A Topological Explication of Empirical Simplicity

{Kevin T. Kelly, Konstantin Genin}

Ockham's Razor

- "Presume no more complexity than necessary."

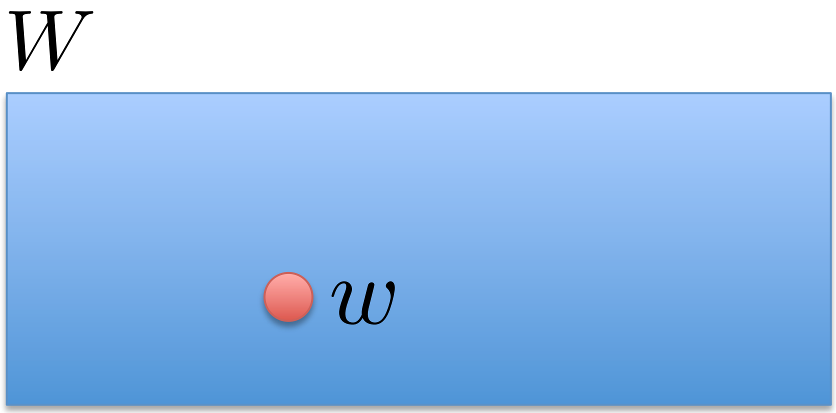
Three Fundamental Questions

1. What is **simplicity**?
2. What is **Ockham's razor**?
3. What is its **epistemic justification**?

1. INFORMATION TOPOLOGY

Worlds

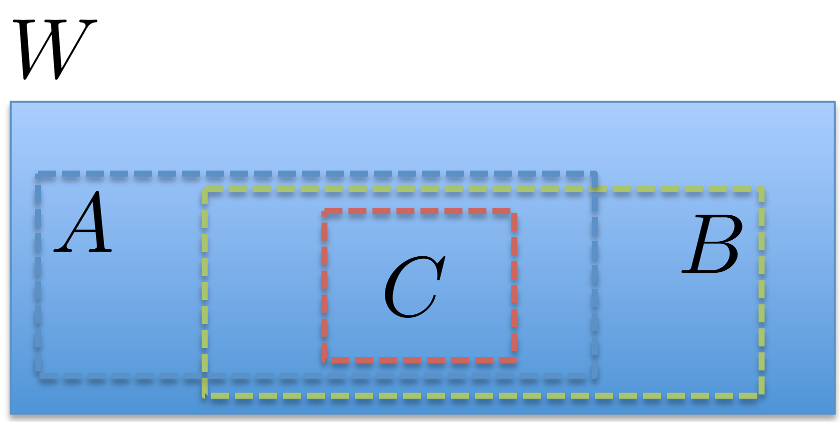
- The points in W are **possible worlds**.



The Structure of Information

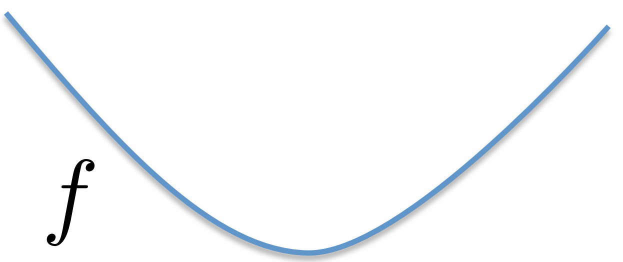
An **information basis** I is a countable set of information states such that :

1. each world makes some information state true;
2. each consistent pair of information states is entailed by a stronger information state.



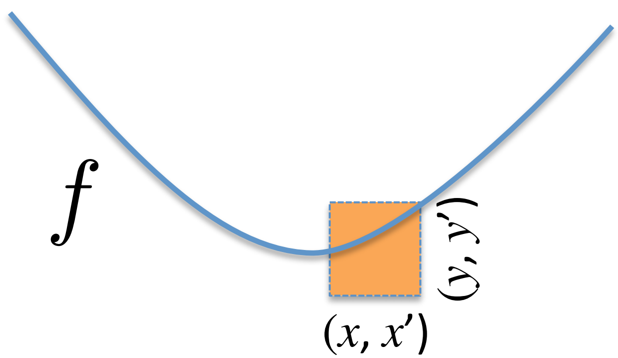
Example: Equations

- **Worlds** = functions $f : \mathbb{R} \to \mathbb{R}$.



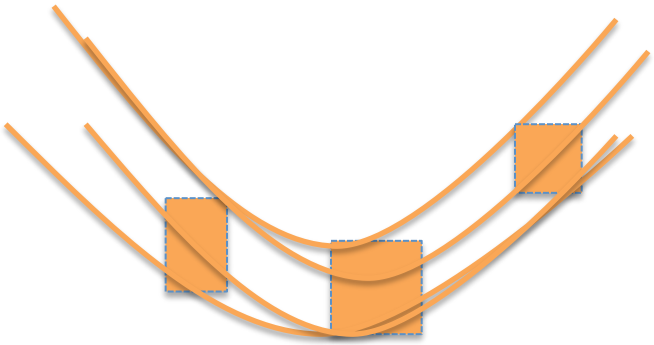
Example: Possible Laws

- An **observation** is a joint measurement.



Example: Possible Laws

- The **information state** is the set of all functions that touch each observation.

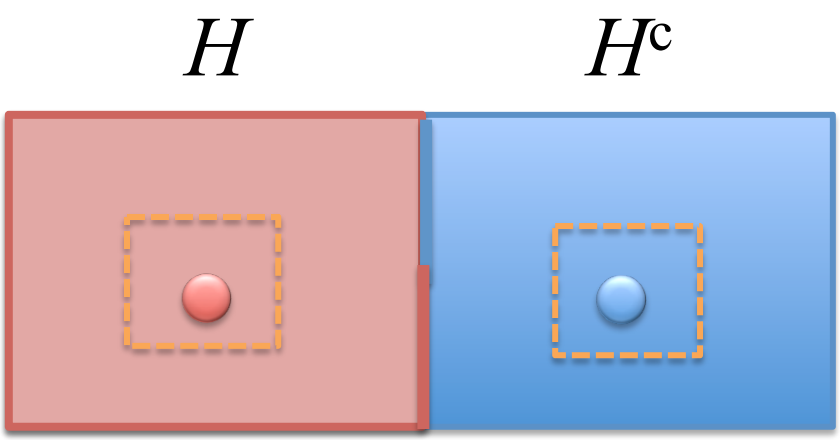


H will be Verified in w

w is an **interior [exterior] point** of H

iff H **will be** verified [refuted] in w

iff there is $E \in \mathcal{I}(w)$ s.t. H is verified [refuted] by E.

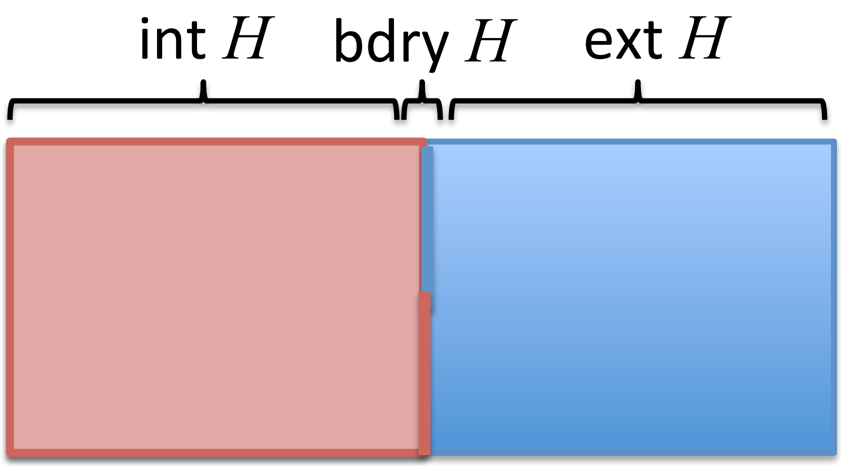


H will be Verified/Refuted

int H := the proposition that H **will be verified.**

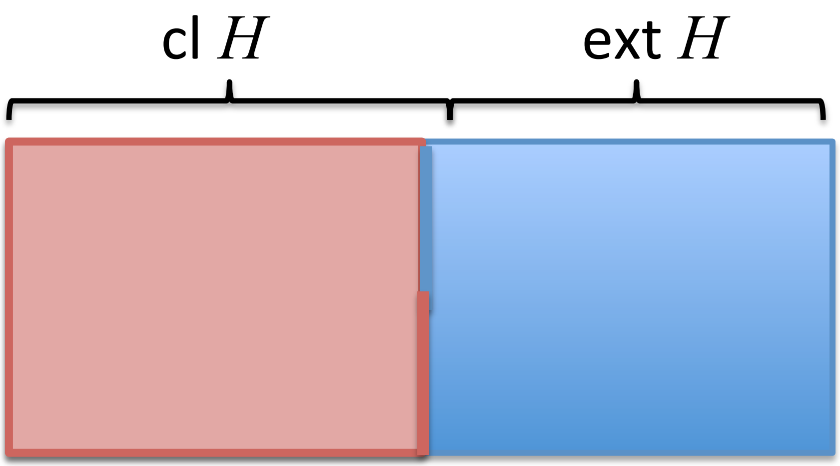
ext H := the proposition that H **will be refuted.**

bdry H := the proposition that H **will never be decided.**



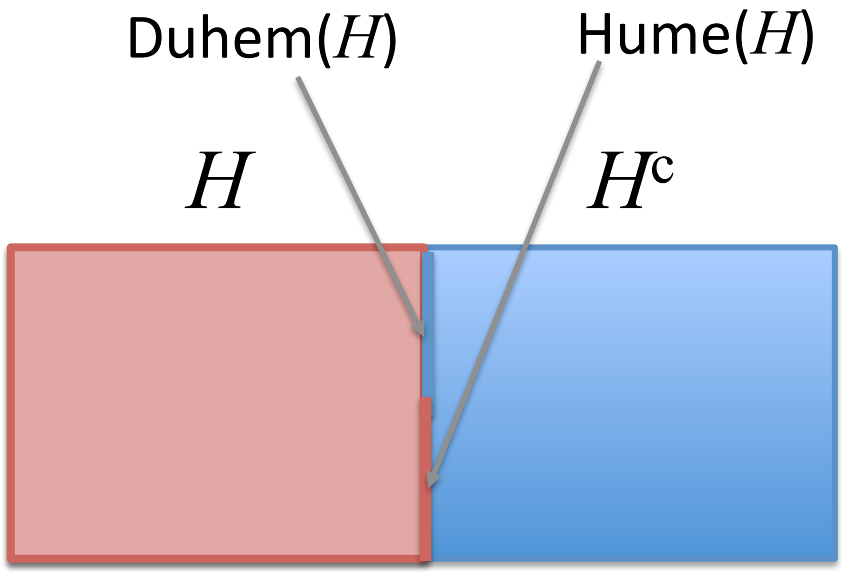
H will Never be Decided

cl H := the proposition that H **will never be refuted.**



Hume and Duhem

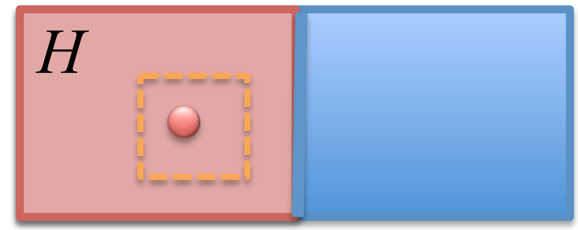
- bdry(H) ∩ H = "you face **Hume's problem** w.r.t. H";
- bdry(H) ∩ H^c = "you face **Duhem's problem** w.r.t. H"



Verifiability, Refutability, Decidability

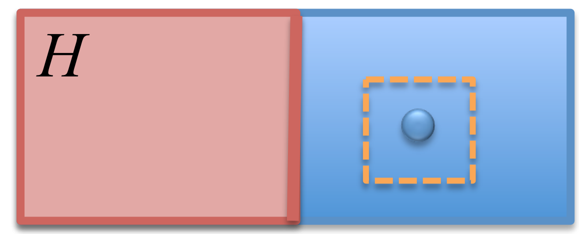
H is **verifiable (open)** iff $H \subseteq \mathrm{int}(H)$.

i.e., iff H will be verified however H is true.

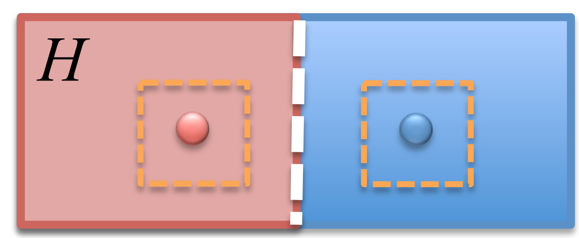


H is **refutable (closed)** iff $\mathrm{cl}(H) \subseteq H$.

i.e., iff H will be refuted however H is false.

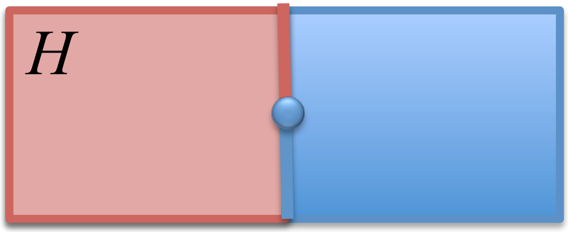


H is **decidable (clopoen)** iff H is both verifiable and refutable.



Veri-futability

H is **veri-futable (locally closed)** iff H will be verified to be refutable, however H is true.



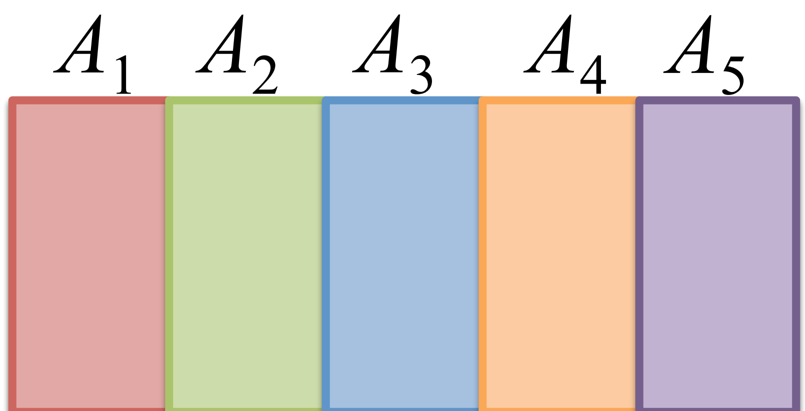
Scientific models.

- E.g., “linear”, “quadratic”.

2. INDUCTIVE METHODS

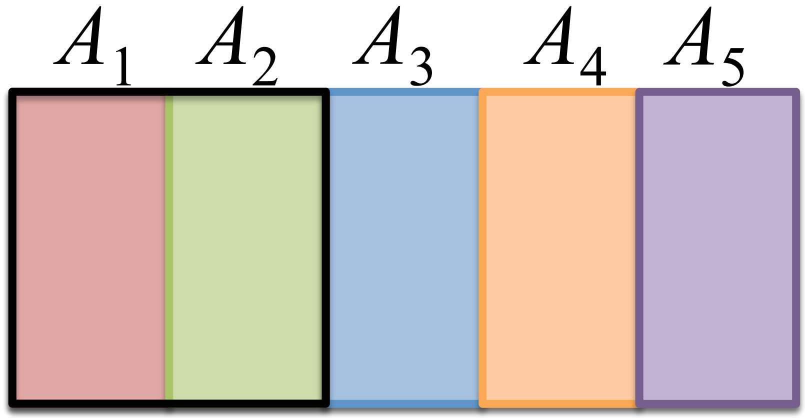
Questions

- A question Q partitions W into a countable set of possible answers.
- Inquiry seeks the true answer.



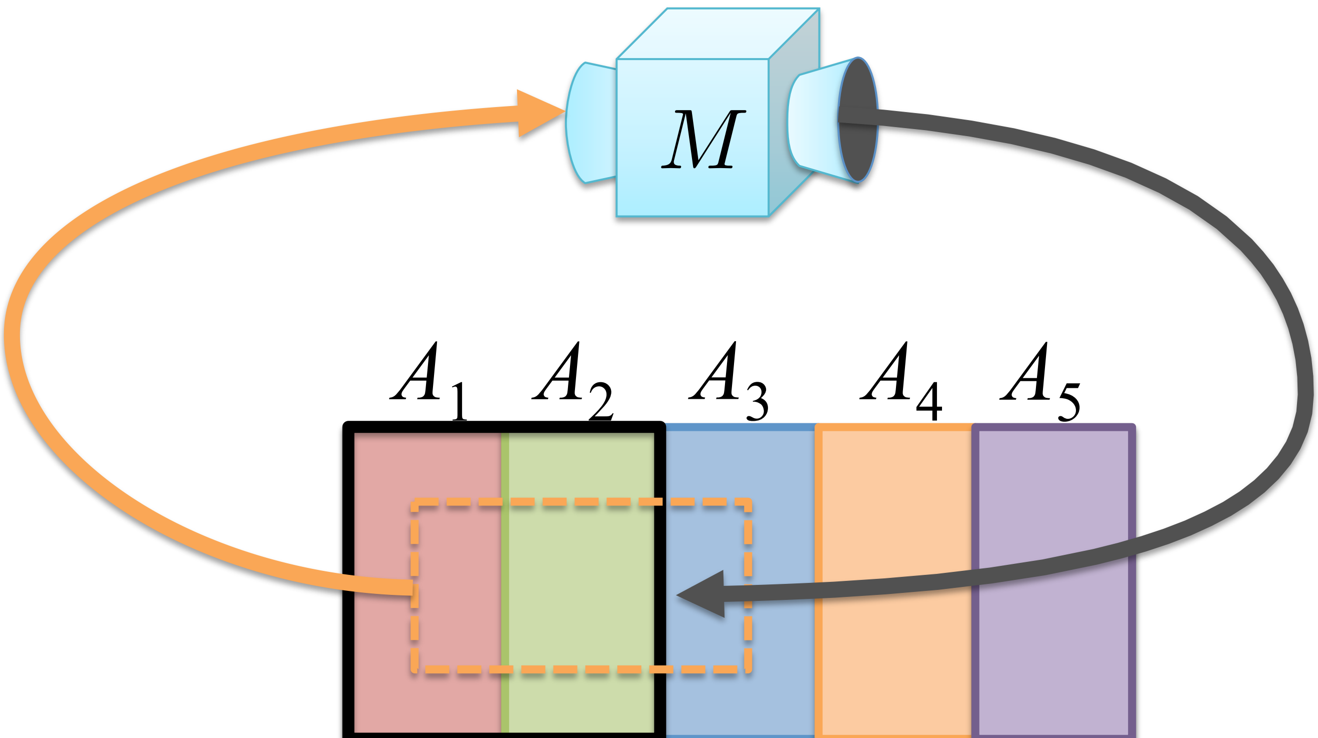
Relevant Response

- A disjunction of answers.



Inductive Methods

Information in, relevant response out.



Verification Methods

Verification method. In every world w:

$1.\ w \in H:$ M converges to H without error.

$2.\ w \in H^c:$ M always concludes W.

Refutation method. In every world w:

$1.\ w \in H:$ M always concludes W.

$2.\ w \in H^c:$ M converges to H^c without error.

Decision method. does both.

Fundamental Correspondence

Proposition.

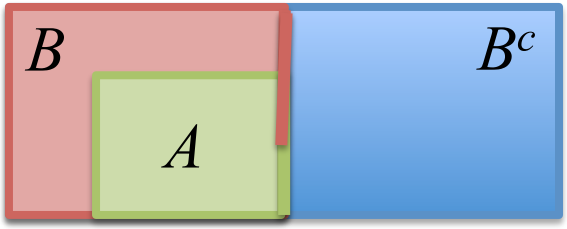
$$\begin{aligned}
\text{open} &\quad=\quad \text{verifiable} &\quad=\quad& \text{meth. verifiable;}\\
\text{closed} &\quad=\quad \text{refutable} &\quad=\quad& \text{meth. refutable;}\\
\text{clopen} &\quad=\quad \text{decidable} &\quad=\quad& \text{meth. decidable.}
\end{aligned}$$

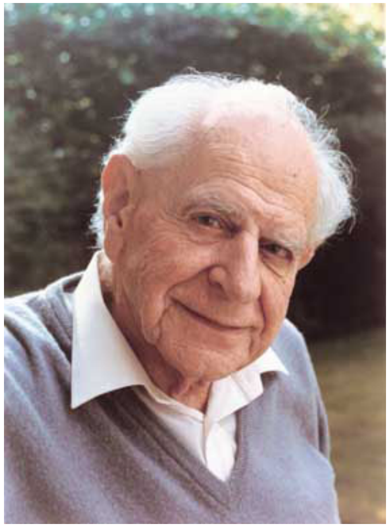
3. EMPIRICAL SIMPLICITY

Popper's Simplicity Order

- Every information state that refutes B refutes A.
- **Equivalently:** every information state compatible with A is compatible with B.

$$A \preceq B \text{ iff } A \subseteq \mathrm{cl} B.$$



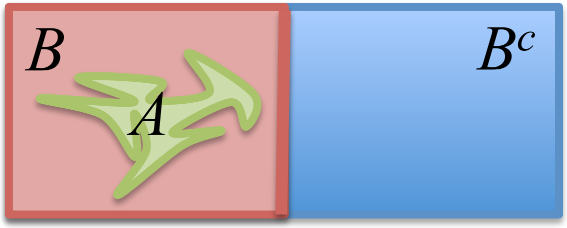


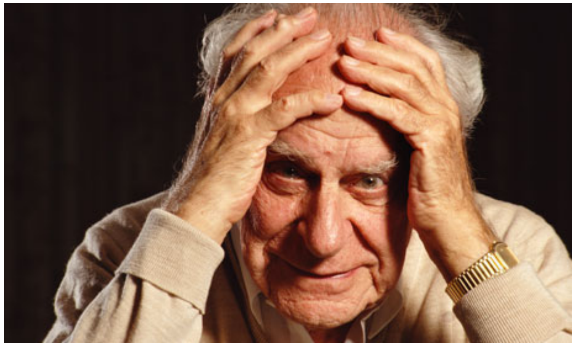
Karl Popper

The “Tack-on” Objection

- It’s wrong that stronger theories are simpler.

Clark Glymour





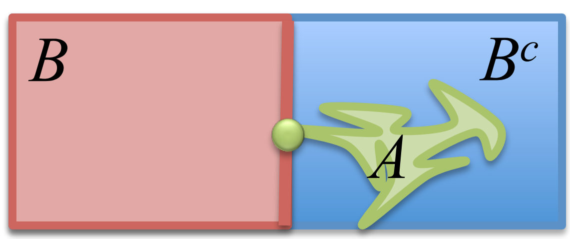
Karl Popper

Our Slight Revision

- It's possible that you face the problem of induction from A to B.

$$A \lhd B \text{ iff } \; A \cap \mathrm{cl}\, B \setminus B \neq \varnothing.$$

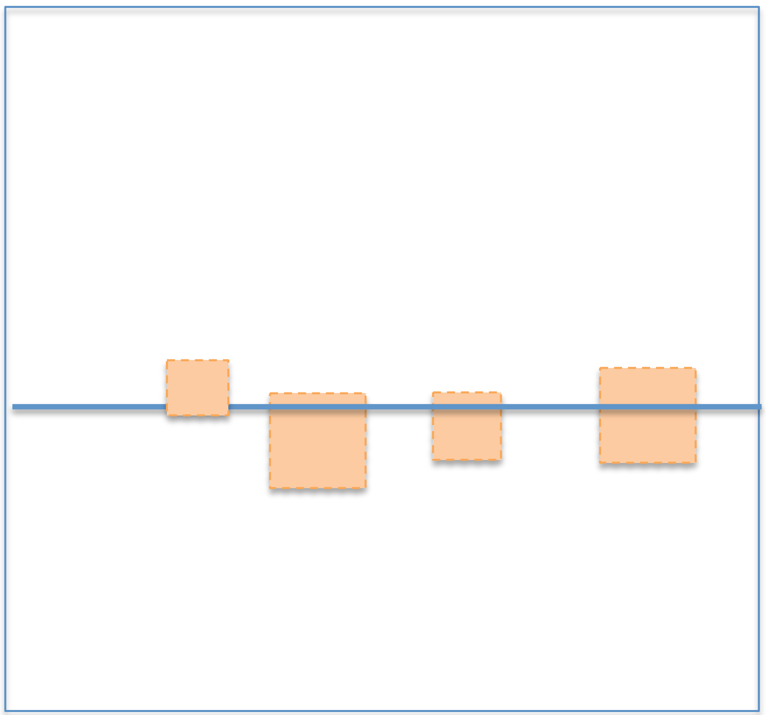
- Strict order if every answer is ver-ifutable.

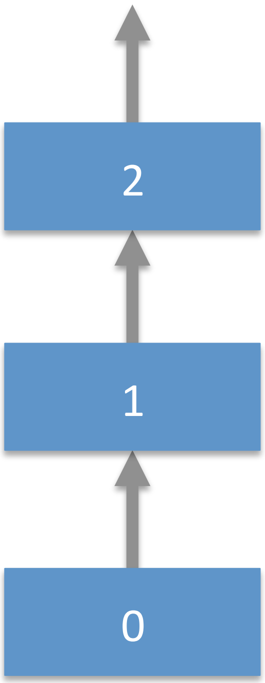


Example: Quantitative Laws

$\mathcal{Q}=$ What is the true polynomial degree?

$\mathcal{I}=$ finitely many inexact measurements.

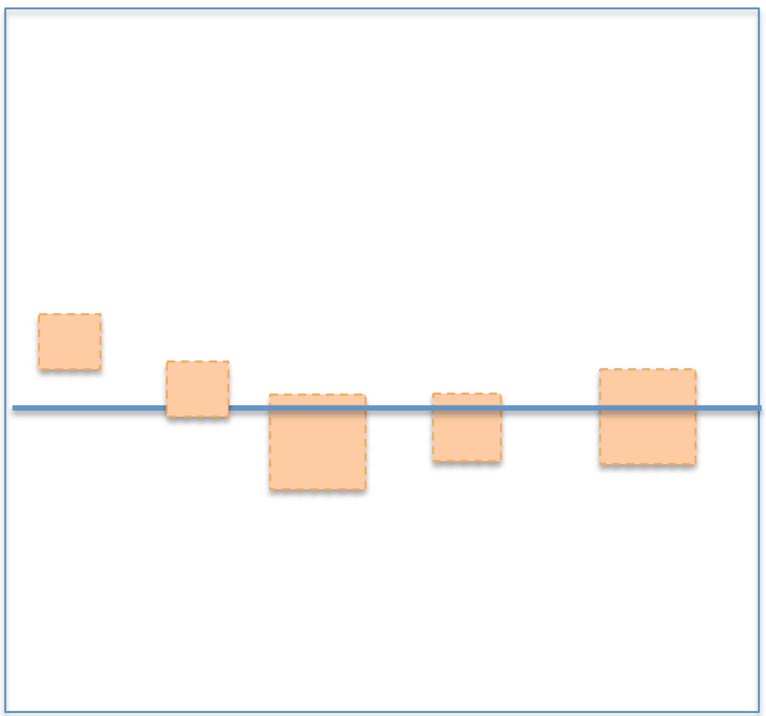


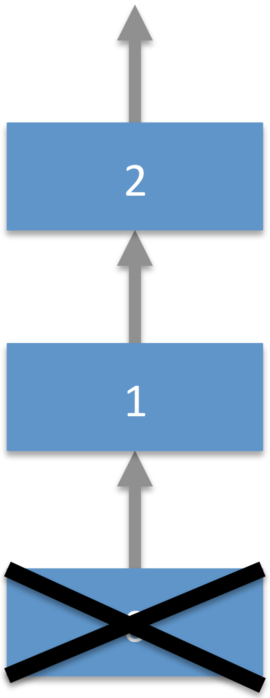


Example: Quantitative Laws

$\mathcal{Q} =$ What is the true polynomial degree?

$\mathcal{I} =$ finitely many inexact measurements.

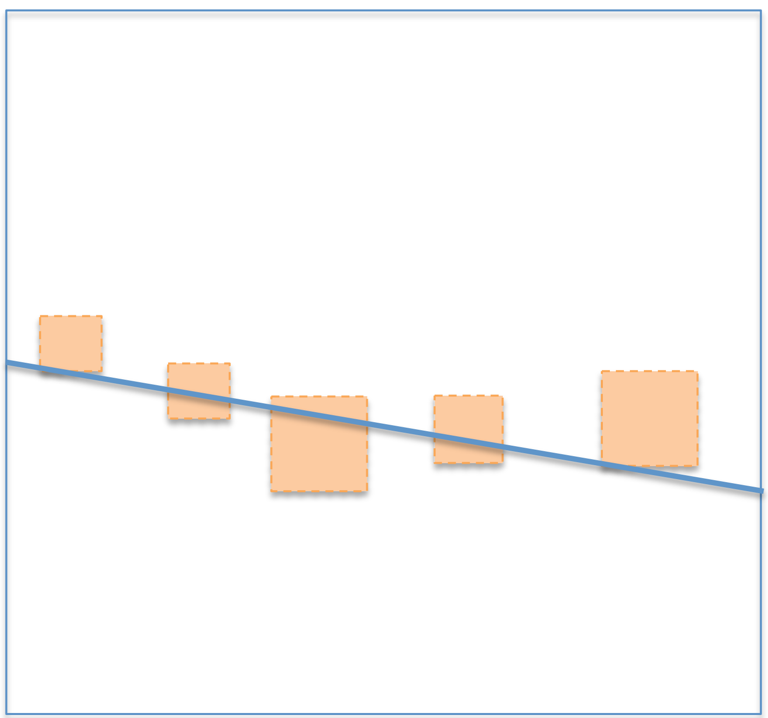


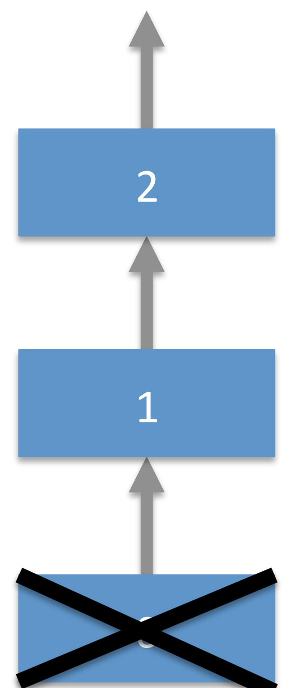


Example: Quantitative Laws

$\mathcal{Q}$ = What is the true polynomial degree?

$\mathcal{I}$ = finitely many inexact measurements.

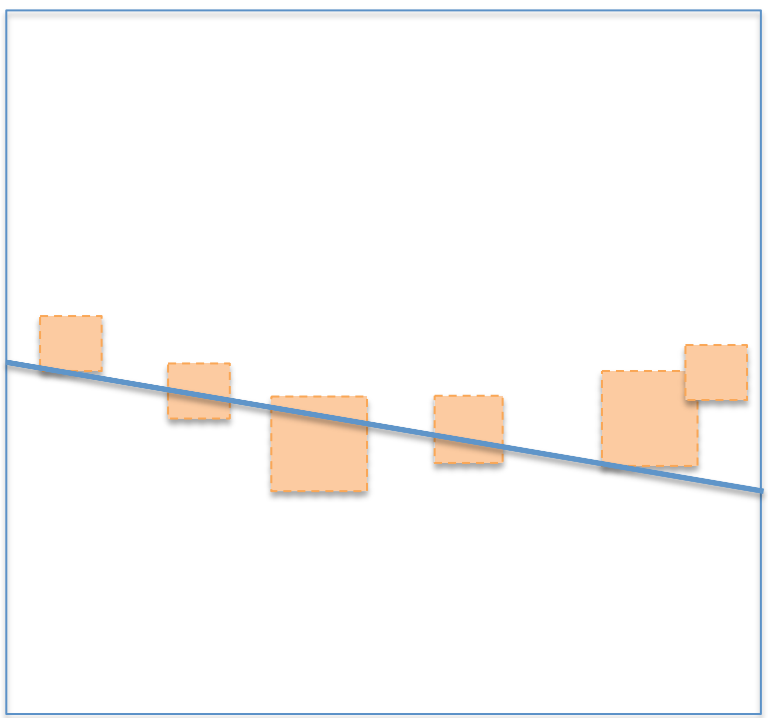


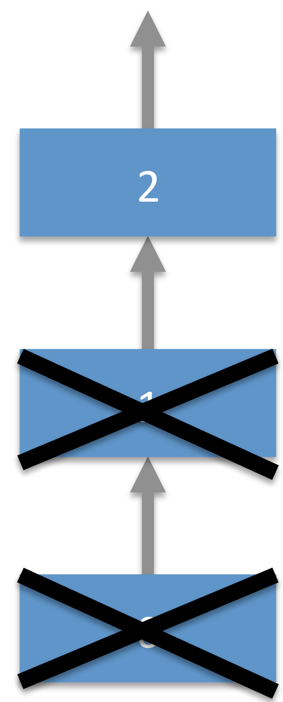


Example: Quantitative Laws

$\mathcal{Q} =$ What is the true polynomial degree?

$\mathcal{I} =$ finitely many inexact measurements.

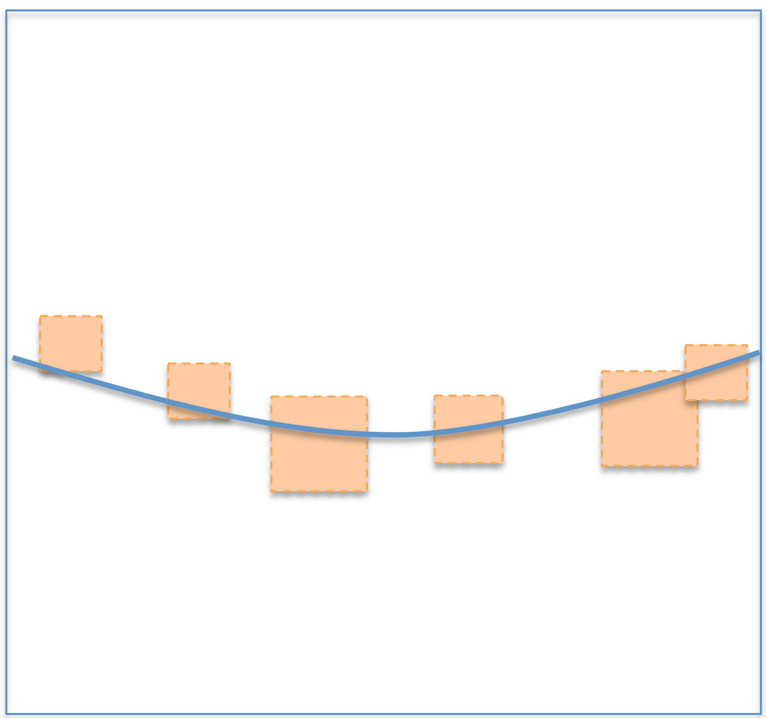


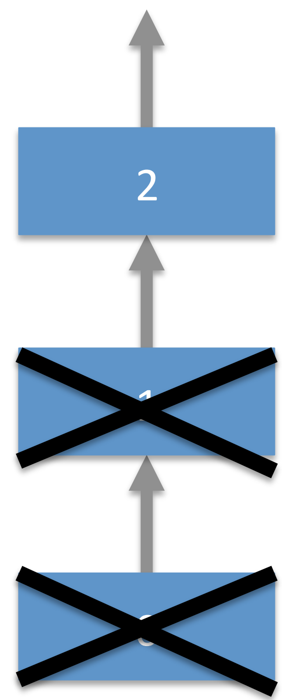


Example: Quantitative Laws

$\mathcal{Q} =$ What is the true polynomial degree?

$\mathcal{I} =$ finitely many inexact measurements.





Topological Simplicity

1. Motivated by the problem of induction.
2. Depends only on the structure of possible information.
3. Independent of notation.
4. Independent of parameterization.
5. Independent of prior probabilities.
6. Non-trivial in 0-dimensional spaces.

Example: EM Unification

Actual history (M. Morrisson).

Ockham

~~EM waves~~
~~EM waves ≠ light~~

Maxwell { EM waves
EM waves = light

Fizeau

~~No EM waves~~

Hertz

Example: EM Unification

Hypothetical history (H. Lin).

EM waves
EM waves ≠ light

?

Maxwell
~~EM waves~~
~~EM waves = light~~

Fizeau

No EM waves

Example: EM Unification

- If simplicity is a ranking, then Hertz is pre-empted by Ockham (AGM, Spohn).

Example: EM Unification

- But Hertz can settle the question, so wait.

Hertz?

EM waves
EM waves ≠ light

No EM waves

4. OCKHAM’S RAZOR

Ockham's Razor

- Output a simplest relevant response given E.
  - Allows for suspension of judgment.
  - Works for infinite descending simplicity chains.

Popper's Razor

- Output a relevant response that is refutable (closed) given E.

Error Razor

- "Err on the side of simplicity".
- In arbitrary world w, never produce a relevant response B such that the true answer A_w is strictly simpler than B.

All the Same Razor!

Proposition.

Ockham's razor = Popper's razor = error razor.

5. OCKHAM’S RAZOR JUSTIFIED

Inductive Justification

Infer straight to the truth

Too **strong**!

Convergence in the limit

Too **weak**!

Inductive Justification

Infer straight to the truth

Too **strong**!

Straightest possible convergence

Just right!

- Feasible;
- Mandates short-run norms.

Convergence in the limit

Too **weak**!

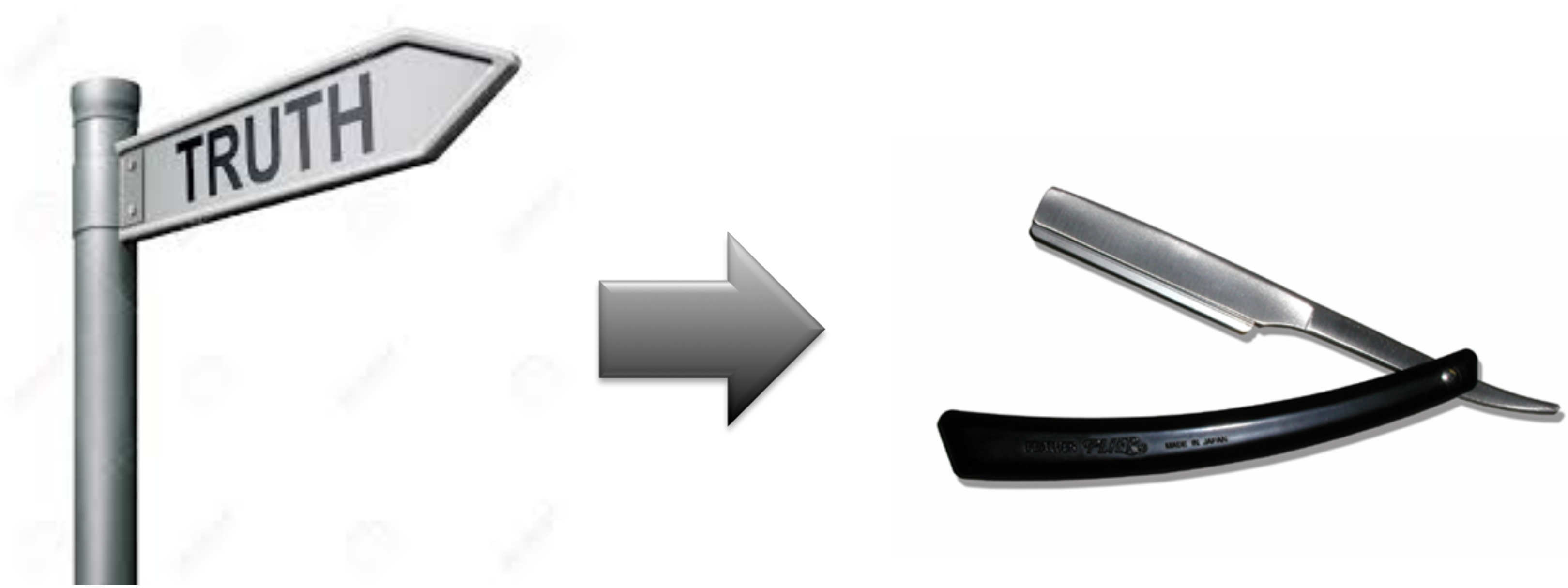
Departures from Straightness

Bad

Worse!

Thesis

- Ockham's razor is necessary for avoiding doxastic cycles.



Doxastic Cycles

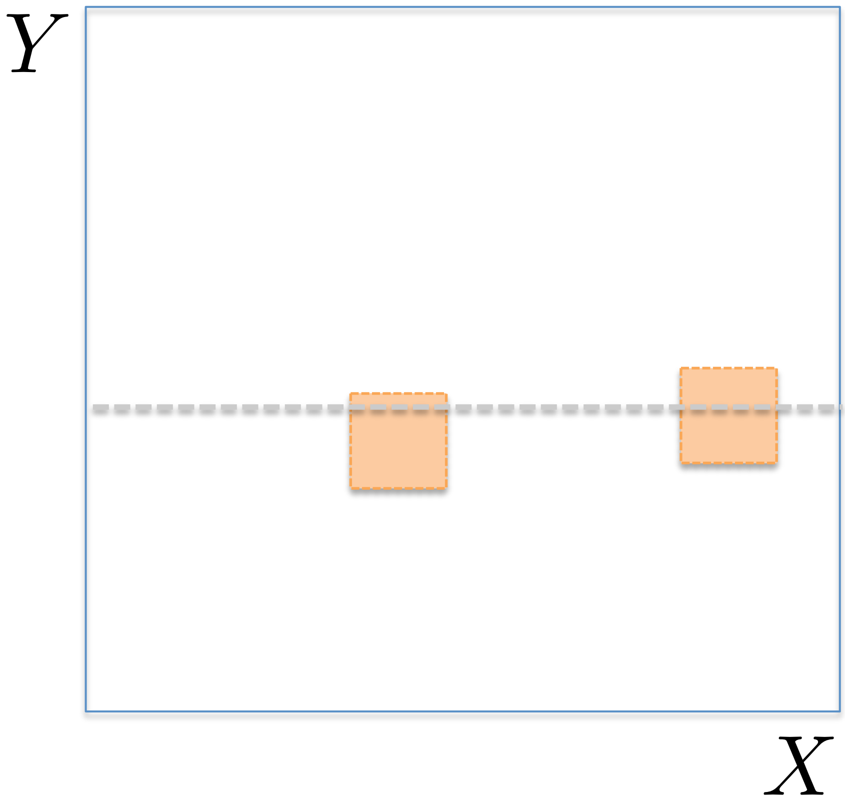
- Each relevant response contradicts the preceding.
- The last response entails the first.

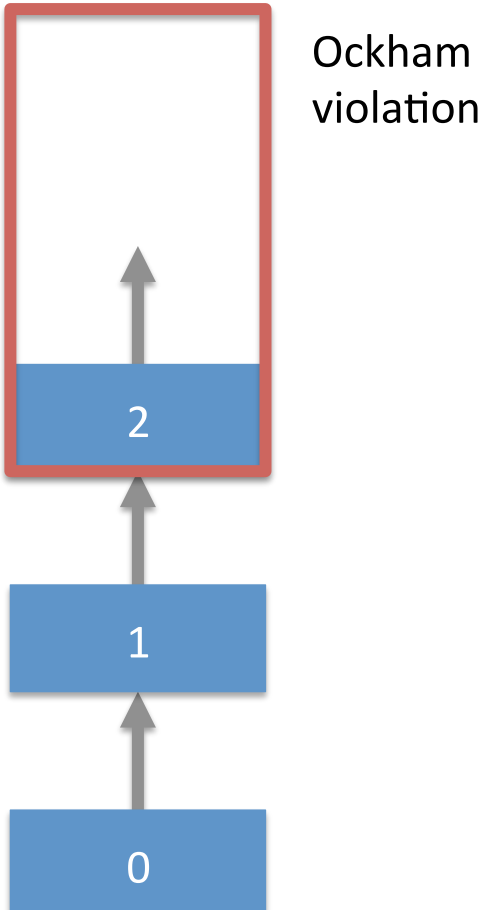
A B A

Main Result 1

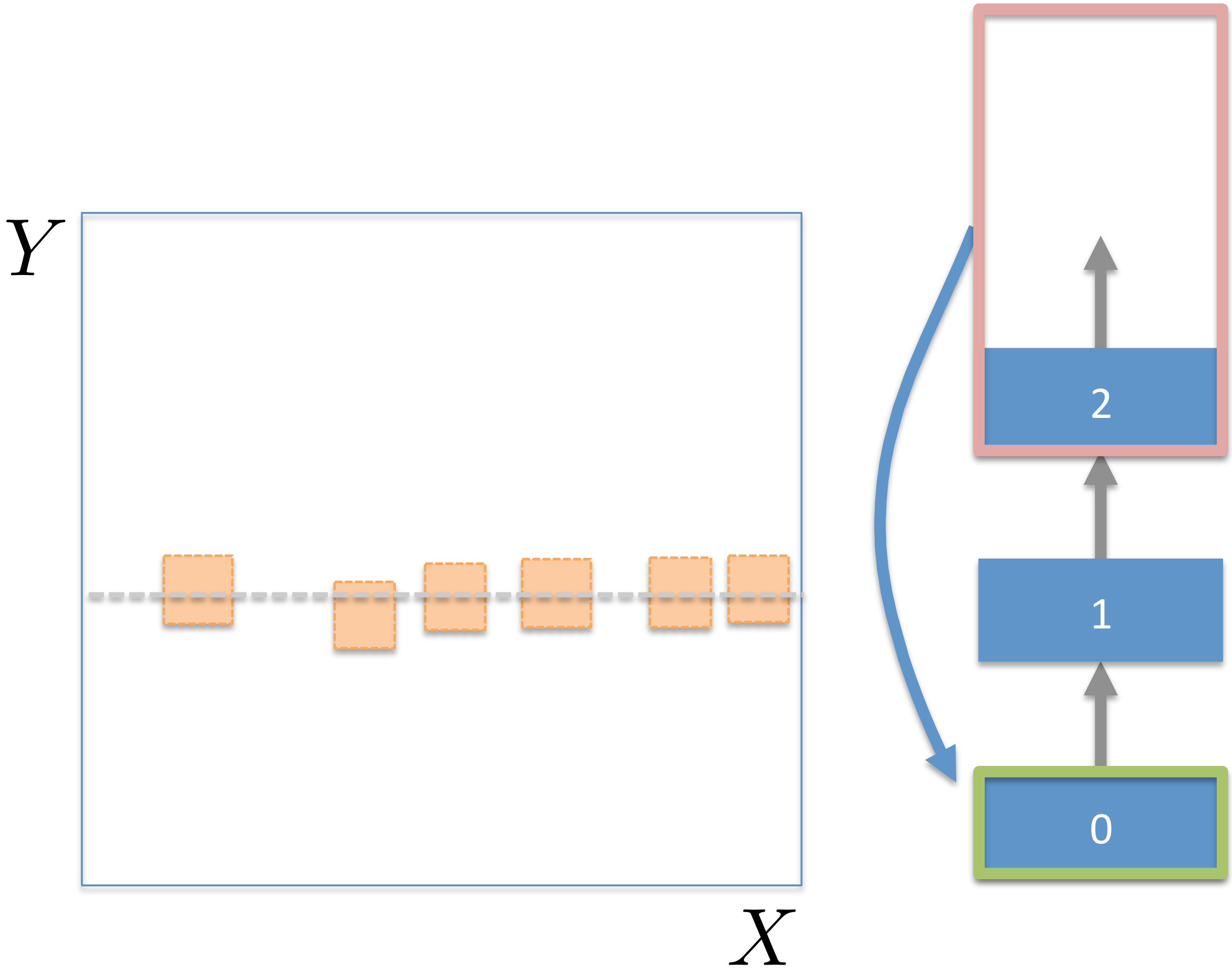
- **Proposition:** Every cycle-free solution satisfies Ockham's razor.

The Idea



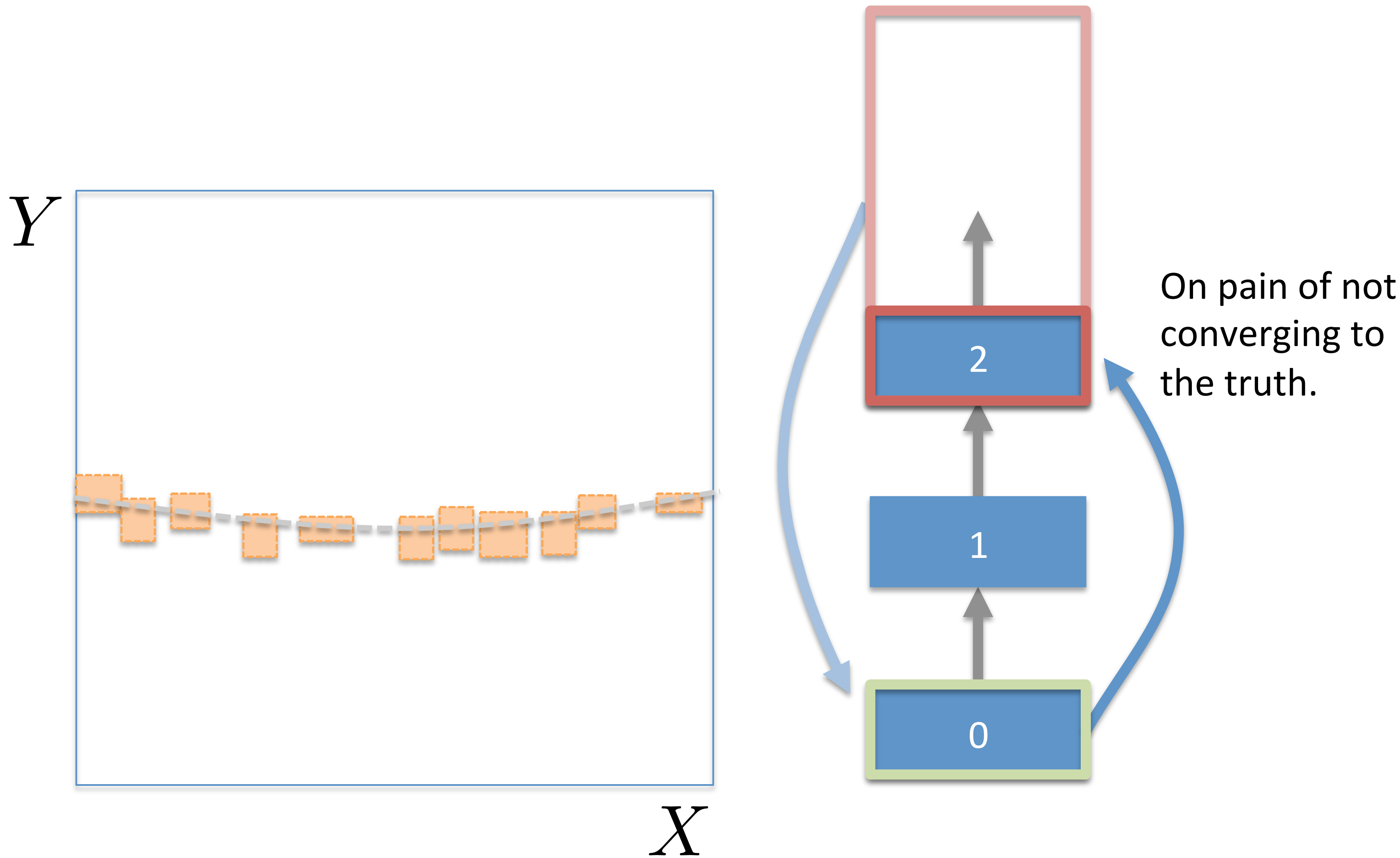


The Idea



On pain of not converging to the truth.

The Idea



Main Result 2

- **Proposition** (Baltag, Gierasimczuk, and Smets)**:** Every solvable question is refinable to a verifutable question with a cycle-free solution.

6. OCKHAM'S STATISTICAL RAZOR

Skepticism

The above account...

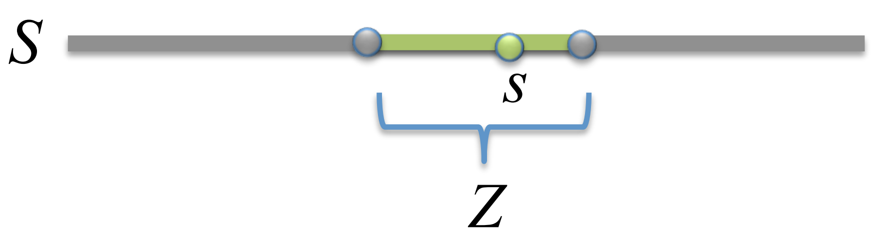
"may be okay if the candidate theories are **deductively** related to observations, but when the relationship is **probabilistic**, I am **skeptical** ...".

Eliott Sober, *Ockham's Razors*, 2015

Statistics

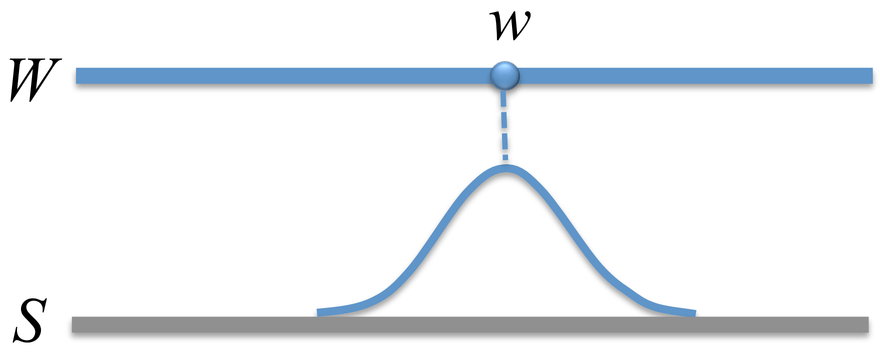
- The sample space S always comes with its own topology $\mathcal{T}$.
- $\mathcal{T}$ reflects what is verifiable about the **sample** itself.

s definitely falls within open interval Z .



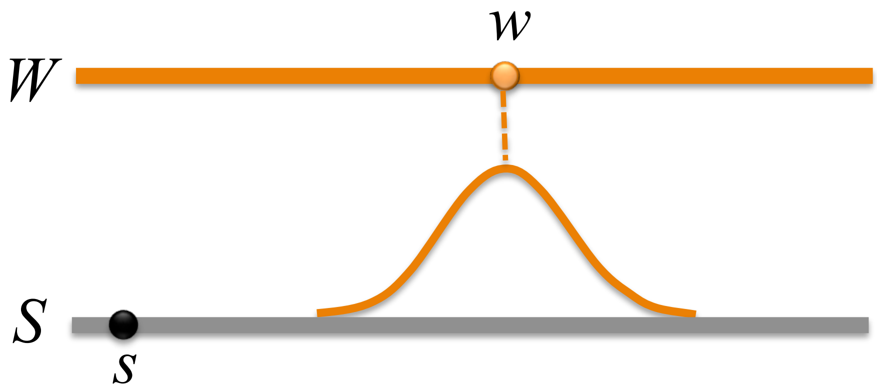
Statistics

- Worlds are probability measures over $\mathcal{T}$.



The Difficulty

- Every sample is logically consistent with all worlds!
- So it seems that statistical information states are all trivial!



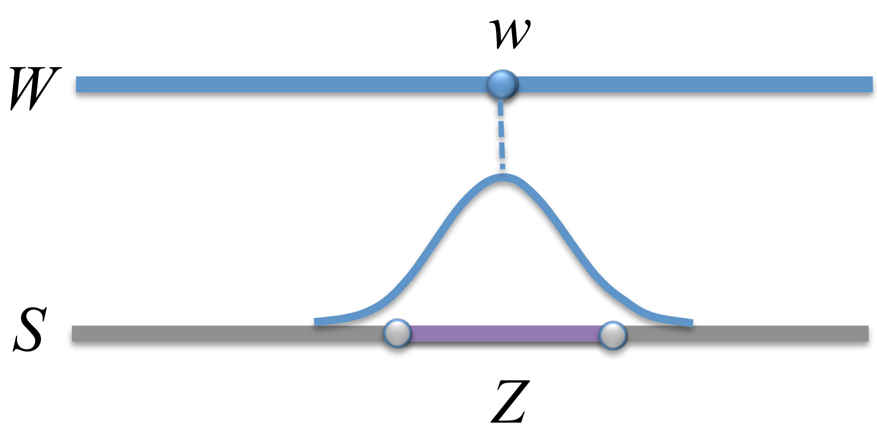
Response

- Solve for the unique topology whose open sets are exactly the statistically verifiable propositions.

Topology  Statistics

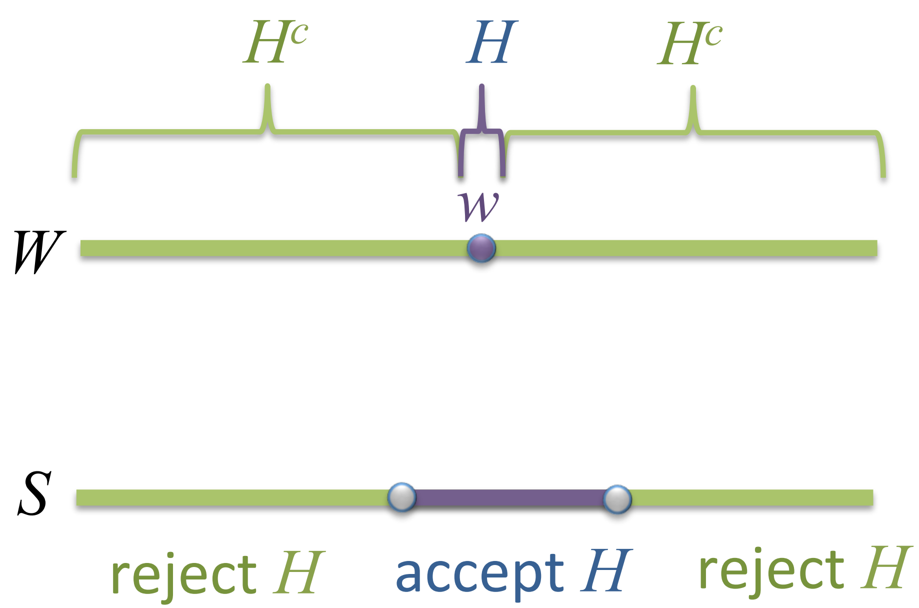
Feasible Sample Events

- It's impossible to tell whether a point right on the boundary of Z is in or out of Z.
- Z is feasible iff the chance of the boundary is zero in every world.



Feasible Tests

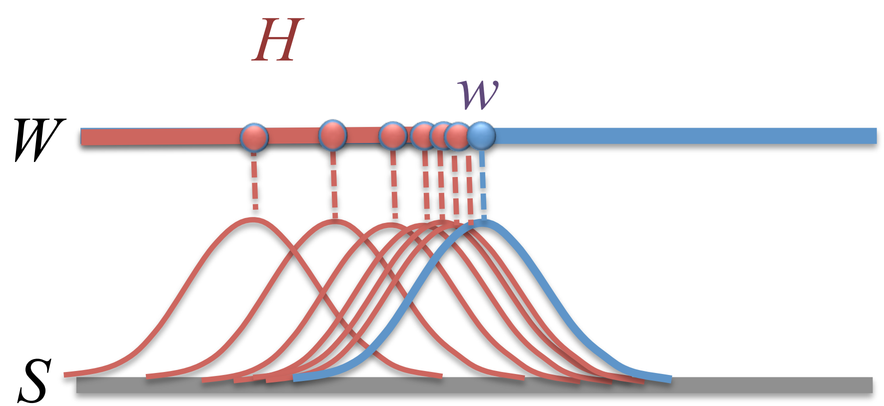
A feasible test M of a statistical hypothesis H is a measurable function from samples to {not-H, W} with a feasible rejection region.



Statistical Information Topology

$w \in \mathrm{cl}\, H$ iff there exists sequence (w_n) in H, such that for all feasible tests M:

$$\lim_{n\to\infty} p_{w_n}(M \text{ rejects}) \to p_w(M \text{ rejects}).$$



Weak Topology

Proposition: If $\mathcal{T}$ has a basis of feasible zones, then statistical information topology = the standard, weak topology.

Statistical Verification Methods

- A **statistical verification method** for H at level $\alpha > 0$ is a sequence (M_n) of feasible tests of not-H such that for all n:

  1. if $w \in H$: M_n converges in probability to H;
  2. If $w \in H^c$: M_n concludes W with probability at least 1-α.

- H is **statistically verifiable** iff H has a statistical verification method at each $\alpha > 0$.

The Topology of Statistical Methodology

Proposition. If $\mathcal{T}$ has a basis of feasible regions,

1. **open** = statistically verifiable.
2. **closed** = statistically refutable.
3. **clopen** = statistically decidable.

Topology

Statistics

Simplicity

Same as before!

I.e., it is possible that A is true but B is never statistically refuted.

Ockham’s Razor

- “Simplest compatible with the data” is trivial since every answer is logically compatible with every sample.

Ockham's Razor

- "Simplest compatible with the data" is trivial since every answer is logically compatible with every sample.

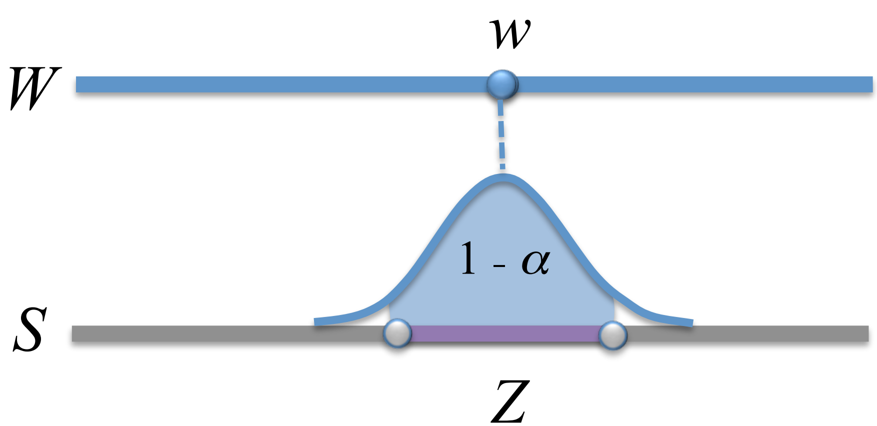
- **Solution.** The **error razor** is defined in terms of truth rather than **compatibility** with current information, so it still makes sense!

Ockham's α-Razor

Probabilistic version of the error-razor:

> A statistical method is α-Ockham iff the chance that it outputs an answer more complex than the true answer is bounded by α.

Agrees with significance for simple vs. complex binary questions!



Ockham's α-Razor

Statistical method (M_n) is α-Ockham iff for all worlds w, sample sizes n and relevant responses A:

$$\text{if } Q_w \lhd A, \text{ then } p_w^n[M_n = A] \leq \alpha.$$

Reversals in Chance

Method (M_n) performs the sequence (A, B) at α iff there is a world w and two sample sizes such that:

the **gain** in chance of outputting B

pro-rated by

the **loss** in chance of outputting A,

is at least α.

Main Result

- **Proposition:** Ockham's α razor is necessary for avoiding α cycles in chance.
- Valid for all solvable problems.

Conjecture (with Simulations)

- Every solvable question is refinable to a verifutable question such that has an α-cycle-free solution, for all $\alpha > 0$.

Summary and Discussion

1. Simplicity is a **topological** feature of problems.
2. Topological system is **notation-independent**.
3. Ockham's razor is necessary for **optimally straight** convergence to the truth.
4. The same holds for **statistical** inductive inference.
5. Optimally straight convergence is weak, but its implications for scientific method are **strong**.

THE BAYESIAN MIRACLE

It Would be a Miracle if...

...the parameters of the complex theory were tuned to mimic the predictions of the simple theory.

The Miracle is in You

On simple data E there is parameter setting θ such that:

$$p(E \mid Comp(\theta)) \approx p(E \mid Simp).$$

So the **miracle** is your own **prior prejudice**.

$$p(Simp) \gg p(Comp(\theta)).$$

But that **is** Ockham's razor, not an epistemic **justification** of it.

THE "OVER-FITTING ARGUMENT"

Accuracy

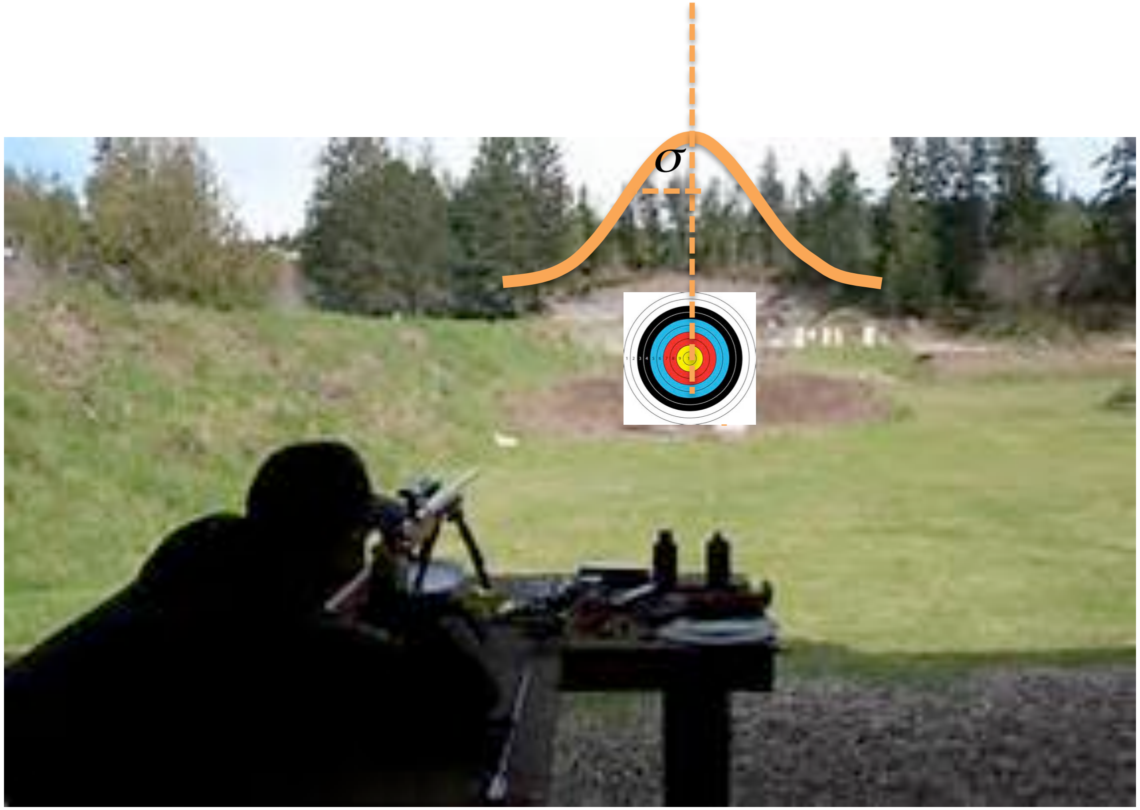
- Our national pastime.

Analysis of Inaccuracy

- MSE $=$ **bias**^{2} $+$ **variance**.

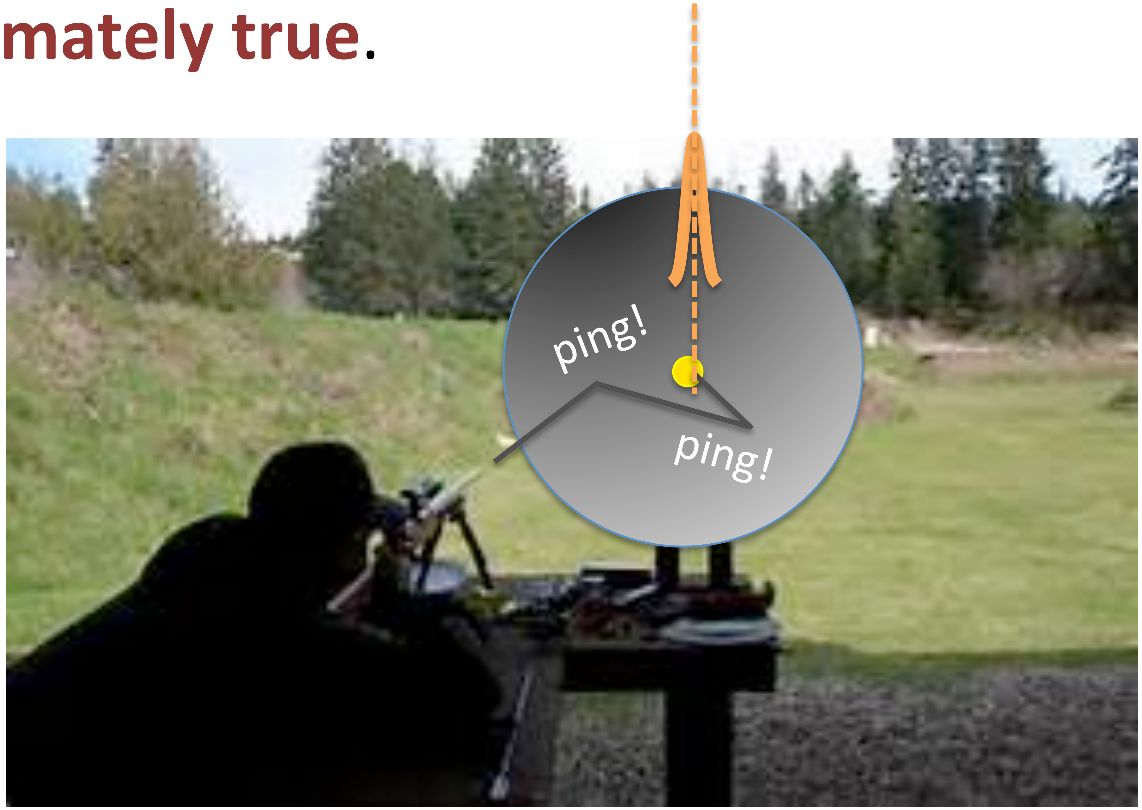
Non-Ockham Empirical Estimates

Variance but no **bias**.



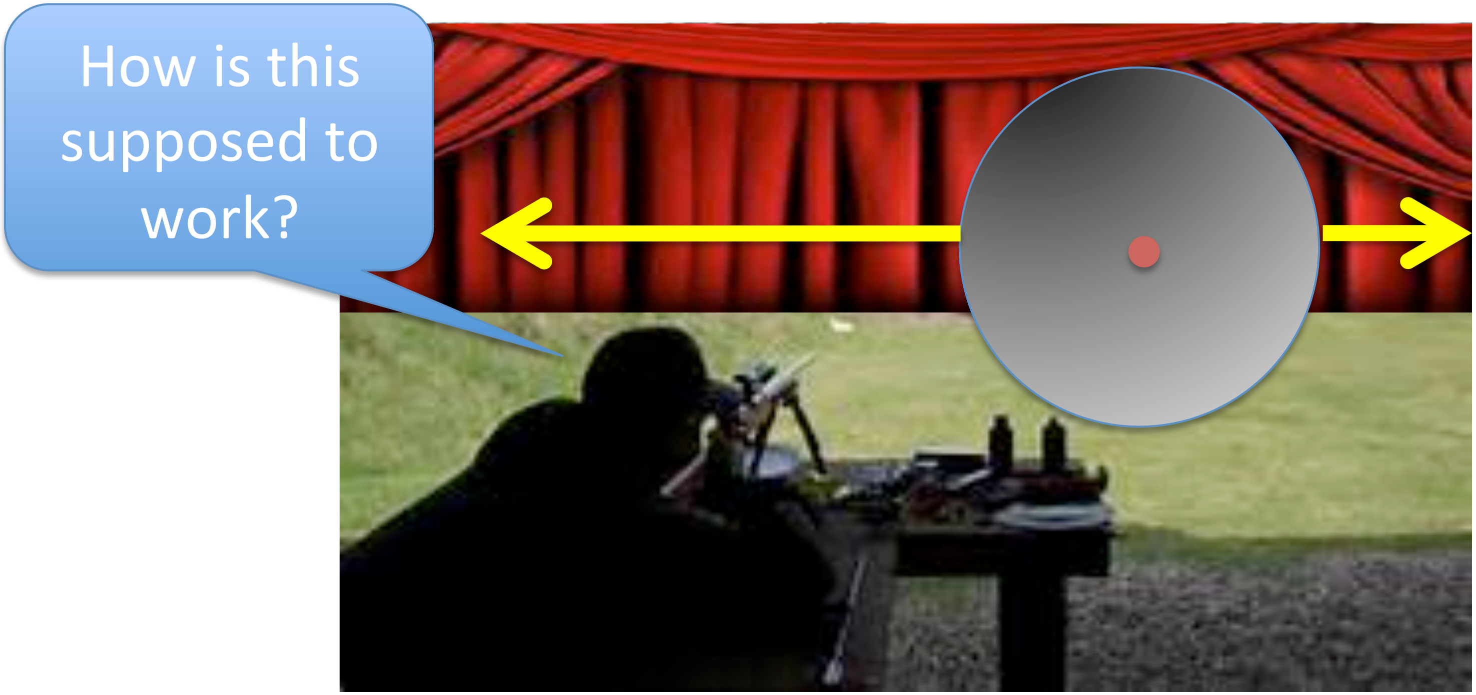
Ockham Estimates

- Like shooting through a **funnel**.
- Small variance and bias **if** the simple theory is **approximately true**.



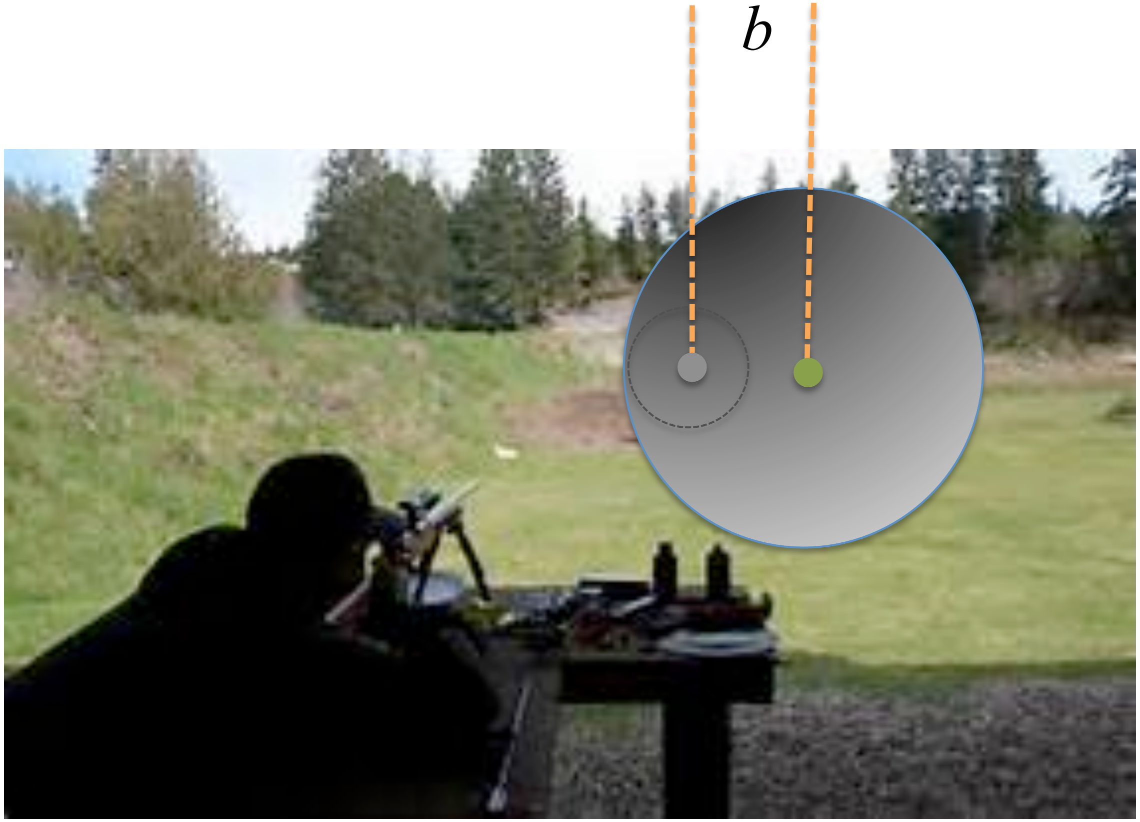
But in Science...

- The funnel must be installed **before** you see the target!



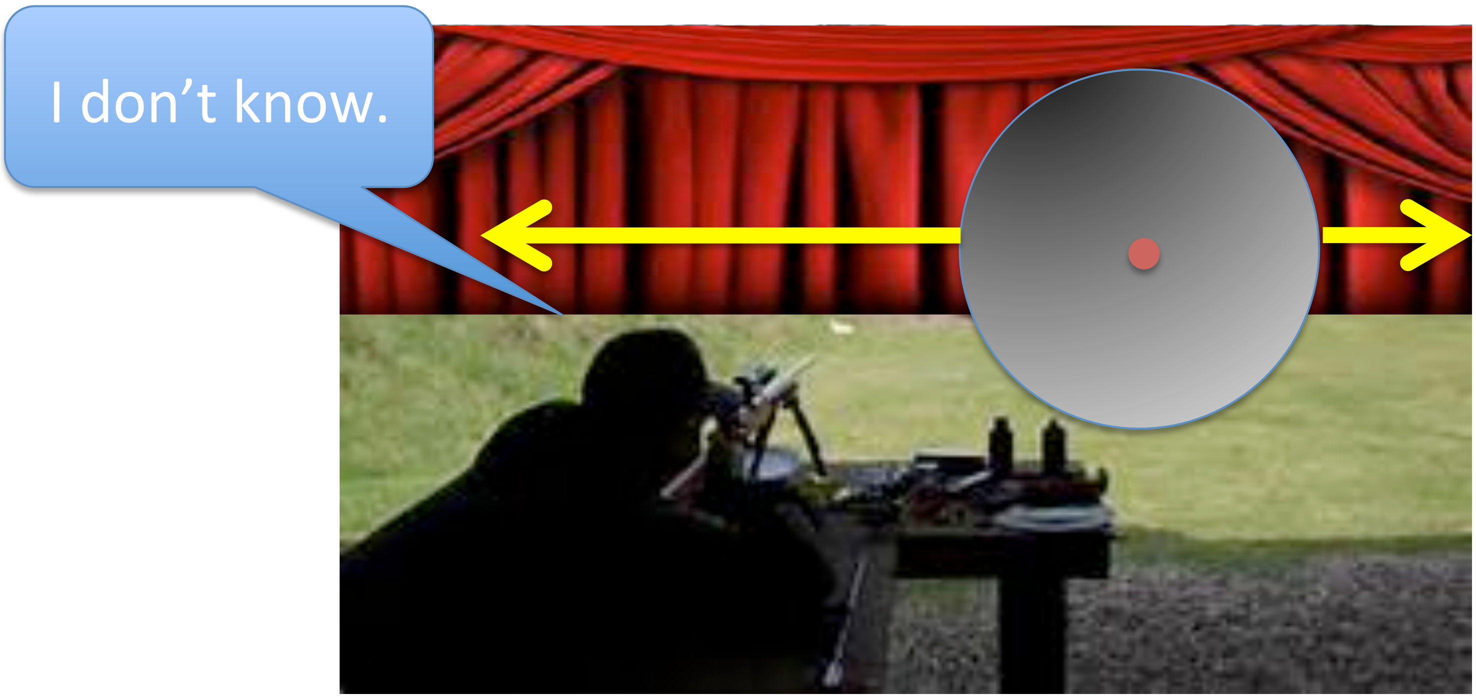
The Curtain Rises

- If the funnel is not nearly centered, it makes good shots **worse**.



The Elusive Overfitting Argument

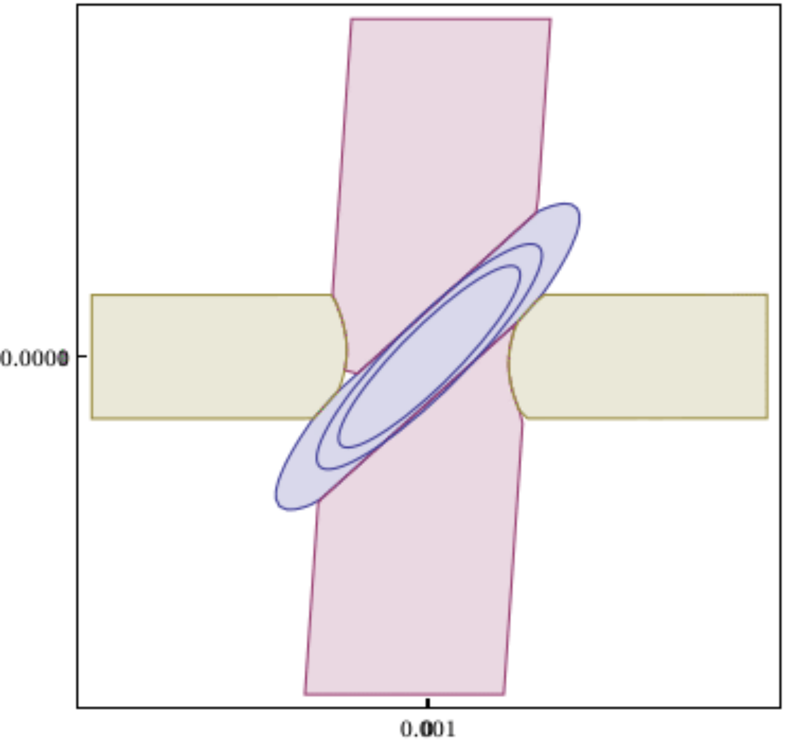
- So how does blindly installing the funnel make you more accurate?

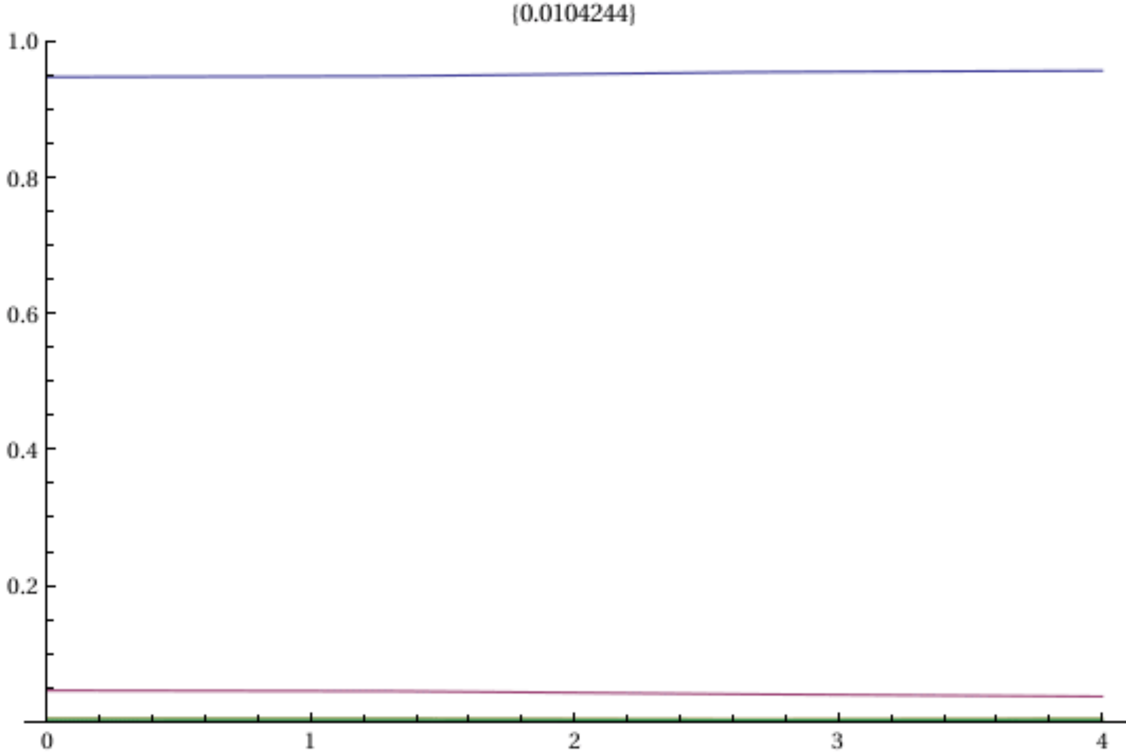


SIMULATIONS

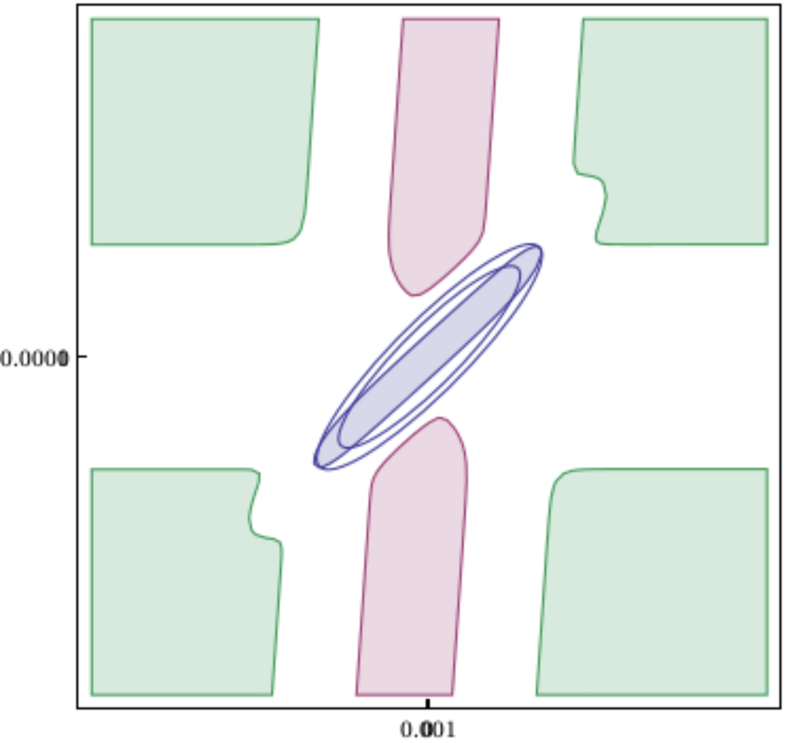
Bayesian Mode

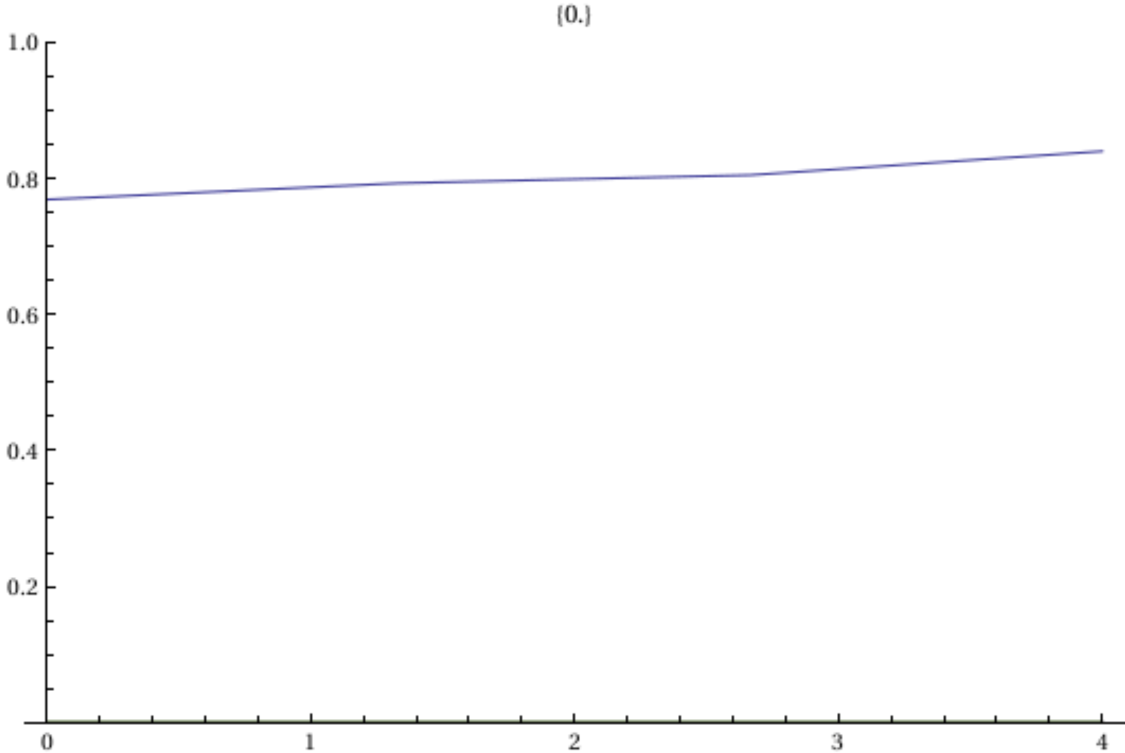
- Method: Bayesian mode.
- Prior bias toward simplicity, Gaussian priors on parameters.





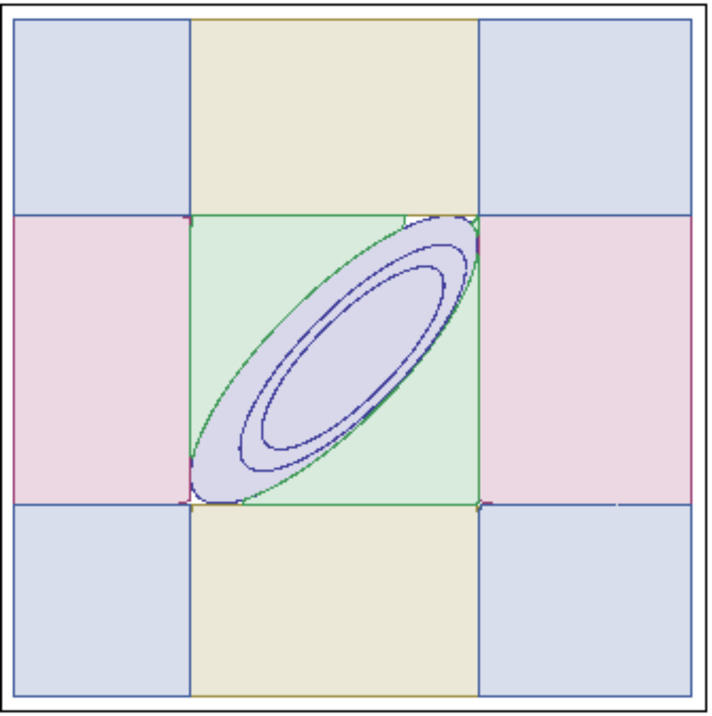
Bayesian, 95% Threshold

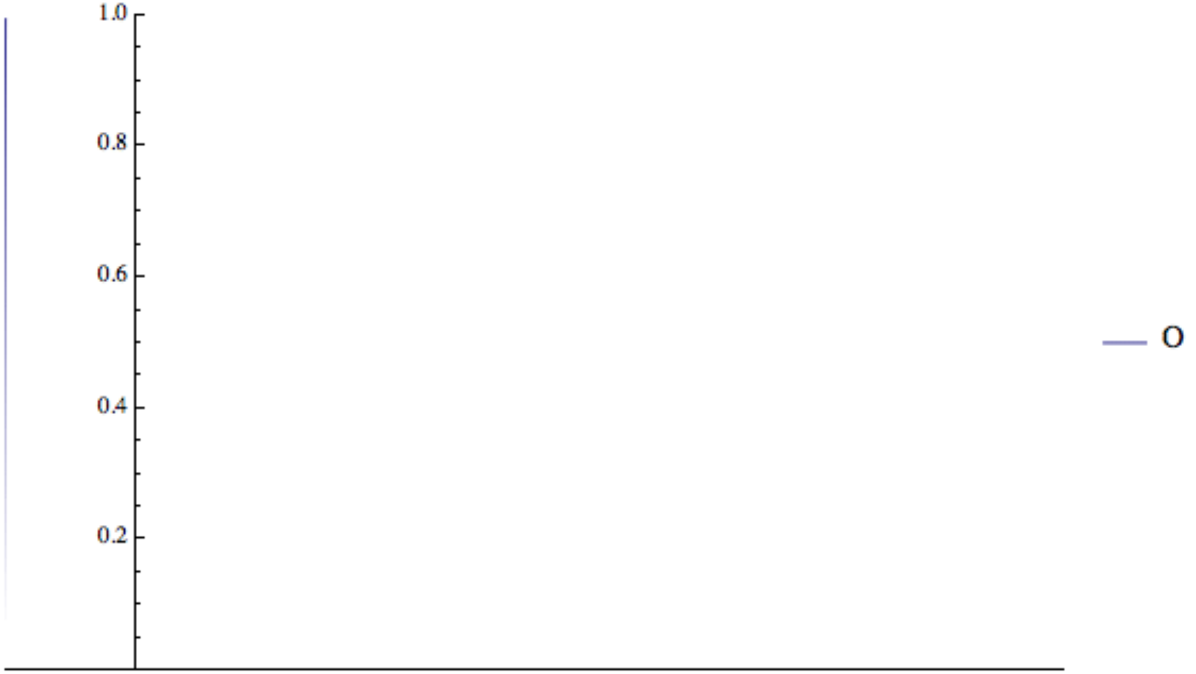




Frequentist Ockham

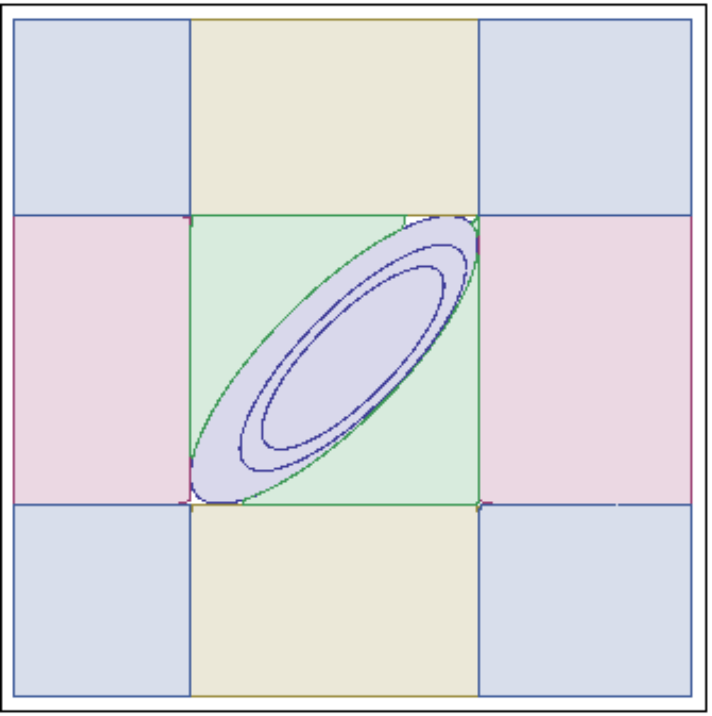
- Nested tests

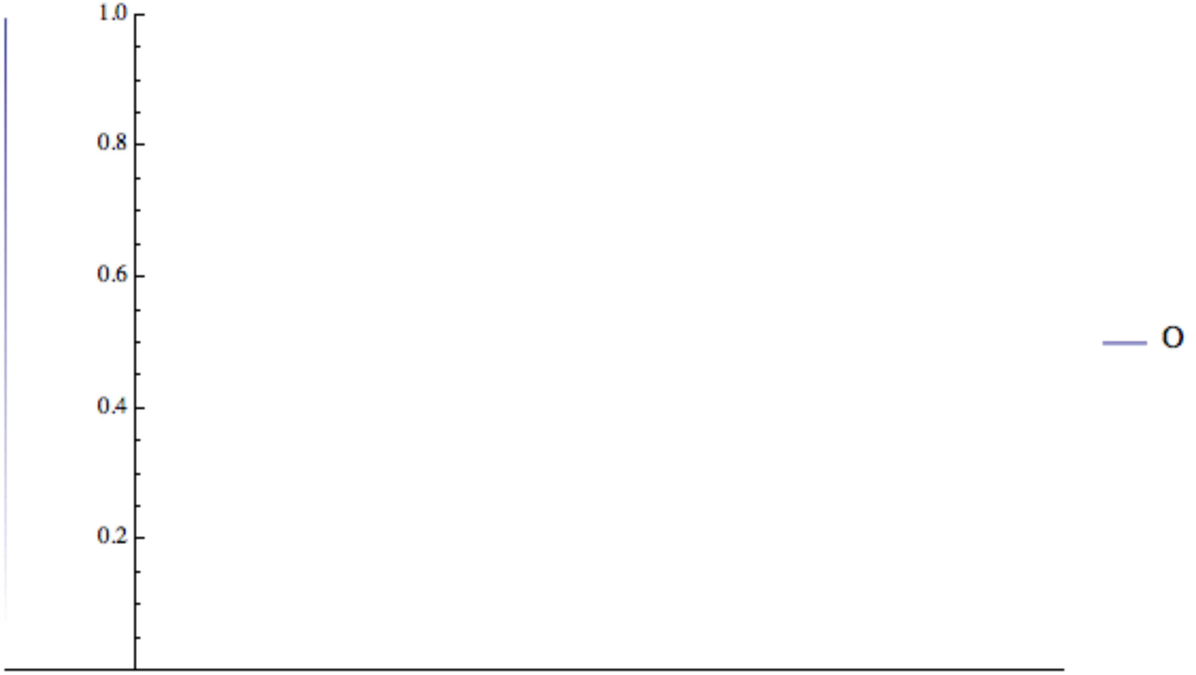




Error Statistics Reinterpreted

- "Significance" = tolerance on cycles and reversals in chance.
- "Power" = if you are destined to drop a model, get it over with a.s.a.p.





SIMPLICITY AND PARADIGMS

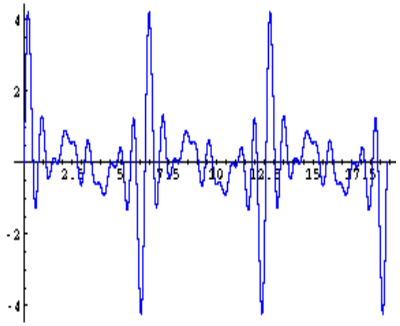
Example: Competing Paradigms

Polynomial paradigm

$$Y = \sum_{i=0}^{N} a_i X^i.$$

Trigonometric polynomial paradigm

$$Y = \sum_{i=0}^{N} a_i \sin(iX) + b_i \cos(iX).$$



Example: Competing Paradigms

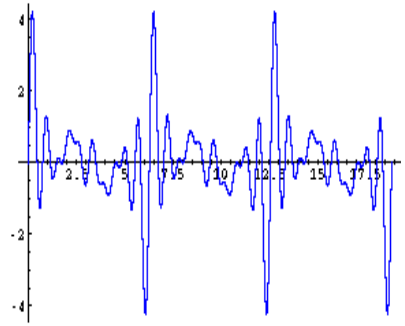
Polynomial paradigm

$$Y = \sum_{i=0}^{N} a_i X^i.$$

degree

Trigonometric polynomial paradigm

$$Y = \sum_{i=0}^{N} a_i \sin(iX) + b_i \cos(iX).$$



Example: Competing Paradigms

$\mathcal{Q} =$ which degree and which paradigm is true?

$\mathcal{I} =$ finitely many inexact measurements.