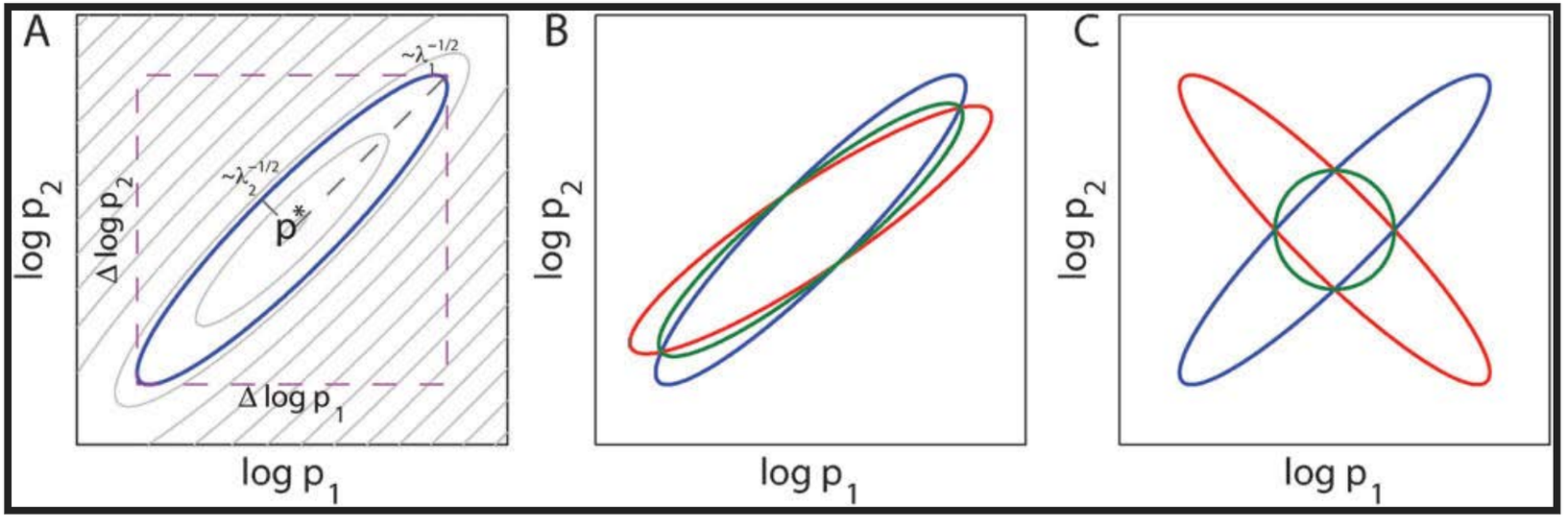
OPTIMAL EXPERIMENTAL DESIGN

PROBLEM STATEMENT

- Models are often practically unidentifiable.
- The data carry little information about the parameters.
- But the parameters are important!
  - At least we think they are. That's why we put them in the model.
  - Sloppiness?
- Can we find new data (new experiments) that carry more information about parameters?

COMPLEMENTARY EXPERIMENTS



- Models are often sloppy.
- Different data can carry different information.
- Predictions without parameters occur when the predicted experiments carry the same information as the existing data.
- If different experiments carry *complementary* information, perhaps we can learn all the parameters.

Apgar, Joshua F., et al. "Sloppy models, parameter uncertainty, and the role of experimental design." Molecular BioSystems 6.10 (2010): 1890-1900.

OED GENERAL STRATEGY (D-OPTIMAL)

1. Choose an ensemble of potential experiments.
2. Calculate the FIM for combinations of these experiments.
3. Choose the subset of experiments that *maximize* $|\mathcal{I}|$.
   (Maximizes the determinant.)

- Long history of this practice
- Lots of variations
- Frequentist/Bayesian Methods
- Nonlocal methods (don't use FIM)

PREDICTIONS VS. PARAMETERS

- One possible variation
- Minimize the uncertainty in a particular prediction, rather than in parameters.
- Does not require as many new experiments.
- Get new predictions, often without constraining any parameters.

Casey, Fergal P., et al. "Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model." IET systems biology 1.3 (2007): 190-202.

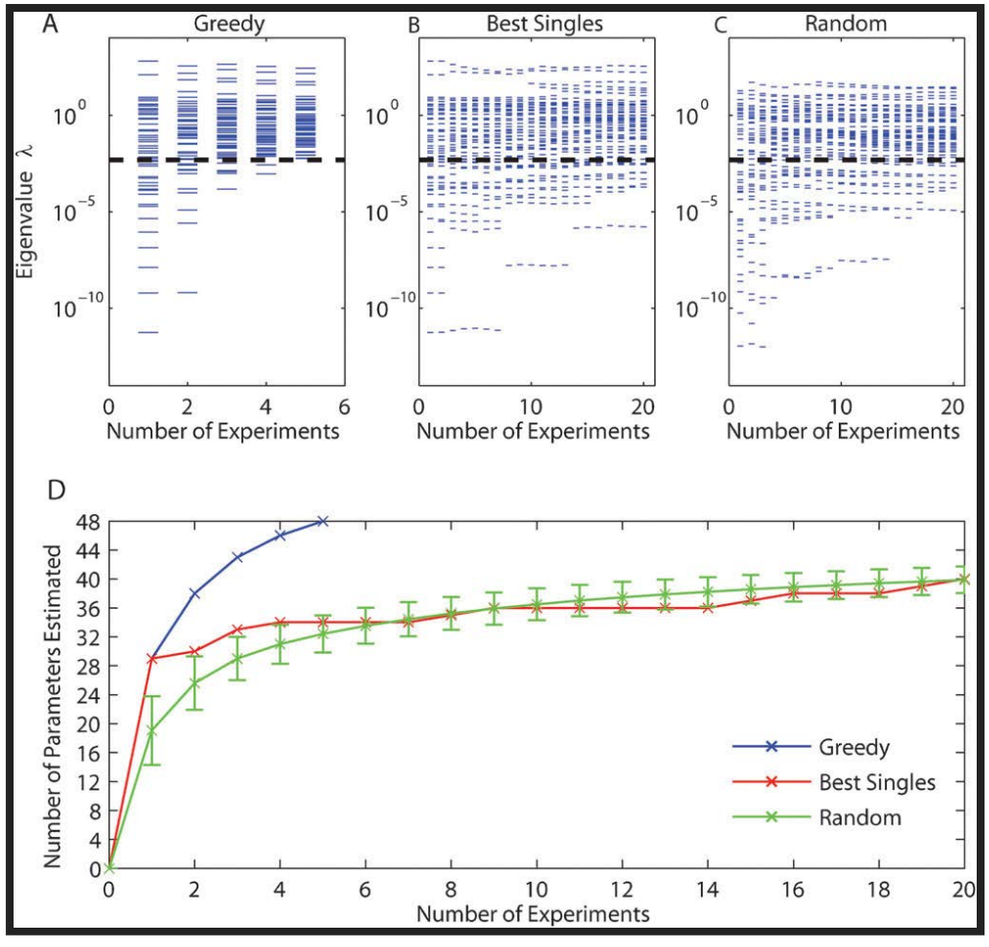
SLOPPINESS AND THE ROLE OF

EXPERIMENTAL DESIGN

- Apgar et al. use OED to estimate parameters in the signaling model of Brown et al.
- Considered a pool of 164,500 experiments
- Results:
  1. All parameters can be estimated to high accuracy.
  2. The optimal experiments are biochemically interesting.
  3. No fundamental limits to how accurately parameters can be estimated (with enough experimental effort)

Apgar, Joshua F., et al. "Sloppy models, parameter uncertainty, and the role of experimental design." Molecular BioSystems 6.10 (2010): 1890-1900.

ESTIMATING PARAMETERS OF BROWN ET AL.



- 5 Experiments could constrain parameters to 10%.
- These experiments had to be chosen carefully.
- Original data constrained parameters (best case) to within a factor of 50.

Apgar, Joshua F., et al. "Sloppy models, parameter uncertainty, and the role of experimental design." Molecular BioSystems 6.10 (2010): 1890-1900.

HOW MUCH DATA IS NECESSARY?

<u>Original Experiments</u>

- 68 Data points

<u>Optimal Experiments</u>

- No error bars (assumed 20%)
- Only observed a few of the species in the network

- 5 Experiments
  - Measured all species
  - 100 time points
  - 10% errror bars
- 200 times as many *measurements*
- Eigenvalues span 5 orders of magnitude (still sloppy)

The three primary conclusions remain unchanged.

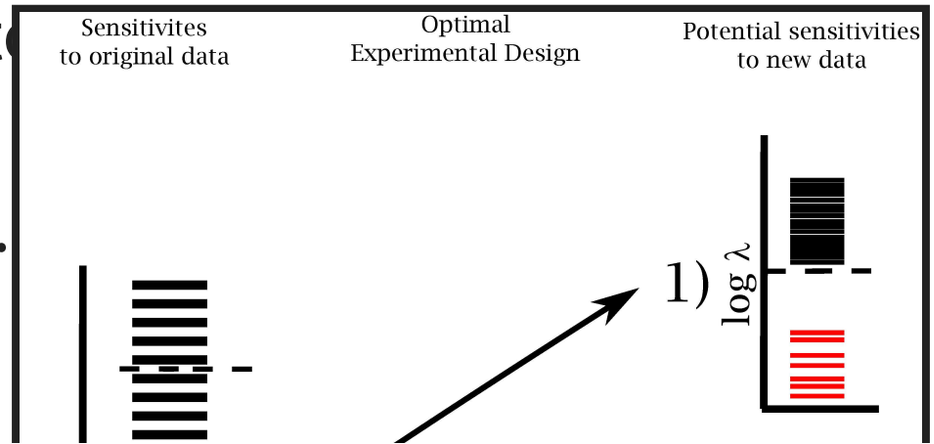
Chachra, Ricky, Mark K. Transtrum, and James P. Sethna. "Comment on “Sloppy models, parameter uncertainty, and the role of experimental design”." Molecular BioSystems 7.8 (2011): 2522-2522.

THE CAUSE AND CURE OF SLOPPINESS

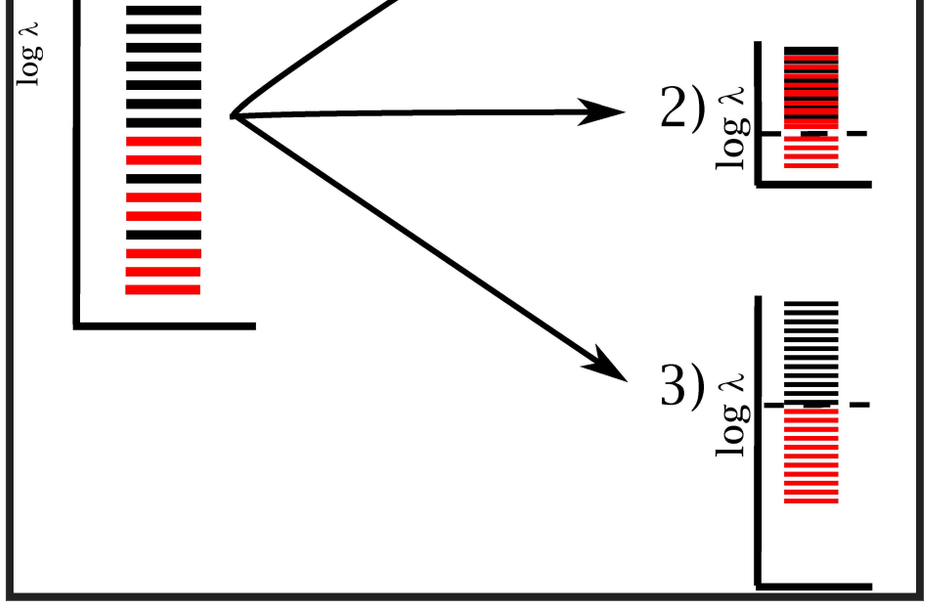
- Random Matrix Theory
- Cause of Sloppiness:
  - Systematic Correlations in the affect of parameters on predictions
- Cure of Sloppiness
  - Choose experiments that reduce horizontal structure in J.
  - Biologically motivated experiments
  - Generally more dense time points are not effective
- Possible to choose experiments t
  1. Minimize parameter errors.
  2. Minimize range of eigenvalues.

Tönsing, Christian, Jens Timmer, and Clemens Kreutz. "Cause and cure of sloppiness in ordinary differential equation models." Physical Review E 90.2 (2014): 023303.

THE LIMITATIONS OF OED



- "All models are wrong."
- Hypothesis: Predictions are possible *because* of sloppiness.
  - It is not necessary to model every little detail in the system.
  - As long your models in the same "universality class" as the physical system you will have a good "effective theory."



White, Andrew, et al. "The Limitations of Model-Based Experimental Design and Parameter Estimation in Sloppy Systems." PLoS Computational Biology, 12(12): e1005227 (2016).

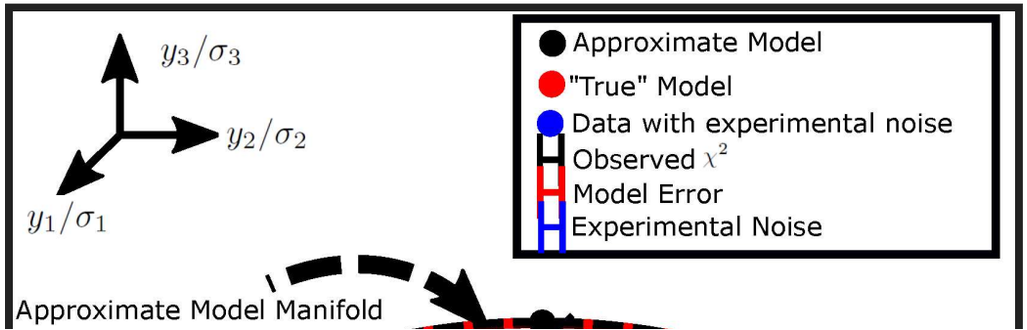
DNA REPAIR

- Modeling the dynamics of DNA after radiation.
- Dual dose experiments, asymmetric response
  - Intensity of each dose
  - Duration of each dose
  - Length of rest time
- Six parameter model
- Selected 19 new optimal experiments from a pool of 21,870
- Anticipated an accuracy of about 10% in parameter

estimates

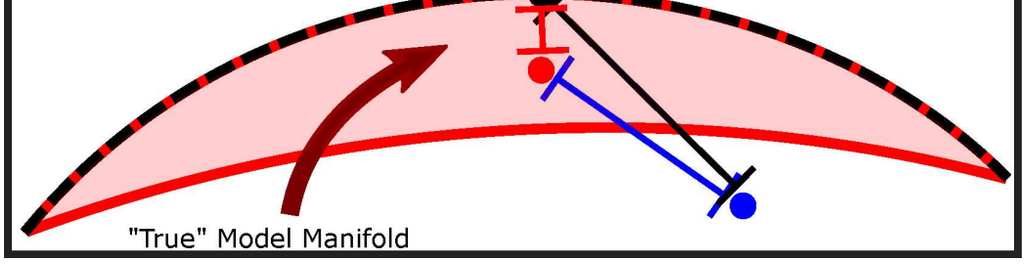
Result: Model could not give a good fit to all experiments simultaneously

UNCERTAINTY QUANTIFICATION



Modify the model assumptions:

- $d_i = y_i(\theta) + \sigma_i \xi_i + \delta_i$
- Measurements error: $\sigma_i \xi_i$
- Model error: δ_i



White, Andrew, et al. "The Limitations of Model-Based Experimental Design and Parameter Estimation in Sloppy Systems." PLoS Computational Biology, l 12(12): e1005227 (2016).

MODELING MODEL ERROR

Hyper-model ansatz: $\delta_i = f \sigma_i \xi_i'$

Estimate f by fitting to data:

$$\hat{f} = \sqrt{\frac{2C(\theta)}{M-N} - 1}$$

$$\delta f = \frac{1 + \hat{f}^2}{\hat{f}\sqrt{2(M-N)}}$$

If $\hat{f} \approx \delta f$, then the model error is small.

Modified estimator variance on original parameters:

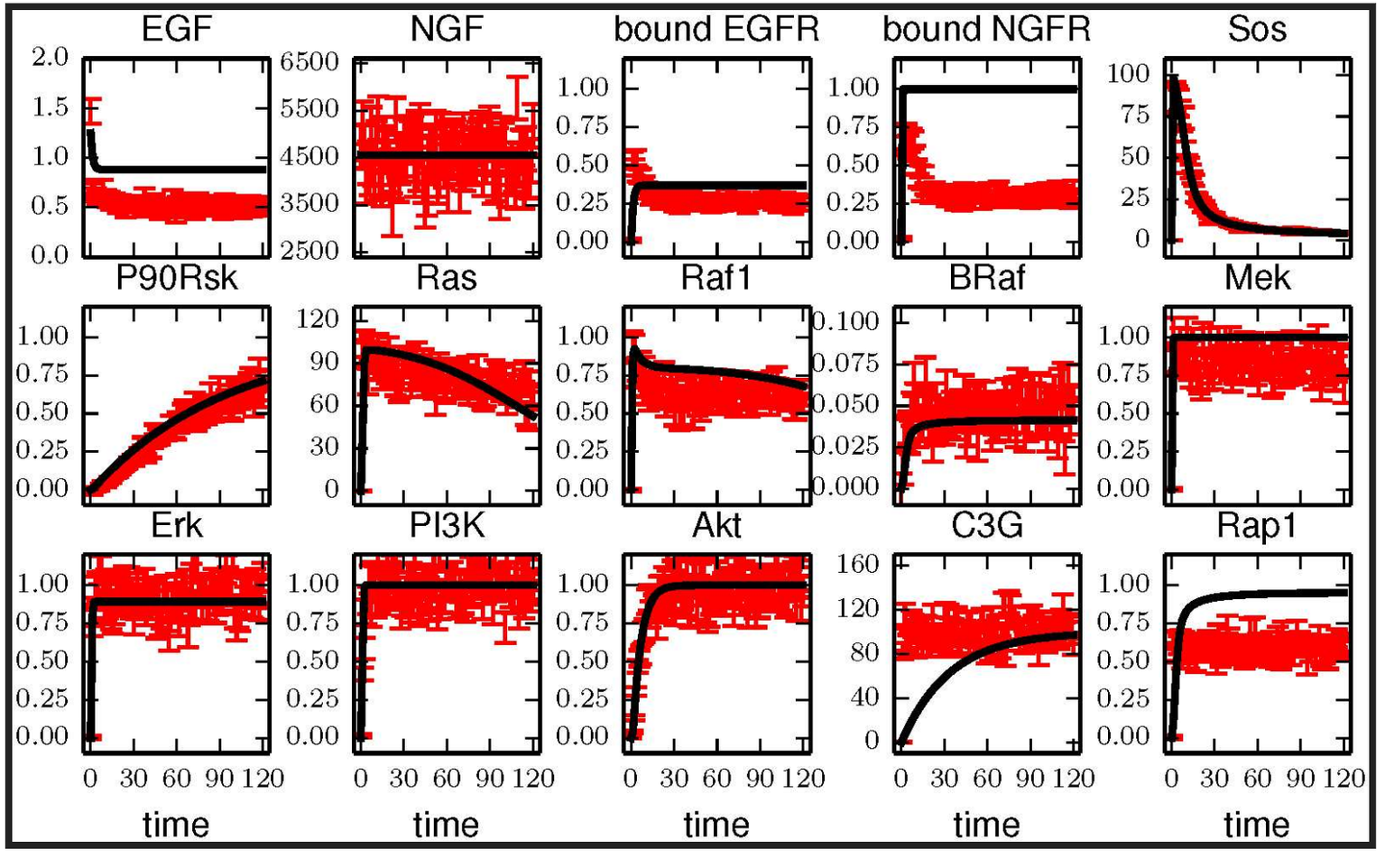
$$Cov(\theta) = \frac{2C(\theta)}{M - N}\mathcal{I}^{-1}$$

EGFR SIGNALING REVISITED

One approach to exploring model error: Use multiple models

1. Replace Brown model (Michaelis Menten) with more detailed reactions (Mass action)
2. Calibrate mass-action model to Brown's original experiments (Sloppy)
3. Create data to the 5 Apgar experiments using the mass-action model
4. Fit this data with original Brown (Michaelis-Menten) model

EGFR SIGNALING REVISITED



PARAMETERS WITHOUT PREDICTIONS

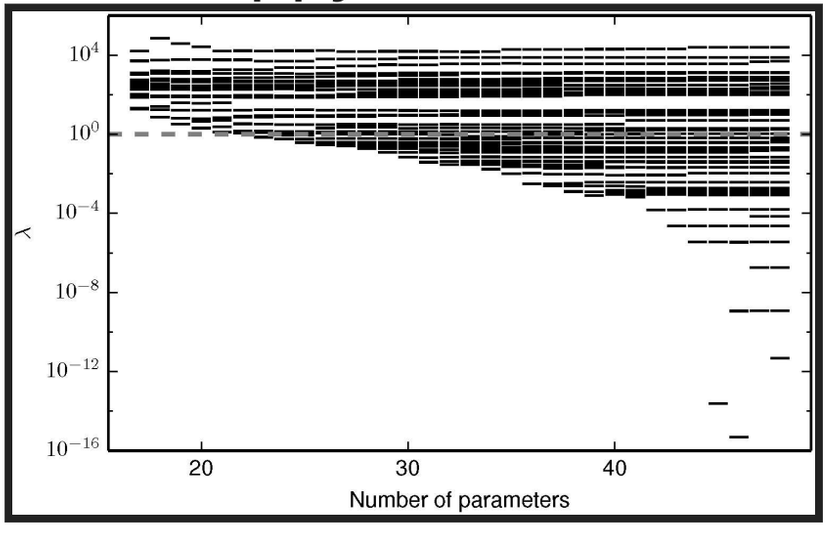
Hyper-model for Brown model fit to Apgar Experiments:

- $\hat{f} = 3.7, \delta f = 0.03$
- Parameter estimates: 10% $\rightarrow$ 40%.
  - Not Bad!
- Uncertainty in predictions?
  - Huge. Full range of physically allowed values.
  - Dominated by model error.

FUNDAMENTAL LIMITS TO PARAMETER ESTIMATION

Introduce a useful concept: *Sloppy System:*

A physical system and a set of experimental protocols that can be approximated by a hierarchy of mechanistic, mathematical models of growing complexity that become sloppy in the limit of microscopic accuracy



ESTIMATING MODEL ERROR IN SLOPPY SYSTEM

- Approximate the eigenvalues as a geometric series with ratio r
  .
- Let λ_0 be the smallest eigenvalue in the model
- Cost due to the model error: $\lambda_0 r/(1-r)$
- Expected Cost: $\frac{1}{2}(M-N) + \lambda_0 r/(1-r)$
- Requiring $\hat{f} < 1$ gives:

$$\lambda_0 < \frac{1-r}{r}(M-N)$$

Key Result: If a system is sloppy, then a model cannot be both predictive and have arbitrarily large eigenvalues.
There is a fundamental limit to the accuracy of estimated parameters.

OUTLOOK

REDUCTIONISM, MODELING, AND OED

- All models are wrong
- The model selection problem cannot be divorced from the model calibration problem.
- "Every good model starts from a question."
- OED assumes that it can ask as many other (non-target) questions as it needs to calibrate the model.
- If the model was not designed to be able to answer these additional questions, it may break.

RELEVANT VS. IRRELEVANT PARAMETERS

- There is no theory of complex systems.
- Given a scientific objective and a reasonably complete microscpic model (up to unknown parameters), there is no way to know what is the right simple model to use.
- Given a simple model, there is also no way to know which questions it can answer.
- There is no effective bridge between mechanism and phenomenology that is scalable.