The Topology of Empirical Simplicity

{Kevin T. Kelly, Konstantin Genin}

Ockham’s Razor

- **Ockham:** “Pluralitas non est ponenda sine neccesitate.”
- **Science:** “Presume no more complexity than necessary.”

Ockham’s Razor

- **Ockham:** “Pluralitas non est ponenda sine neccesitate.”
- **Science:** “Presume no more complexity than necessary.”
- But what is simplicity?
- And why rely on it?

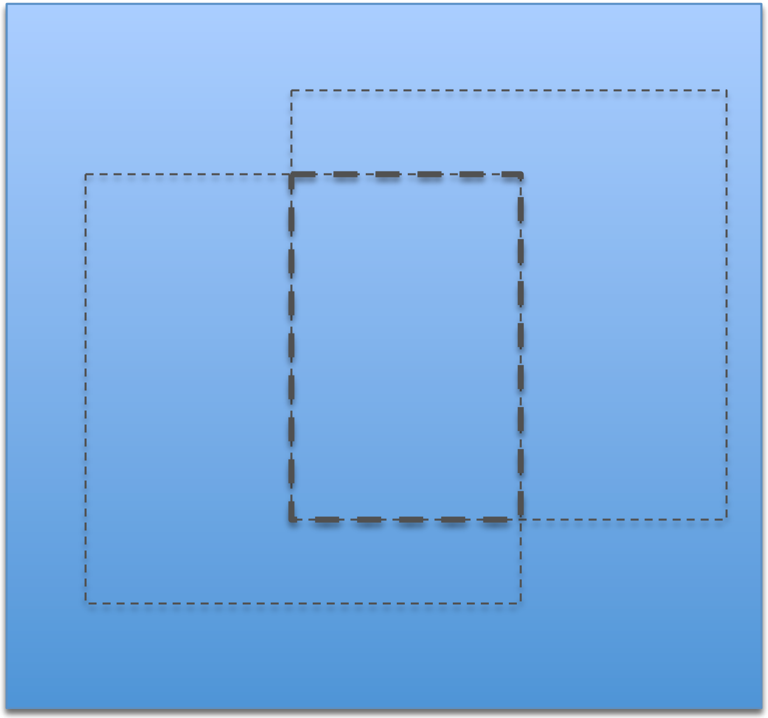
Indispensable for:

- Theory choice
- Inductive inference
- Statistical model selection
- Causal discovery from non-experimental statistical data.

1. EMPIRICAL PROBLEMS

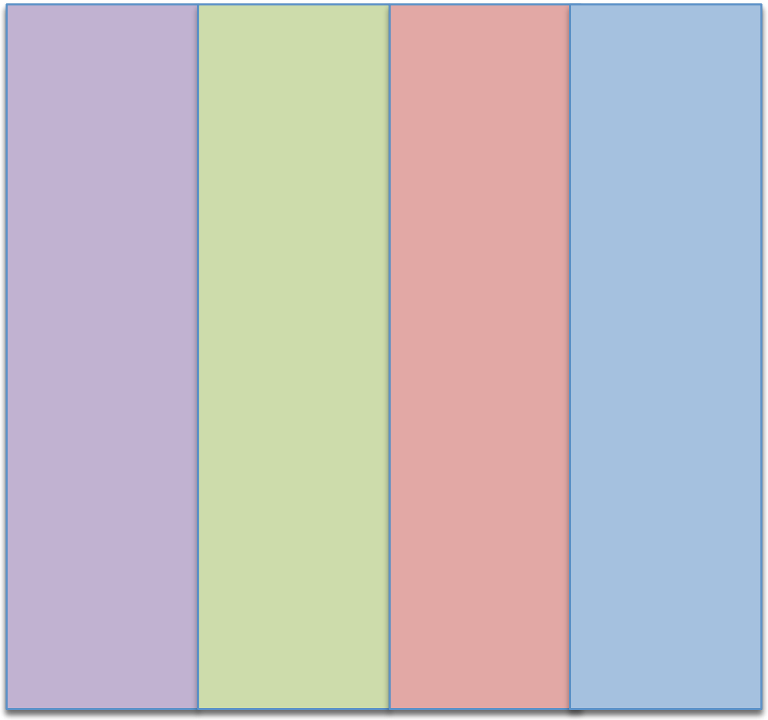
Information Spaces

- W is a set of possible worlds.
- $\mathcal{I}$ is a set of propositional information states.
  1. Covers W.
  2. Closed under finite conjunction.



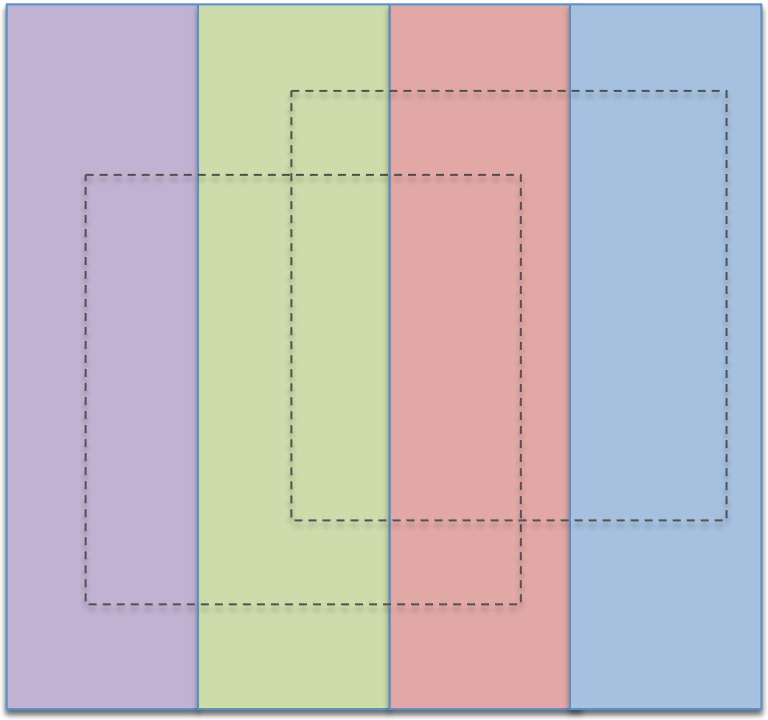
Questions

- A question partitions W into countably many possible answers (Hamblin 1958)
- Relevant responses are disjunctions of answers.

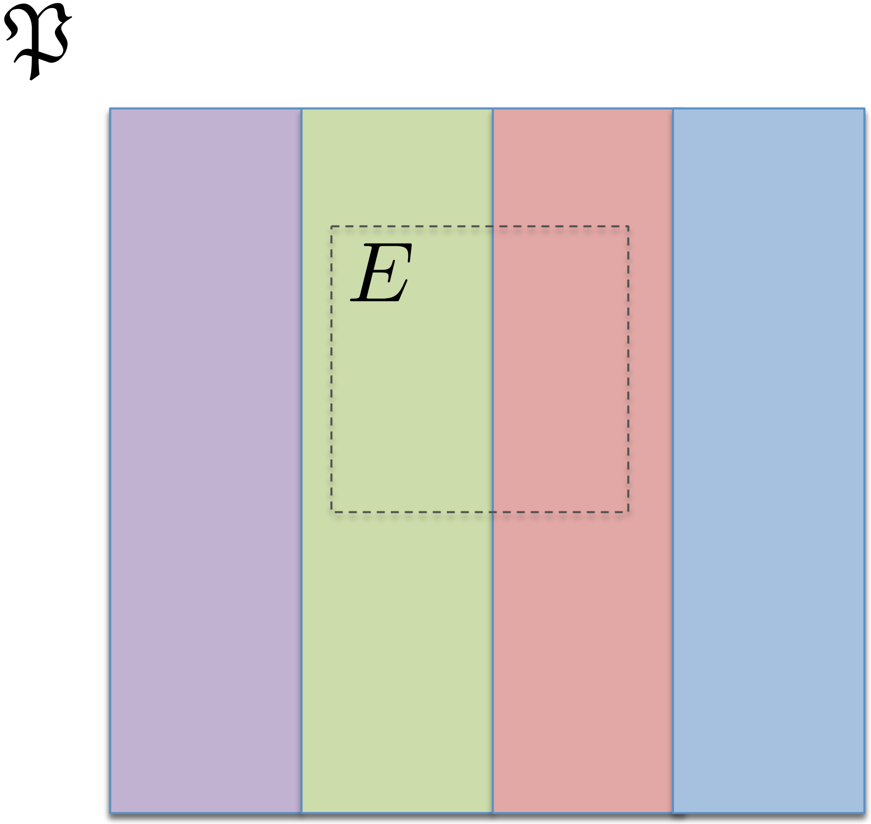


Empirical Problem

$$\mathfrak{P} \;=\; (W, \mathcal{I}, \mathcal{Q}).$$

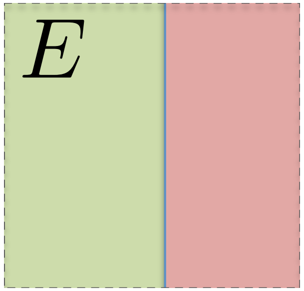


Problem Restriction



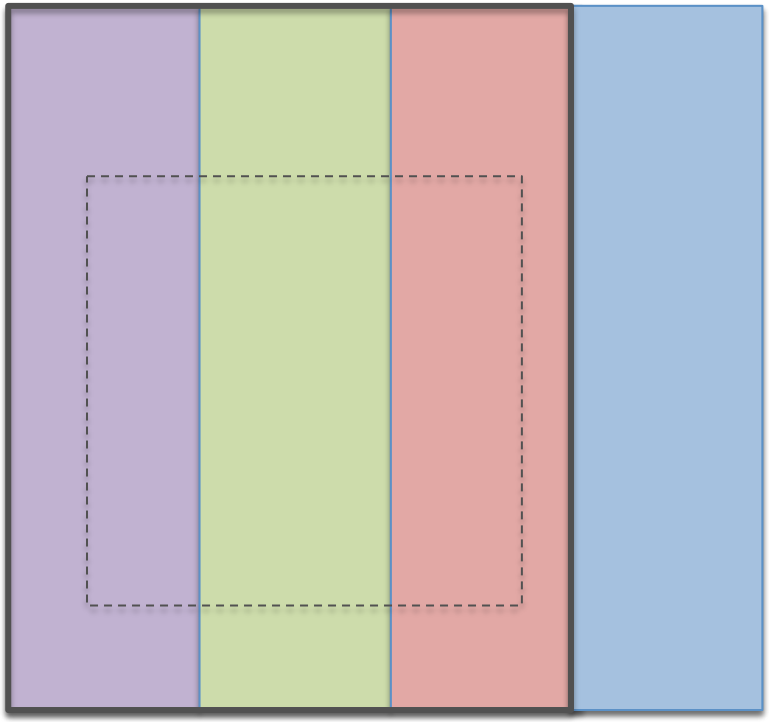
Problem Restriction

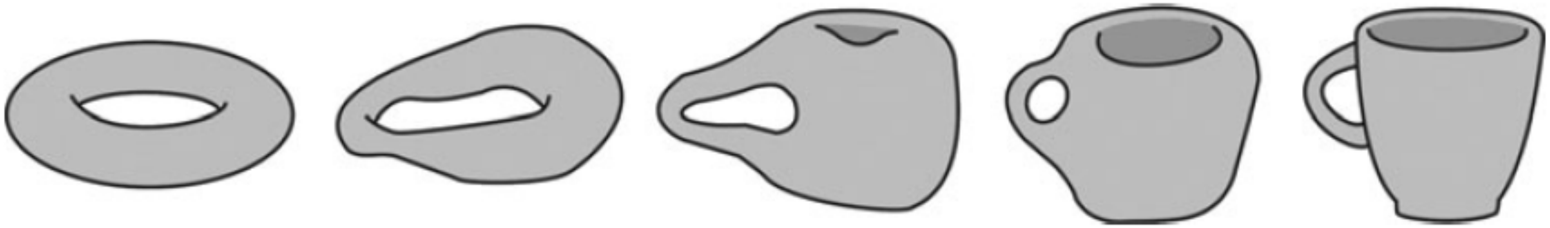
$\mathfrak{P}|_E$



Relevant Response Given E

Disjunction of answers compatible with E.

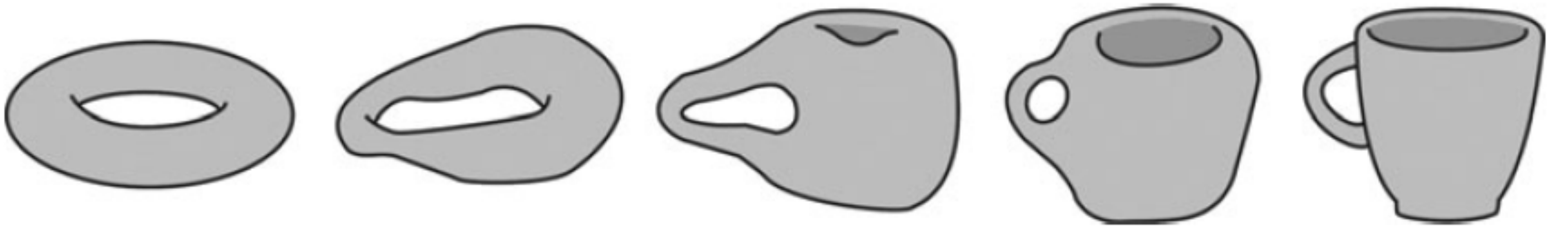




2. INFORMATION TOPOLOGY

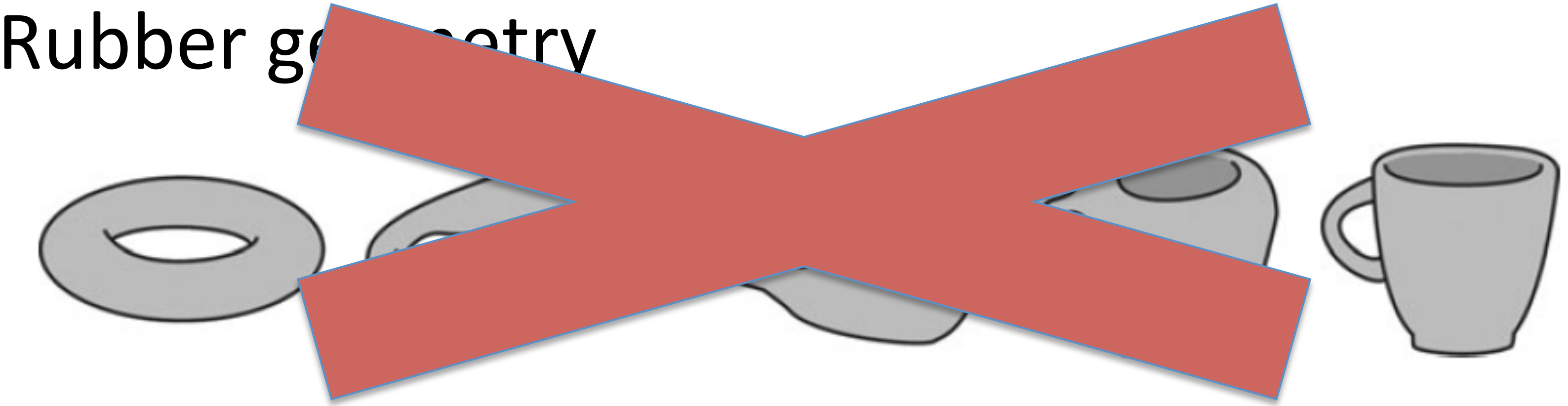
Topology

Rubber geometry



Topology

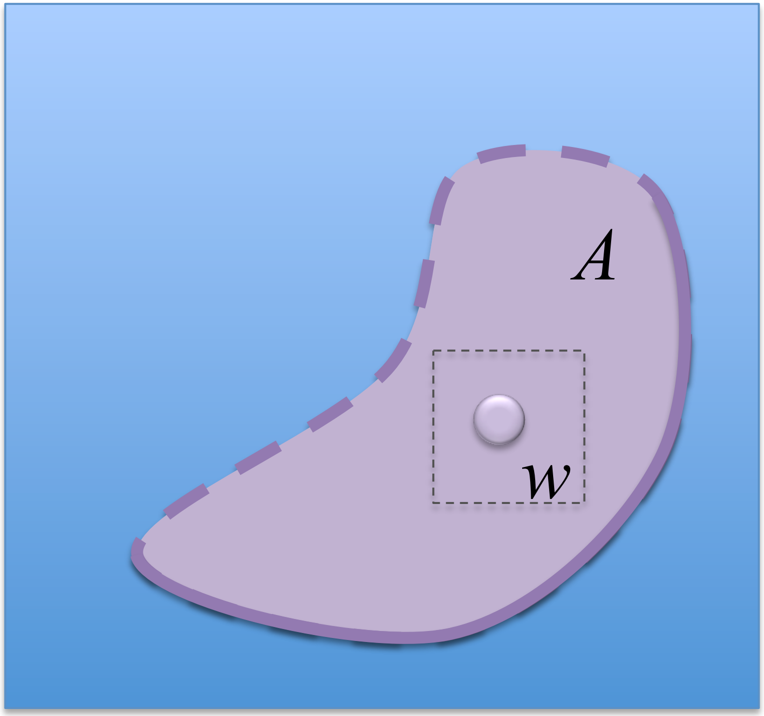
~~Rubber geometry~~



The logic of verification.

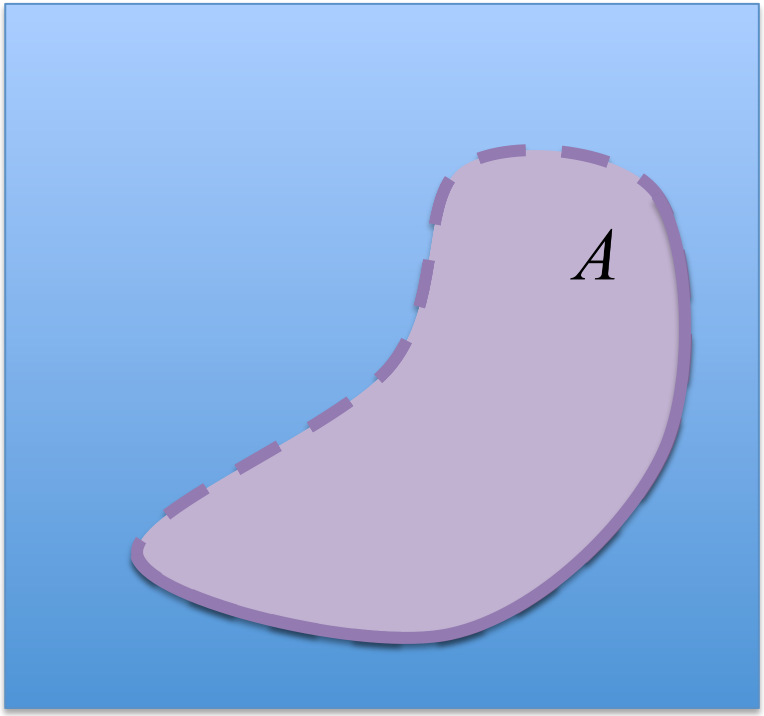
w is an Interior Point of A

W presents information that verifies A.



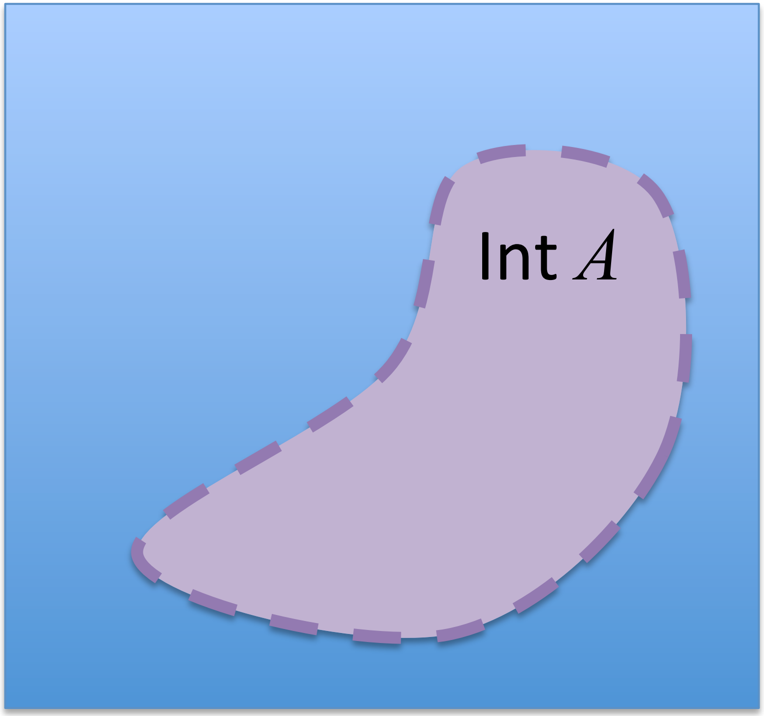
Interior of A

Int A = it will be verified that A.



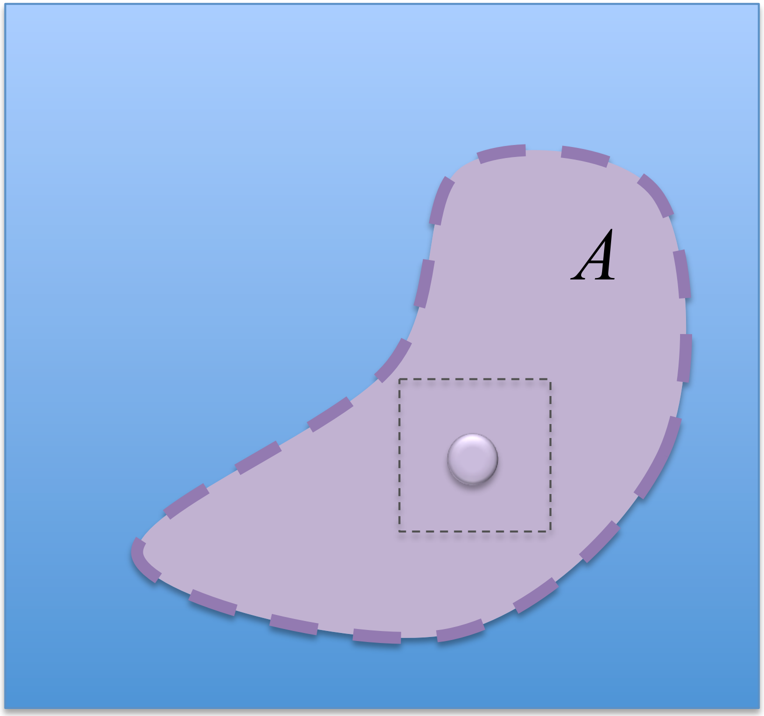
Interior of A

Int A = it will be verified that A.



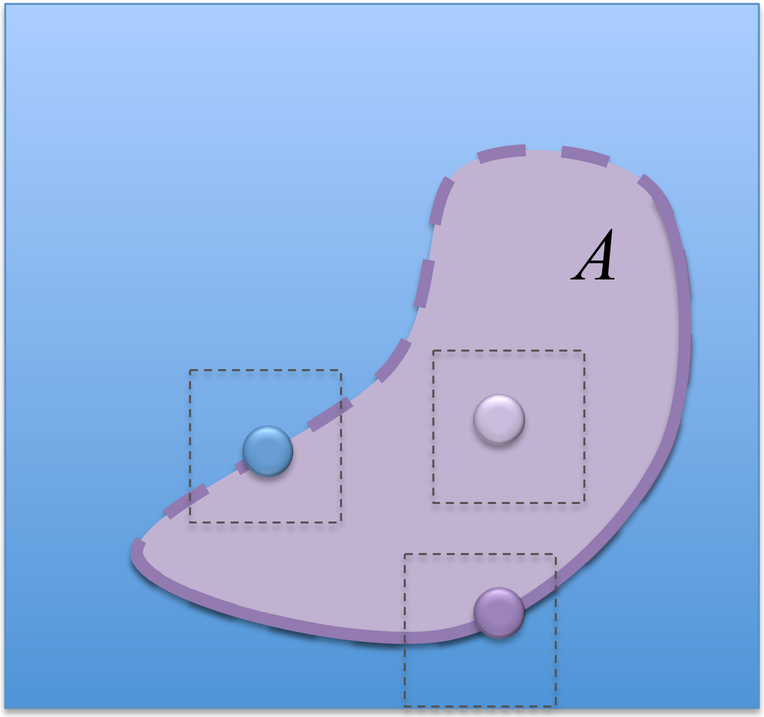
Open = Verifiable

A is open iff A entails that A will be verified.



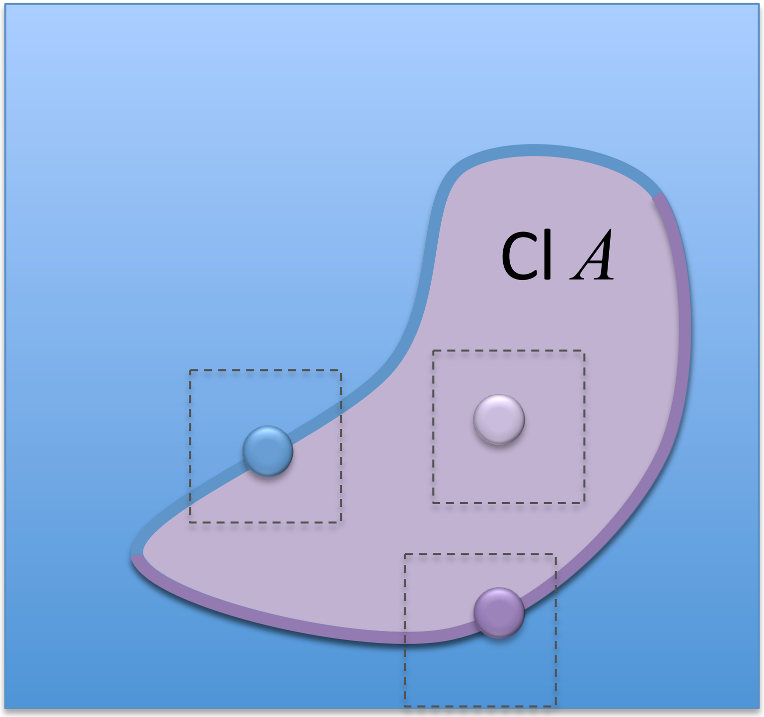
Closure of A

Cl A = A will never be refuted.



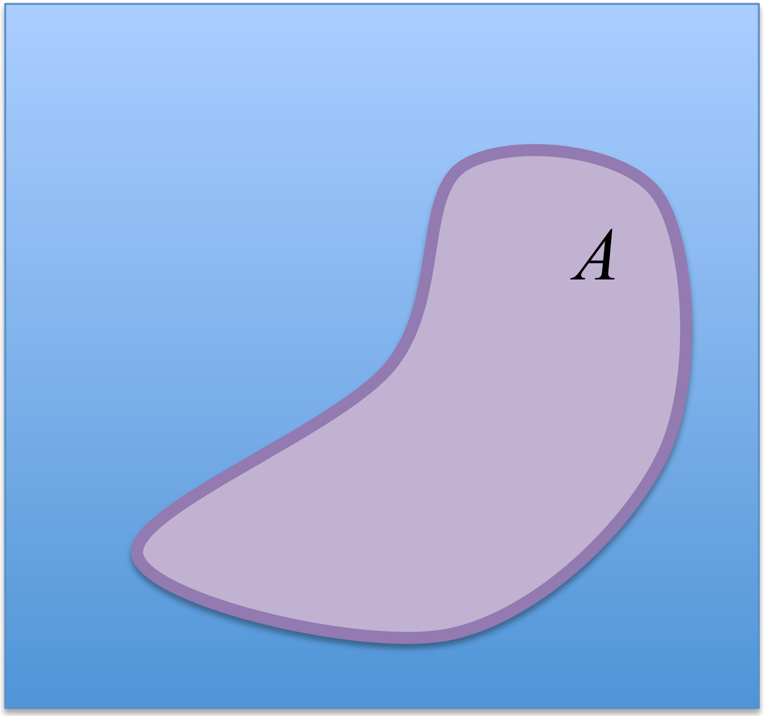
Closure of A

Cl A = A will never be refuted.



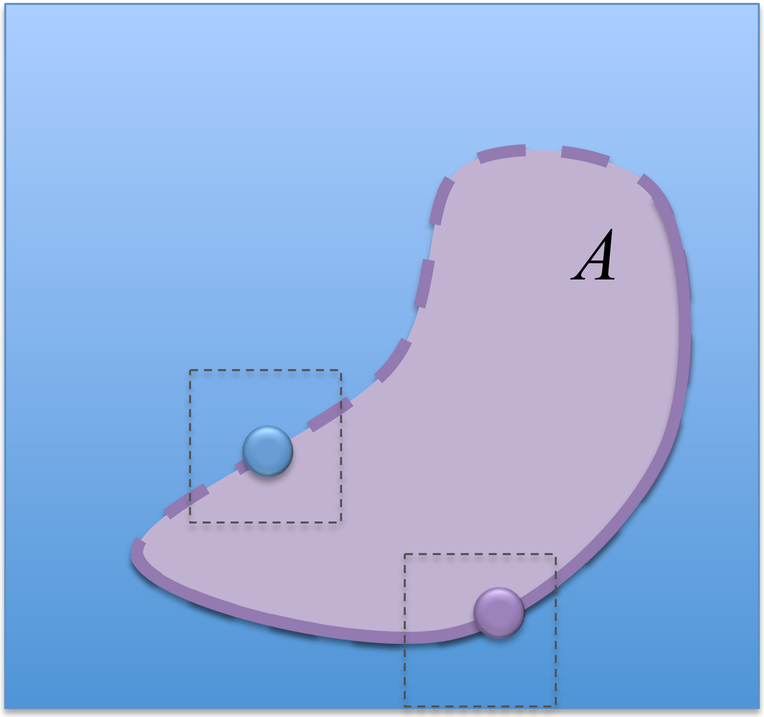
Closed = Refutable

A is closed iff not-A entails that A will be refuted.



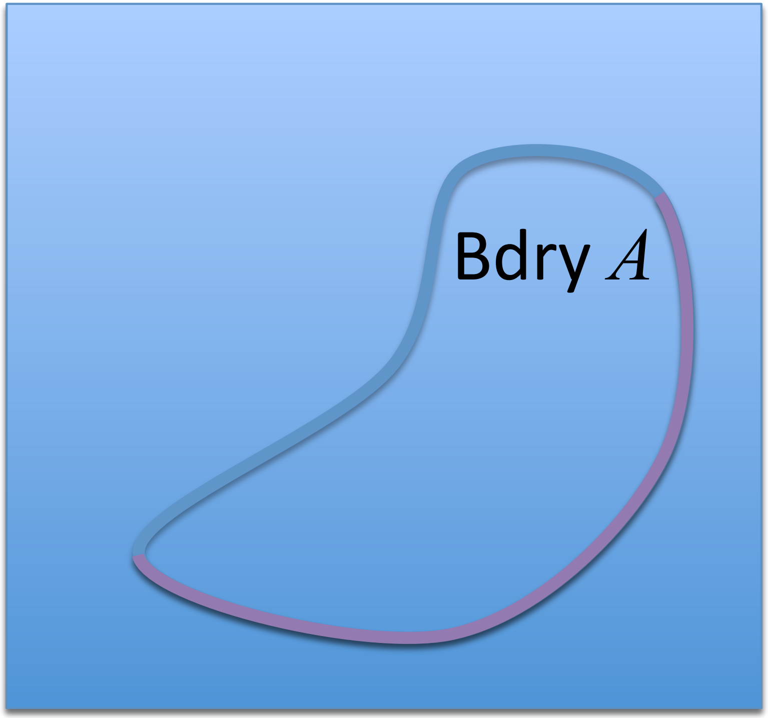
Boundary of A

$\mathrm{Bdry}\,A$ = "A will never be decided".



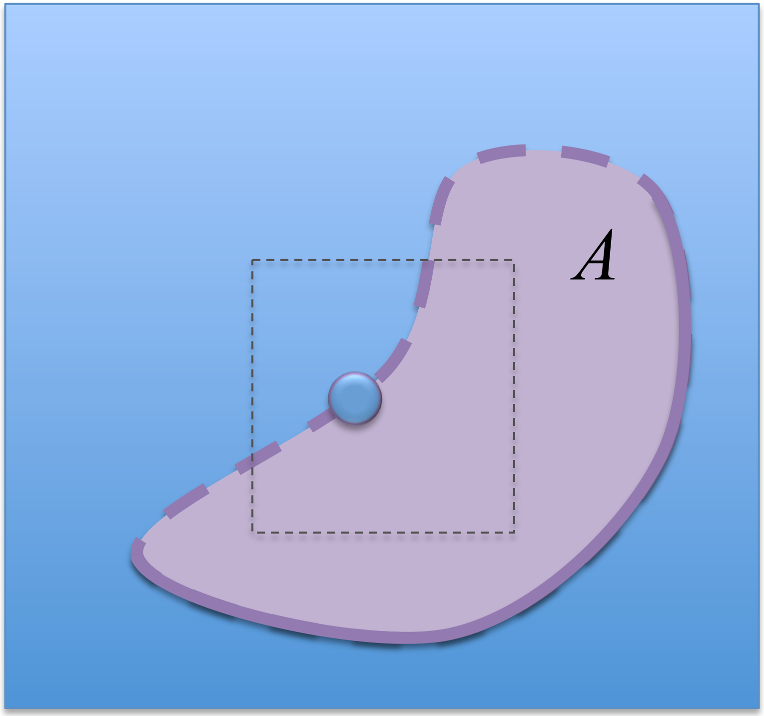
Boundary of A

$$\text{Bdry}A \;=\; \text{“}A\text{ will never be decided”}.$$



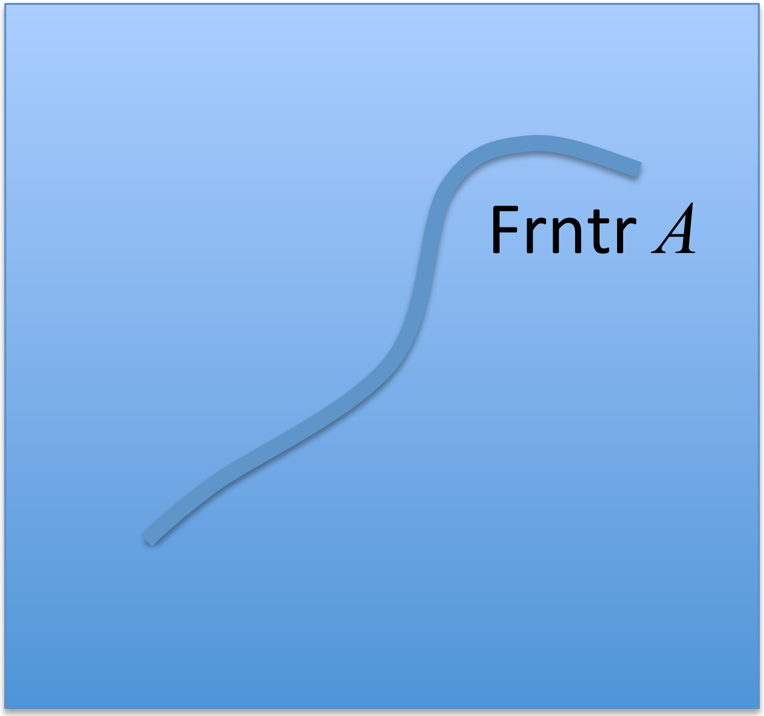
Frontier of A

Frntr A = A is false, but will never be refuted.



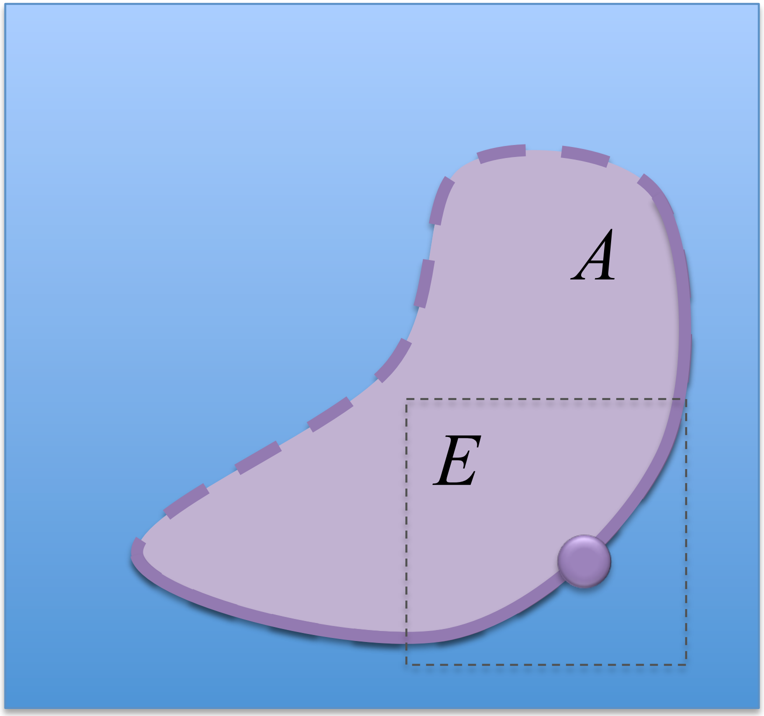
Frontier of A

Frntr A = A is false, but will never be refuted.



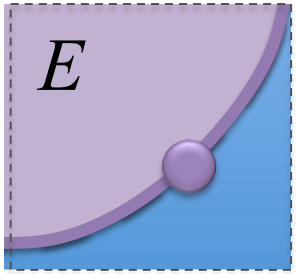
Locally Closed

A is locally closed iff A entails that A will become refutable.



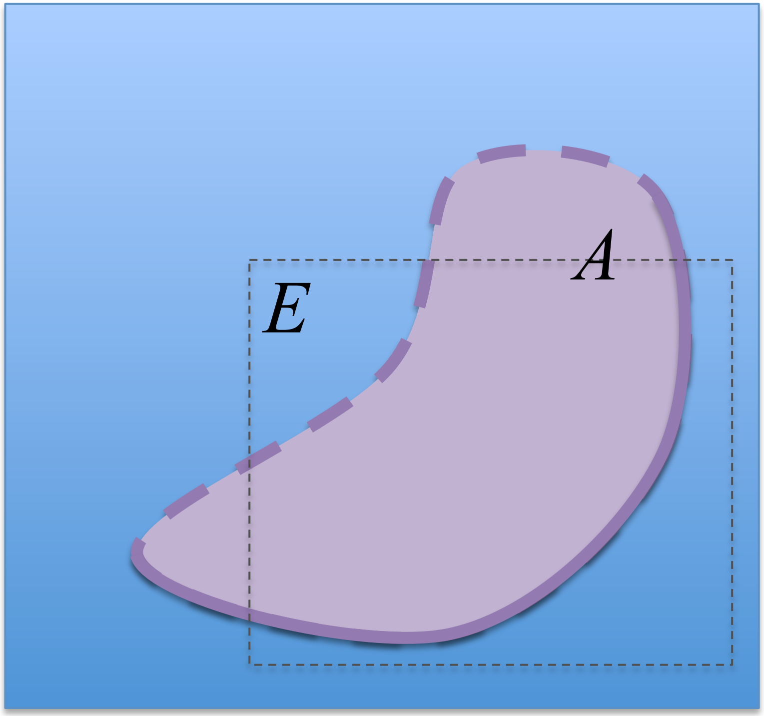
Locally Closed

A is locally closed iff A entails that A will become refutable.



Local Closure and Default Reasoning

E is a default reason for A iff
A is refutable but not refuted given E.

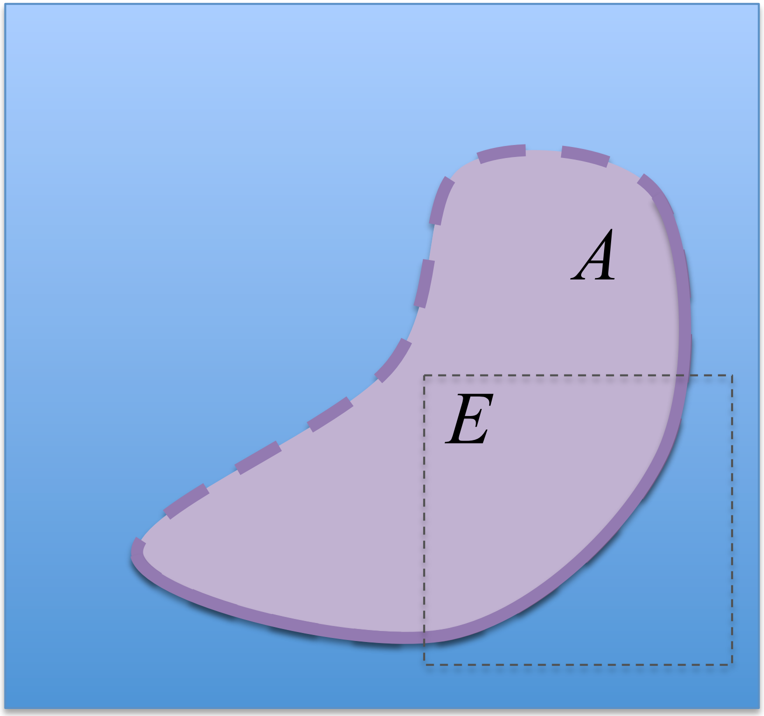


E is not a default reason for A.

Locally Closed

E is a default reason for A iff

A is refutable but not refuted given E.

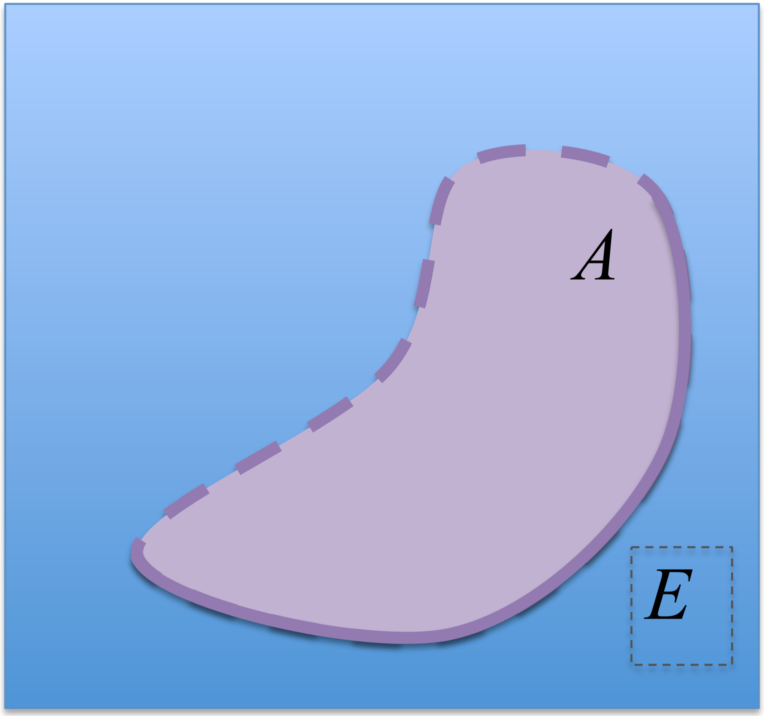


E is a default reason for A.

Locally Closed

E is a default reason for A iff

A is refutable but not refuted given E.



E refutes A.

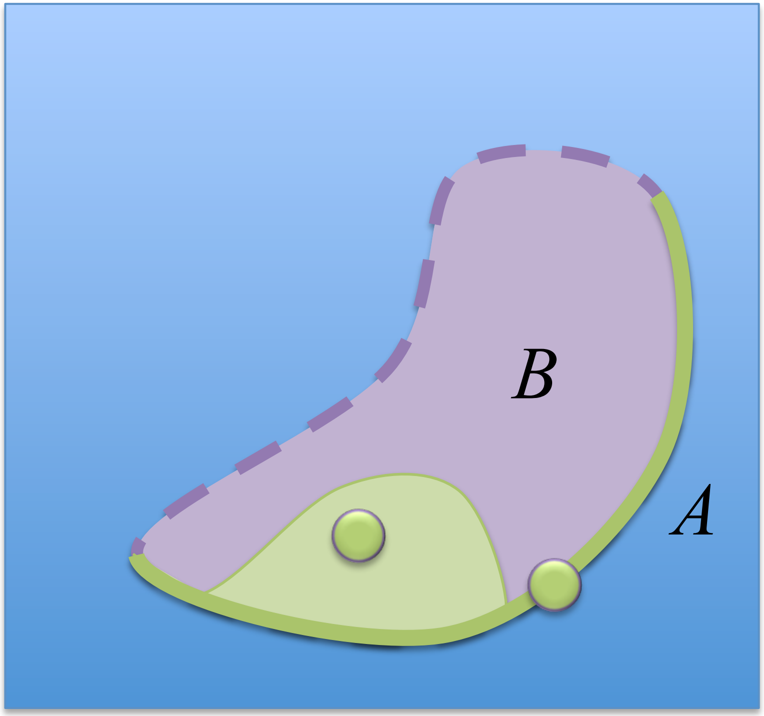
3. EMPIRICAL SIMPLICITY

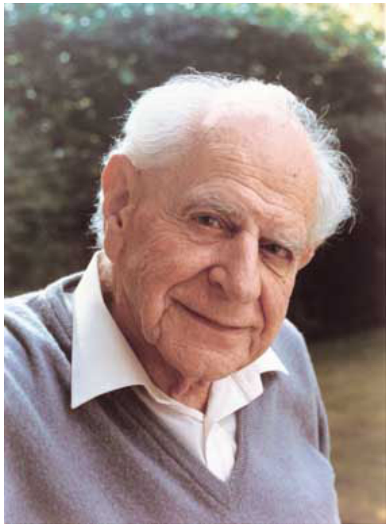
Simplicity = Specialization Preorder

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

iff all information compatible with A is compatible with B

iff all information refuting B also refutes A.

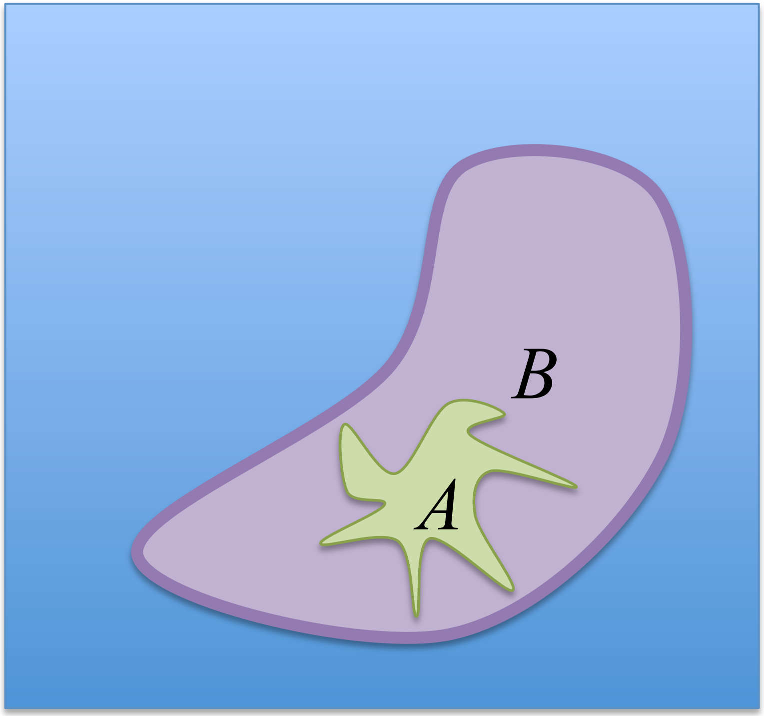


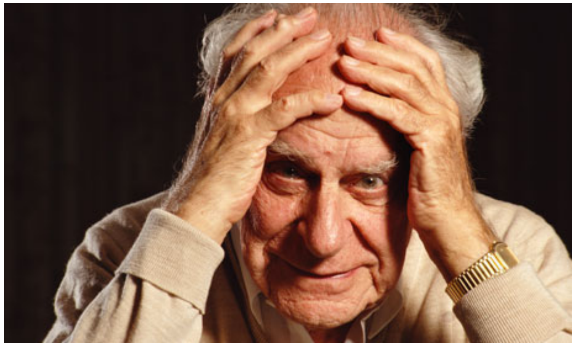


Sir Karl Popper

The “Tack-on” Objection

- Adding complex principles to a simple theory doesn’t make it simpler (Glymour 1980).

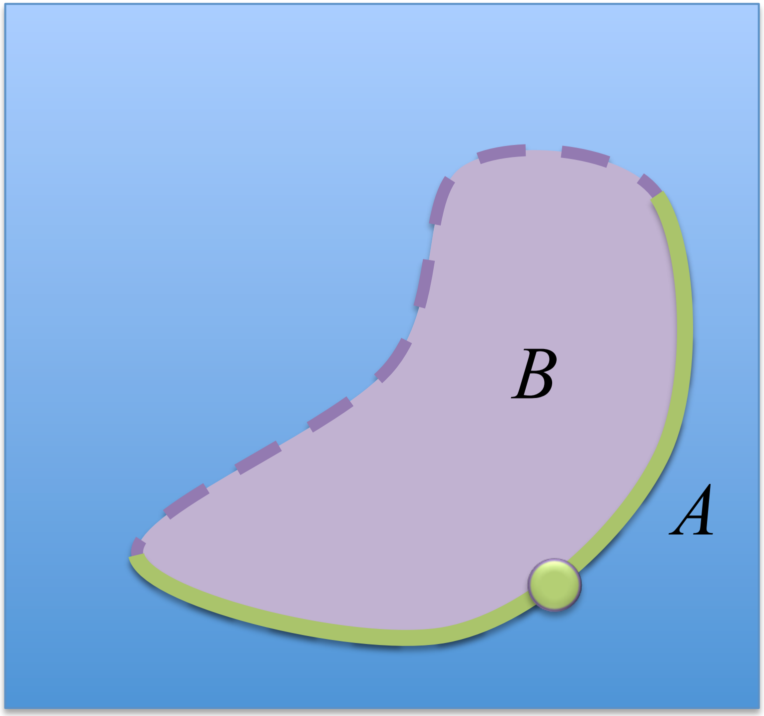




Empirical Simplicity

$A \lhd B$ iff $A \cap \mathsf{Frntr} B \neq \varnothing$.

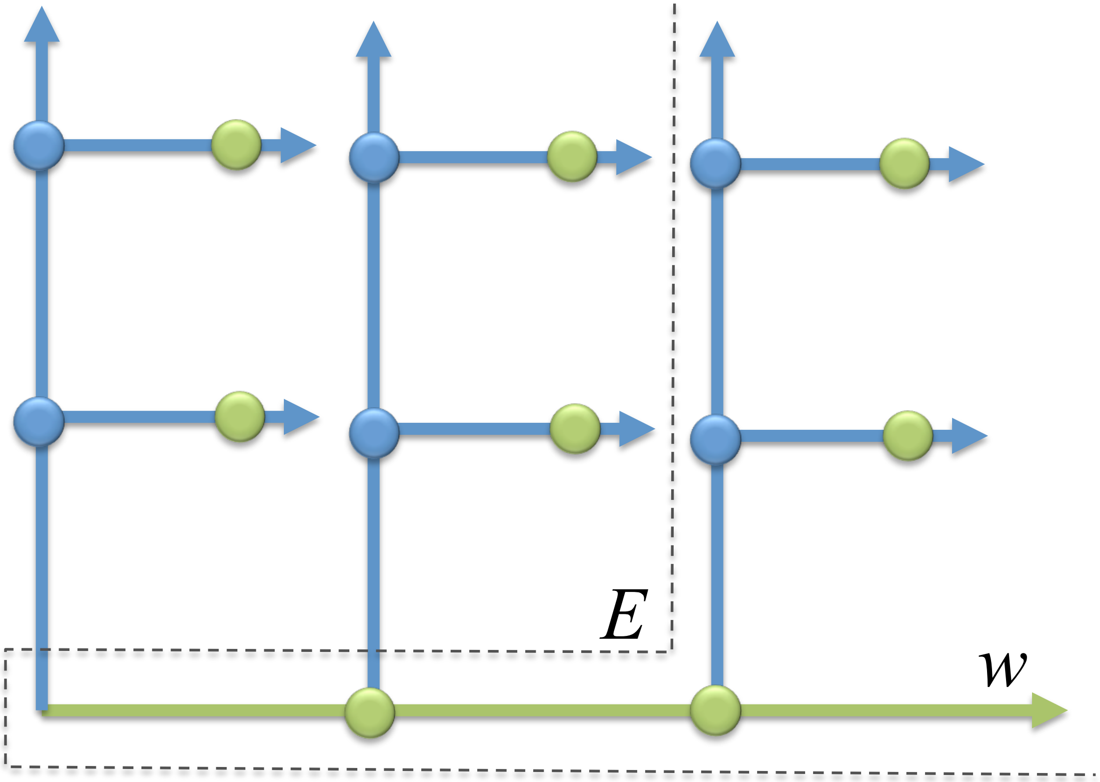
iff A entails that B is false but will never be refuted.

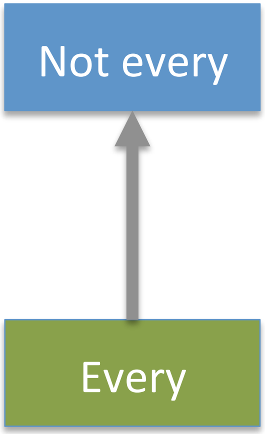


Example: The Baire Space

$\mathcal{Q} =$ Will every outcome be green?

$\mathcal{I} =$ observation histories.





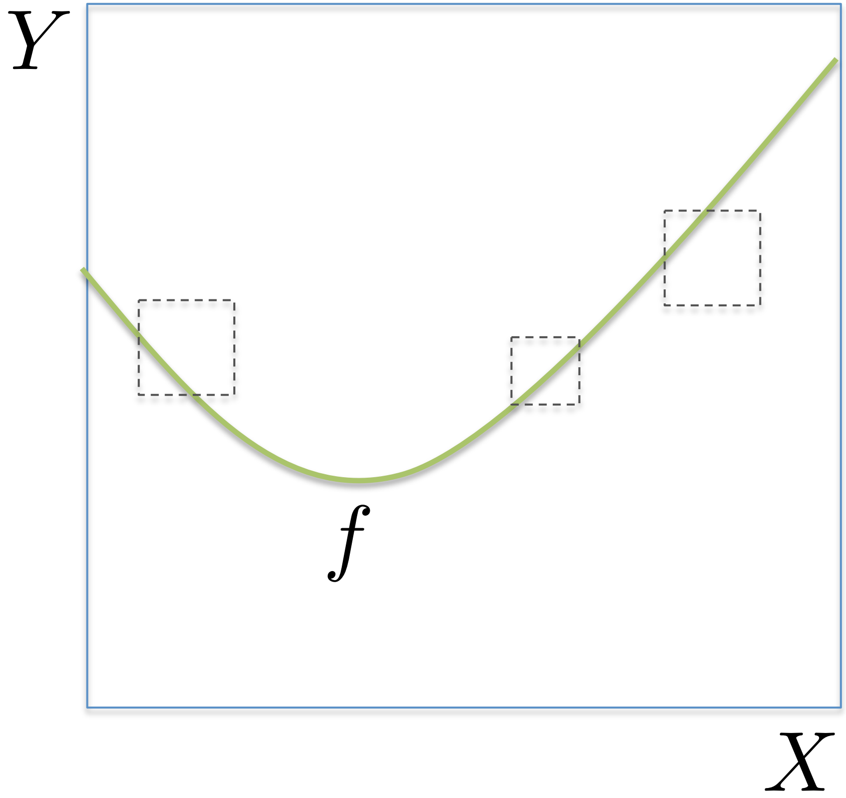
Not every — open

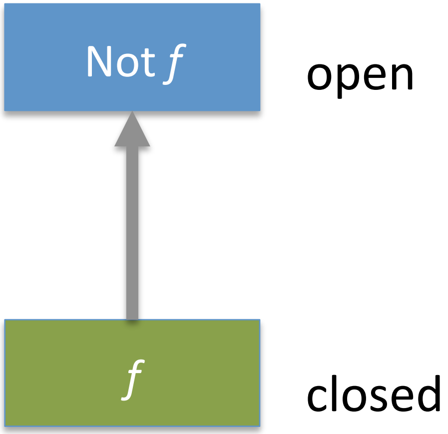
Every — closed

Example: Continuous Laws

$\mathcal{Q} =$ Does $Y = f(X)$?.

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





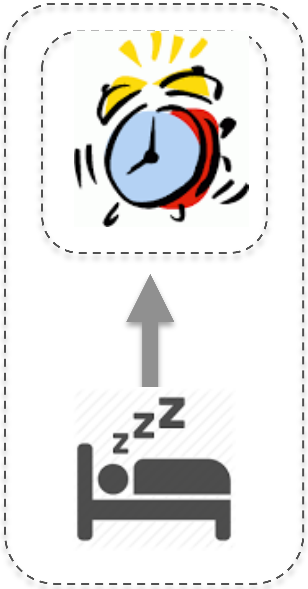
Not f open

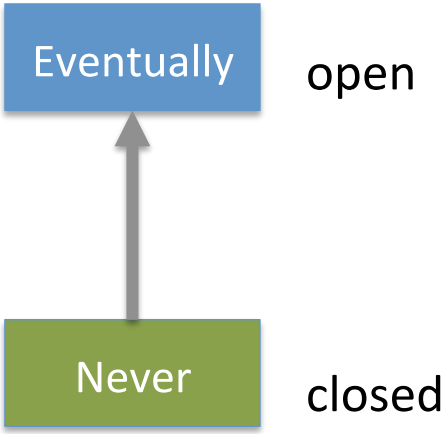
f closed

Example: Sierpinski Space

$\mathcal{Q} =$ Will it ever ring?

$\mathcal{I} =$ alarm or no alarm yet.

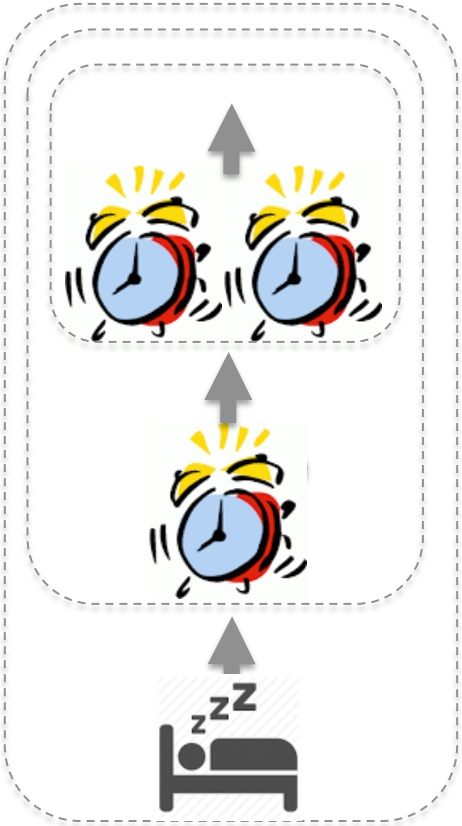


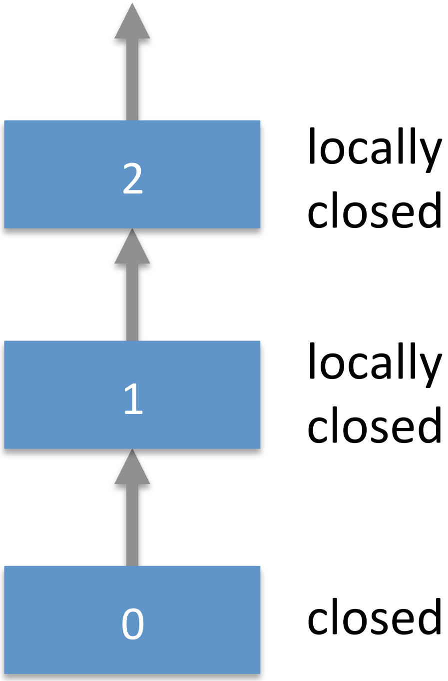


Example: Upward Topology

$\mathcal{Q} =$ How many alarms?

$\mathcal{I} =$ cumulative alarms.





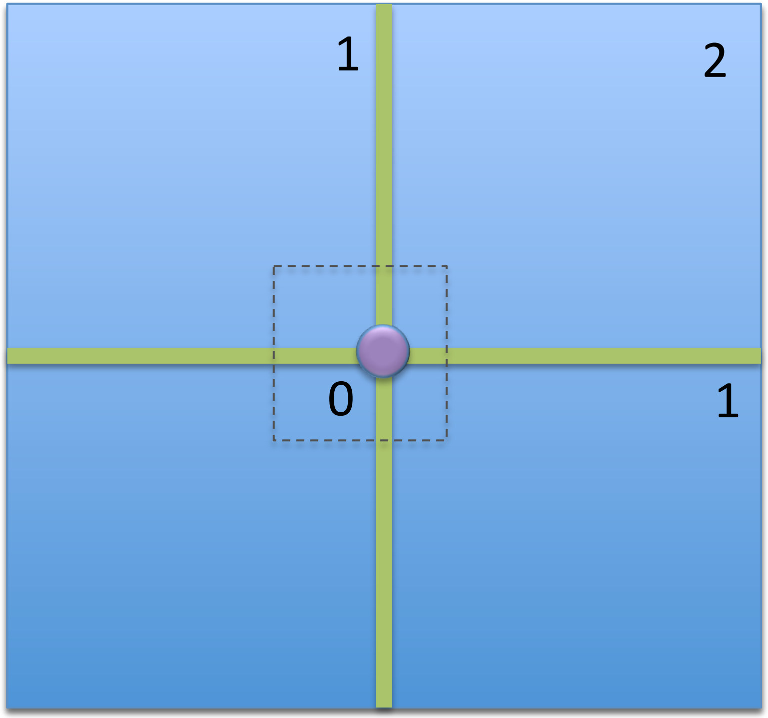
locally closed

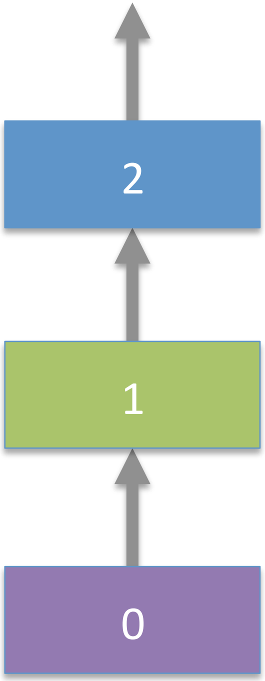
locally closed

closed

Example: Euclidean Metric Topology

$\mathcal{Q} =$ How many parameters are free?





locally closed

locally closed

closed

Example: Quantitative Laws

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

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

degree

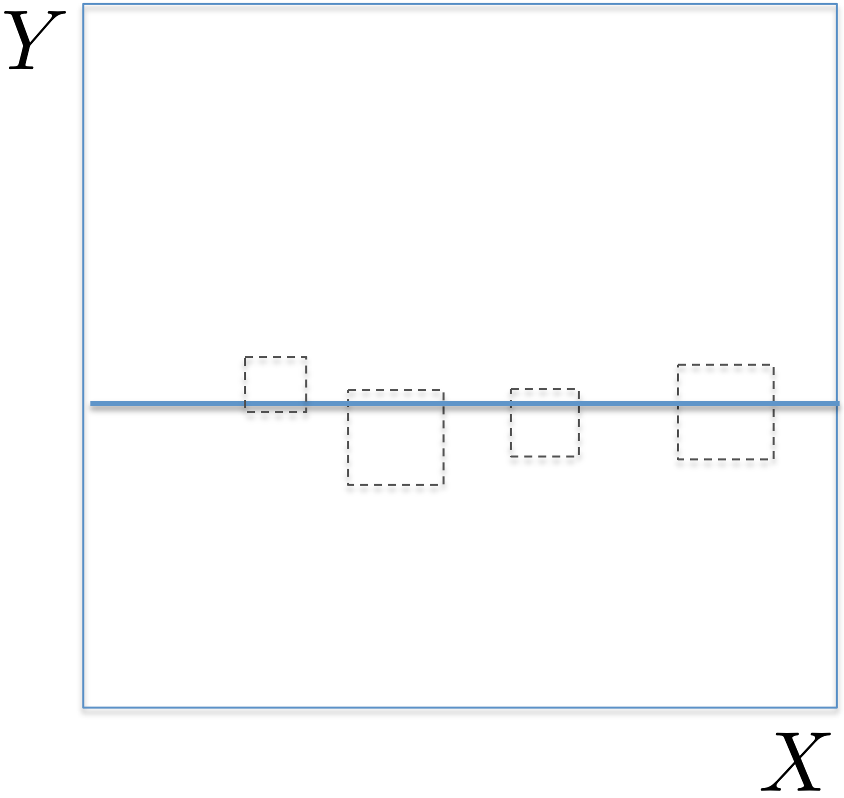
$$Y \;=\; \textstyle\sum_{i=0}^{N} a_i X^i.$$

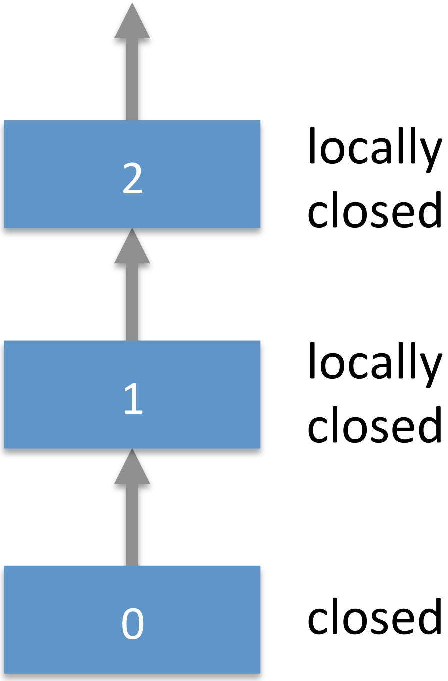
$$\begin{aligned}
Y &= a_0. \\
Y &= a_0 + a_1 X. \\
Y &= a_0 + a_1 X + a_2 X^2. \\
&\vdots
\end{aligned}$$

Example: Quantitative Laws

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

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





2 locally closed

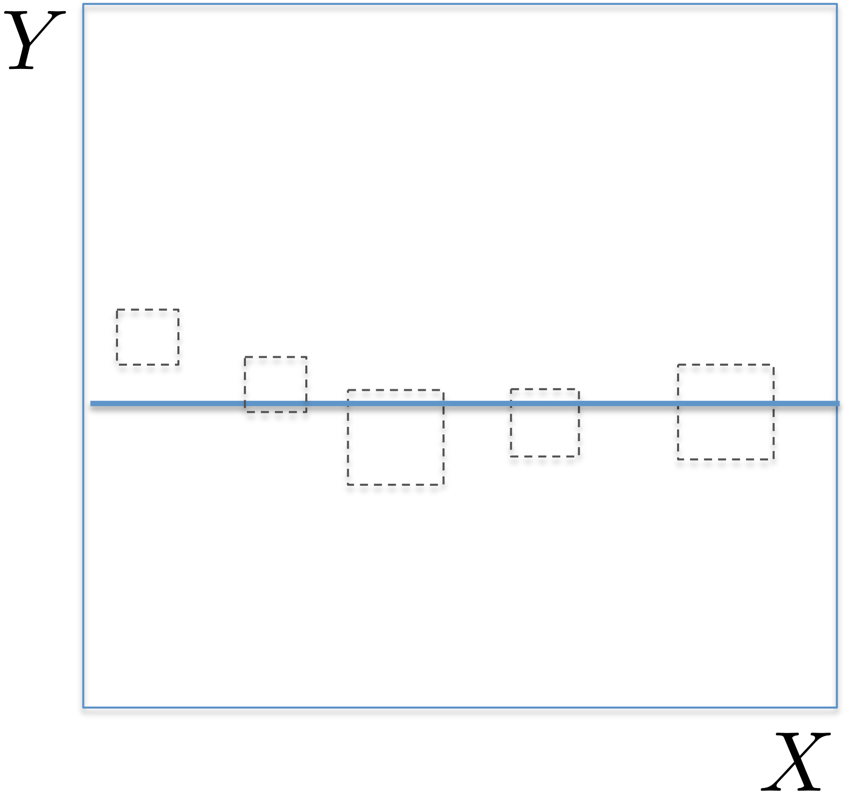
1 locally closed

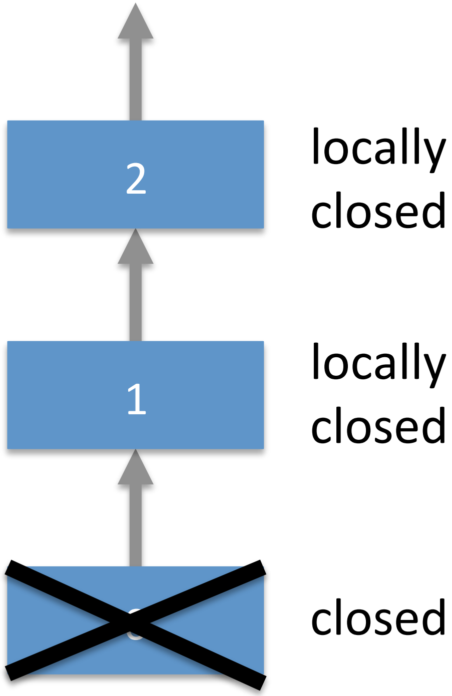
0 closed

Example: Quantitative Laws

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

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





locally closed

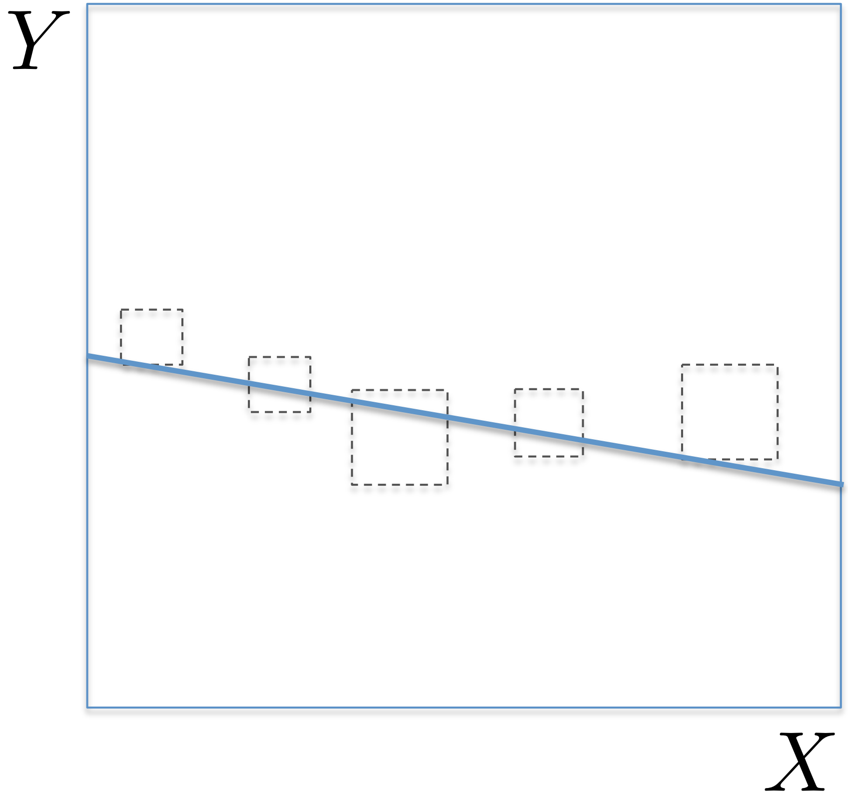
locally closed

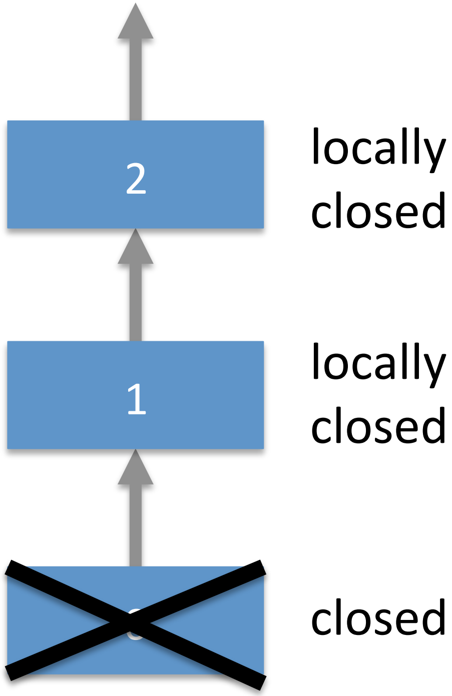
closed

Example: Quantitative Laws

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

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





locally closed

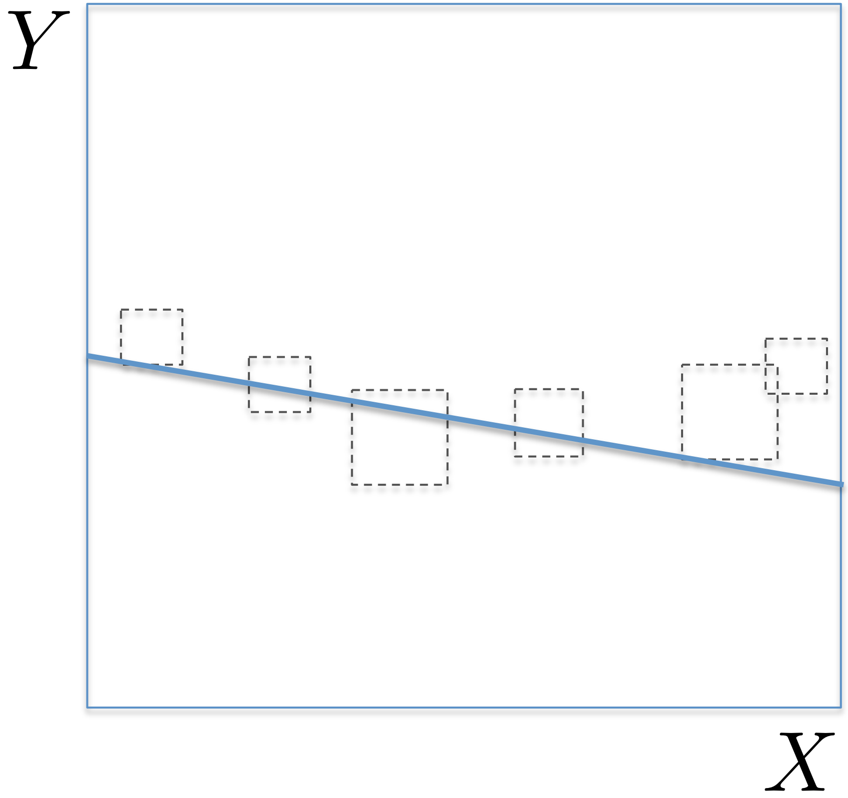
locally closed

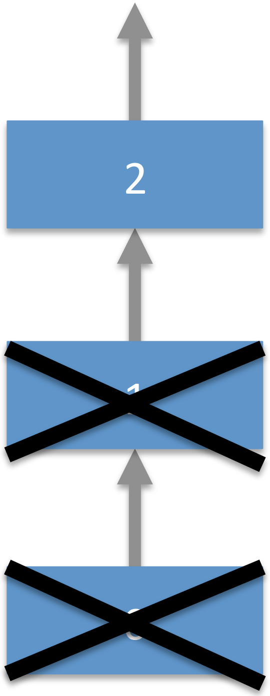
closed

Example: Quantitative Laws

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

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





locally closed

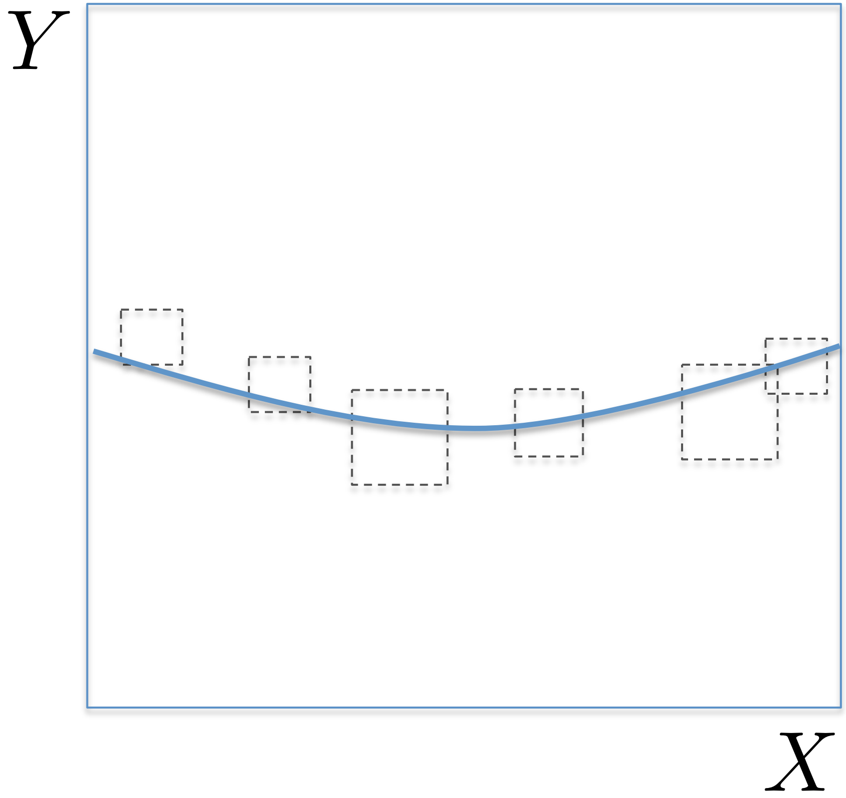
locally closed

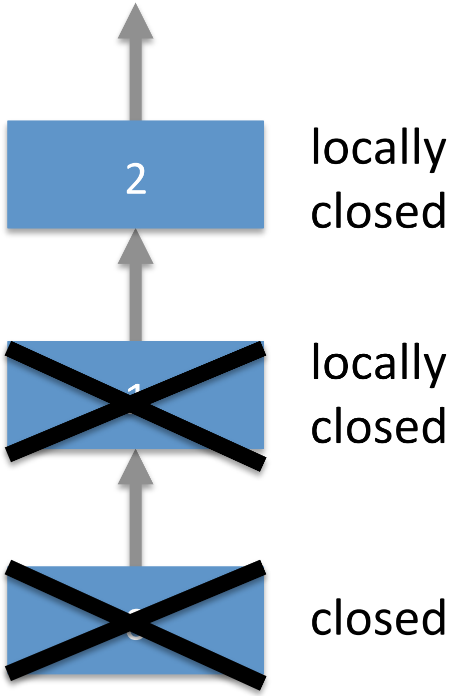
closed

Example: Quantitative Laws

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

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





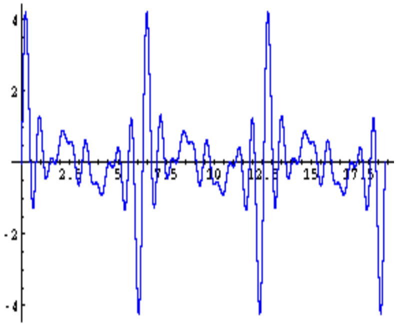
Example: Competing Paradigms

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

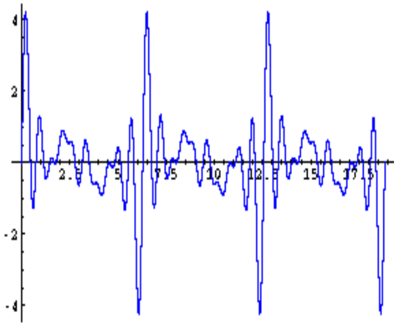
Polynomial paradigm

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

degree

Trigonometric polynomial paradigm

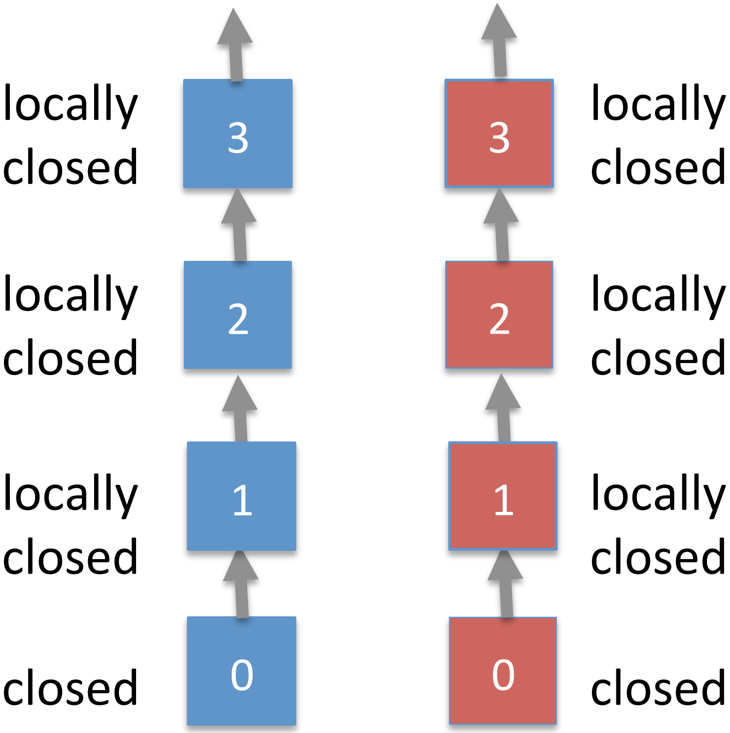
$$Y = \sum_{i=0}^{N} a_i \sin(iX) + b_i \cos(iX).$$



Example: Competing Paradigms

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

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



locally closed

locally closed

locally closed

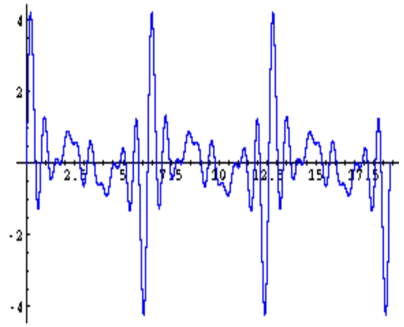
closed

locally closed

locally closed

locally closed

closed



4. INDUCTIVE METHODS

Reasoning

Deductive

- Monotonic

Reasoning

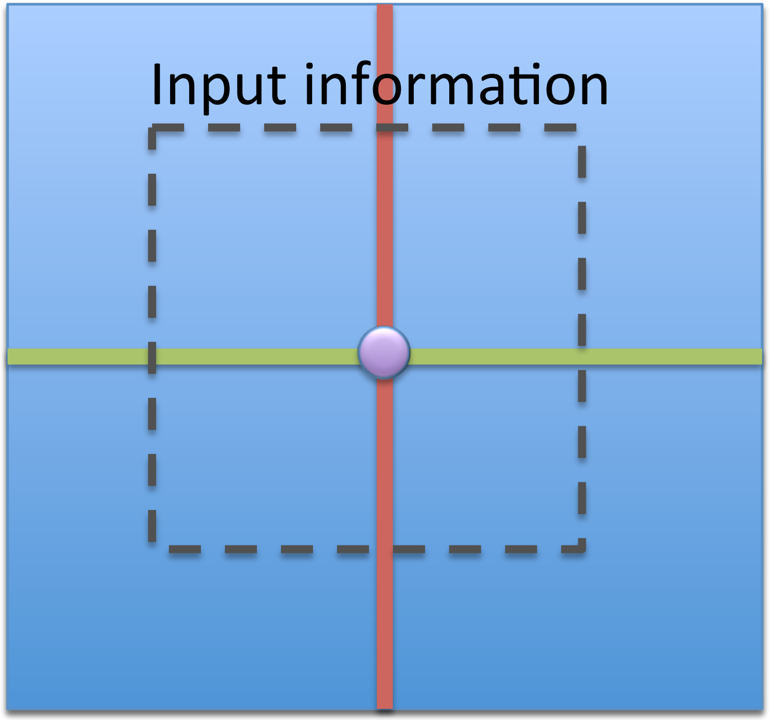
Deductive

- Monotonic

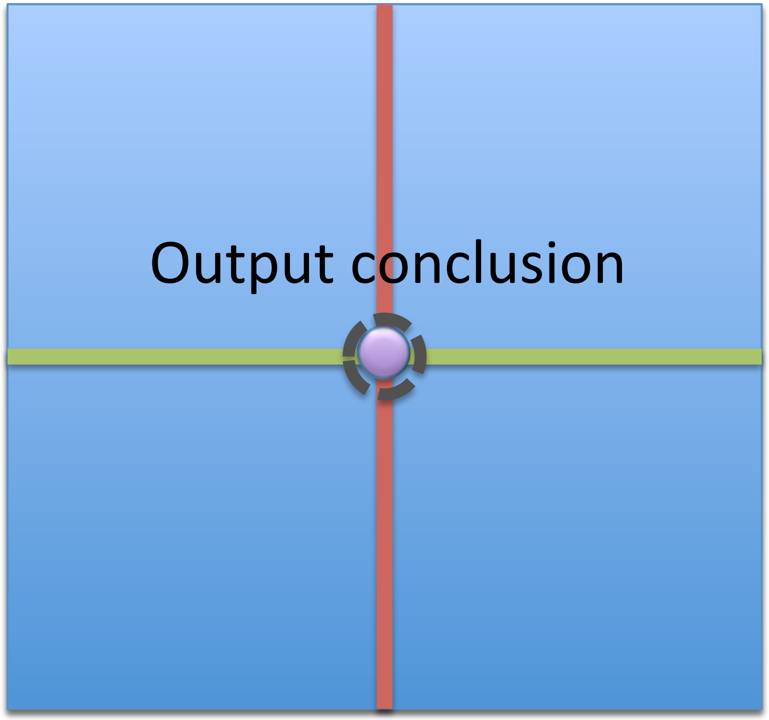
Inductive

- Non-monotonic

Inductive Inference

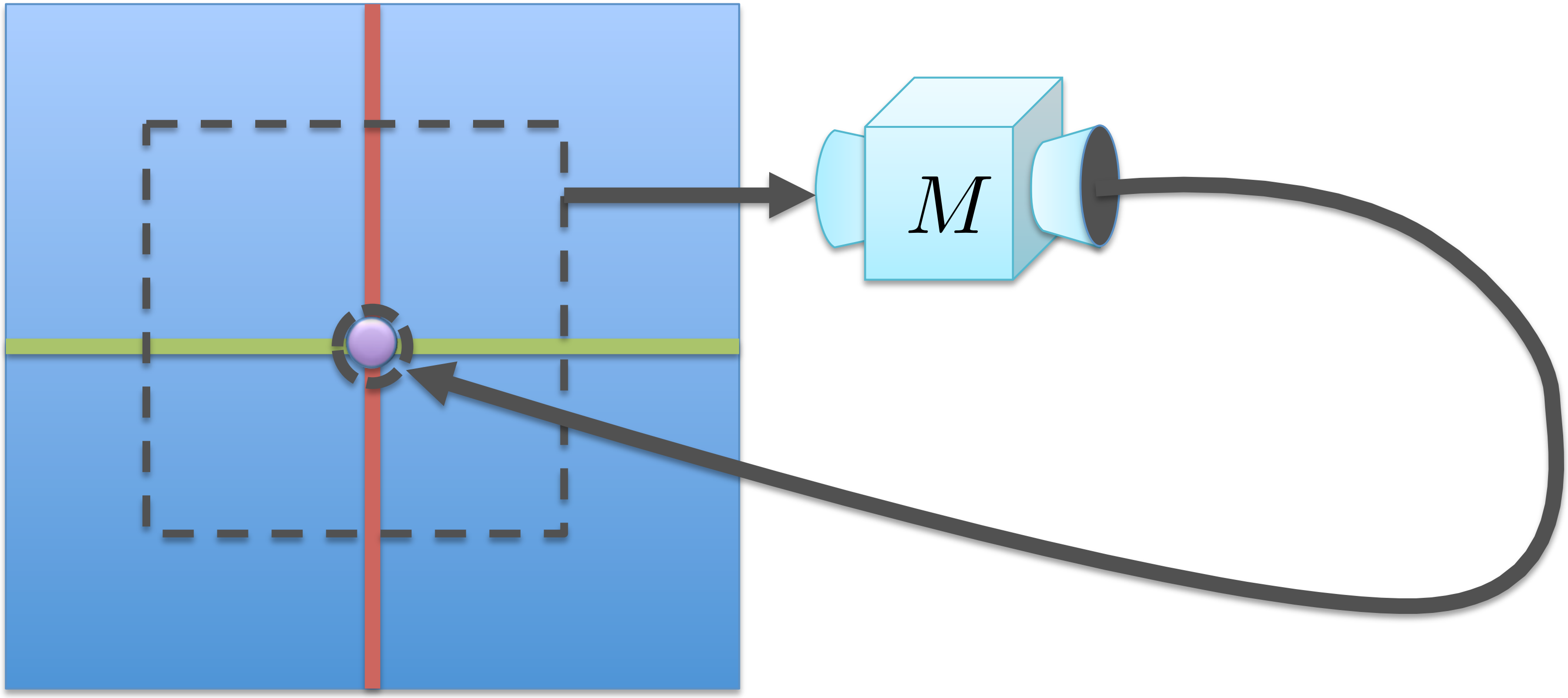


Inductive Inference



Inductive Methods

Information in, relevant response out.

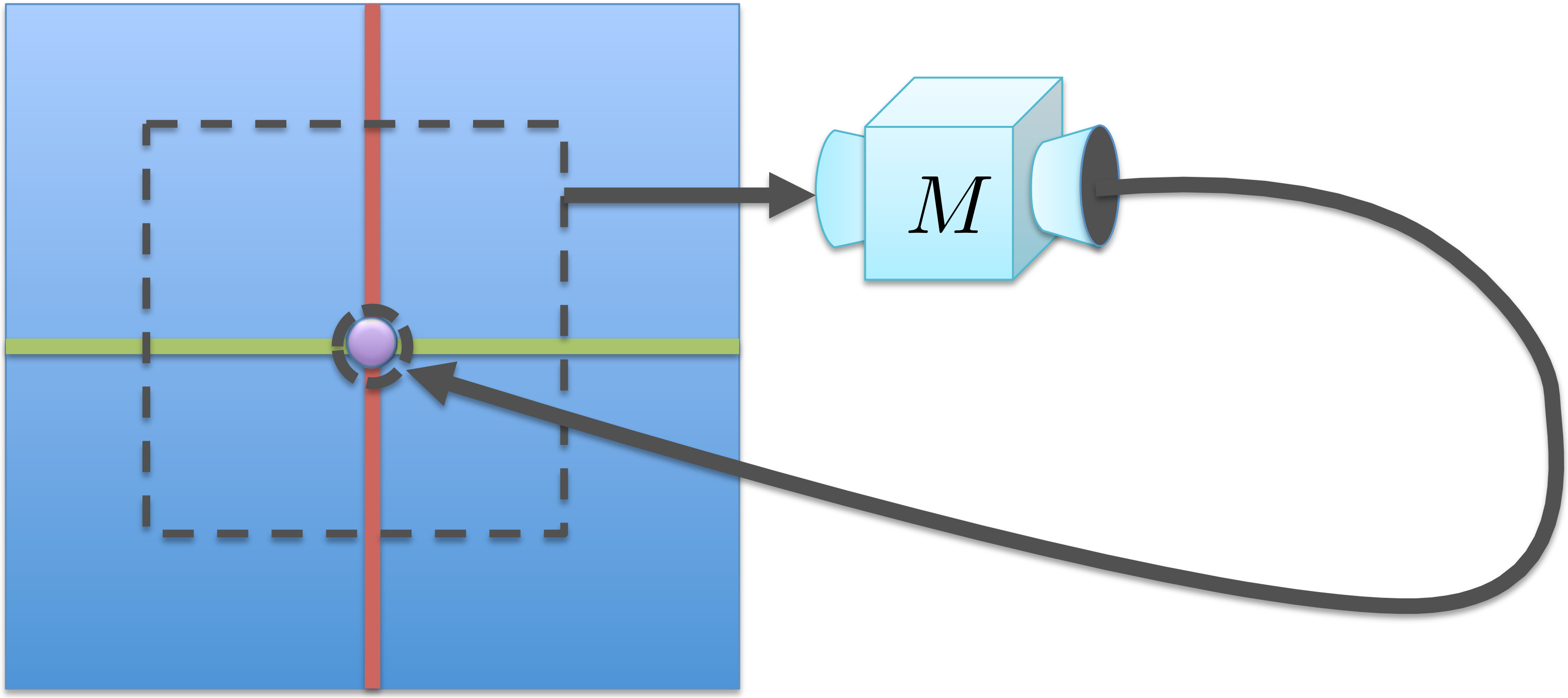


Solution in the Limit

M solves a problem in the limit iff

each world w presents information such that

M produces the true answer in w on any further information in w.

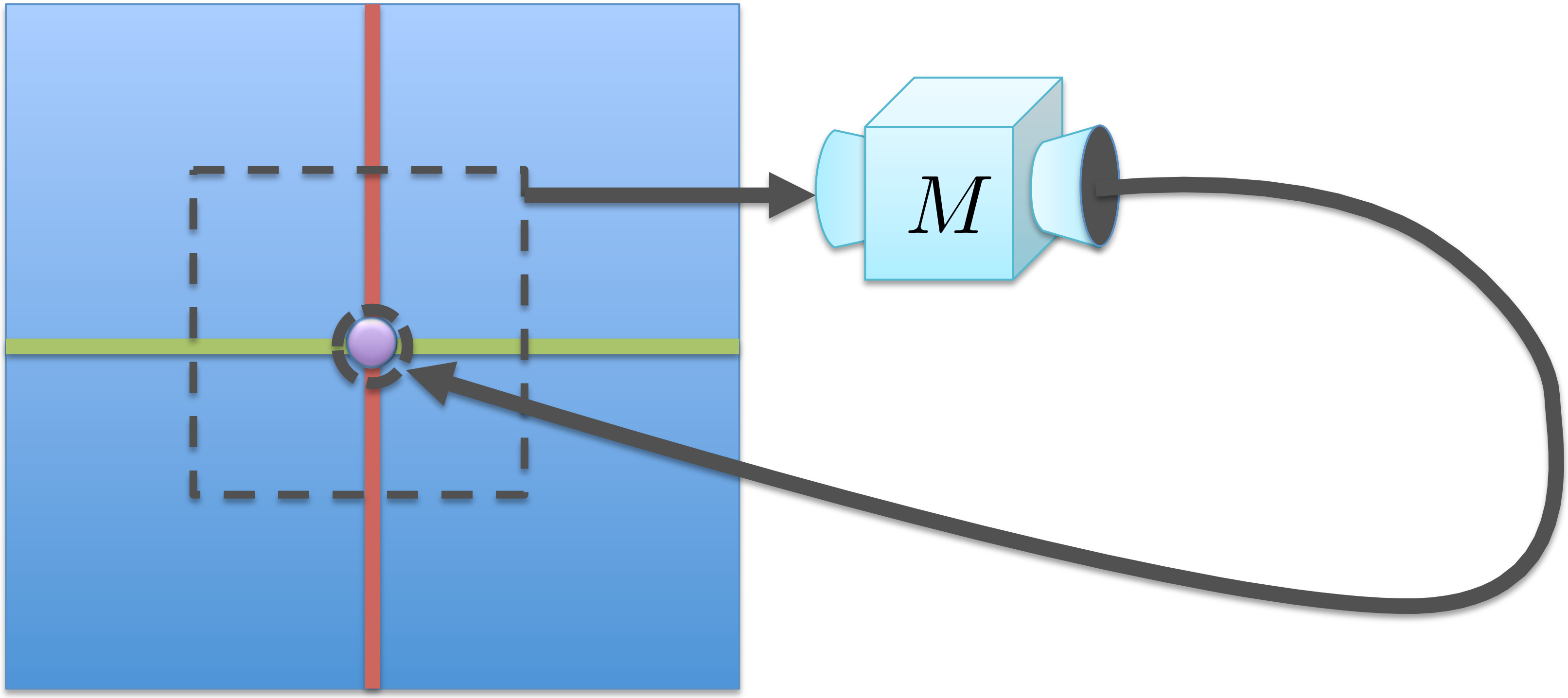


Solution in the Limit

M solves a problem in the limit iff

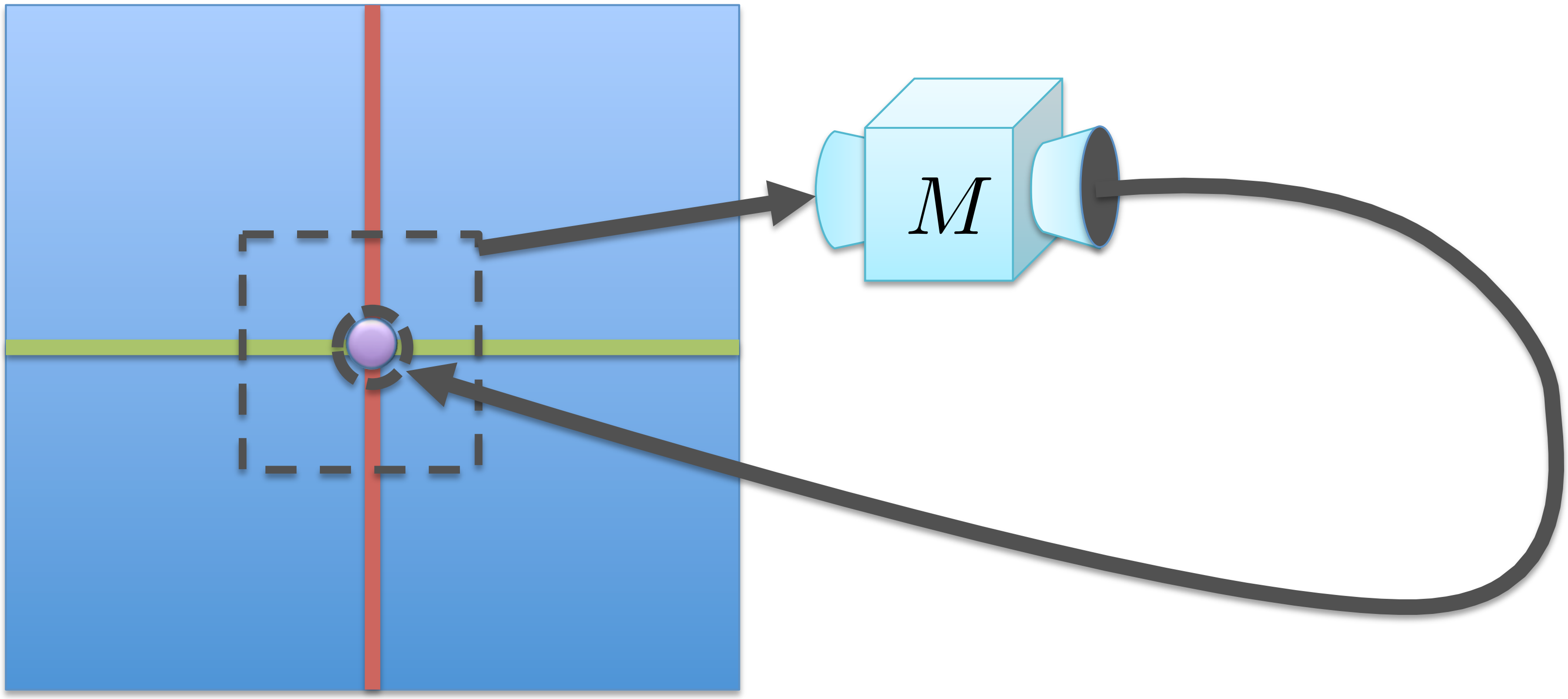
each world w presents information such that

M produces the true answer in w on any further information in w.



Solution in the Limit

M solves a problem in the limit iff
each world w presents information such that
M produces the true answer in w on any further information in w.

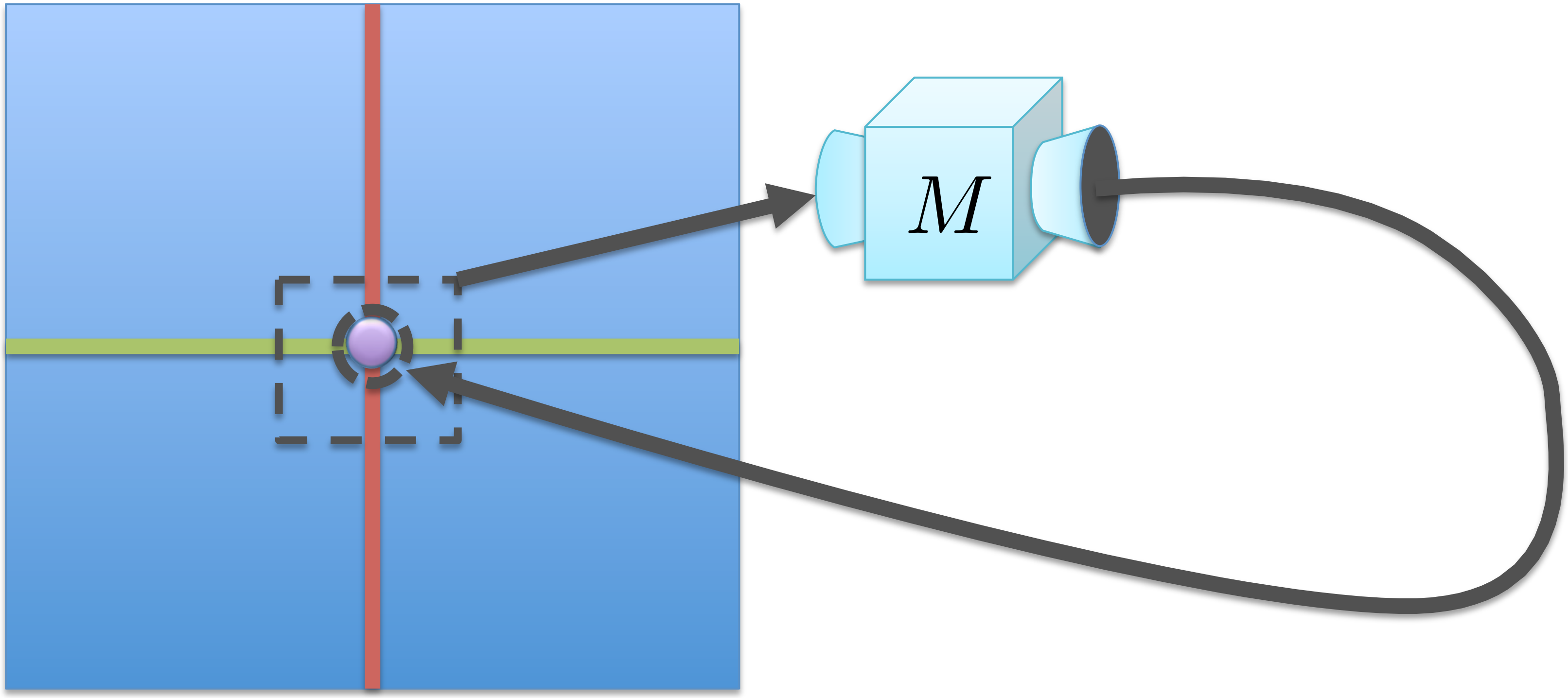


Solution in the Limit

M solves a problem in the limit iff

each world w presents information such that

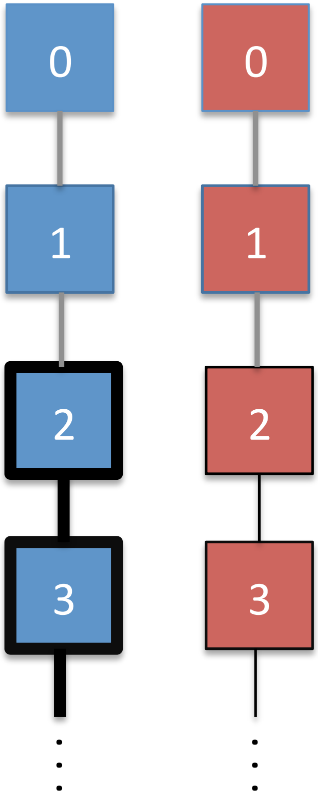
M produces the true answer in w on any further information in w.



5. OCKHAM’S RAZOR

Ockham’s Razor

- Output a simplest relevant response given E.
  - Allows for suspension of judgment.
  - Makes sense for infinite descending 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.

Equivalence

Proposition. In natural problems:

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

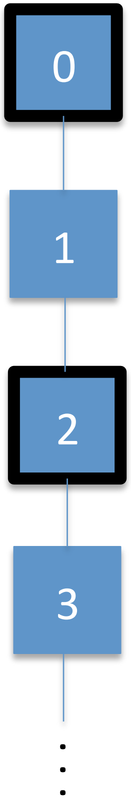
=

=

?

Patience

- Never rule out a simplest relevant response given E.
  - Says that Ockham's razor is the only reason for inductive leaps beyond experience.
  - Logically independent of Ockham's razor.



Patient but
not Ockham

Patience

- Never rule out a simplest relevant response given E.
  - Says that Ockham's razor is the only reason for inductive leaps beyond experience.
  - Logically independent of Ockham's razor.

Ockham but
not patient

🟦 🟥

Error patience

- In arbitrary world w, never output a relevant response that rules out all answers as simple as A_w.

Equivalence

- **Proposition:** Error patience is equivalent to patience.

6. OCKHAM’S RAZOR JUSTIFIED

Reliability

Deductive

- Converge to the truth directly

Reliability

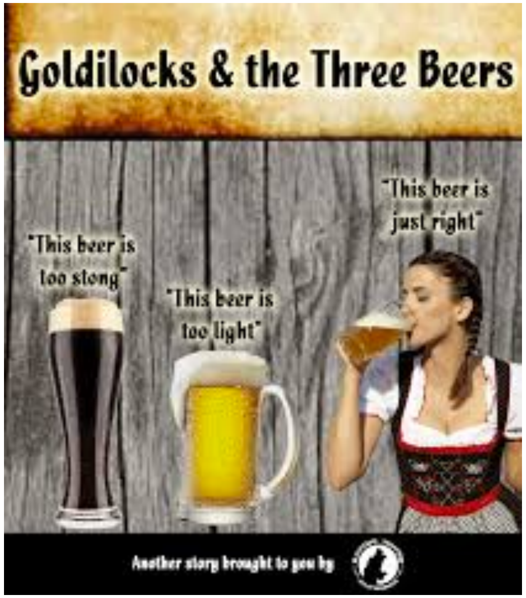
Deductive

- Converge to the truth directly

Inductive

- Converge to the truth indirectly

Goldilocks Philosophy



Straight

Too strong!

Just right?

Arbitrarily crooked

Too weak!

Straightest Possible Convergence

Straight

Too strong!

Couldn’t be straighter

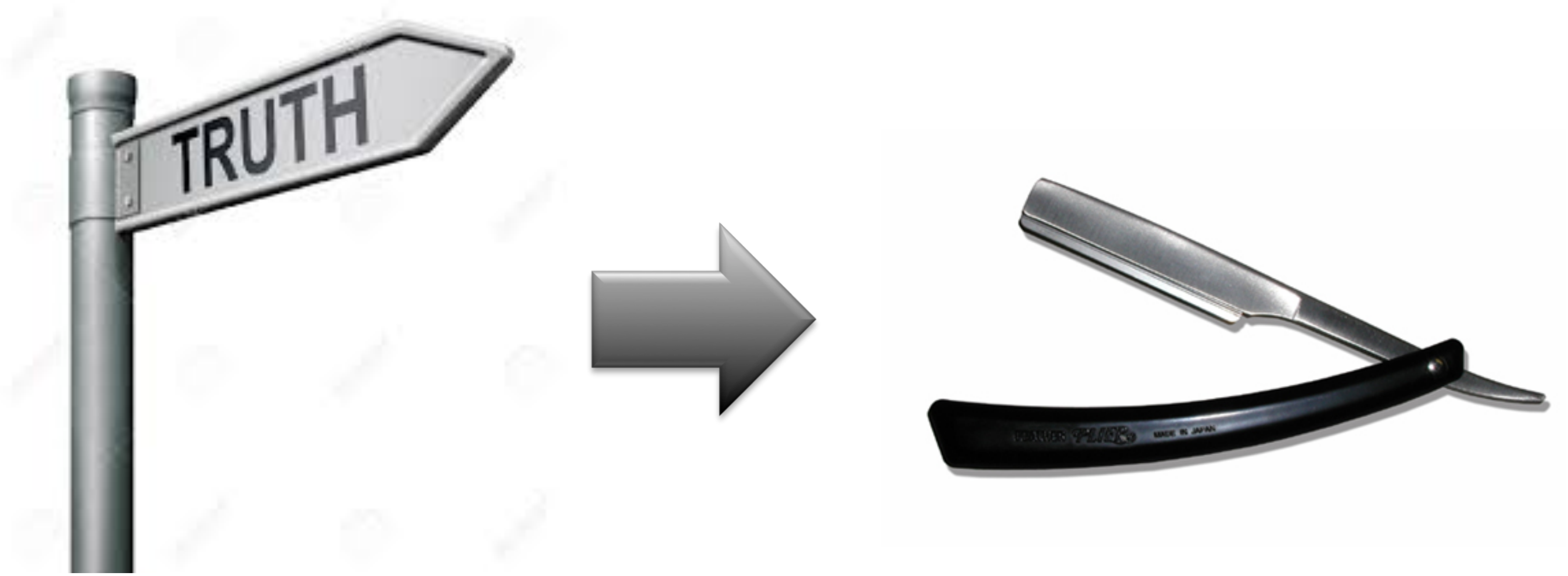
Just right!

Arbitrarily crooked

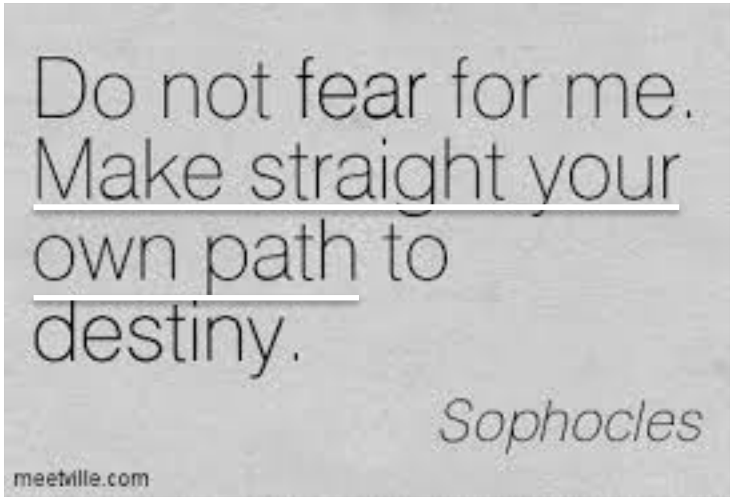
Too weak!

Thesis

- Ockham's razor is necessary for straightest convergence to the truth.



The Straightest Path



Do not fear for me. Make straight your own path to destiny.

Sophocles

meetville.com

sperience

Trust in the LORD with all your heart
and lean not on your own understanding;
in all your ways acknowledge him,
and he will make your paths straight.

Proverbs 3:5-6

اَهْدِنَا الصِّرَاطَ الْمُسْتَقِيمَ ٦

GUIDE US TO THE STRAIGHT PATH

AL-FATIHAH : 6

Fools dwelling in darkness, but thinking themselves wise, go round and round, by tortuous paths, like the blind led by the blind.

Katha Upanishad

Two Departures from Straightness

Course-reversals

Cycles

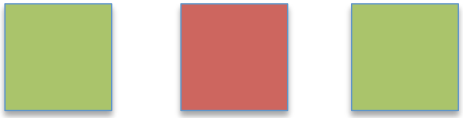
Doxastic Reversal Sequence

- A finite sequence of relevant responses in which each entry contradicts its predecessor.

🟩 🟥 🟦

Doxastic Cycle Sequences

- A reversal sequence whose terminal entry entails its first entry.



Straightest Convergence

- Solution M is reversal optimal iff:

every solution can be forced to produce each reversal sequence produced by M.

Straightest Convergence

- Solution M is cycle optimal iff:

every solution can be forced to produce each cycle sequence produced by M.

Main Result 1

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

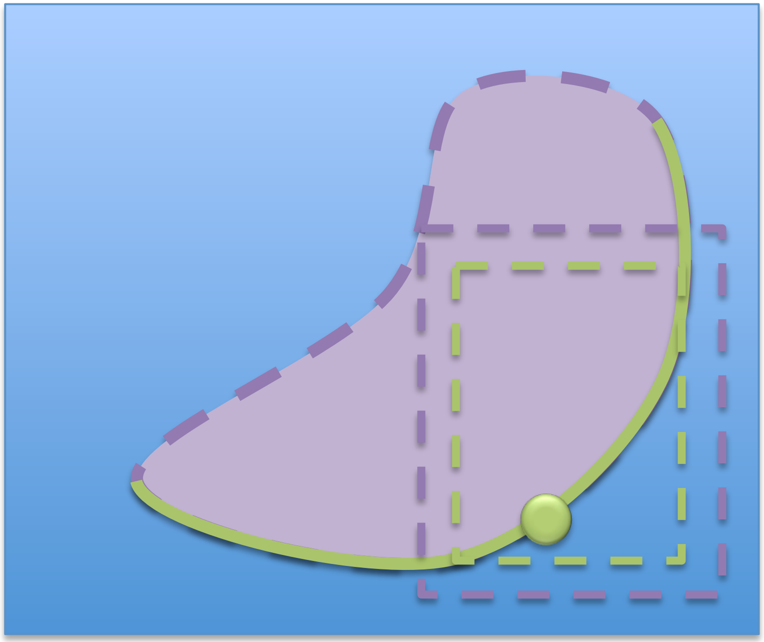
Main Result 2

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

The Idea

The Idea

🟪 🟩



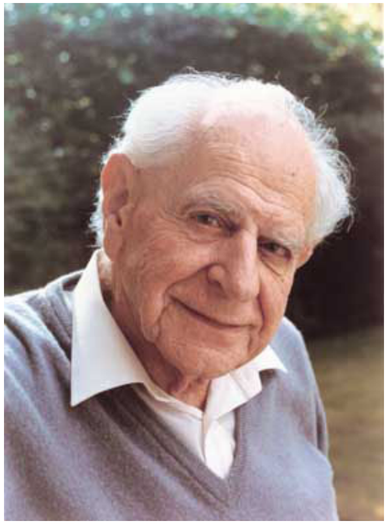
The Idea

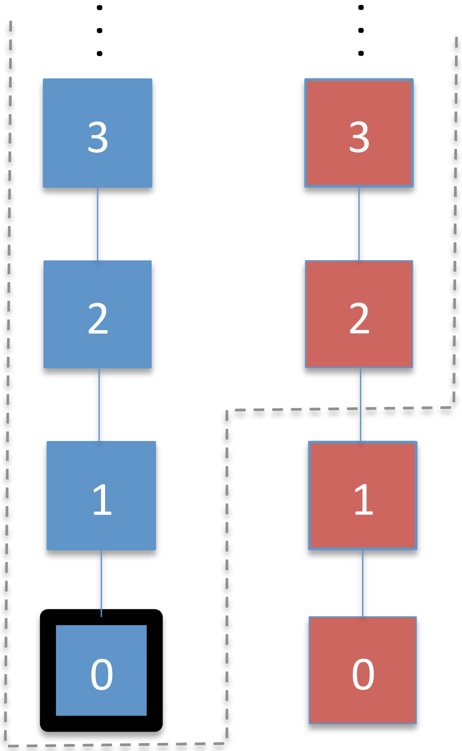
Main Result 3

- We can characterize the solvable problems that have reversal optimal solutions.
- The characterization depends on the basis, so it is not topological.

No Reversal Optimal Solution

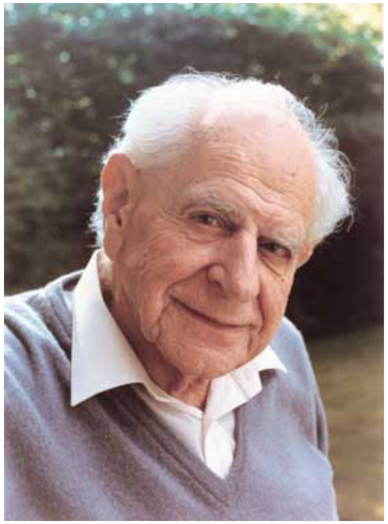
"Popper": choose the paradigm with fewer free parameters.

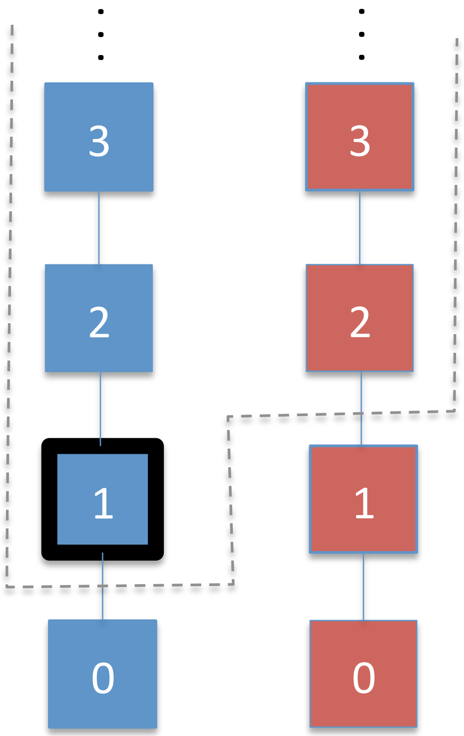




No Reversal Optimal Solution

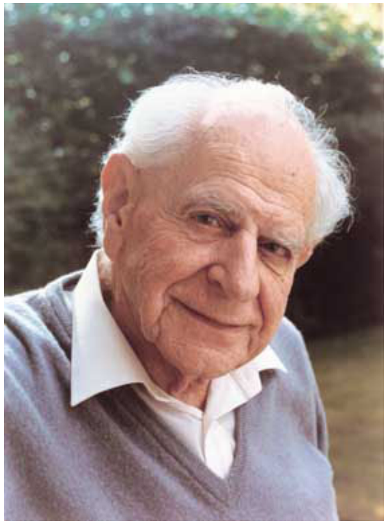
"Popper":
choose the paradigm with fewer free parameters.

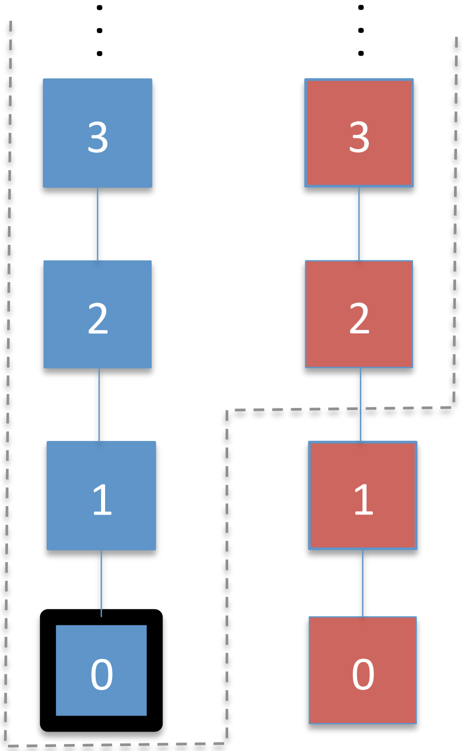




No Reversal Optimal Solution

"Popper":
choose the paradigm with fewer free parameters.

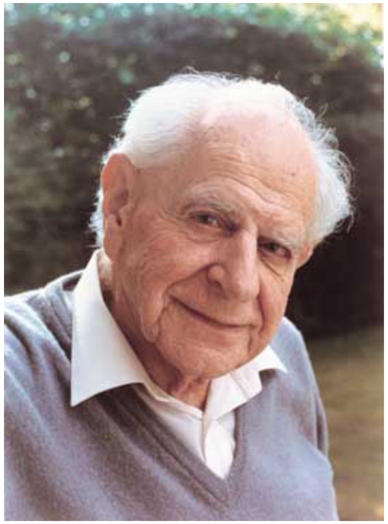


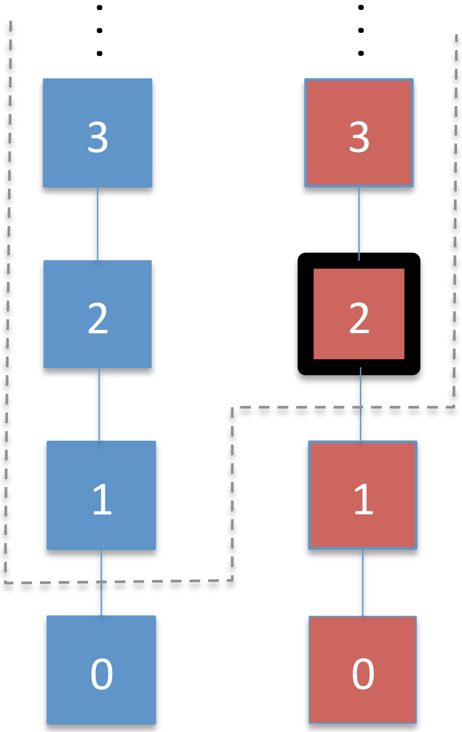


Lakatos:
choose the paradigm that was adjusted least recently.

No Reversal Optimal Solution

“Popper”:
choose the paradigm with fewer free parameters.





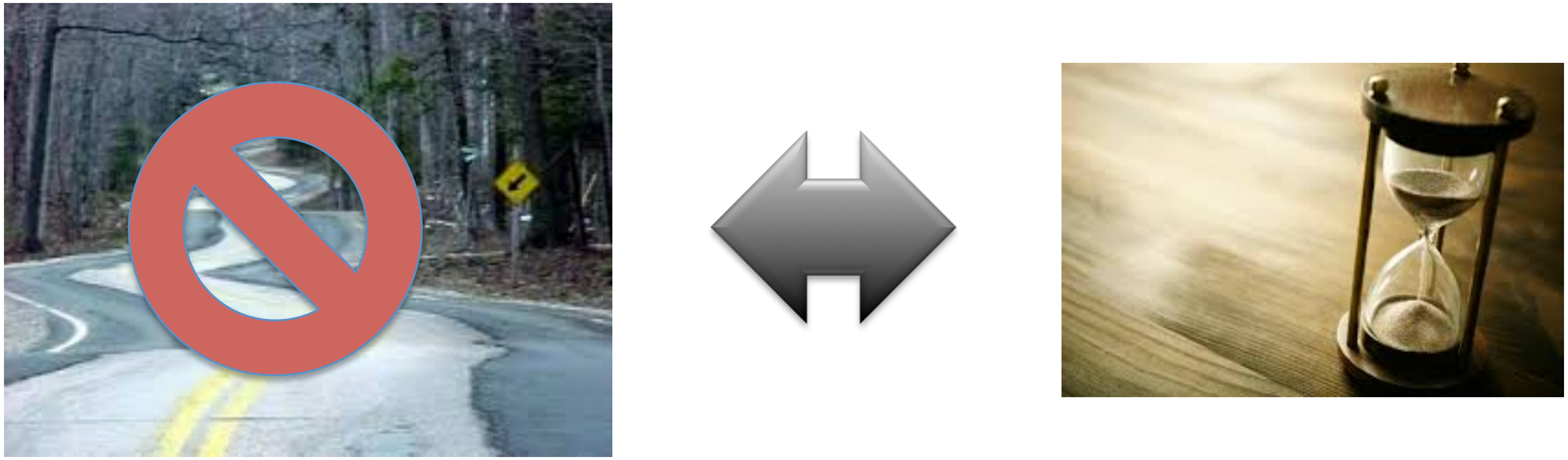
Lakatos:
choose the paradigm that was adjusted least recently.

Contextual Justification

- If patience is truth-conducive in your problem, its feasibility in some other problem is irrelevant.

Main Result 4

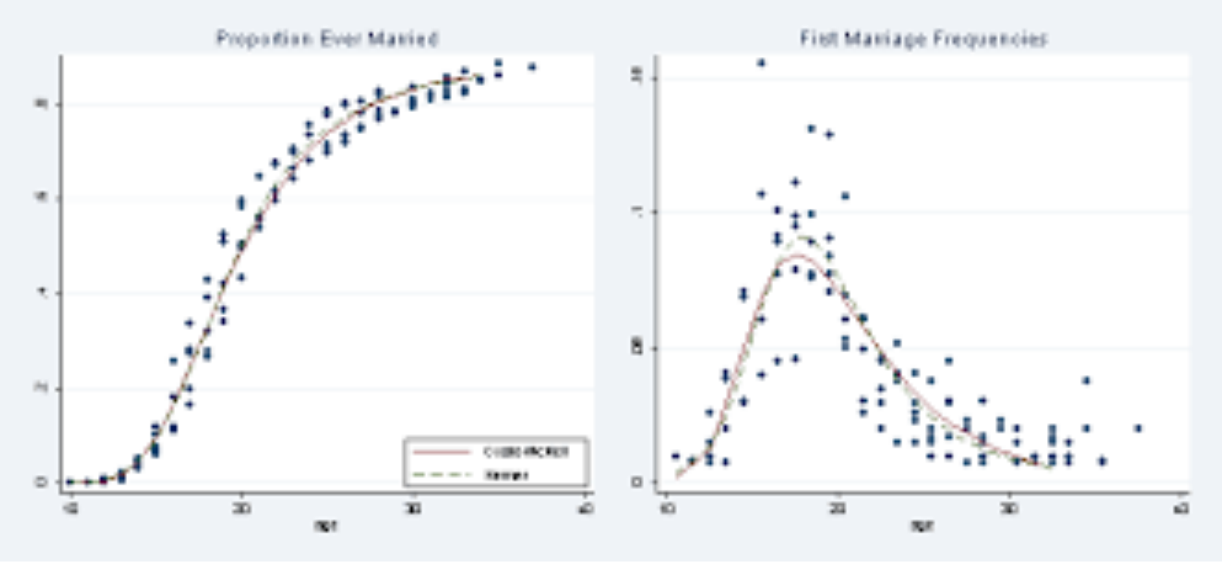
- **Proposition:** a solution is reversal optimal only if it is patient.



9. OCKHAM'S STATISTICAL RAZOR

Stochastic Theories

- Every theory is stochastic due to measurement error.
- Do the preceding results extend to stochastic theories?



Stats Wars

Bayesianism

- (+) Induction
- (-) Unreliable

Frequentism

- (-) No induction
- (+) Reliable

Come to the coherence side, Luke, and together we will believe a complete theory of the universe!

Never! Without reliability, I'll just use theories for prediction without believing them!

Darth Bayeser

Luke Estimator

Stats Peace

Frequentist theory of inductive inference

- (+) Induction
- (+) Optimal inductive reliability
- (+) Bayesian methods are an option

Darth Bayeser Luke Estimator

Short Story

The preceding concepts and results extend naturally to statistical inference.

Statistical Problems

Statistics

- W = a set of joint probability distributions over $\mathbf{X}$.
- $\mathbf{X}_N$ = random sample of size N.
- Method M_N is a measurable function of $\mathbf{X}_N$ that is almost surely continuous (the chance of landing on the boundary of a zone is zero, for all w).

Information Topology

- There are no information states!
- A statistical method draws conclusions based on whether a sample point hits an open region of sample space.
- Two distributions are less distinguishable if they produce smaller differences in the chance that such a region is hit.
- So the information topology is the initial topology for:

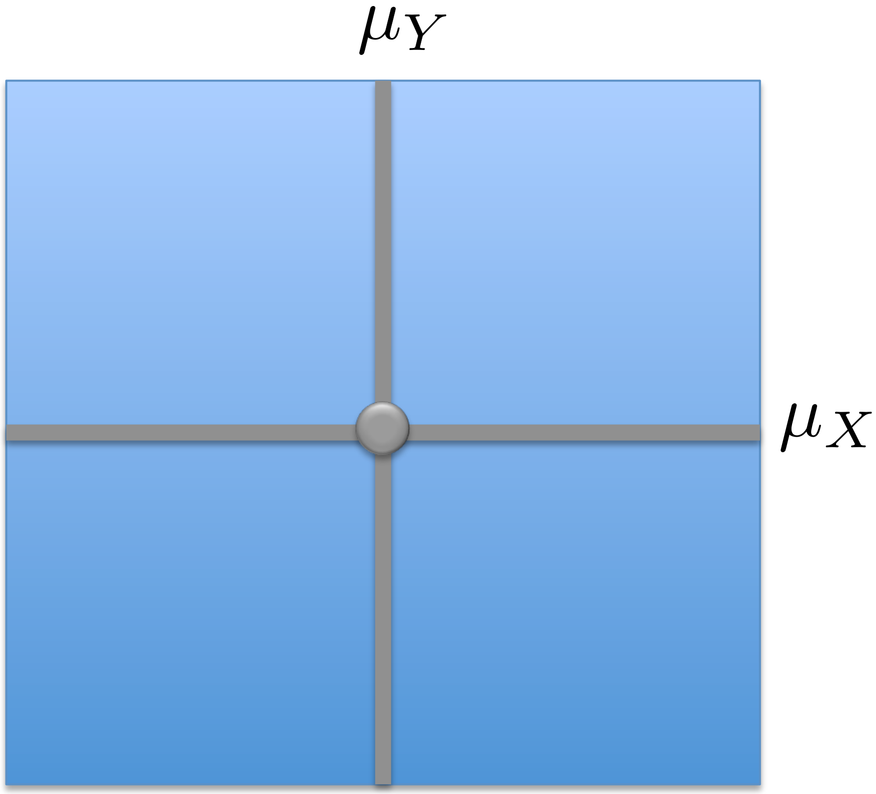
$$\{f_S^N(w) = p_w^N(\mathbf{X}_N \in S) : N \in \omega; S \text{ open}\}.$$

“In Chance” Translation

Topology	**Statistics**
Simplicity	Simplicity
Convergence	Convergence **in chance**
Reversal sequence	α-reversal sequence
Ockham’s razor	α-razor
Patience	α- patience

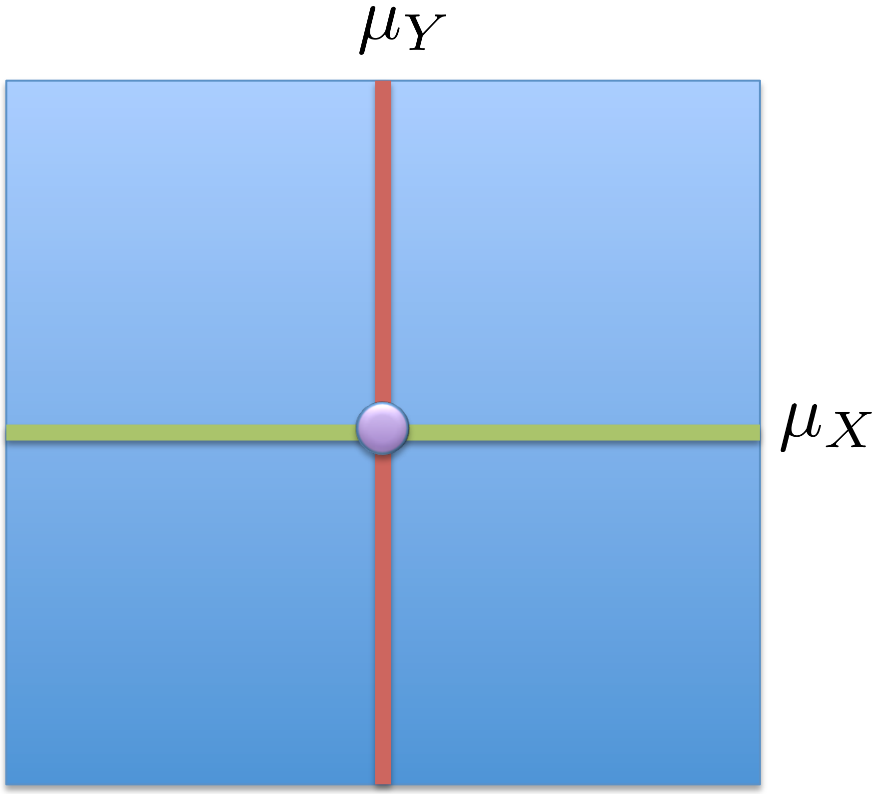
Example

- Worlds: bivariate independent normal distributions.



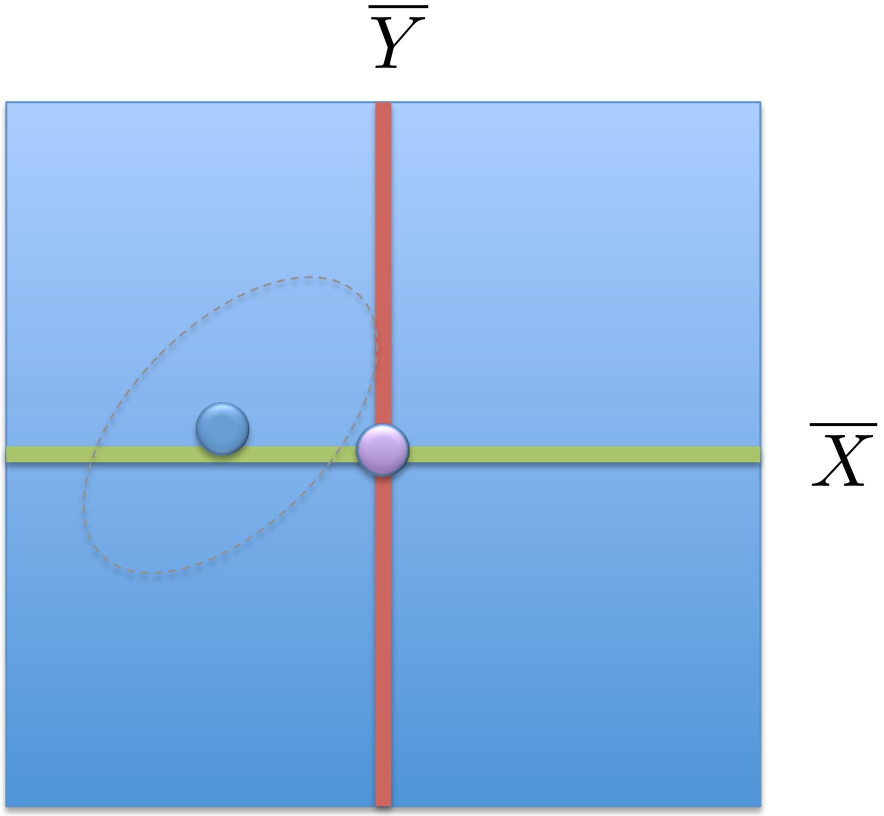
Example

- Worlds: bivariate independent normal distributions.
- Question: which mean components are non-zero?



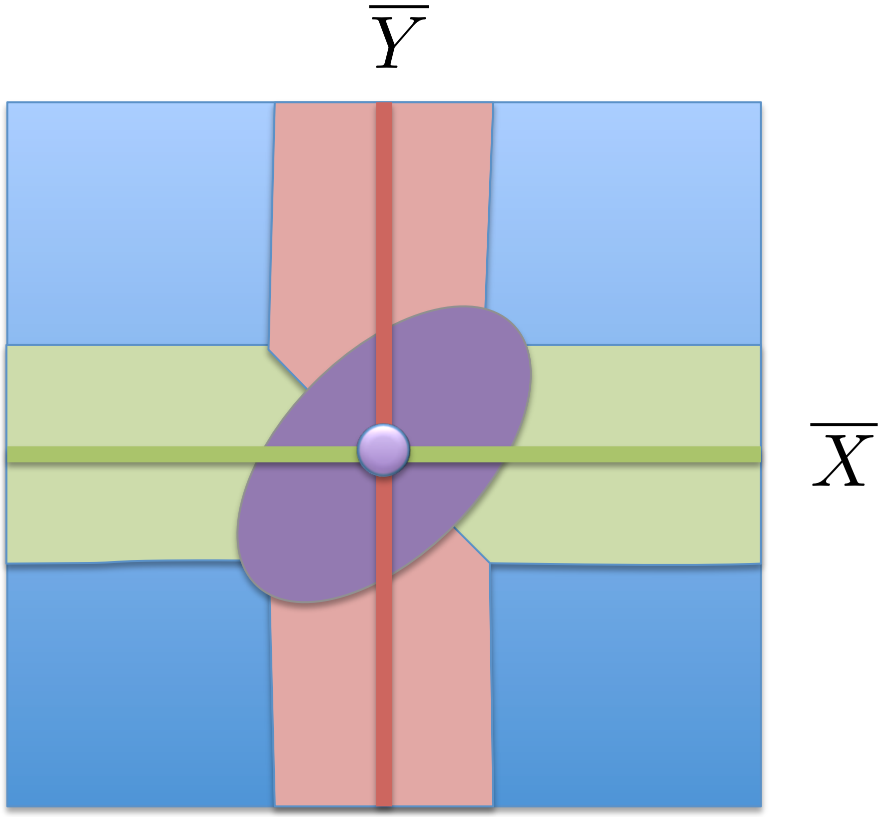
Example

- **Worlds:** bivariate independent normal distributions.
- **Question:** which mean components are non-zero?
- **Input:** sample mean vector at sample size N.



Example

- Method: maps possible samples to relevant responses.



Information Topology on W

The information topology on W is the weakest topology for which the function

$$f(p) = p(\mathbf{X}_N \in S)$$

is continuous from W to R, for arbitrary, fixed N and Borel event S in R^N.

Thus $g(p) = p(M_N = B)$ is continuous.

Reversals in Chance

Reversal sequence: $(A_1, \ldots, A_n)$.

M produces the sequence with chance $\geq \alpha$ iff

$$p_w^{N_i}(M = A_i) \geq \alpha,$$

for some world *w* and sample sizes $N_1, \ldots, N_n$:

Reversal Efficiency

- Solution *M* is α cycle efficient iff *M* produces no α cycles.

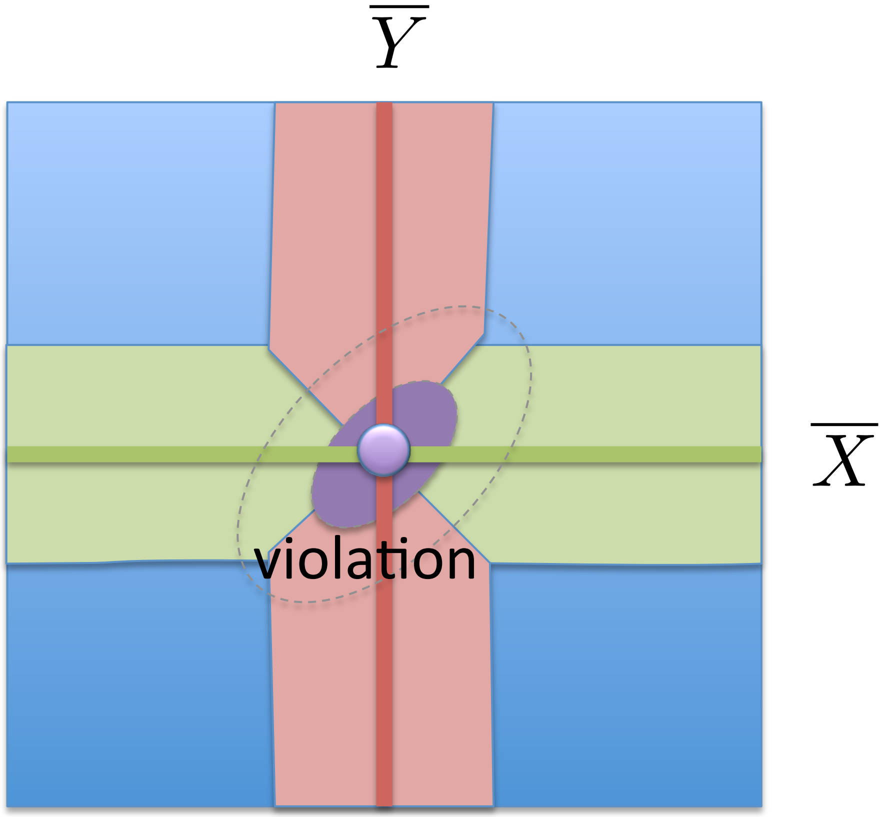
Reversal Efficiency

- Solution *M* is α reversal efficient iff every α reversal sequence produced by *M* is also produced by an arbitrary solution.

Ockham's α-Razor

Violate the error razor with at most chance α.

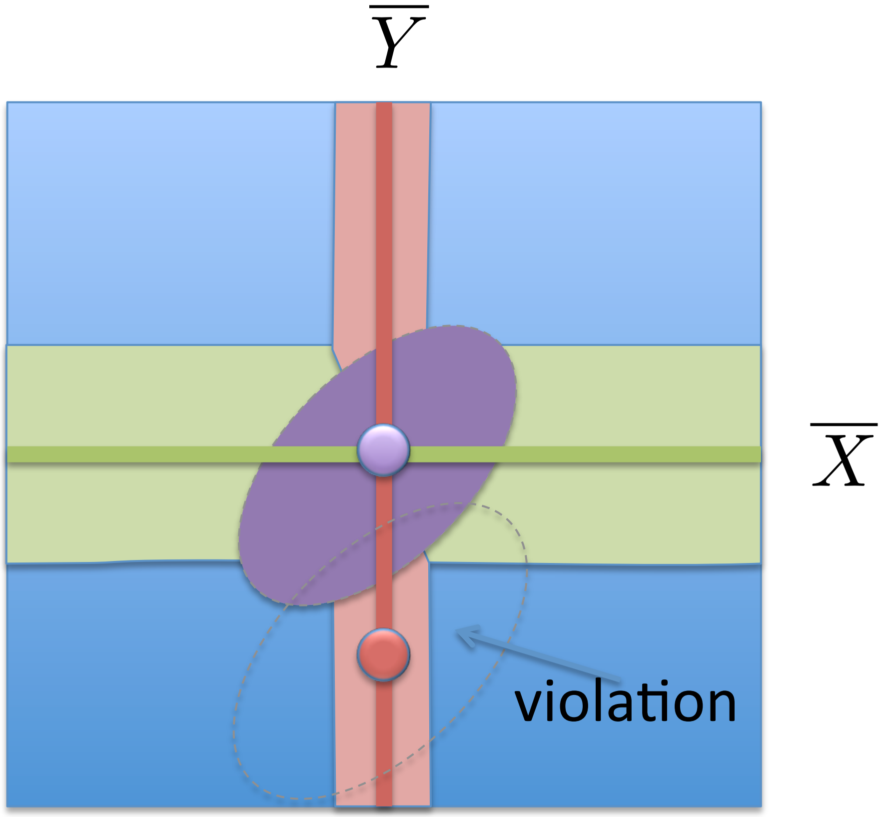
M_N produces no answer more complex than the true answer with chance ≥ α .



Ockham's α-Razor

Violate the error razor with at most chance α.

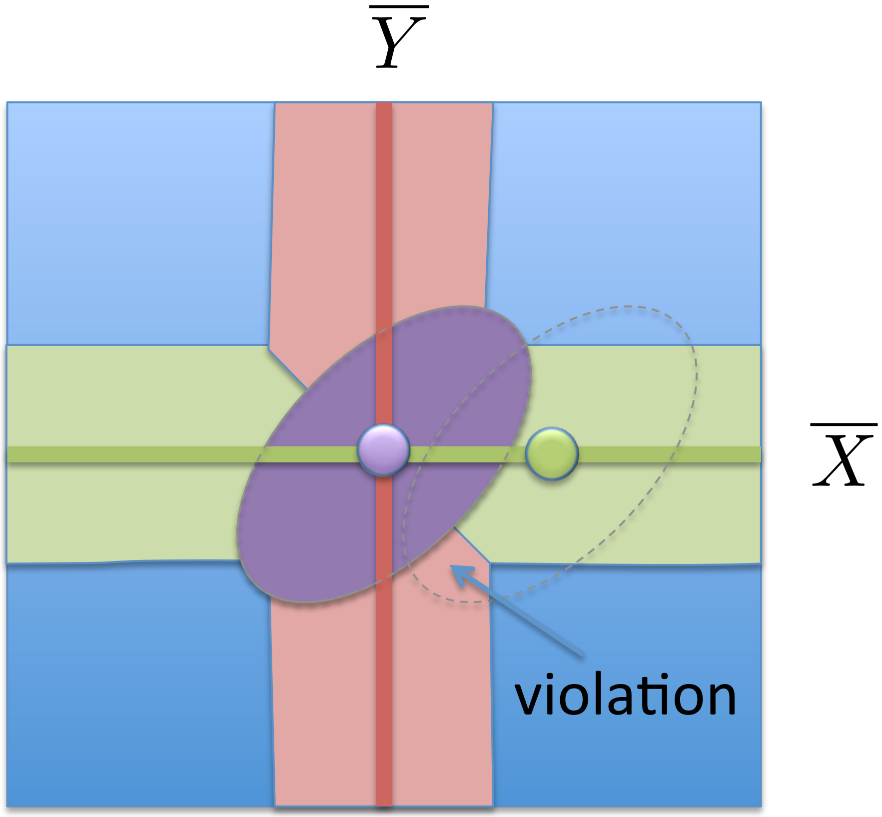
M_N produces no answer more complex than the true answer with chance ≥ α .



α-Patience

Violate error patience with at most chance α.

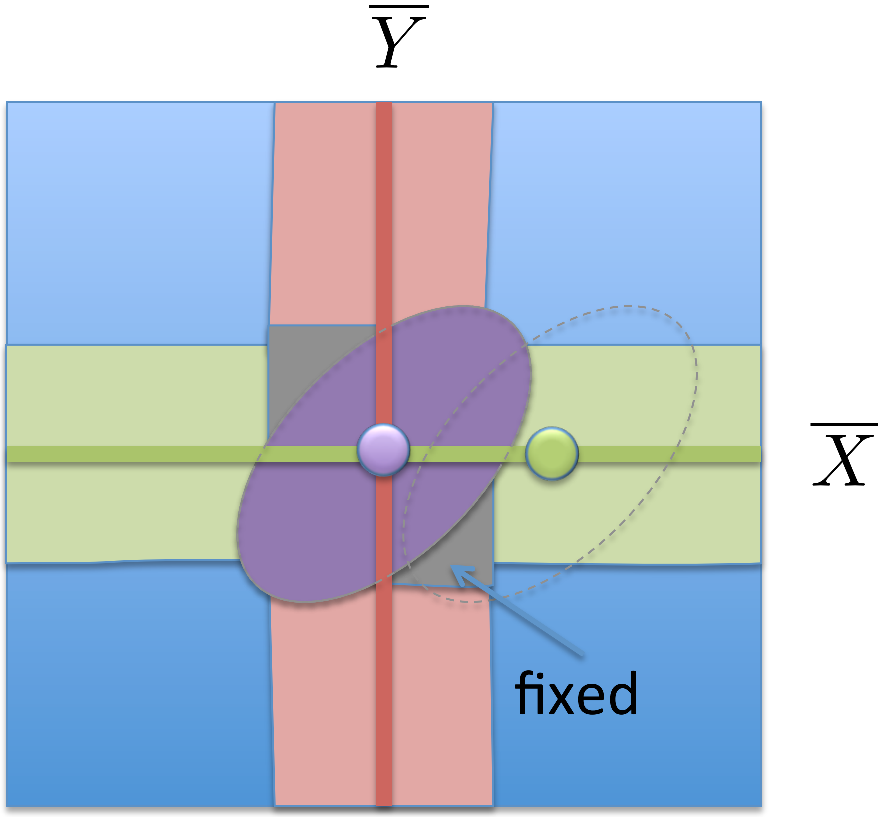
M_N produces no answer that is not as simple as the true answer with chance ≥ α .



α-Patience

Violate error patience with at most chance α.

M_N produces no answer that is not as simple as the true answer with chance ≥ α .



Main Result 5

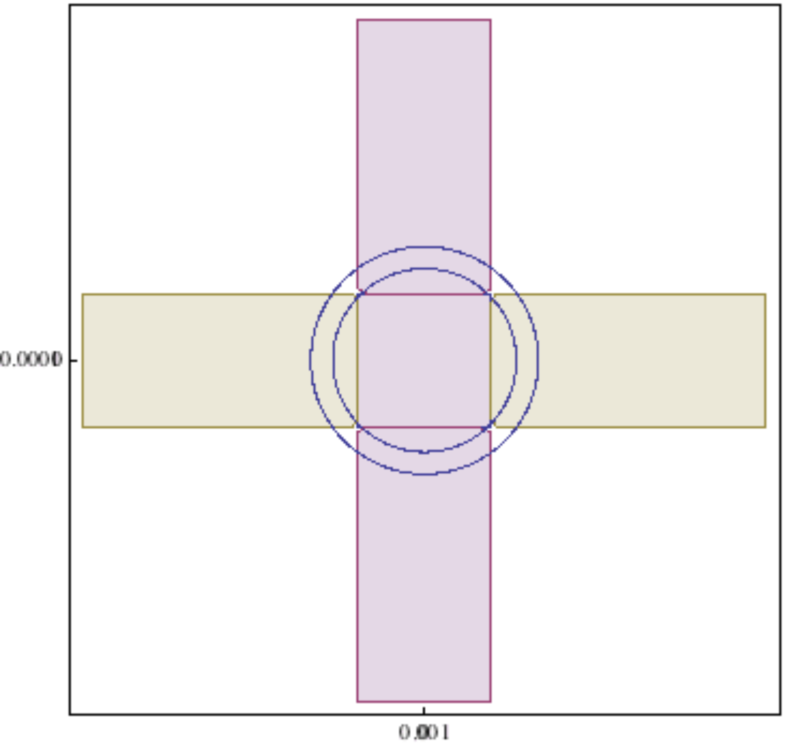
- **Proposition:** Every α-cycle free solution in chance to a statistical problem satisfies Ockham's α- razor.

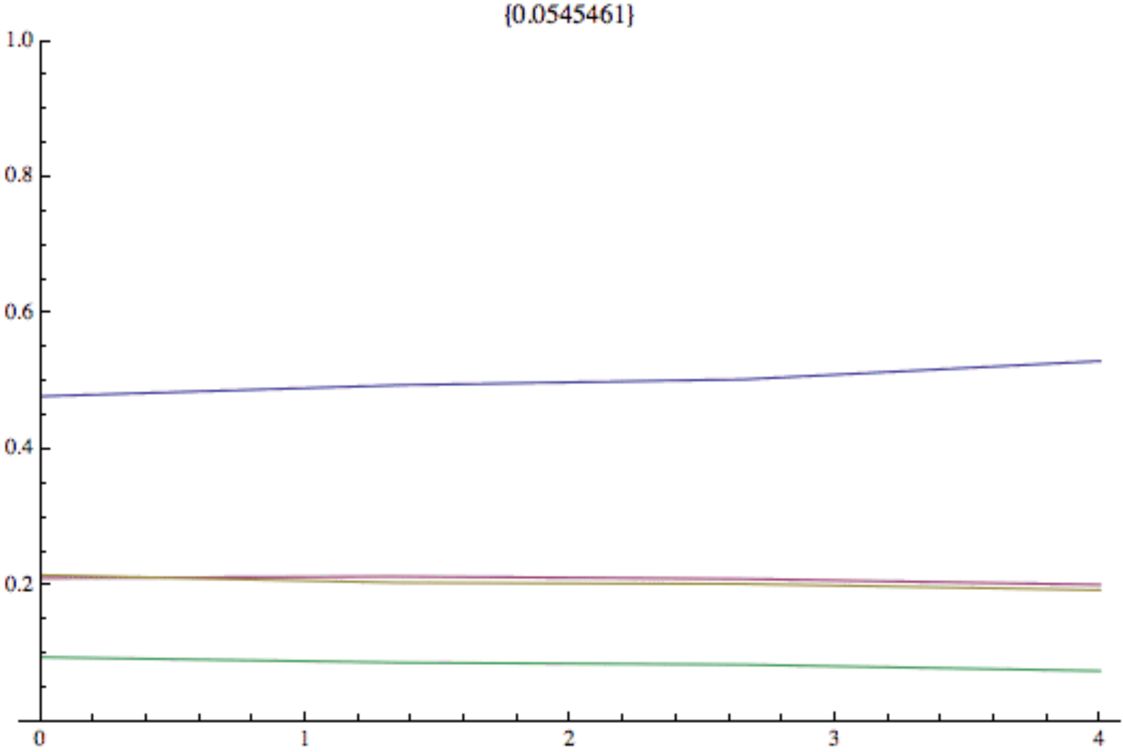
Main Result 6

- **Conjecture:** Every α-reversal-efficient solution satisfies α- patience.

Bayes

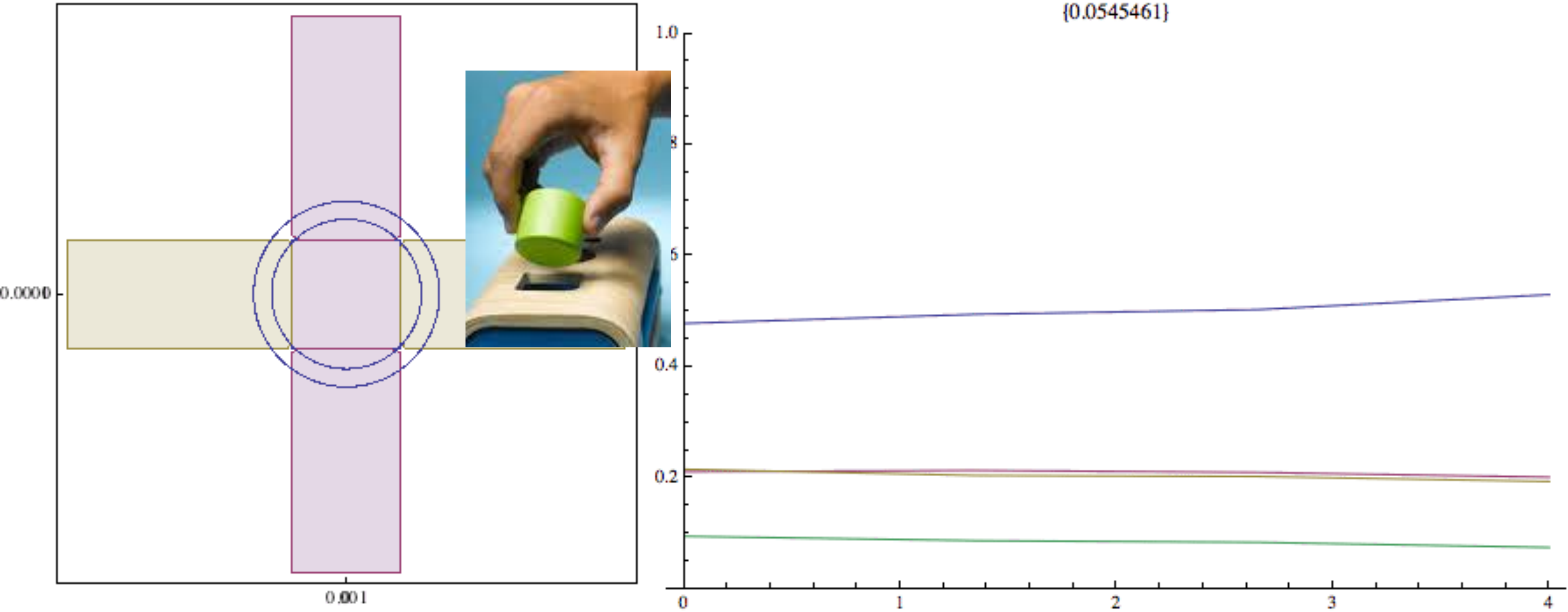
- Method: Maximize Bayesian credence,
- Even priors on models, Gaussian priors on parameters.





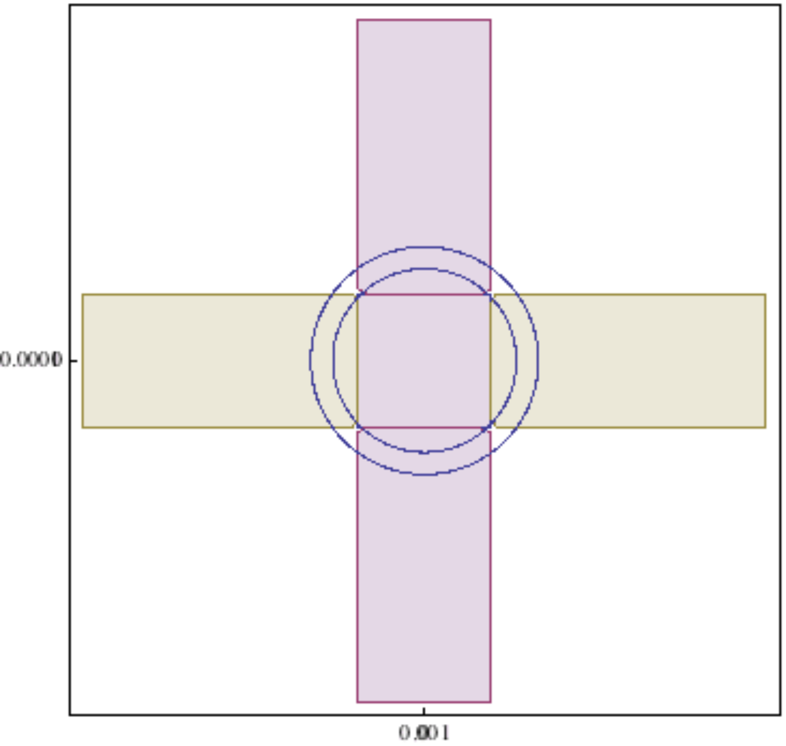
Bayes Simulation

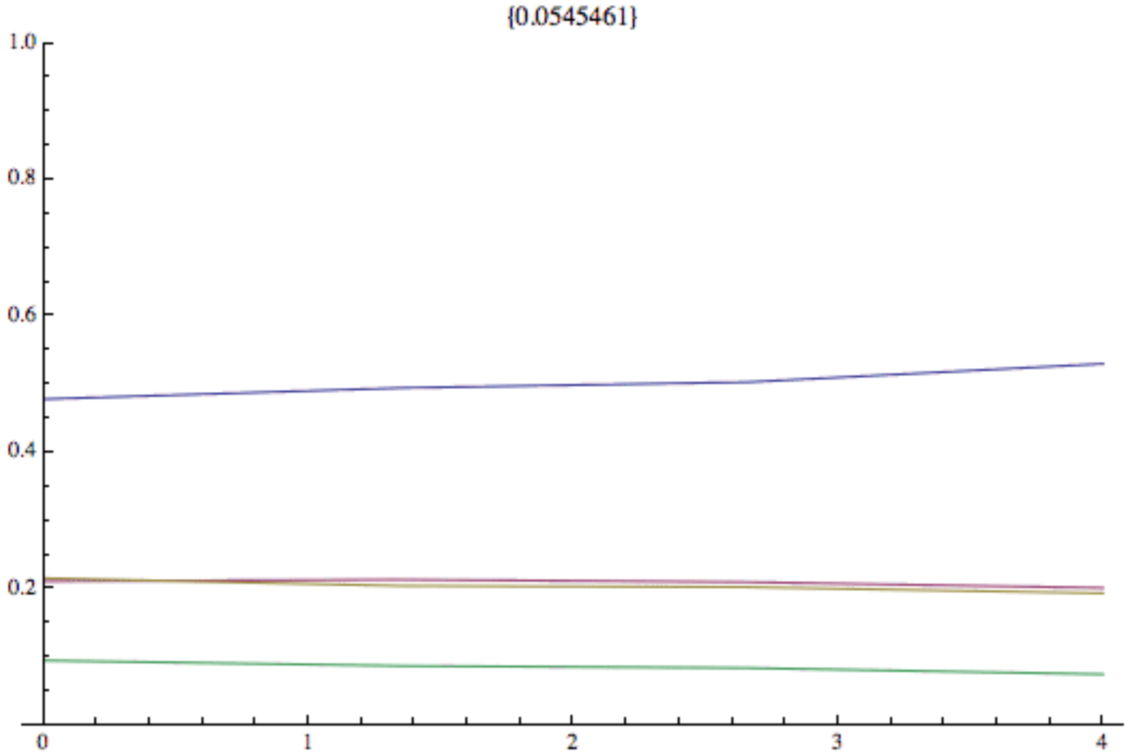
- Method: Maximize Bayesian credence,
- Even priors on models, Gaussian priors on parameters.



Bayes Simulation

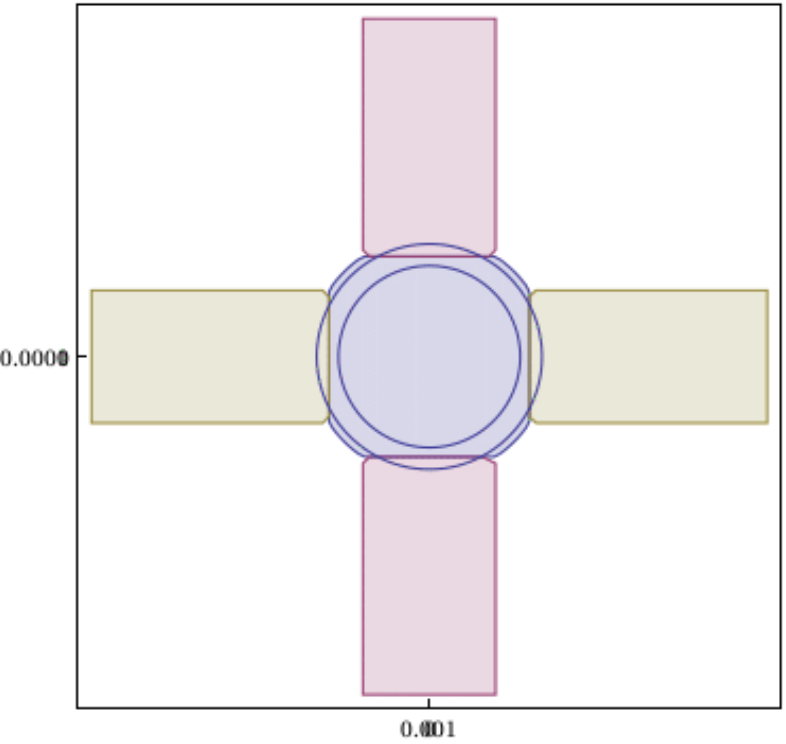
- Method: Maximize Bayesian credence,
- Even priors on models, Gaussian priors on parameters.
- 20% Ockham violation, 20% impatience, bad power.

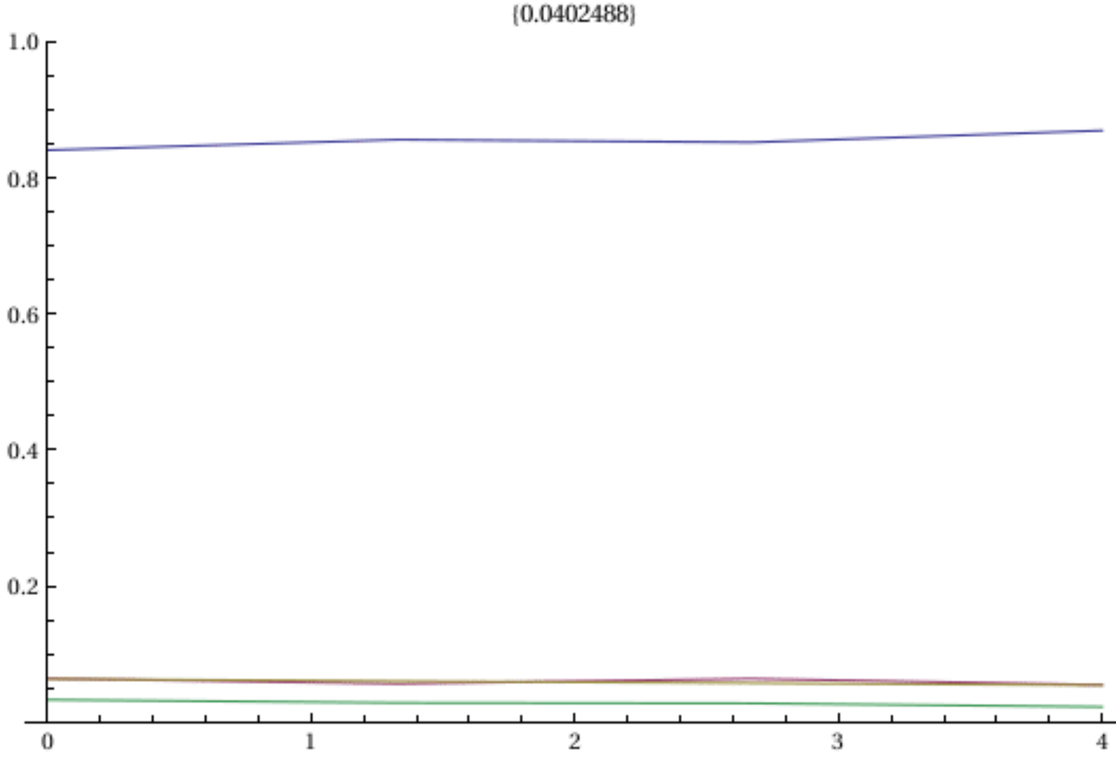




Bayes Simulation

- Method: Maximize Bayesian credence,
- Prior bias toward simplicity, Gaussian priors on parameters.
- 7% Ockham violation, 7% impatience, bad power.



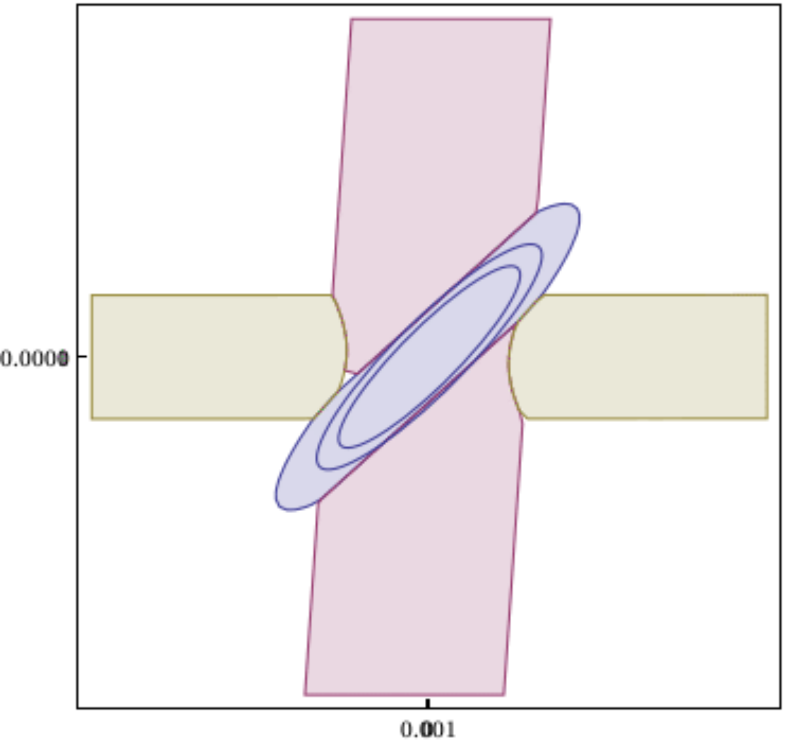


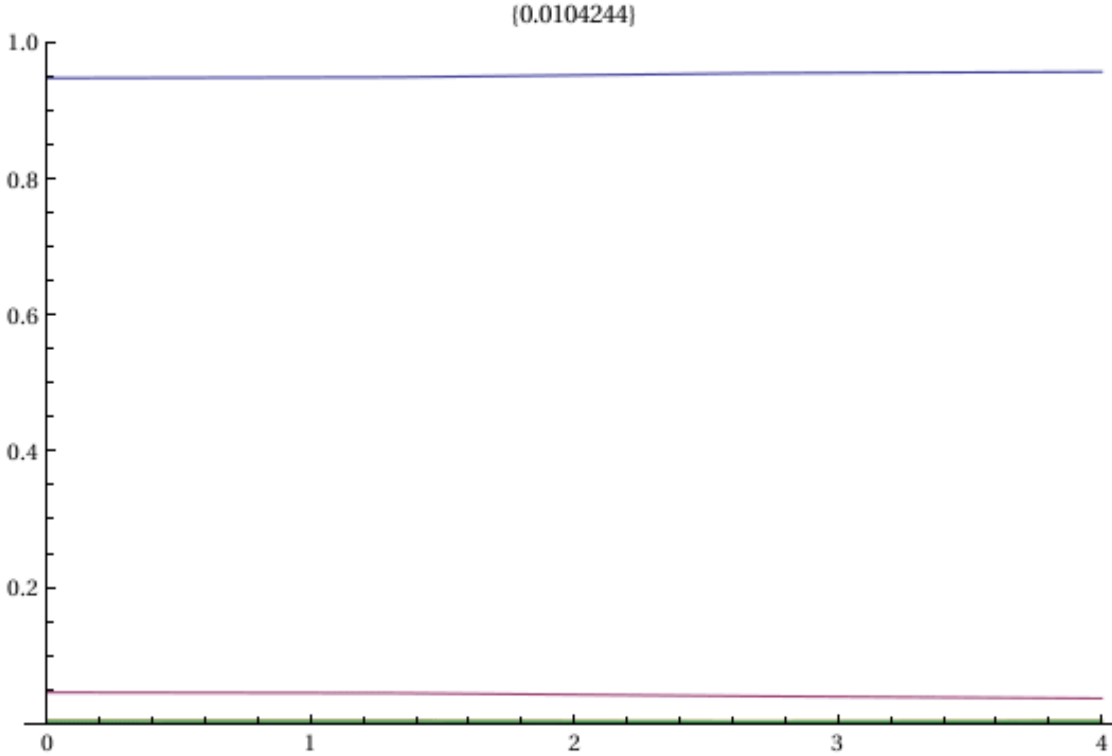
Friendly Advice

- Boost the prior on simple models to eliminate α-cycles in chance.
- Correction to GES already made independently by Oliver and Joe!
- Here is a reliabilist explanation.

Confirmation vs. Model Selection

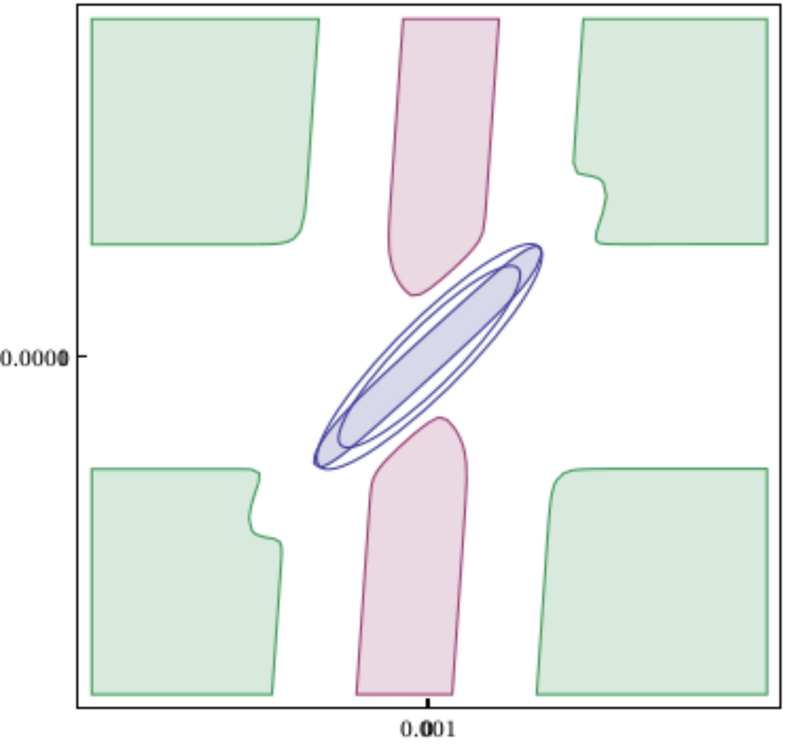
- Method: Maximize Bayesian credence,
- Prior bias toward simplicity, Gaussian priors on parameters.
- 40% impatience, bad power.

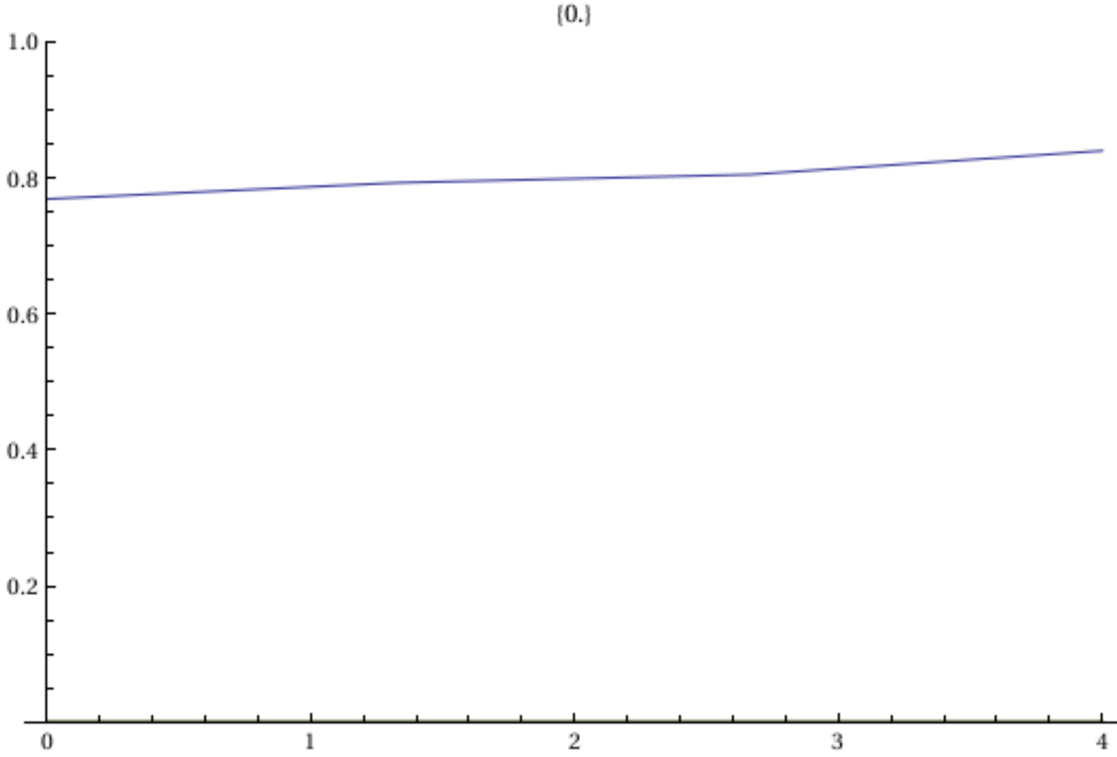




Confirmation Reconceived

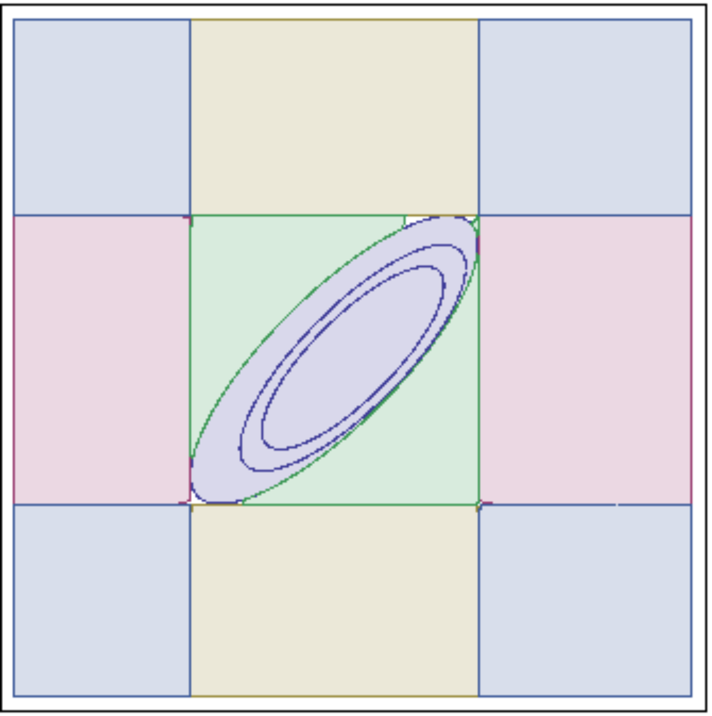
- Method: Accept at 95% threshold
- Waiting for "confirming data" brings reversals in chance down to around 5%.

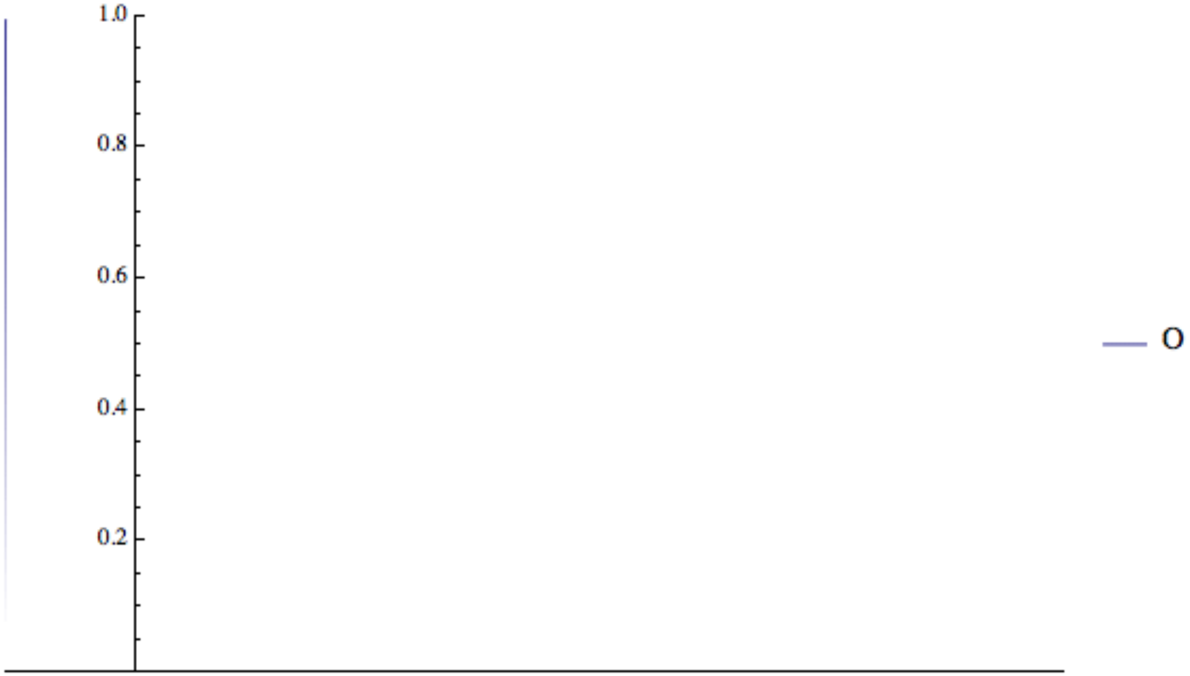




Frequentist Patient Ockham

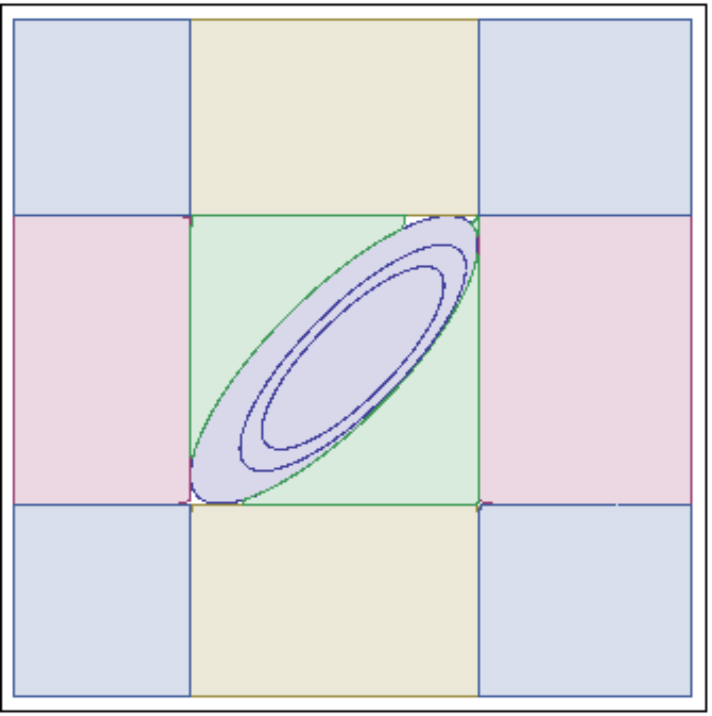
- Choose powerful nested tests at significance α .
- Tune down α with sample size to make the tests consistent.
- Disjoin the simplest models whose tests do not reject.

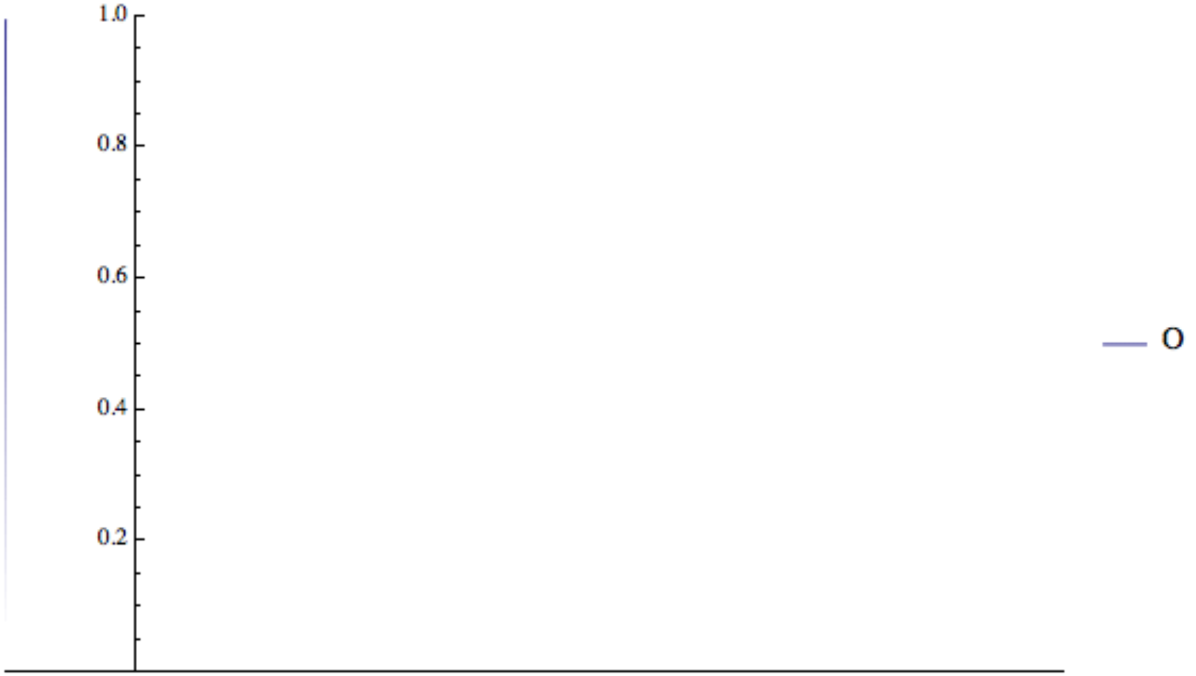




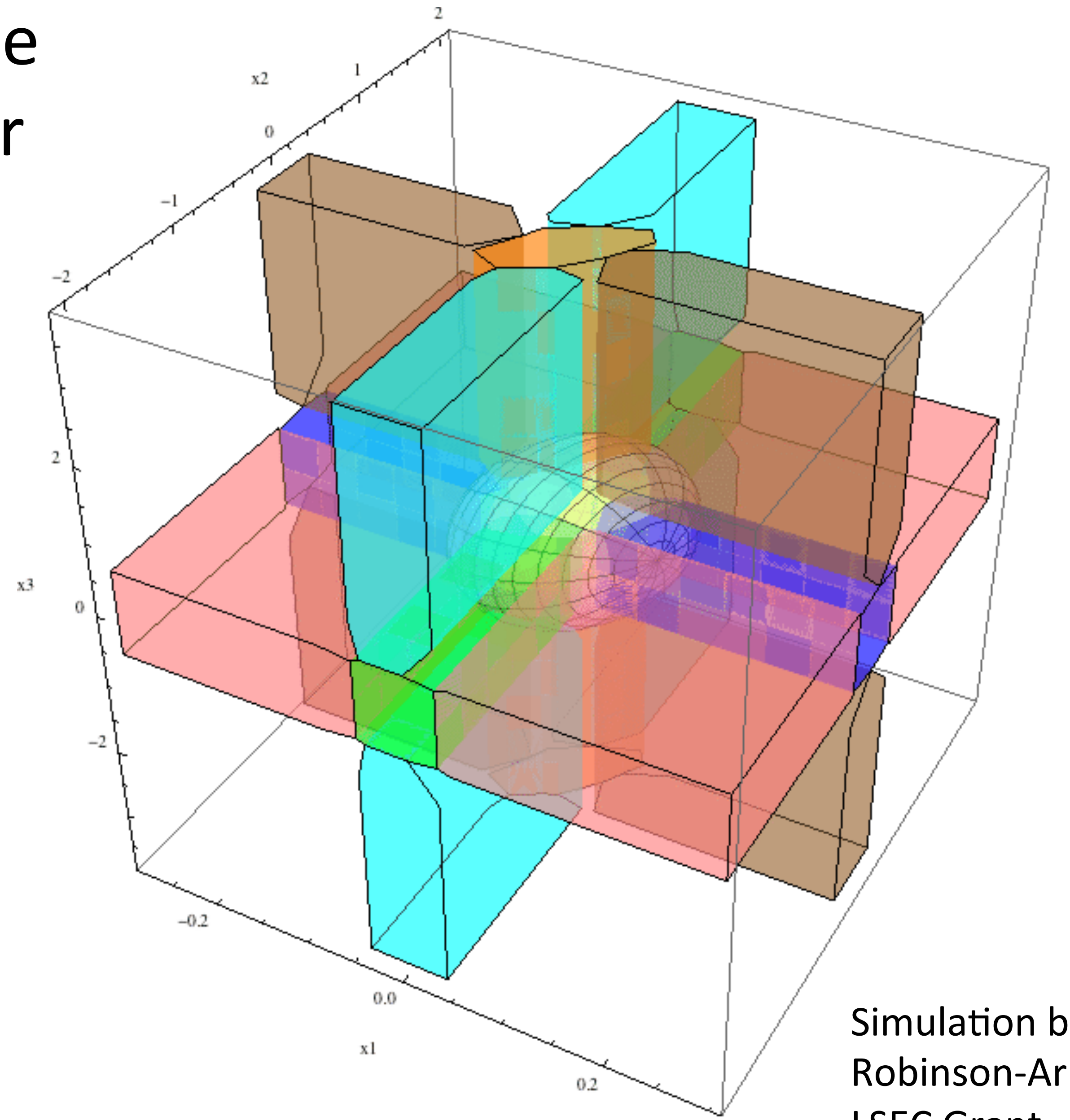
Error Statistics Reinterpreted

- “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.



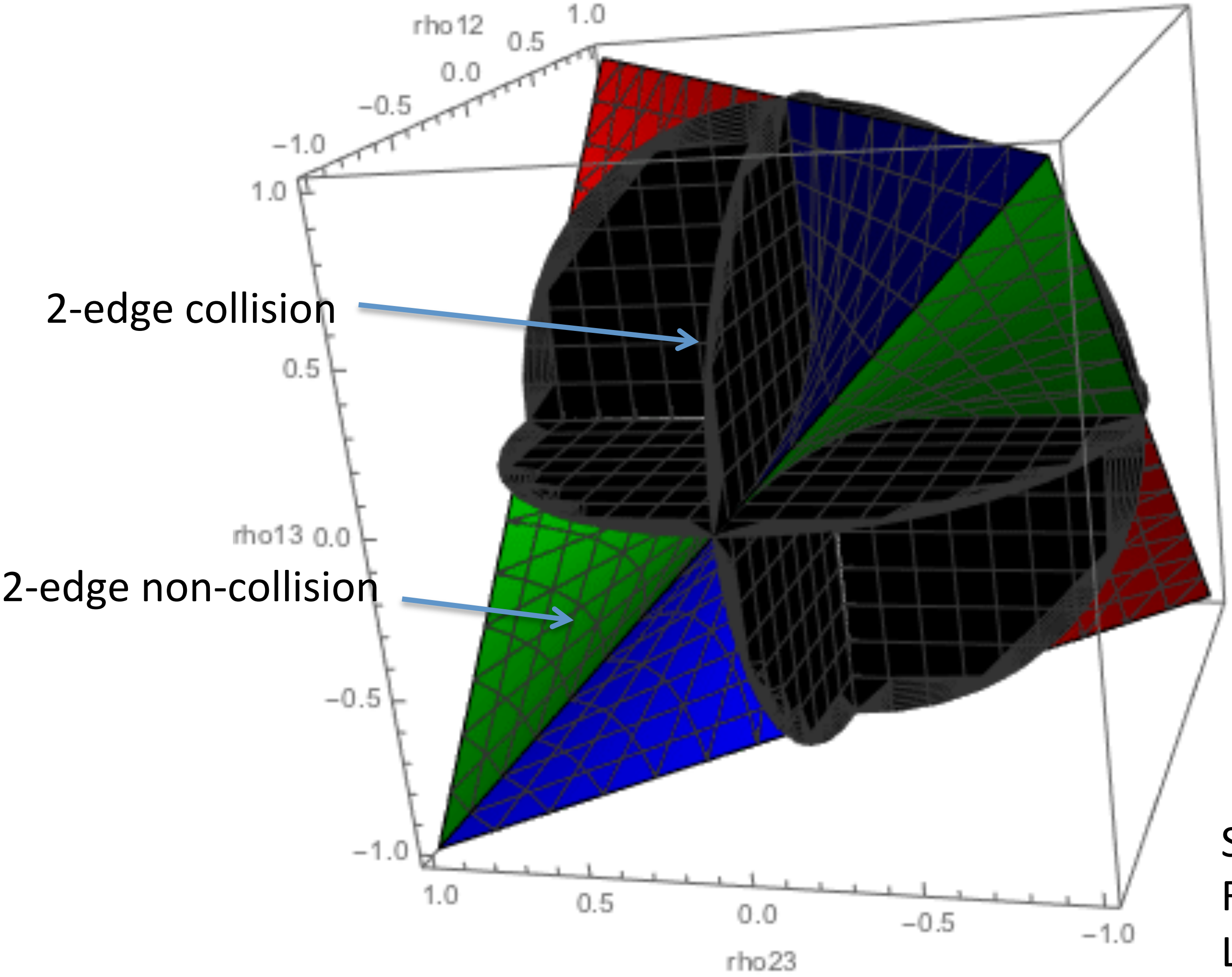


Maximize Posterior



Simulation by Connor Robinson-Arnull
LSEC Grant

Causal Inference on 3 Variables

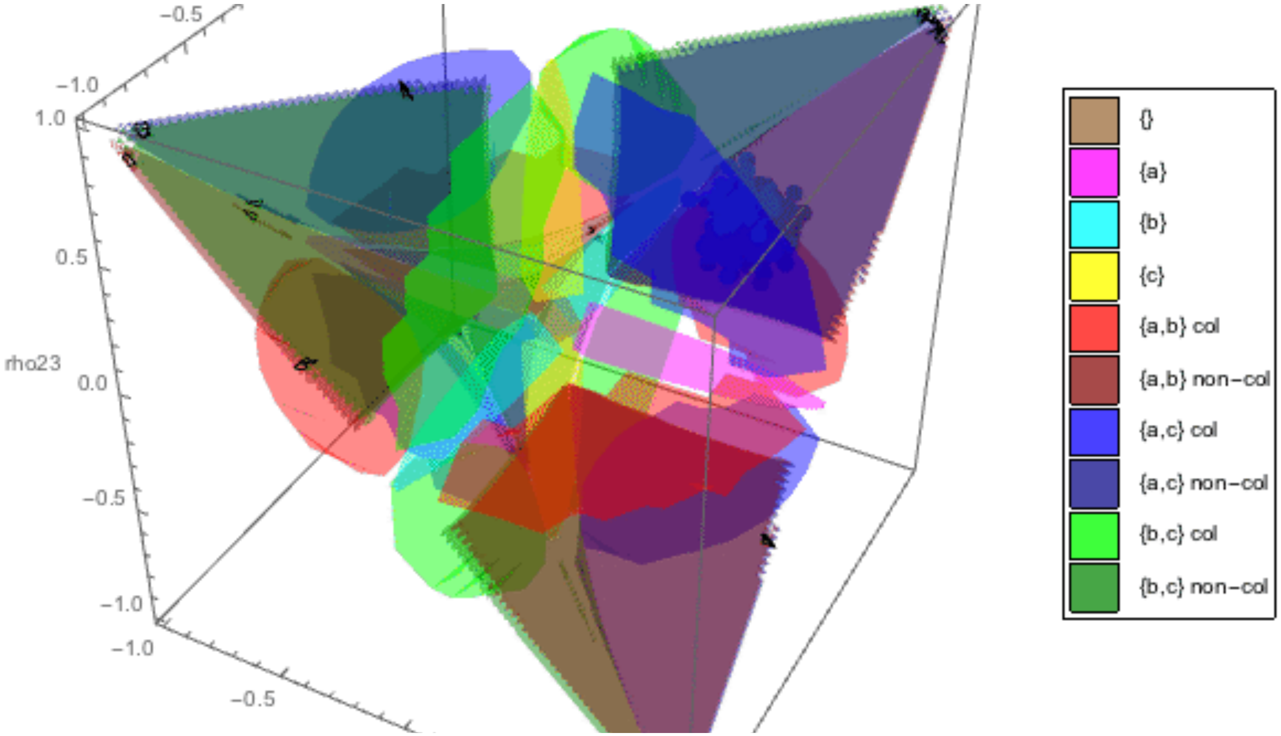


2-edge collision

2-edge non-collision

Simulation by Connor Robinson-Arnull
LSEC Grant

α-Optimal Fishing with Tests



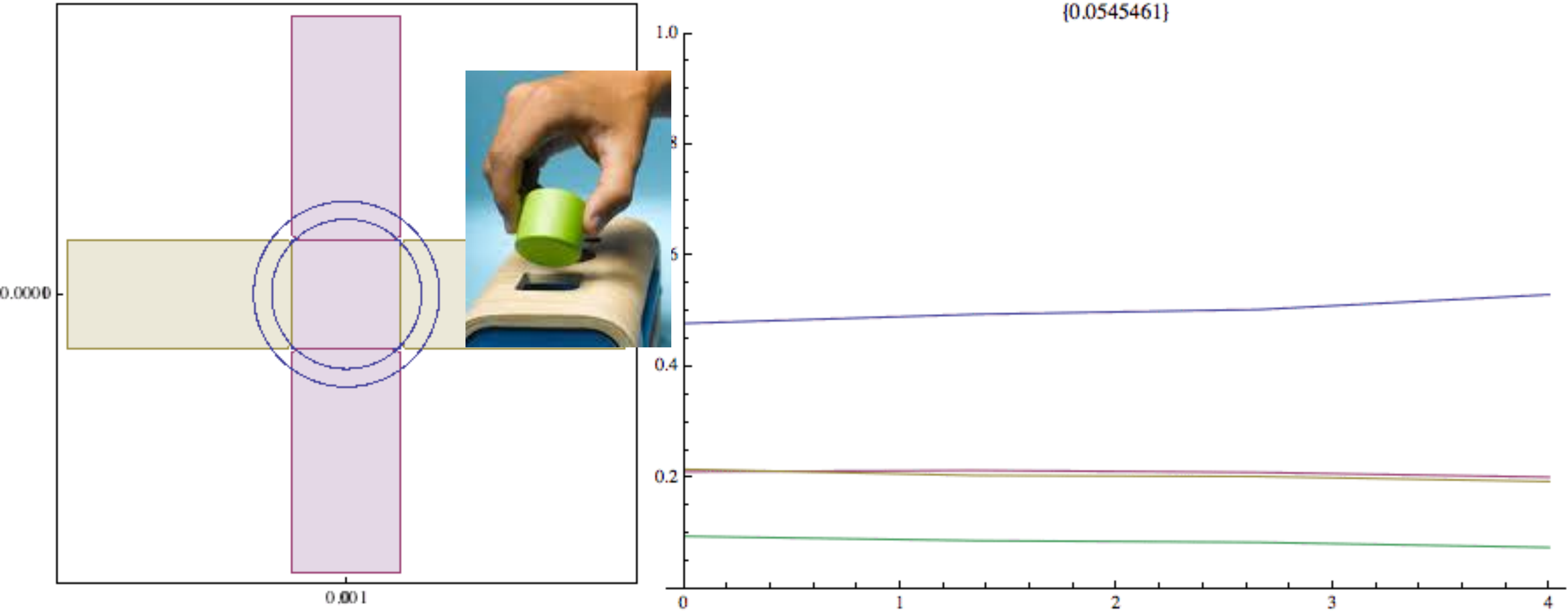
Simulation by Connor Robinson-Arnull
LSEC Grant

Summary and Discussion

1. Simplicity is a topological feature of problems.
2. Ockham's razor is necessary for cycle-optimal convergence to the true answer.
3. Patience is necessary for reversal-optimal convergence to the true answer.
4. Optimally straight convergence is weak, but its implications for scientific method are strong.
5. The same holds for statistical inductive inference.
    1. Significance → a small tolerance for reversals and cycles.
    2. Power → drop theories you are destined to drop sooner.

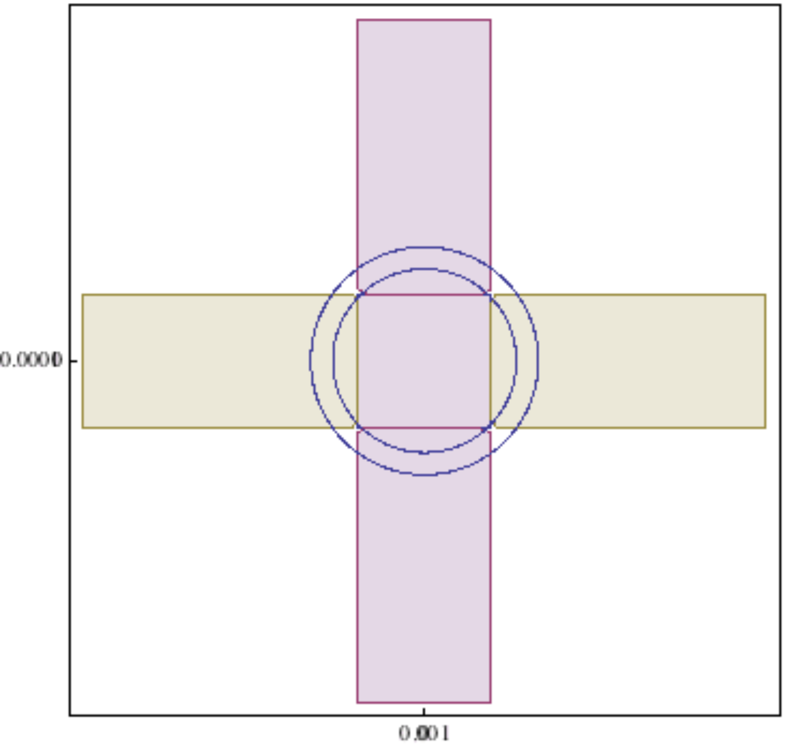
Critique of Bayes

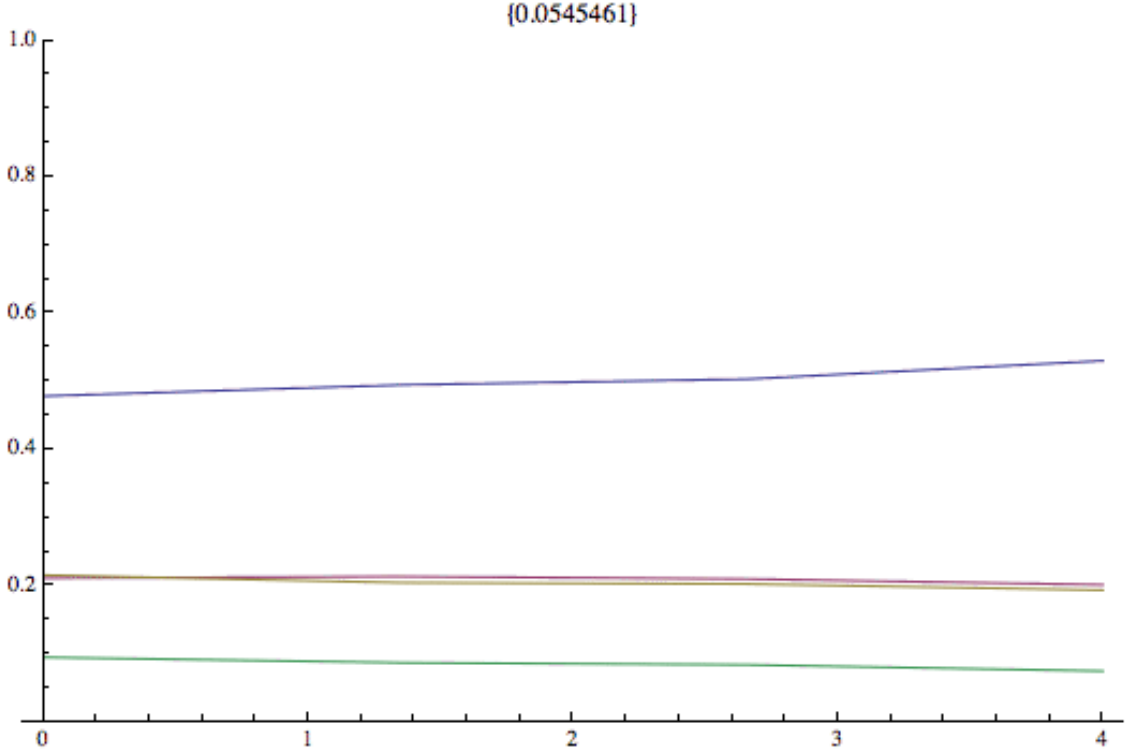
- Method: Maximize Bayesian credence,
- Even priors on models, Gaussian priors on parameters.
- 20% Ockham violation, 20% impatience, bad power.



Critique of Bayes

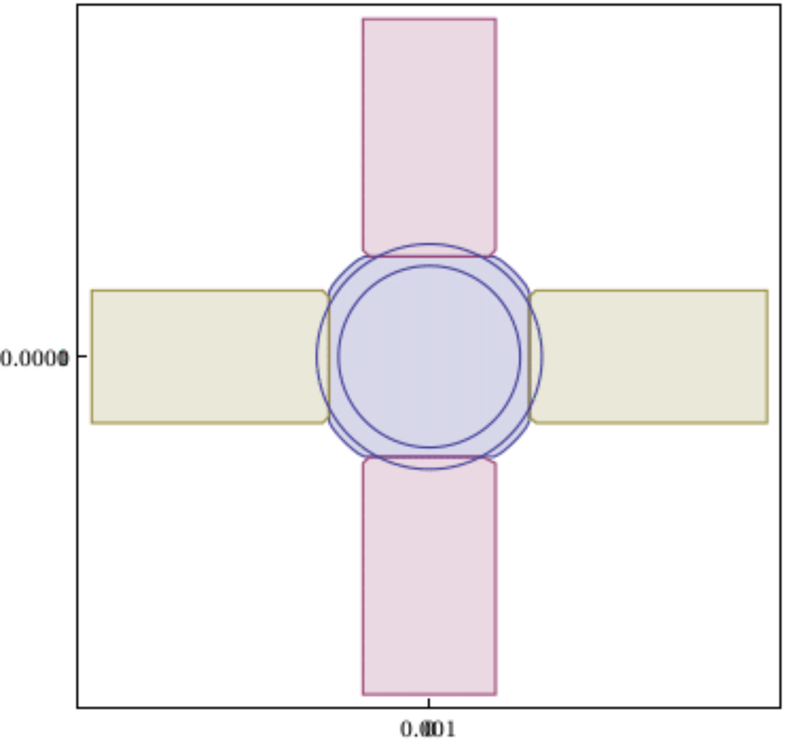
- Method: Maximize Bayesian credence,
- Even priors on models, Gaussian priors on parameters.
- 20% Ockham violation, 20% impatience, bad power.

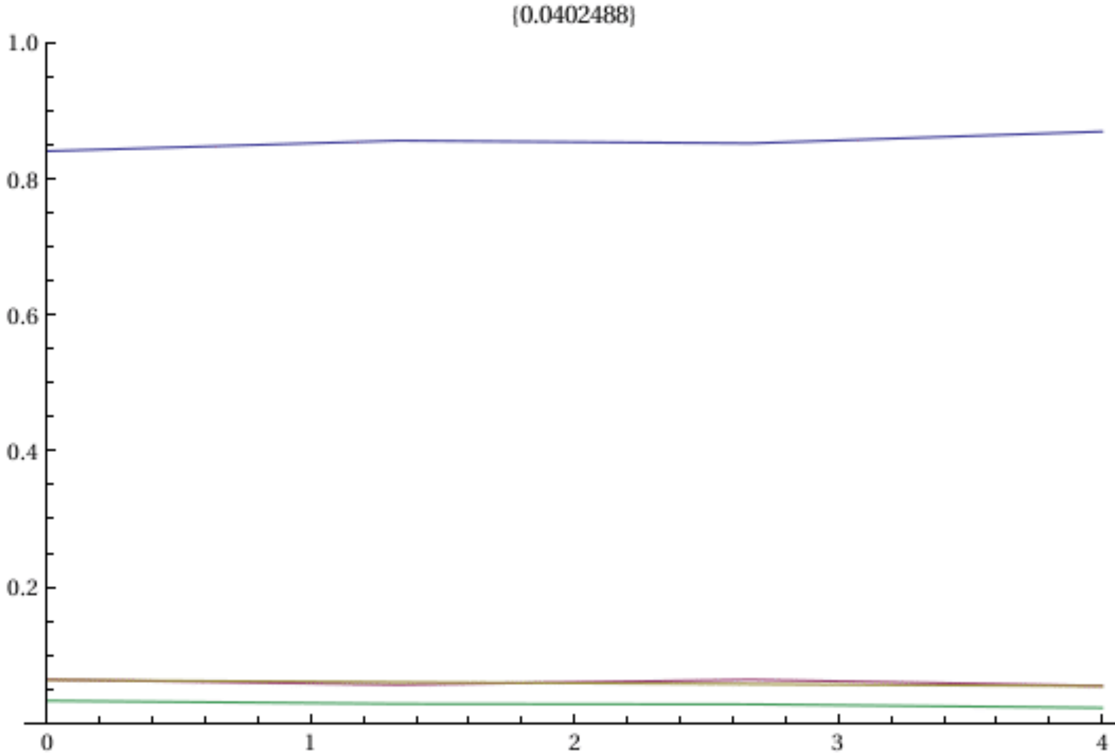




Critique of Bayes

- Method: Maximize Bayesian credence,
- Prior bias toward simplicity, Gaussian priors on parameters.
- 7% Ockham violation, 7% impatience, bad power.



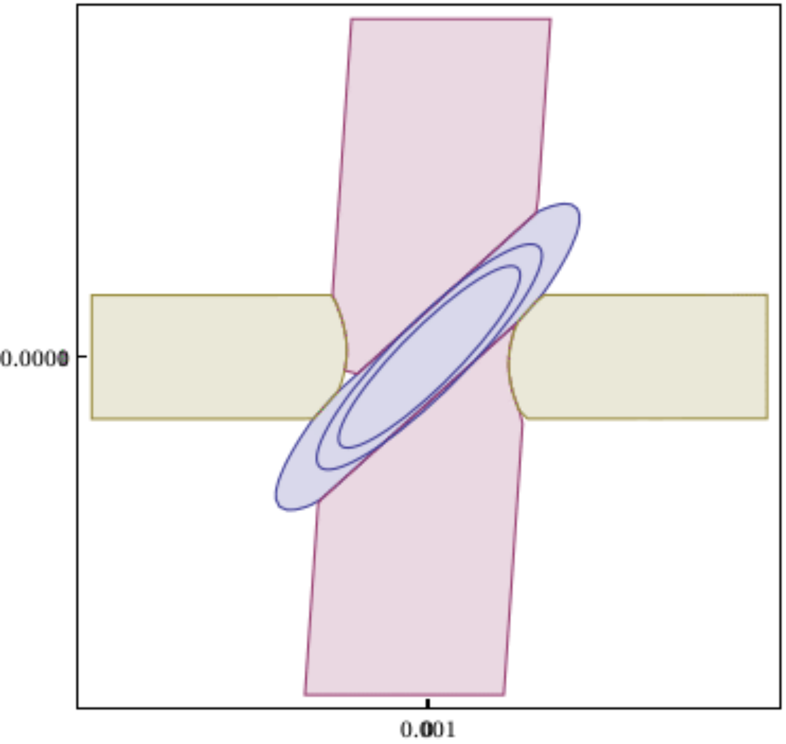


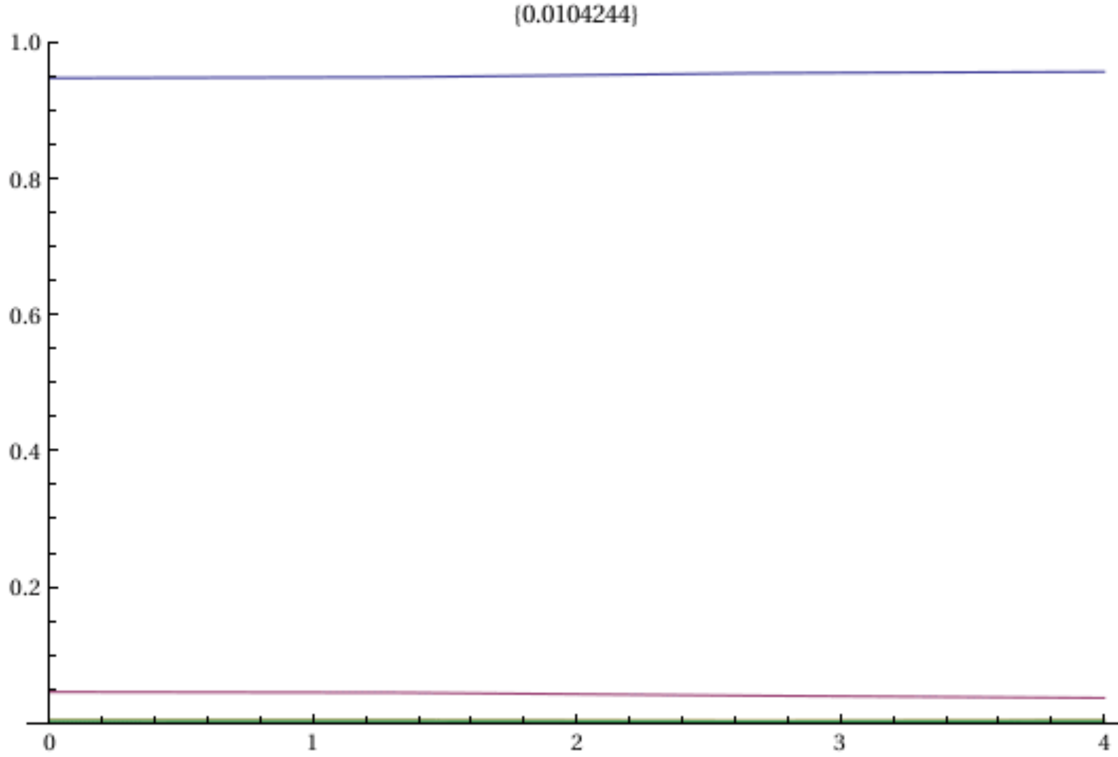
Plausible Advice

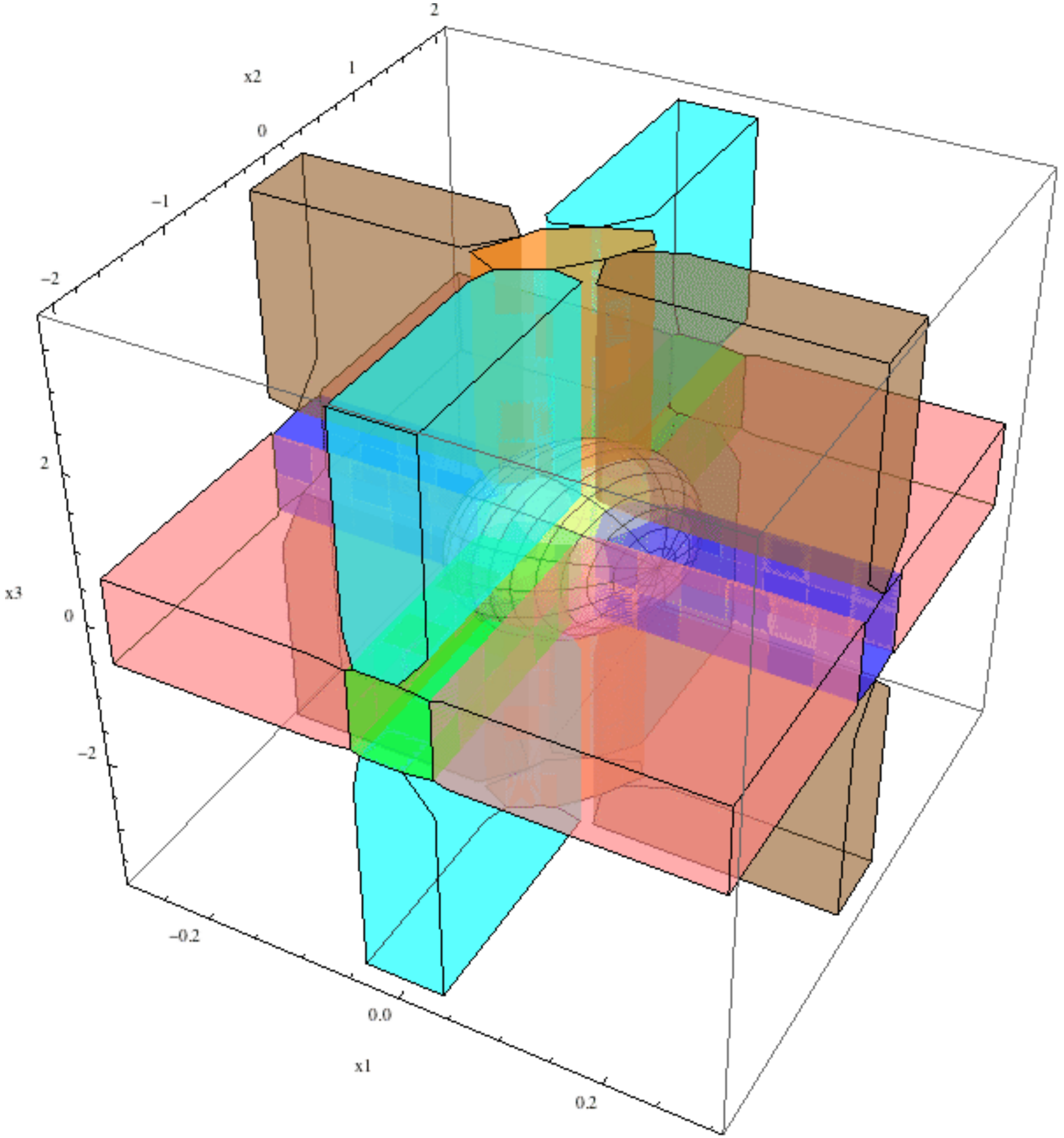
- Boost the prior on simple models to eliminate α cycles in chance.
- Then optimize power to get reversals over a.s.a.p.

Confirmation vs. Model Selection

- Method: Maximize Bayesian credence,
- Prior bias toward simplicity, Gaussian priors on parameters.
- 40% impatience, bad power.



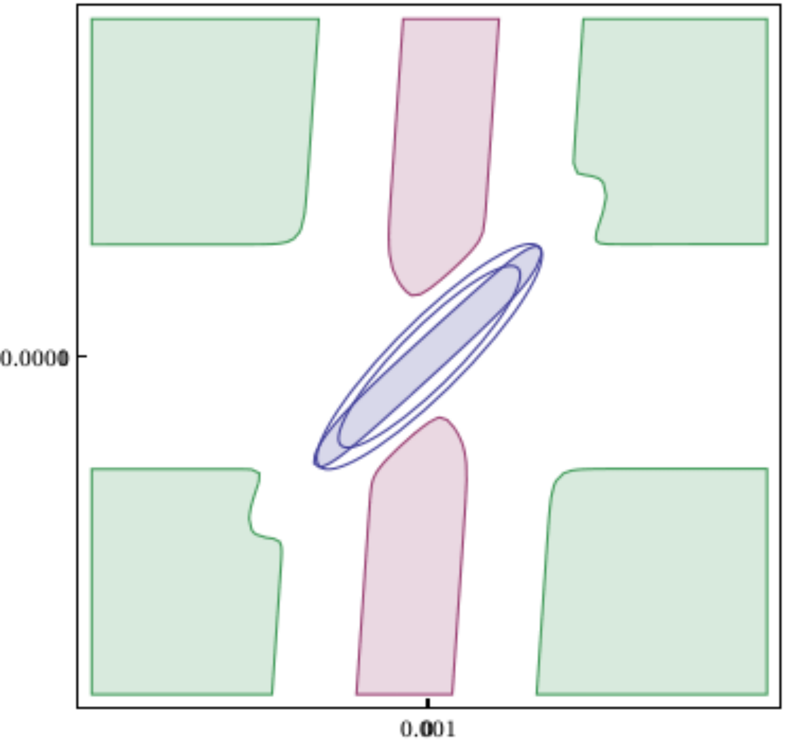


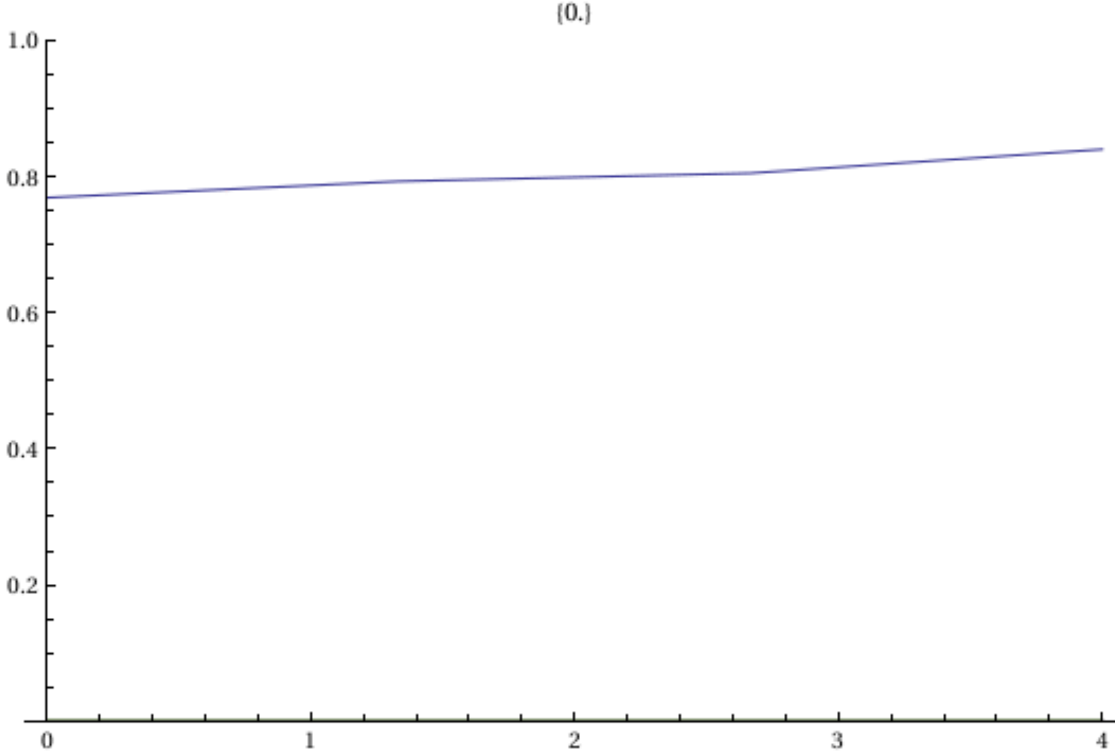


x2
2
1
0
−1
−2
x3
2
0
−2
−0.2
0.0
0.2
x1

Confirmation Reconceived

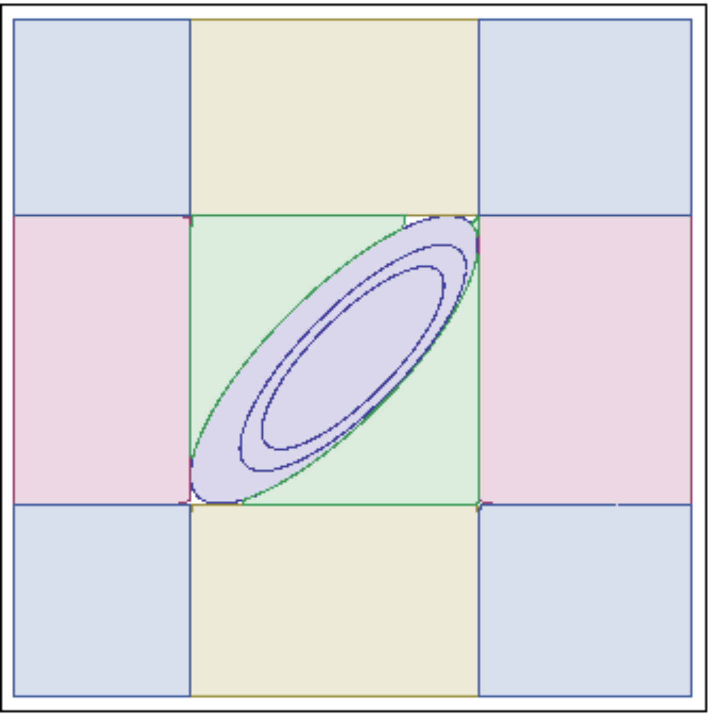
- Method: 95% threshold for Bayes Posterior
- Waiting for “confirming data” brings reversals in chance down to around 5%.

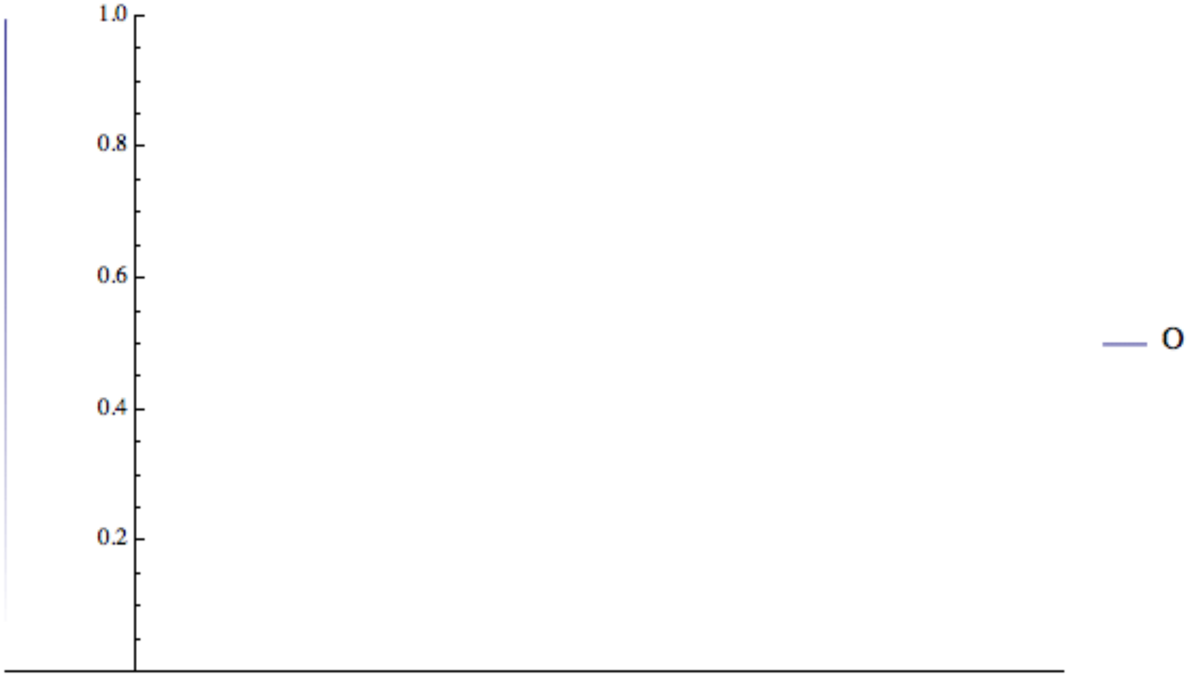




Frequentist Patient Ockham

- Choose powerful nested tests at significance α .
- Tune down α with sample size to make the tests consistent.
- Disjoin the simplest models whose tests do not reject.





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.