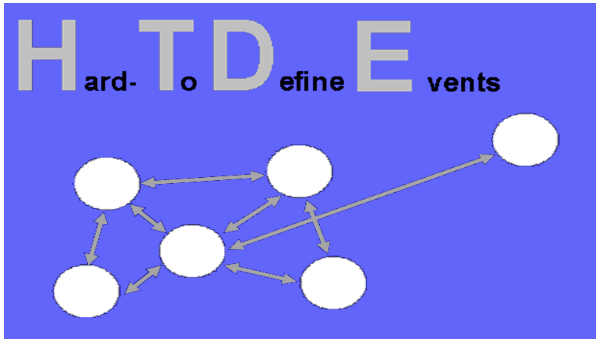
Welcome to the Hard-to-Define Events Workshop



HTDE (Hard-to-Define Events) Workshop 2012,
Artificial Life XIII, East Lansing, MI   USA

Bradly Alicea
Michigan State University
http://www.msu.edu/~aliceabr   http://syntheticdaisies.blogspot.com

What is this Workshop About?

Philosophy of Science?

Computational Theory?

Cognitive Science?

Measurement/Analysis?

Science as Practice?

Inspired by a need to synthesize a methodology of getting at "what we least know"

* experimentation relies on observables.

* good theory relies on experimental verification.

What if we cannot define our problem scope or variables very well?

What if they are highly context-dependent?

Hard-to-Define Events are actually two interrelated problems

Measurement (how to quantify things, determine causality, interpret anomalous results) Defining complex variables (many dimensions, highly nonlinear, self-referential) Abstraction (how real-world phenomena get represented in a model) Discovery (how to incorporate new variables into an existing model)	

Hard-to-Define Events are actually two interrelated problems

Measurement (how to quantify things, determine causality, interpret anomalous results)	**Analysis** (how to characterize non-uniform distributions and rare events)
Defining complex variables (many dimensions, highly nonlinear, self-referential)	Rare events (things that occur at an extremely low frequency, hard to observe)
Abstraction (how real-world phenomena get represented in a model)	No parameterization possible (summary statistics result in more questions than answers)
Discovery (how to incorporate new variables into an existing model)	Is formal analysis appropriate (do the appropriate tools exist)?

Different Fields, Different Perspectives

Mathematics: Modeling of distributions (long tails, failure rate).

Philosophy: Fuzzy logic, causality.

Computer Science: Novel representational schemes, computational complexity, hidden variables, anomaly/outlier detection (KDD).

Biology/Medicine: Diagnosing disorders, Causation of disease (rare variants/disease states), high-throughput analysis, automated assays.

Economics: Risk management.

Social Sciences: Constructs (variables), Qualitative phenomena.

Physics: Chaos, power laws, and uncertainty.

Is that a fly in my science????

* unexpected results (unpredictability).

The "UFO" category…..

* everything is generic until understood.

It was a one-in-a-million shot (or luck):

* low-frequency, large-scale events.

Is that a fly in my science????

* unexpected results (unpredictability).

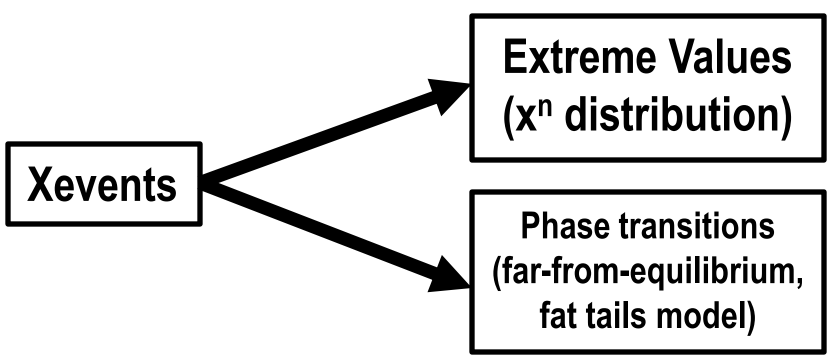
The "UFO" category…..

* everything is generic until understood.

It was a one-in-a-million shot (or luck):

* low-frequency, large-scale events.

Xevents model:



Albeverio, S., Jentsch, V., and Kantz, H. (2007). Extreme Events in Nature and Society. Springer, Berlin.

Is that a fly in my science????

* unexpected results (unpredictability).

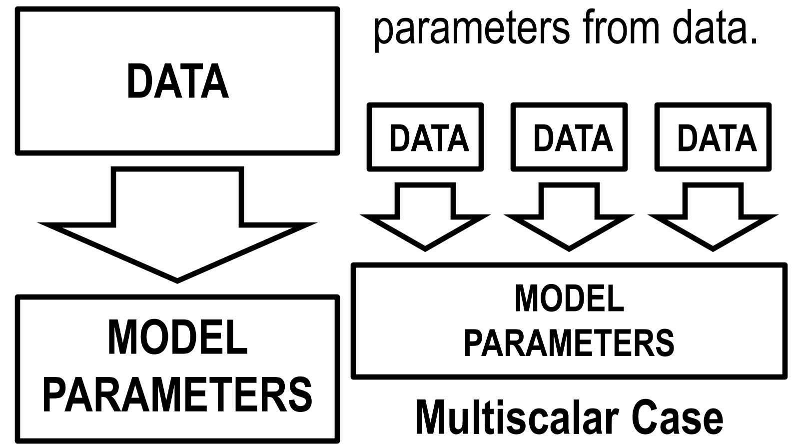
The "UFO" category.....

* everything is generic until understood.

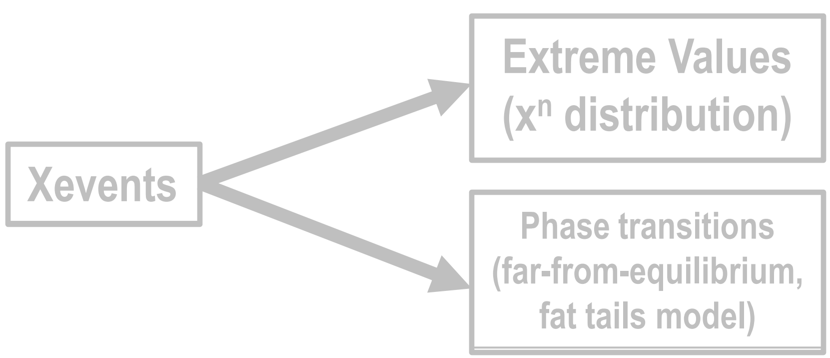
It was a one-in-a-million shot (or luck):

* low-frequency, large-scale events.

Inverse Problem: Infer model parameters from data.



Xevents model:



Albeverio, S., Jentsch, V., and Kantz, H. (2007). Extreme Events in Nature and Society. Springer, Berlin.

Is that a fly in my science????

* unexpected results (unpredictability).

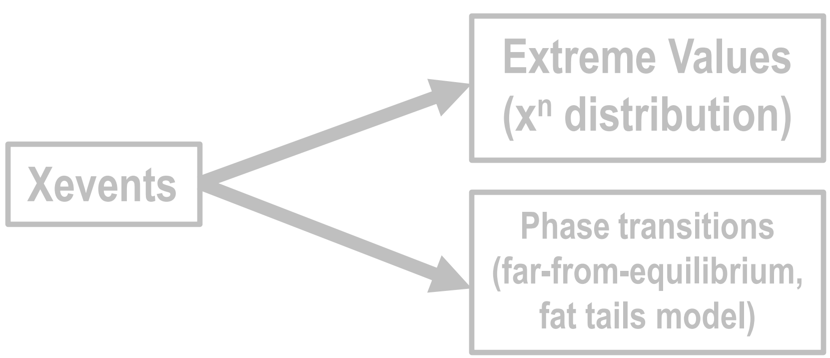
The “UFO” category.....

* everything is generic until understood.

It was a one-in-a-million shot (or luck):

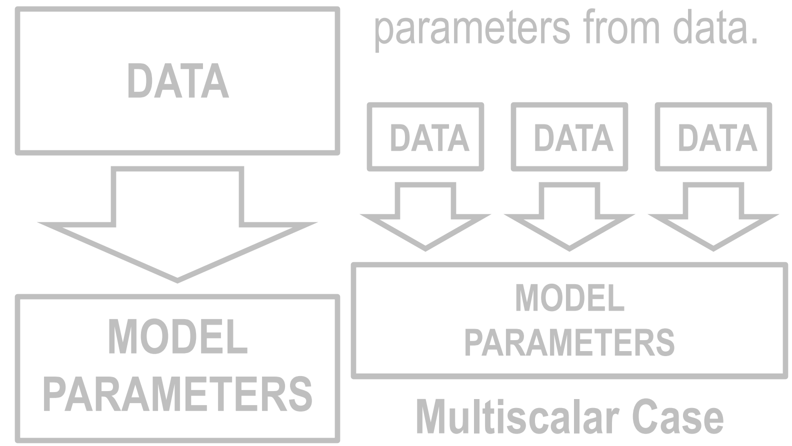
* low-frequency, large-scale events.

Xevents model:



Albeverio, S., Jentsch, V., and Kantz, H. (2007). Extreme Events in Nature and Society. Springer, Berlin.

Inverse Problem: Infer model parameters from data.



Ill-posed Problem:

1) No unique solution exists.

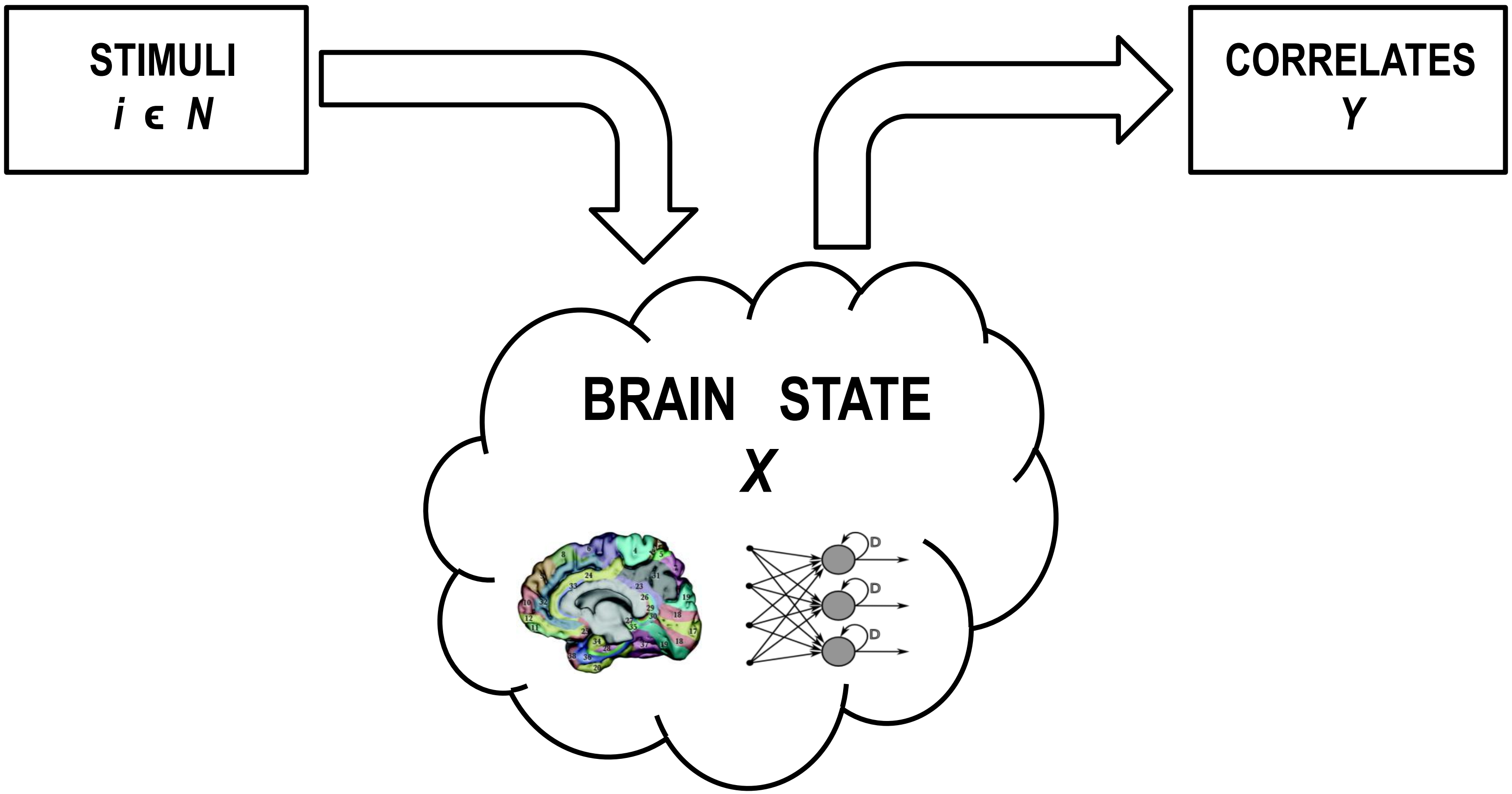
* learning and memory.

* collective behavior.

2) Solution is not continuous w.r.t. data (does not map to a topological space).

* high-dimensional fitness landscape?

Classic Empirical Example of a "Hard-to-Define" Event

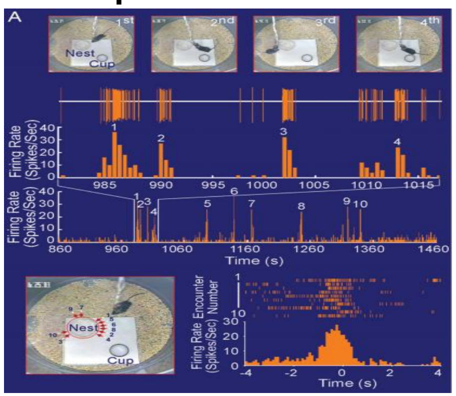


Classic Empirical Example of a “Hard-to-Define” Event

STIMULI
$i \in N$

CORRELATES
Y

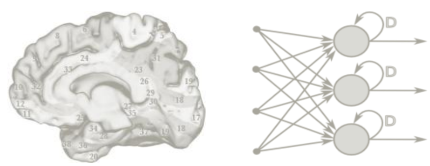
Concept of a Nest:



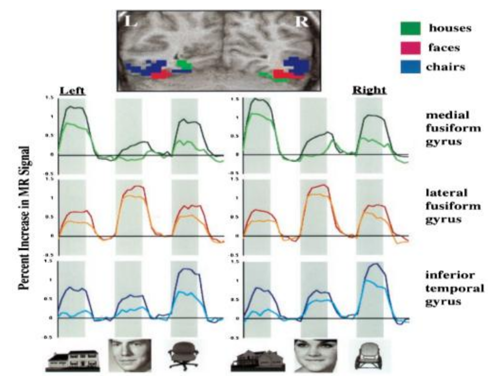
Lin et.al Neural encoding of the concept of nest in the mouse brain. PNAS, 104(14), 6066-6071 (2007).

Figure 1.A: transient “on” cells.

BRAIN STATE
X



Distributed representations:



Ishai et.al Distributed representation of objects in the human ventral visual pathway. PNAS, 96, 9379-9384 (1999).

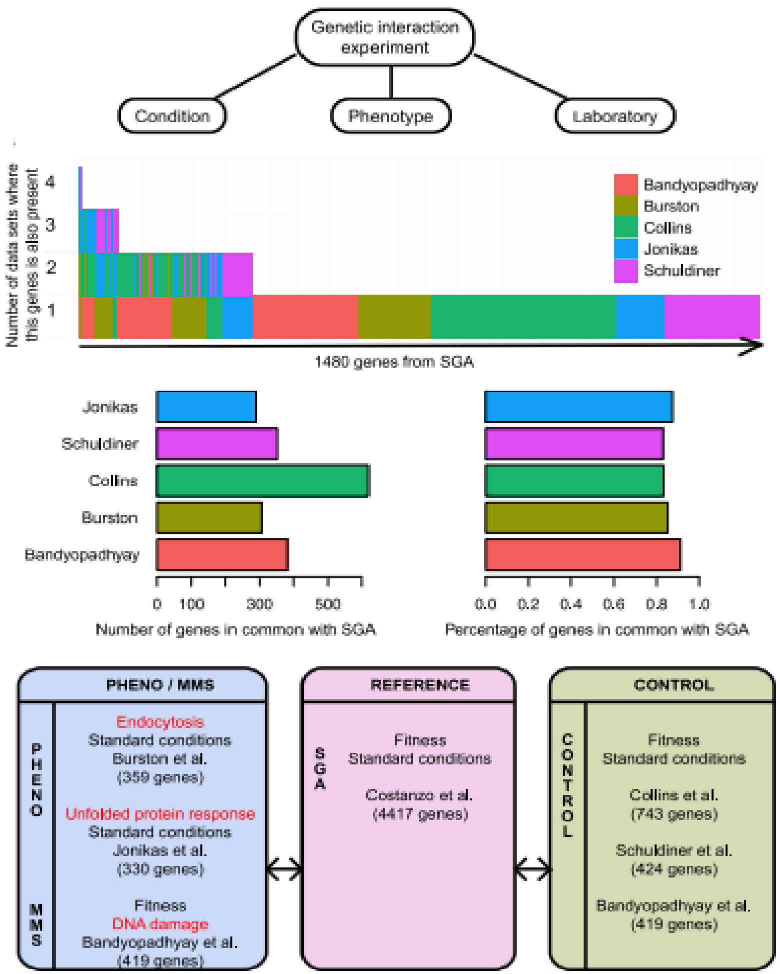
Figure 1.

What makes this “hard-to-define”?

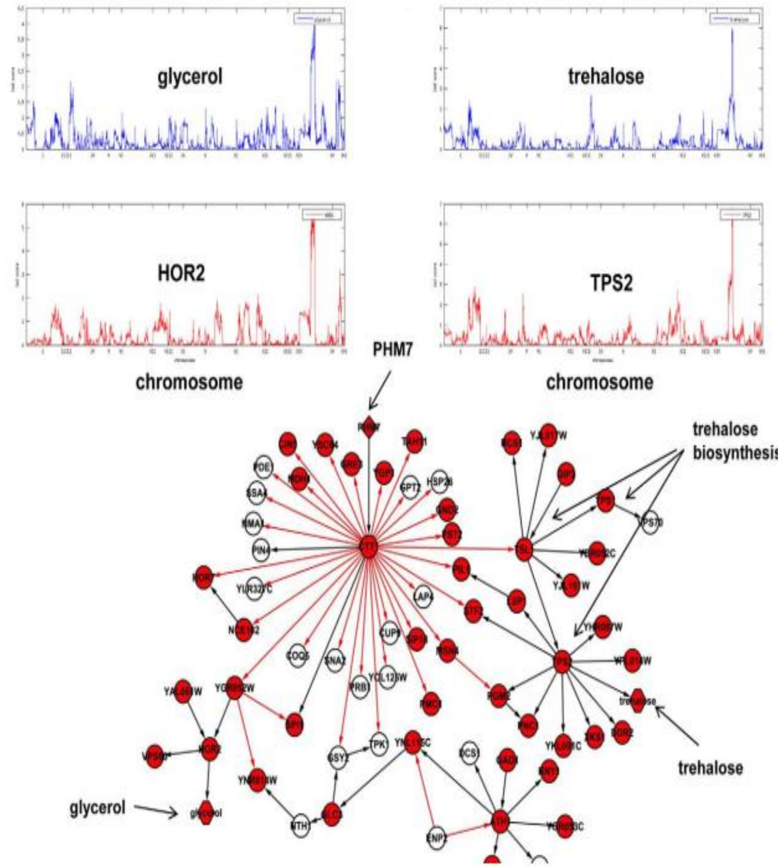
* lack of appropriate measures, analytical techniques?

* lack of context, understanding w.r.t. what results mean (synthesis)?

Does more data get us closer to an objective set of variables (empirically-speaking)?



COURTESY: Figure 7, PLoS Biology, 10(4), E1001301 (2012).



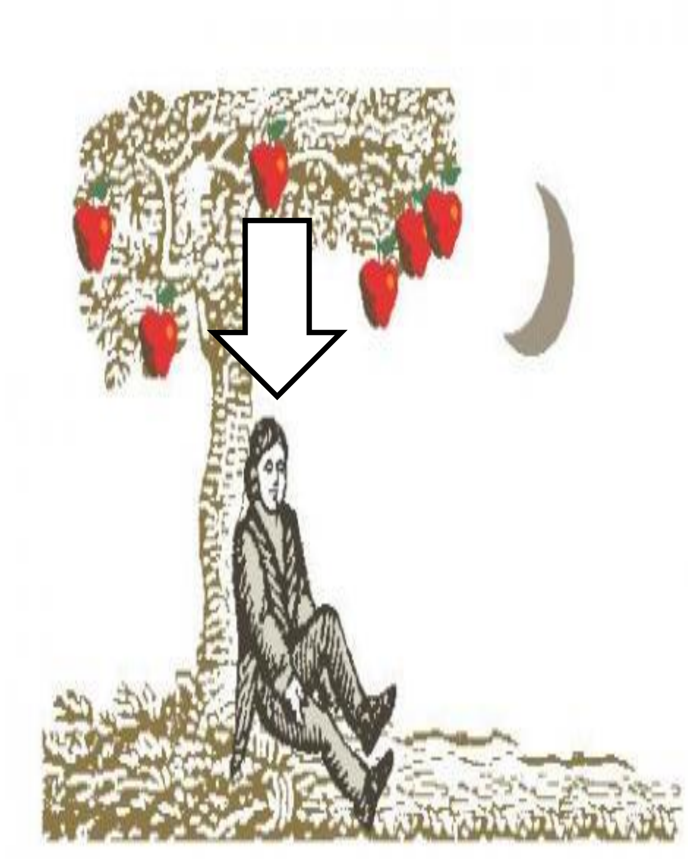
COURTESY: Figure 1, PLoS Computational Biology, 8(6), e1002559.

LEFT: Merging multiple types, sources of data.

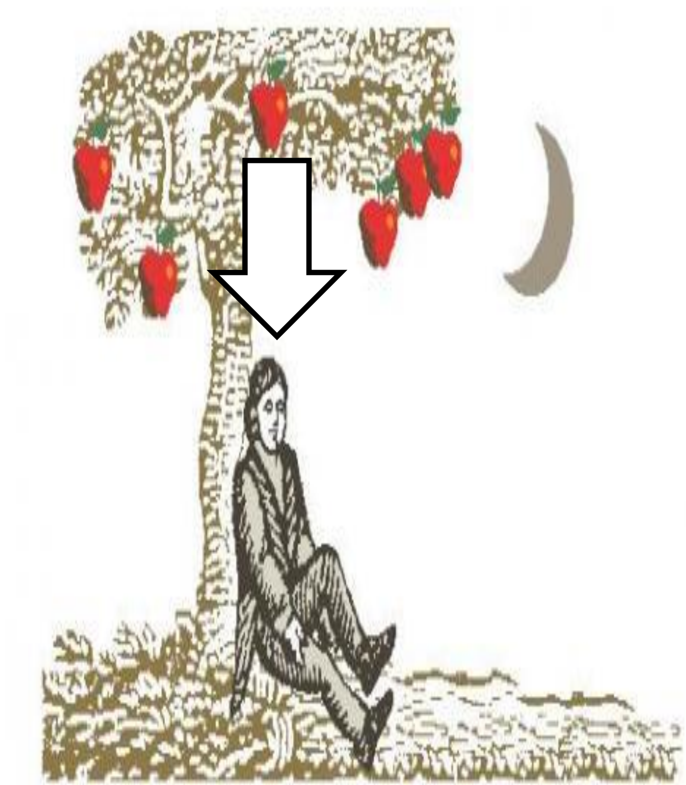
ABOVE: Complementary information (gene-gene interactions).

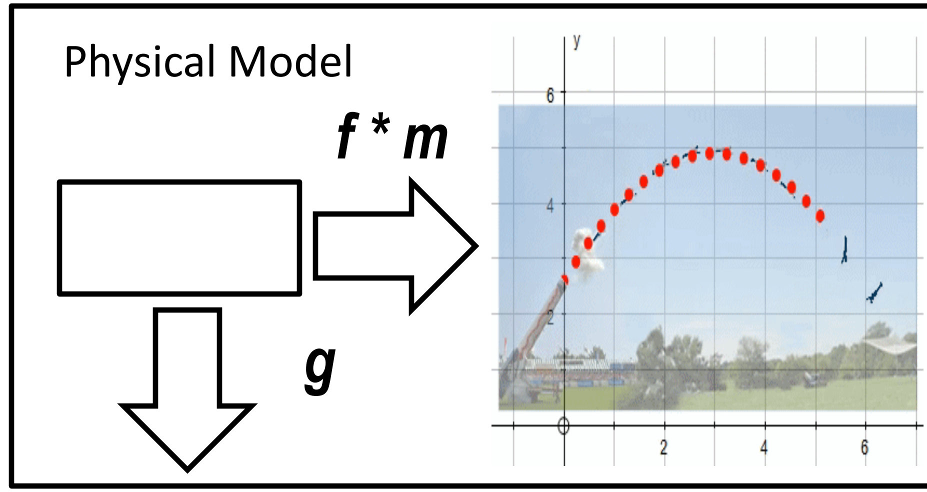
How “hard” is the problem?

Newtonian Physics (not so hard?)

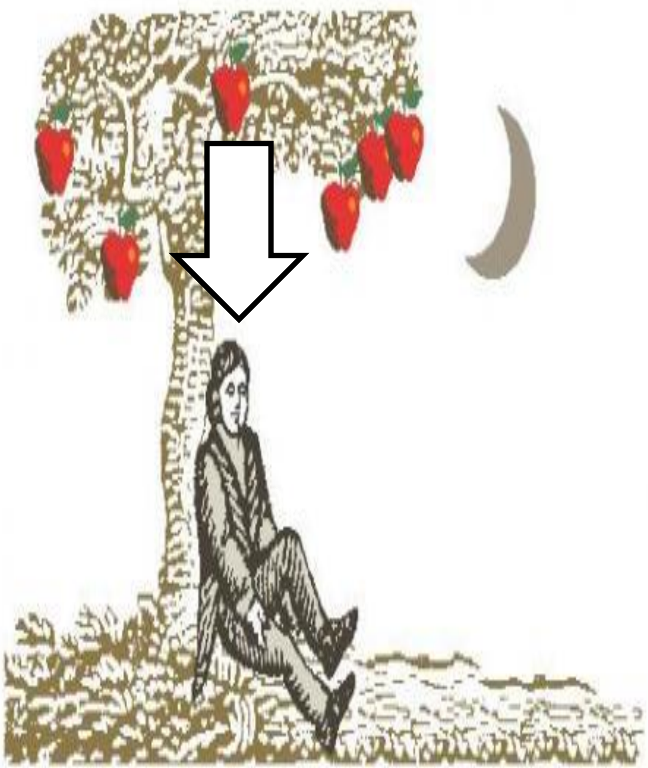


Newtonian Physics (not so hard?)





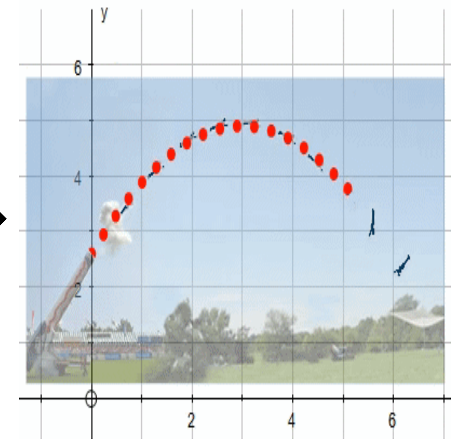
Newtonian Physics (not so hard?)



Physical Model

f * m

g



Mental Model

Experience of catching a ball

=

implicit knowledge + inference + prediction.

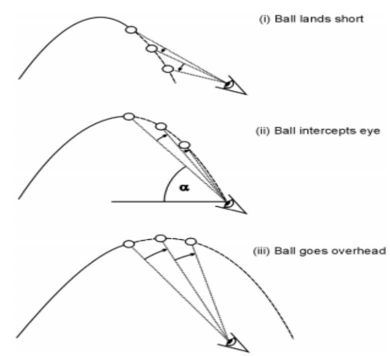
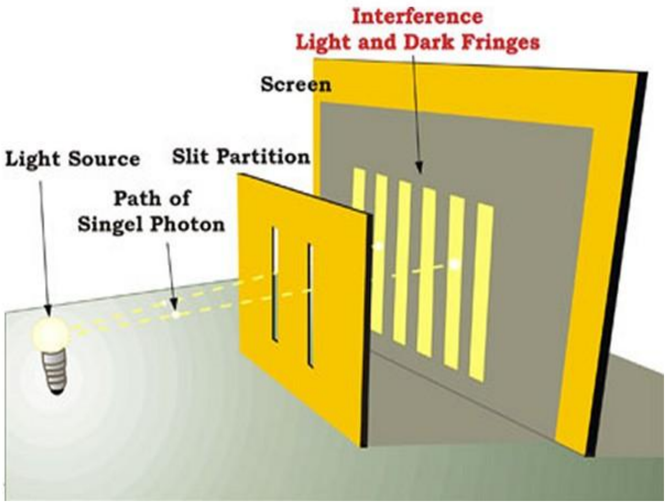
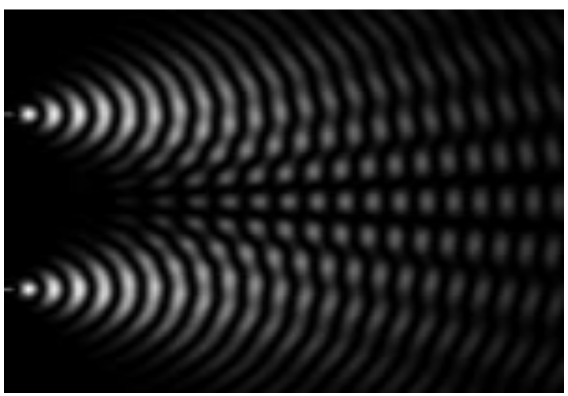


Figure 1: Consciousness and Cognition,19, 63-76 (2010).

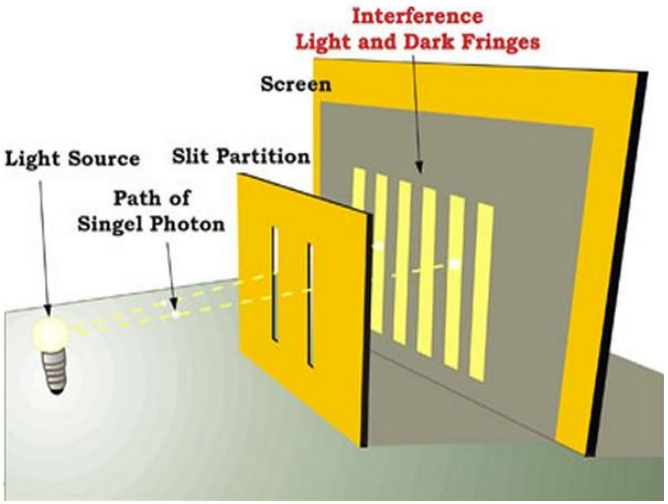
Quantum Physics (now that's hard!)



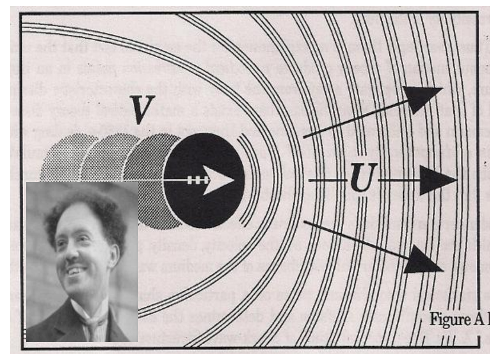


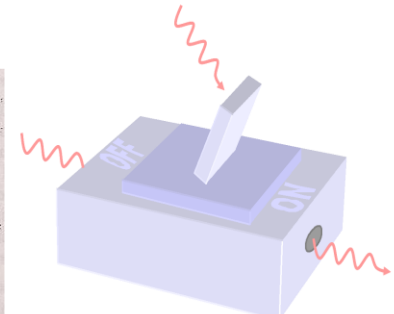
"Mechanics" = implicit knowledge + inference + prediction (in different amounts)

Quantum Physics (now that's hard!)

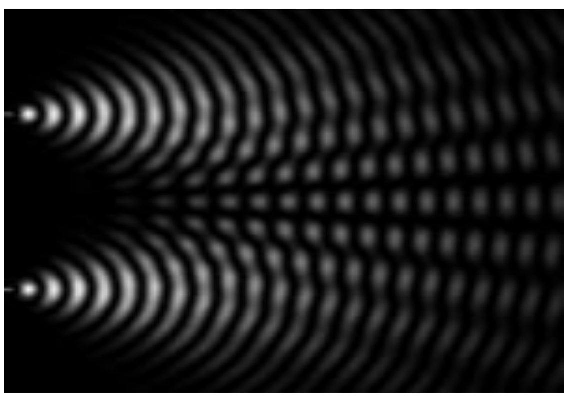


Physical Model



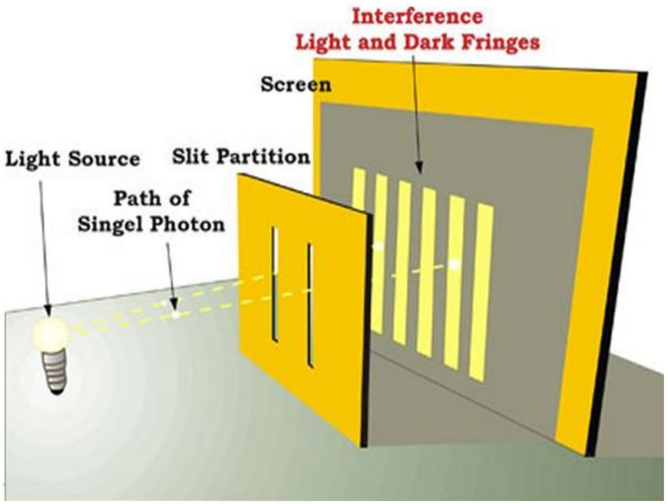


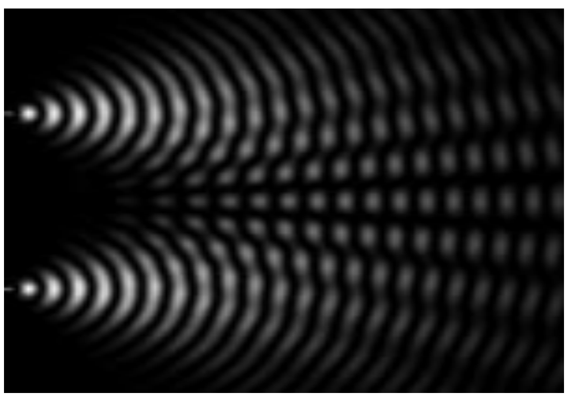
DeBroglie Model (left) and
Quantum gate (above)



"Mechanics" = implicit knowledge + inference + prediction (in different amounts)

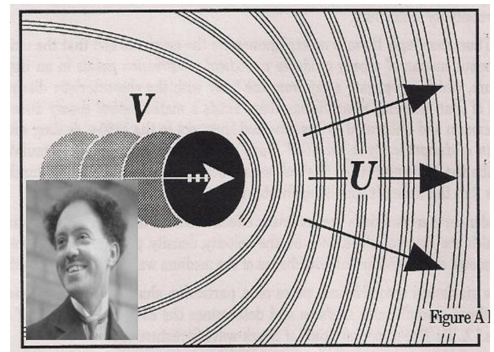
Quantum Physics (now that's hard!)

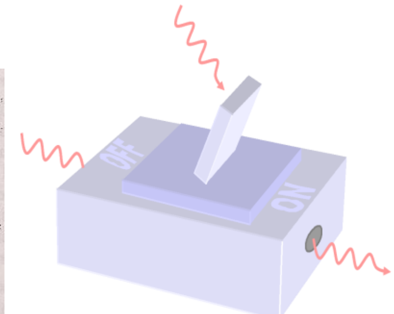




"Mechanics" = implicit knowledge + inference + prediction (in different amounts)

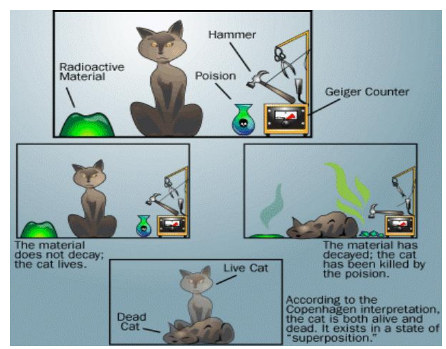
Physical Model

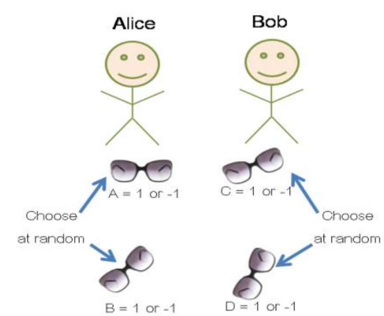




DeBroglie Model (left) and Quantum gate (above)

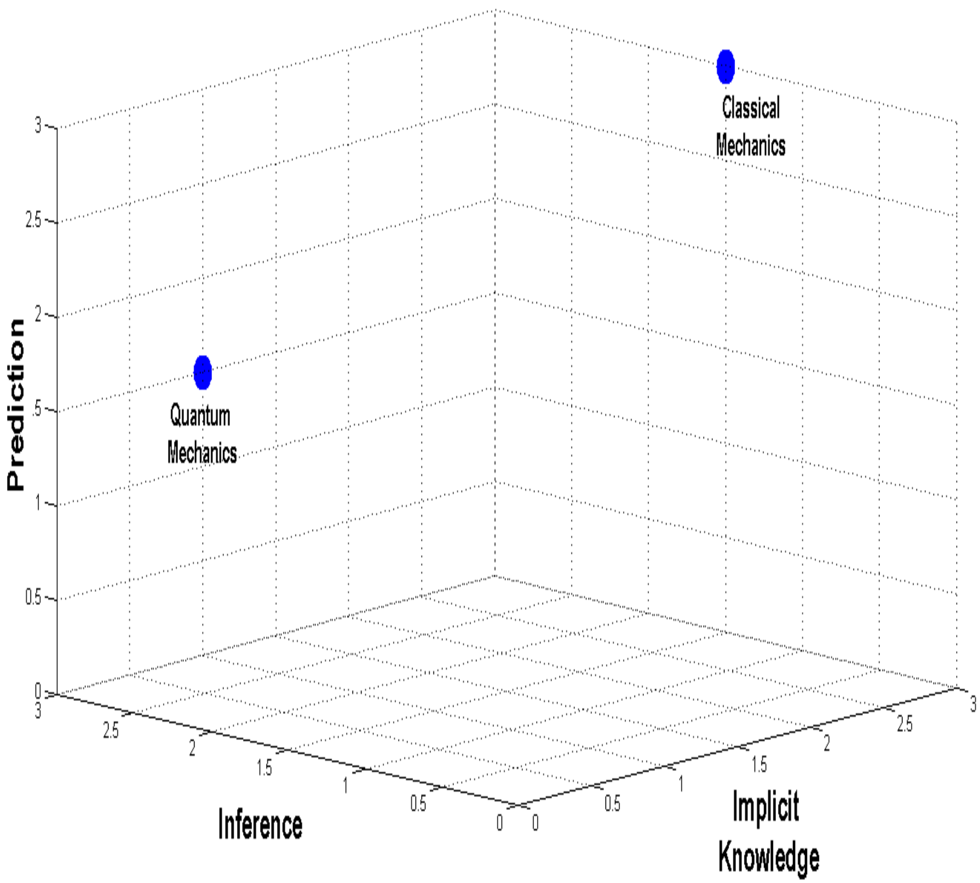
Mental Models





Schroedinger's Cat (left) and Quantum Information (above)

How do they compare?

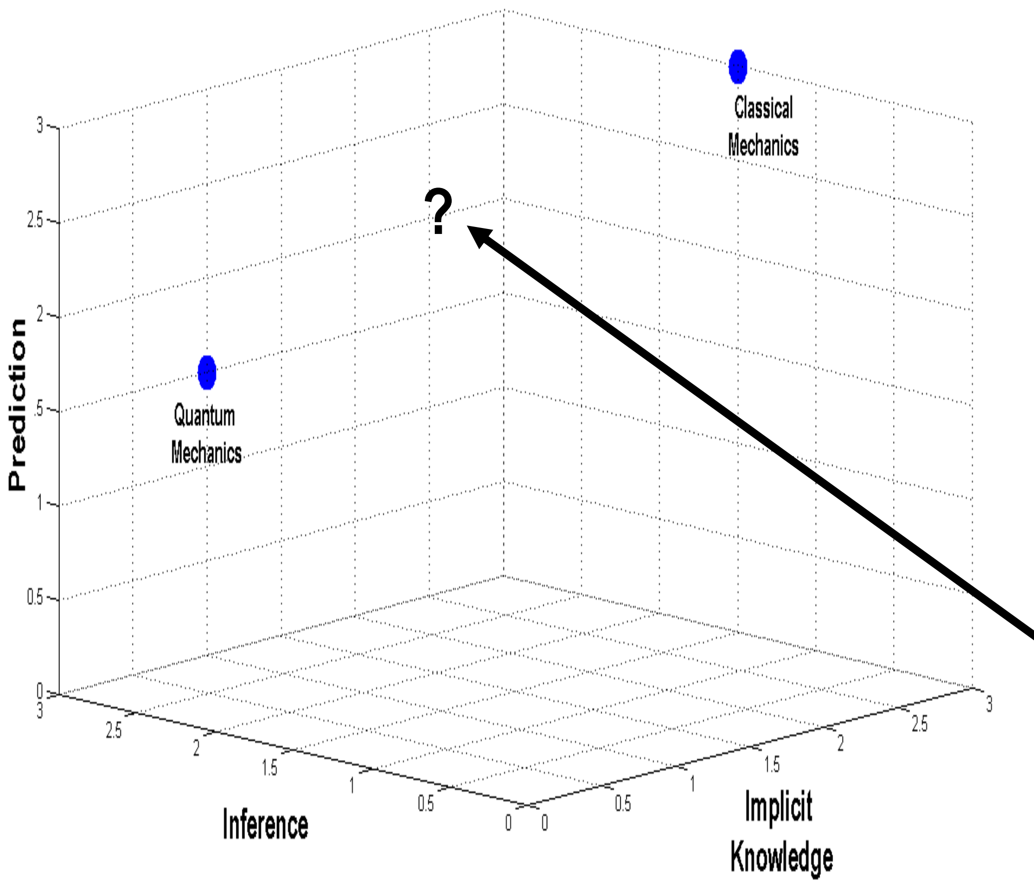


“Identity of Science” metric space.

What is “hard” about these two examples?

* definition of “variables” and “states”.

How do they compare?



"Identity of Science" metric space.

What is "hard" about these two examples?

* definition of "variables" and "states".

Now consider the problem of the origins of life……

* is this "harder" than quantum physics? Why?

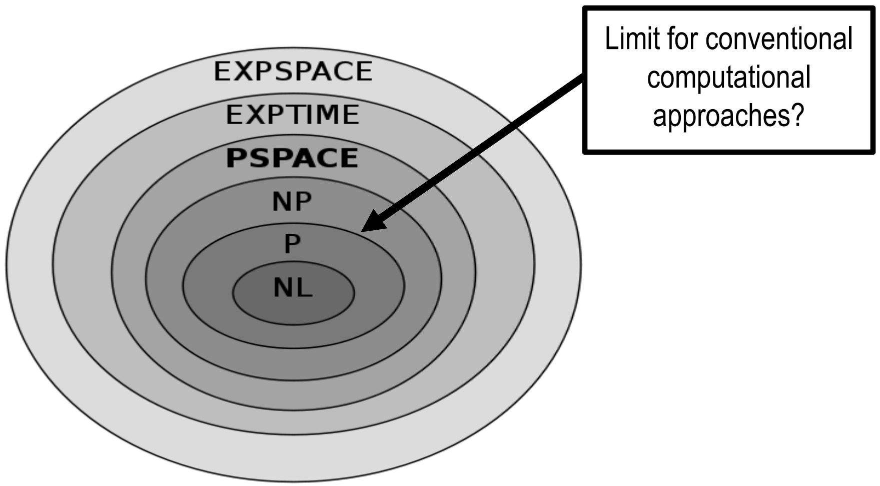
Can this approach be used to represent such domains computationally?

* provides them with a common currency.

A Perspective from Problem Hardness (Math and CS)….

When we claim that a problem is computationally “hard”, it is generally not solvable in polynomial time on a computer processor.

* claim related to problem structure, algorithm design, information about possible solutions.



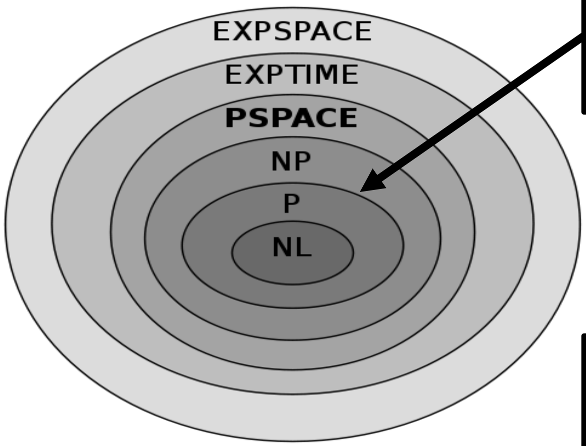
A Perspective from Problem Hardness (Math and CS)….

When we claim that a problem is computationally "hard", it is generally not solvable in polynomial time on a computer processor.

* claim related to problem structure, algorithm design, information about possible solutions.

Not a typical way to think about problem hardness in science

* consider if there were an optimal search time for good measures, solutions, and theories.



Limit for conventional computational approaches?

Where do scientific problems fit into this scheme?

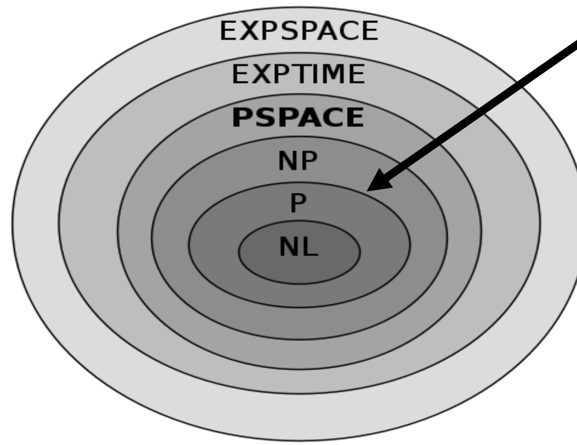
A Perspective from Problem Hardness (Math and CS)….

When we claim that a problem is computationally "hard", it is generally not solvable in polynomial time on a computer processor.

* claim related to problem structure, algorithm design, information about possible solutions.

Not a typical way to think about problem hardness in science

* consider if there were an optimal search time for good measures, solutions, and theories.



Limit for conventional computational approaches?

Where do scientific problems fit into this scheme?

Research in various fields can be convergent, but also divergent as well:

* consequence of people not talking between fields, or due to the "identity of science" space of hard scientific problems?

Does lead us into discussions of thinking of a qualitative concept ("soft" vs. "hard" science) algorithmically.

How do we “find” variables?

CASE STUDY: Cell and Developmental Biology

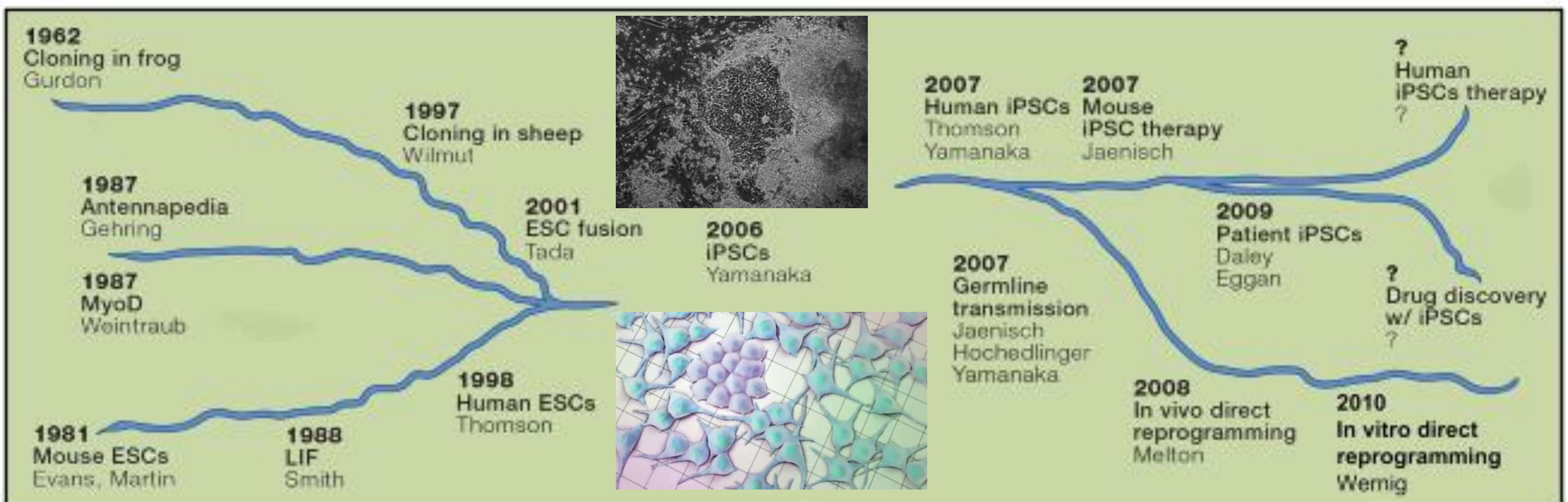
Convergence, divergence among research groups that study "pluripotency".

* scientific problems converge on a particular milestone, technology.

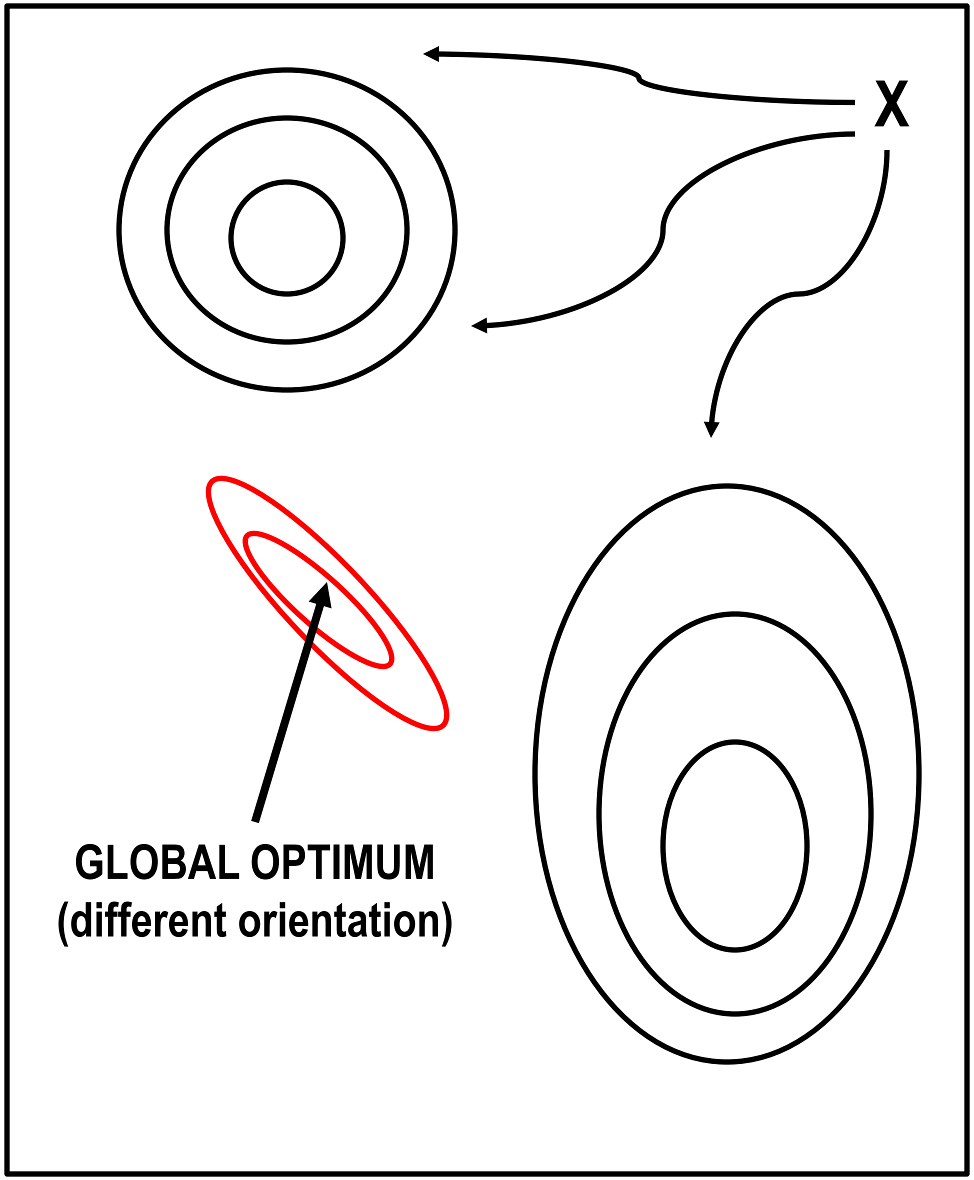
* fields coalesce and change, but major challenges do not.

Variables: applied from one context to another (e.g. cloning, cell culture → reprogramming).

* do variables need to change as questions, methods change?



Adapted from Figures 1 and 2, Yamanaka, Cell Stem Cell, 10, 679 (2012).



Two-dimensional non-convex surface

Finding Hidden Variables in a Dynamic System

* stable optima that exist in phase space.

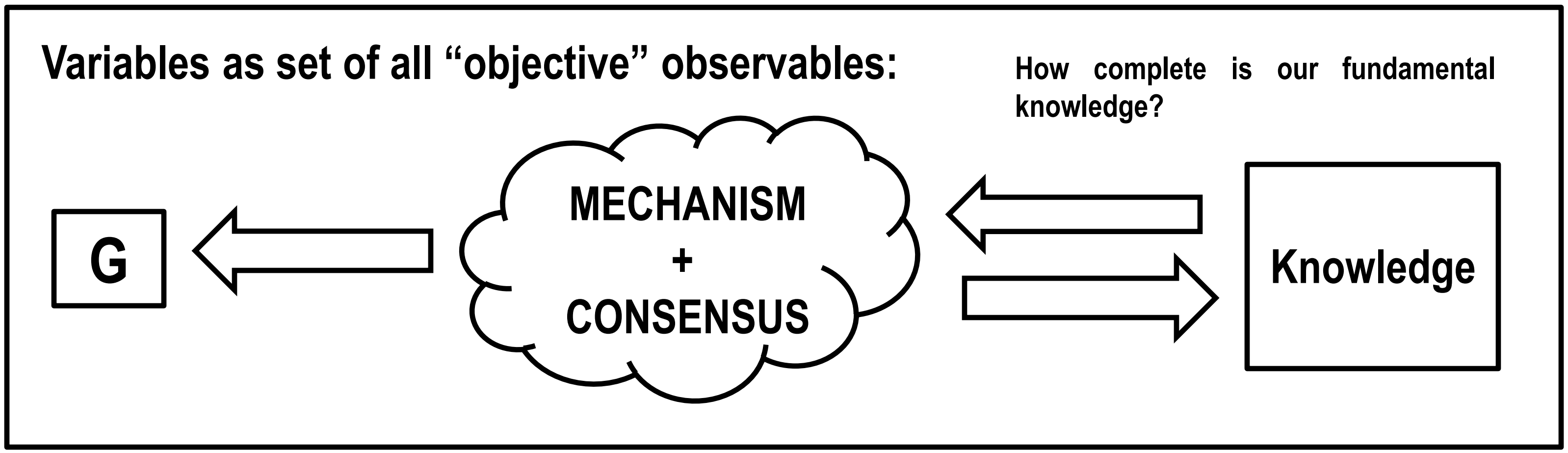
* local optima (black circles) are larger than smaller global optimum (red circles).

* harder to find using a search algorithm).

* hard-to-define variables are related to the non-convexity of this space.

Two approaches to variable construction, problem definition (objectivity *vs.* enculturation)

How do you “create” a set of complex variables? Example using phenomenon *G*.

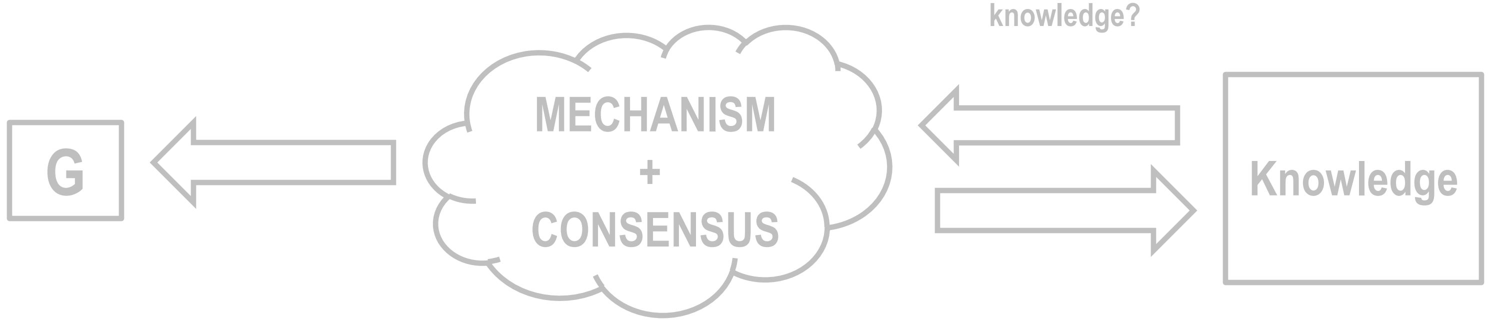


Two approaches to variable construction, problem definition (objectivity *vs.* enculturation)

How do you “create” a set of complex variables? Example using phenomenon *G*.

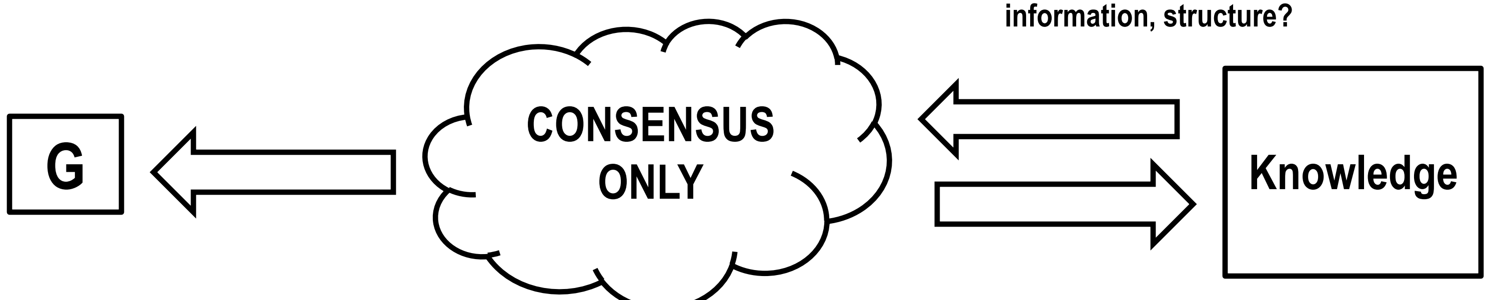
Variables as set of all “objective” observables:

How complete is our fundamental knowledge?



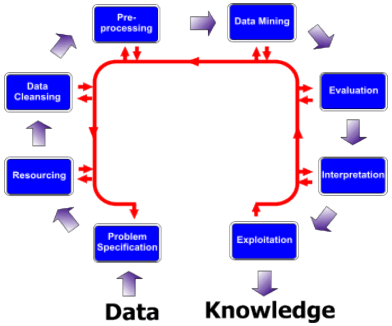
Variables as set of all things we learn are important:

What have we learned to be valuable? What can be learn from "hidden" information, structure?



Potential Solutions

Josef Zurada (University of Louisville): KDD approach

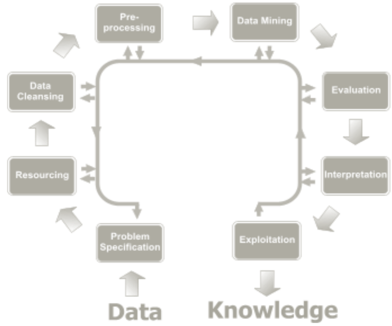


CASE STUDY: Rare Event Detection

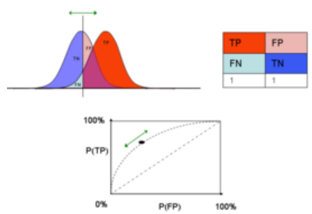
JZ) unsupervised learning (clustering, anomaly/outlier detection).

JZ) association rules (rule confidence/support).

Josef Zurada (University of Louisville): KDD approach



Arindam Banerjee (University of Minnesota): ROC Curve approach



CASE STUDY: Rare Event Detection

JZ) unsupervised learning (clustering, anomaly/outlier detection).

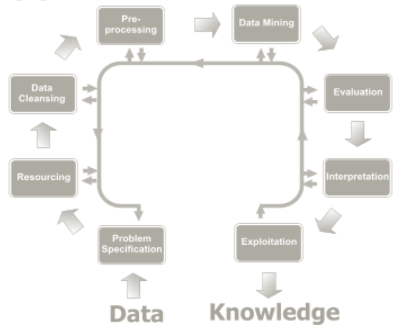
JZ) association rules (rule confidence/support).

AB) can be objects, or unexpected bursts of activity.

AB) analogous to a “needle in a haystack”.

CASE STUDY: Rare Event Detection

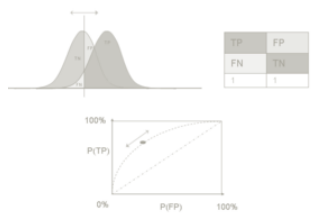
Josef Zurada (University of Louisville): KDD approach



JZ) unsupervised learning (clustering, anomaly/outlier detection).

JZ) association rules (rule confidence/support).

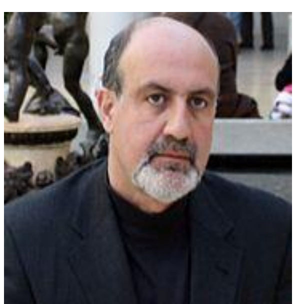
Arindam Banerjee (University of Minnesota): ROC Curve approach

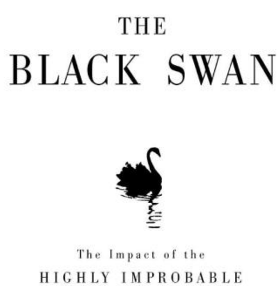


AB) can be objects, or unexpected bursts of activity.

AB) analogous to a “needle in a haystack”.

Nicholas Nassim Taleb: “Black Swan” approach





NNT) black swan = improbable, rare events that cannot be predicted.

NNT) “anti-fragility” – systems that benefit from random events, errors, and volatility.

CASE STUDY: Extreme Value Theory

How do you characterize long-tail (extreme) data, including those that have not been directly observed yet? For example, earthquakes, floods, a .500 batting average……….

* assume a probability distribution for the observed data, then model tail for subsequent data points. Extreme values will occur at some infinitesimal rate.

* what if data do not conform well to a known probability distribution? What if extreme events has different effects on your system (robustness)?

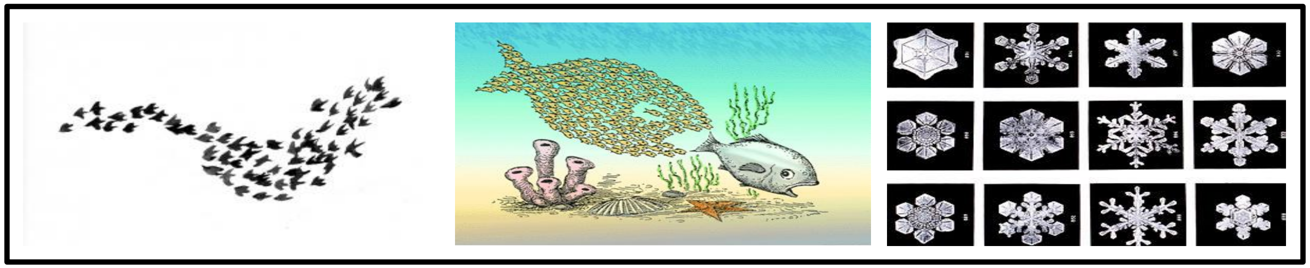
Are extreme values predictable? What about processes?

EVOLUTION: hopeful monsters, rare variants, complex traits?

EVOLUTIONARY PROCESSES: adaptive radiations, speciation events, survival of the fittest?

CASE STUDY: Emergence: an elusive phenomenon?

Emergence: whole is greater than the sum of its parts, out of this we can get highly-ordered systems.



What is it that we need to define?

* reductionist viewpoint: determine causality among pairwise relationships, establish "simplest" unit of action.

* complexity viewpoint: characterize interactions, higher-level patterns in phase space.

Key components: 1) multiplicative interactions between agents, 2) no exact solution.

* not deterministic but not spontaneous, either. What we lose in control, we gain in "life-like" behavior.

Evolutionary Solution for Solving "Hard" Problems?

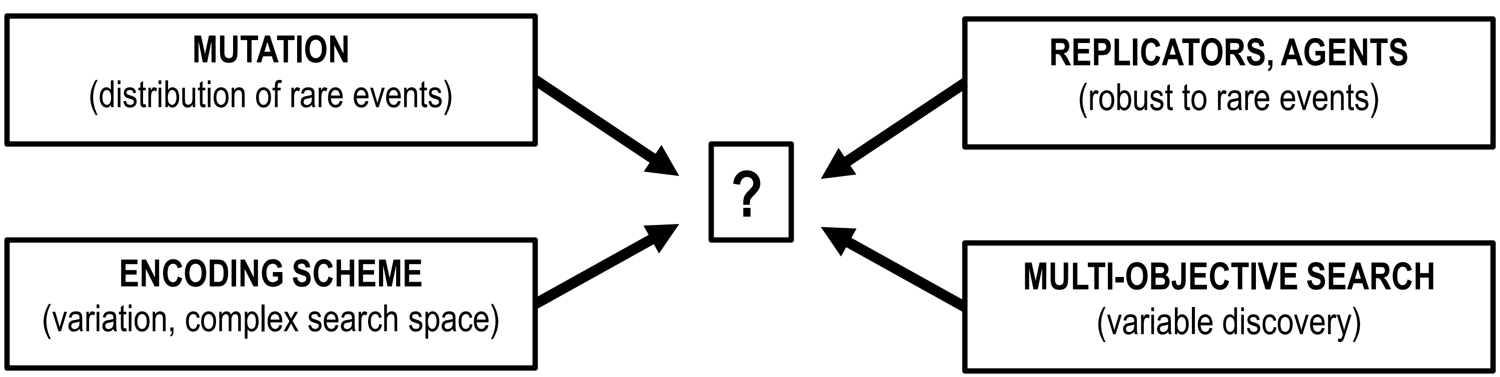
Using an algorithmic approach to "hard-to-define" problems:

> * what would a GA look like that deals effectively with the following problems? What would fitness function look like?

1) rare events, low-frequency occurences (for practical purposes, unpredictable).

2) searching for variables that provides more explanatory power (*vs*. relying on tradition, intuition).

3) nonlinear, non-deterministic problem spaces (where multiplicative interactions are common, no exact solution exists).



PARTICIPANTS (in no particular order):

Laura Grabowski (University of Texas-Pan American):
Toward Robotic Intelligence: Evolution of Memory Use in Digital Organisms

Bill Punch (Institute for Cyber-enabled Research, Michigan State):
Parallel Processing and Why it Matters to Everyone

Nicholas Keeney (Oceanography and Coastal Sciences, Louisiana State):
Drawing Conclusions from Drunk Fish in Dynamic Environments

Bradly Alicea (Cellular Reprogramming Laboratory, Michigan State):
Multiscale and Rare Events in Physiology

OTHER CONTRIBUTORS (in no particular order):

Michael Levin (Center for Developmental and Regenerative Medicine, Tufts):
Identifying Hard-to-Define Problems in Regenerative Biology

Anne Buchanan (Department of Anthropology, Penn State):
Rare and Hard-to-predict Events in Human Genetics and Disease