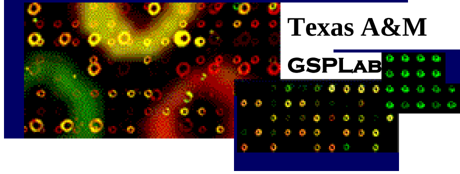


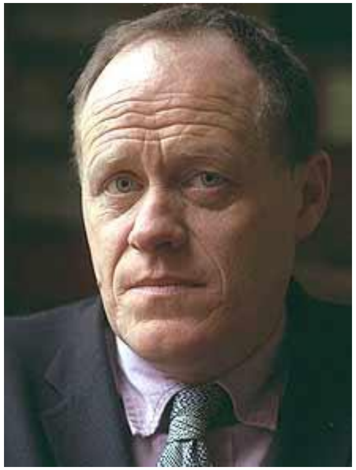
Limitations of Biological Knowledge

Ivan Ivanov
Department of Veterinary Physiology and Pharmacology,
Genomic Signal Processing Lab,
Intercollegiate Faculty of Toxicology,
Institute for Applied Mathematics and Computational Science
Texas A&M University

With Permission From My Mentor and Friend: Edward R. Dougherty
Department of Electrical and Computer Engineering
Center for Bioinformatics and Genomic Systems Engineering
Texas A&M University

Two Views: Limited Prospects for Medicine

- **James Le Fanu** (*The Rise and Fall of Modern Medicine*): “Medicine, like any field of endeavor, is bounded by its concerns – the treatment of disease – so success necessarily places a limit on further progress…. As of the moment, it is not clear whether or how the last challenge left – the discovery of the causes of disease like multiple sclerosis and leukemia – is indeed ‘soluble.’… The limited prospects of future medical advance should by now be well recognized.”



Two Views: A New Translational Medicine

- **Siddhartha Mukherjee** (*The Emperor of Maladies*): "Gene by gene, and now pathway by pathway, we have an extraordinary glimpse into the biology of cancer. The complete maps of mutations in many tumor types… will soon be complete, and the core pathways that are mutated fully defined... Once the mutations have been identified, the mutant genes will need to be assigned functions in cellular physiology. We will need to move through a renewed cycle of knowledge that recapitulates a past cycle – from anatomy to physiology to therapeutics."

Far to Early to Judge

- As biological science turns to its natural home in control, communication, and information, it will have open to it a vast store of systems theory gained over the last 75 years.
- To judge it a failure before it has hardly begun would be akin to arguing that humans cannot walk on the moon before the arrival of Newton.

Understanding Limitations

- There are limitations. Understanding these can serve as a guide to the kinds of problems that confront us and as a prescription for the kinds of auxiliary advances that need to occur to mitigate those limitations.
  - Limitations need not be permanent.
- Limitations arise from epistemological requirements:
  - (1) knowledge represented by a mathematical model,
  - (2) operational definitions to tie the model to experiments,
  - (3) agreement (in some defined sense) of model-based predictions and experimental outcomes.

Mathematical (Computational) Complexity

- One may write down a large number of equations to finely describe behavior, but the ability to derive solutions deteriorates with increasing model complexity.
- Owing to the vast number of quantitative variables and their involved relationships, a tractable system cannot incorporate more than a small portion of them.
- Model reduction is necessary
  - Latent variables make the system stochastic
- Fundamental trade-off: model tractability versus phenomenal predictability.

Complexity and Model Representation

- Methods need to be developed to reduce models while preserving important information for the task at hand, such as therapeutic intervention.
- This requires appropriate (canonical) model representation so as to separate out unneeded structure, a classic problem in signal processing.
- It also requires characterization of approximation accuracy for the reduction, where approximation is related to the goal of the modeling and the reduction.
- These are difficult problems requiring biological insight and a strong background in stochastic processes.

Complexity and the Translational Mission

- A scientist might be satisfied with the original network and judge it superior to the reduced one because it provides better prediction, but the translational scientist (engineer or physician) requires the reduced network in order to accomplish the translational mission, albeit, perhaps with decreased performance than would be the case with the original network.
- A trade-off must be made or else one is paralyzed.

Experimentation Bedeviled by Complexity

- Experiments need to ask precise questions.
  - How does one construct a precise experiment amid complexity?
- Experiments confront enormous regulatory complexity.
  - Transcription factors function jointly, binding in complexes with other proteins or via combinatorial binding to promoter regions.
  - Feedback mechanisms modulate dynamics of gene expression.
  - microRNAs produce transcript degradation, forming auxiliary networks interacting with transcription-factor networks.
  - Proteins modify chromatin via methylation, phosphorylation, or acetylation, altering accessibility of DNA – epigenetic regulation.
  - Such post-translational form dynamical protein-protein interaction networks

Complexity and the Translational Mission

- Despite improved ability to measure the abundances, states, and interactions of biomolecules on a global scale, current experimental capabilities are still inadequate for constructing all but the simplest and reduced models of biomolecular networks in a cell.
- There are several fundamental experimental limitations.

Fundamental Experimental Limitations

- Sensitivity.
  - Most biological measurements are performed on populations of cells, including mRNAs, microRNAs, and proteins.
  - Population level measurements can only yield average behaviors, which can be highly misleading.
  - Single cell dynamics can be measured using fluorescent-based imaging, only a few genes/proteins can be measured this way.
- Transcriptional regulation is over multiple timescales.
  - Since processes are coupled, they need to be measured jointly.
- Multi-scale characteristics in the spatial domain.
  - Intracellular molecular dynamics are nonhomogeneous, with most molecular species localized to subcellular compartments.

Validity in a Stochastic World

- In a deterministic environment, one makes a prediction from the model, conducts an experiment in which an observable corresponds to the prediction and checks for agreement.
  - Owing to experimental variability, one allows for some disagreement between the predicted and observed values. Assuming perfect accuracy, the conclusion would either be agreement or disagreement.
  - If disagreement, we reject the model – the theory is falsified.
  - If agreement, the model is accepted insofar as the particular prediction (characteristic) is concerned, but remains open to rejection. The model is contingently validated.

Validity in a Stochastic World

- In a deterministic environment, one makes a prediction from the model, conducts an experiment in which an observable corresponds to the prediction and checks for agreement.
  - Owing to experimental variability, one allows for some disagreement between the predicted and observed values. Assuming perfect accuracy, the conclusion would either be agreement or disagreement.
  - If disagreement, we reject the model – the theory is falsified.
  - If agreement, the model is accepted insofar as the particular prediction (characteristic) is concerned, but remains open to rejection. The model is contingently validated.

Scientific Model Validity

- $\mu(\mathcal{M}, \mathcal{H})$ compares observable variables from $\mathcal{M}$ to variables from $\mathcal{H}$, an hypostasized "real-world" model that we never know.
- Values of variables from $\mathcal{M}$ are compared via μ to observed values of $\mathcal{H}$ to form an estimate of $\mu(\mathcal{M}, \mathcal{H})$.
  - For instance, the steady-state distribution of $\mathcal{M}$ is compared to measurements coming from the steady-state of $\mathcal{H}$.

Prior Knowledge

- A scientist must come to the table with sufficient knowledge of the problem to formulate a small class of models for which it remains only necessary to utilize data to estimate some set of parameters to instantiate the model.
- Data in the absence of knowledge leaves one with a virtually unbounded model space in which to configure mathematical theories.
- Prior knowledge is inextricably tied to experimental design – the greater the prior knowledge the better experiment one can design.

Texas A&M GSPLAB

Forms of Prior Knowledge

- Biological organizational principles: general principles that constrain and focus the conceptualization on feasible systems and constrain the model space for inference.
  - Connectivity constraints (e.g. distribution laws), dynamics (e.g. criticality), functionality requirements (e.g. robustness to environmental perturbations), energy efficiency, etc.
- Existing relations among the variables that constrain the model by requiring it to be consistent with these prior relations.
  - Regulatory mechanisms among transcription factors and their target genes, protein-protein interactions, and targets of microRNAs..

Intellectual Capacity

- In addition to the limitations imposed by mathematical theory, computation, statistics, and experimentation, we face fundamental limitations in terms of human intellectual capacity.

Upper Bounds

- A hard upper bound exists – consider an analogy.
  – Throw a ball to Maggie. She assesses the trajectory, runs to the area where she expects the ball to bounce, and adjusts to catch the ball in her mouth before it hits the ground.
  – She learns all of this with a very small training set.
  – But she does not formalize her learning into theory.
  – She is limited by her hardwiring and has no idea of what lies beyond it.

Human Limitations – Education

- Our limits lay farther out. We can understand the requisite differential equations but, like Maggie, we have no idea of what lies beyond our hardwiring. If there are categories of understanding beyond our hardwiring that are useful for scientific modeling, we are both incapable of using them and unaware of what it is that we cannot do.
- Besides the hard-wired limits, there are conceptualization limits and these can only be overcome by education. Absent a solid education, one cannot go beyond the superficial, especially in a subject such as biology (and medicine) that requires mathematical depth and breadth in difficult areas like stochastic processes.