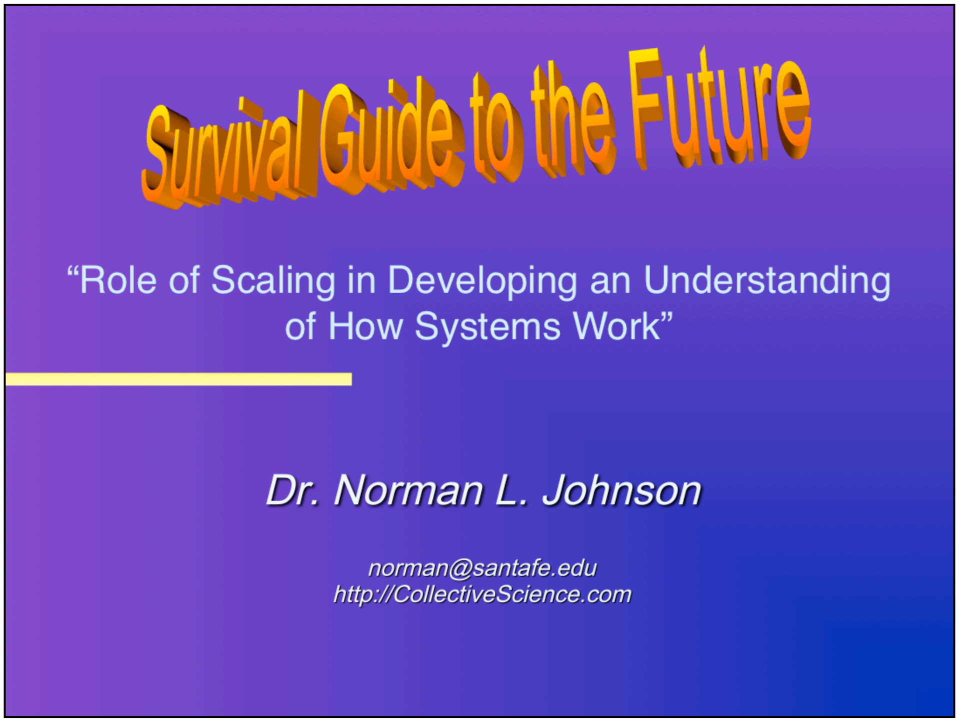
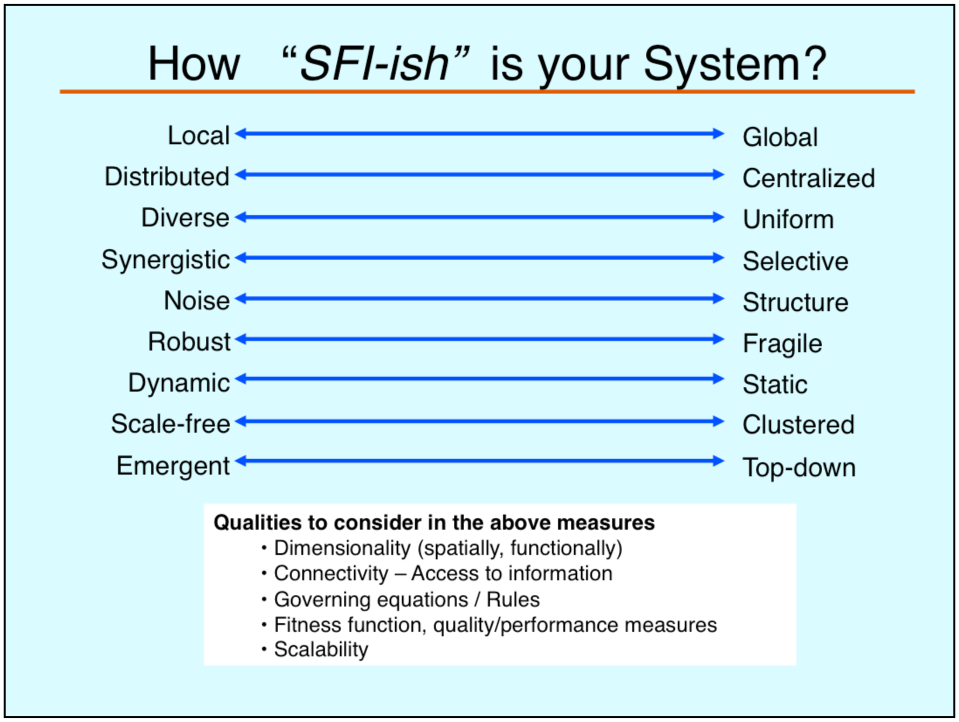


Goal is to talk about Scaling under the rubric of survival in a fast changing, increasingly complex world.

Introduction: This is not an area that I'm deeply familiar with, so I present the following much in the vain that you might do after you've spent some time reading about what's been said and how it relates to your area of interest.



Different ways to think about your system of interest. There is no real judgment as to which side of the distribution a "SFI-ish" system might be located, but certainly the list captures many research and application areas that SFI is currently or in the past focused on.

Note how these distributions are not independent.

For example a robust system often has diverse and distributed components that are locally controlled, and have some degree of chaos or noise and often express emergent properties and are dynamic rather than static systems.

Similarly a fragile system often has uniform and centralized components that are globally controlled and have a high degree of structure and top-down functions, and are often appear static until they break under stress or change.

Power Laws and Complexity:
Not without some controversy

"Over the last decade or so, it has become reasonably common to see people (**especially physicists**) claiming that some system or other is complex, because it exhibits a power law distribution of event sizes. Despite its popularity, this is simply a fallacy. No one has demonstrated any relation between power laws and any kind of formal complexity measure. Nor is there any link tying power laws to our intuitive idea of complex systems as ones with strongly interdependent parts."

In "METHODS AND TECHNIQUES OFCOMPLEX SYSTEMS SCIENCE: AN OVERVIEW" by Cosma Shalizi

No one at the SFI meeting was guilty of this.

Nor any SFI researchers that I know.

But it does illustrate that this isn't a subject that is just academic and obvious.

There is a lot of mystique about this topic that has been exploited in the popular literature, leading to this observation.

Lesson: beware of selling complexity and justifying it by making associations with technical observations.

(See the viewgraph near the end for the full reference.)

My Background

Polymer physics

Star Wars

Novel fusion device

Combustion modeling

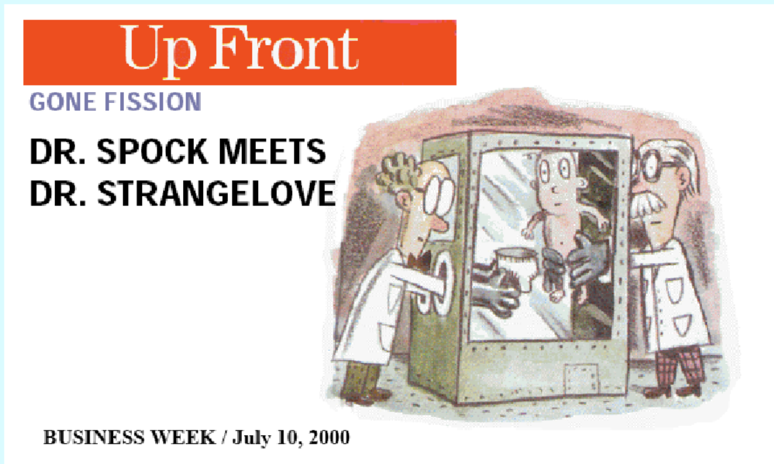
Hydrogen Fuel Program

P&G multi-phase flows

The next couple of viewgraphs represent some of my background - The main point is that I've been working closely with industry for the last 15 years at the interface between science and industry, particularly in the area of development of tools to reduce costs - mostly in the area of fluid dynamics modeling.

The P&G project is a great example of how to close the gap between application and deep science: P&G came to the multiphase fluid dynamics group twice to ask for collaboration. They were turned away. On the third visit they made their pitch that they have rich data in need of theory (a major lesson on how data is becoming more available - more on this in a second). Secondly they suggested starting small ($10ks). Ultimately the project became a million dollars a year at LANL and ended up saving P&G about 20 times more than they spend - with a continued return long after the end of the project: They now market the resources under P&G's name.

P&G Application Area?



There was a time that I was very happy that the project with P&G was highly proprietary and I couldn’t talk about what I was working on. Until this came out.

My Background

Polymer physics

Star Wars

Novel fusion device

Combustion modeling

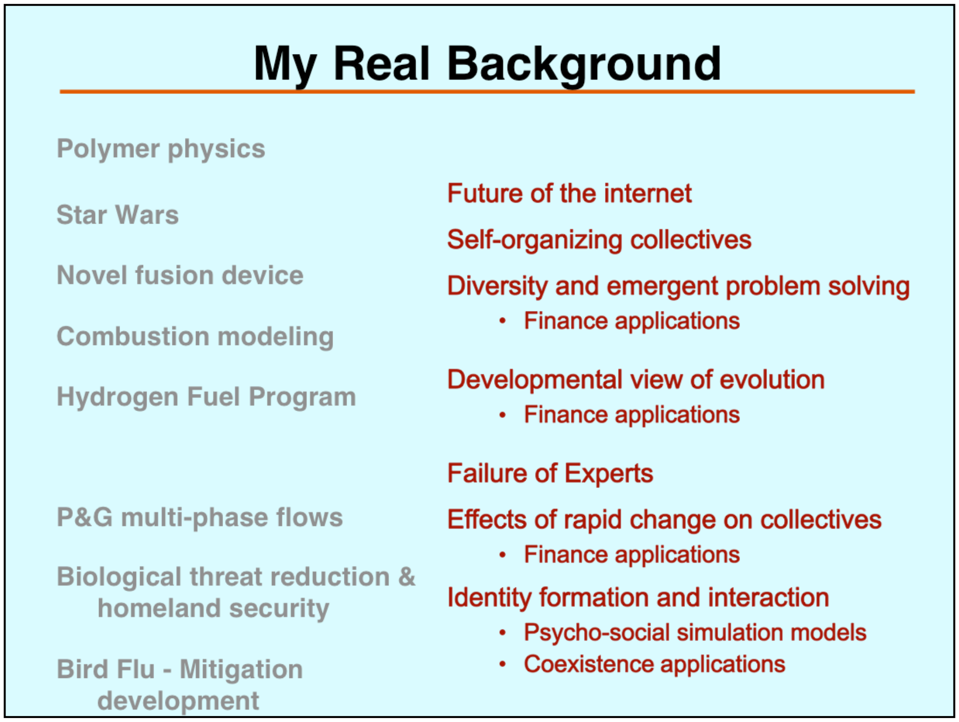
Hydrogen Fuel Program

P&G multi-phase flows

Biological threat reduction & homeland security

Bird Flu - Mitigation development

Recent problems I've worked on have shown me that much of the complexity of problems facing society cannot be solved by technological solutions, but require an exceptional integration of science and policy awareness. One of the major challenges is not to solve new problems in the same way we tried to solve old problems. That is why we are here today.



These topics speak very much to the reason that the interface between science and industry in this new age is so important. Much is changing and requiring tools that weren't available just a few years ago.

In my interactions with business this is the list that I hear are the biggest challenges - and is the reason that I think the first item is often the summary for all the items that follow. An why SFI is the interface to the future of doing business.

•I think this is the core problem - we see it in every facet of our personal and work lives. But Faster change refers not just to increased rates, but also the changes in the ordering of scales of change: for example, major changes occur in a worker's life when the average lifespan of a company because much shorter than the average work lifespan of a worker - as has strongly happened over the last few decades (see the book Creative Destruction by Foster)

•The data rich environment has two consequences: 1) we may feel we have too much data, much irrelevant, to make decisions (information overload) and 2) we don't have the tools to deal with the extra information. A relevant observation that connects item 1 and 3: computers are good processing large amounts of low complexity information, humans a good at processing small amounts of high complexity information. We have yet to develop the tool/resources/procedures for dealing with large amounts of highly complex information (See the papers on symbiotic intelligence at http://ishi.lanl.gov)

•"Fall of the house of experts" refers to the title of a talk I gave at an SFI public lecture - on how experts are failing us because of the complexity of the problems and how collective solutions a filling in the gap. Video available from SFI (a bit out of date)

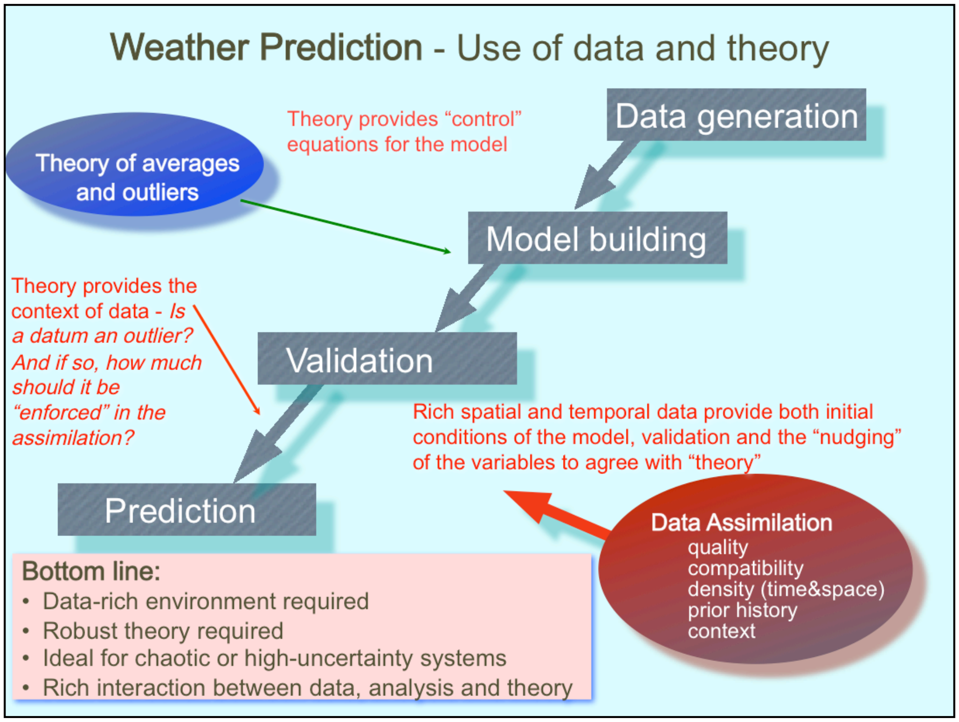
•Globalization - well documented.   An observation from the audience was that this is a complex issue that cuts many ways.

•Technology surprise: a major innovation, maybe not even technology but possibly a way of doing business (Amazon), can make rich companies poor quickly. Geof West observations about resetting the growth curve is quite relevant. I would add that some technologies can change the infrastructure, which may start a completely different growth curve.

•?? Pick your own. - Much discussion on how it's not just data rich, but knowing what data is important. I believe the issue also is the context of data - which gets lost in a fast

As with all challenges there are often opportunities. These are just a quick list, but this is an area for rich discussion among the business network.

Note that these are not just additive - a small improvement in a couple of the above areas and have a multiplicative or exponential increase in opportunity.



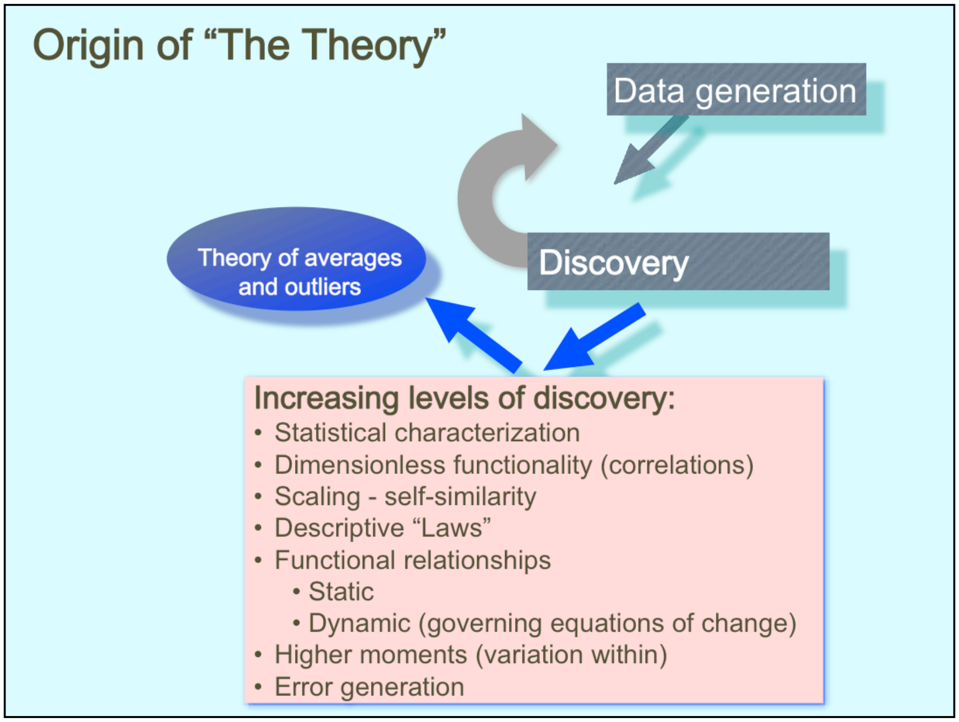
One example of how prediction of complex-chaotic system (the weather) made great advances:

By the combination of three advances:

- •rich data sources world wide
- •Theories or models that can give the data context
- •Fast computers

With these, we've been able to predict the worst of chaotic systems. I believe this is the paradigm for the future of business: as we develop better data and models, we'll be able to make better predictions and decisions.

Note that in order to deal with the chaotic nature of the system, **data assimilation** is required (using data real time to adjust the arrow to the future states)



This viewgraph illustrates the context and role of scaling or power laws in science (and business).

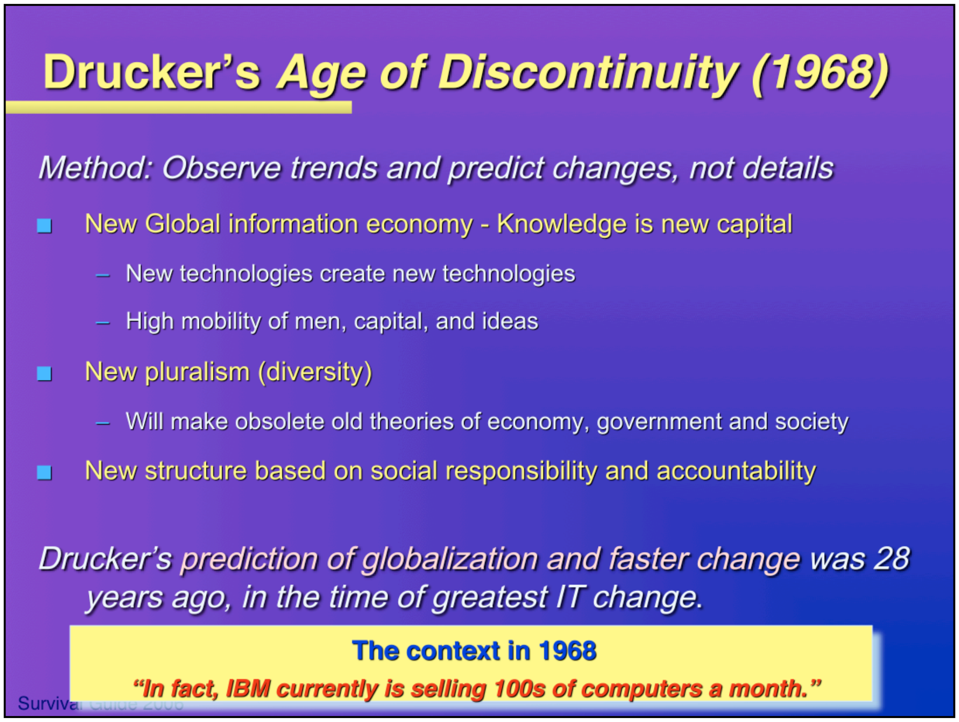
Observations:

•Most businesses stop at correlations in dealing with large amounts of data. The challenge and big payoffs are from driving further down in the list. My view is that this is why we are all here today.

•The last two items are rarely touched even in well developed sciences, but are proving to be the real resources needed for decision makers in dealing with complex systems with potentially severe unintended consequences of decisions. Much of this can be captured under the rubric of UNCERTAINTY MANAGEMENT.

•Higher moments refer to the variation of the data around the mean

•Error generation refers to the tracking of uncertainty in systems or of the noise in a system. (search on infodynamics on the web for background)



Ducker was the master at prediction of the future, in the most challenging time of our history.   I higlyh recommend reading the first chapter of this book on his approach to prediction. In many ways this is what scaling is all about: you focus on predicting trends, not details (technologies).

Just to put Ducker's achievement of predicting the information age in context: he includes the quote above as proof that computers are important. The personal computer didn't even exist at this time.

Power Laws and Complexity (con't)

"… it has been known for half a century that there are many, many ways of generating power laws, just as there are many mechanisms which can produce Poisson distributions, or Gaussians.

"Perhaps the simplest one is that recently demonstrated by Reed and Hughes, namely exponential growth coupled with random observation times.

"The observation of power laws alone thus says nothing about complexity (except in thermodynamic equilibrium), and certainly is not a reliable sign of some specific favored mechanism, such as self-organized criticality or highly-optimized tolerance."

In "METHODS AND TECHNIQUES OFCOMPLEX SYSTEMS SCIENCE: AN OVERVIEW" by Cosma Shalizi, Center for the Study of Complex Systems, University of Michigan, Ann Arbor

Major point: observation of a power law (or more generally scaling) is of little use by itself. When supported by a model or theory, then it gets exciting.

The cited paper by Reed and Hughes is useful because it addresses a simple process that causes powerlaw behavior that has not been appreciated. (See the viewgraph near the end for the full reference.)

Importance of Power Laws - Farmer

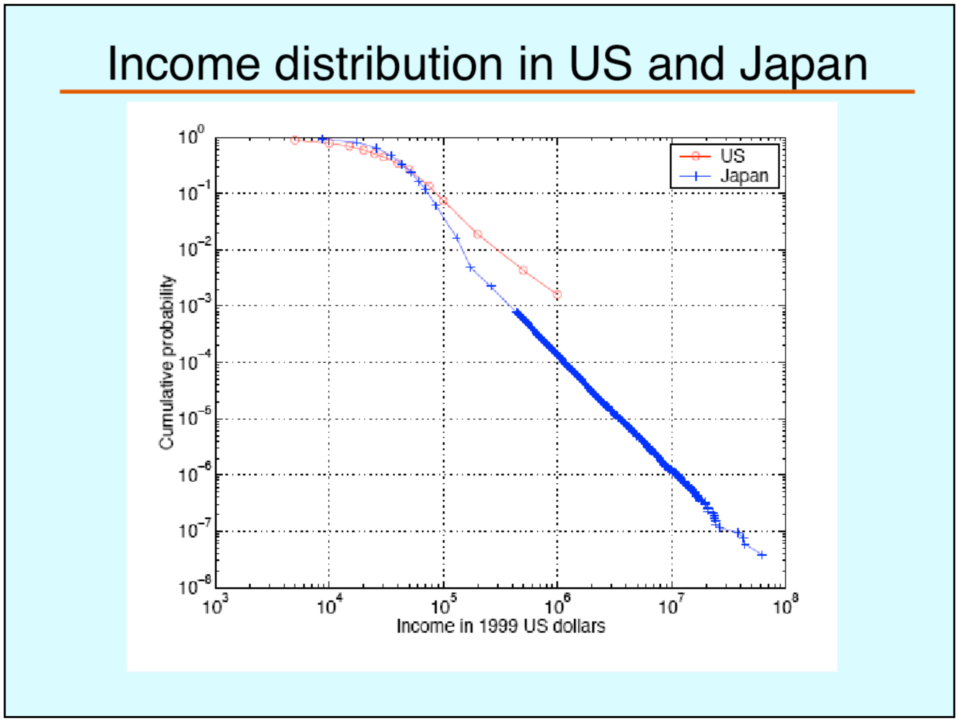
"Scale free behavior has important scientific implications because it strongly suggests that the same mechanism is at work across a range of different scales."

The best technical summary for power law behavior is by Doyne Farmer (see references at end).

A question was: how can you resolve the above statement with the second viewgraph on scaling and complexity?   There are two different points:

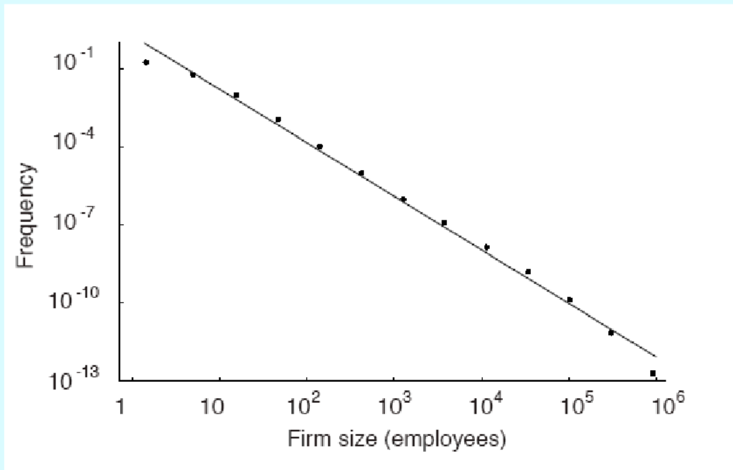
•The statement that power laws or scale-free behavior implies complexity is not correct

•But generally power laws or scale-free behavior often, but not always (see the Reed reference in the previous viewgraph) imply a fundamental mechanism controls behavior over many scales.

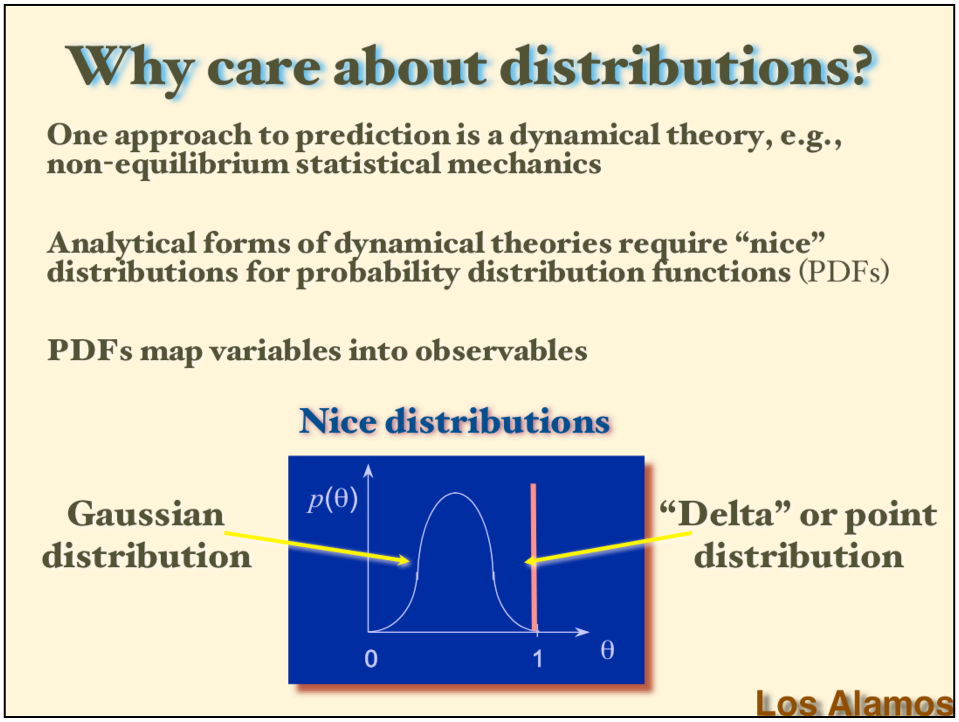

Income distribution in US and Japan
US
Japan
Cumulative probability
Income in 1999 US dollars
10^0
10^-1
10^-2
10^-3
10^-4
10^-5
10^-6
10^-7
10^-8
10^3
10^4
10^5
10^6
10^7
10^8

An example

US Firm Size by Employees



Another example that is particularly remarkable.



Just a quick reason why we care about distributions - from a science viewpoints. (We also care about them because they tell us about the behavior of the systems.)

The problem is that because we often try to develop analytical theories, we often force these distributions into unrealistic but "nice" ones. This has been a major source of controversy in many fields.

Probability Distributions (ref. Farmer)

Normal (Gaussian) distribution

$$P(x) \sim \exp(-\frac{x^2}{2\sigma^2}).$$

Defined on (−∞, ∞)

Natural explanation: Central limit theorem: Sum of many random variables (second moment finite).

Many applications: Maxwell: velocity distribution of particles in a gas
Heights of individuals, IQ, ...

Distribution is thin tailed: No one is 10 feet tall

One common distribution.

Probability Distributions (ref. Farmer)

Normal (Gaussian) distribution

$$P(x) \sim \exp(-\frac{x^2}{2\sigma^2}).$$

Defined on (−∞, ∞)

Natural explanation: Central limit theorem: Sum of many random variables (second moment finite).

Many applications: Maxwell: velocity distribution of particles in a gas
Heights of individuals, IQ, ...

Distribution is thin tailed: No one is 10 feet tall

Main point: thin tailed (more on this in a second)

Probability Distributions (ref. Farmer)

Exponential distribution

$$P(x) \sim \exp(-x/x_0)$$

Defined on [0, ∞)

Natural explanation (1): Survival times for constant probability of decay

Natural explanation (2): Equilibrium statistical mechanics
(Maximum entropy subject to constraint on mean)

Many applications: Radioactive decay, Energy distribution at equilibrium, ...

Distribution is also thin tailed: characteristic scale x_0.

Ditto on thin tailed.

Probability Distributions (ref. Farmer)

Power law

$$P(x) \sim x^{-\alpha}$$

Defined on [a, ∞). α, a > 0.

Natural explanation?

Distribution is heavy tailed, no characteristic scale.

Many applications –
is there a common link?

A power law is a linear relation between logarithms

$$\begin{aligned} f(x) &= Kx^{-\alpha} \\ \log f(x) &= -\alpha \log x + \log K \end{aligned}$$

Major note: Not thin tailed, but heavy tailed. These distributions can cause many problems in analytical treatment in science.

Difference between Thin and Thick Tailed distributions

Similar for frequent events

1/1000 event is twice as large for a power law

1/10,000 event is three and a half times as large

"The probability of observing a fluctuation of 21% (the size of the famous negative S&P return on October 19, 1987) under the normal hypothesis is less than 10^{-16}, whereas the probability under the power law distribution is 0.08%. Under the normal distribution it is essentially impossible that this event could ever have occurred, whereas under a power law distribution such an event is to be expected." - Farmer

What's the difference between thin and thick tails. An example.

Many Examples of Power-Law Distributions

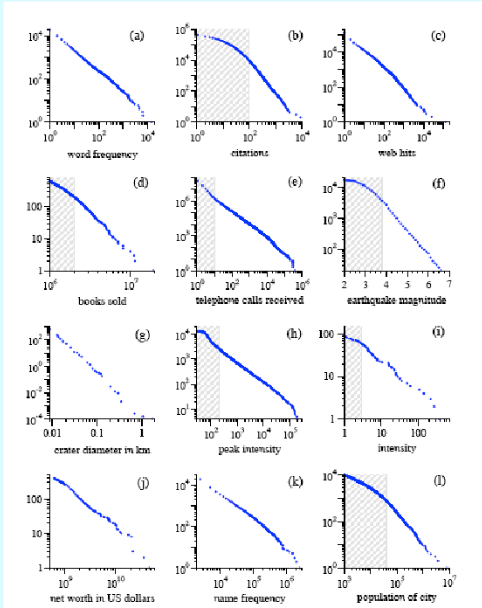
Frequency vs. measure

earthquakes	DoD cost overruns
flood levels	frequency of word usage
rainfall	name frequency
craters	authored papers citations to papers
insurance claims	patent profitability
income	music sales
price changes	movie sales
firm size	book sales
transaction volume	telephone calls
price for order placement	number of email messages
city size	size of computer files
intensity of wars	hits on web page
length of strikes	links to web sites

Lots of examples.

Many Examples of Power-Law Distributions

Frequency
vs.
measure



Pictures. Note the range on the horizontal axis: For ones with only a few decades of range, these are questionable power laws (see later viewgraphs). Ones with many decades are pretty solid.

What's Special about Power Laws

Power laws are scale invariant

- Retains functional form under scaling

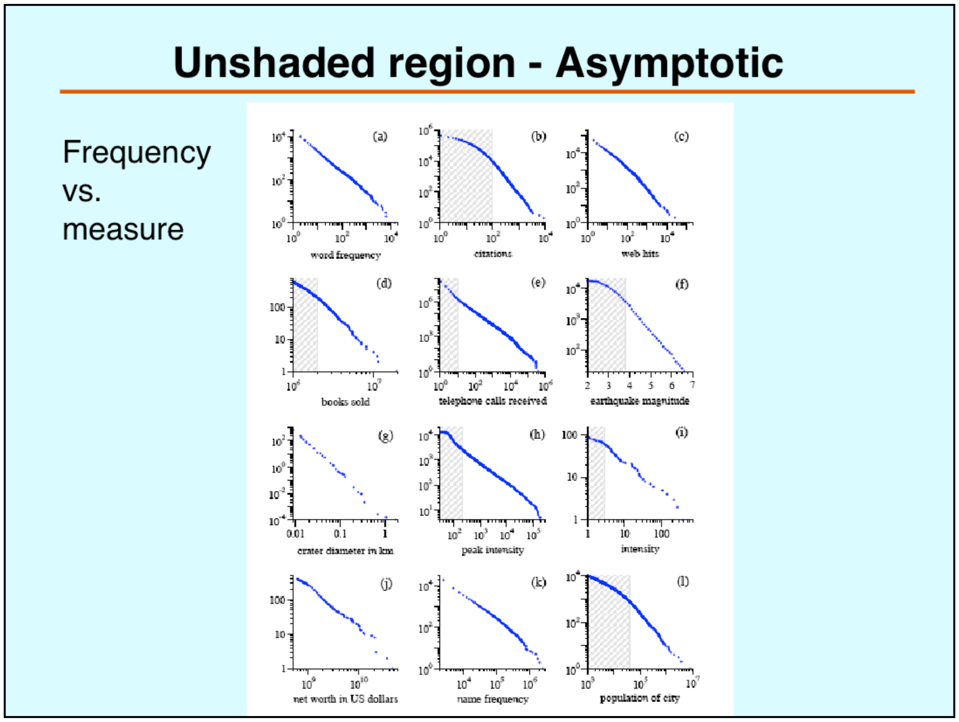
Power law is an asymptotic relation

- Only meaning full in a bounded region
- Practically, real problems always have cut-offs

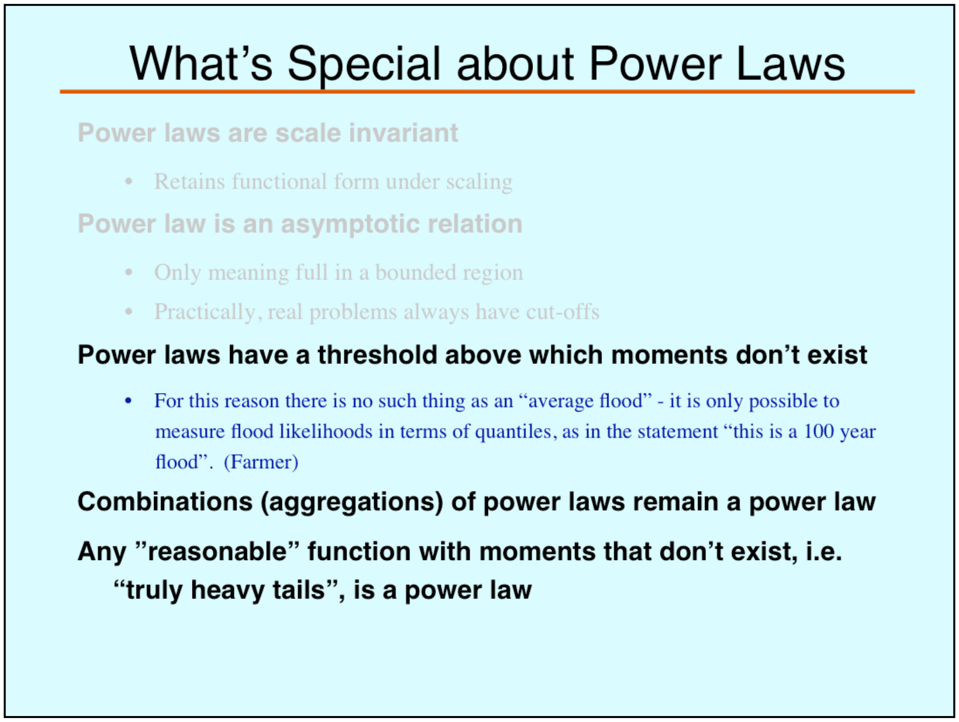
Previous speakers have addressed these:

-the first is important to applications.

-The second is a reminder that the real world has bookends that may not allow the power law region to be very broad.



The shaded portion illustrated the non-powerlaw regions.



These are more esoteric properties of paper laws - but very important in a general treatment and not generally appreciated. For a complete discussion see the paper (chapter) by Farmer in the references at at end.

Power Laws vs. Long-Memory

Difficult to determine whether power laws or long-memory are present by empirical sampling

Tails by definition only have a small number of events. May not have enough data to probe tails.

Can have slow convergence to tail (slowly varying functions)

Can have cutoffs of tail due to finite size effects (e.g. physical limits)

Thin tailed distributions may mimic power law behavior over a wide range

Statistical convergence for long-memory processes is very slow

Doyne Farmer has a section on long memory effects that can also lead to power law behavior. The point is that a system with no memory (or little) memory effects often are the systems that are identified to have power law behavior. But the long memory effect should also be a consideration when thick tail behavior is observed.

This viewgraph also speaks to the difficulties around sampling systems with memory effects.

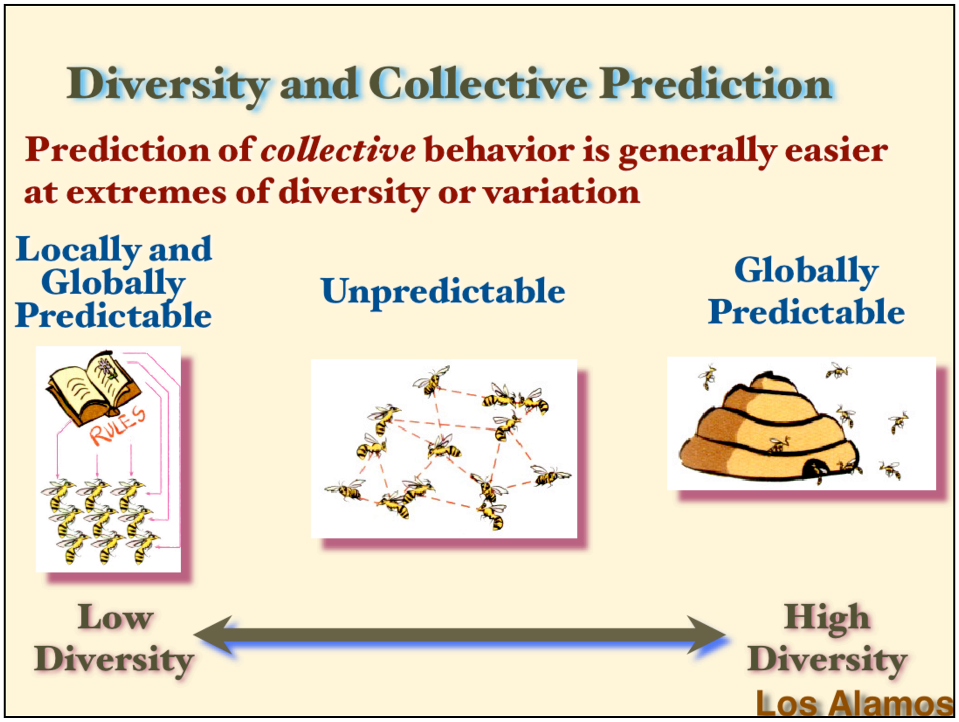
Importance of Power Laws - Farmer

"Scale free behavior has important scientific implications because it strongly suggests that the same mechanism is at work across a range of different scales."

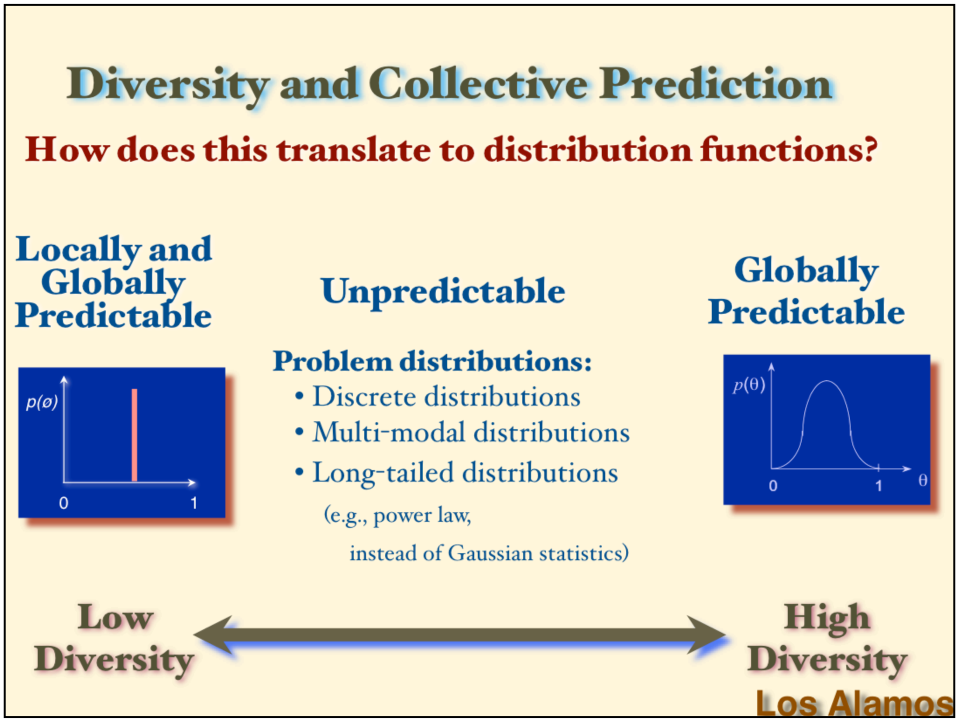
"The real test is whether power laws can improve our predictive or explanatory power by leading to better models. Self-similarity is such a strong constraint that, even if only an approximation over a finite range, it is an important clue about mechanism."

"Ultimately, the best method to demonstrate that power laws are applicable is to construct theories that also have more detailed testable predictions."

The over-all perspective: Yes, they are useful, but their full utility only comes out when supported with a theory or model. Note that this is not only to test the origin of the power aw behavior but also to help in the sampling strategy for data acquisition that is a challenge for these systems. Here the point is that incorrect sampling of distributions may lead to incorrect conclusions.



Back to distributions and prediction – with respect to diversity or heterogeneity of the system. Turns out that a little or a lot of diversity (that is well sampled) is good for prediction. The qualifier “well-sampled” diversity is required because some systems have lots of diversity that is poorly interconnected and therefore the diversity really doesn’t really get sampled, which has a major effect on the dynamics or robustness of the system – a prime example is a senescent ecosystem: lots of diversity but very restrained interactions. Same is true for old economies.



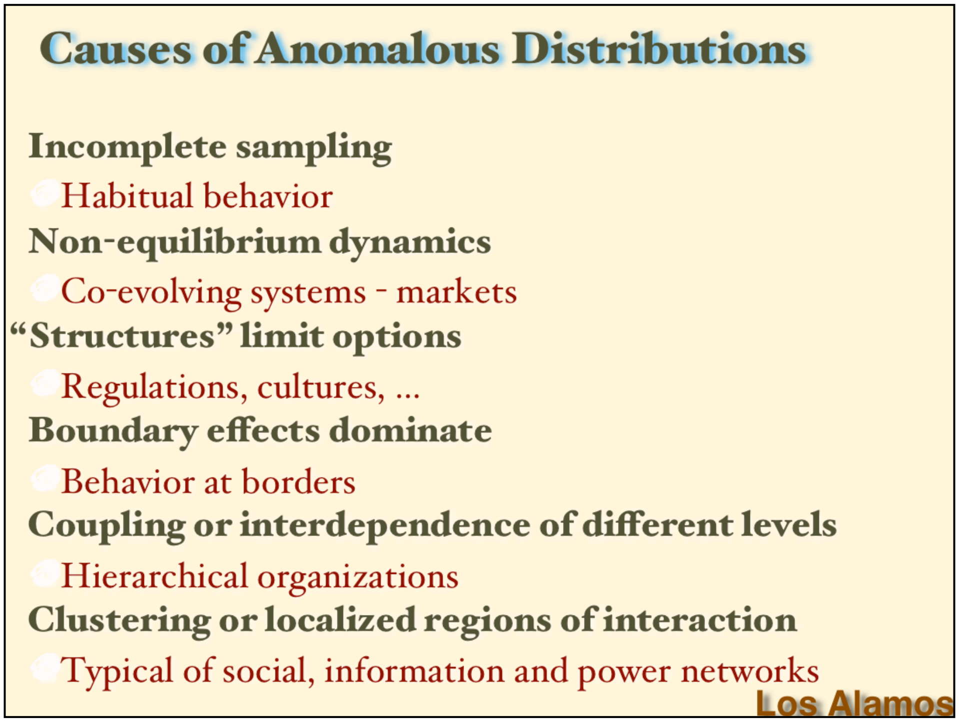
So what causes distributions to be not "nice"? One list is given above. You can read lots on this looking at the work by Tsallis (more on this in a bit).

Mechanisms for Generating Power Laws

- Critical points and deterministic dynamics
- Random processes
- Sampling from (terminated) exponential growth (Reed and Hughes)
- Mixtures
- Dimensional constraints
- Maximization principles
- Preferential attachment
- Non-equilibrium statistical mechanics

From Farmer.

This is an "academic" list of mechanisms for generation power laws. Books have been written on each of these topics.



This is a old viewgraph of mine (circa 2001) that describes the mechanism for generating anomalous distributions and mapping it over to real systems. This list has many similarities to the previous list for powerlaws and may be more general.

Statistical Issues Relating to Power Laws

"some common mistakes"

Parameter Estimation

Use linear regression to find the line of best fit to the points on the log-log plot. But the line minimizing the sum of squared errors is not a valid probability distribution, and so this is simply not a reliable way to estimate the distribution.

Error Estimation

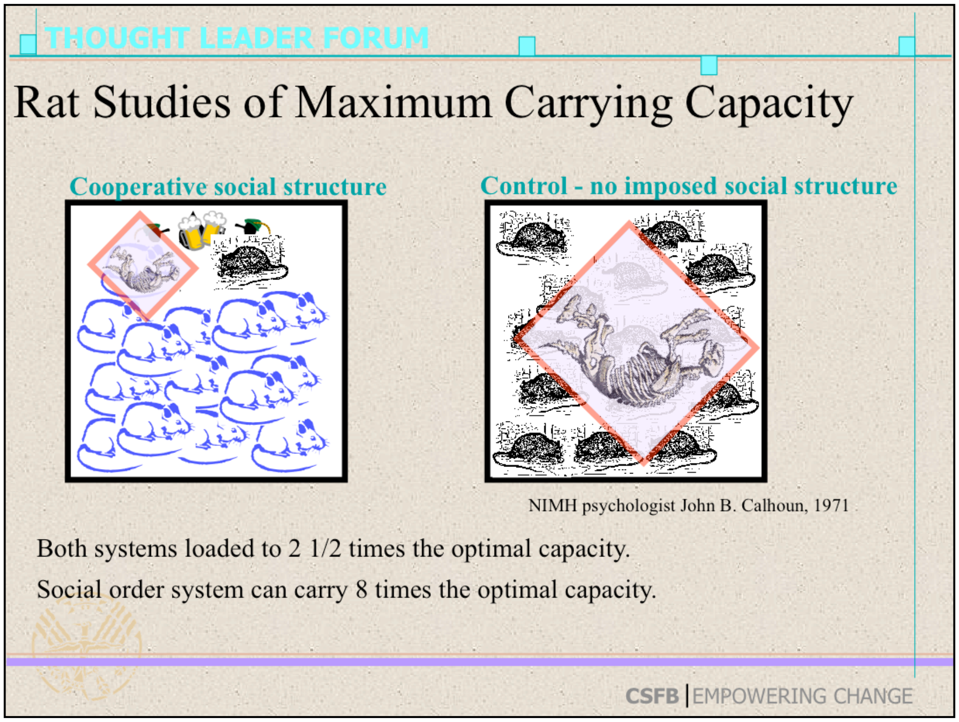
Estimate of the standard error in the estimated slope and report this as the uncertainty in the power law. "This is an entirely unacceptable procedure." On a log-log plot this violates the assumption that measured values have Gaussian fluctuations around their true means.

Validation

"The basic problem here is that any smooth curve looks like a straight line, if you confine your attention to a sufficiently small region and for some non-power-law distributions, such sufficiently small regions can extend over multiple orders of magnitude.

In "METHODS AND TECHNIQUES OFCOMPLEX SYSTEMS SCIENCE: AN OVERVIEW" by Cosma Shalizi

If you find yourself looking at data with apparent power law behavior, this advice is for you.



Because of a discussion during the break I added this viewgraph - to illustrate impact of habitual behavior on social systems.

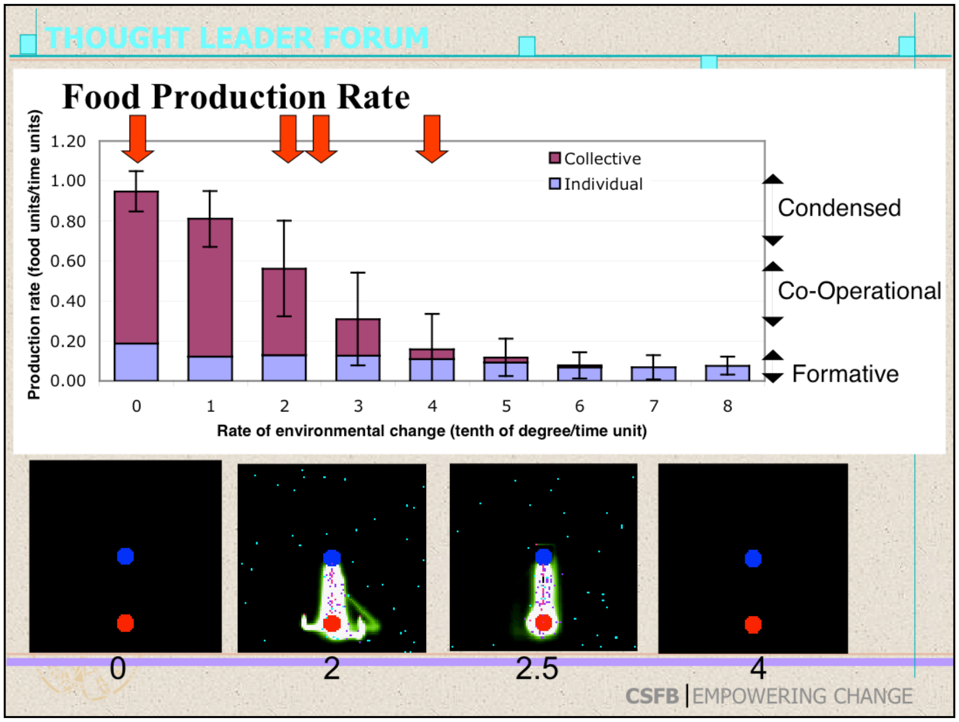
You can read more about this experiment in: http://www.capatcolumbia.com/CSFB%20TLF/2002/johnson_sidecolumn.pdf

I'd generally recommend looking at the other talks at:

http://www.leggmason.com/thoughtleaderforum/2006/index.asp for 2003-2006

http://www.capatcolumbia.com/CSFB%20Thought%20Leader%20Forum.htm for 2000-2003

Contact me for the reference.



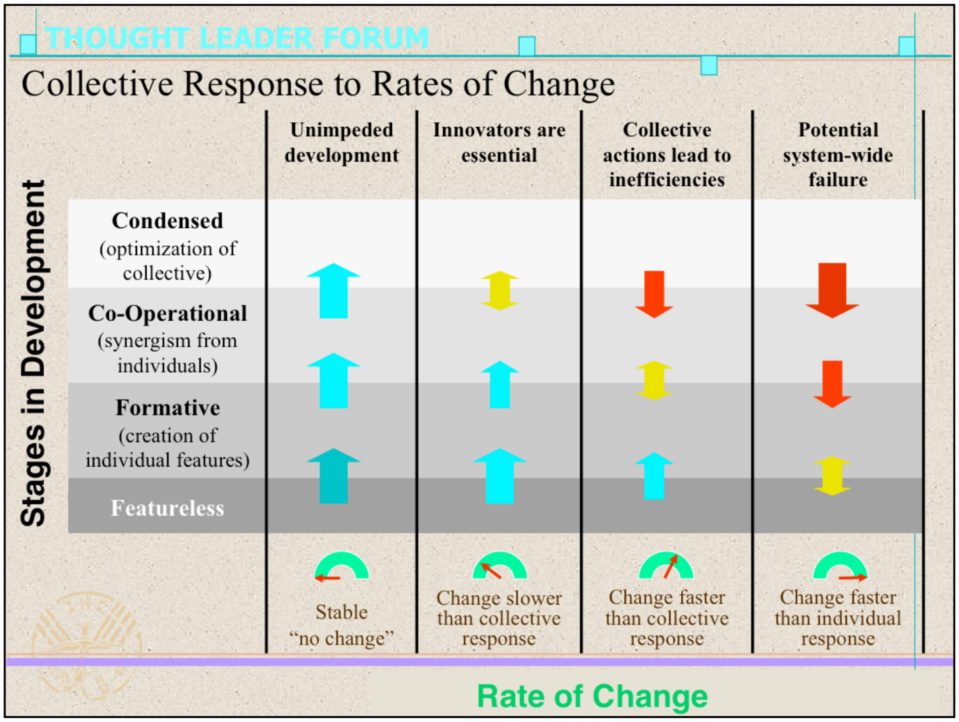
The system is food foraging by ants with a moving food supply. It illustrate different stages of development.

This illustrates the effect of rate of change on a self organizing system.

Note how as the food moves faster, the collective contribution declines and becomes more erratic (the "error" bars for each stack).

The take away is that rates of change do drive systems into different stages, each with their characteristics,properties, and different scaling behavior.

Contact me for the movies and a paper on this topic. Also net-logo model for playing with it yourself.



Summary of the effect of change on the collective system;
increased change forces the system to earlier stages in its development.

The major lesson here is that most of the systems we have discussed today have had time to develop - a long time for biosystems. But many financial systems are relative young and potentially unstable. Are they as likely to show scaling or power law behavior? This is a major unanswered question.

On Constructing Theories (Models) and Testing

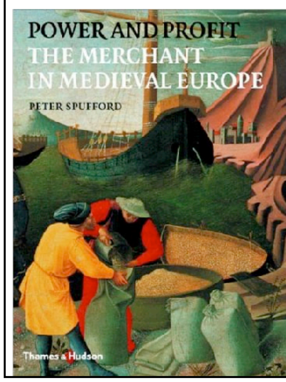
1. Replication is essential.
2. It is a good idea to share not just data but programs.
3. Always test the robustness of your model to changes in its parameters. (This is fairly common.)
4. Always test your model for robustness to small changes in qualitative assumptions. If your model calls for a given effect, there are usually several mechanisms which could accomplish it. If it does not matter which mechanism you actually use, the result is that much more robust. Conversely, if it does matter, the over-all adequacy of the model can be tested by checking whether that mechanism is actually present in the system. Altogether too few people perform such tests.

In "METHODS AND TECHNIQUES OFCOMPLEX SYSTEMS SCIENCE: AN OVERVIEW" by Cosma Shalizi

Ok, so you have a model of the power law behavior, what should you do to investigate the robustness of the behavior and is accuracy as an explanation?

European Conference on Complex Systems Paris, 14-18 November 2005

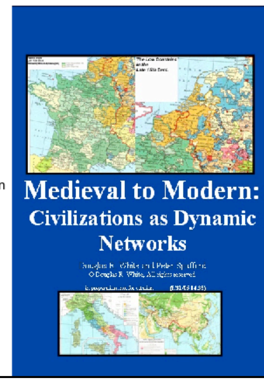
Civilizations as dynamic networks



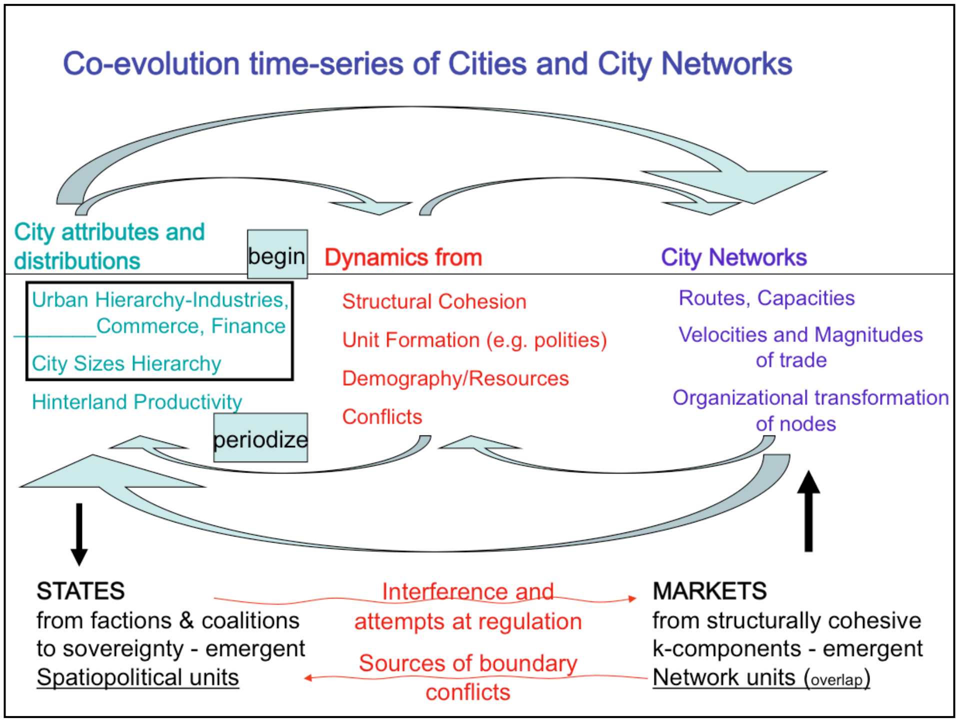
Cities, hinterlands, populations, industries, trade and conflict

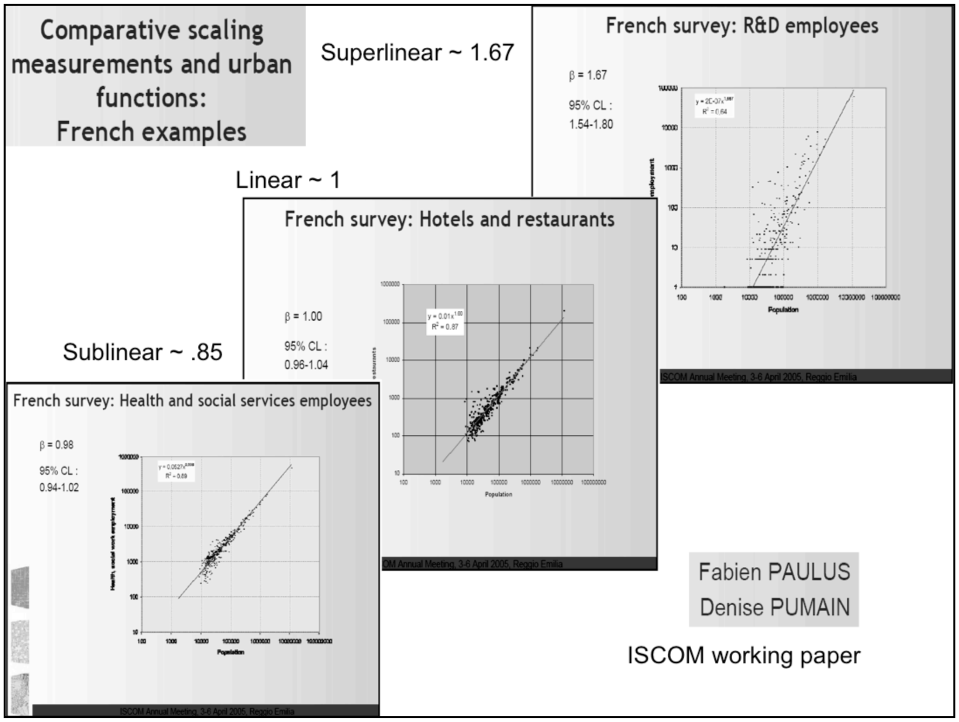
Douglas R. White

50 slides - also viewable on drw conference paper website version 1.3 of 11/12/2005



A complex model of human city and city network formation and dynamics. From an SFI regular. This is a excellent set of viewgraphs on the development of networks and dynamics on those networks


Co-evolution time-series of Cities and City Networks
City attributes and distributions
begin
Dynamics from
City Networks
Urban Hierarchy-Industries, Commerce, Finance
City Sizes Hierarchy
Hinterland Productivity
periodize
Structural Cohesion
Unit Formation (e.g. polities)
Demography/Resources
Conflicts
Routes, Capacities
Velocities and Magnitudes of trade
Organizational transformation of nodes
STATES
from factions & coalitions
to sovereignty - emergent
Spatiopolitical units
Interference and attempts at regulation
Sources of boundary conflicts
MARKETS
from structurally cohesive
k-components - emergent
Network units (overlap)

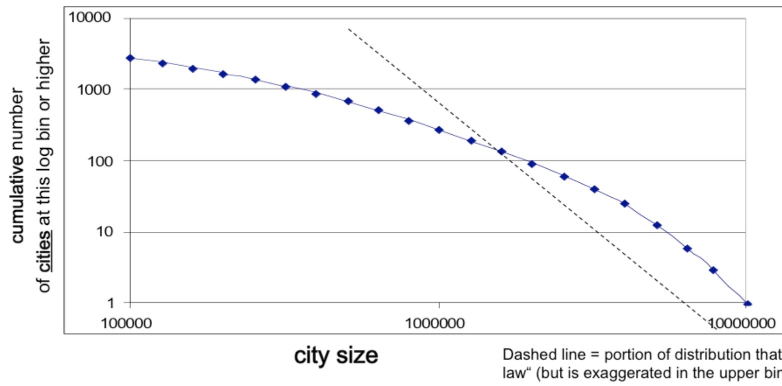


Doug's examples in his area of scale free behavior. Note the distribution around the mean.

Urban Scaling: City Sizes - 1950 United Nations data for world cities

Compare the scale K and α coefficients of the power-law $y(x) \approx K x^{-\alpha}$ (and Pareto β=_α+1) with the *q*-exponential parameters for *q* slope and scale κ in $y(x) \sim [1 + (1-q)\, x/\kappa)]^{1/(1-q)}$, fitted to entire size curves

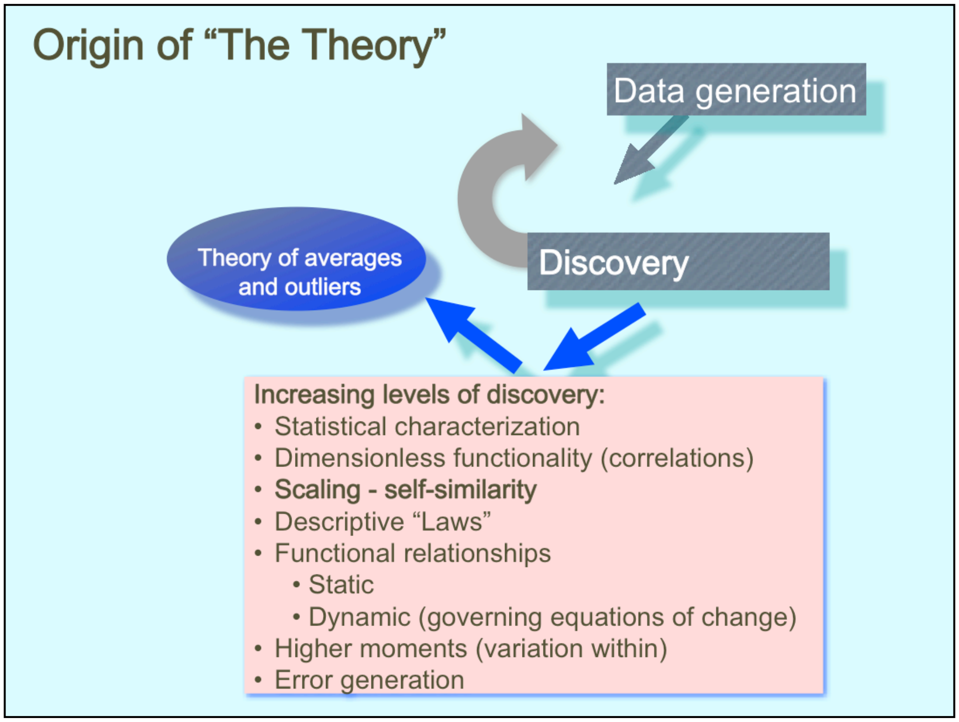
Power laws and Zipf's law might fit upper bin frequencies for city sizes but not the whole curve



Dashed line = portion of distribution that is "power-law" (but is exaggerated in the upper bins)

(White, Kejžar, Tsallis, and Rozenblat © 2005 working paper)

This plot is the main reason for showing excerpts from Doug's set of viewgraphs. Doug makes the point that many of the curves he observes that might initially seem to be powerlaw are better fitted by an alternative "Q-exponential" fit that was developed by Tsallis. This in not a scale free distribution in general (although can reduce to one) and is argued to occur for many of the reasons that I listed in the previous viewgraph on the origins of anomalous distributions.



A summary: So here is where power law distribution fit into the bigger picture of discover of trends in the data. It is very important to not stop there, but to continue down the list if possible. And to develop theories if possible.

Summary

Why true power law observations are exciting

- Essential to discovery of governing "laws"
- Often from a critical process in a complex system

But power laws are only part of the process to developing a broader understanding. Not meaningful without a model or theory.

Power laws behavior doesn't mean complexity

Power laws are

- (Idealized) asymptotic behavior - beware of the boundaries!
- Represent thick tails, rather than thin tailed distributions

In general, anomalous distributions may dominate real systems

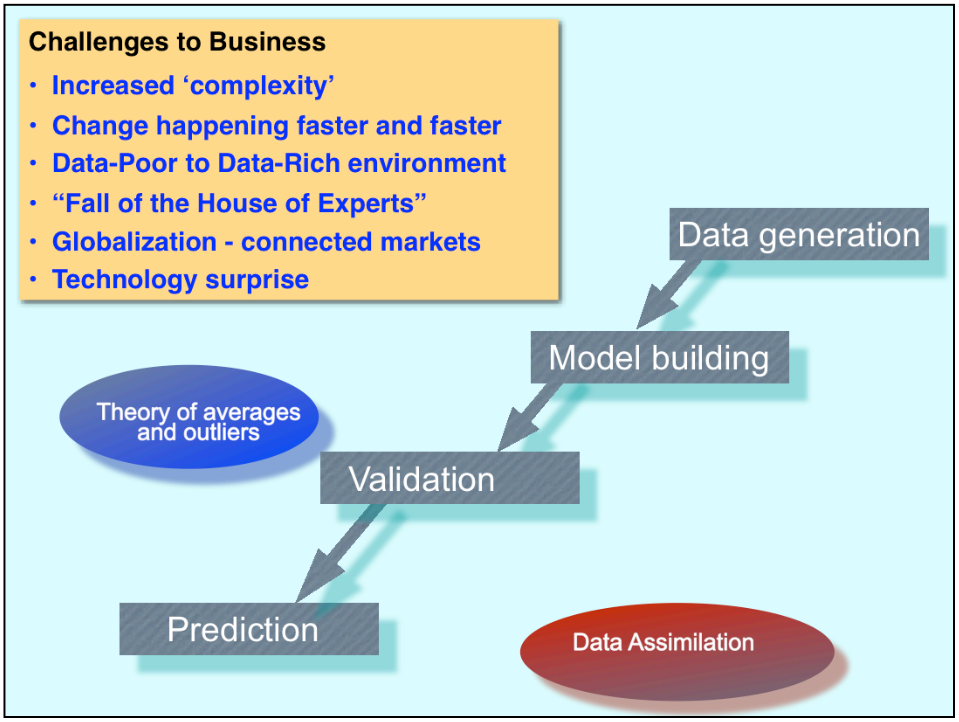
Consider maturity of system evolution - "young" systems may be very different

- Or mixed maturity of sub-systems

Consider optimization versus robustness

Best resources on the subject (appear to be) from SFI

A summary of what we observed about power law behavior and what should be considered.



My speculation is that the future is all about data-rich, complex systems and developing tools that help us predict their behavior. The above process is the path that will be taken, and much of what has been discussed at this workshop address the discovery of the processes that will enable us ultimately to predict these systems.

References

Shalizi, Cosma R., “METHODS AND TECHNIQUES OF COMPLEX SYSTEMS SCIENCE: AN OVERVIEW”, Chapter 1 (pp. 33-114) in Thomas S. Deisboeck and J. Yasha Kresh (eds.), Complex Systems Science in Biomedicine (New York:Springer, 2006) http://arxiv.org/abs/nlin/0307015

Farmer, J. Doyne & John Geanakoplos, “Power laws in economics and elsewhere”, DRAFT April 4, 2005 (chapter from a preliminary draft of a book called “Beyond equilibrium and efficiency”) - Contact the authors for a copy.

Farmer, J. Doyne, “Power laws”, Santa Fe Institute Summer School June 29, 2005. Contact the author for a copy.

Holbrook, Morris B.. 2003. "Adventures in Complexity: An Essay on Dynamic Open Complex Adaptive Systems, Butterfly Effects, Self-Organizing Order, Co-evolution, the Ecological Perspective, Fitness Landscapes, Market Spaces, Emergent Beauty at the Edge of Chaos, and All That Jazz http://www.amsreview.org/articles/holbrook06-2003.pdf

White, Douglas R., “Civilizations as dynamic networks: Cities, hinterlands, populations, industries, trade and conflict”, European Conference on Complex Systems Paris, 14-18 November 2005. http://eclectic.ss.uci.edu/~drwhite/ ppt/CivilizationsasDynamicNetworksParis.ppt

For exceptional talks on Complexity in financial systems, see the Thought Leaders Forums:

- http://www.leggmason.com/thoughtleaderforum/2006/index.asp for 2003-2006
- http://www.capatcolumbia.com/CSFB%20Thought%20Leader%20Forum.htm for 2000-2003

Topics for Discussion?

What are the implications in your application area of the shift to being "data rich"?
Are the moments around the norm important? Are you driven by the outliers or norms?
How much can the results presented today be generalized?

- Stephanie's table of correspondence? Worth more development / study?
- Are global properties always determined by the "supply chain" - as Geoff observed in organisms?
- How do physical/operational restrictions dominate the growth of a system?
- In complex composite systems, does the host always dominate the dynamics?

Can human-designed systems be better planned by using general "laws"?
Bio-inspired solutions - where/when valid?

- Are ant colonies more like bio organisms or human societies?
- What about material versus information differences?
  - (If I give you an apple, I don't have an apple; is information the same, why not?)
- How does basic scales affect the system: e.g., cell size is fixed, but computer size is not

What can be concluded about network scaling and global system performance?
Social systems and innovative reset (Singularity horizon?)

- Assumes infrastructure is largely unchanged
- Can innovation reset infrastructure?

What about optimization of Efficiency versus Robustness?

- In times of faster change, robustness is maybe more important

Stages of development and maturity

- How much what has been observed today is only for mature, well developed systems?
- How do early innovative systems behave? How do opportunities and approaches differ?

Others?