Theoretical Biology: Not Even Wrong?

David C. Krakauer

August 3, 2007

1. Arguably the most successful disciplines in the natural sciences making use of formal languages to explain natural phenomena have been the physical sciences. In the physical sciences, theoretical physics and mathematical physics in particular, research involves the application of mathematics to data sets displaying deep regularities in pursuit of minimal, mathematical expressions describing important relationships among quantifiable observables. These are often referred to as physical laws. The great conceptual revolutions in the physical sciences have tyopically been associated with formal theory: Keplers laws, Newtons classical mechanics, Gallileos theories of motion, Einsteins field equations, quantum mechanics, and more recently quantum field theory and string theory.

2. The success and application of mathematics in physics has not been without its detractors. Recently string theory has come under scrutiny for failing to make predictions that can be tested using current experimental technologies, and for its development of a mathematics of such sophistication that it can be understood by only a small community of predominantly mathematicians. String theory is also faulted for generating a bewildering number of equally plausible but very different solutions from a common set of assumptions and equations.

3. Similar criticism have been made about areas of theoretical biology, albeit that the theories in these cases have been far more modest and local in their application than those in physics. Theory in biology often takes the form of making explicit a counter-intuitive insight, where mathematics is in the service of establishing the credibility of this insight. Frequently, little attempt is made to test the unique predictions of the theory, and often the theory provides no unique predictions. In these cases the purpose of the theory - it is argued - is the improved organization of a nominally unconnected body of evidence and research under a common set of principles. One often has the sense that one is being told what one already knows but in a more rigorous language. One might say that the theory is 'not even wrong'.

4. In the biological sciences some of the more significant conceptual increments have been expressed through a combination of empirical data and natural language narrative. Perhaps the most far-reaching discovery is the Darwinian theory of evolution through the process of natural selection. This theory that was not a formal theory when first devised, but became partially formalized through the application of the calculus to the dynamics of changing gene frequencies, in the process, reconciling selection with particulate inheritance. However, while the mathematics provided a means of quantifying the strength of selection acting on simple genes, the exposition of the core theory did not yield to mathematics, and arguably, was oversimplified in the early stages of formalization. It remains true even today that the most com-

mon way of explaining natural selection is through natural language and schematic images precisely because the theory is silent on many phenomena of great interest.

5. One prevailing trend in theoretical biology has been toward the development of toy models of very little general interest and almost no possibility of refutation. These models exist by virtue of an intellectual saprotrophic strategy, scavenging the carcasses of similar if not identical models, and generating rather modest value to the research ecosystem.

6. I can not help but note a trend in which biology becomes increasingly formalized as the collection and analysis of large, well curated data-sets grows, but with a concomitant diversion of resources from an emphasis on inclusiveness to reduction. Systems biology is one of the greatest proponents of the engineering approach to theory which somehow has managed to confuse insight with quantification. Moreover by remaining firmly rooted in genetics and molecular biology, the link to the phenotype and function or behavior is almost invariably neglected - at some insight cost one suspects.

7. A partial list of the research areas where I think these ideas and concerns might be discussed at the meeting include:

   (a) Considering the target of evolutionary change and Darwins theory and its relation to ideas of emergence etc.

   (b) The role of our understanding of the excitable membrane and dynamical systems in neuroscience and the historical connection to connectionism and cognitive science. (case study in reductionism)

   (c) The problem of cooperation and the (non)utility of game theory as what I call 'worst case modeling'.

   (d) Community structure and Theoretical Ecology - issues of functional course graining.

   (e) Morphogenesis and Pattern formation and its vitiation by crude genetic regulation ;)

   (f) Methods of dimension reduction for gene expression and regulatory networks.

   (g) The quantification and analysis of social organization using social network properties. (methods driven research)

   (h) Agent based modeling as the application of, let's call it, a 'science objective function' where the time required to generate results is traded-off against the number of parameters and analytical tractability of the models.

   (i) This touches on the thorny issue of insights we might be able to provide into highly contingent or correlated systems where the effective dimensionality of a problem remains fairly large.