Steven Frank's Vision Statement

I study the evolutionary processes that shape organismal design. My projects emphasize the synergism between concepts and empirical analysis.

Microbes and the forces of design

A recent review of bacterial metabolism concluded

> In the past, changes in gene expression and metabolic strategies across growth conditions have often been attributed to the optimization of steady-state growth rates. However, mounting evidence suggests that cells are capable of significantly faster growth rates in many conditions, including supposedly 'poor' carbon sources. Based on these observations, it is clear that objectives other than optimization of steady-state growth rates must be considered to explain these phenotypes. (Basan 2018)

The great progress in microbial studies continually brings us up against this mismatch. We have increasingly rich data on key microbial traits. The data reveal the many weakness in our tools for understanding the evolutionary design of those traits.

Evolutionary biologists have of course been thinking about adaptation and design since Darwin. We have very many theoretical predictions and tools for inference. But the reality is not so good when we look over the microbiologists' shoulders at what they are actually seeing.

Why, in fact, don't microbial cells always grow as fast as possible? Yes, there are tradeoffs. But the likely candidates form a long list.

There are no clear guidelines for turning the multiple vague theories into directly workable tools. We need three steps. First, what explicitly are the fundamental forces of design? Second, how do those fundamental forces translate into testable predictions that one can actually study in a simple and direct way? Third, what tools do we need to connect observations to inferences of causation?

My preliminary manuscript sets out a complete vision for how to approach each of these three steps. I have made progress on some aspects. More importantly, I know where we need to go.

The common patterns of nature

I study cancer for many intellectual and applied reasons. But, in my personal vision, the understanding of design is my primary concern. Cancer arises from the failures of the fundamental protective and error-correcting designs of our tissues, cells, and molecular regulatory processes. It is through the study of failure that we understand biological design. Mutational screens reveal genetic and molecular design, brain lesions reveal neural design, and so on.

In my book, *Dynamics of Cancer*, I analyzed how the breakdown of the individual protective designs generates particular patterns of failure. The age of cancer onset defines the classic aggregate pattern of failure. I explored how age-onset curves could be used to test hypotheses about alternative underlying molecular and physiological generative mechanisms.

That is an inverse problem. It is hard. But there is no choice. The generative mechanisms alter the kinetics by which cancer arises. Those kinetics leave their trace only through the aggregate pattern of the age of onset.

The point here concerns the general relations between generative process and aggregate pattern. In particular, I showed that failure processes have an invariant Gompertzian geometry. That generic invariant geometry determines the broad pattern of cancer onset. Particular biological mechanisms modulate that generic pattern. To understand biological pattern, one must always study the duality of the generic and the particular.

That duality between the generic and the particular arises in nearly all studies of pattern and process. We typically focus on the particular biological processes in our studies, as I initially did in my cancer work. I realized that to understand the commonly observed patterns of nature, I had to learn a lot more about the generic side.

That challenge set me on a ten-year study of the common patterns of nature and their underlying structure. I believe that I now have the most comprehensive understanding of the common probability patterns and their relations to each other, through my analysis of measurement and invariance.

To call that an arguable claim is an understatement. I have rarely encountered anything but argument. But there can hardly be a more important argument to have. In my view, I possess the broadest, most coherent and most defensible argument. Maybe not the ultimate truth, but better than any going alternative.

You will have to decide for yourself. Have a look at my first major attempt at applying these ideas to a key problem in biology in my recent manuscript Invariance in ecological pattern.

With that new manuscript, I am now moving on to the essential task of showing how my vision helps to solve important biological puzzles. In this era of big data, almost every discipline presents an important challenge.

Returning to my theme of design, the invariant geometry of pattern must play a central role in evolution. Biological traits arise by the aggregation of multiple underlying processes. That aggregation sets strong generic tendencies for pattern.

Design by natural selection modulates those generic tendencies with particular tunings for component processes. The simplest designs will follow the natural generic contours. Less often, natural selection will favor designs that work against the generic contours of pattern.

To understand design, one must study the interplay between the generic and the particular. I am not yet certain how to study that duality. But it must be important.

I will need excellent collaborators to make progress in genetics and also in other fields. Each problem will differ and will require new theory and new tools of data analysis.

Evolutionary design of regulatory control

Error-correcting feedback is the single greatest principle of systems design. Error measures the difference between a system's actual output and its target. By feeding back the error as an input, the system can move in the direction that reduces the error.

Error correction compensates robustly for misinformation about system dynamics and for perturbations to system components. Excellent performance often follows in spite of limited information and noisy signals.

A robust error-correcting feedback system compensates for sloppy, error-prone components. That robust compensation weakens the pressure of natural selection on the components, leading to what I have called the paradox of robustness. The evolution of each additional error-correcting feedback loop at the system level will tend to associate with the evolution of cheaper, lower performing system components. Those protected components may also tend to accumulate greater genetic variability and stochasticity of expression.

The layering of higher-level robustness mechanisms and the decay of underlying components seem like they must be constant processes throughout the evolutionary history of design. Yet, it has not been easy to translate this profound theory into a program with broad empirical insight.

To build out the theoretical framework, I am developing a series of articles:

- Design tradeoffs and control theory: combining evolutionary analysis with engineering control theory provides the essential methods.
- Genetic variability and stochasticity of trait expression: the paradox of robustness increases component variability and the heritability of disease.
- Decay of costly components: the paradox of robustness favors substitution of cheaper, lower performing components within systems.
- Learning as a robustness mechanism: systems that acquire information and adjust control have an additional robustness layer with further consequences from the paradox of robustness.
- Wiring of control architecture: the evolutionary process of building layered control architectures yields seemingly haphazard, complex wiring of control.

The challenge becomes how to turn the theory into testable predictions for specific systems. One likely path is to find simple regulatory control systems that vary in architecture between closely related populations or species. Microbial systems often provide the best opportunities.

Additionally, the theory may potentially be applied to modern datasets on genetic variability and single-cell stochasticity of gene expression. It should be possible to make comparative predictions about the relative levels of variability of particular genes in relation to the function of those genes within particular regulatory control architectures. There is much room here for new insights and approaches.