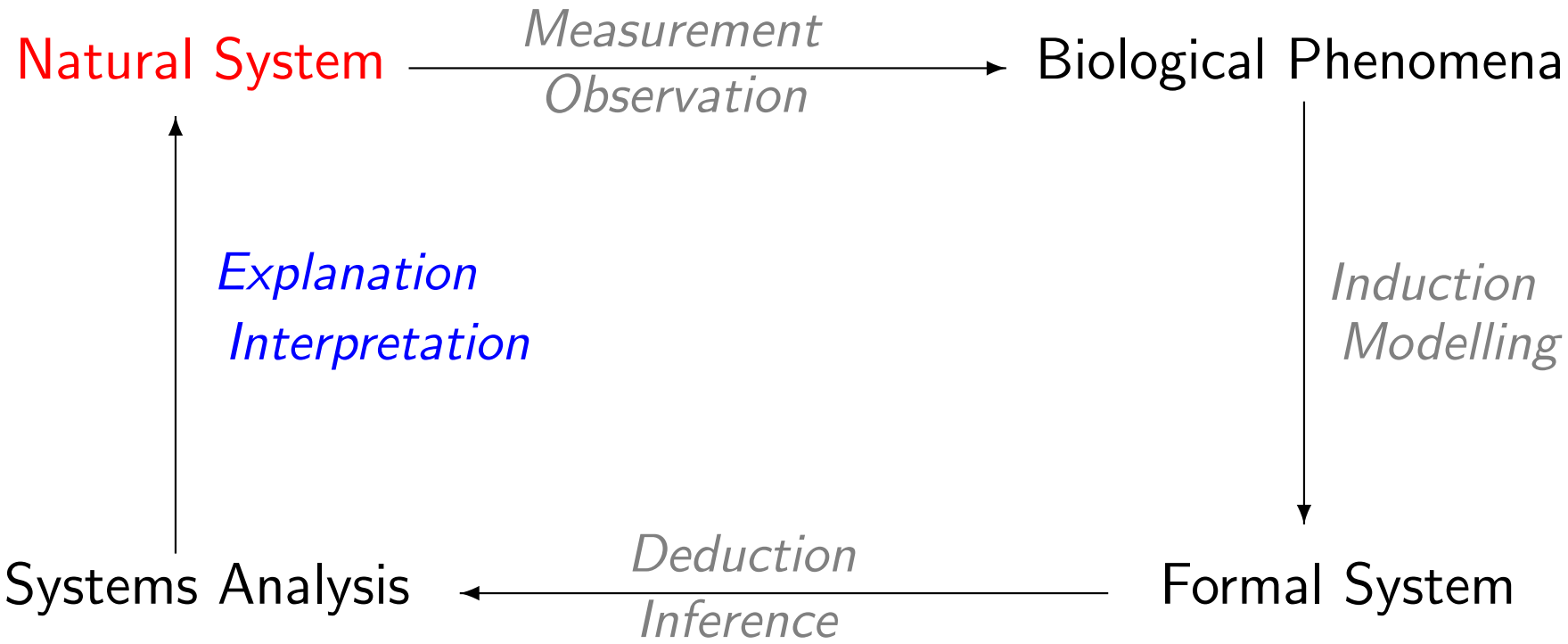
Systems Biology Methodology



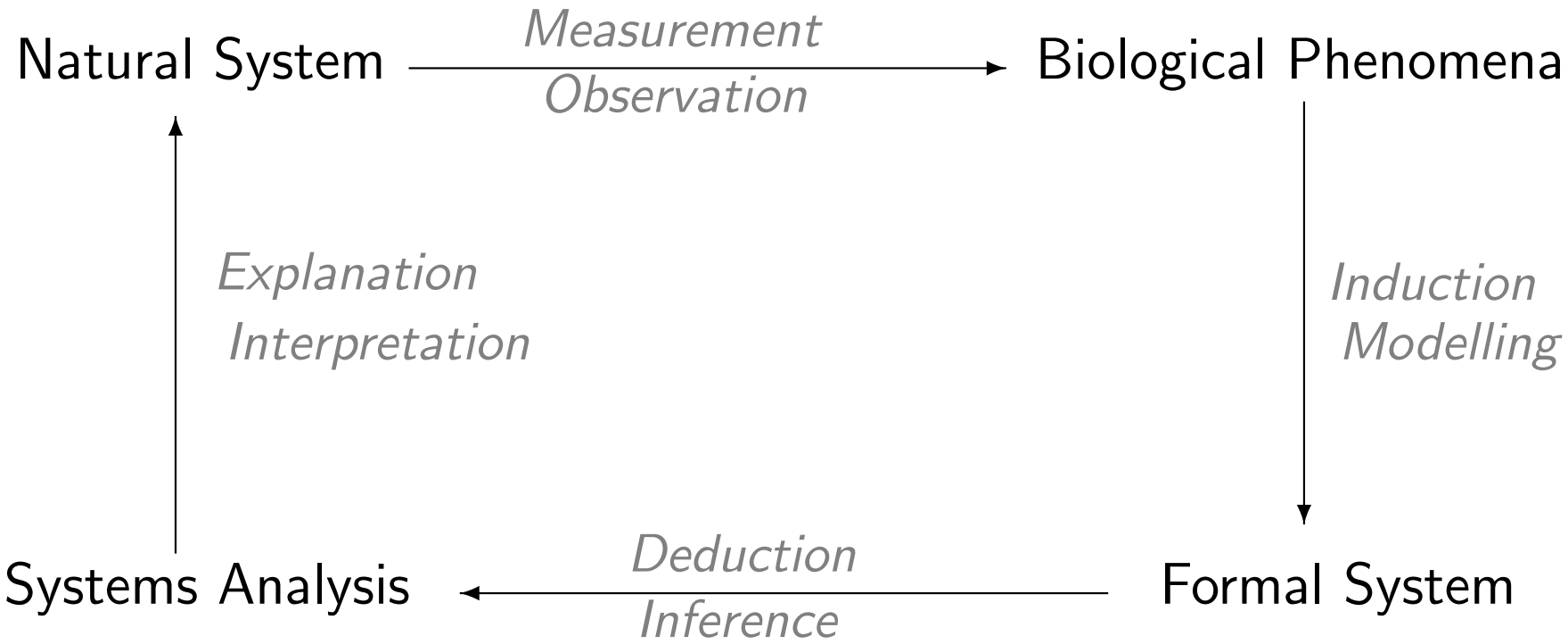
Systems Biology Methodology



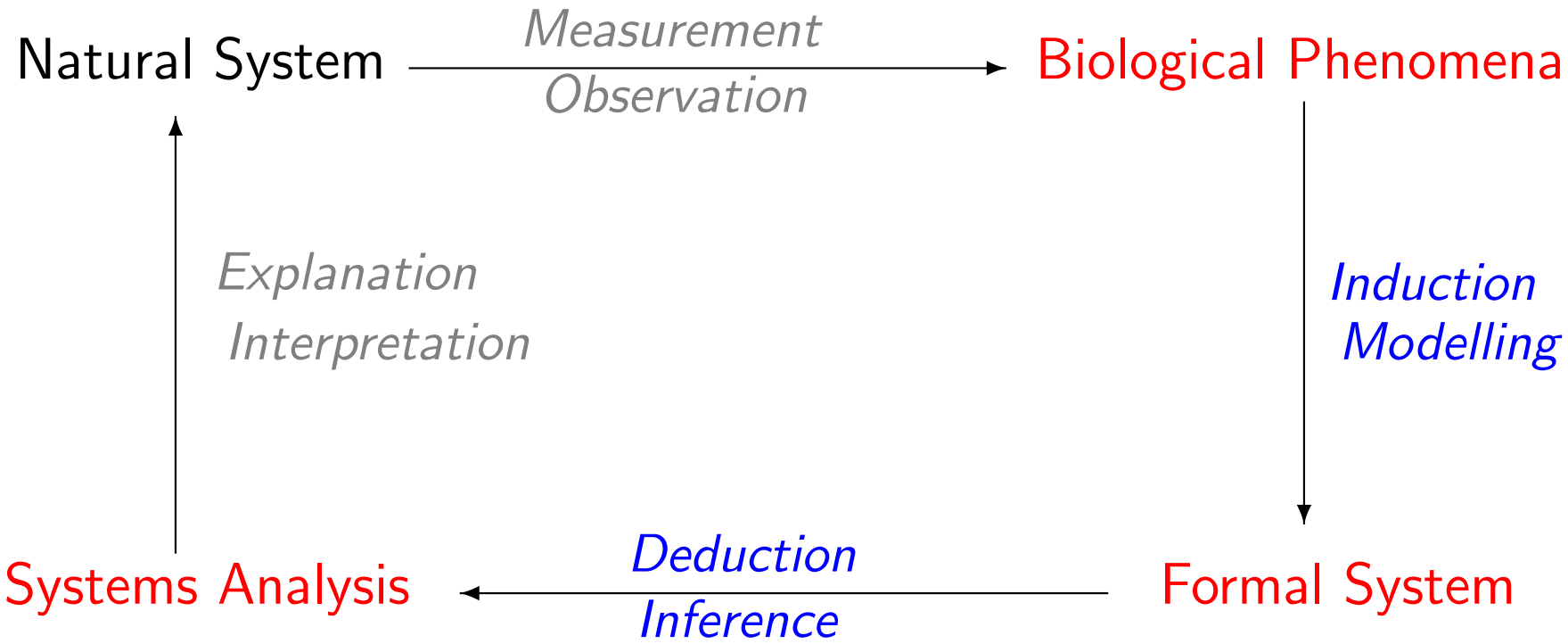
Systems Biology Methodology



Systems Biology Methodology



Systems Biology Methodology



Biological Phenomena

- As an approach systems biology can be applied to many different biological systems at different scales.

Biological Phenomena

- As an approach systems biology can be applied to many different biological systems at different scales.
- For example, from gene regulation within the nucleus of a cell, to whole organs, or even complete organisms.

Biological Phenomena

- As an approach systems biology can be applied to many different biological systems at different scales.
- For example, from gene regulation within the nucleus of a cell, to whole organs, or even complete organisms.
- The biological phenomena to be studied will clearly depend on the type of system being investigated.

Biological Phenomena

- As an approach systems biology can be applied to many different biological systems at different scales.
- For example, from gene regulation within the nucleus of a cell, to whole organs, or even complete organisms.
- The biological phenomena to be studied will clearly depend on the type of system being investigated.
- A grand challenge for systems biology is to develop multi-scale models which seek to account for high-level behaviour (at the level of the whole organisms) at all levels down to the intra-cellular processes.

Biochemical Pathways

At the intra-cellular level we can distinguish three distinct types of pathways or networks

Biochemical Pathways

At the intra-cellular level we can distinguish three distinct types of pathways or networks

Gene networks: Genes control the production of proteins but are themselves regulated by the same or different proteins.

Biochemical Pathways

At the intra-cellular level we can distinguish three distinct types of pathways or networks

Gene networks: Genes control the production of proteins but are themselves regulated by the same or different proteins.

Signal transduction networks: External stimuli initiate messages that are carried through a cell via a cascade of biochemical reactions.

Biochemical Pathways

At the intra-cellular level we can distinguish three distinct types of pathways or networks

Gene networks: Genes control the production of proteins but are themselves regulated by the same or different proteins.

Signal transduction networks: External stimuli initiate messages that are carried through a cell via a cascade of biochemical reactions.

Metabolic pathways: The survival of the cell depends on its ability to transform nutrients into energy.

Biochemical Pathways

At the intra-cellular level we can distinguish three distinct types of pathways or networks

Gene networks: Genes control the production of proteins but are themselves regulated by the same or different proteins.

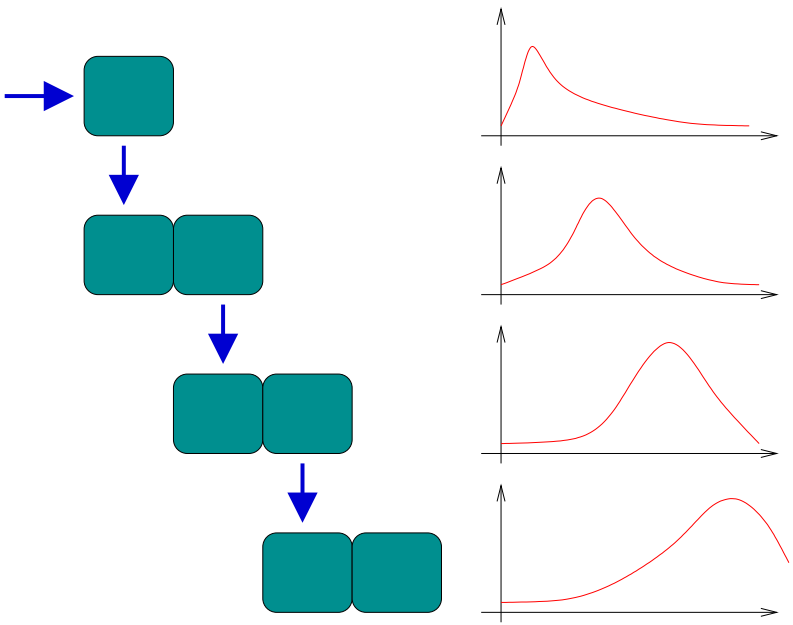
Signal transduction networks: External stimuli initiate messages that are carried through a cell via a cascade of biochemical reactions.

Metabolic pathways: The survival of the cell depends on its ability to transform nutrients into energy.

But these distinctions are to some extent arbitrary as models may include elements of more than one pathway type.

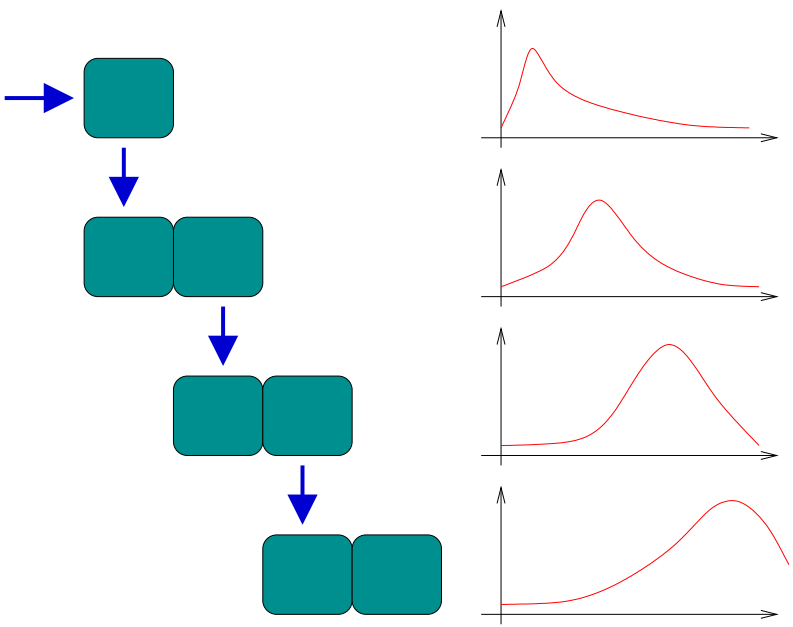
Signal transduction pathways

- All signalling is biochemical:



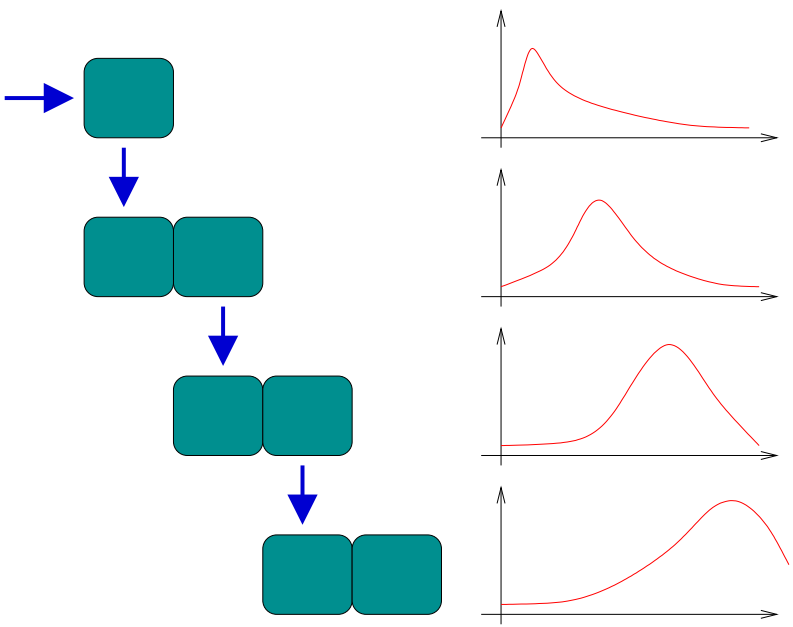
Signal transduction pathways

- All signalling is biochemical:
- Increasing protein concentration broadcasts the information about an event.



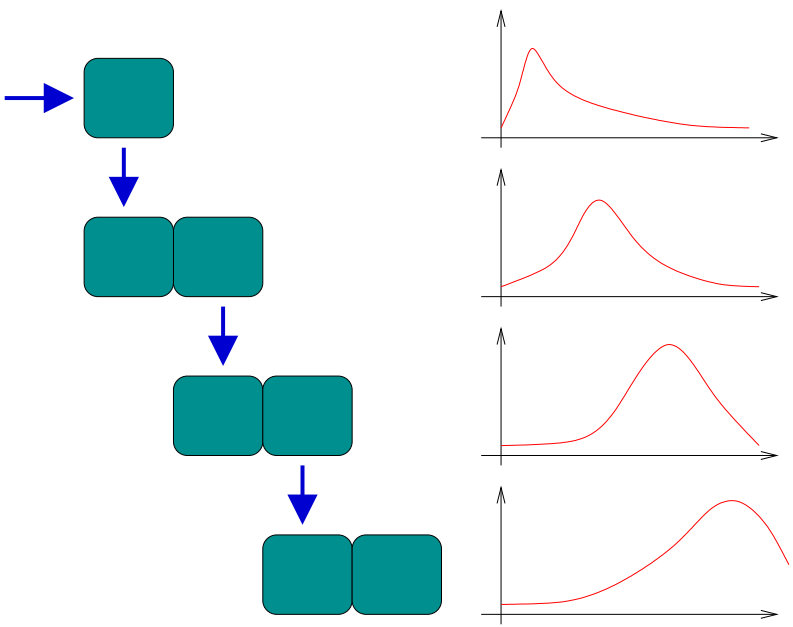
Signal transduction pathways

- All signalling is biochemical:
- Increasing protein concentration broadcasts the information about an event.
- The message is "received" by a concentration dependent response at the protein signal's site of action.



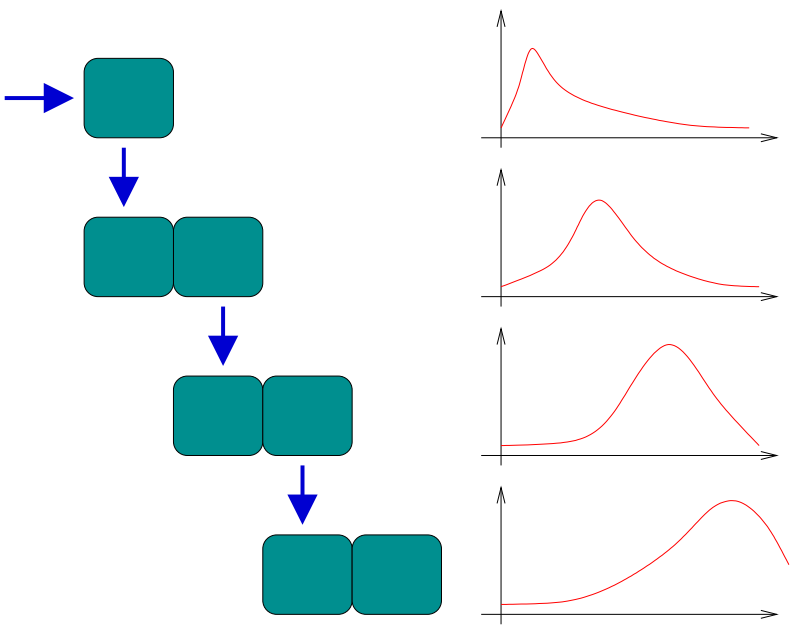
Signal transduction pathways

- All signalling is biochemical:
- Increasing protein concentration broadcasts the information about an event.
- The message is "received" by a concentration dependent response at the protein signal's site of action.
- This stimulates a response at the signalling protein's site of action.



Signal transduction pathways

- All signalling is biochemical:
- Increasing protein concentration broadcasts the information about an event.
- The message is "received" by a concentration dependent response at the protein signal's site of action.
- This stimulates a response at the signalling protein's site of action.
- Signals propagate through a series of protein accumulations.



Formal Systems

There are two alternative approaches to contructing dynamic models of biochemical pathways commonly used by biologists:

Formal Systems

There are two alternative approaches to contructing dynamic models of biochemical pathways commonly used by biologists:

- Ordinary Differential Equations:
  - continuous time,
  - continuous behaviour (concentrations),
  - deterministic.

Formal Systems

There are two alternative approaches to contructing dynamic models of biochemical pathways commonly used by biologists:

- Ordinary Differential Equations:
  - continuous time,
  - continuous behaviour (concentrations),
  - deterministic.
- Stochastic Simulation:
  - continuous time,
  - discrete behaviour (no. of molecules),
  - stochastic.

Ordinary Differential Equations

- This deterministic approach has at its core the law of mass action. This states that for a reaction in a homogeneous, free medium, the reaction rate will be proportional to the concentrations of the individual reactants involved.

Ordinary Differential Equations

- This deterministic approach has at its core the law of mass action. This states that for a reaction in a homogeneous, free medium, the reaction rate will be proportional to the concentrations of the individual reactants involved.

For example, for a reaction $A + B \xrightarrow{k} C$:

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = -k[A][B]$$

$$\frac{d[C]}{dt} = k[A][B]$$

Limitations of Ordinary Differential Equations

- Given knowledge of initial molecular concentrations, the law of mass action provides a complete picture of the component concentrations at all future time points.

Limitations of Ordinary Differential Equations

- Given knowledge of initial molecular concentrations, the law of mass action provides a complete picture of the component concentrations at all future time points.
- This is based on the assumption that chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic.

Limitations of Ordinary Differential Equations

- Given knowledge of initial molecular concentrations, the law of mass action provides a complete picture of the component concentrations at all future time points.
- This is based on the assumption that chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic.
- This is a simplification, because in reality chemical reactions involve discrete, random collisions between individual molecules.

Limitations of Ordinary Differential Equations

- Given knowledge of initial molecular concentrations, the law of mass action provides a complete picture of the component concentrations at all future time points.
- This is based on the assumption that chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic.
- This is a simplification, because in reality chemical reactions involve discrete, random collisions between individual molecules.
- As we consider smaller and smaller systems, the validity of a continuous approach becomes ever more tenuous.

Stochastic: Propensity function

As explicitly derived by Gillespie, the stochastic model uses basic Newtonian physics and thermodynamics to arrive at a form often termed the propensity function that gives the probability a_μ of reaction μ occurring in time interval $(t, t + \mathrm{d}t)$.

$$a_\mu \mathrm{d}t = h_\mu c_\mu \mathrm{d}t$$

where the M reaction mechanisms are given an arbitrary index μ $(1 \leq \mu \leq M)$, h_μ denotes the number of possible combinations of reactant molecules involved in reaction μ, and c_μ is a stochastic rate constant.

Stochastic: Chemical Master Equation

Applying this, and re-arranging the former, leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

$$\frac{\partial \Pr(\mathbf{X}; t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\mathbf{X} - \mathbf{v}_\mu) \Pr(\mathbf{X} - \mathbf{v}_\mu; t) - a_\mu(\mathbf{X}) \Pr(\mathbf{X}; t)$$

Does not lend itself to either analytic nor numerical solutions.

Stochastic: Chemical Master Equation

Applying this, and re-arranging the former, leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

$$\frac{\partial \Pr(\mathbf{X}; t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\mathbf{X} - \mathbf{v}_\mu) \Pr(\mathbf{X} - \mathbf{v}_\mu; t) - a_\mu(\mathbf{X}) \Pr(\mathbf{X}; t)$$

Does not lend itself to either analytic nor numerical solutions.

Stochastic: Chemical Master Equation

Applying this, and re-arranging the former, leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

$$\frac{\partial \Pr(\mathbf{X}; t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\mathbf{X} - \mathbf{v}_\mu) \Pr(\mathbf{X} - \mathbf{v}_\mu; t) - a_\mu(\mathbf{X}) \Pr(\mathbf{X}; t)$$

Does not lend itself to either analytic nor numerical solutions.

Stochastic: Chemical Master Equation

Applying this, and re-arranging the former, leads us to an important *partial differential equation* (PDE) known as the Chemical Master Equation (CME).

$$\frac{\partial \Pr(\mathbf{X}; t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\mathbf{X} - \mathbf{v}_\mu) \Pr(\mathbf{X} - \mathbf{v}_\mu; t) - a_\mu(\mathbf{X}) \Pr(\mathbf{X}; t)$$

Does not lend itself to either analytic nor numerical solutions.

Stochastic simulation algorithms

Gillespie's Stochastic Simulation Algorithm (SSA) is essentially an exact procedure for numerically simulating the time evolution of a well-stirred chemically reacting system by taking proper account of the randomness inherent in such a system.

Stochastic simulation algorithms

Gillespie's Stochastic Simulation Algorithm (SSA) is essentially an exact procedure for numerically simulating the time evolution of a well-stirred chemically reacting system by taking proper account of the randomness inherent in such a system.

It is rigorously based on the same microphysical premise that underlies the chemical master equation and gives a more realistic representation of a system's evolution than the deterministic reaction rate equation (RRE) represented mathematically by ODEs.

Stochastic simulation algorithms

Gillespie's Stochastic Simulation Algorithm (SSA) is essentially an exact procedure for numerically simulating the time evolution of a well-stirred chemically reacting system by taking proper account of the randomness inherent in such a system.

It is rigorously based on the same microphysical premise that underlies the chemical master equation and gives a more realistic representation of a system's evolution than the deterministic reaction rate equation (RRE) represented mathematically by ODEs.

As with the chemical master equation, the SSA converges, in the limit of large numbers of reactants, to the same solution as the law of mass action.

Systems Analysis

- In biochemical signalling pathways the events of interests are
  - when reagent concentrations start to increase;
  - when concentrations pass certain thresholds;
  - when a peak of concentration is reached.

Systems Analysis

- In biochemical signalling pathways the events of interests are
  - when reagent concentrations start to increase;
  - when concentrations pass certain thresholds;
  - when a peak of concentration is reached.
- For example, delay from the activation of a gene promoter until reaching an effective level to control the next promoter in a pathway depends on the rate of protein accumulation.

Systems Analysis

- In biochemical signalling pathways the events of interests are
  - when reagent concentrations start to increase;
  - when concentrations pass certain thresholds;
  - when a peak of concentration is reached.
- For example, delay from the activation of a gene promoter until reaching an effective level to control the next promoter in a pathway depends on the rate of protein accumulation.
- These are the data that can be collected from wet lab experiments.

Systems Analysis

- In biochemical signalling pathways the events of interests are
  - when reagent concentrations start to increase;
  - when concentrations pass certain thresholds;
  - when a peak of concentration is reached.
- For example, delay from the activation of a gene promoter until reaching an effective level to control the next promoter in a pathway depends on the rate of protein accumulation.
- These are the data that can be collected from wet lab experiments.
- The accumulation of protein is a stochastic process affected by several factors in the cell (temperature, pH, etc.).

Systems Analysis

- In biochemical signalling pathways the events of interests are
  - when reagent concentrations start to increase;
  - when concentrations pass certain thresholds;
  - when a peak of concentration is reached.
- For example, delay from the activation of a gene promoter until reaching an effective level to control the next promoter in a pathway depends on the rate of protein accumulation.
- These are the data that can be collected from wet lab experiments.
- The accumulation of protein is a stochastic process affected by several factors in the cell (temperature, pH, etc.).
- Thus it is more realistic to talk about a distribution rather than a deterministic time.

Case Study: Circadian Rhythms – Overview

The study by Locke *et al.* (Extension of a genetic network model by iterative experimentation and mathematical analysis. *Molecular Systems Biology*) focuses on the circadian rhythms in plants, combining mathematical models and molecular biology.

Case Study: Circadian Rhythms – Overview

The study by Locke *et al.* (Extension of a genetic network model by iterative experimentation and mathematical analysis. *Molecular Systems Biology*) focuses on the circadian rhythms in plants, combining mathematical models and molecular biology.

Their objective is to identify the genes (and proteins) responsible for maintaining the daily rhythms observed in the plants.

Case Study: Circadian Rhythms – Overview

The study by Locke *et al.* (Extension of a genetic network model by iterative experimentation and mathematical analysis. *Molecular Systems Biology*) focuses on the circadian rhythms in plants, combining mathematical models and molecular biology.

Their objective is to identify the genes (and proteins) responsible for maintaining the daily rhythms observed in the plants.

The research exploits an interplay between mathematical models, experiments in the laboratory and literature search.

Case Study: Circadian Rhythms – Overview

The study by Locke *et al.* (Extension of a genetic network model by iterative experimentation and mathematical analysis. *Molecular Systems Biology*) focuses on the circadian rhythms in plants, combining mathematical models and molecular biology.

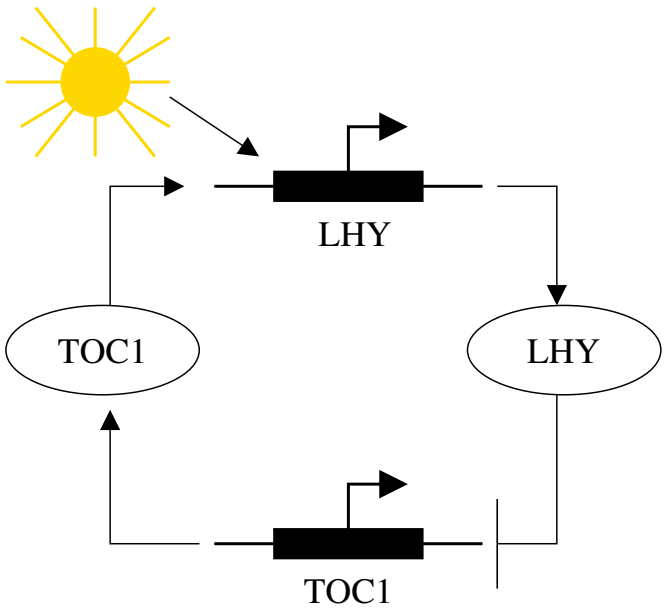
Their objective is to identify the genes (and proteins) responsible for maintaining the daily rhythms observed in the plants.

The research exploits an interplay between mathematical models, experiments in the laboratory and literature search.

It is held up as an exemplar of what systems biology is trying to achieve, and the breakthroughs that it can bring about when it is successful.

Case Study: Circadian Rhythms – Initial Model

From initial experiments Locke *et al.* identified a two genes and two proteins which appeared to operate in a simple loop:



An initial mathematical model (ODEs) was constructed to capture this model.

Case Study: Circadian Rhythms – Role of Mathematics

Initial simulations with the mathematical model showed good agreement with the experimental data for some of the observed phenomena but significant discrepancies for others.

Case Study: Circadian Rhythms – Role of Mathematics

Initial simulations with the mathematical model showed good agreement with the experimental data for some of the observed phenomena but significant discrepancies for others.

Experiments were then undertaken with the mathematical model to find an alternative model which was biologically plausible but produced a better fit.

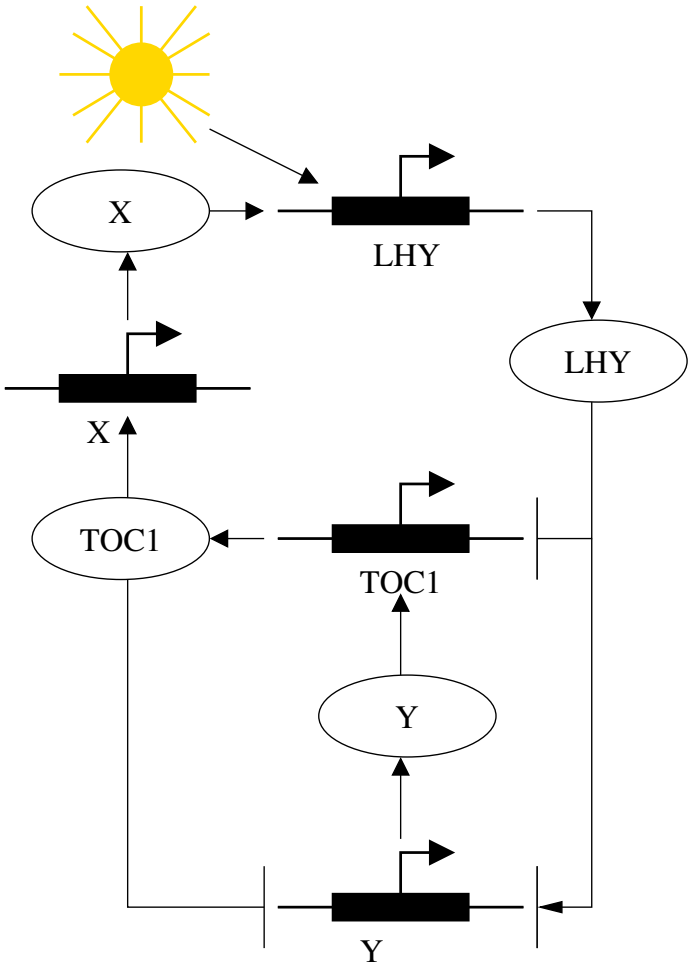
Case Study: Circadian Rhythms – Role of Mathematics

Initial simulations with the mathematical model showed good agreement with the experimental data for some of the observed phenomena but significant discrepancies for others.

Experiments were then undertaken with the mathematical model to find an alternative model which was biologically plausible but produced a better fit.

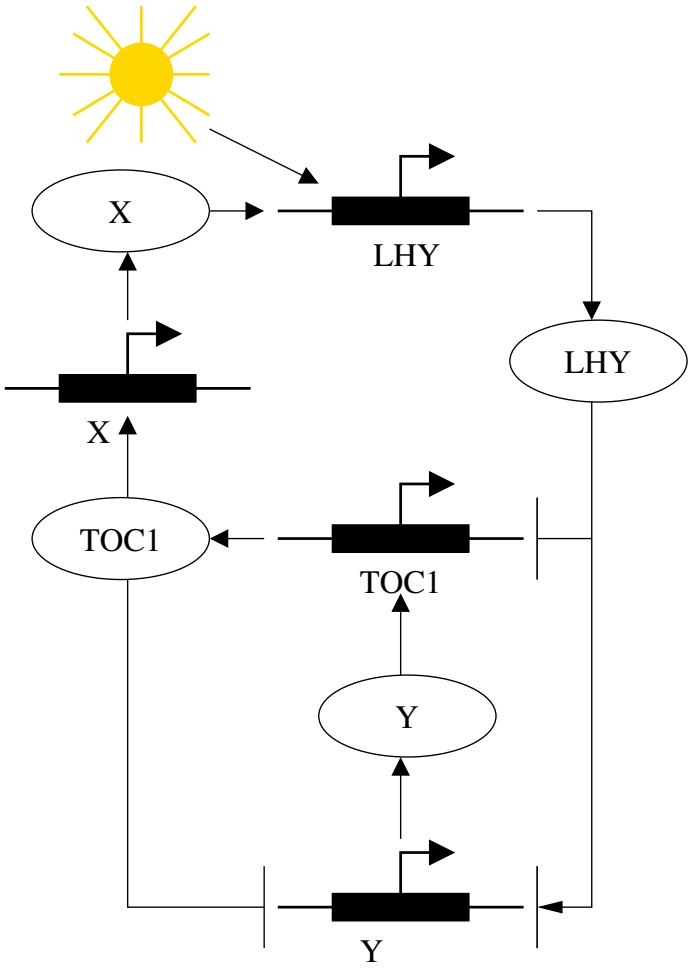
These mathematical experiments conjectured a network with two interacting loops.

Case Study: Circadian Rhythms – Elaborated Model



Two "new" genes were introduced to the model which now has interlocking loops and more complex feedback.

Case Study: Circadian Rhythms – Elaborated Model



Two "new" genes were introduced to the model which now has interlocking loops and more complex feedback.

The simulation results from this model showed much better agreement with the observed data.

Case Study: Circadian Rhythms – Validating the Model

The researchers then sought to identify the "new" genes X and Y.

Searching the literature elicited several candidate genes which previous experimental studies had suggested were implicated in the circadian rhythm.

In particular, "knockout" data for one, GIGANTEA (GI), coincided with the pattern from simulation experiments of the original model with a single loop.

Subsequent wet lab experiments have reinforced this impression that GI is gene Y.

Case Study: The VICE project – Overview

- There is an open problem in evolutionary theory to find LUCA — the Last Unknown Common Ancestor.

Case Study: The VICE project – Overview

- There is an open problem in evolutionary theory to find LUCA — the Last Unknown Common Ancestor.
- Based on various bacterial genomes previous work had conjectured a Minimal Gene Set (MGS) containing 254 genes.

Case Study: The VICE project – Overview

- There is an open problem in evolutionary theory to find LUCA — the Last Unknown Common Ancestor.
- Based on various bacterial genomes previous work had conjectured a Minimal Gene Set (MGS) containing 254 genes.
- Chiarugi *et al.* carried out in silico experiments to test the viability of this gene set.

Case Study: The VICE project – Overview

- There is an open problem in evolutionary theory to find LUCA — the Last Unknown Common Ancestor.
- Based on various bacterial genomes previous work had conjectured a Minimal Gene Set (MGS) containing 254 genes.
- Chiarugi *et al.* carried out in silico experiments to test the viability of this gene set.

D. Chiarugi, M. Curti, P. Degano and R. Marangoni
VICE: A VIrtual CEll
in *Proceedings of the 2nd International Workshop on Computational Methods in Systems Biology* Paris, France, April 2004.

Case Study: The VICE project – Initial model

- An initial model was developed using the *pi*-calculus.

Case Study: The VICE project – Initial model

- An initial model was developed using the *pi*-calculus.
- The model focussed on the metabolic behaviour of the cell (no reproduction). Thus the π-calculus model was intended to behave as a simplified single cell organism.

Case Study: The VICE project – Initial model

- An initial model was developed using the *pi*-calculus.
- The model focussed on the metabolic behaviour of the cell (no reproduction). Thus the π-calculus model was intended to behave as a simplified single cell organism.
- Each reactant was modelled as a distinct component in the model, capturing how each reaction changed the state of that reactant.

Case Study: The VICE project – Initial model

- An initial model was developed using the *pi*-calculus.
- The model focussed on the metabolic behaviour of the cell (no reproduction). Thus the π-calculus model was intended to behave as a simplified single cell organism.
- Each reactant was modelled as a distinct component in the model, capturing how each reaction changed the state of that reactant.
- The experiments showed that MGS was not viable: the cell could not survive in simulation.

Case Study: The VICE project – Initial model

- An initial model was developed using the *pi*-calculus.
- The model focussed on the metabolic behaviour of the cell (no reproduction). Thus the π-calculus model was intended to behave as a simplified single cell organism.
- Each reactant was modelled as a distinct component in the model, capturing how each reaction changed the state of that reactant.
- The experiments showed that MGS was not viable: the cell could not survive in simulation.
- 76 genes were found to be functionally duplicated and 7 additional genes were added to form VICE.

Case Study: The VICE project – In silico experimentation

- Many different alternative gene sets were possible on the basis of the original model and experiments.

Case Study: The VICE project – In silico experimentation

- Many different alternative gene sets were possible on the basis of the original model and experiments.
- Subsequent models were developed in stochastic *pi*-calculus so that comparison with dynamic biological data was possible.

Case Study: The VICE project – In silico experimentation

- Many different alternative gene sets were possible on the basis of the original model and experiments.
- Subsequent models were developed in stochastic *pi*-calculus so that comparison with dynamic biological data was possible.
- A bespoke simulator was written to simulate the behaviour of the alternative gene sets and the VICE gene set was chosen as the most promising.

Case Study: The VICE project – In silico experimentation

- Many different alternative gene sets were possible on the basis of the original model and experiments.
- Subsequent models were developed in stochastic *pi*-calculus so that comparison with dynamic biological data was possible.
- A bespoke simulator was written to simulate the behaviour of the alternative gene sets and the VICE gene set was chosen as the most promising.
- The steady state distribution of the concentrations of virtual metabolites was similar to that measured for bacteria experimentally.

Formal Systems Revisited

- Currently mathematics is being used directly as the formal system — even the work with the stochastic π-calculus only uses the π-calculus to describe a continuous time Markov chain (CTMC) for simulation.

Formal Systems Revisited

- Currently mathematics is being used directly as the formal system — even the work with the stochastic π-calculus only uses the π-calculus to describe a continuous time Markov chain (CTMC) for simulation.
- Previous experience in the performance arena has shown us that there can be benefits to interposing a formal model between the system and the underlying mathematical model.

Formal Systems Revisited

- Currently mathematics is being used directly as the formal system — even the work with the stochastic π-calculus only uses the π-calculus to describe a continuous time Markov chain (CTMC) for simulation.
- Previous experience in the performance arena has shown us that there can be benefits to interposing a formal model between the system and the underlying mathematical model.
- Moreover taking this "high-level programming" style approach offers the possibility of different "compilations" to different mathematical models.

Outline

Individual vs. Population behaviour

- Biochemistry is concerned with the reactions between individual molecules and so it is often more natural to model at this level.

Individual vs. Population behaviour

- Biochemistry is concerned with the reactions between individual molecules and so it is often more natural to model at this level.
- However experimental data is usually more readily available in terms of populations rather than individual molecules cf. average reaction rates rather than the forces at play on an individual molecule in a particular physical context.

Individual vs. Population behaviour

- Biochemistry is concerned with the reactions between individual molecules and so it is often more natural to model at this level.
- However experimental data is usually more readily available in terms of populations rather than individual molecules cf. average reaction rates rather than the forces at play on an individual molecule in a particular physical context.
- These should be regarded as alternatives, each being appropriate for some models. The challenge then becomes when to use which approach.

Individual vs. Population behaviour

- Biochemistry is concerned with the reactions between individual molecules and so it is often more natural to model at this level.
- However experimental data is usually more readily available in terms of populations rather than individual molecules cf. average reaction rates rather than the forces at play on an individual molecule in a particular physical context.
- These should be regarded as alternatives, each being appropriate for some models. The challenge then becomes when to use which approach.
- Note that given a large enough number of molecules an "individuals" model will (in many circumstances) be indistinguishable from the a "population" level model.

Noise vs. Determinism

- With perfect knowledge the behaviour of a biochemical reaction would be deterministic.

Noise vs. Determinism

- With perfect knowledge the behaviour of a biochemical reaction would be deterministic.
- However, in general, we do not have the requisite knowledge of thermodynamic forces, exact relative positions, temperature, velocity etc.

Noise vs. Determinism

- With perfect knowledge the behaviour of a biochemical reaction would be deterministic.
- However, in general, we do not have the requisite knowledge of thermodynamic forces, exact relative positions, temperature, velocity etc.
- Thus a reaction appears to display stochastic behaviour.

Noise vs. Determinism

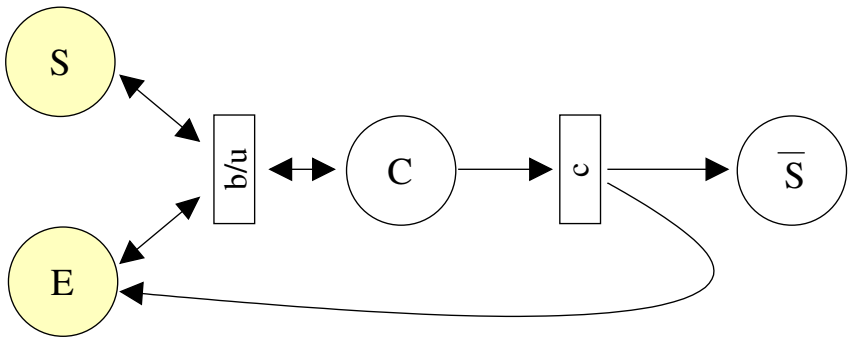
- With perfect knowledge the behaviour of a biochemical reaction would be deterministic.
- However, in general, we do not have the requisite knowledge of thermodynamic forces, exact relative positions, temperature, velocity etc.
- Thus a reaction appears to display stochastic behaviour.
- When a large number of such reactions occur, the randomness of the individual reactions can cancel each other out and the apparent behaviour exhibits less variability.

Noise vs. Determinism

- With perfect knowledge the behaviour of a biochemical reaction would be deterministic.
- However, in general, we do not have the requisite knowledge of thermodynamic forces, exact relative positions, temperature, velocity etc.
- Thus a reaction appears to display stochastic behaviour.
- When a large number of such reactions occur, the randomness of the individual reactions can cancel each other out and the apparent behaviour exhibits less variability.
- However, in some systems the variability in the stochastic behaviour plays a crucial role in the dynamics of the system.

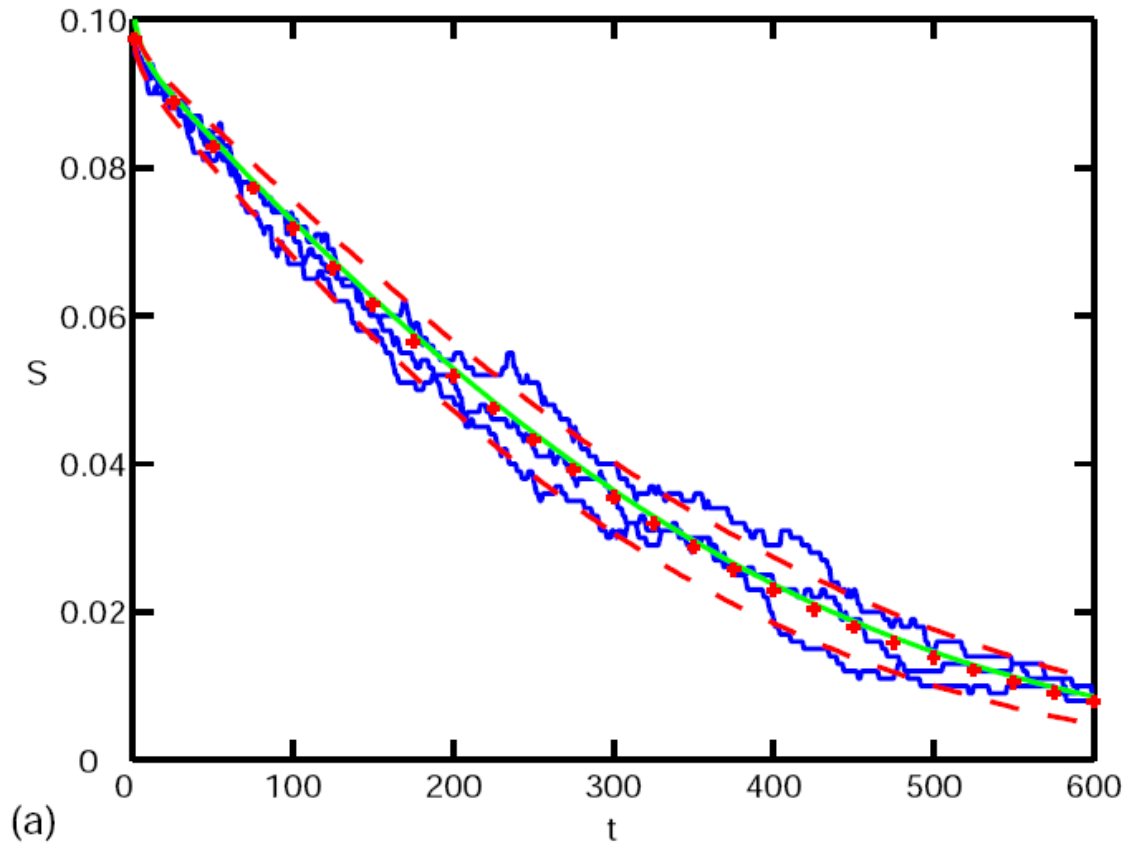
Comparing stochastic simulation and ODEs

Consider a Michaelis Menten reaction in which a substrate S is transformed to a product $\overline{S}$ via a complex C formed with an enzyme E.

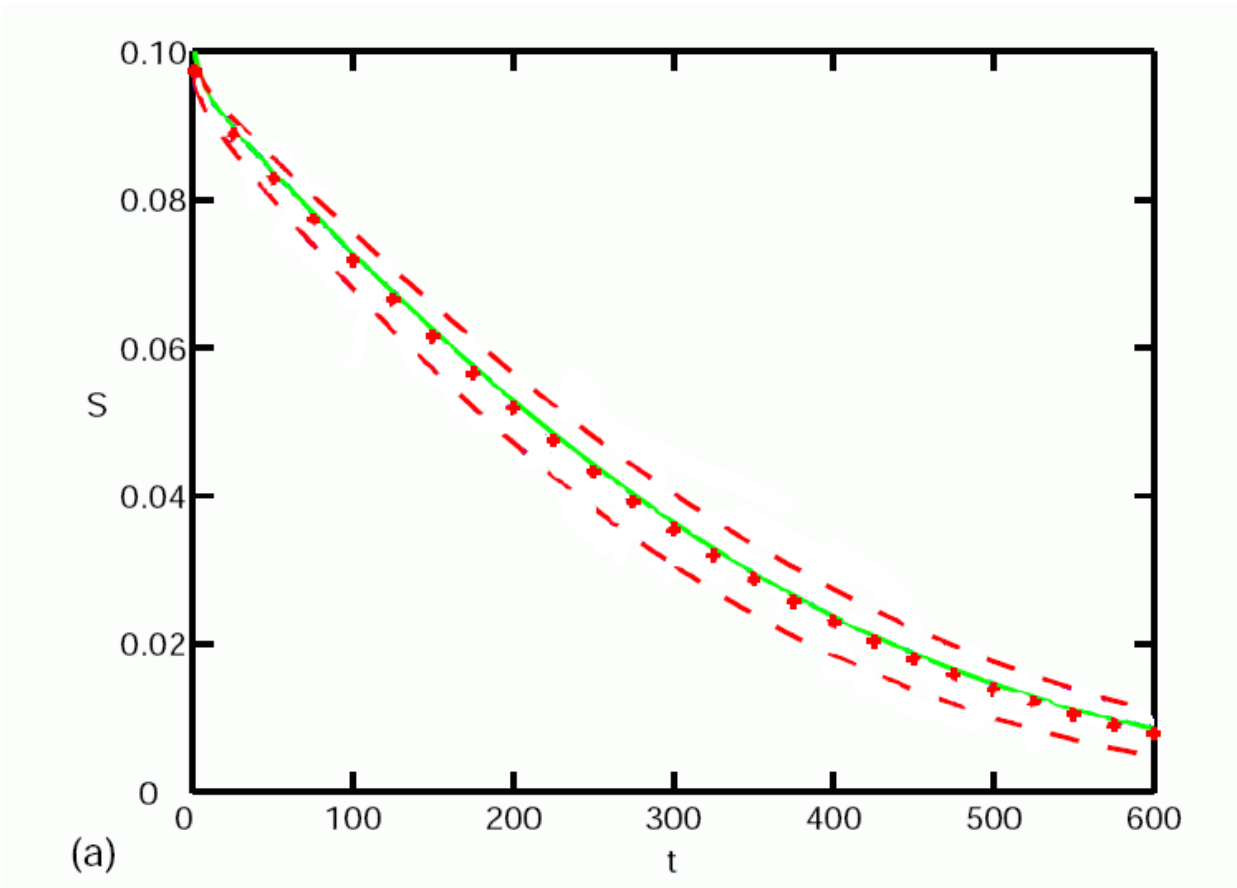


It is relatively straightforward to contrast the results of the two methods. We compare the results of 2000 runs of the stochastic algorithm simulating a system with initial molecular populations $S_0 = 100, E_0 = 10, C_0 = 0, \overline{S}_0 = 0$ and a volume of 1000 units.

Results for $S_0 = 100, E_0 = 10, C_0 = 0, \overline{S}_0 = 0$ (vol 1000)



Results for $S_0 = 100, E_0 = 10, C_0 = 0, \overline{S}_0 = 0$ (vol 1000)



In vivo behaviour compared with model behaviour

However, it is worth bearing in mind that an actual *in vivo* biochemical reaction would follow just one of the many random curves that average together producing the closely fitting mean. This curve may deviate significantly from that of the deterministic approach, and thus call into question its validity.

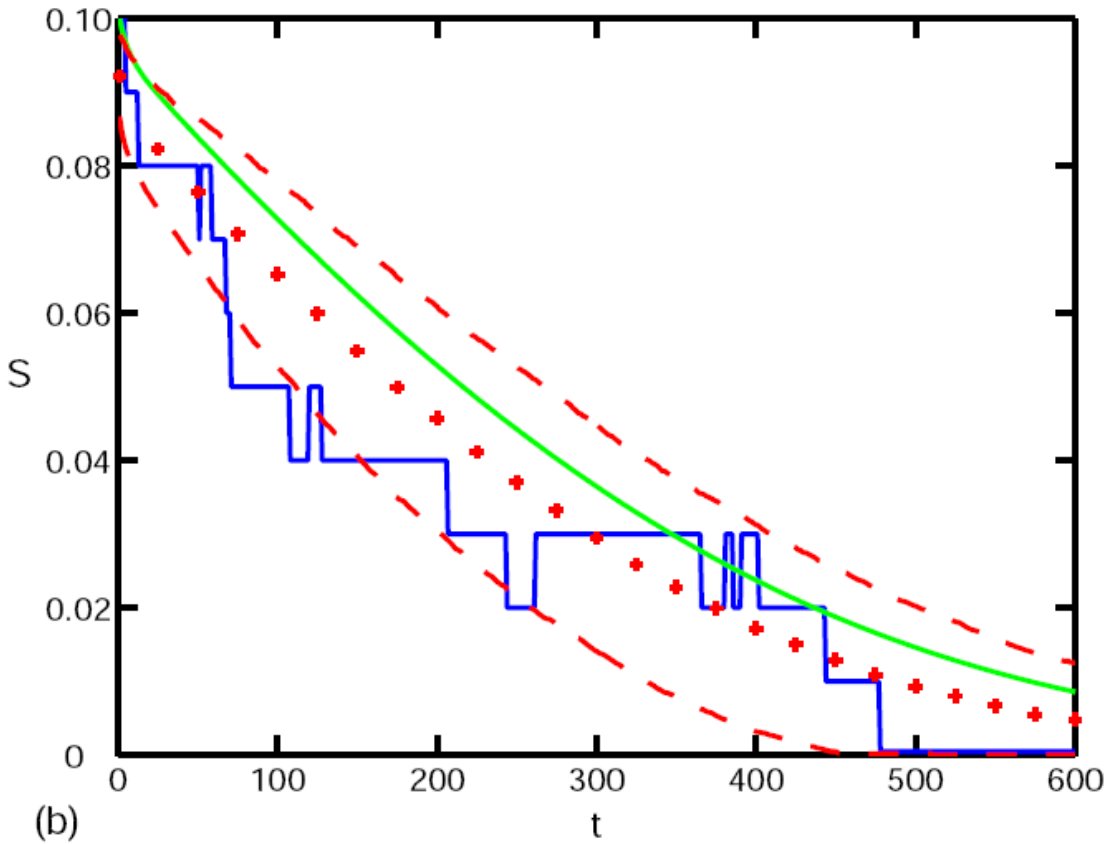
In vivo behaviour compared with model behaviour

However, it is worth bearing in mind that an actual *in vivo* biochemical reaction would follow just one of the many random curves that average together producing the closely fitting mean. This curve may deviate significantly from that of the deterministic approach, and thus call into question its validity.

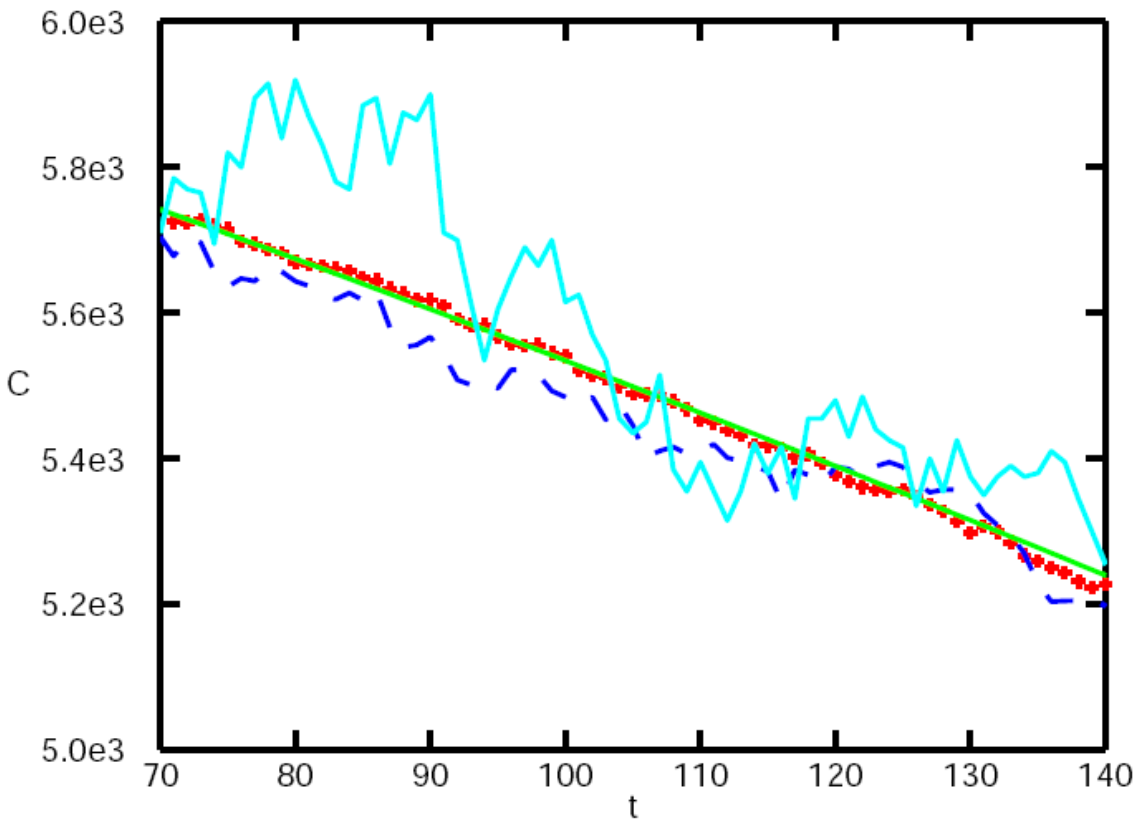
But this does not mean that the randomness exhibited by a particular stochastic simulation trajectory will be the same as the randomness of a particular *in vivo* reaction. Indeed, a set of stochastic simulation trajectories (ensemble) is usually averaged before any conclusions are drawn.

Comparing results at lower population sizes

$S_0 = 10, E_0 = 1, C_0 = 0, \overline{S}_0 = 0$ (vol 100)



Mean results for 11, 110 and 1100 molecules



Circadian clock

- The Vilar-Kueh-Barkai-Leibler (VKBL in short) description of the circadian oscillator incorporates an abstraction of a minimal set of mechanisms for a circadian system.

Circadian clock

- The Vilar-Kueh-Barkai-Leibler (VKBL in short) description of the circadian oscillator incorporates an abstraction of a minimal set of mechanisms for a circadian system.
- The VKBL model involves two genes, an activator A and a repressor R, which are transcribed into $mRNA$ and subsequently translated into proteins.

Circadian clock

- The Vilar-Kueh-Barkai-Leibler (VKBL in short) description of the circadian oscillator incorporates an abstraction of a minimal set of mechanisms for a circadian system.
- The VKBL model involves two genes, an activator A and a repressor R, which are transcribed into *mRNA* and subsequently translated into proteins.

Circadian clock

- The Vilar-Kueh-Barkai-Leibler (VKBL in short) description of the circadian oscillator incorporates an abstraction of a minimal set of mechanisms for a circadian system.
- The VKBL model involves two genes, an activator A and a repressor R, which are transcribed into *mRNA* and subsequently translated into proteins.
- The activator A binds to the A and R promoters and increases their expression rate. Thus, A implements a positive loop acting on its own transcription.

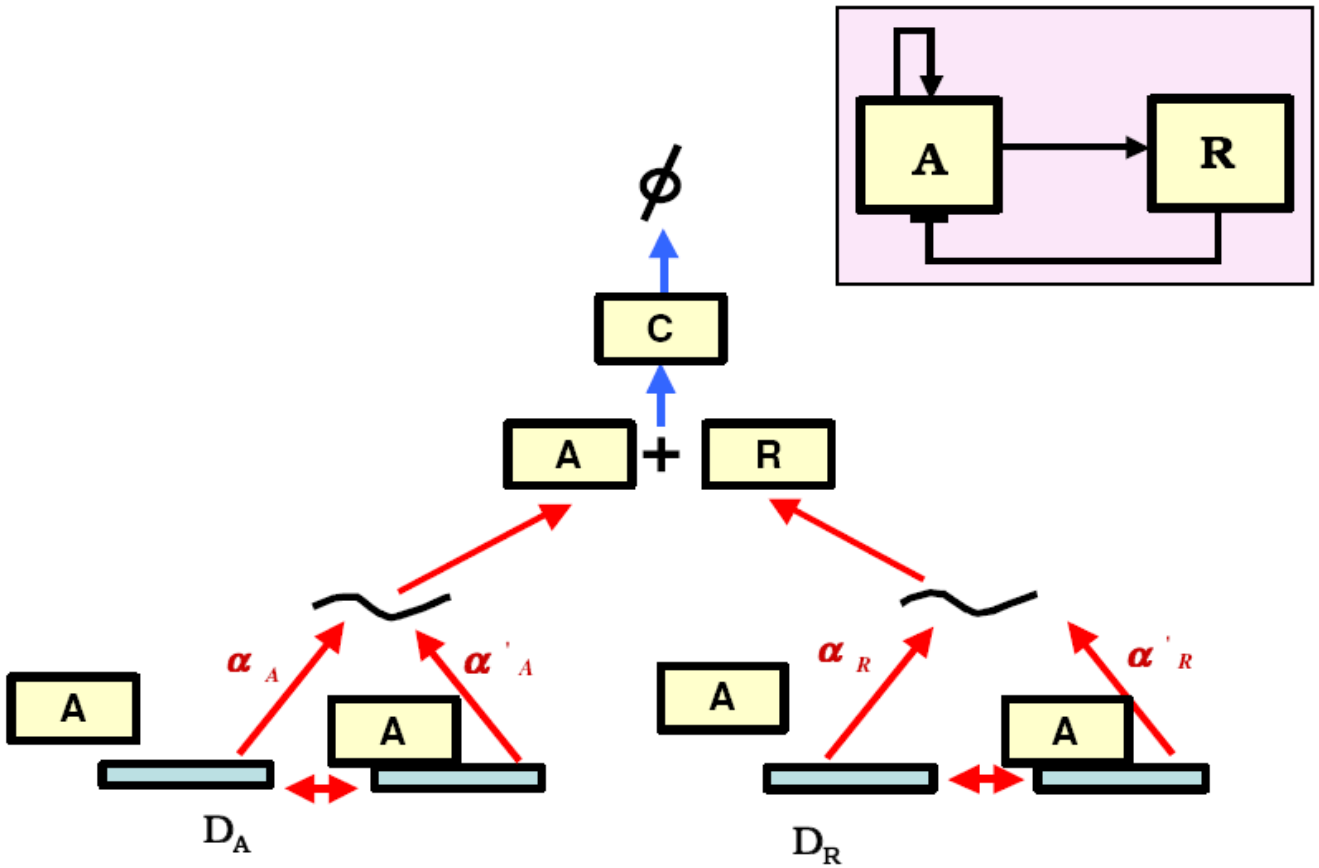
Circadian clock

- The Vilar-Kueh-Barkai-Leibler (VKBL in short) description of the circadian oscillator incorporates an abstraction of a minimal set of mechanisms for a circadian system.
- The VKBL model involves two genes, an activator A and a repressor R, which are transcribed into *mRNA* and subsequently translated into proteins.
- The activator A binds to the A and R promoters and increases their expression rate. Thus, A implements a positive loop acting on its own transcription.

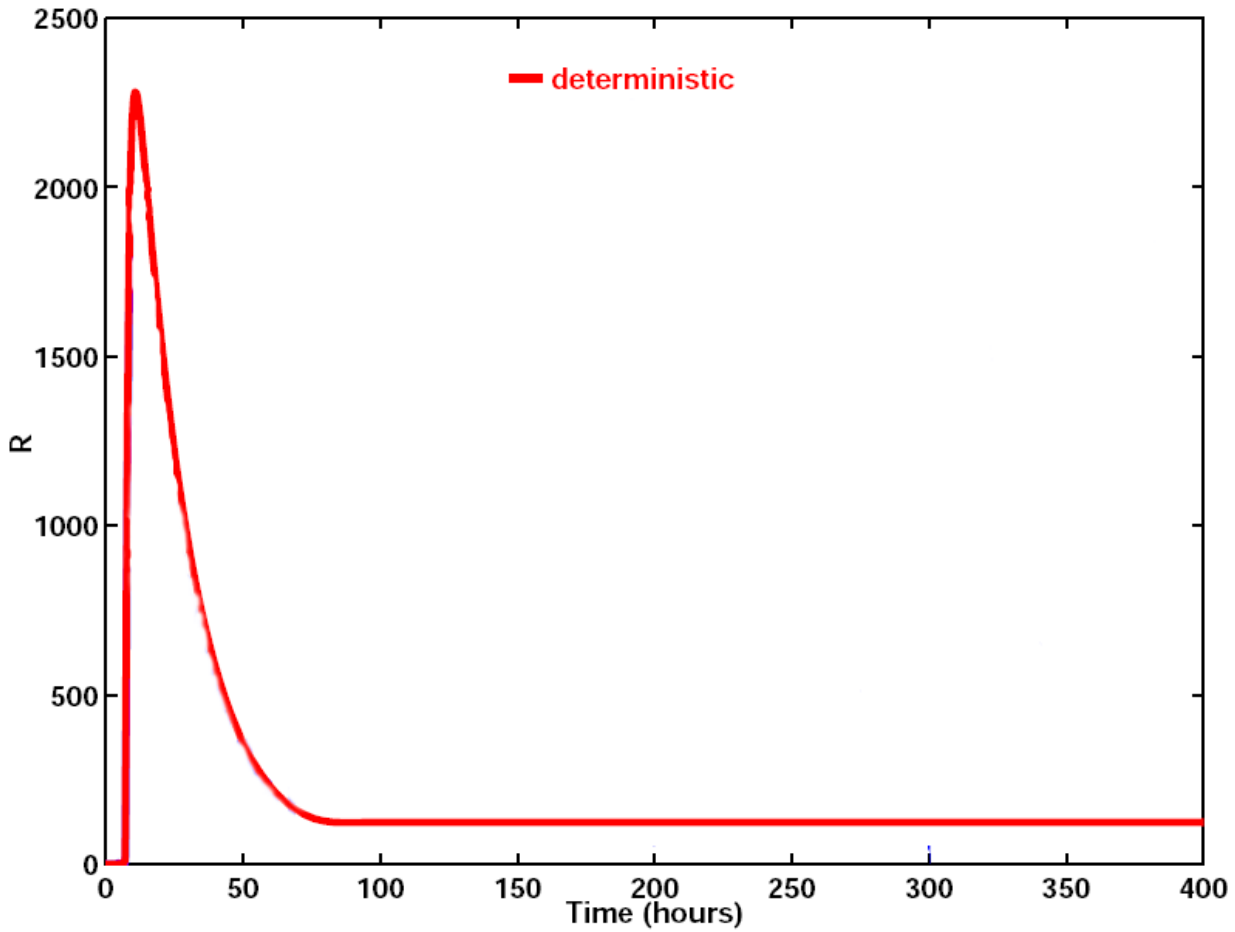
Circadian clock

- The Vilar-Kueh-Barkai-Leibler (VKBL in short) description of the circadian oscillator incorporates an abstraction of a minimal set of mechanisms for a circadian system.
- The VKBL model involves two genes, an activator A and a repressor R, which are transcribed into *mRNA* and subsequently translated into proteins.
- The activator A binds to the A and R promoters and increases their expression rate. Thus, A implements a positive loop acting on its own transcription.
- Conversely, R sequesters A to form a complex C, therefore inhibiting it from binding to the gene promoter and acting as a negative feedback loop.

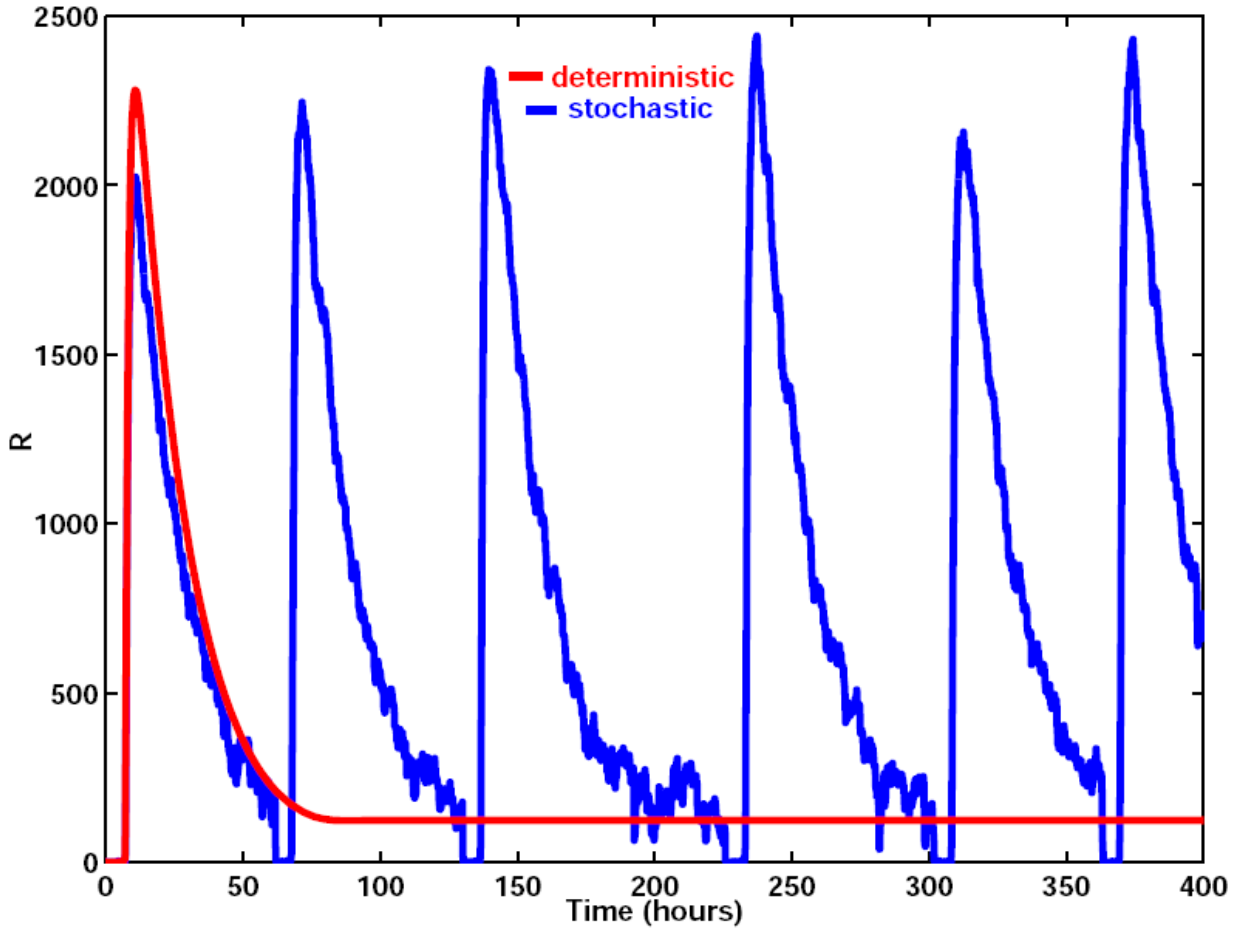
Circadian clock (cartoon)



Circadian clock (deterministically . . .)



Circadian clock (. . . and stochastically)



Conclusions from the Circadian Clock

- For some parameter values a differential equation model exhibits autonomous oscillations.
- These oscillations disappear from the deterministic model as the degradation rate of the repressor δ_R is decreased.
- The system of ODEs undergoes a bifurcation at this point and exhibits a unique stable deterministic equilibrium.
- However, if the effects of molecular noise are incorporated the oscillations in the stochastic system pertain.
- This phenomenon is a manifestation of coherence resonance, and illustrates the crucial interplay between noise and dynamics.

Modularity vs. Infinite Regress

As computer scientists we are firm believers in modularity and compositionality. When it comes to biochemical pathways opinion amongst biologists is divided about whether is makes sense to take a modular view of cellular pathways.

Modularity vs. Infinite Regress

As computer scientists we are firm believers in modularity and compositionality. When it comes to biochemical pathways opinion amongst biologists is divided about whether is makes sense to take a modular view of cellular pathways.

Some biologists (e.g. Leibler) argue that there is modularity, naturally occuring, where they define a module relative to a biological function.

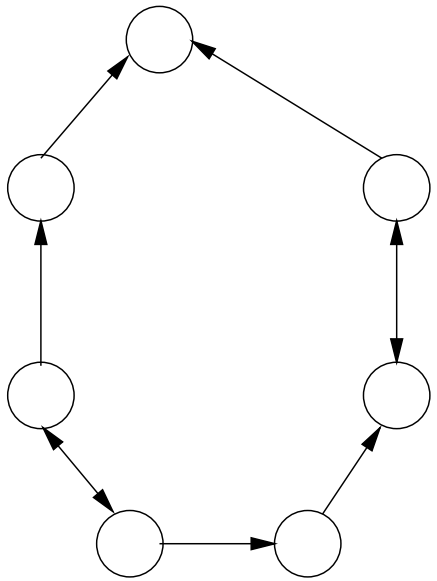
Modularity vs. Infinite Regress

As computer scientists we are firm believers in modularity and compositionality. When it comes to biochemical pathways opinion amongst biologists is divided about whether is makes sense to take a modular view of cellular pathways.

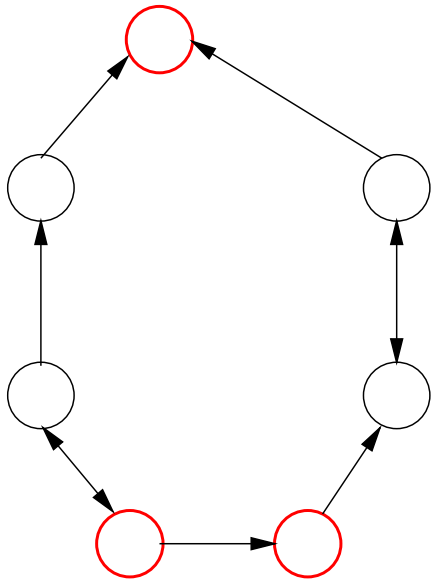
Some biologists (e.g. Leibler) argue that there is modularity, naturally occuring, where they define a module relative to a biological function.

Others such as Cornish-Bowden are much more skeptical and cite the problem of infinite regress as being insurmountable.

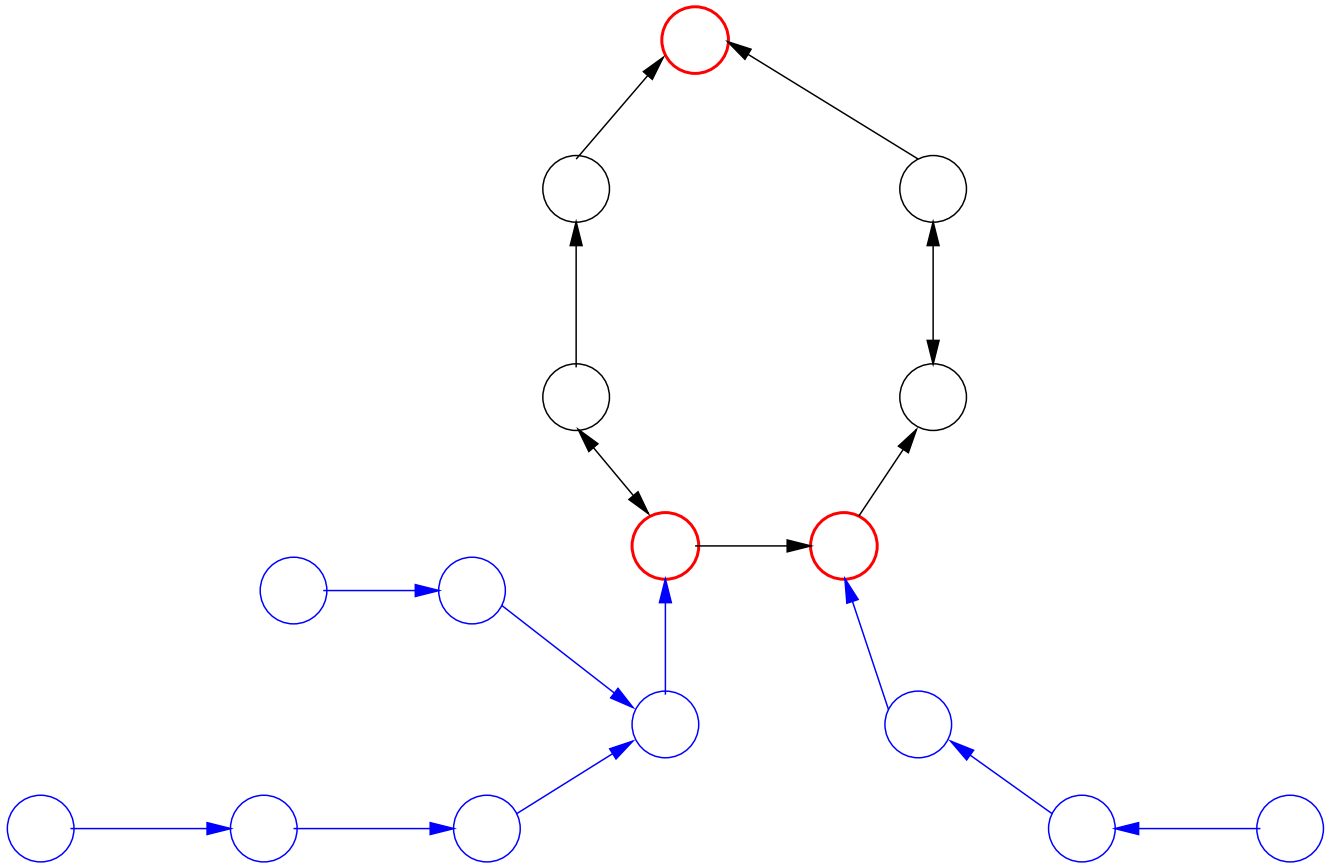
The problem of *Infinite Regress*



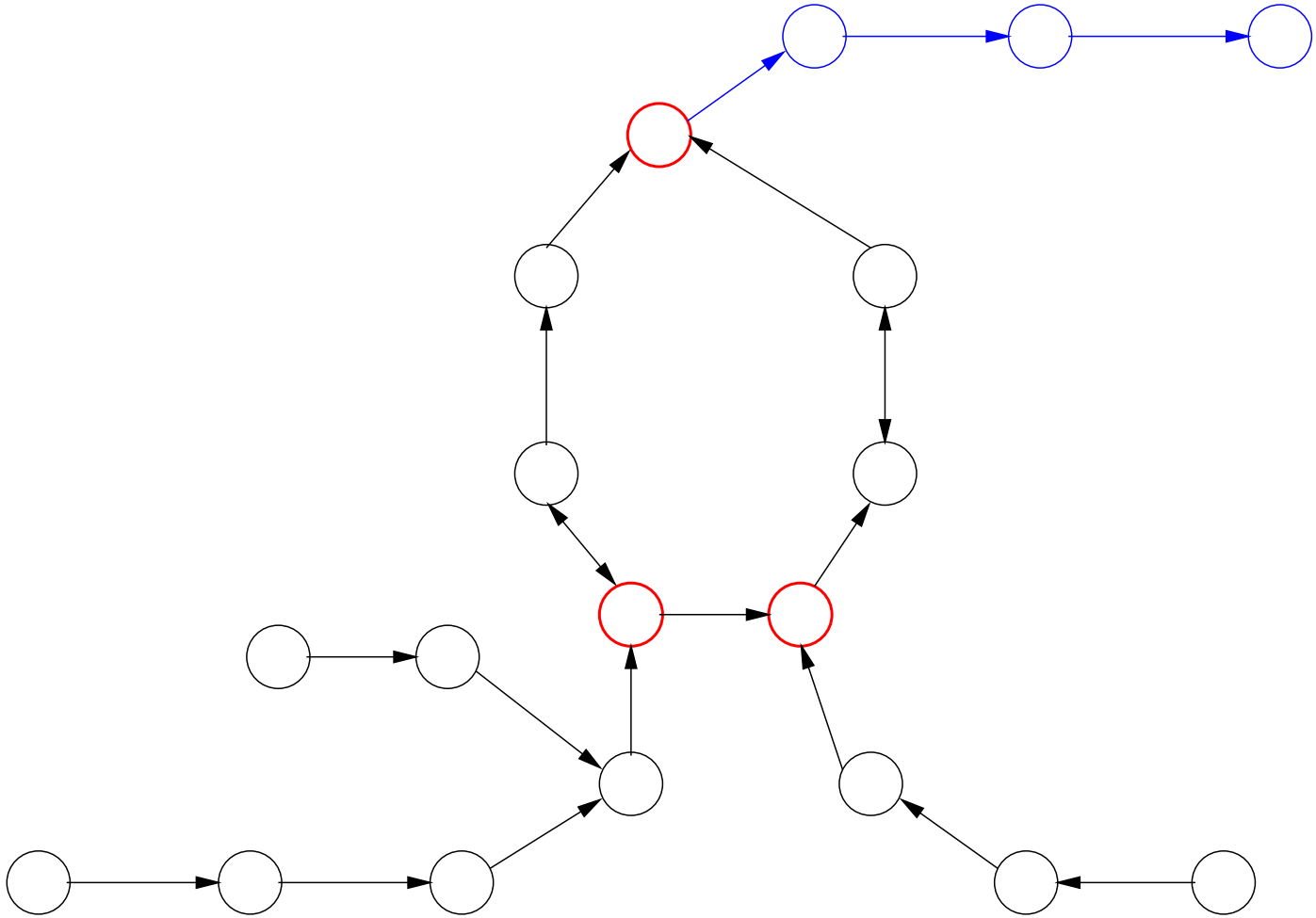
The problem of *Infinite Regress*



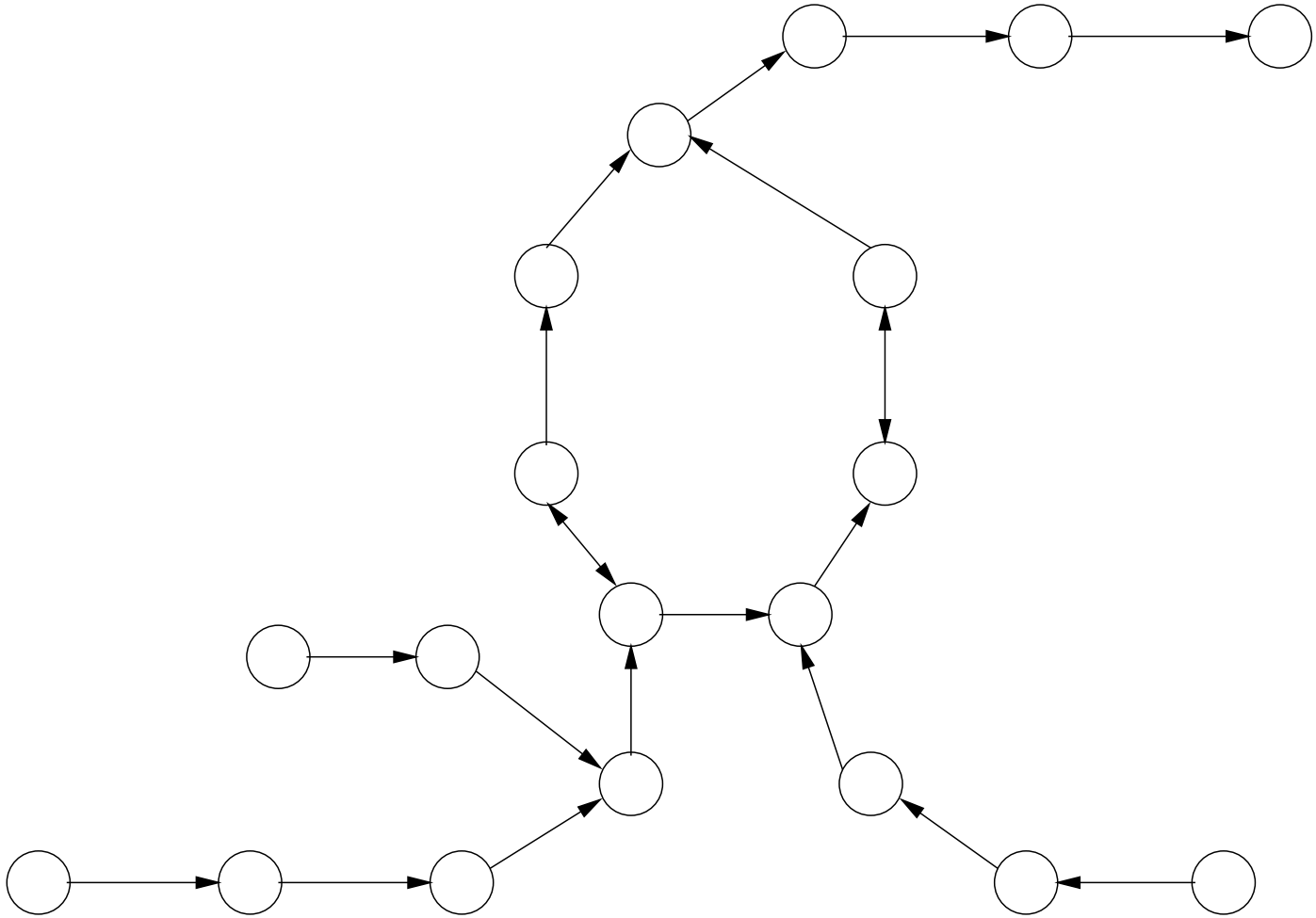
The problem of *Infinite Regress*



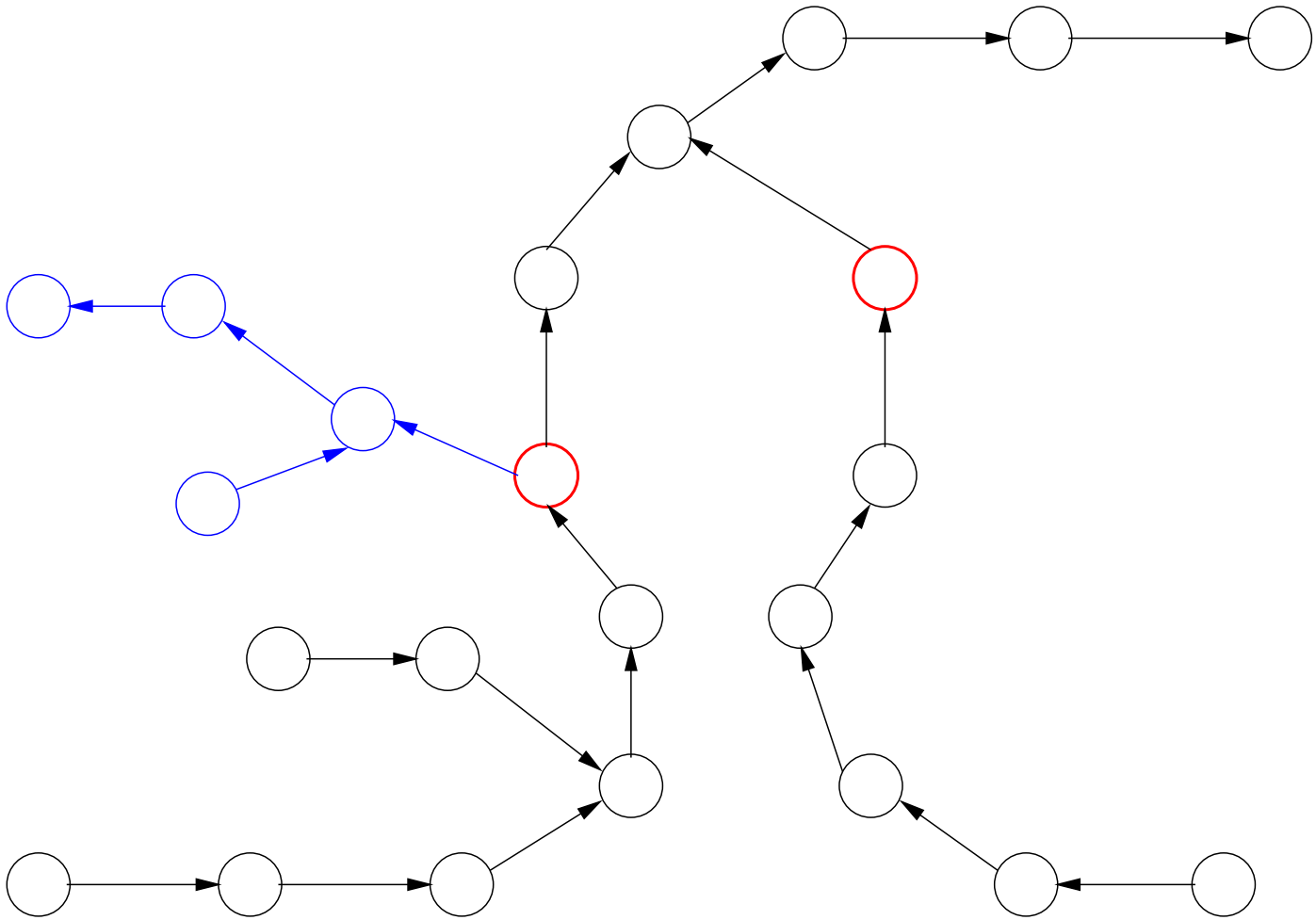
The problem of *Infinite Regress*



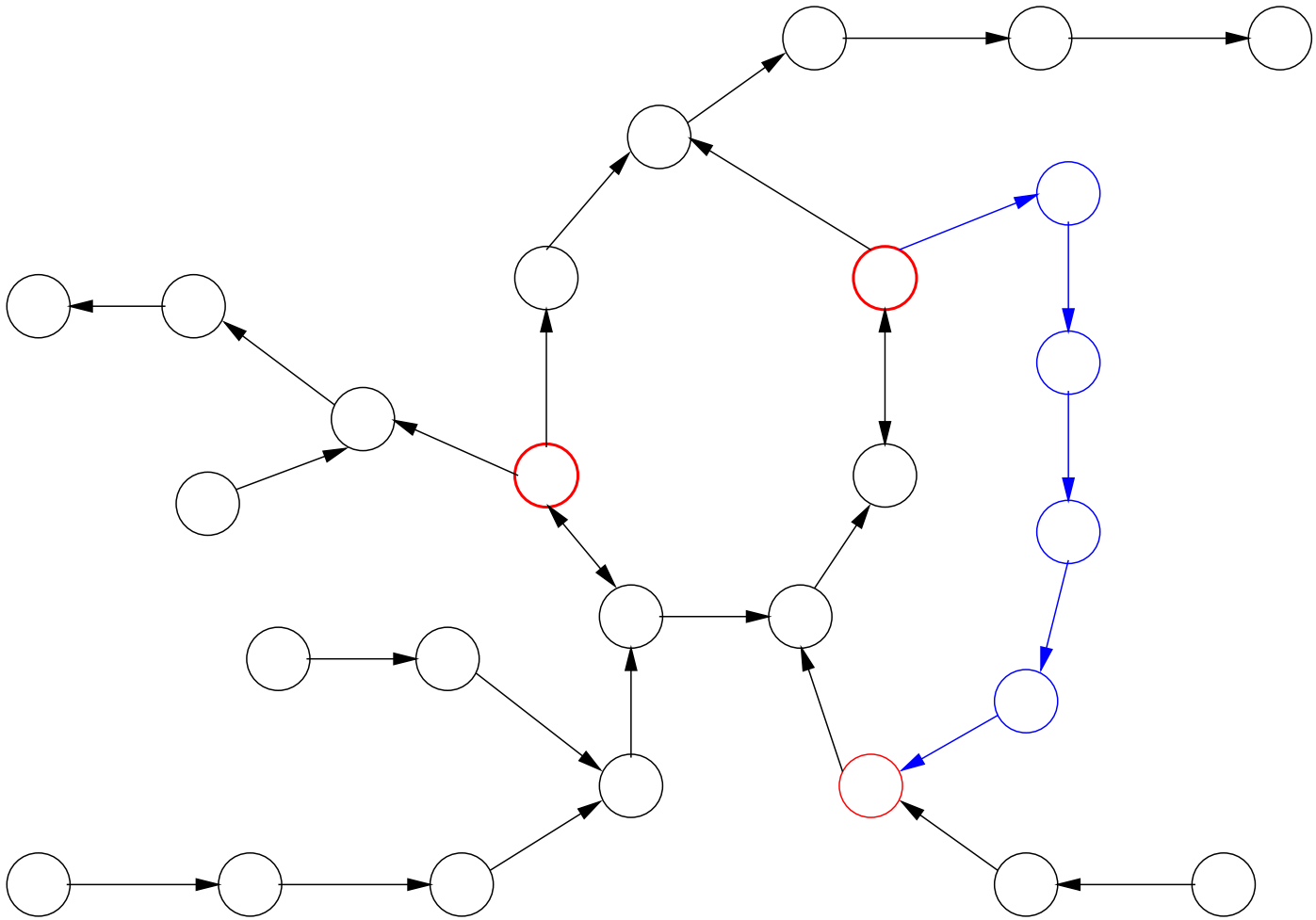
The problem of *Infinite Regress*



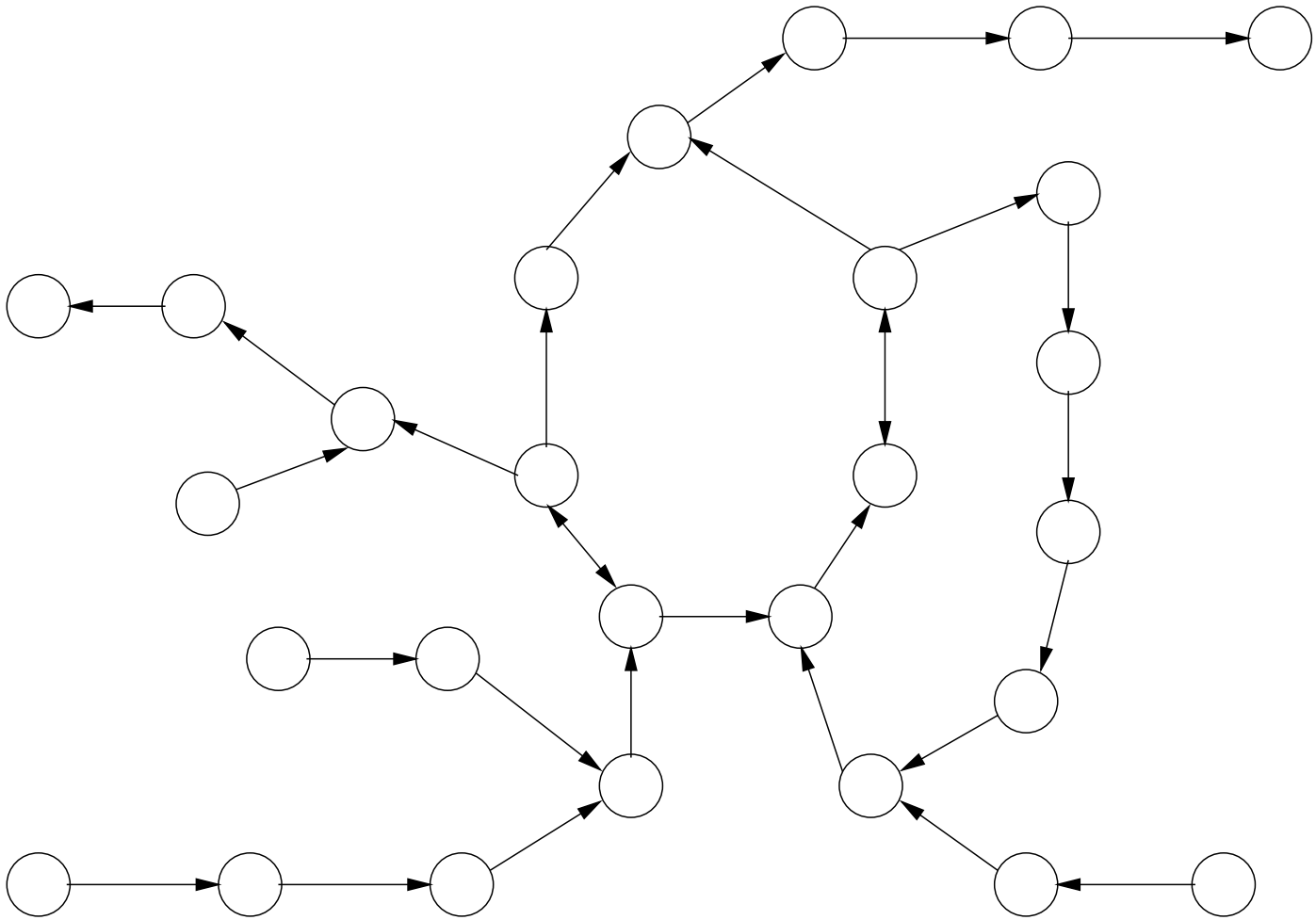
The problem of *Infinite Regress*



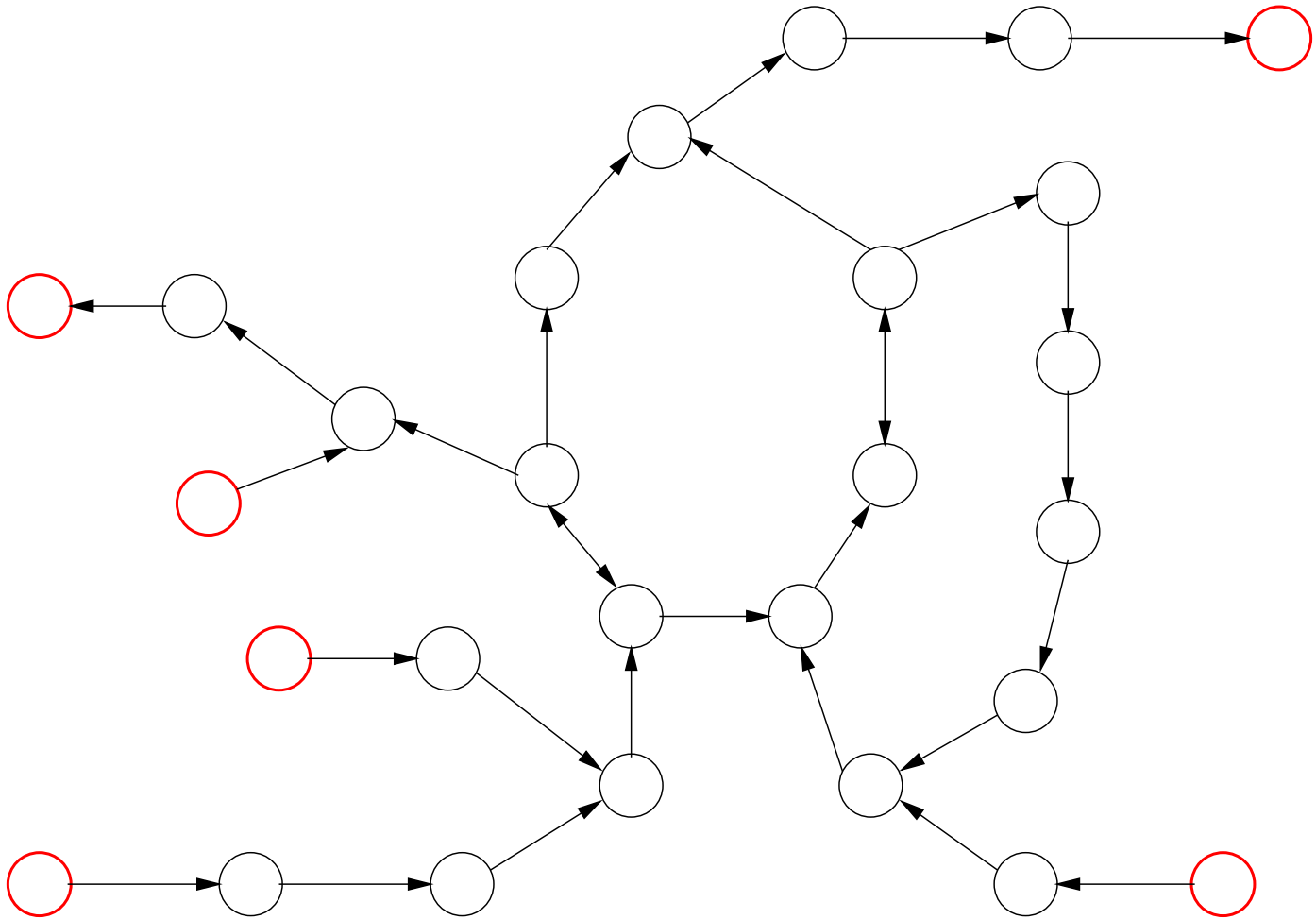
The problem of *Infinite Regress*



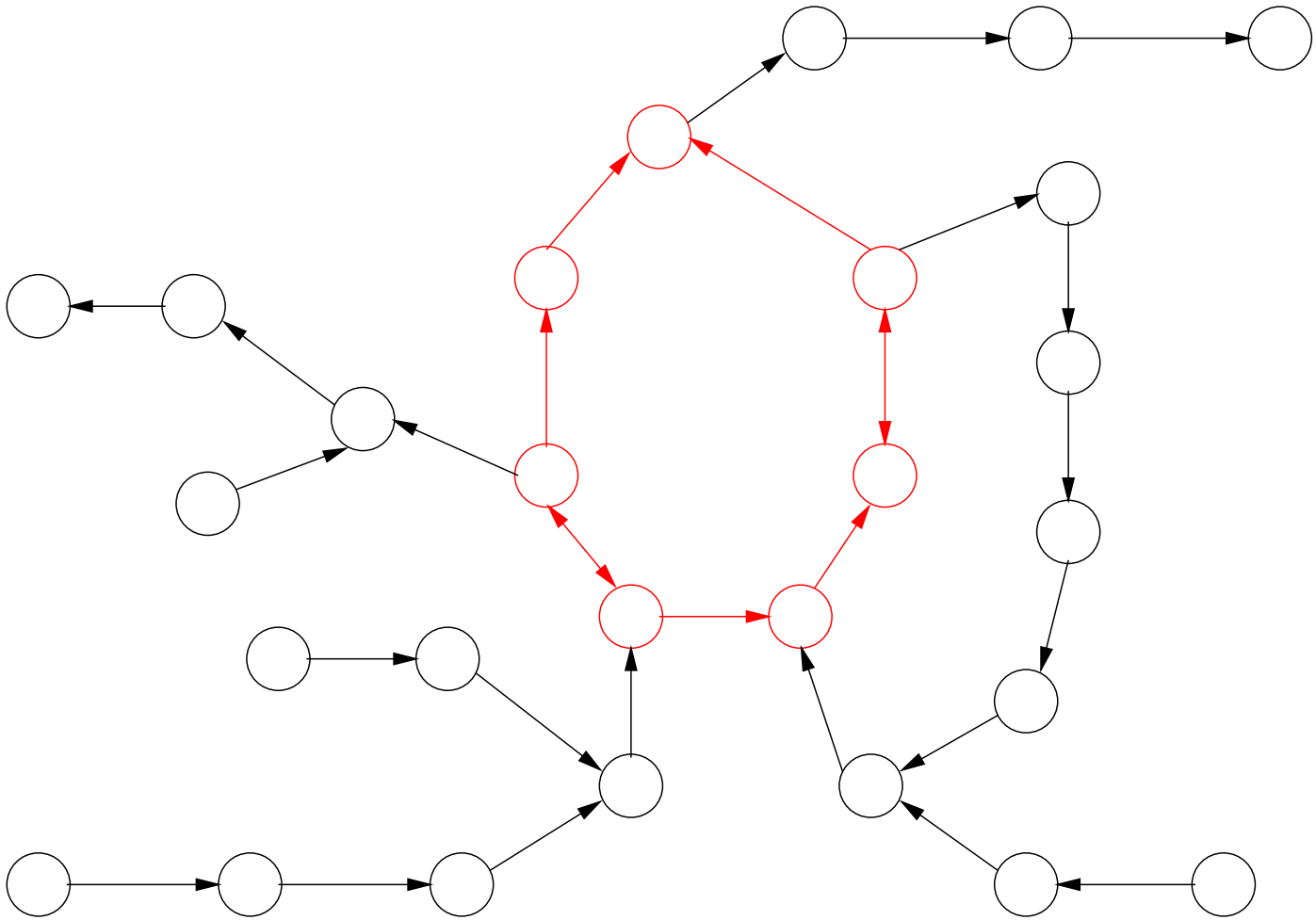
The problem of *Infinite Regress*



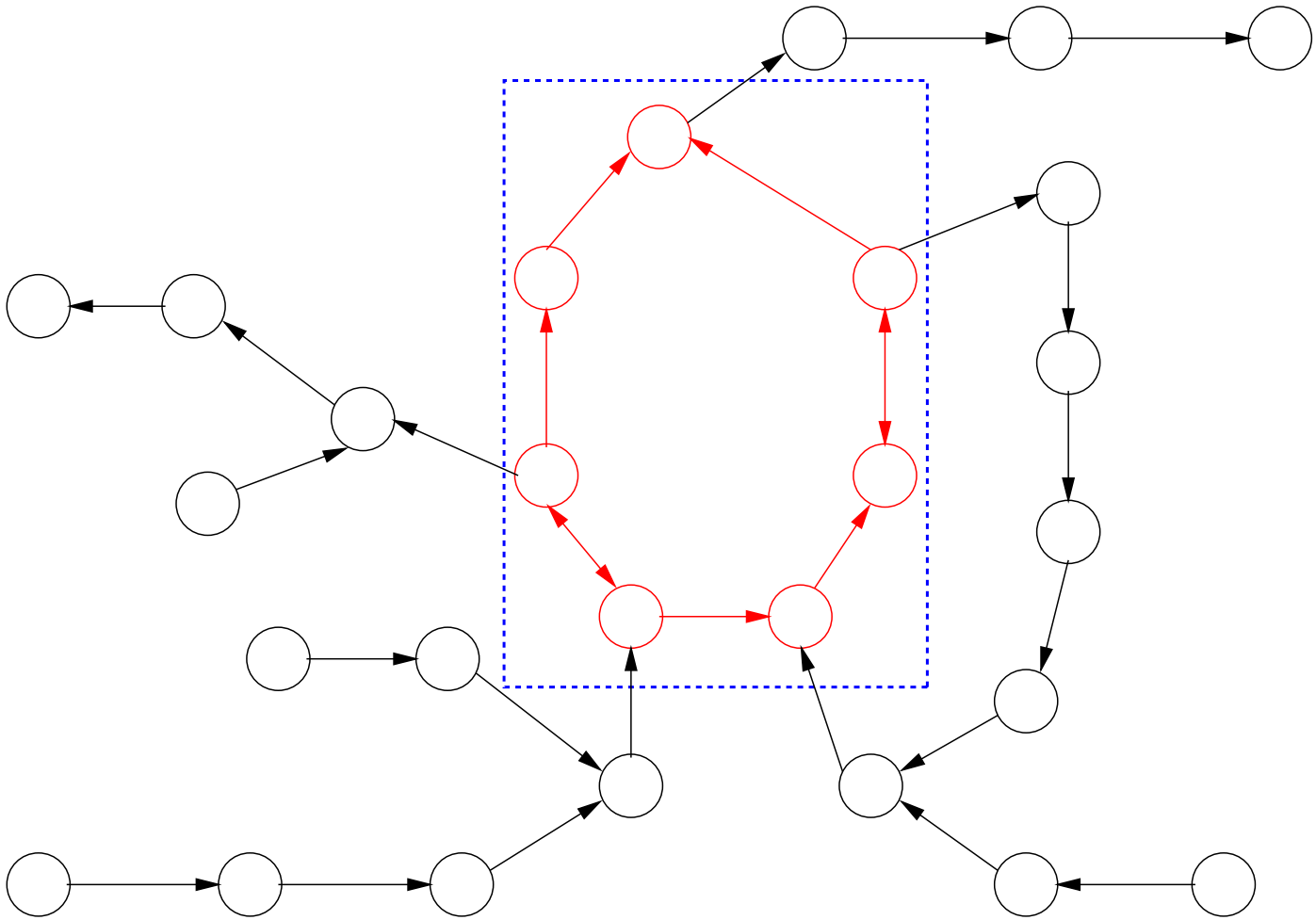
The problem of *Infinite Regress*



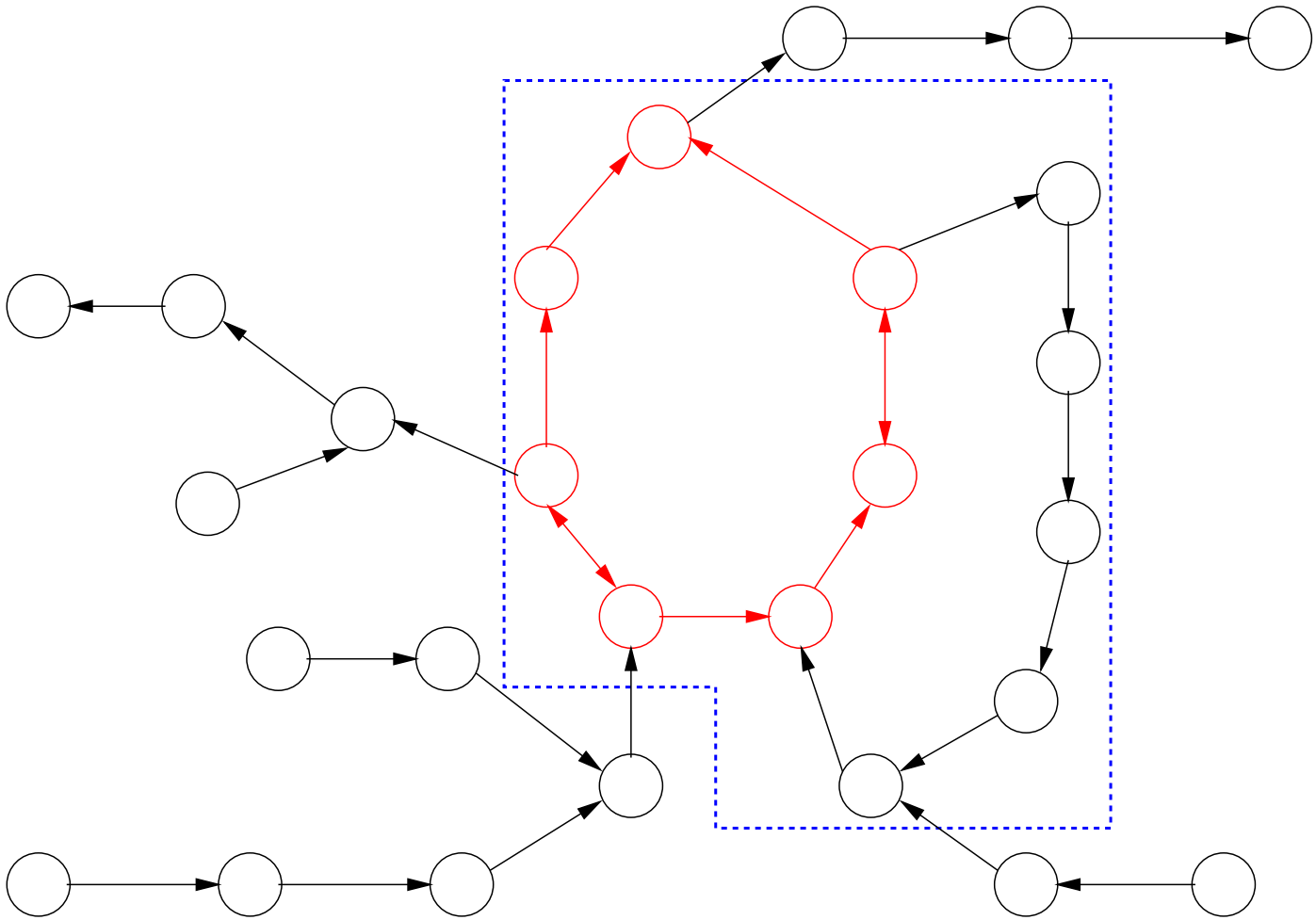
The problem of *Infinite Regress*



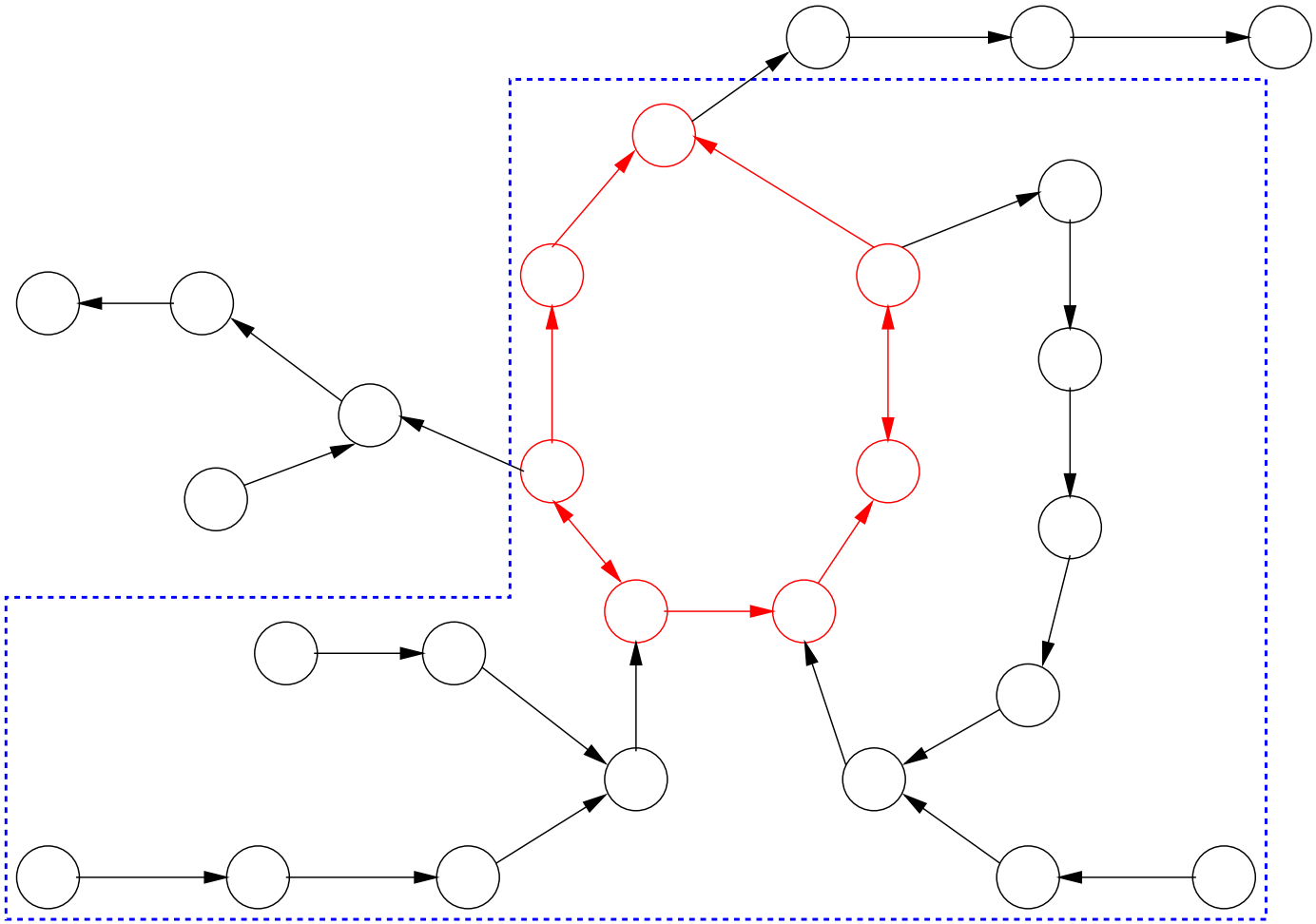
The problem of *Infinite Regress*



The problem of *Infinite Regress*



The problem of *Infinite Regress*



Dealing with the Unknown

There is a fundamental challenge when modelling cellular pathways that little is known about some aspects of cellular processes.

Dealing with the Unknown

There is a fundamental challenge when modelling cellular pathways that little is known about some aspects of cellular processes.

In some cases this is because no experimental data is available, or that the experimental data that is available is inconsistent.

Dealing with the Unknown

There is a fundamental challenge when modelling cellular pathways that little is known about some aspects of cellular processes.

In some cases this is because no experimental data is available, or that the experimental data that is available is inconsistent.

In other cases the data is unknowable because experimental techniques do not yet exist to collect the data, or those that do involve modification to the system.

Dealing with the Unknown

There is a fundamental challenge when modelling cellular pathways that little is known about some aspects of cellular processes.

In some cases this is because no experimental data is available, or that the experimental data that is available is inconsistent.

In other cases the data is unknowable because experimental techniques do not yet exist to collect the data, or those that do involve modification to the system.

Even when data exists the quality is often very poor.

Outline

Using Stochastic Process Algebras

Process algebras have several attractive features which could be useful for modelling and understanding biological systems:

Using Stochastic Process Algebras

Process algebras have several attractive features which could be useful for modelling and understanding biological systems:

- Process algebraic formulations are compositional and make interactions/constraints explicit.

Using Stochastic Process Algebras

Process algebras have several attractive features which could be useful for modelling and understanding biological systems:

- Process algebraic formulations are compositional and make interactions/constraints explicit.
- Structure can also be apparent.

Using Stochastic Process Algebras

Process algebras have several attractive features which could be useful for modelling and understanding biological systems:

- Process algebraic formulations are compositional and make interactions/constraints explicit.
- Structure can also be apparent.
- Equivalence relations allow formal comparison of high-level descriptions.

Using Stochastic Process Algebras

Process algebras have several attractive features which could be useful for modelling and understanding biological systems:

- Process algebraic formulations are compositional and make interactions/constraints explicit.
- Structure can also be apparent.
- Equivalence relations allow formal comparison of high-level descriptions.
- There are well-established techniques for reasoning about the behaviours and properties of models, supported by software. These include qualitative and quantitative analysis, and model checking.

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

The language may be used to generate a Markov Process (CTMC).

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

The language may be used to generate a Markov Process (CTMC).

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

The language may be used to generate a Markov Process (CTMC).

PEPA
MODEL

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

The language may be used to generate a Markov Process (CTMC).

PEPA MODEL $\xrightarrow{\text{SOS rules}}$

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

The language may be used to generate a Markov Process (CTMC).

PEPA MODEL $\xrightarrow{\text{SOS rules}}$ LABELLED TRANSITION SYSTEM

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

The language may be used to generate a Markov Process (CTMC).

PEPA MODEL $\xrightarrow{\text{SOS rules}}$ LABELLED TRANSITION SYSTEM $\xrightarrow{\text{state transition diagram}}$

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

The language may be used to generate a Markov Process (CTMC).

PEPA MODEL $\xrightarrow{\text{SOS rules}}$ LABELLED TRANSITION SYSTEM $\xrightarrow[\text{diagram}]{\text{state transition}}$ CTMC **Q**

PEPA: Performance Evaluation Process Algebra

$$
\begin{aligned}
S &::= (\alpha, r).S \mid S + S \mid A \\
P &::= S \mid P \underset{L}{\bowtie} P \mid P/L
\end{aligned}
$$

The language may be used to generate a Markov Process (CTMC).

PEPA MODEL $\xrightarrow{\text{SOS rules}}$ LABELLED TRANSITION SYSTEM $\xrightarrow[\text{diagram}]{\text{state transition}}$ CTMC **Q**

Q is the infinitesimal generator matrix characterising the CTMC.

Markovian analysis

- Analysis of the Markov process can yield quite detailed information about the dynamic behaviour of the model.

Markovian analysis

- Analysis of the Markov process can yield quite detailed information about the dynamic behaviour of the model.
- A steady state analysis provides statistics for average behaviour over a long run of the system, when the bias introduced by the initial state has been lost.

Markovian analysis

- Analysis of the Markov process can yield quite detailed information about the dynamic behaviour of the model.
- A steady state analysis provides statistics for average behaviour over a long run of the system, when the bias introduced by the initial state has been lost.
- A transient analysis provides statistics relating to the evolution of the model over a fixed period. This will be dependent on the starting state.

Markovian analysis

- Analysis of the Markov process can yield quite detailed information about the dynamic behaviour of the model.
- A steady state analysis provides statistics for average behaviour over a long run of the system, when the bias introduced by the initial state has been lost.
- A transient analysis provides statistics relating to the evolution of the model over a fixed period. This will be dependent on the starting state.
- Note, transient Markovian analysis is exact because it takes account of all possible evolutions, unlike a stochastic simulation which considers only one possible evolution in each run.

Molecular processes as concurrent computations

Concurrency	**Molecular Biology**	**Metabolism**	**Signal Transduction**
Concurrent computational processes	Molecules	Enzymes and metabolites	Interacting proteins
Synchronous communication	Molecular interaction	Binding and catalysis	Binding and catalysis
Transition or mobility	Biochemical modification or relocation	Metabolite synthesis	Protein binding, modification or sequestration

[Regev *et al* 2000]

Molecular processes as concurrent computations

Concurrency	**Molecular Biology**	**Metabolism**	**Signal Transduction**
Concurrent computational processes	Molecules	Enzymes and metabolites	Interacting proteins
Synchronous communication	Molecular interaction	Binding and catalysis	Binding and catalysis
Transition or mobility	Biochemical modification or relocation	Metabolite synthesis	Protein binding, modification or sequestration

[Regev *et al* 2000]

Motivations for Abstraction

Our motivations for seeking more abstraction in process algebra models for systems biology comes from both key aspects of modelling:

Motivations for Abstraction

Our motivations for seeking more abstraction in process algebra models for systems biology comes from both key aspects of modelling:

- The data that we have available to parameterise models is sometimes speculative rather than precise.

Motivations for Abstraction

Our motivations for seeking more abstraction in process algebra models for systems biology comes from both key aspects of modelling:

- The data that we have available to parameterise models is sometimes speculative rather than precise. This suggests that we should use semiquantitative models rather than quantitative ones.

Motivations for Abstraction

Our motivations for seeking more abstraction in process algebra models for systems biology comes from both key aspects of modelling:

- The data that we have available to parameterise models is sometimes speculative rather than precise. This suggests that we should use semiquantitative models rather than quantitative ones.
- Process algebra based analyses such as comparing models (e.g. for equivalence or simulation) and model checking are only possible is the state space is not prohibitively large.

Motivations for Abstraction

Our motivations for seeking more abstraction in process algebra models for systems biology comes from both key aspects of modelling:

- The data that we have available to parameterise models is sometimes speculative rather than precise. This suggests that we should use semiquantitative models rather than quantitative ones.
- Process algebra based analyses such as comparing models (e.g. for equivalence or simulation) and model checking are only possible is the state space is not prohibitively large.

Mapping biological systems to process algebra

The work using the stochastic π-calculus and related calculi, for modelling biochemical signalling within cells maps a molecule in a pathway to a process in the process algebra description.

Mapping biological systems to process algebra

The work using the stochastic π-calculus and related calculi, for modelling biochemical signalling within cells maps a molecule in a pathway to a process in the process algebra description.

This is an inherently individuals-based view of the system and assumes analysis will be via stochastic simulation.

Mapping biological systems to process algebra

The work using the stochastic π-calculus and related calculi, for modelling biochemical signalling within cells maps a molecule in a pathway to a process in the process algebra description.

This is an inherently individuals-based view of the system and assumes analysis will be via stochastic simulation.

In the PEPA modelling we have been doing we have experimented with more abstract mappings between process algebra constructs and elements of signalling pathways.

Mapping biological systems to process algebra

The work using the stochastic π-calculus and related calculi, for modelling biochemical signalling within cells maps a molecule in a pathway to a process in the process algebra description.

This is an inherently individuals-based view of the system and assumes analysis will be via stochastic simulation.

In the PEPA modelling we have been doing we have experimented with more abstract mappings between process algebra constructs and elements of signalling pathways.

In our mapping we focus on species (c.f. a type rather than an instance, or a class rather than an object).

Discretising the population view

We can discretise the continuous range of possible concentration values into a number of distinct states. These form the possible states of the component representing the reagent.

Reagent-centric modelling [CGH04]

Reagent role	*Impact on reagent*	*Impact on reaction rate*
Producer	decreases concentration	has a positive impact, i.e. proportional to current concentration
Product	increases concentration	has no impact on the rate, except at saturation
Enzyme	concentration unchanged	has a positive impact, i.e. proportional to current concentration
Inhibitor	concentration unchanged	has a negative impact, i.e. inversely proportional to current concentration

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a Markov Process (CTMC).

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a Markov Process (CTMC).

SPA
MODEL

Deriving quantitative data

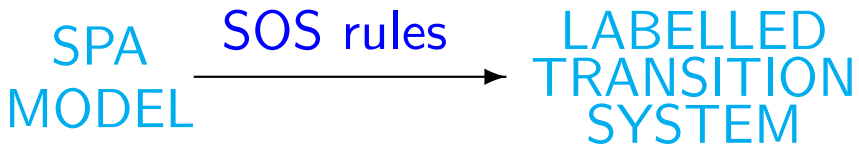
PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a Markov Process (CTMC).

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a Markov Process (CTMC).



Deriving quantitative data

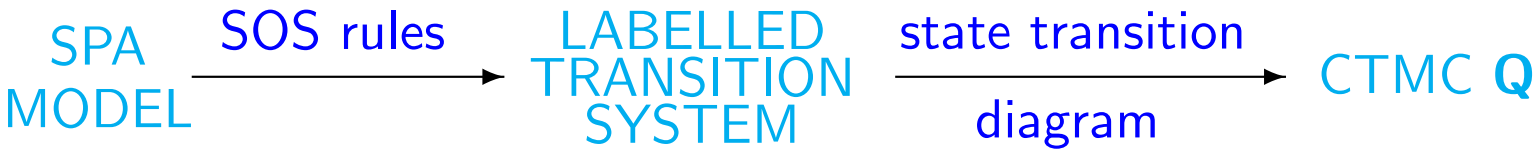
PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a Markov Process (CTMC).

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a Markov Process (CTMC).



Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a system of ordinary differential equations (ODEs).

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

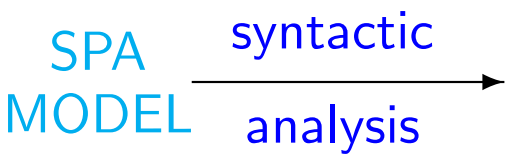
The language may be used to generate a system of ordinary differential equations (ODEs).

SPA
MODEL

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

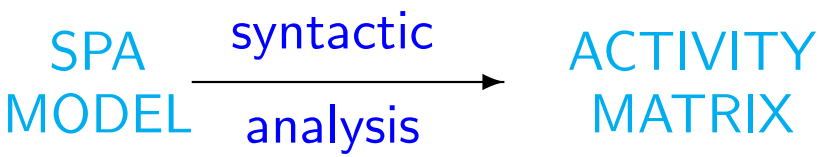
The language may be used to generate a system of ordinary differential equations (ODEs).



Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a system of ordinary differential equations (ODEs).



Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a system of ordinary differential equations (ODEs).

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language may be used to generate a system of ordinary differential equations (ODEs).

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language also may be used to generate a stochastic simulation.

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

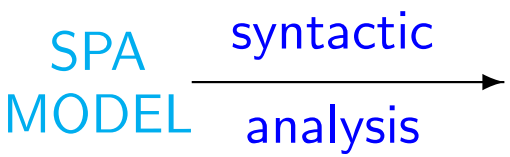
The language also may be used to generate a stochastic simulation.

SPA
MODEL

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

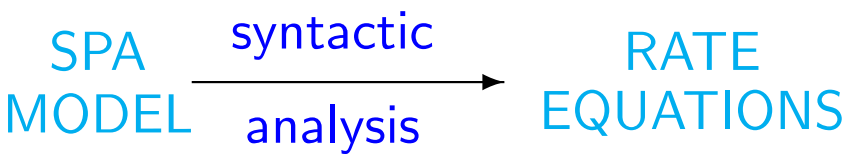
The language also may be used to generate a stochastic simulation.



Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language also may be used to generate a stochastic simulation.



Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language also may be used to generate a stochastic simulation.

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

The language also may be used to generate a stochastic simulation.

Deriving quantitative data

PEPA models can be analysed for quantified dynamic behaviour in a number of different ways.

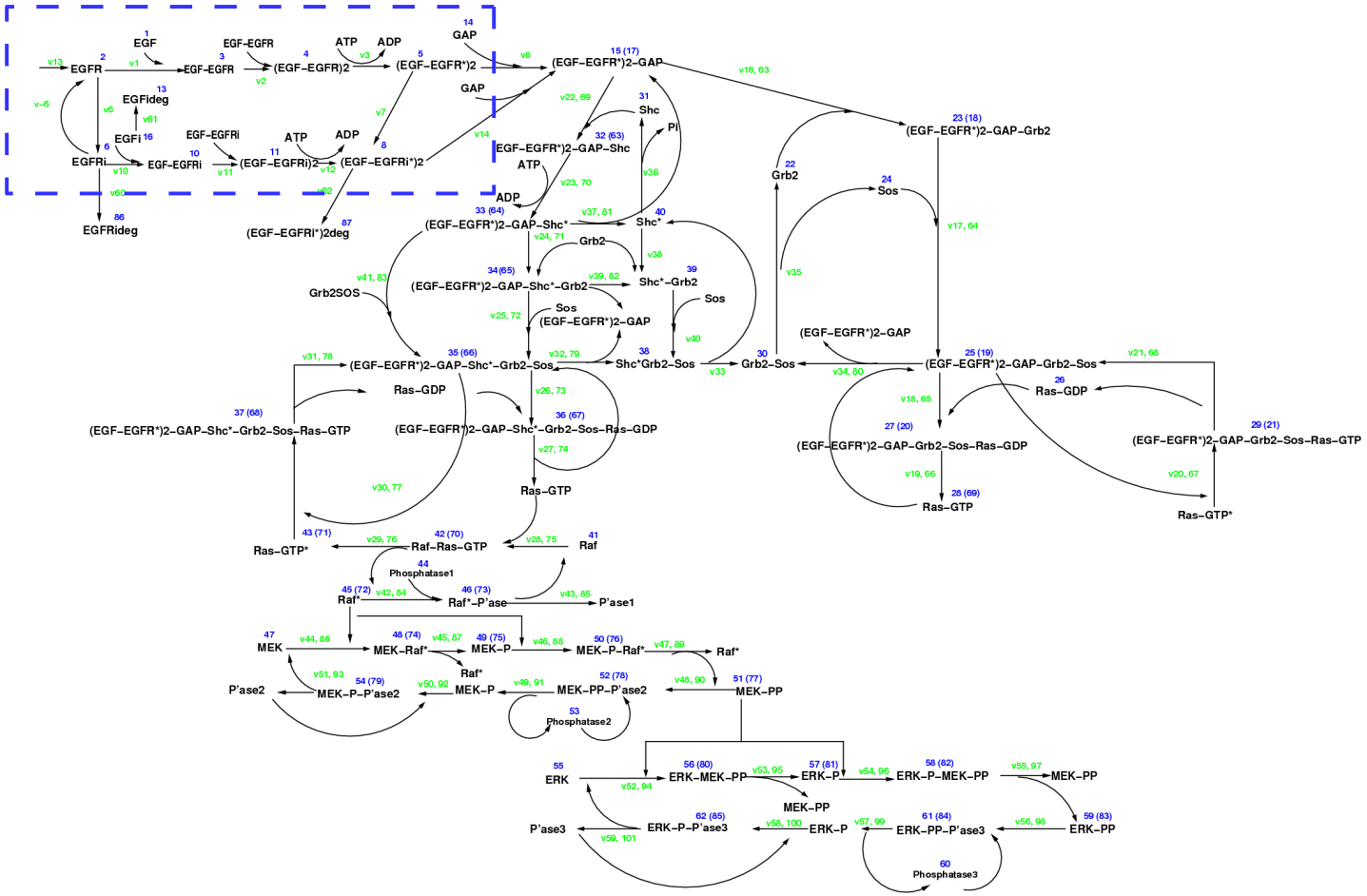
The language also may be used to generate a stochastic simulation.

Each of these has tool support so that the underlying model is derived automatically according to the predefined rules.

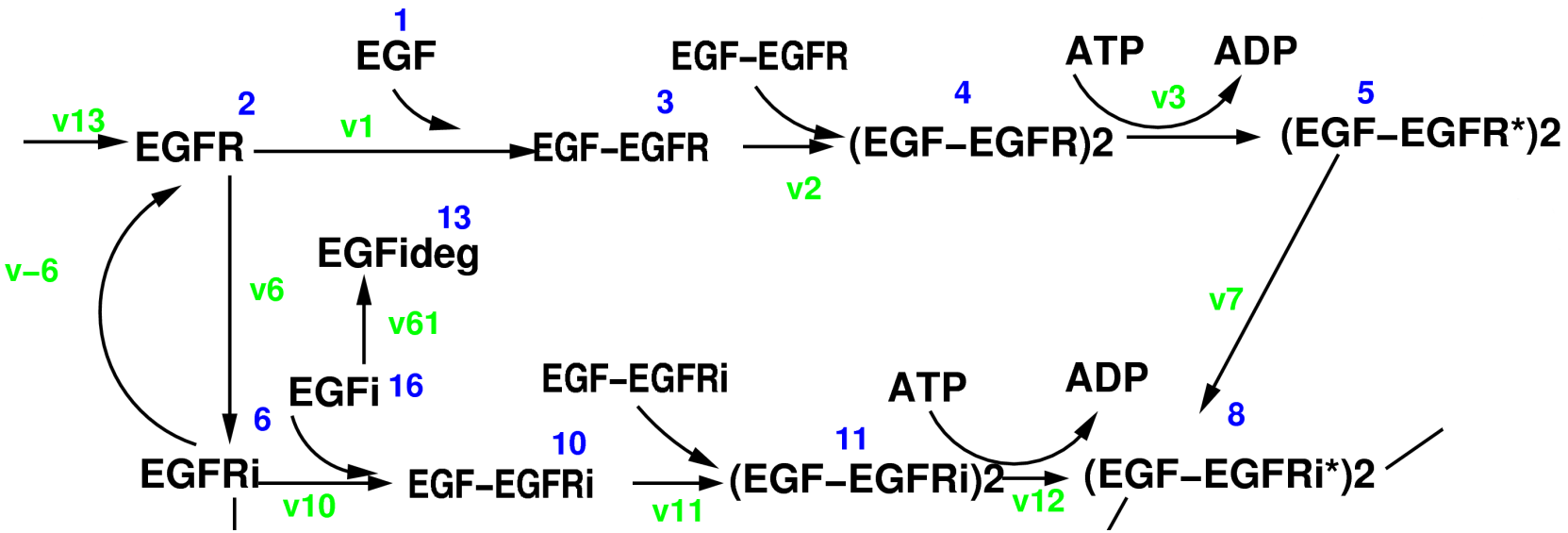
Case Study: Schoeberl *et al.*'s model of the MAPK Cascade [CDGH06]

- Published in *Nature Biotechnology* 20:370-375 in 2002.
- Influential, cited by more than 150 subsequent published papers.
- Consists of 94 reagent species involved in 125 reactions.
- Substantial ODE model consisting of 94 state variables and 95 parameters.
- Original model constructed "by hand", with help of a graphical representation.
- Original analysis based on numerical integration platform of the Matlab numerical computing platform.

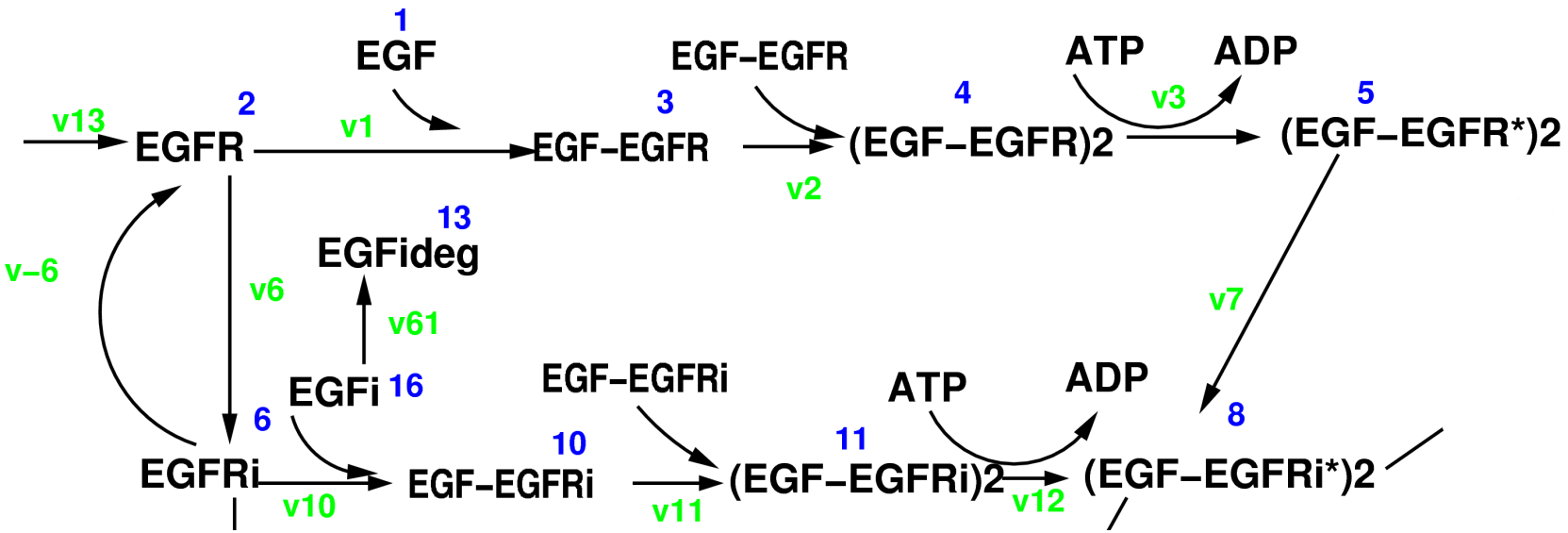
The MAP Kinase Cascade



The MAP Kinase Cascade

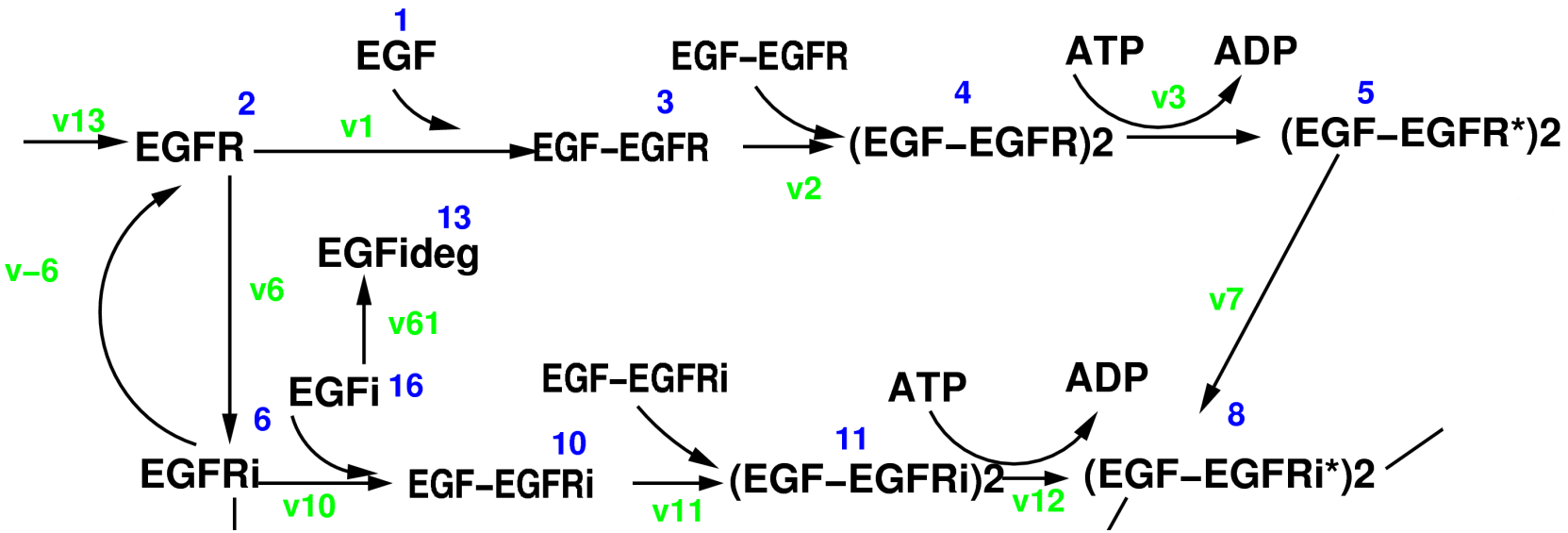


The MAP Kinase Cascade



There are many ambiguities in the graphical representation, e.g.

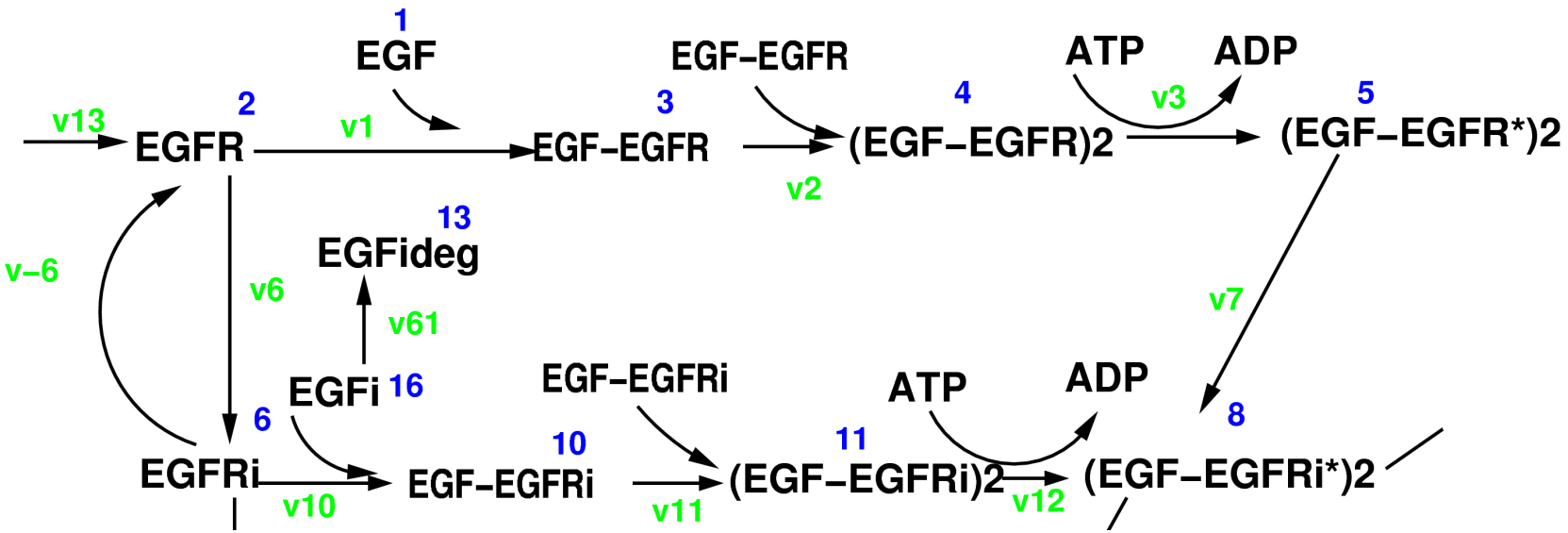
The MAP Kinase Cascade



There are many ambiguities in the graphical representation, e.g.

- An infinite supply of EGF is assumed;

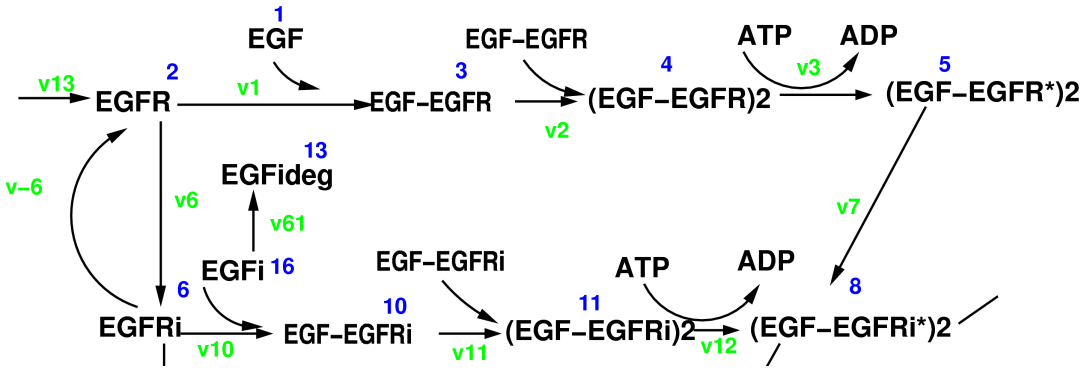
The MAP Kinase Cascade



There are many ambiguities in the graphical representation, e.g.

- An infinite supply of EGF is assumed;
- Reaction *v7* is uni-directional whereas all others are reversible.

Extracts from the model of the MAP Kinase Cascade



$$\mathrm{EGF_H} \stackrel{def}{=} (v_1, k_1).\mathrm{EGF_H}$$

$$\mathrm{EGFR_H} \stackrel{def}{=} (v_1, k_1).\mathrm{EGFR_L} + (v_6, k_6).\mathrm{EGFR_L}$$

$$\mathrm{EGFR_L} \stackrel{def}{=} (v_{\text{-}1},\ k_{\text{-}1}).\mathrm{EGFR_H} + (v_{\text{-}6},\ k_{\text{-}6}).\mathrm{EGFR_H} + (v_{13}, k_{13}).\mathrm{EGFR_H}$$

$$\text{EGF-EGFR}_\mathrm{H} \stackrel{def}{=} (v_2, k_2).\text{EGF-EGFR}_\mathrm{L} + (v_{\text{-}1}, k_{\text{-}1}).\text{EGF-EGFR}_\mathrm{L}$$

$$\text{EGF-EGFR}_\mathrm{L} \stackrel{def}{=} (v_1, k_1).\text{EGF-EGFR}_\mathrm{H} + (v_{\text{-}2}, k_{\text{-}2}).\text{EGF-EGFR}_\mathrm{H}$$

The PEPA model

Similar PEPA definitions were constructed for each of the 94 species in the pathway.

The PEPA model

Similar PEPA definitions were constructed for each of the 94 species in the pathway.

This was tedious, but not difficult, although care was needed to handle the points of ambiguity in the graphical representation.

The PEPA model

Similar PEPA definitions were constructed for each of the 94 species in the pathway.

This was tedious, but not difficult, although care was needed to handle the points of ambiguity in the graphical representation.

In order to complete the model we also needed to capture the interactions (i.e. cooperations) between the reagents. In this case we assumed that whenever reagents participated in reactions with the same name they did so in cooperation.

Validation of the PEPA model

- Once the PEPA model was constructed, we wanted to ensure that it was generating the same mathematical representation of the system.

Validation of the PEPA model

- Once the PEPA model was constructed, we wanted to ensure that it was generating the same mathematical representation of the system.
- In the first instance we derived a set of ODEs in a format suitable for Matlab.

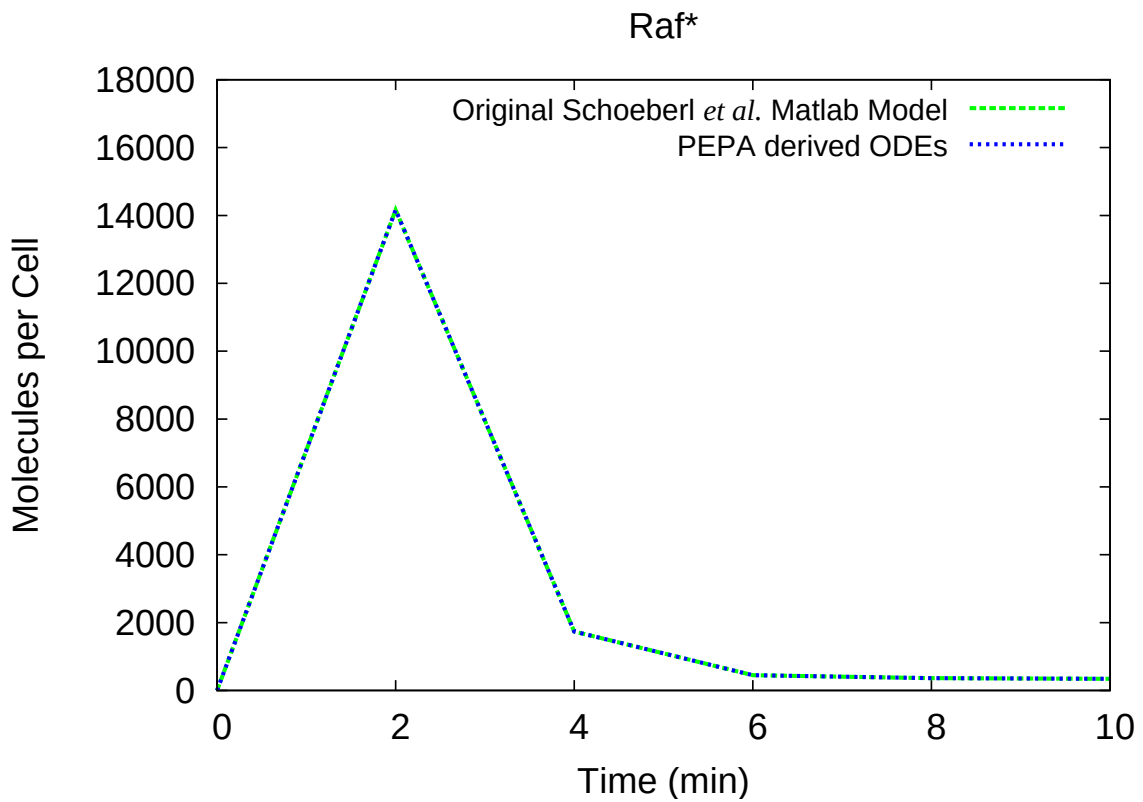
Validation of the PEPA model

- Once the PEPA model was constructed, we wanted to ensure that it was generating the same mathematical representation of the system.
- In the first instance we derived a set of ODEs in a format suitable for Matlab.
- These could not be compared directly with Schoeberl *et al*'s ODEs due to different representations being used, but we compared them empirically in terms of the results.

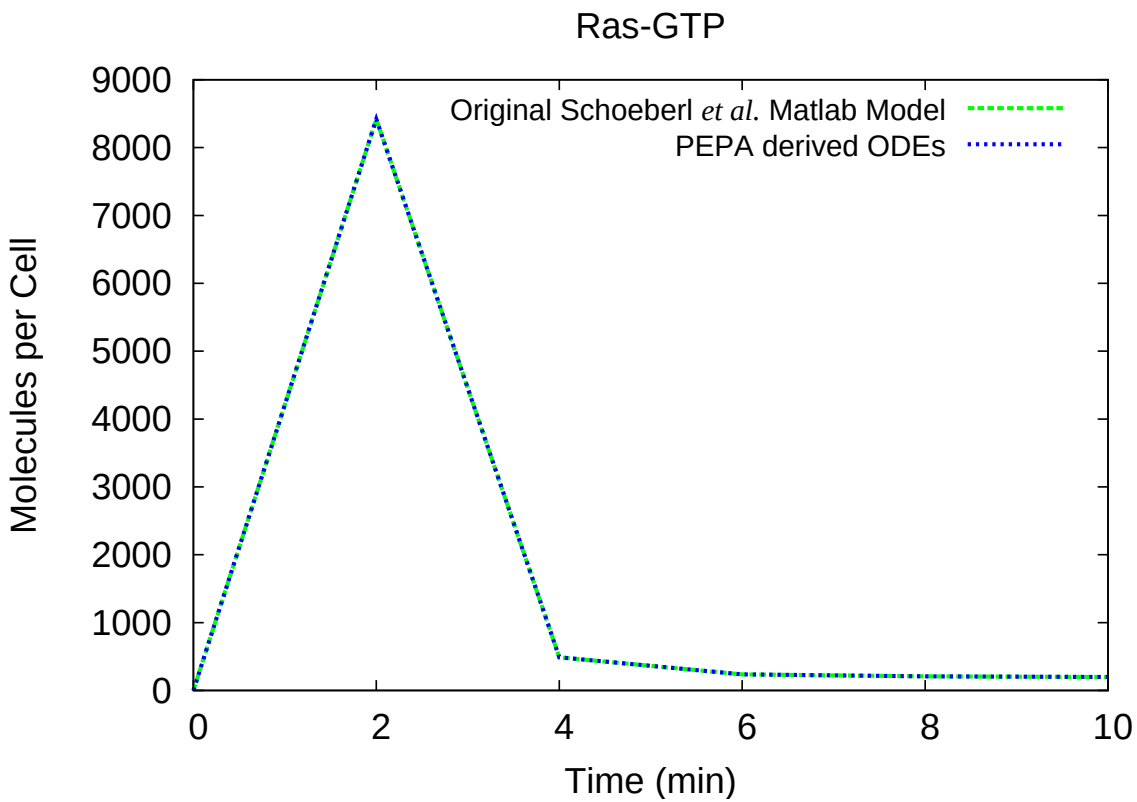
Validation of the PEPA model

- Once the PEPA model was constructed, we wanted to ensure that it was generating the same mathematical representation of the system.
- In the first instance we derived a set of ODEs in a format suitable for Matlab.
- These could not be compared directly with Schoeberl *et al*'s ODEs due to different representations being used, but we compared them empirically in terms of the results.
- Then we used an alternative mapping from the PEPA to generate a stochastic simulation of the system.

Comparing Original Results and PEPA Derived ODEs

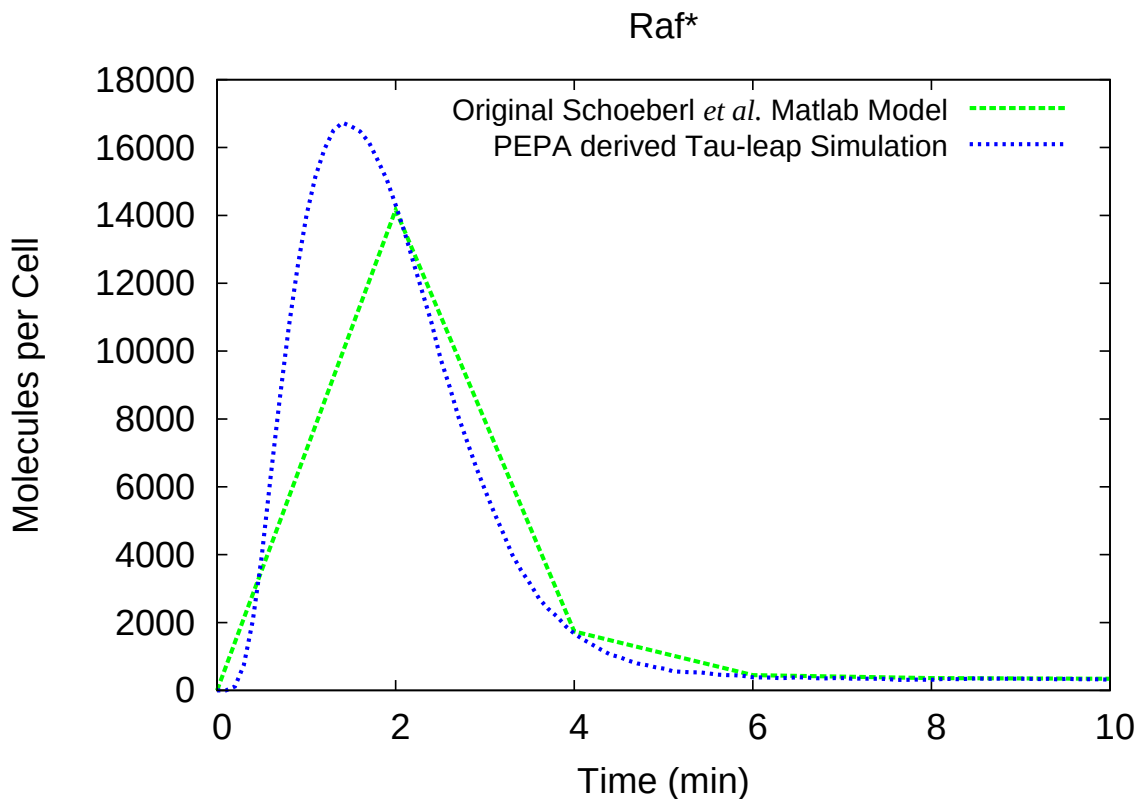


Comparing Original Results and PEPA Derived ODEs

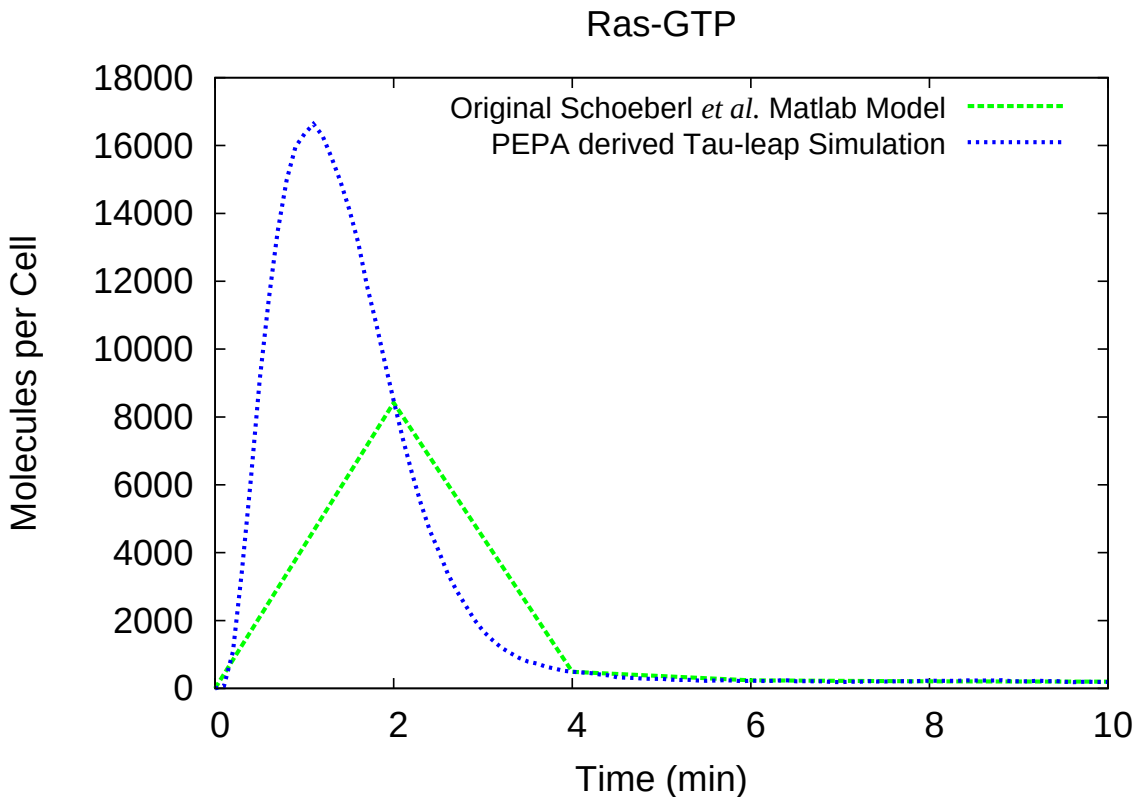


The PEPA derived ODEs return the same results as the Schoeberl *et al.* Matlab model.

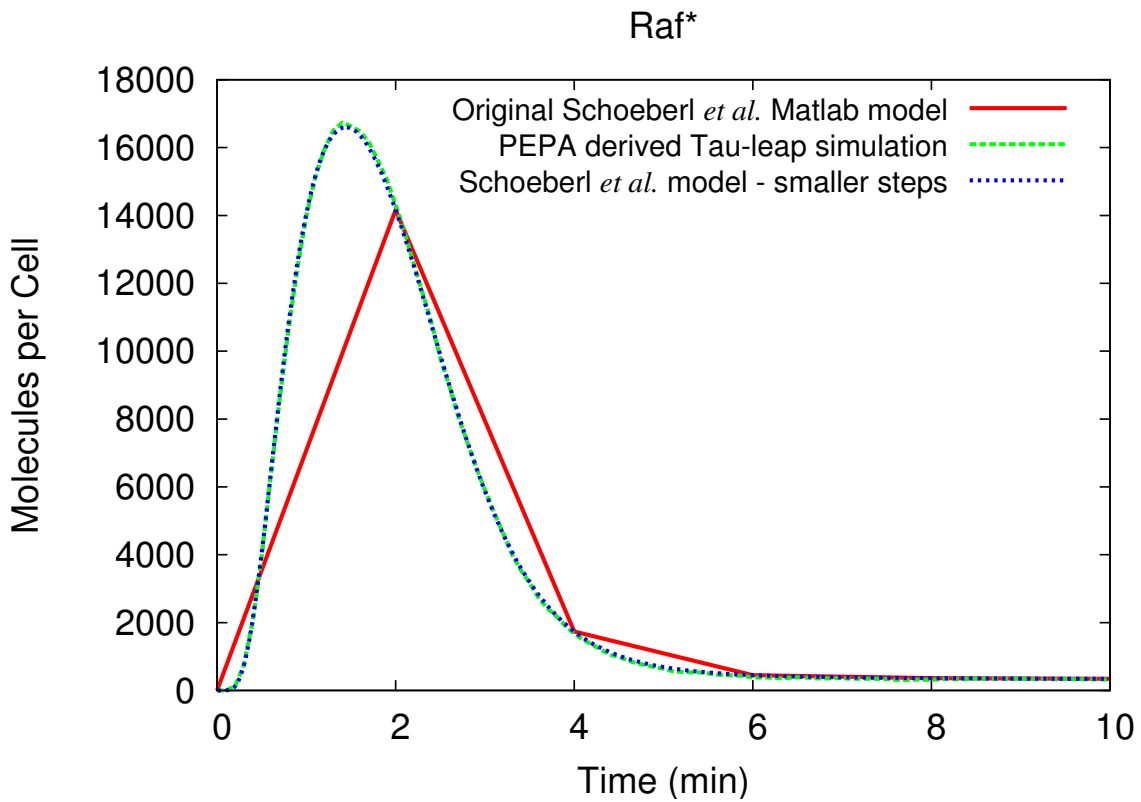
Comparing Original Results and PEPA Derived Stochastic Simulation



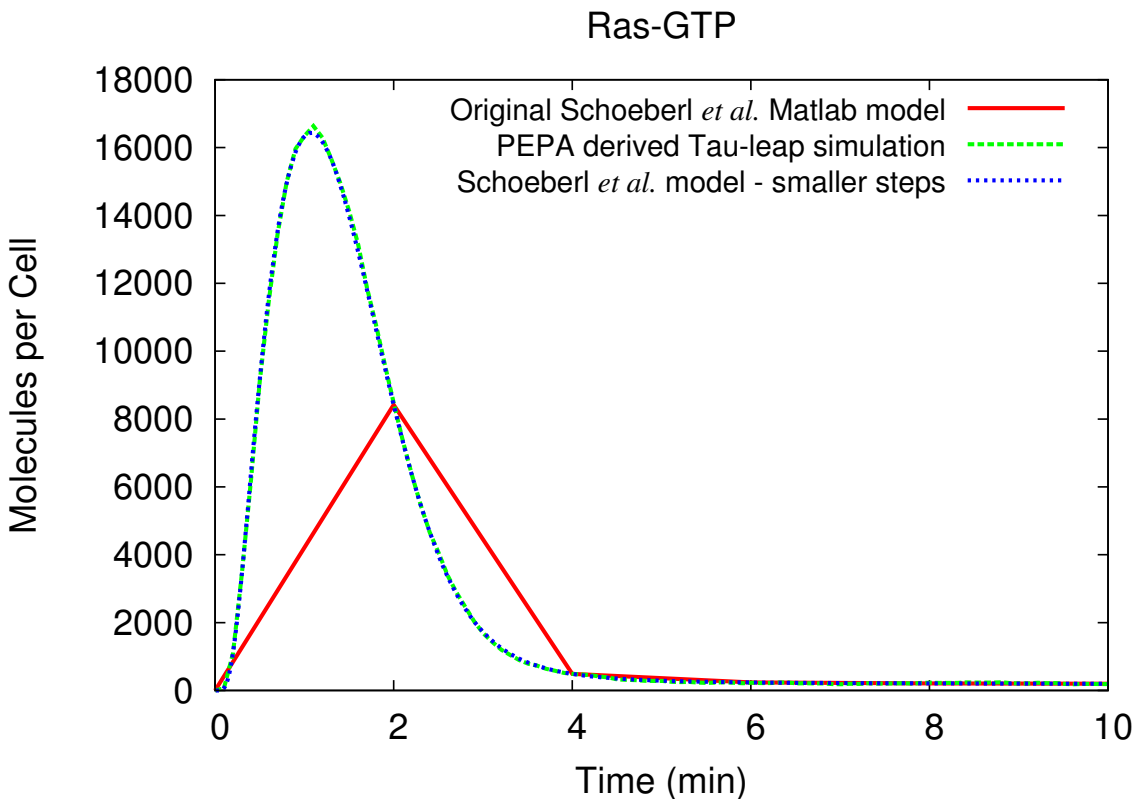
Comparing Original Results and PEPA Derived Stochastic Simulation



Corrected Time Step in Matlab Model

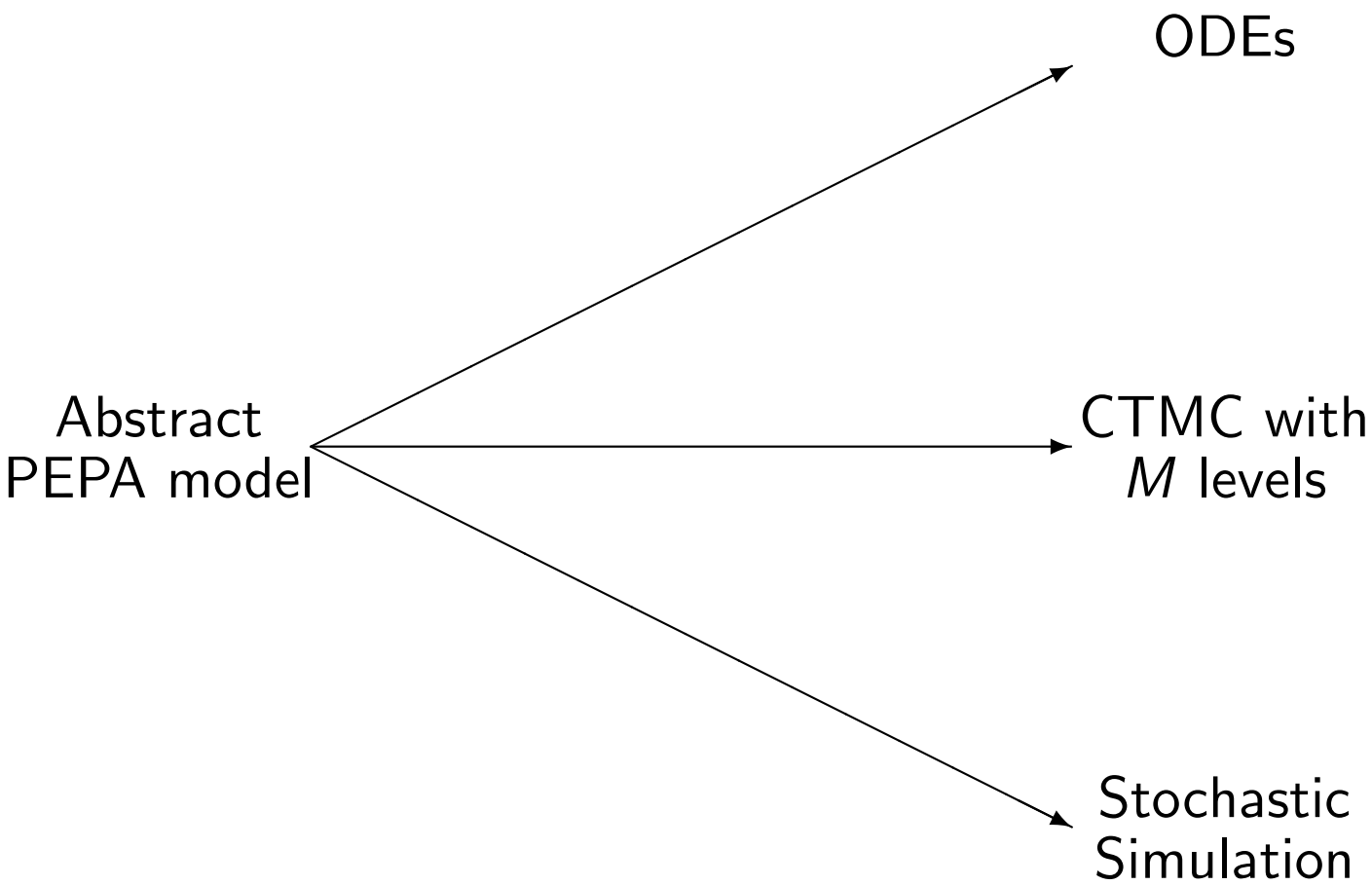


Corrected Time Step in Matlab Model

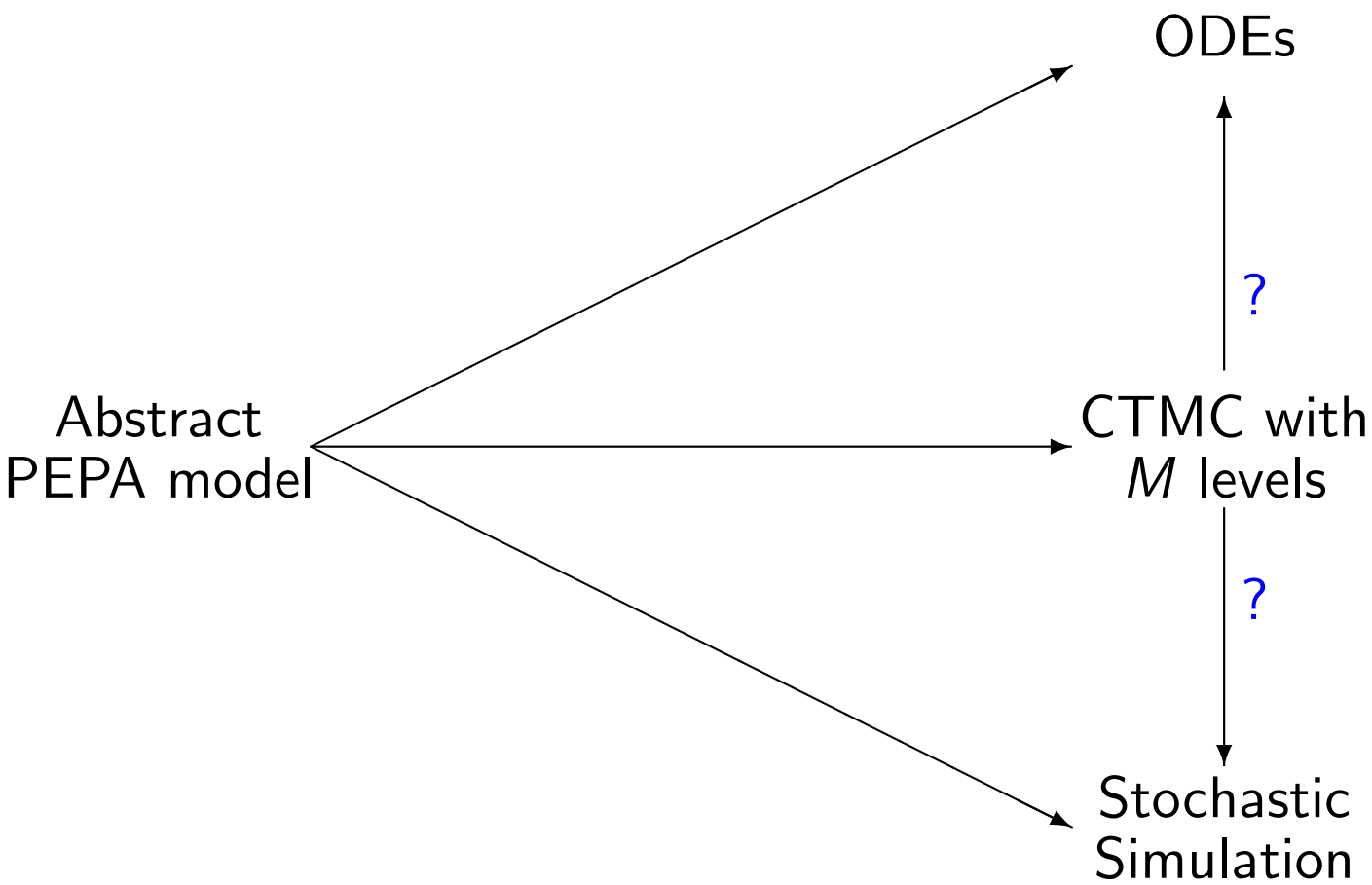


The original parameters for the Matlab model stepped over the true peak.

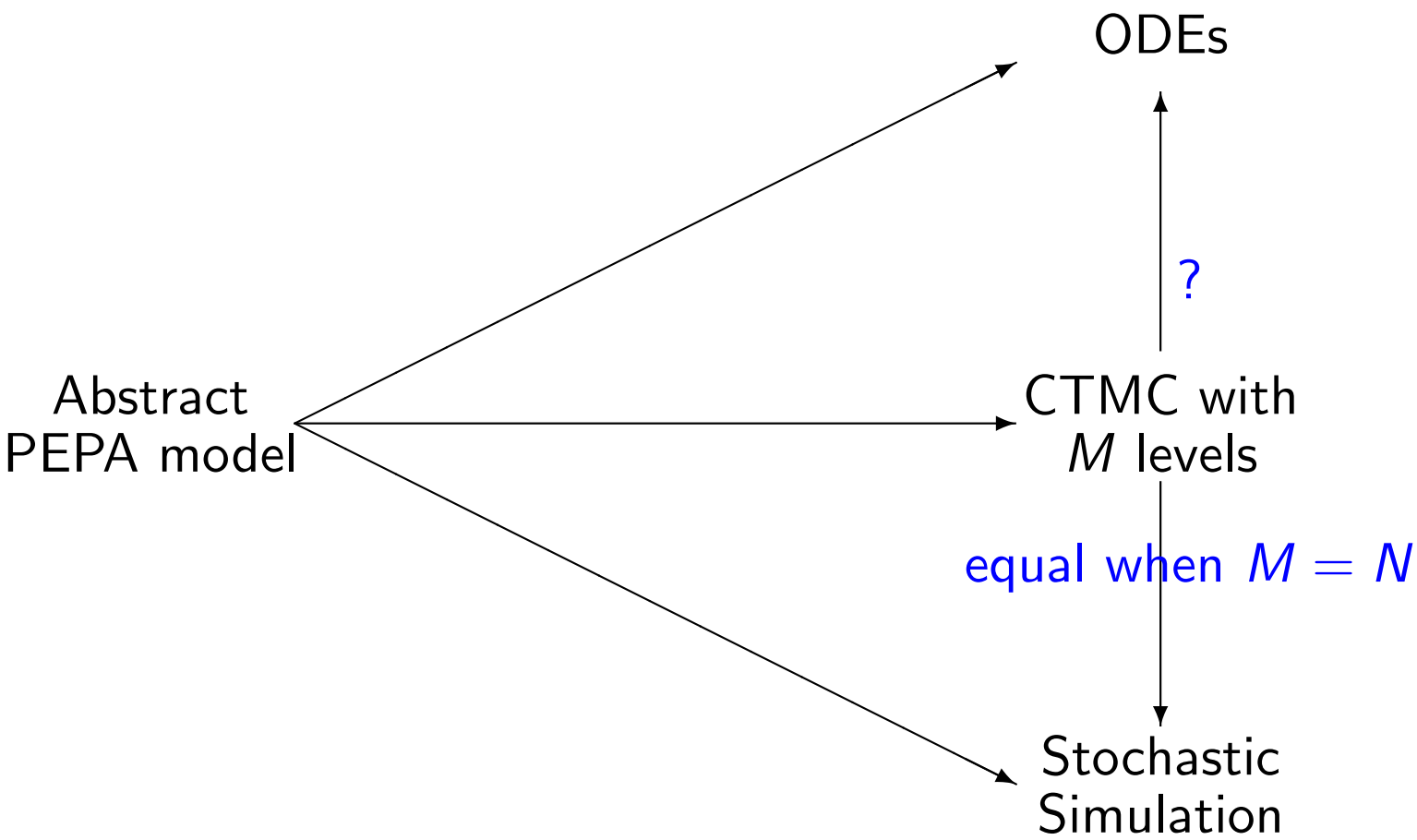
Alternative Representations



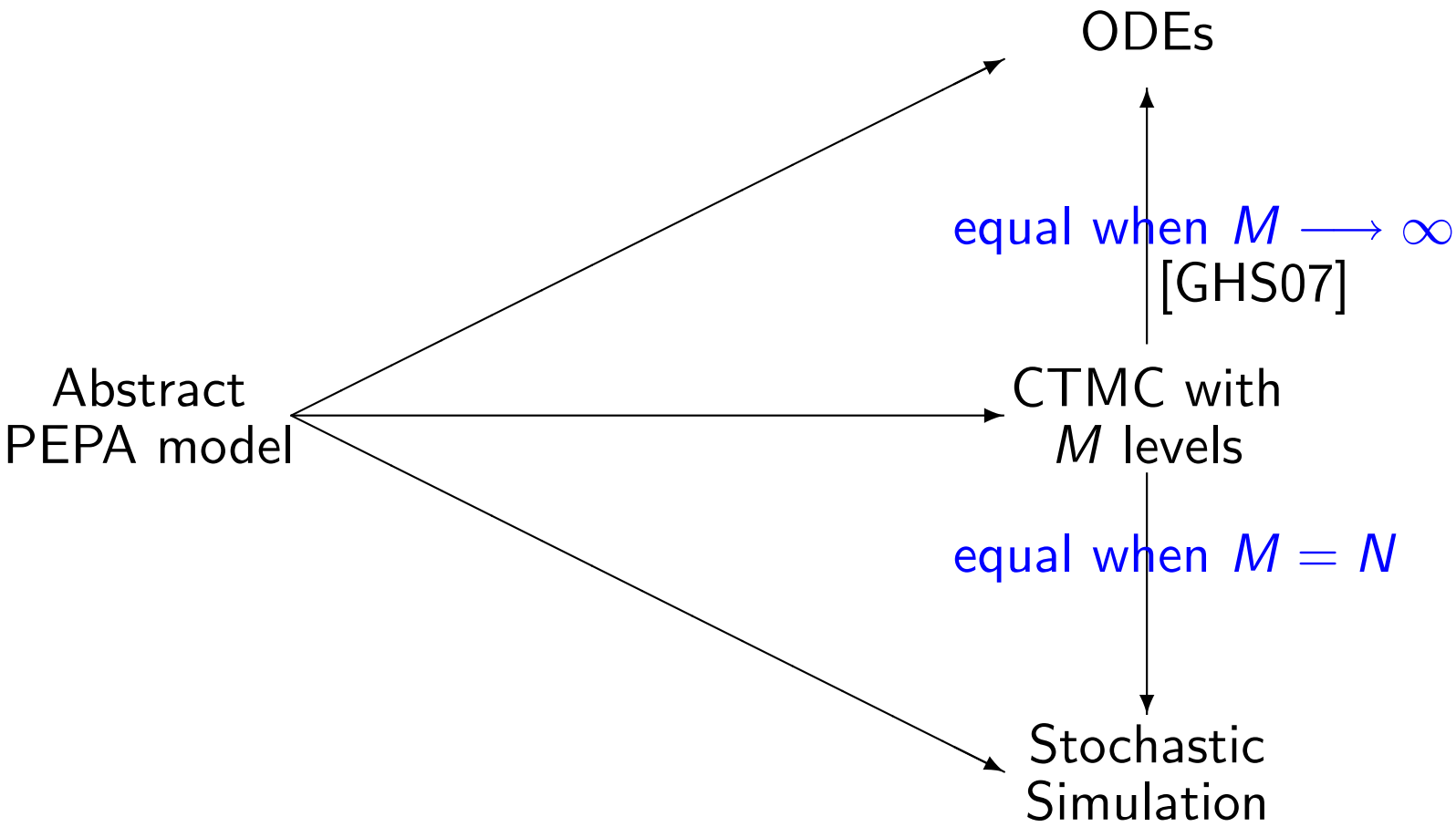
Alternative Representations



Alternative Representations



Alternative Representations



Outline

Challenges

- Abstract modelling offers a compromise between the individual-based and population-based views of systems which biologists commonly take.

Challenges

- Abstract modelling offers a compromise between the individual-based and population-based views of systems which biologists commonly take.
- Moveover we can undertake additional analysis based on the discretised population view.

Challenges

- Abstract modelling offers a compromise between the individual-based and population-based views of systems which biologists commonly take.
- Moveover we can undertake additional analysis based on the discretised population view.
- Further work is needed to establish a better relationship between this view and the population view — empirical evidence has shown that 6 or 7 levels are often sufficient to capture exactly the same behaviour as the ODE model.

Challenges

- Abstract modelling offers a compromise between the individual-based and population-based views of systems which biologists commonly take.
- Moveover we can undertake additional analysis based on the discretised population view.
- Further work is needed to establish a better relationship between this view and the population view — empirical evidence has shown that 6 or 7 levels are often sufficient to capture exactly the same behaviour as the ODE model.
- In the future we hope to investigate the extent to which the process algebra compositional structure can be exploited during model analysis.

Challenges cont.

- The issue of unknown and uncertain data remains to be addressed.

Challenges cont.

- The issue of unknown and uncertain data remains to be addressed.
- The abstract Markovian models allow quantities of interest such as "response times" to be expressed as probability distributions rather than single estimates. This may allow better reflection of wet lab data which showns variability.

Challenges cont.

- The issue of unknown and uncertain data remains to be addressed.
- The abstract Markovian models allow quantities of interest such as "response times" to be expressed as probability distributions rather than single estimates. This may allow better reflection of wet lab data which showns variability.
- Promising recent work by Girolami *et al.* on assessing candidates models which attempt to cover both unknown structure and unknown kinetic rates with respect to experimental data, using Bayesian reasoning.

Conclusions

- Ultimately we want to understand the functioning of cells as useful levels of abstraction, and to predict unknown behaviour.

Conclusions

- Ultimately we want to understand the functioning of cells as useful levels of abstraction, and to predict unknown behaviour.
- It remains an open and challenging problem to define a set of basic and general primitives for modelling biological systems, inspired by biological processes.

Conclusions

- Ultimately we want to understand the functioning of cells as useful levels of abstraction, and to predict unknown behaviour.
- It remains an open and challenging problem to define a set of basic and general primitives for modelling biological systems, inspired by biological processes.
- Achieving this goal is anticipated to have two broad benefits:

Conclusions

- Ultimately we want to understand the functioning of cells as useful levels of abstraction, and to predict unknown behaviour.
- It remains an open and challenging problem to define a set of basic and general primitives for modelling biological systems, inspired by biological processes.
- Achieving this goal is anticipated to have two broad benefits:
  - Better models and simulations of living phenomena

Conclusions

- Ultimately we want to understand the functioning of cells as useful levels of abstraction, and to predict unknown behaviour.
- It remains an open and challenging problem to define a set of basic and general primitives for modelling biological systems, inspired by biological processes.
- Achieving this goal is anticipated to have two broad benefits:
  - Better models and simulations of living phenomena
  - New models of computations that are biologically inspired.

Thank You!

Thank You!

Collaborators: Muffy Calder, Federica Ciocchetta, Adam Duguid, Nil Geisweiller, Stephen Gilmore and Marco Stenico.

Acknowledgements: Engineering and Physical Sciences Research Council (EPSRC) and Biotechnology and Biological Sciences Research Council (BBSRC)