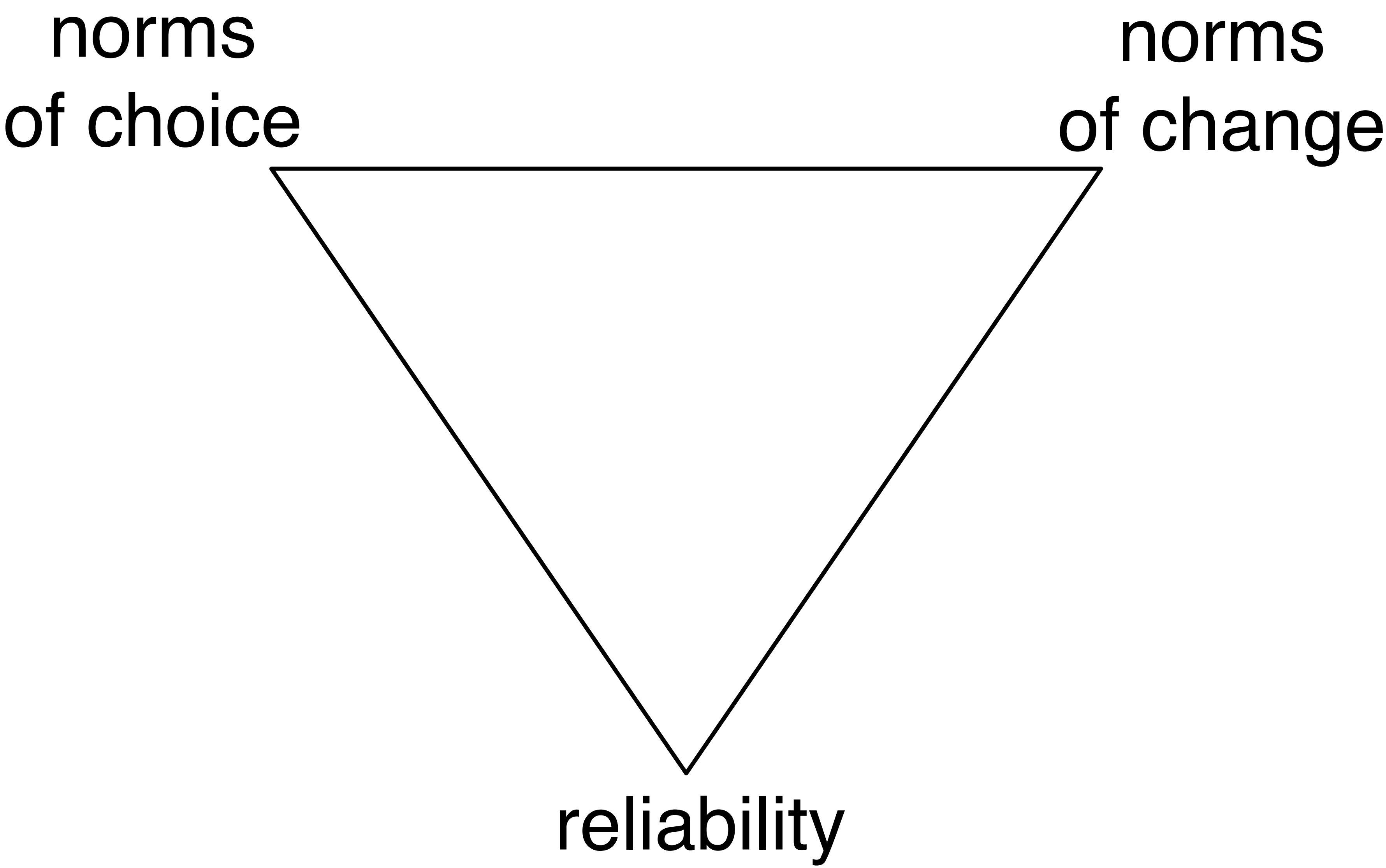
Theory Choice, Theory Change, and Inductive Truth-Conduciveness

Konstantin Genin

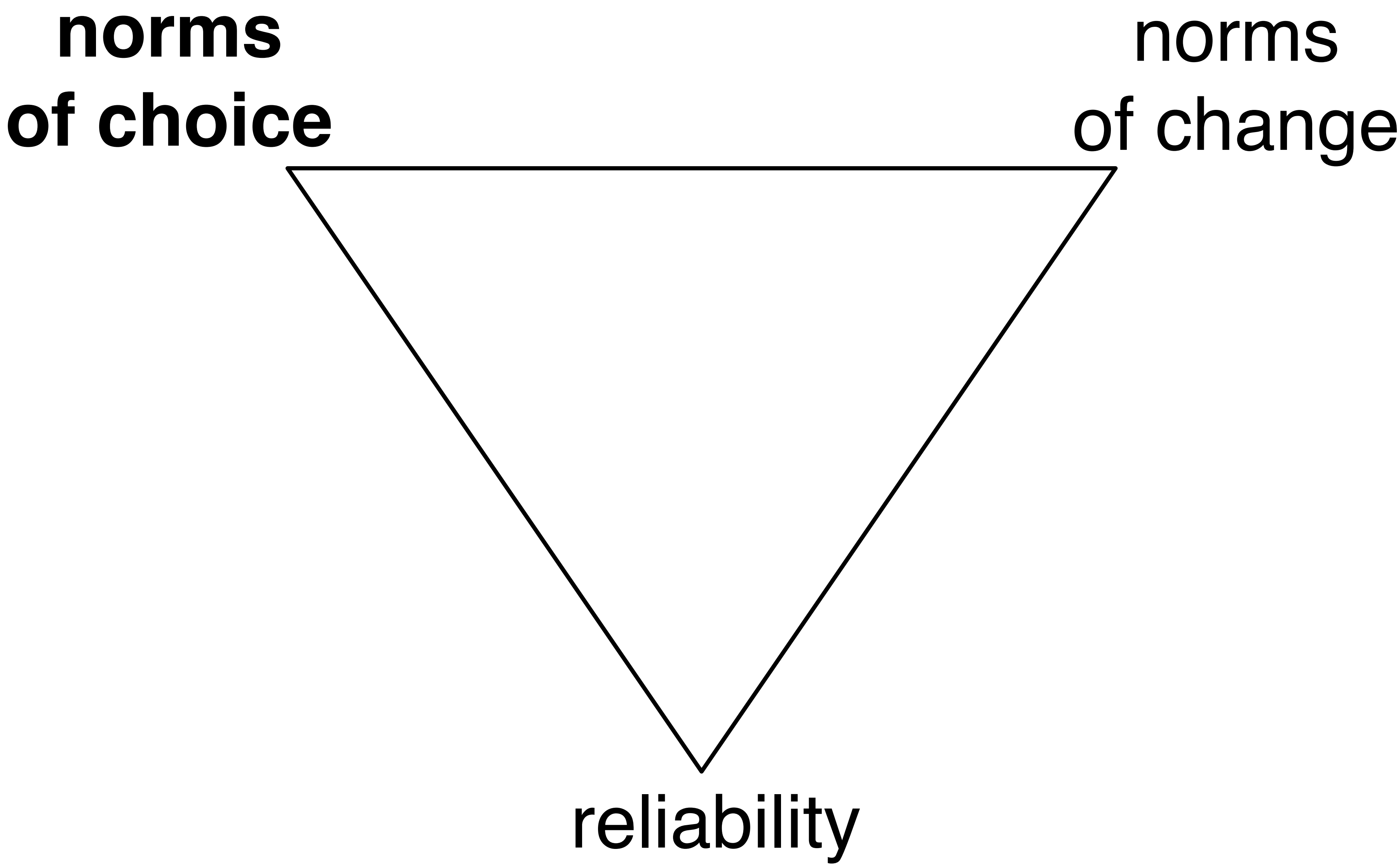
Kevin T. Kelly

Bristol, 2015

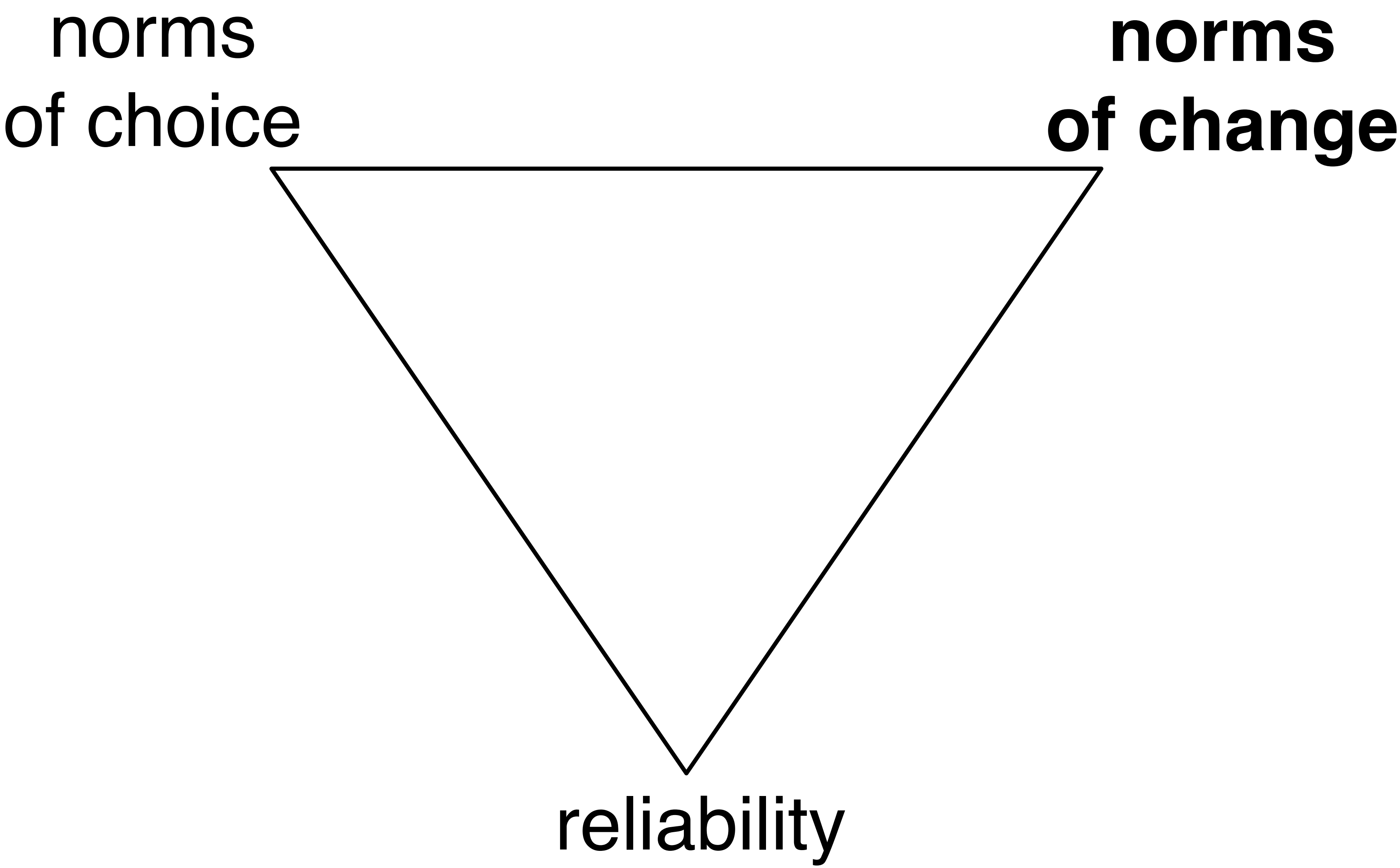
This talk is about ….



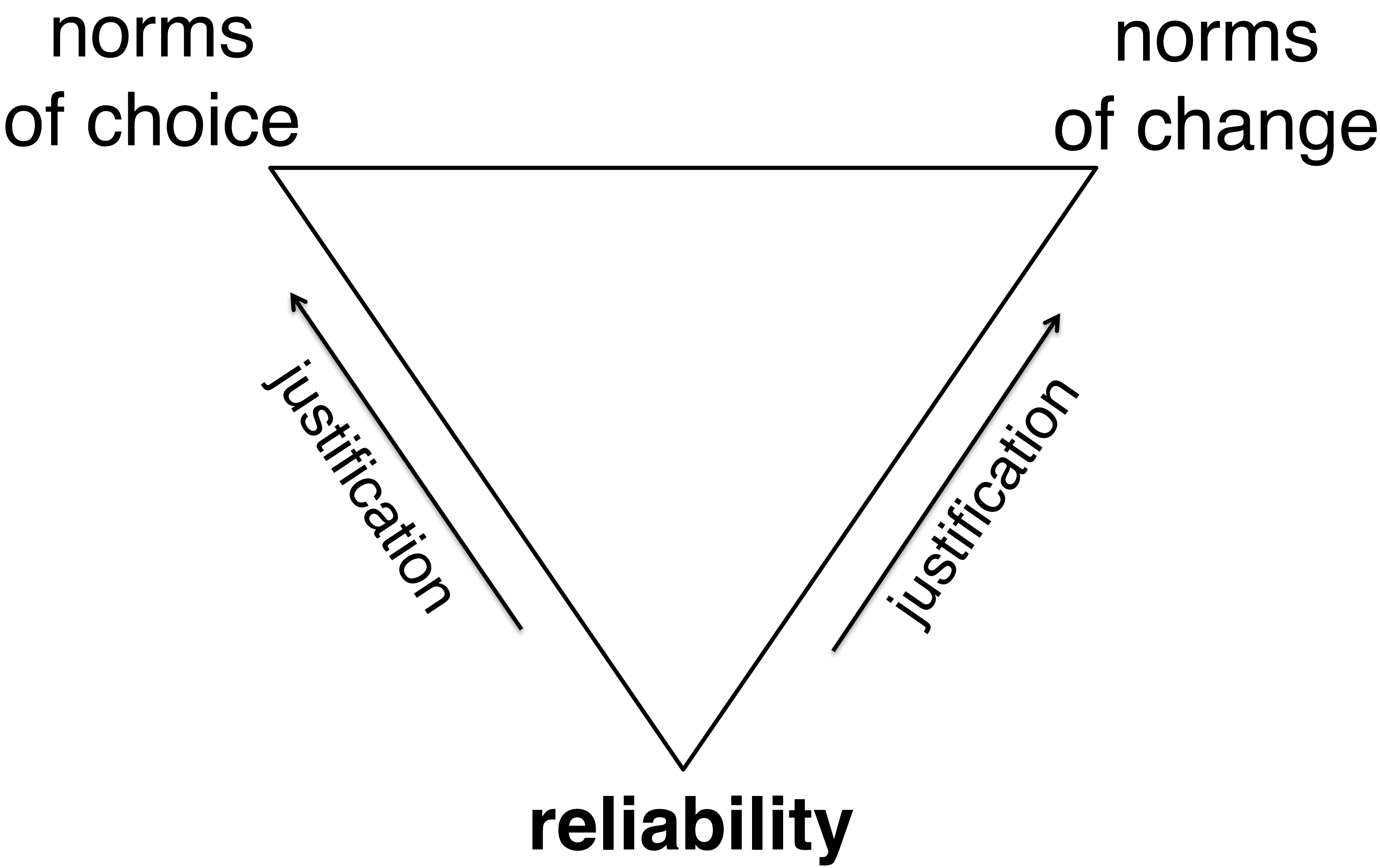
This talk is about ….

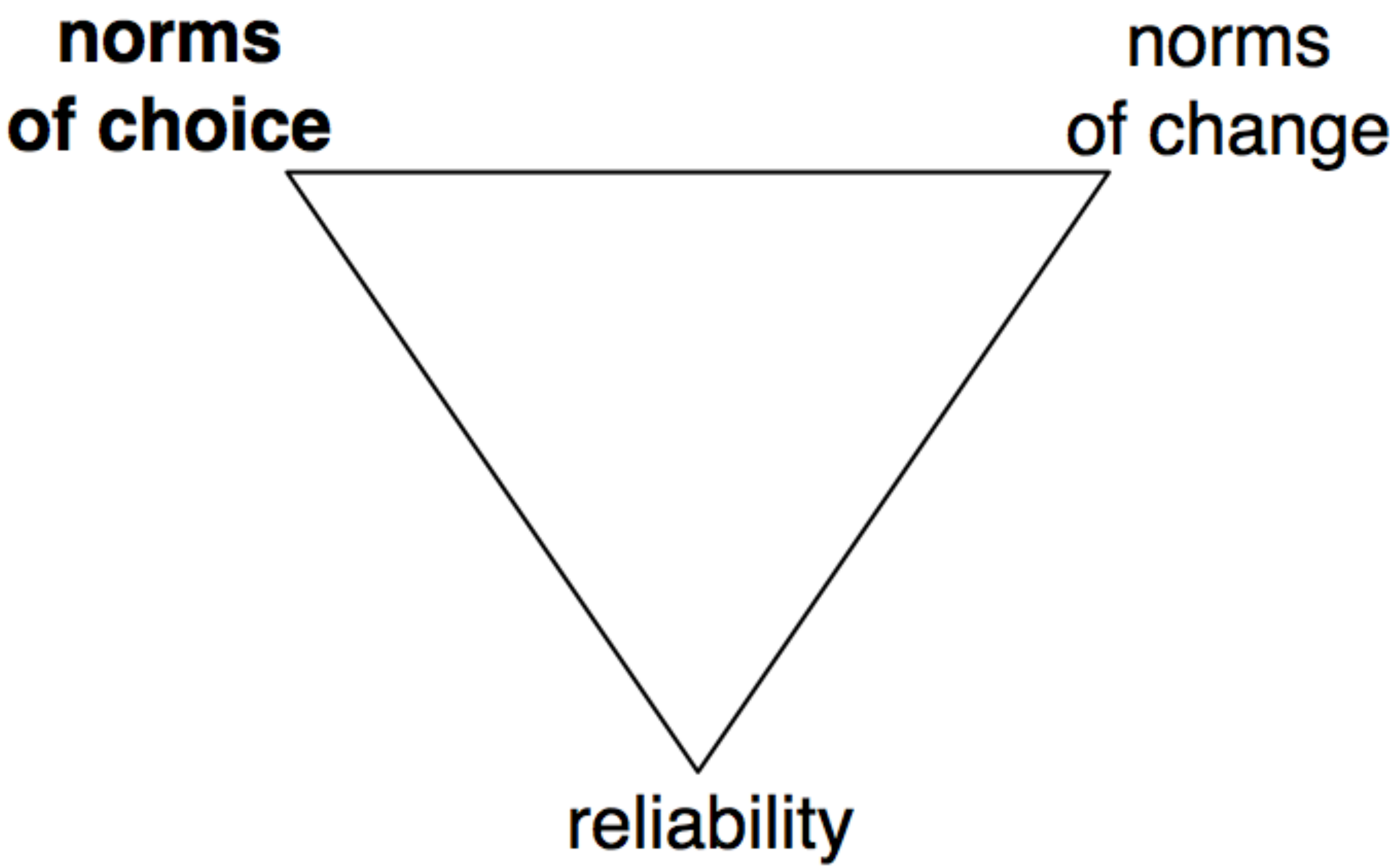


This talk is about ….



This talk is about ....





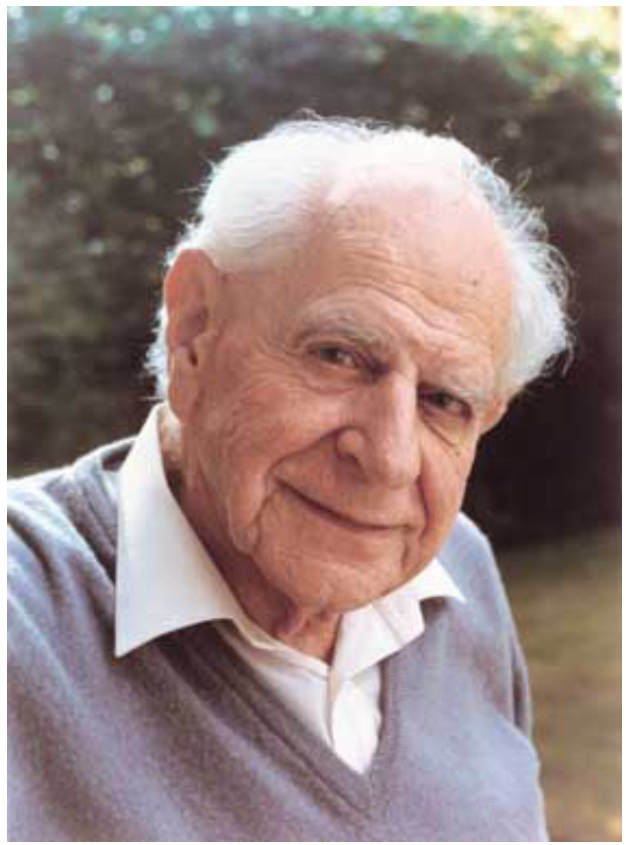
Synchronic norms of theory choice restrict the theories one can choose in light of given, empirical information.

Norm of Choice: Ockham's Razor

- **Ockham:** "Pluralitas non est ponenda sine neccesitate."
- **Science:** "All else equal, prefer simpler theories."
- **Complexity:**
  - Free parameters
  - Multiple mechanisms
  - Coincidences
  - Ad hoc hypotheses

Norm of Choice: Popper's Dictum

- **Popper:** "All else equal, prefer more falsifiable theories."



Reconstruction vs. Reliability

Rational Reconstruction

- Is the simpler theory more plausible?
- Can prior probabilities encode that preference?

Reconstruction vs. Reliability

Rational Reconstruction

- Is the simpler theory more plausible?
- Can prior probabilities encode that preference?
- Of course!

Reconstruction vs. Reliability

Rational Reconstruction

- Is the simpler theory more plausible?
- Can prior probabilities encode that preference?
- Of course!

Reliability

- Does favoring the simpler theory lead one to the truth better than alternative strategies?

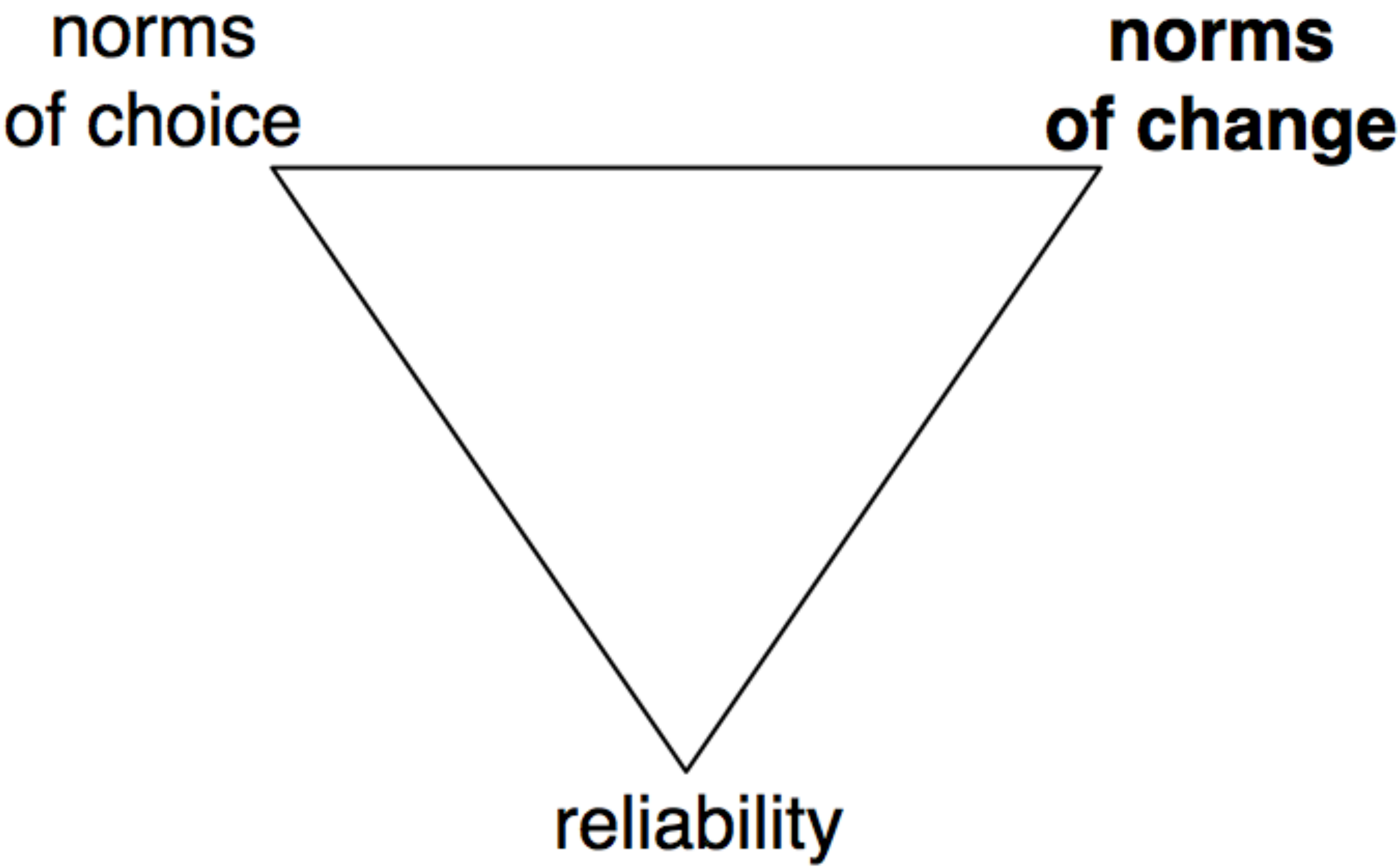
Reconstruction vs. Justification

Rational Reconstruction

- Is the simpler theory more plausible?
- Can prior probabilities encode that preference?
- Of course!

Reliability

- Does favoring the simpler theory lead one to the truth better than alternative strategies?
- How could you show that, without assuming that the world is simple?

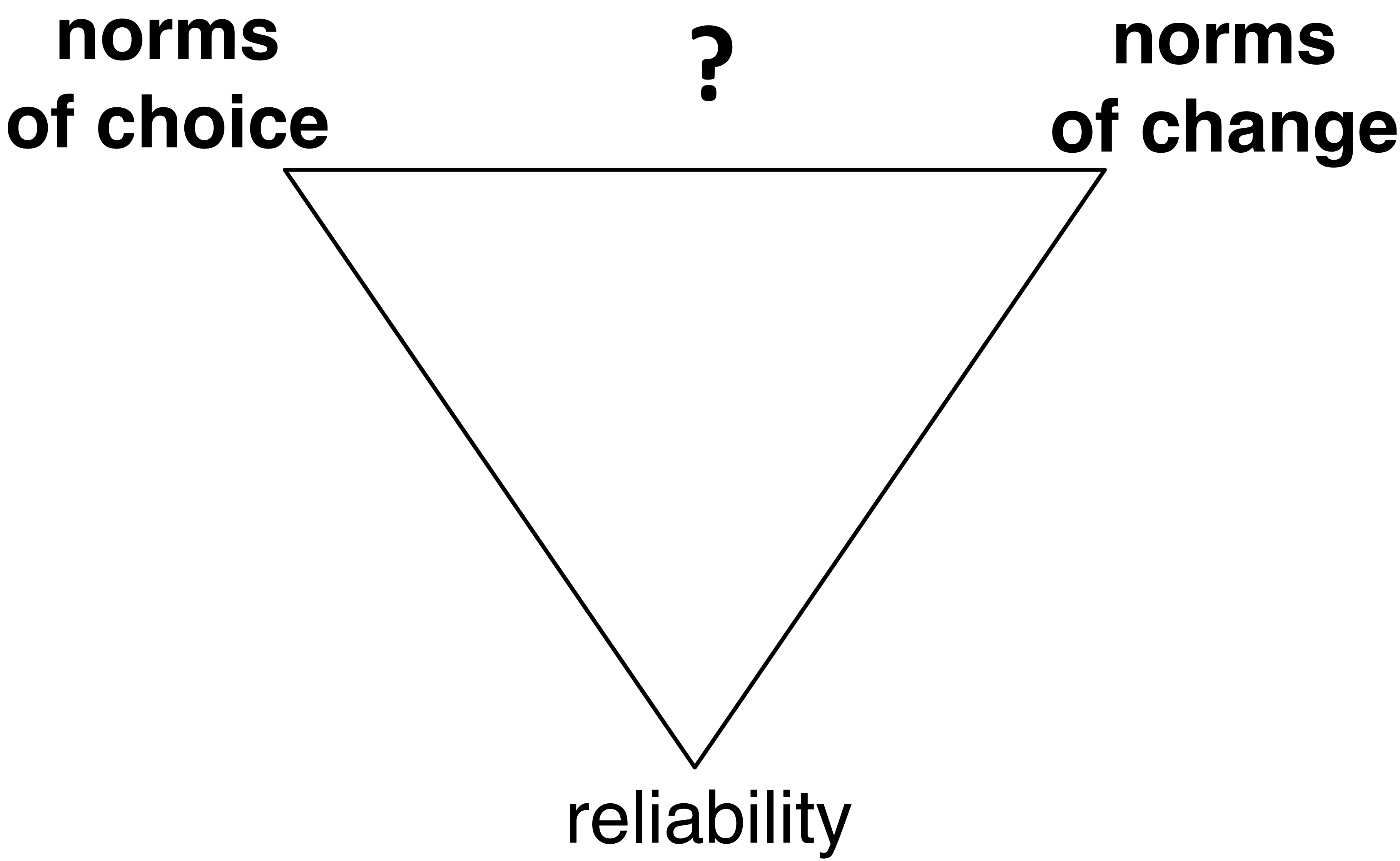


Diachronic norms of theory change govern how one should change one's *current* beliefs, in light of *new* information.

Norm of **Minimal** Change

Alchourrón, Gärdenfors, Makinson:

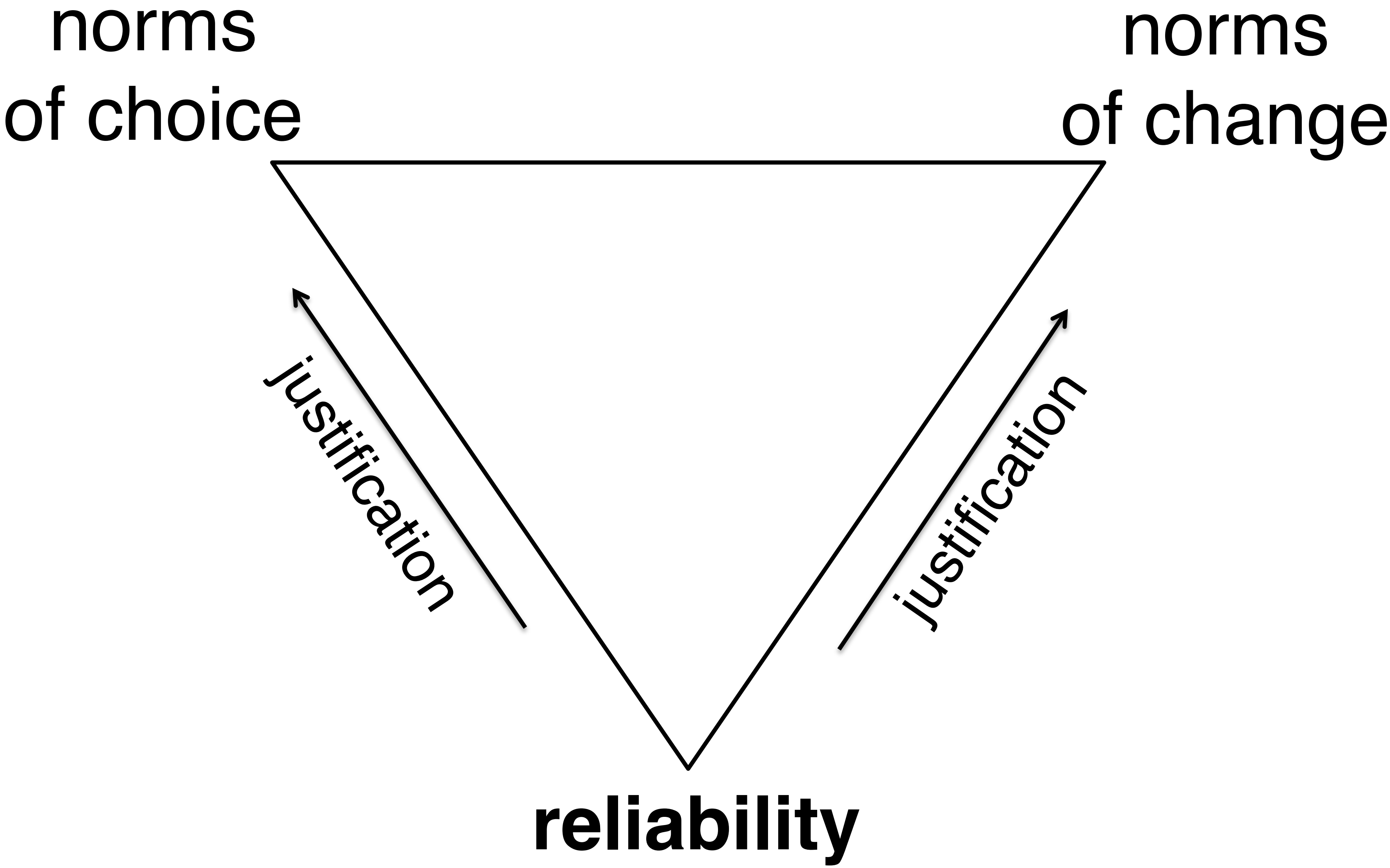
To *rationally* accommodate new evidence, one ought to (1) add only those new beliefs, and (2) remove only those old beliefs, that are *absolutely compelled* by incorporation of new information.



How are the norms of change related to the norms of choice?

norms of choice **?** **norms of change**

It is a strange coincidence that the philosophy of science has focussed on the monadic (nonrelational) features of theory choice, while philosophical logic has emphasized the dyadic (relational) features of theory change. I believe that it is time for researchers in both fields to overcome this separation and work together on a more comprehensive picture (Rott, 2000, p. 15).



Epistemic justification consists in showing that the norms are, in some sense, ***reliable***, or ***truth-conducive***.

Traditionally, truth-conduciveness has been **too strictly** conceived:

Traditionally, truth-conduciveness has been **too strictly** conceived:

> ". . . justifying an epistemic principle requires answering an epistemic question: why are parsimonious theories more likely to be true?" (Baker, 2013)

When your standards are too high, you are led either to metaphysics,

"Nature is pleased with simplicity, and affects not the pomp of superfluous causes" (Newton et al., 1833).

… or despair.

"[N]o one has shown that any of these rules is more likely to pick out true theories than false ones. It follows that none of these rules is epistemic in character" (Laudan, 2004).

Truth Conduciveness: Too Strong

- Theoretical virtues do not ***indicate*** the truth the way litmus paper indicates pH.
- Inductive inferences made in accordance with the rationality principles are still subject to ***arbitrarily high chance of error***.

Truth Conduciveness: Too Strong

We can make progress if we don't demand the impossible:

> "The fact that the truth of the predictions reached by induction cannot be guaranteed does not preclude a justification in a weaker sense" (Carnap, 1945).

Truth Conduciveness: Too Weak

- Truth-indicativeness is too strong a standard. But mere convergence to the truth in the limit is too weak to mandate any behavior in the short run.

> "Reichenbach is right ... that any procedure, which does not [converge in the limit] is inferior to his rule of induction. However, his rule ... is far from being the only one possessing that characteristic. The same holds for an infinite number of other rules of induction. ... Therefore **we need a more general and stronger method for examining and comparing any two given rules of induction** ..." (Carnap, 1945)

Truth Conduciveness: Just Right

Truth-
Indicative

?

Converges
In the limit

Is there something in between?

Reasoning

Deductive Reasoning

- Non-ampliative
- Infallible
- Monotonic

Reasoning

Deductive Reasoning

- Non-ampliative
- Infallible
- Monotonic

Inductive Reasoning

- Ampliative
- Fallible
- Non-monotonic

Reliability

Deductive Reliability

- Converge to the truth directly
- Information determines the right answer

Reliability

Deductive Reliability

- Converge to the truth directly
- Information determines the right answer

Inductive Reliability

- Converge to the truth indirectly
- Anything goes in the short run.

Reliability

Deductive Reliability

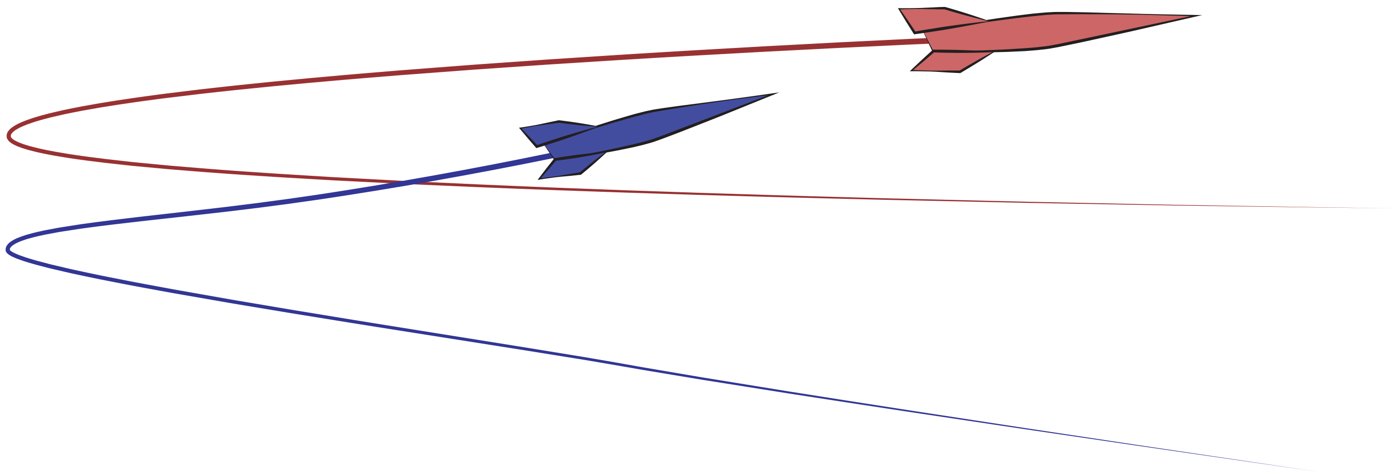
- Converge to the truth directly
- Information determines the right answer

Optimal Inductive Reliability

- Converge to the truth as directly as possible.
- Implies strong short-run norms. E.g....

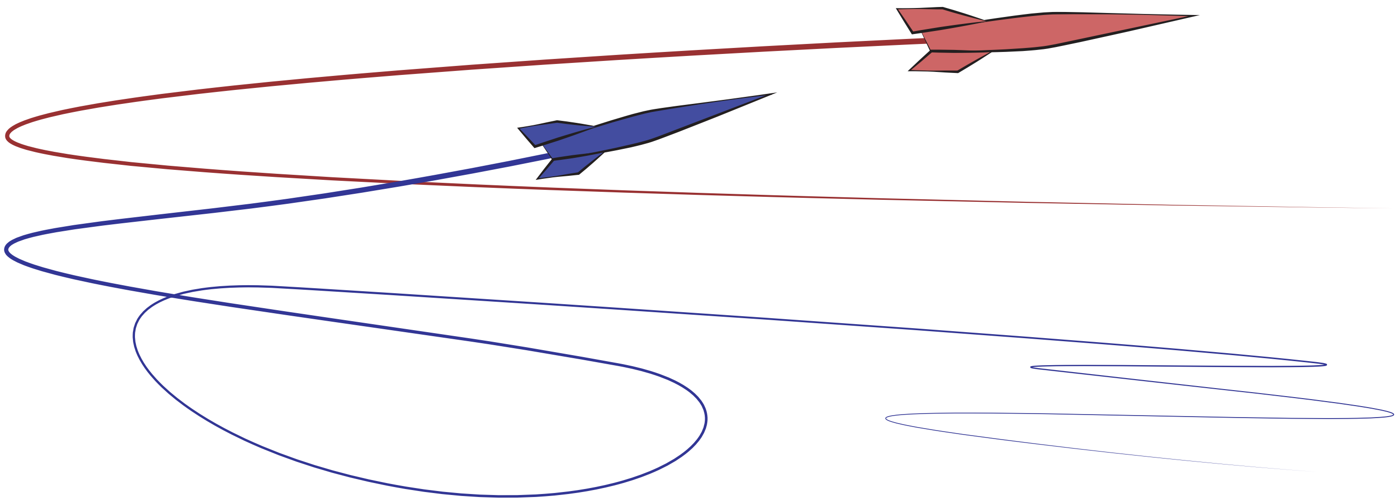
Reliability

Pursuit of truth ought to be as direct as possible.

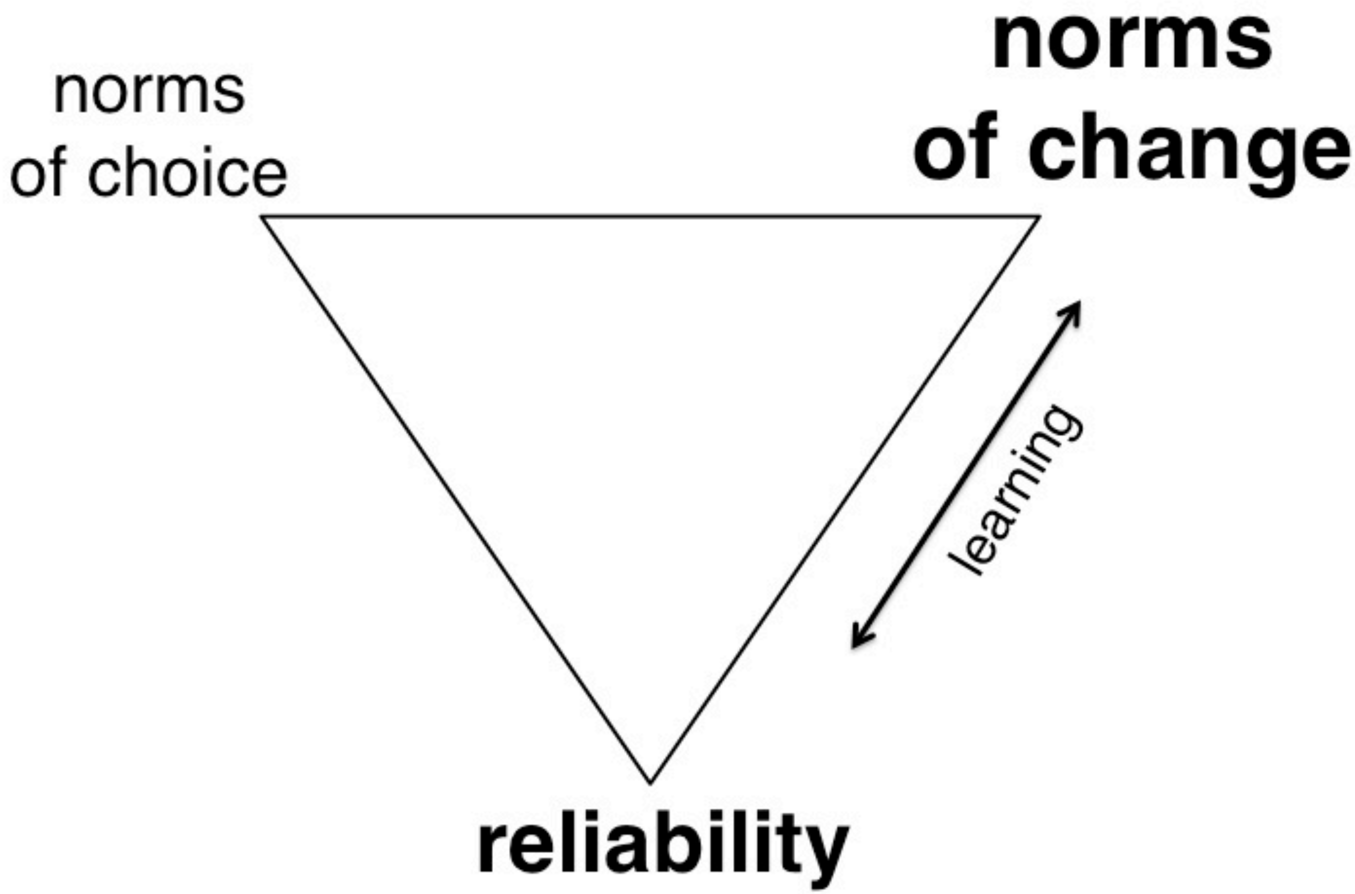


Reliability

Needless cycles and reversals ought to be avoided.

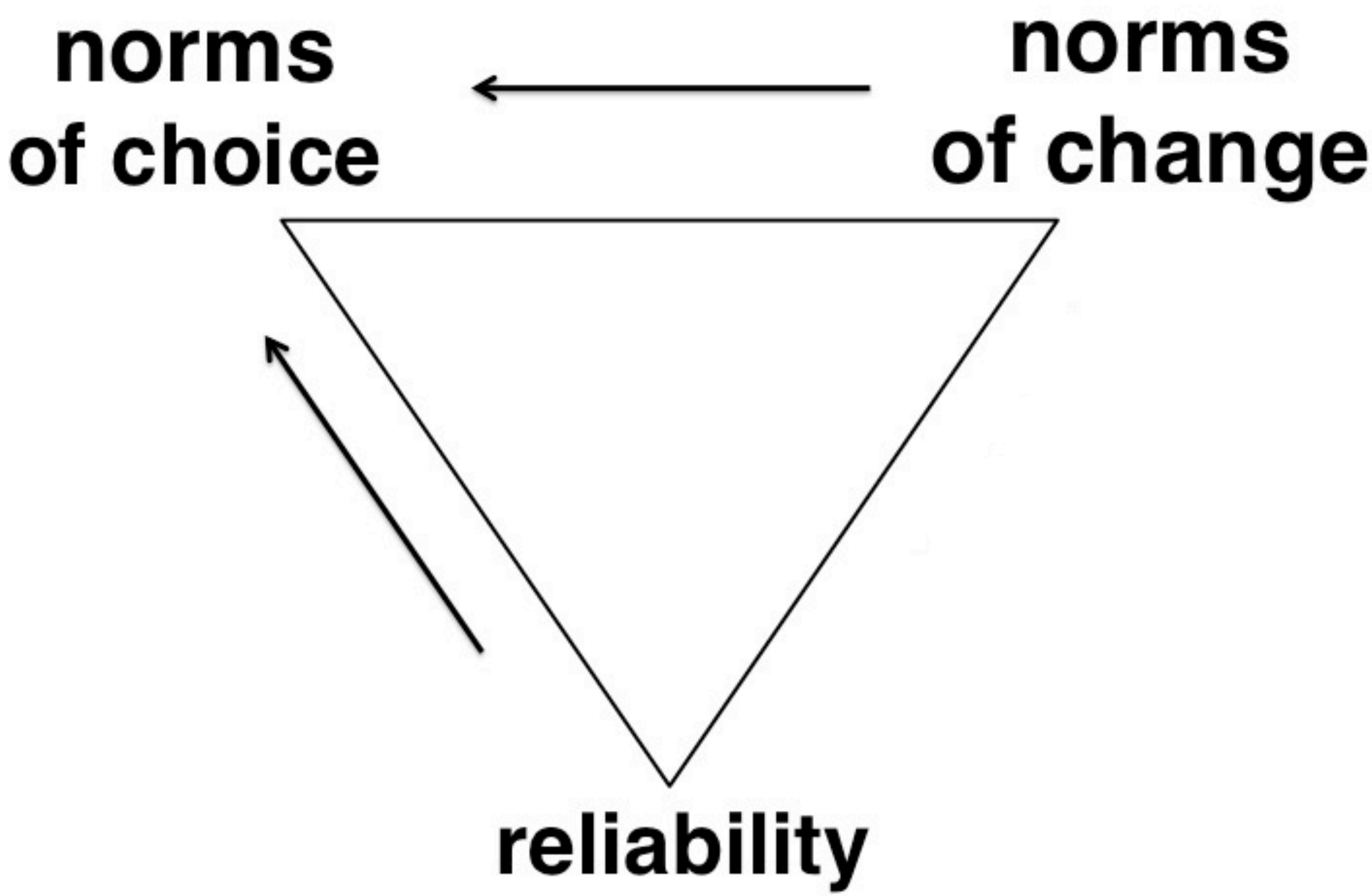


Main Thesis



Once limiting convergence is imposed, cycle-avoidance is **equivalent** to a norm of change.

Main Thesis

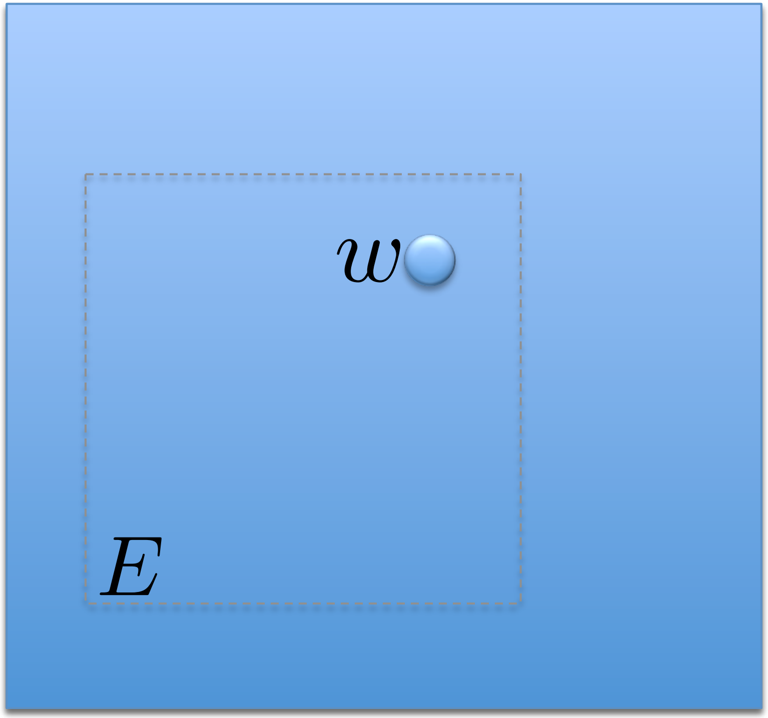


Both necessitate a preference for **simpler,** and **more falsifiable** theories.

1. EMPIRICAL PROBLEMS

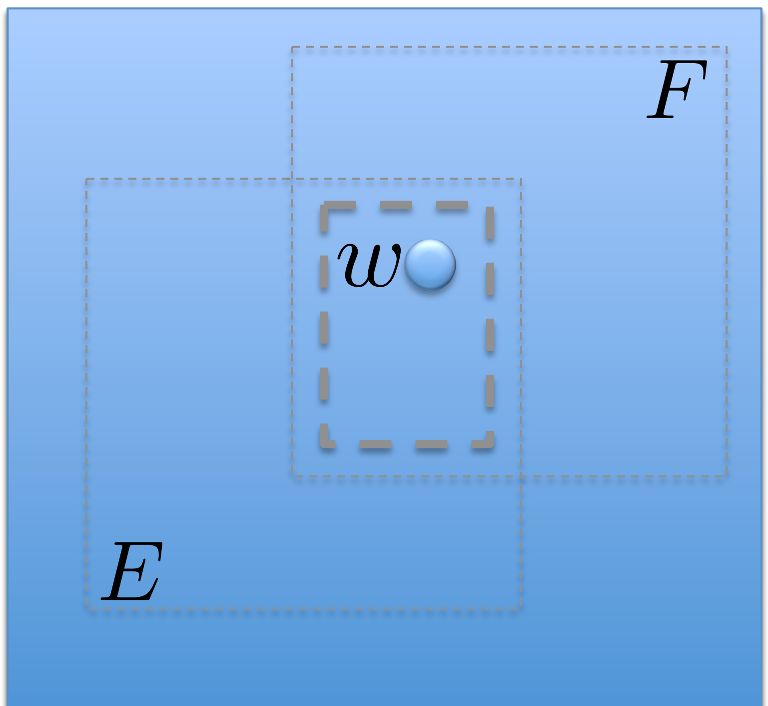
Information Spaces

- W is a set of possible worlds.
- A proposition is a set of possible worlds.
- $\mathcal{I}$ is a countable set of propositional information states.



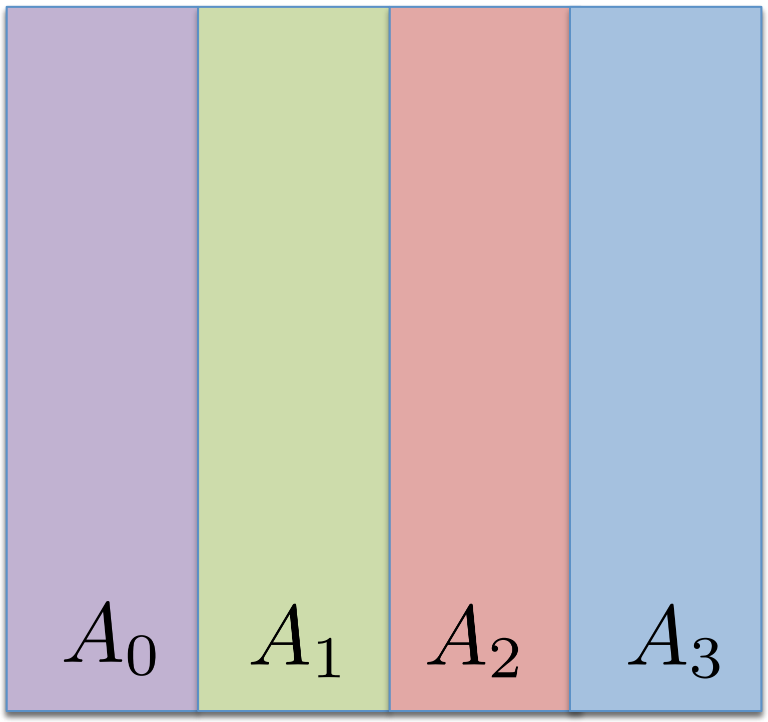
Axioms on Information

1. **Existence:** Each world makes some information state true.
2. **Cumulativity:** Each finite conjunction of information states true in w is entailed by an information state true in w.



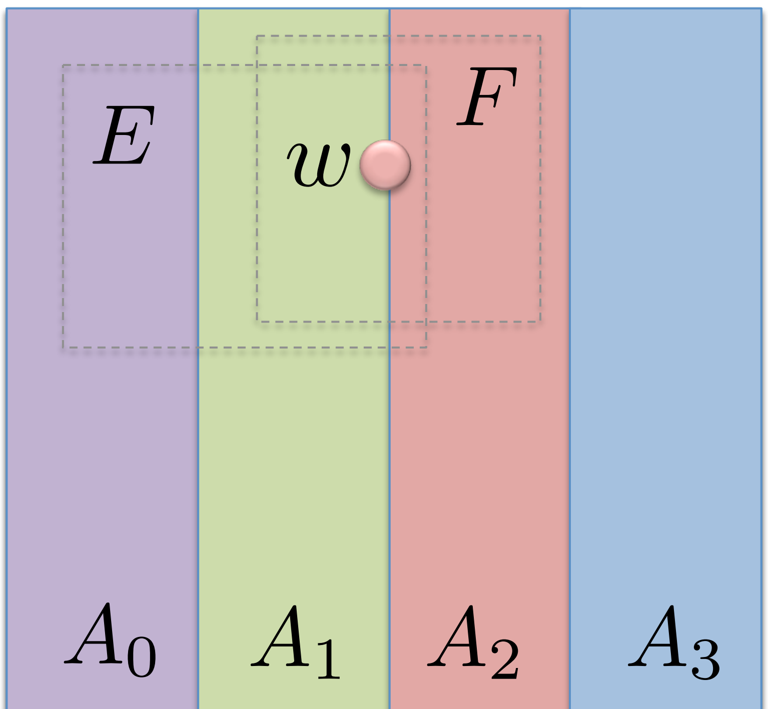
Questions

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

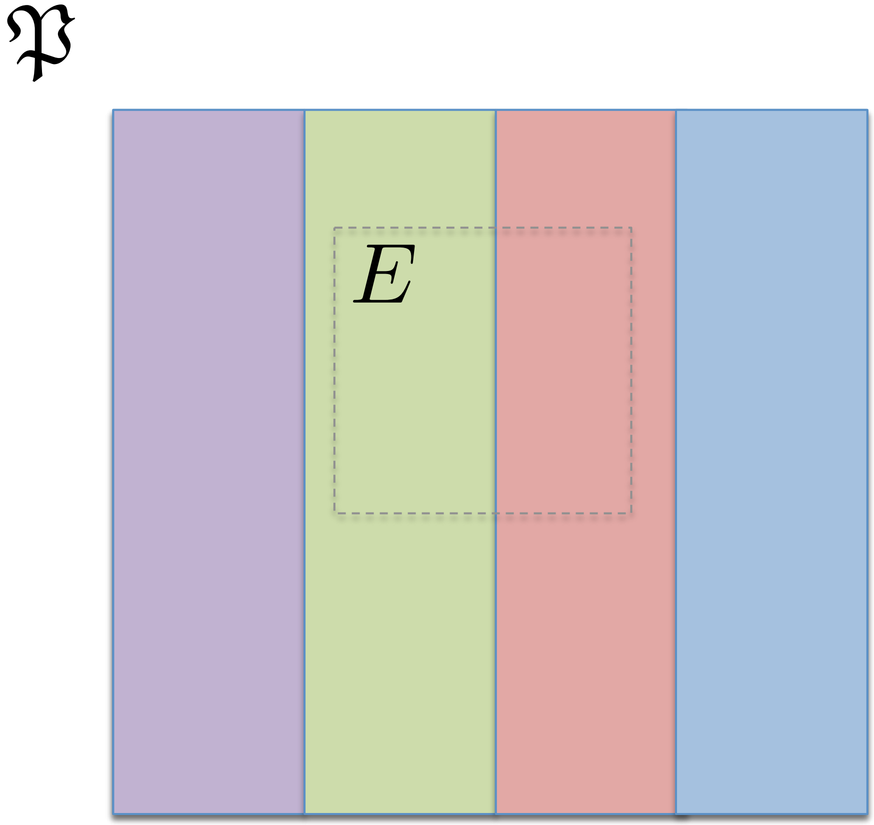


Empirical Problem

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

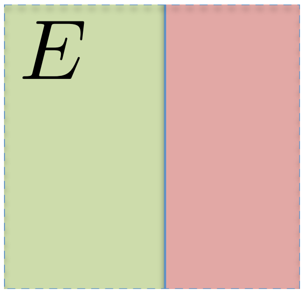


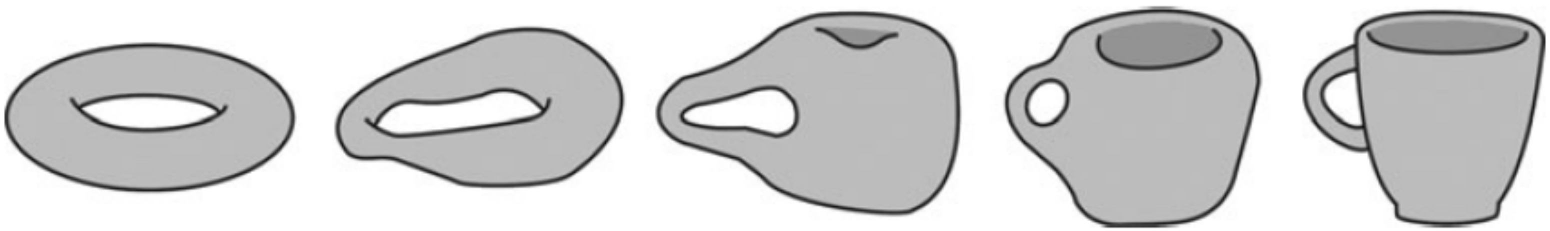
Problem Restriction



Problem Restriction

$\mathfrak{P}|_E$

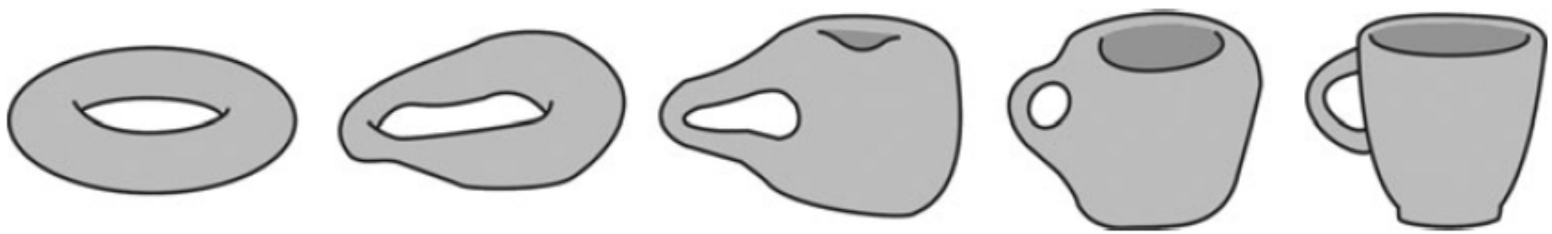




2. INFORMATION TOPOLOGY

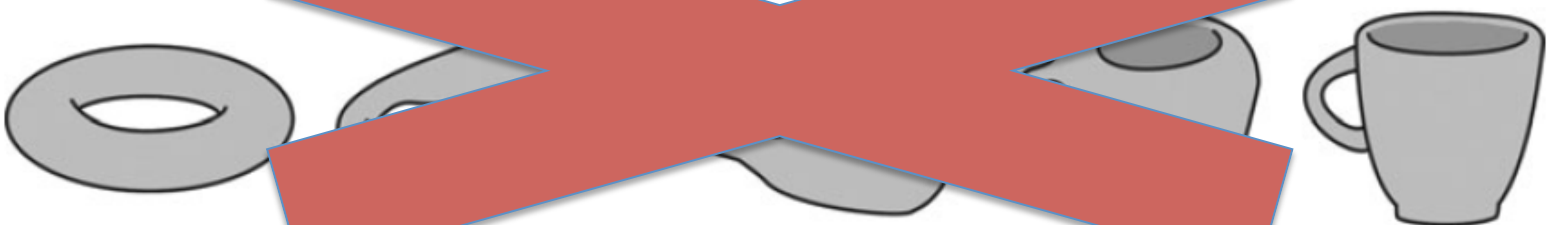
Topology

Rubber geometry



Topology

~~Rubber geometry~~



Logic of verification (Kelly 1996, Vickers 1996).

Topology and Underdetermination

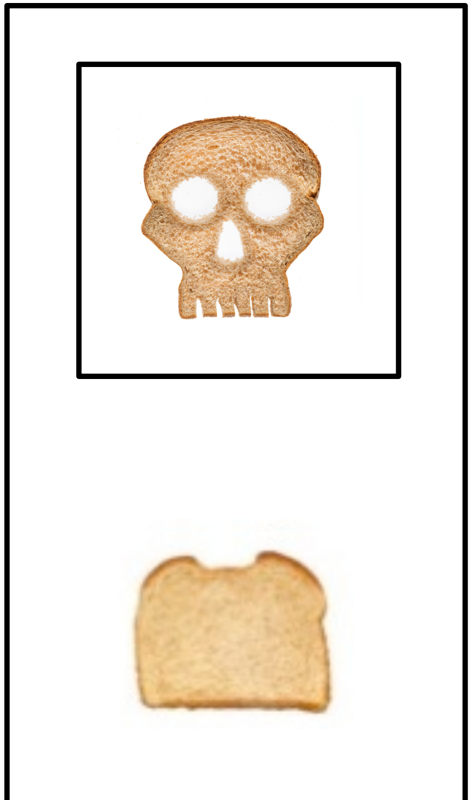
The bread, which I formerly ate, nourished me ... but does it follow, that other bread must also nourish me at another time, and that like sensible qualities must always be attended with like secret powers? The consequence seems nowise necessary (Enquiry Concerning Human Understanding).

Hume, Topologized.

Hume, Topologized.

Hume, Topologized.

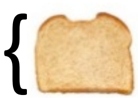
Hume, Topologized.



Interior of A

Int A = it will be verified that A.

Int {  } = {  }

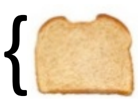
Int {  } = $\varnothing$

Open = Verifiable

A is open iff A entails that A will be verified

iff A entails Int A.

Int {} = {}

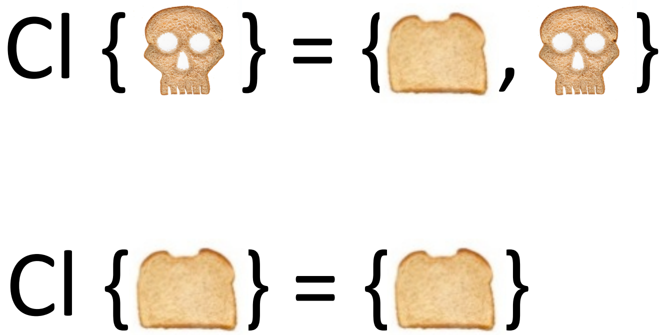
Int {} = $\emptyset$

Closure of A

Cl A = A will never be refuted

= not-A will never be verified

= not Int not A.

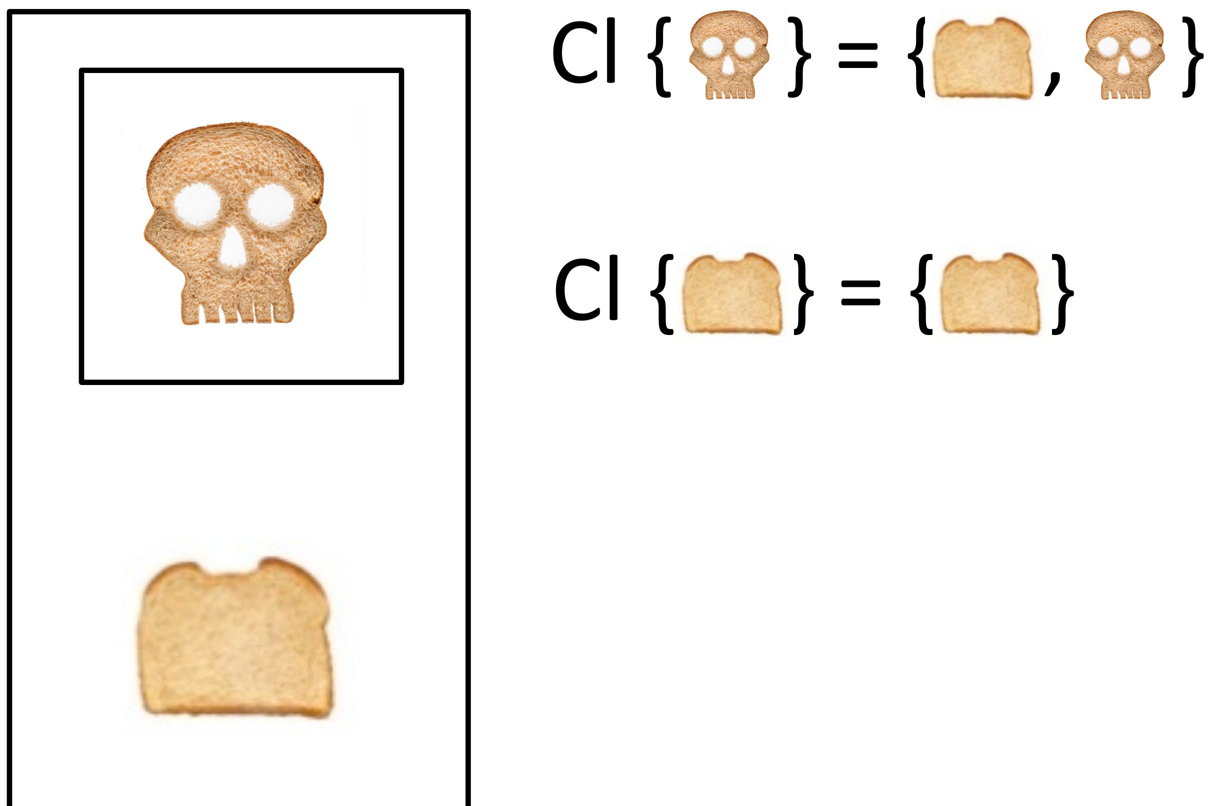


Closed = Refutable

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

iff that A will never be refuted entails A

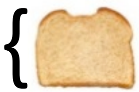
iff Cl A entails A.

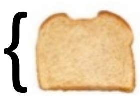


Frontier of A

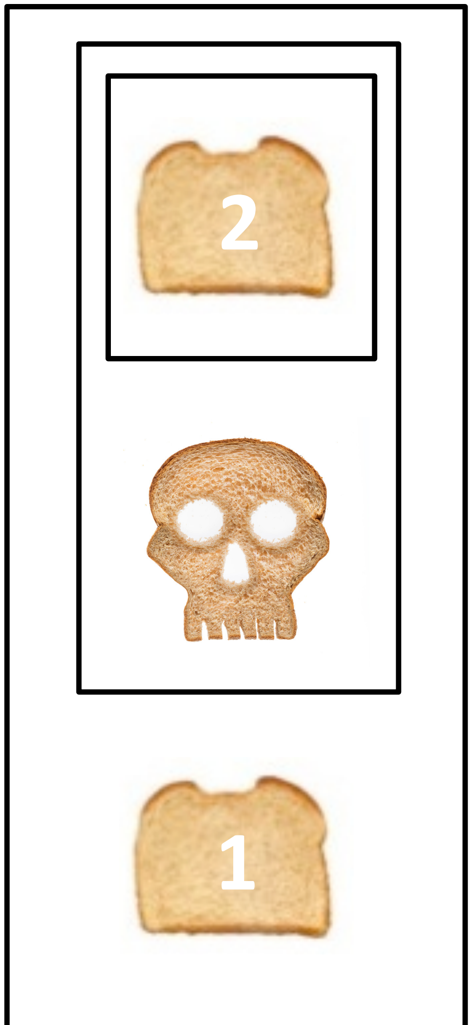
Frntr A = A is false, but will never be refuted

= Cl A, but not A.

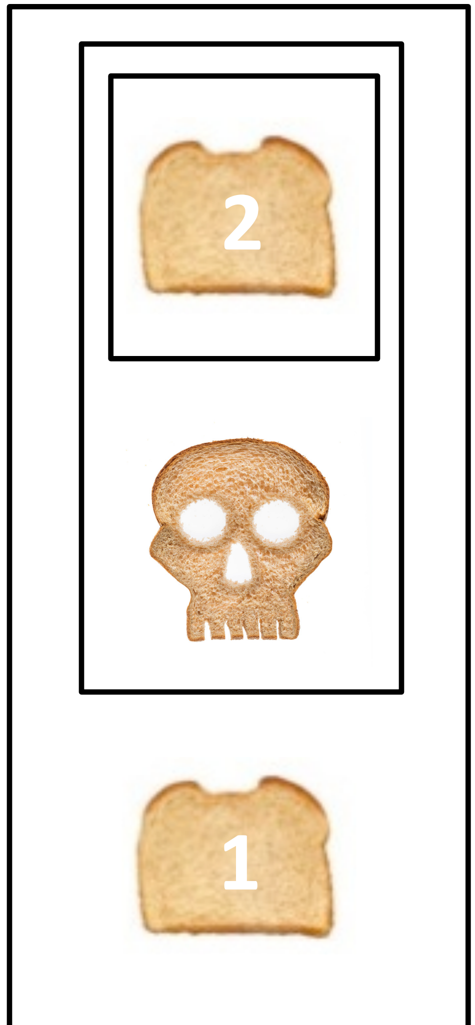
Frntr {  } = {  }

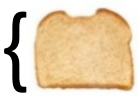
Frntr {  } = $\varnothing$

Hume's Problem, Enhanced.

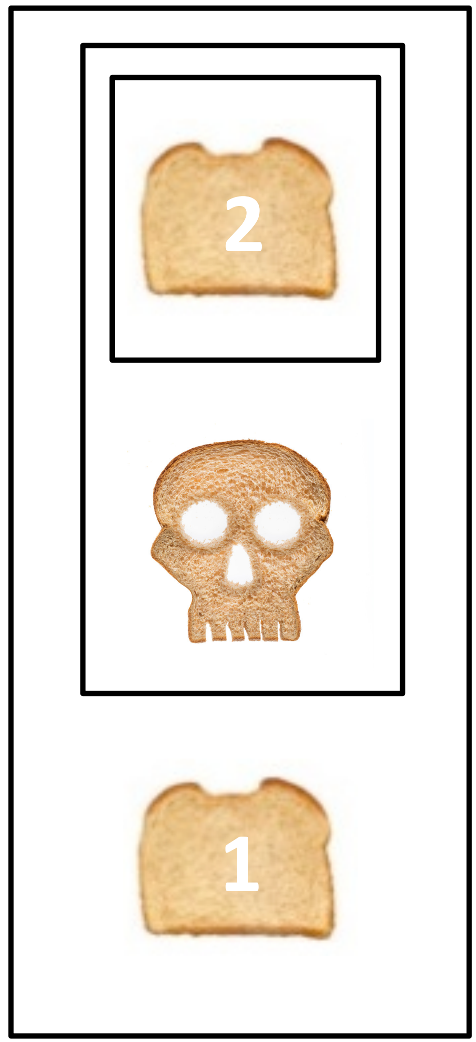


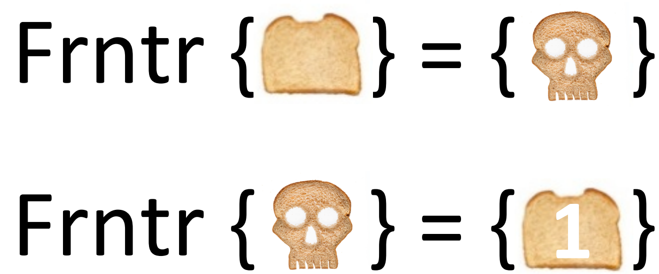
Hume’s Problem, Enhanced.



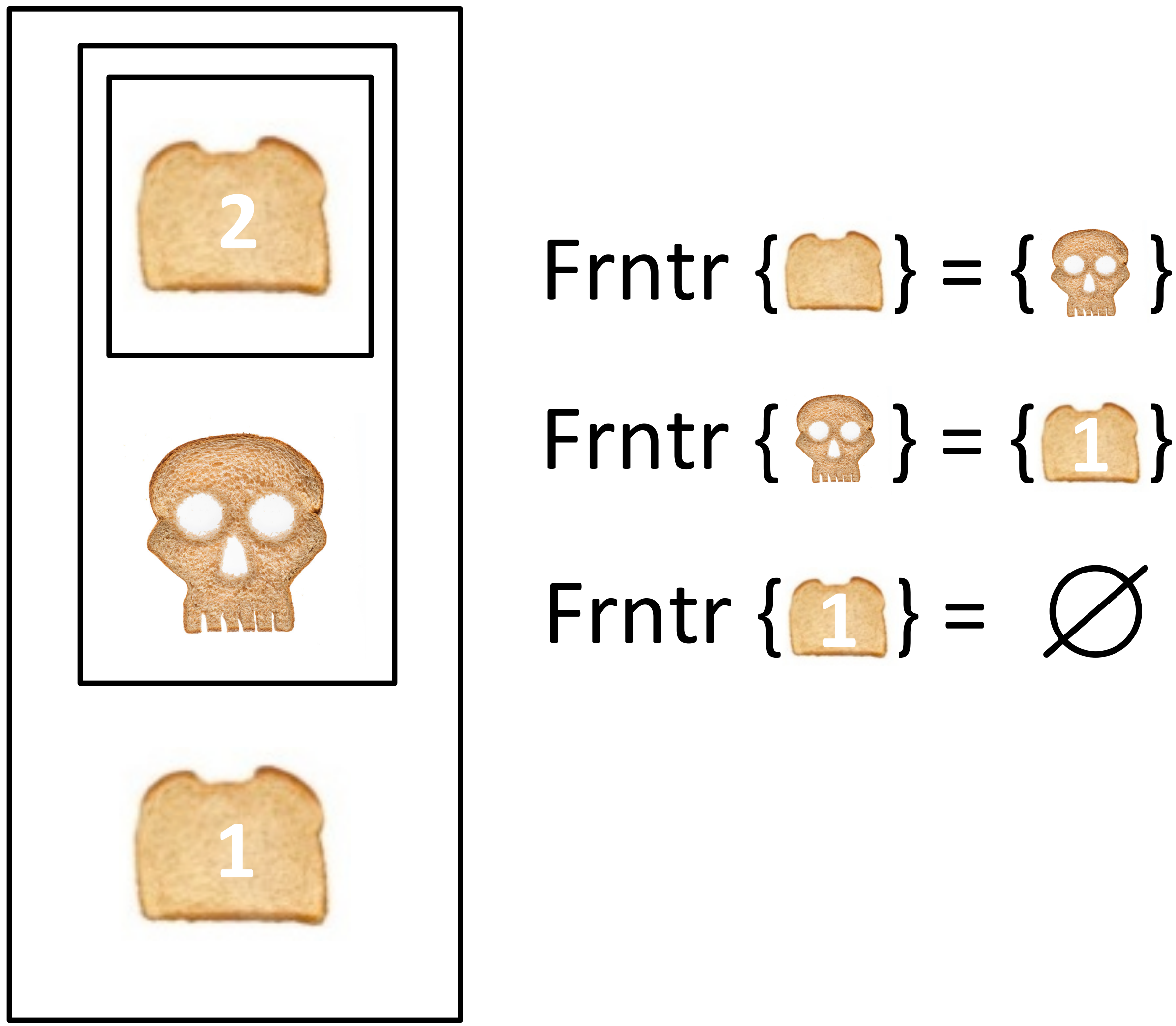
Frntr {} = {}

Hume’s Problem, Enhanced.



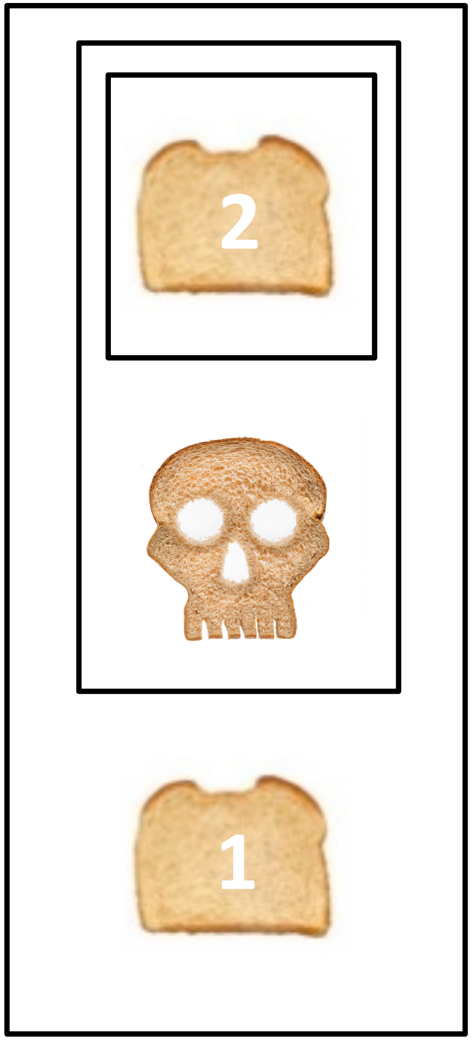


Hume’s Problem, Enhanced.



Locally Closed

A is locally closed iff Frntr A is closed.



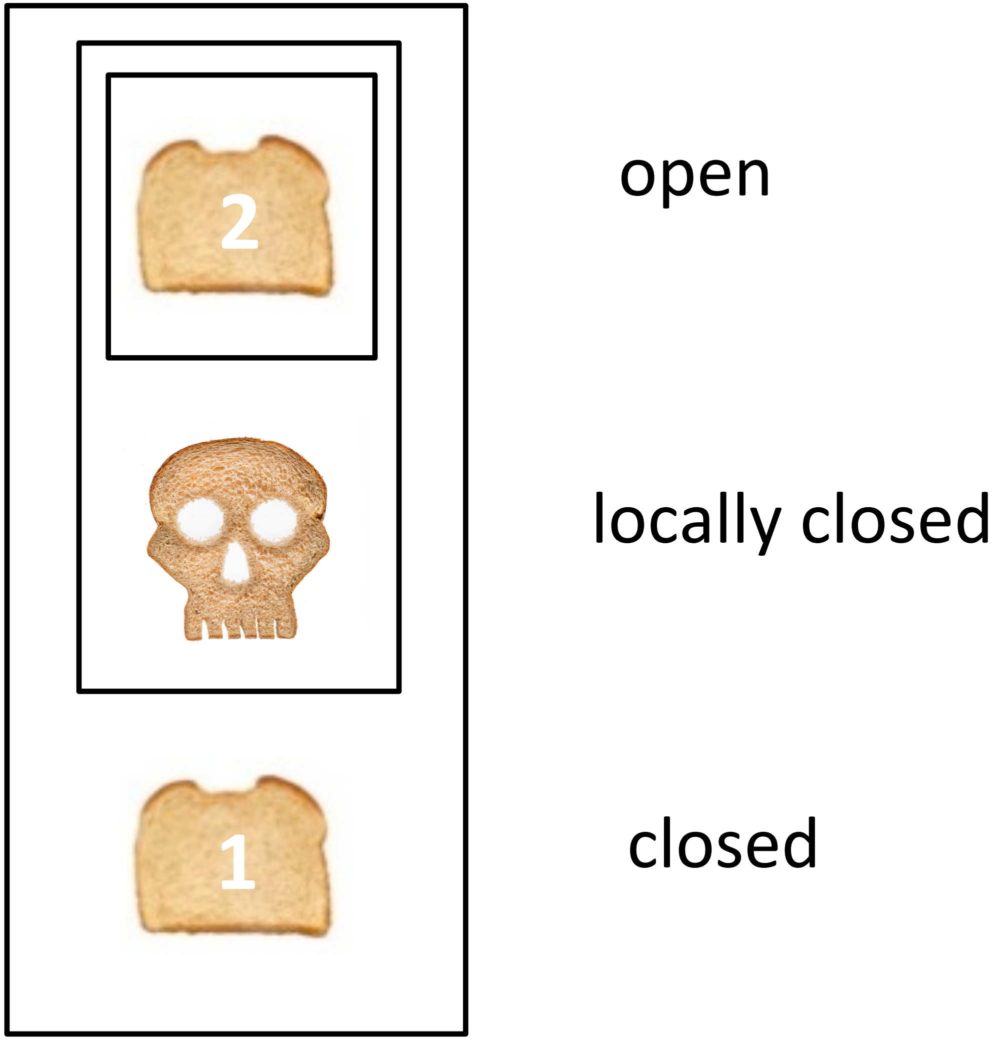
open

locally closed

closed

Locally Closed

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



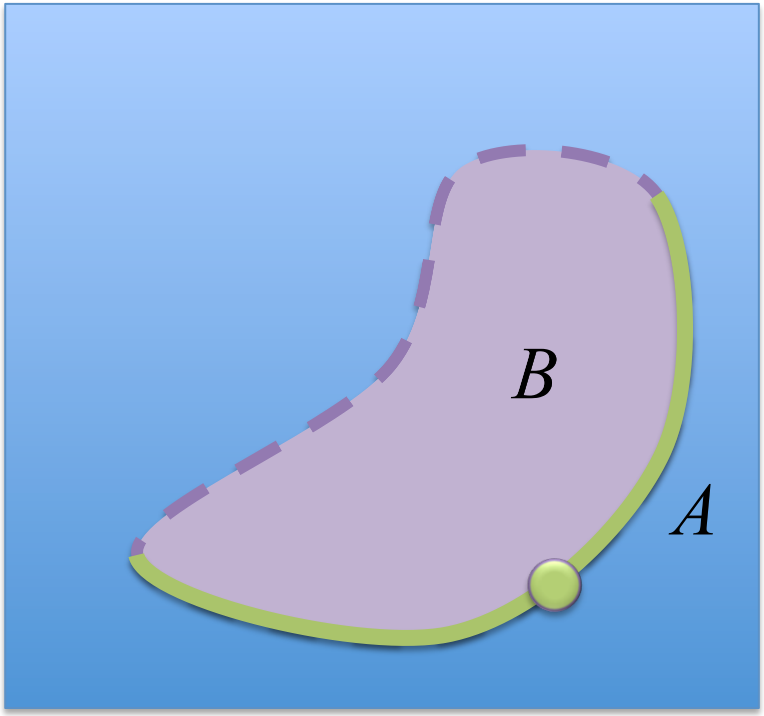
3. EMPIRICAL SIMPLICITY

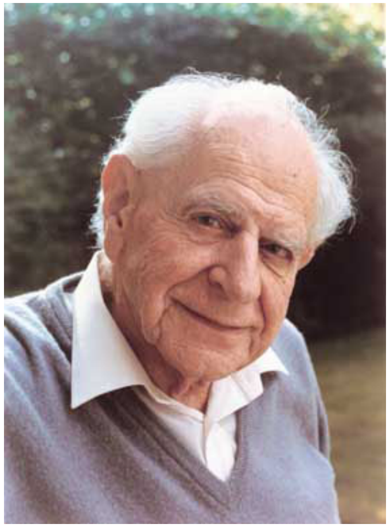
Simpler = More Falsifiable

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

iff all information compatible with A is compatible with B

iff all information refuting B also refutes A.

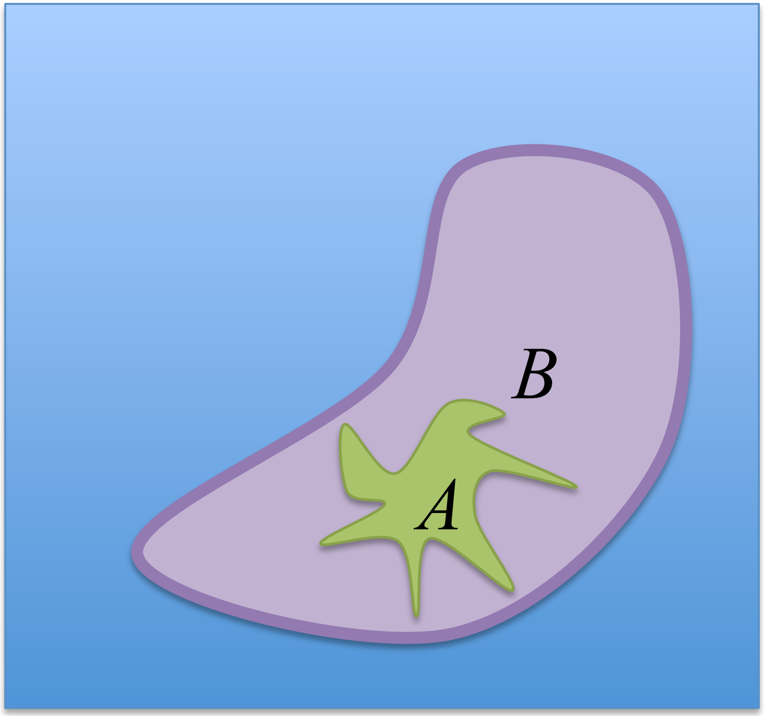


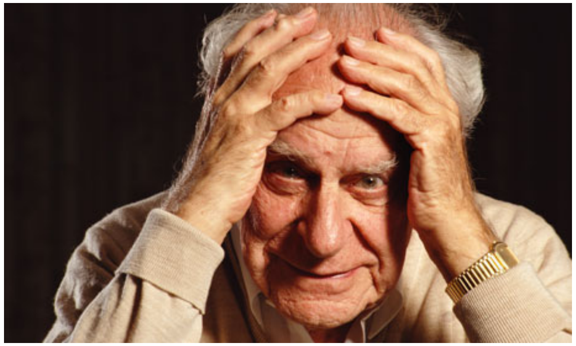


Sir Karl Popper

The "Tack-on" Objection

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



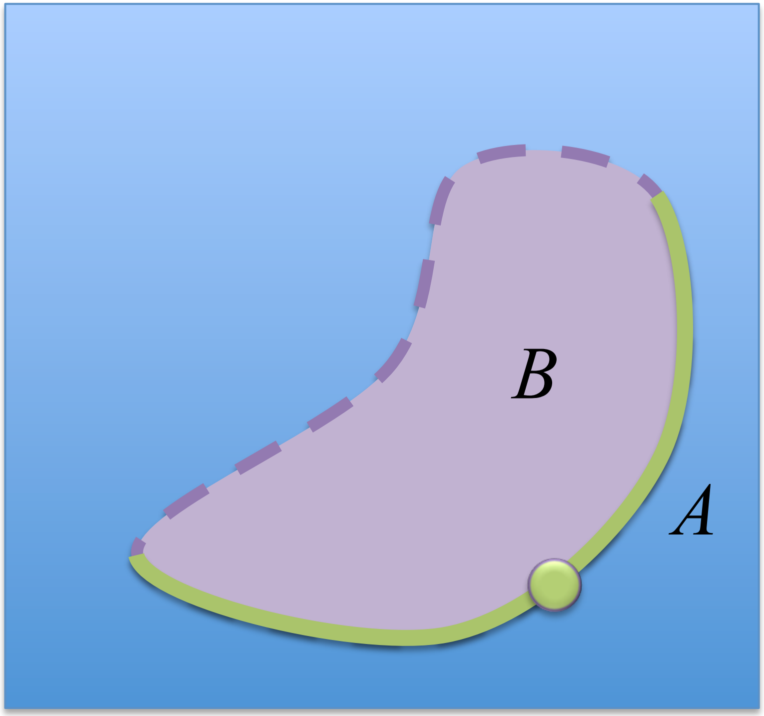


Improved Definition

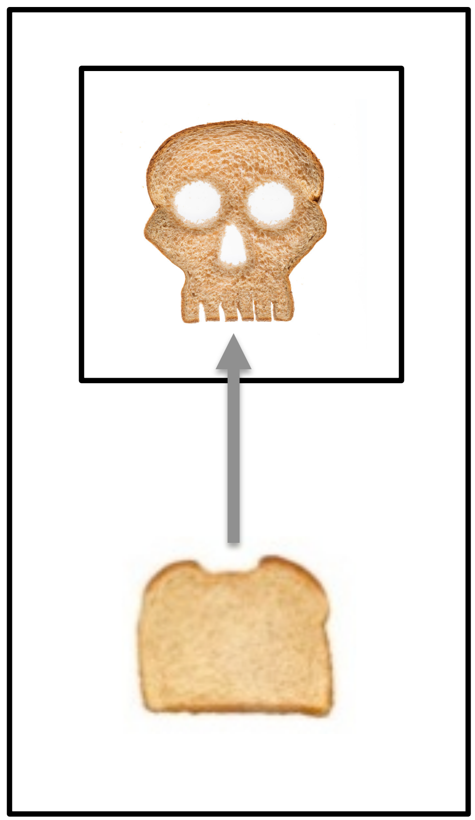
$A \prec B$ iff $A \subseteq \mathsf{Frntr} B$.

iff A entails that B is false

but will never be refuted.



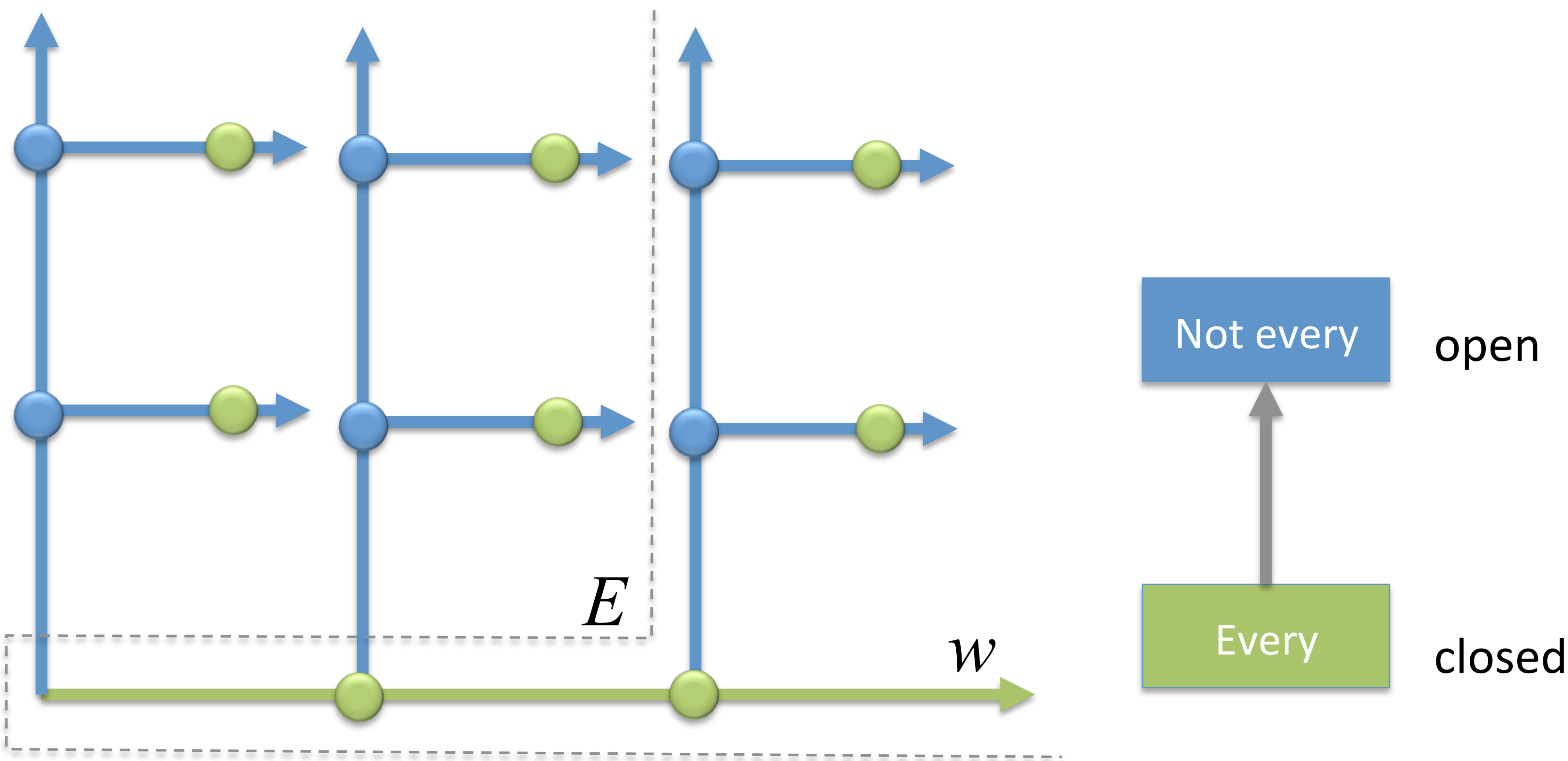
Example: Hume's Problem



Example: Discrete Outcomes

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

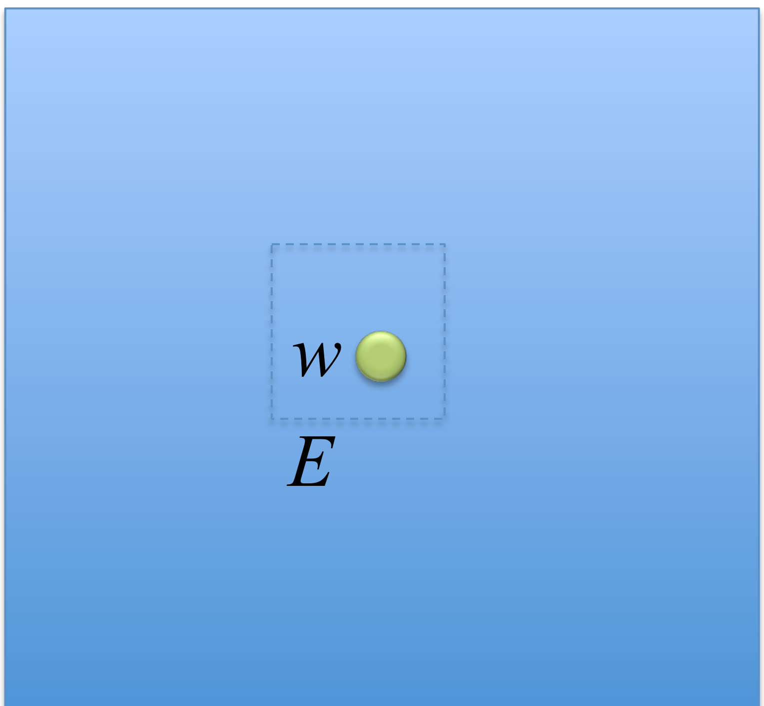
$\mathcal{I} =$ observation histories.

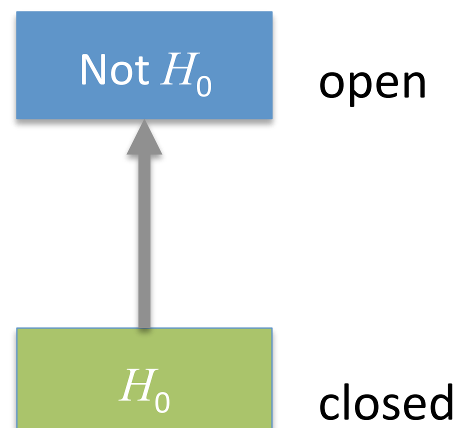


Example: Real Parameter

$\mathcal{Q} =$ Is the sharp null hypothesis true?

$\mathcal{I} =$ open rectangular estimates.

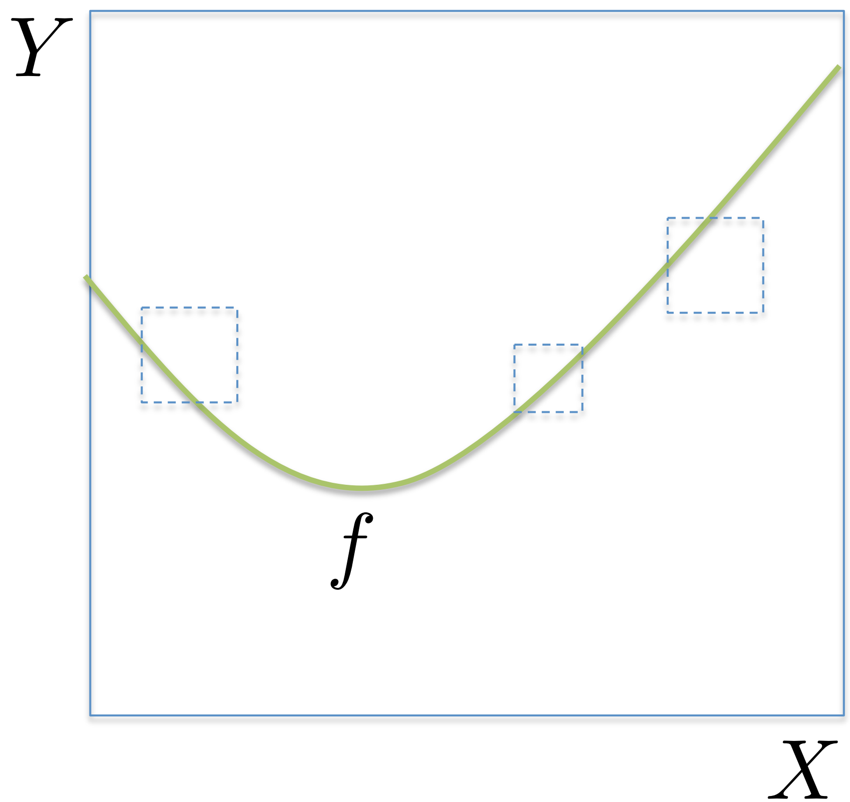


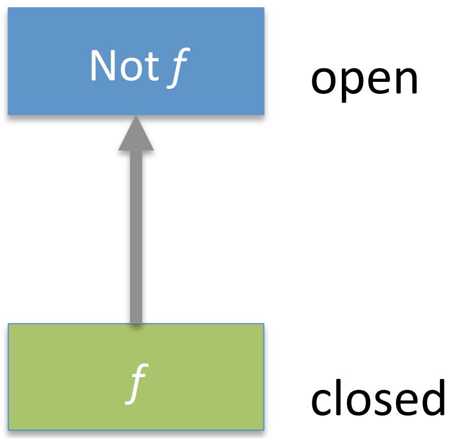


Example: Continuous Laws

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

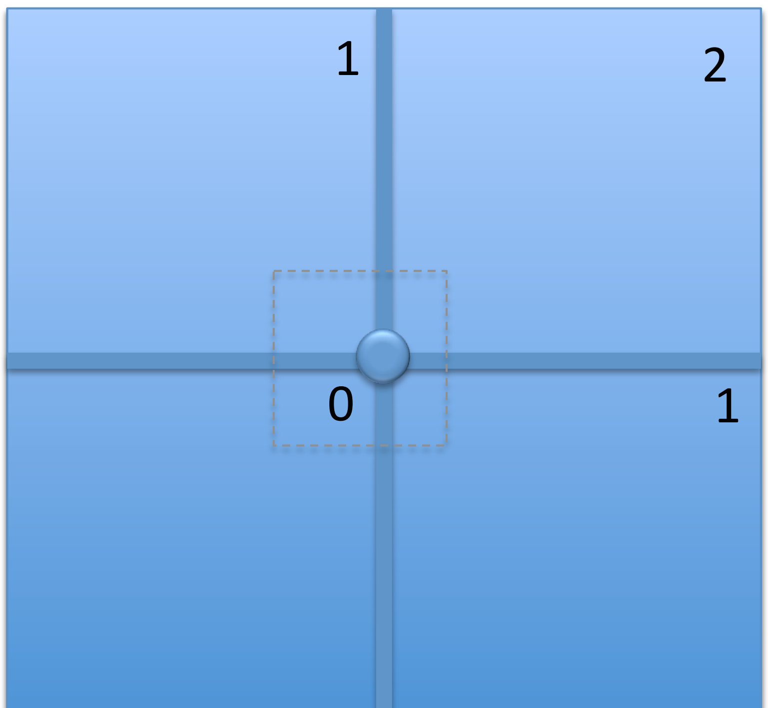
$\mathcal{I} =$ finitely many inexact measurements.

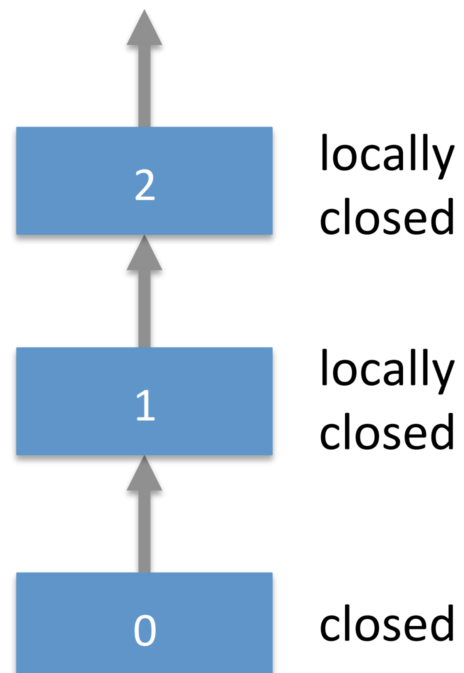




Example: Parametric Models

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





Example: Quantitative Laws

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

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

degree

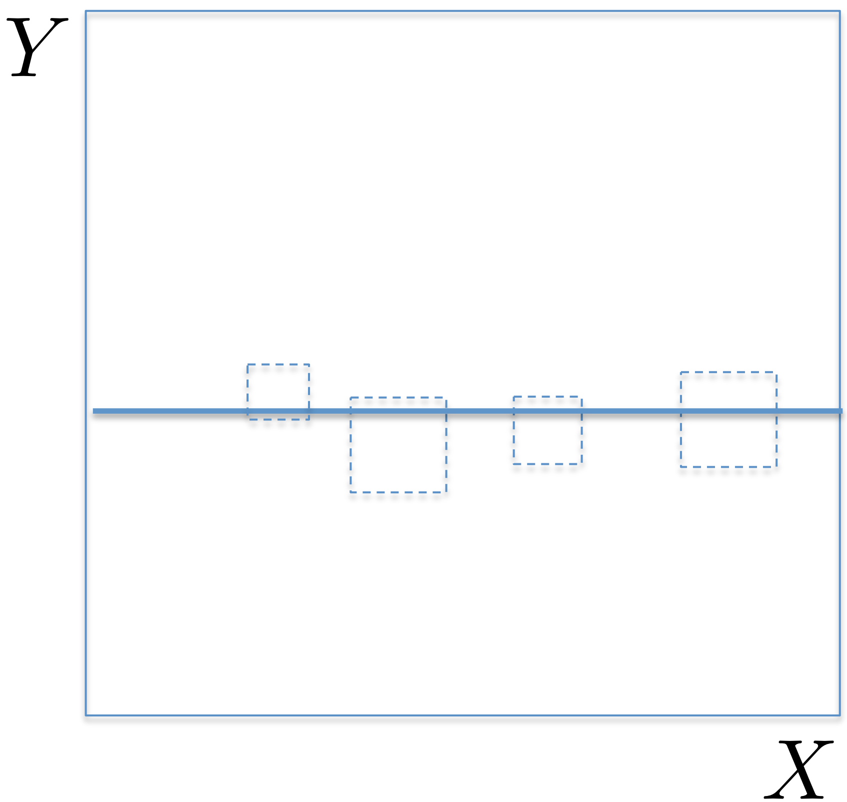
$$Y = \sum_{i=0}^{N} a_i X^i.$$

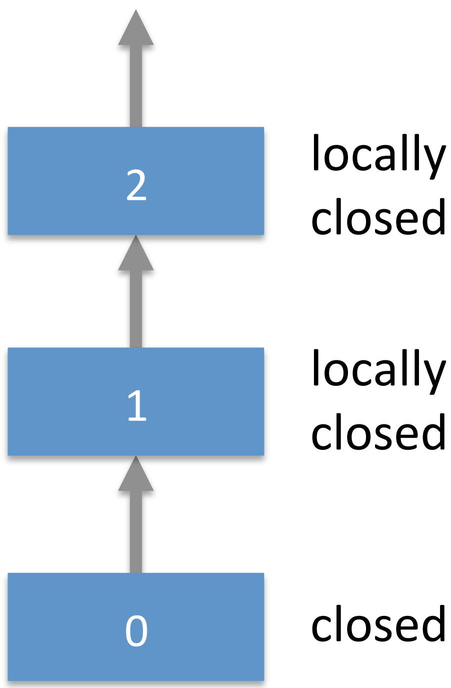
$$Y = a_0.$$
$$Y = a_0 + a_1 X.$$
$$Y = a_0 + a_1 X + a_2 X^2.$$
$$\vdots$$

Example: Quantitative Laws

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

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

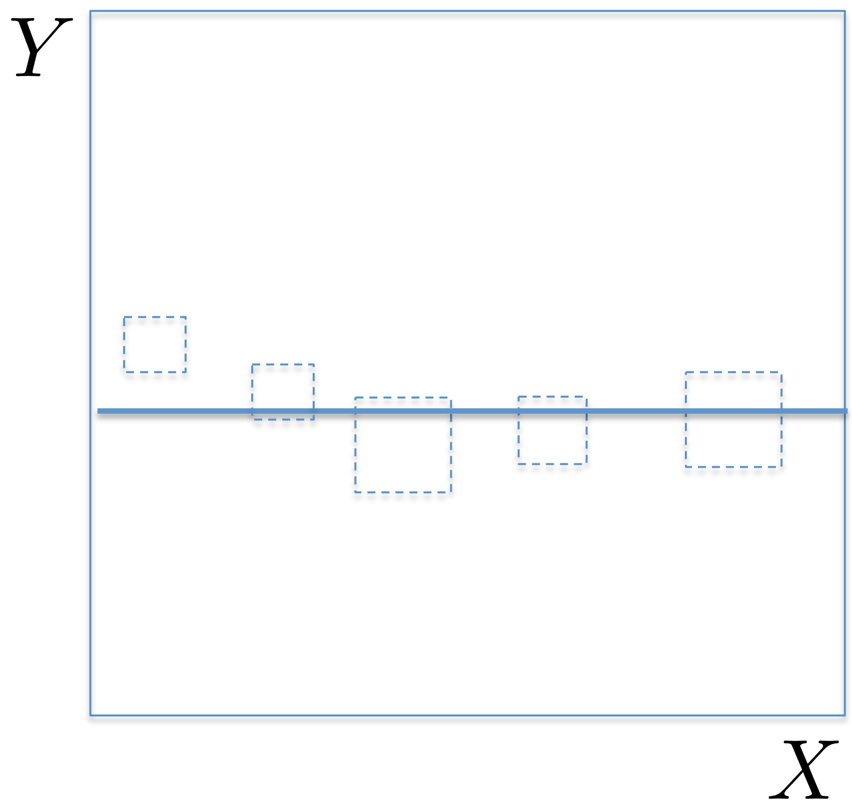


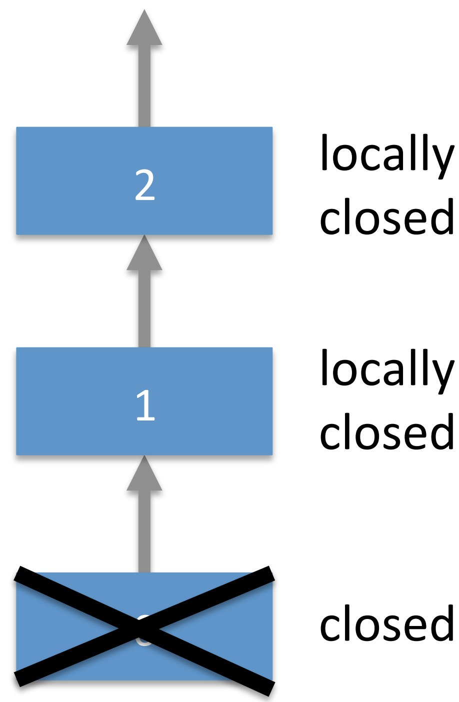


Example: Quantitative Laws

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

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

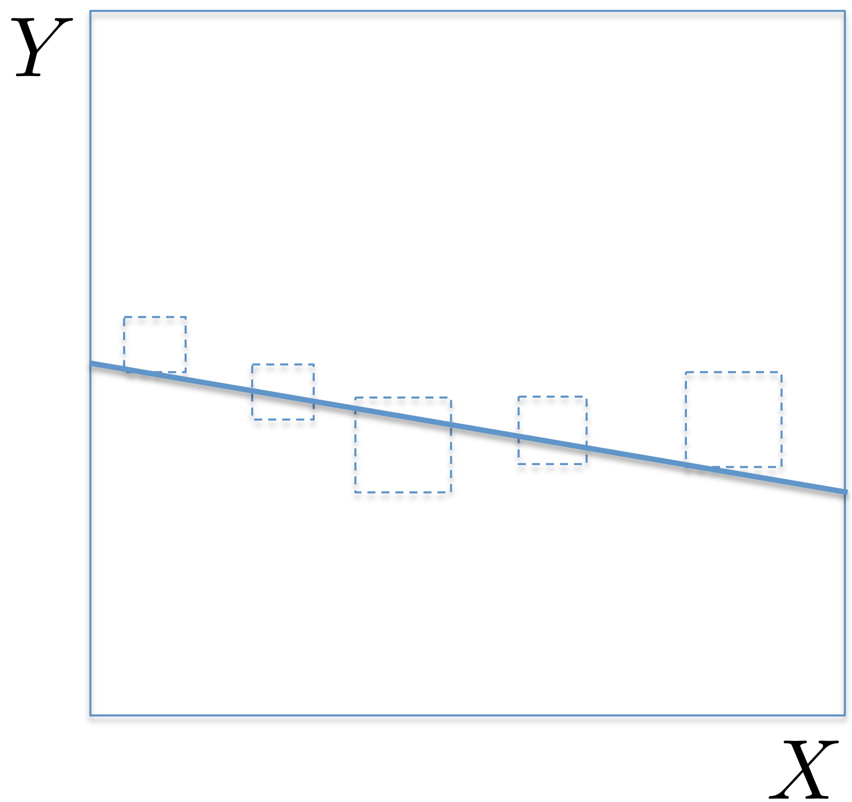


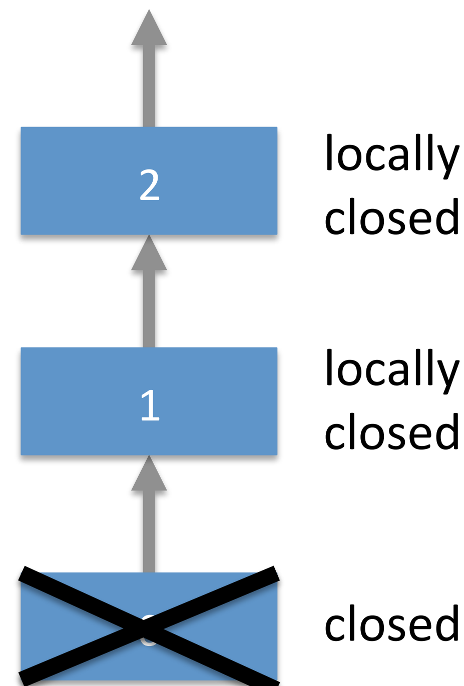


Example: Quantitative Laws

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

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

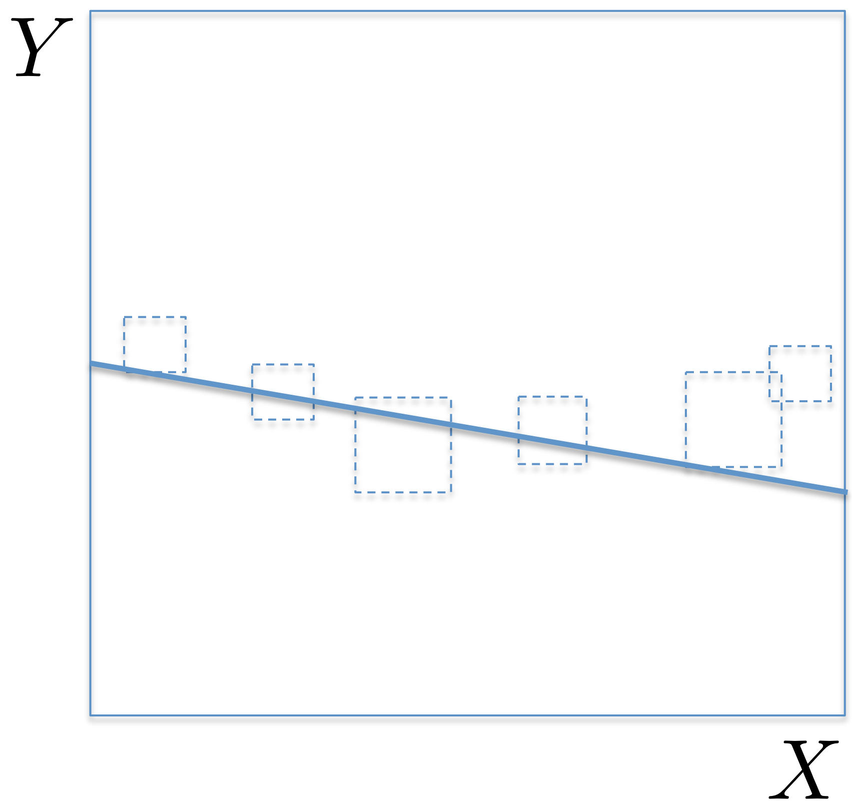


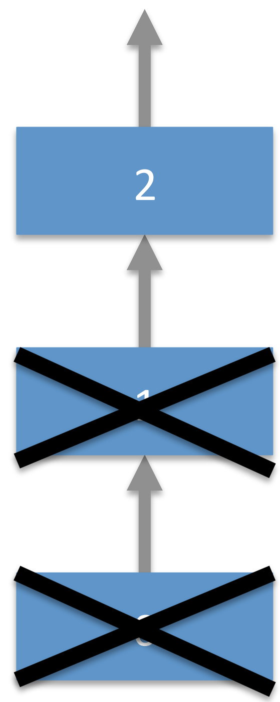


Example: Quantitative Laws

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

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

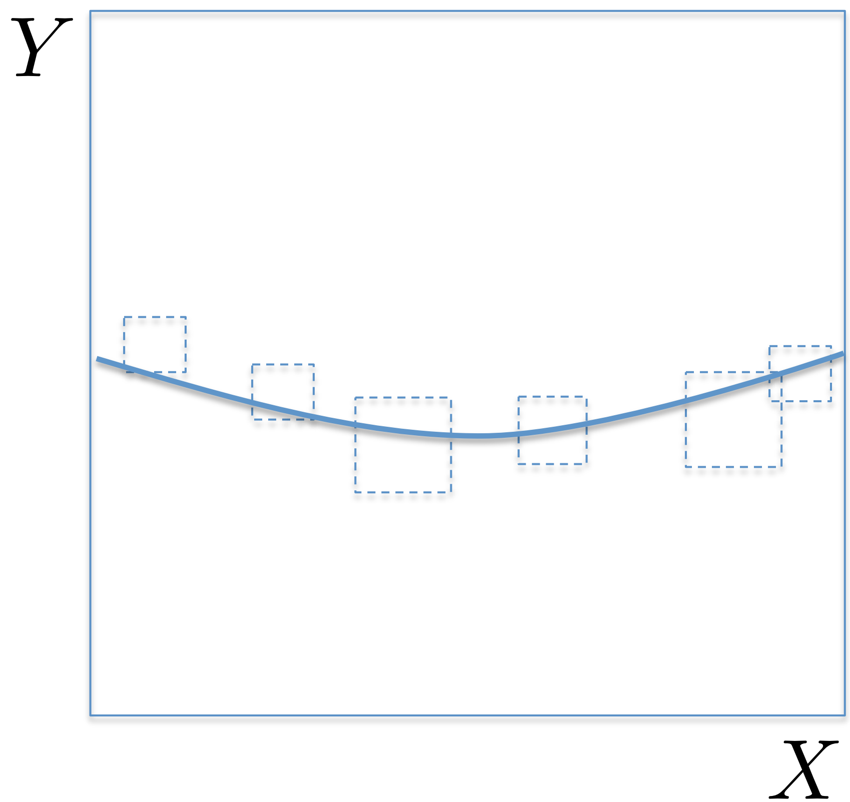


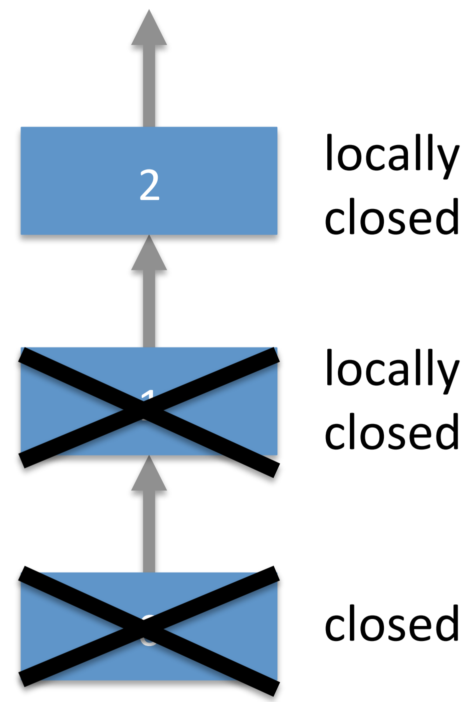


Example: Quantitative Laws

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

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





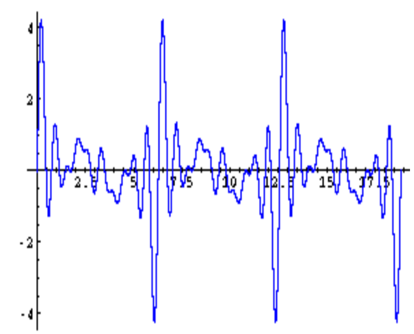
Example: Competing Paradigms

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

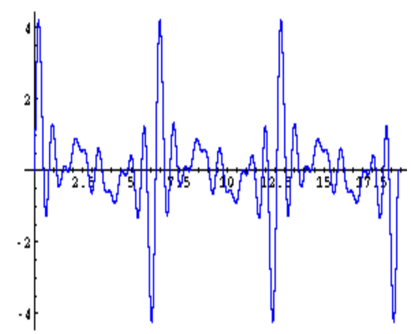
Polynomial paradigm

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

degree

Trigonometric polynomial paradigm

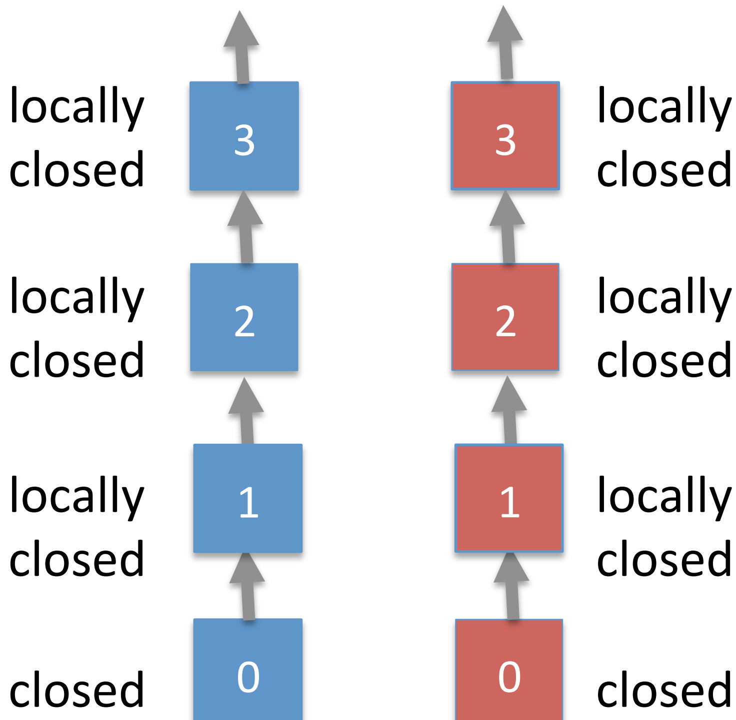
$$Y = \sum_{i=0}^{N} a_i \sin(iX) + b_i \cos(iX).$$

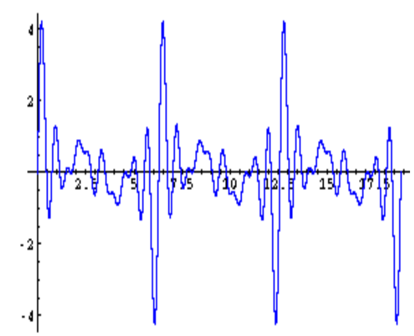


Example: Competing Paradigms

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

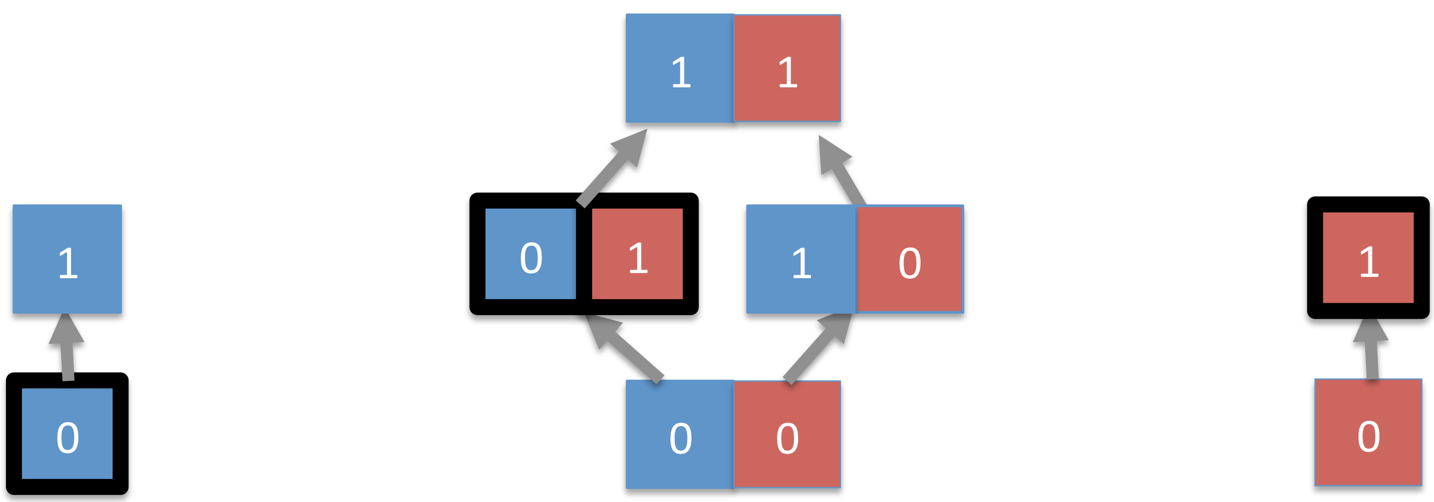
$\mathcal{I} =$ finitely many inexact measurements.





Tack-on Redux

- There is something wrong with tacking a complex answer onto a simple one.
- It is that the tack-on conjunction is more complex than some simpler conjunction (if the joint question is natural).



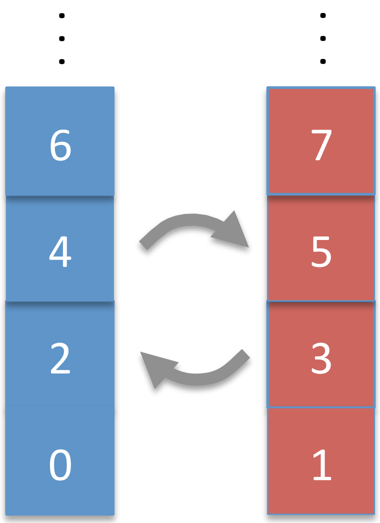
4. NATURAL QUESTIONS

The Question Question

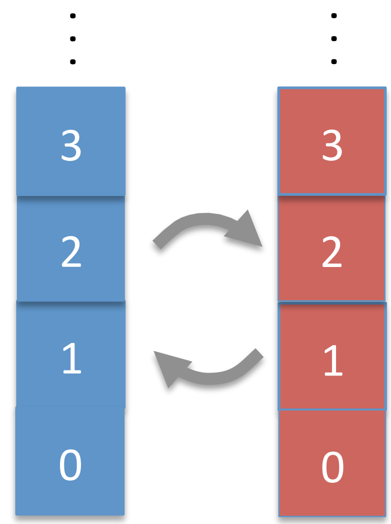
Questions guide inquiry.

What makes some more natural than others?

Intransitive Simplicity



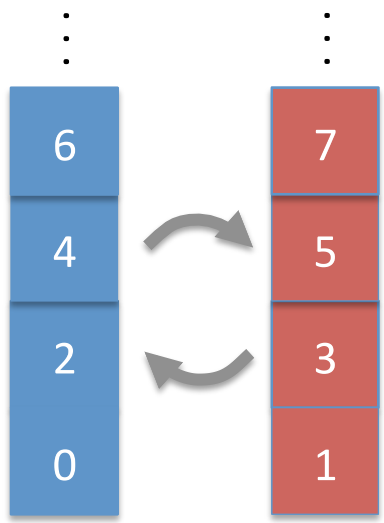
Even/ Odd



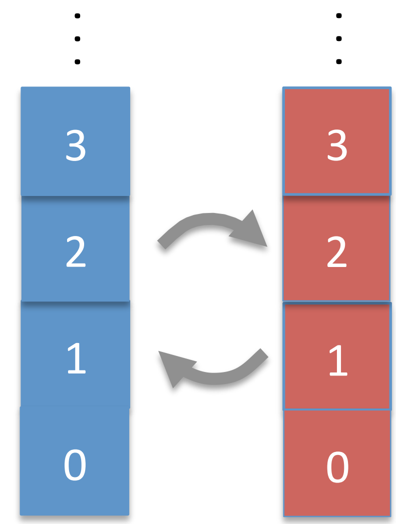
Poly/ Trig poly

Remedy

Proposition. The simplicity relation is transitive if every answer is locally closed.



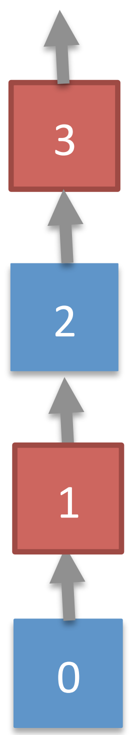
Even/ Odd



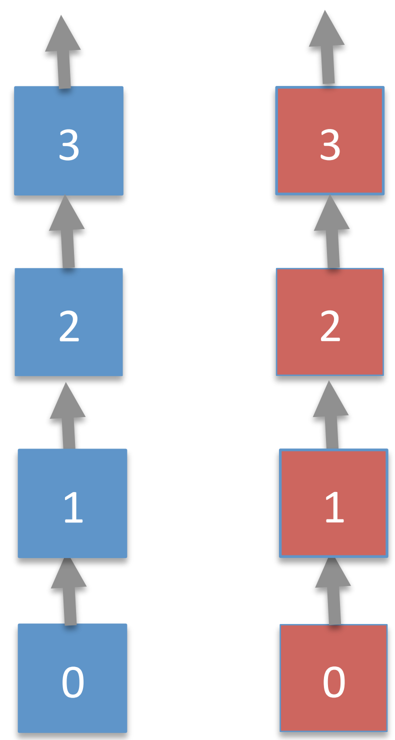
Poly/ Trig poly

Remedy

Proposition. The simplicity relation is transitive if every answer is locally closed.



Even/ Odd



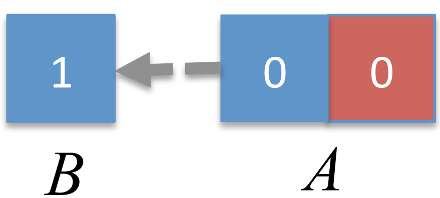
Poly/ Trig poly

Concealed Simplicity

Poly/Trig-poly

Homogeneity

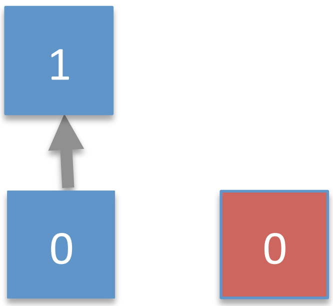
If any part of answer A is simpler than relevant response B, then all of A is simpler than B.



violation

Homogeneity

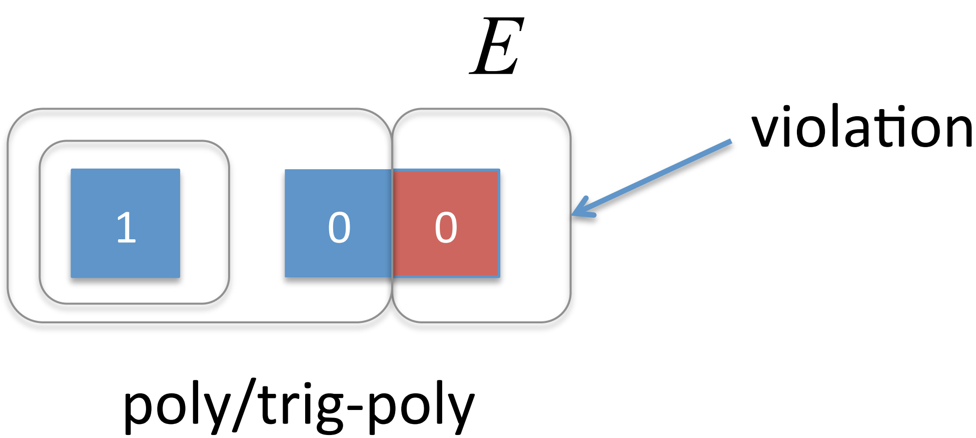
If any part of answer A is simpler than relevant response B, then all of A is simpler than B.



fixed

Homogeneity

Proposition. Homogeneity is equivalent to: the disjunction of the set of all answers compatible with information E is verifiable.



Natural Questions

A question is natural in a problem iff

1. Each answer is locally closed;
2. Each answer is homogeneous.

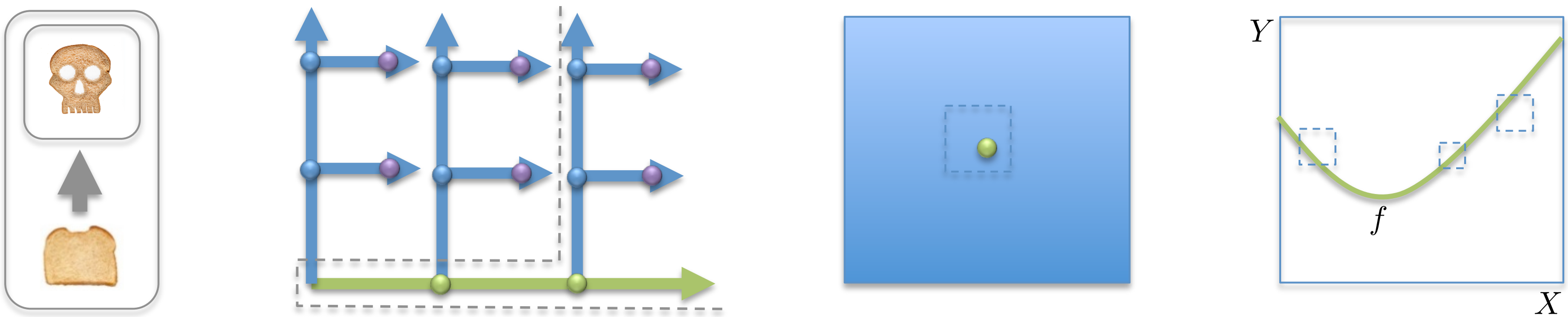
Natural Questions

A question is natural in a problem iff

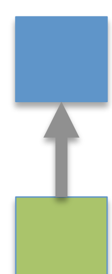
1. Each answer is locally closed;
2. Each answer is homogeneous.

In algebraic geometry, natural questions are thought of geometrically as stratifications of the underlying topology.

Epistemic Equivalence



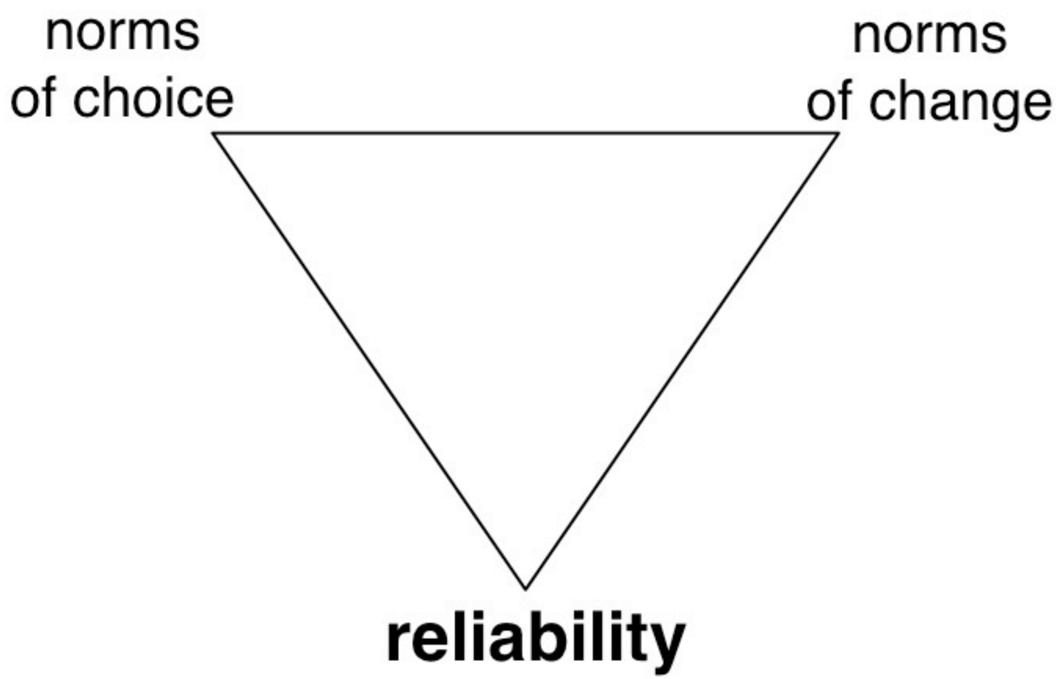
different problems



natural problems

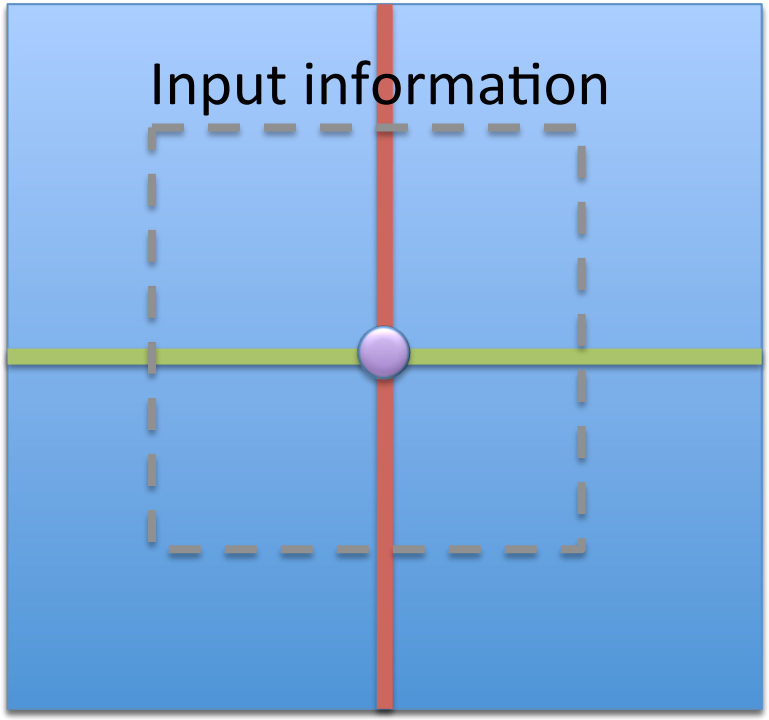
same simplicity structure

same epistemological structure

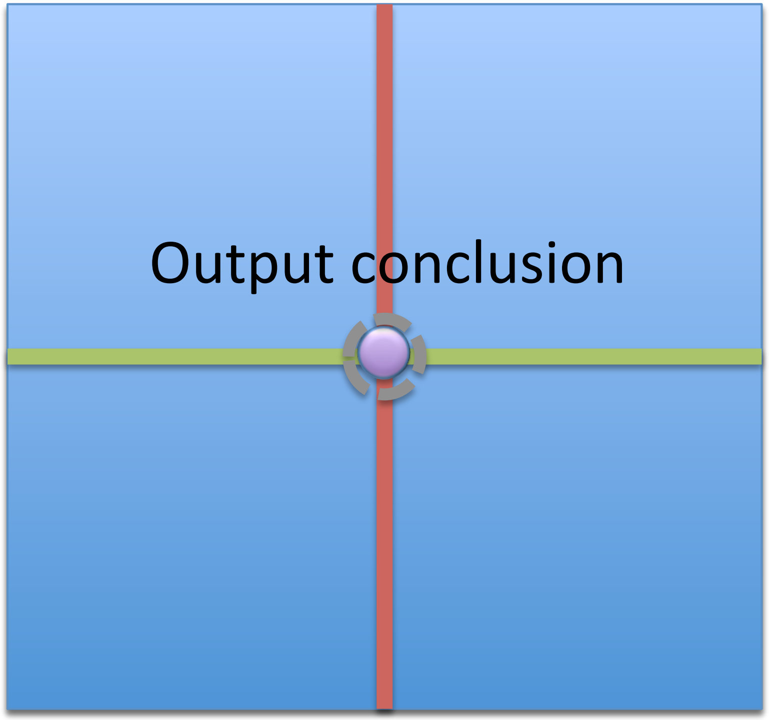


5. METHODS AND CONVERGENCE

Inductive Inference

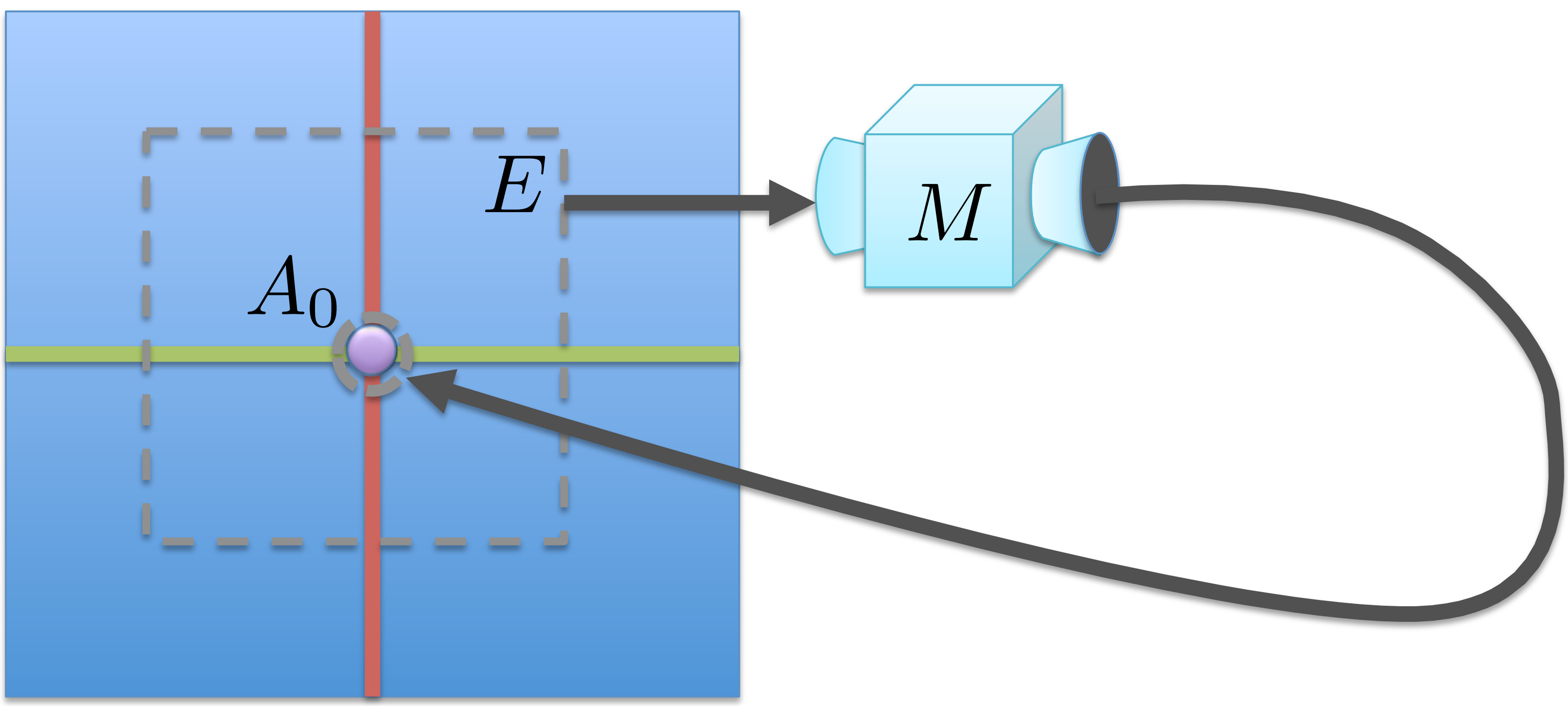


Inductive Inference



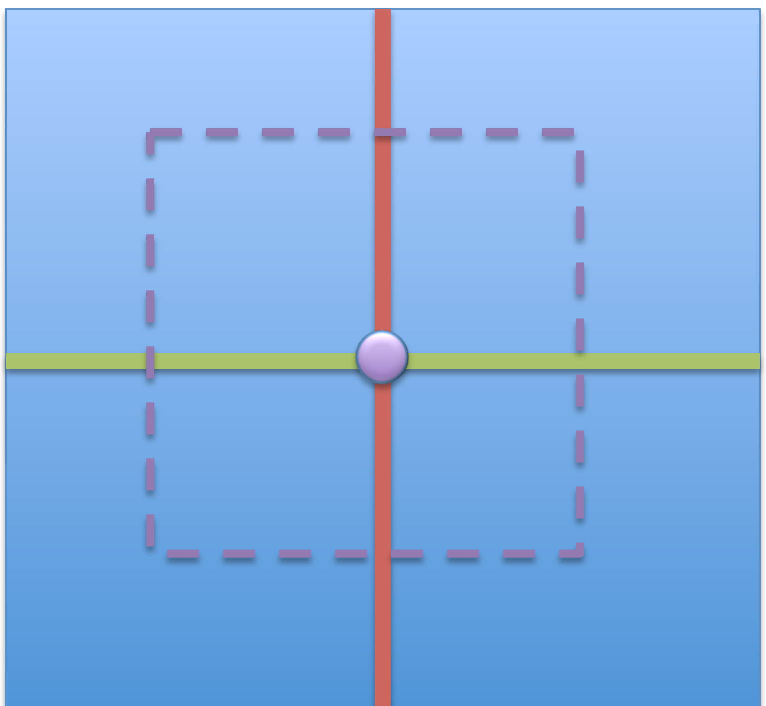
Inductive Methods

Information in, relevant response out.



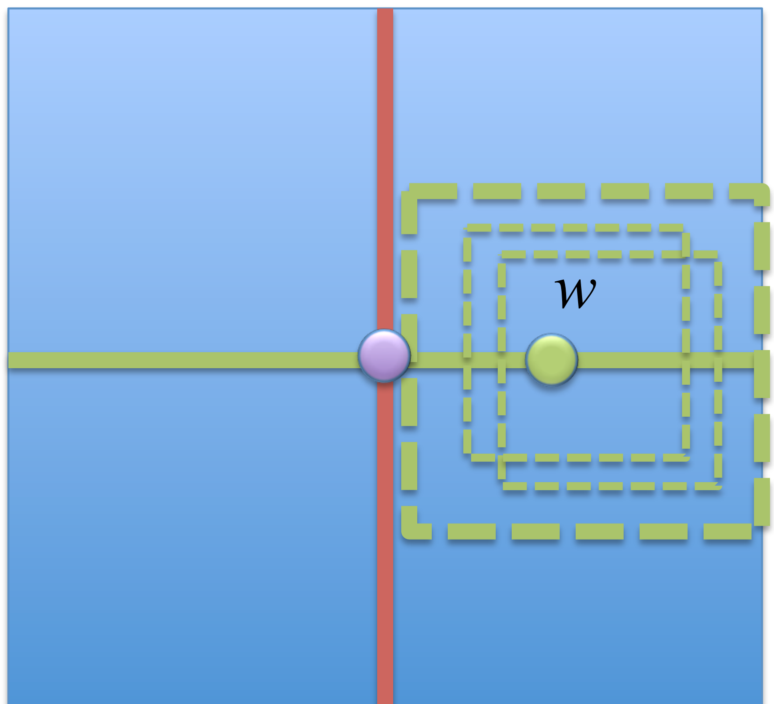
Inductive Methods

Cleaner diagram.



Solution in the Limit

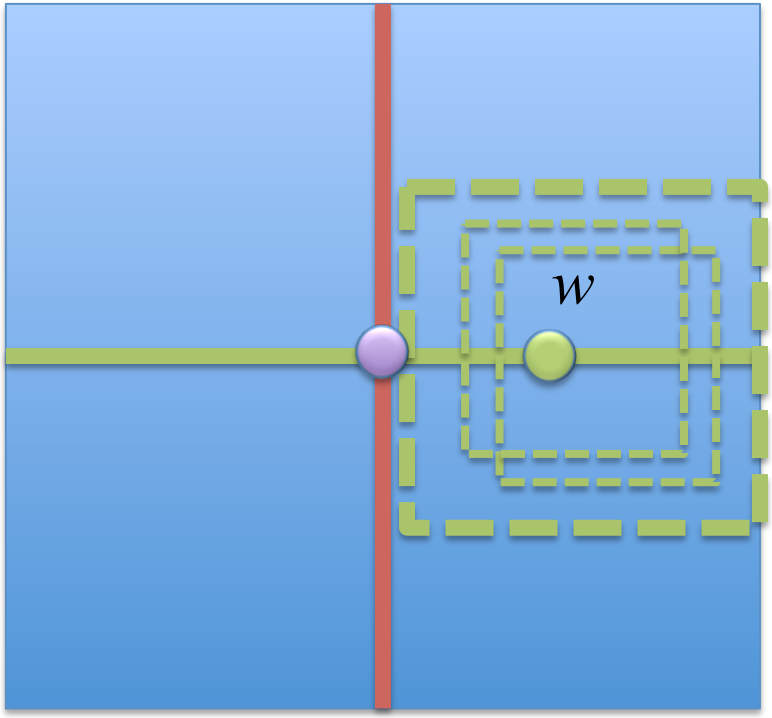
M solves a problem in the limit iff each world w presents information such that M produces the true answer in w on any further information true in w.



Locking information for 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 true in w.



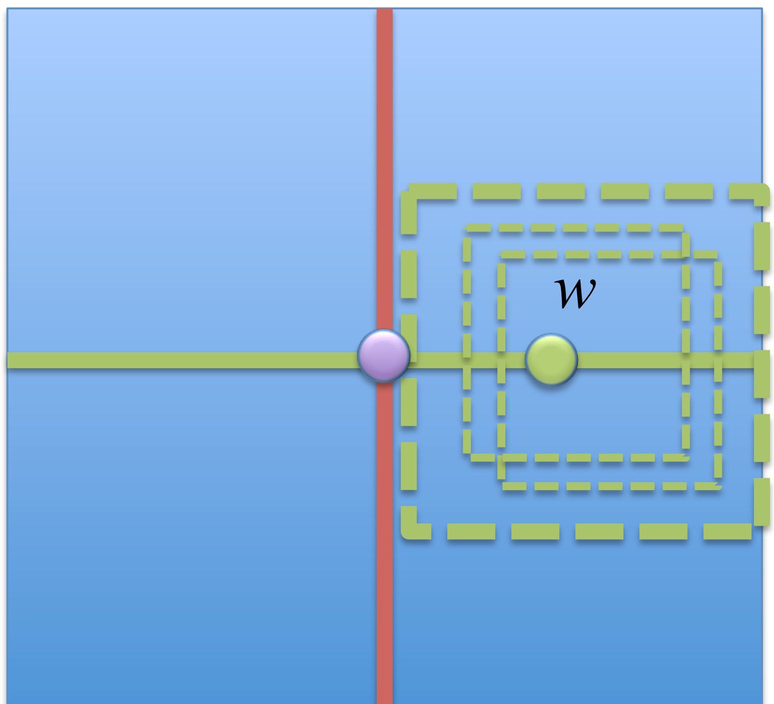
Locking information for w.

Solution in the Limit

M solves $(W, \mathcal{I}, \mathcal{Q})$ in the limit

iff M stabilizes to the true answer in each world

iff $(\forall w \in W)(\exists E \in \mathcal{I}_w)(\forall F \in \mathcal{I}_w)(F \subseteq E \Rightarrow M(F) = \mathcal{Q}_w)$.



Locking information for w.

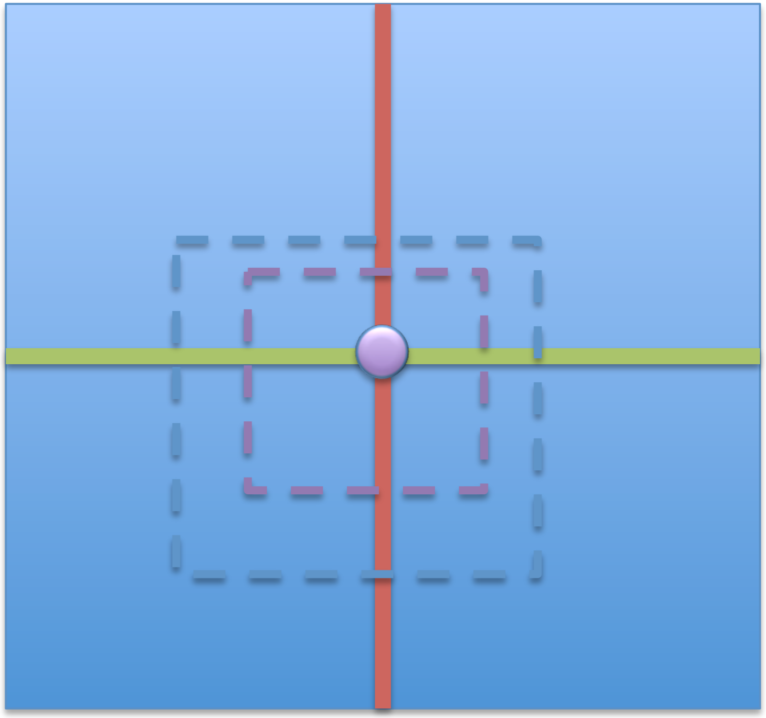
Solvability and Topology

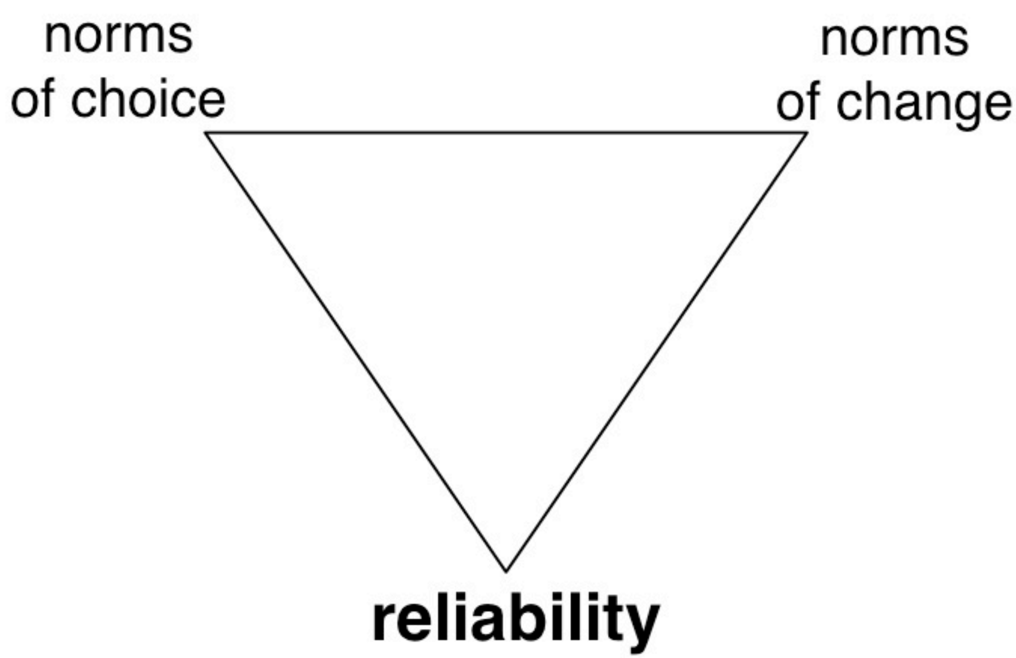
Proposition (Yamamoto and DeBrecht 2010, Kelly 2004, Baltag, Gierasimczuk and Smets 2015):

A problem is solvable in the the limit iff each answer is a countable disjunction of locally closed propositions.

Solution in the Limit

- Solution in the limit implies no constraint on what to say in response to a given information state.
- Convergence can always begin later.





7. STRAIGHTEST CONVERGENCE

Two Departures from Straightness

Course-reversals

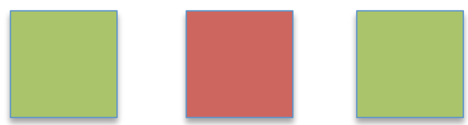
Cycles

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.



Cycle Free Solutions

- Solution M is cycle free iff:

There exists no nested sequence of information states

$$e = (E_i)_{i=1}^{n}$$

such that

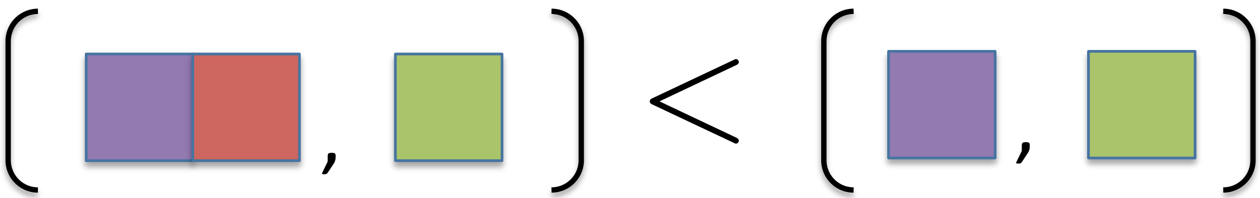
$$M(e) = (M(E_i))_{i=1}^{n}$$

is a cycle sequence.

Reversal Sequence Comparison

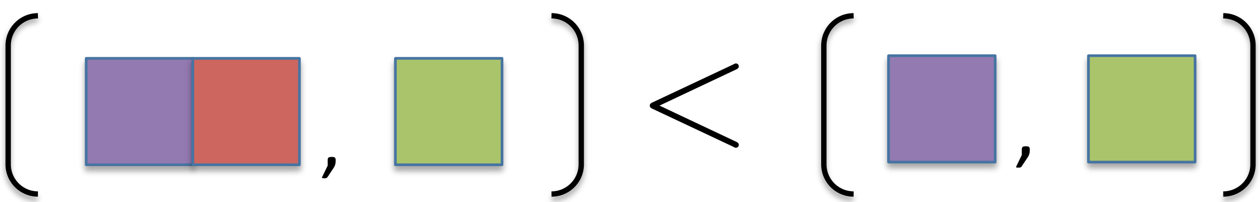
- Reversal sequence s reverses as much as reversal sequence s' (of the same length) iff

  each entry in s entails the corresponding entry in s'.



Forcible Reversal Sequences

- Reversal sequence s is forcible iff every solution to $\mathfrak{P}$ performs a reversal sequence $s' > s$.



Forcible Sequences

In natural problems, $(A_i)_{i=1}^n$ is forcible iff

$$A_1 \prec A_2 \prec \cdots \prec A_{n-1} \prec A_n.$$

Forcible Sequences

In general, $(A_i)_{i=1}^{n}$ is forcible iff

$$A_1 \cap \mathsf{Frntr}(A_2 \cap \mathsf{Frntr}(\cdots \cap \mathsf{Frntr}(A_{n-1} \cap \mathsf{Frntr}(A_n)))) \neq \varnothing$$

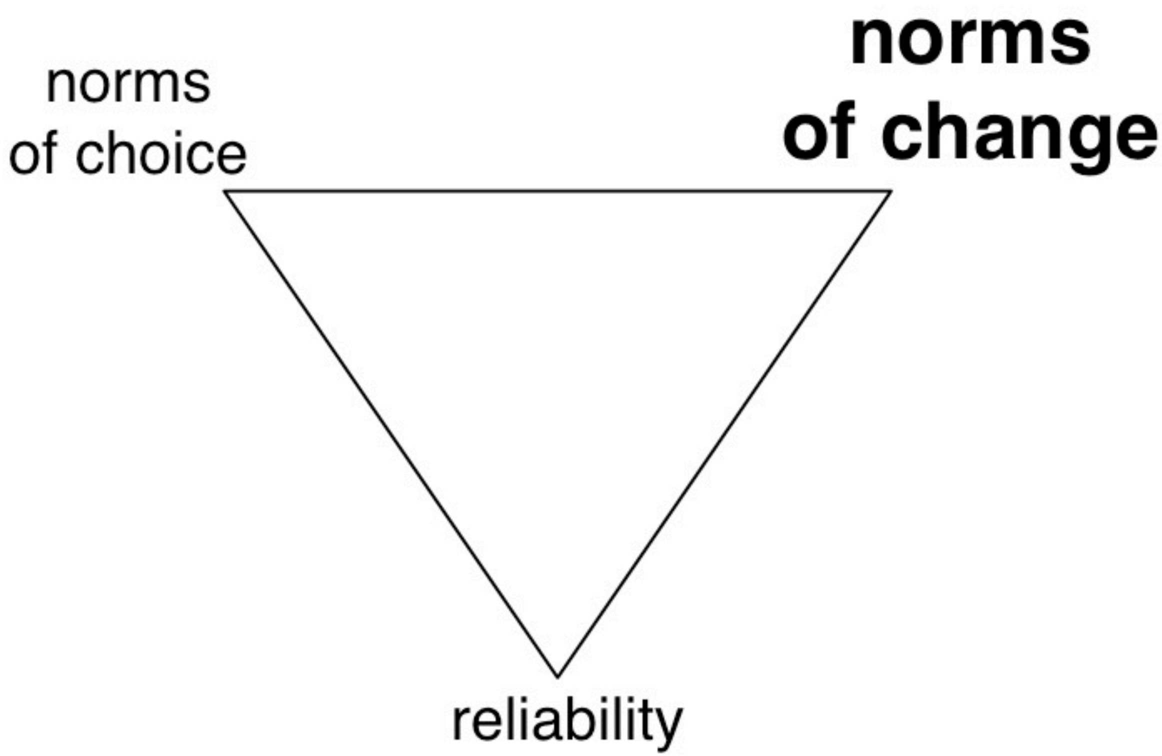
Method Comparison

- Solution M' reverses as much as solution M iff:

  For each reversal sequence generated by solution M, method M' generates a reversal sequence at least as bad in some world.

Optimal Truth Conduciveness

- Solution M is reversal-optimal iff every reversal it performs is forcible.



6. MINIMAL CHANGE

Conditionalization

A method *M* satisfies conditionalization iff for all information states *E,F:*

$$M(E) \cap \mathcal{Q}(E \cap F) \subseteq M(E \cap F).$$

In slogan form:

“no induction without refutation.”

Rational Monotony

A method *M* is rationally monotone iff

$$M(E \cap F) \subseteq M(E) \cap \mathcal{Q}(E \cap F),$$

for all information states *E,F* such that

$$M(E) \cap \mathcal{Q}(E \cap F) \neq \varnothing.$$

In slogan form:

"no retraction without refutation."

Reversal Monotony

A method *M* is reversal monotone iff

$$M(E \cap F) \cap M(E) \neq \varnothing,$$

for all information states *E,F* such that

$$M(E) \cap \mathcal{Q}(E \cap F) \neq \varnothing.$$

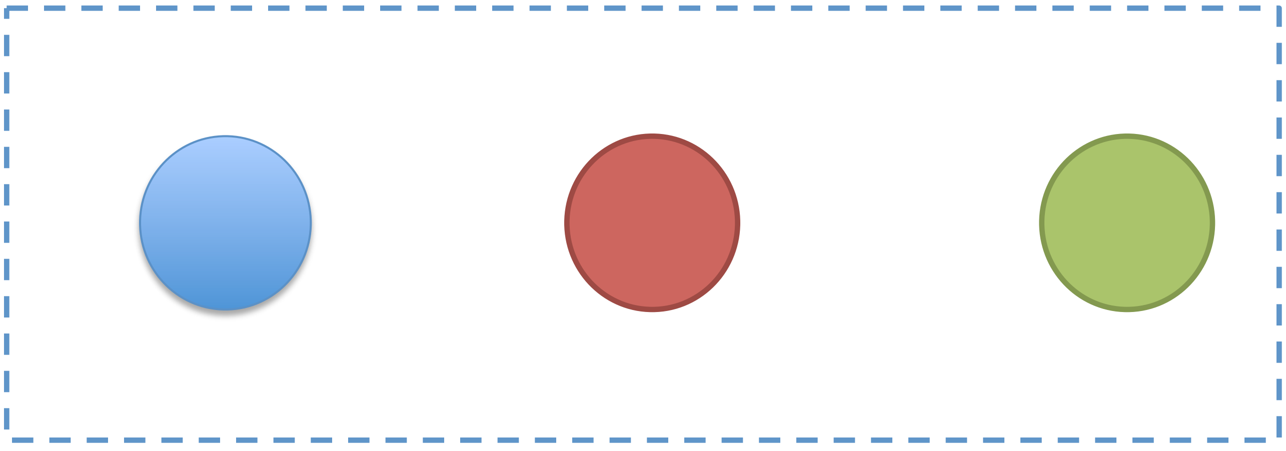
In slogan form:

“no reversal without refutation.”

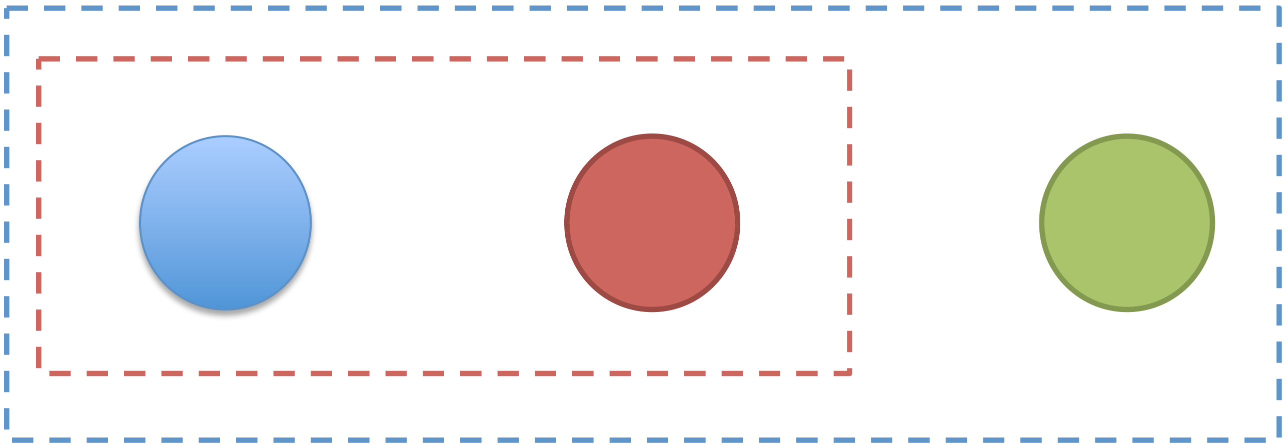
Theorem

If M is a consistent solution to $\mathfrak{P}$, then M cycle-free iff M is reversal monotone.

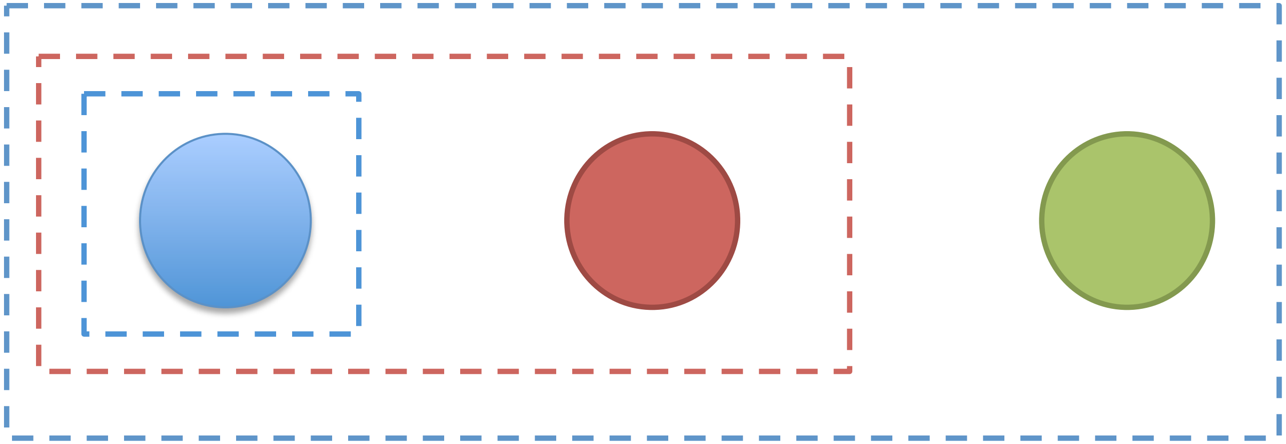
The Proof Idea



The Proof Idea

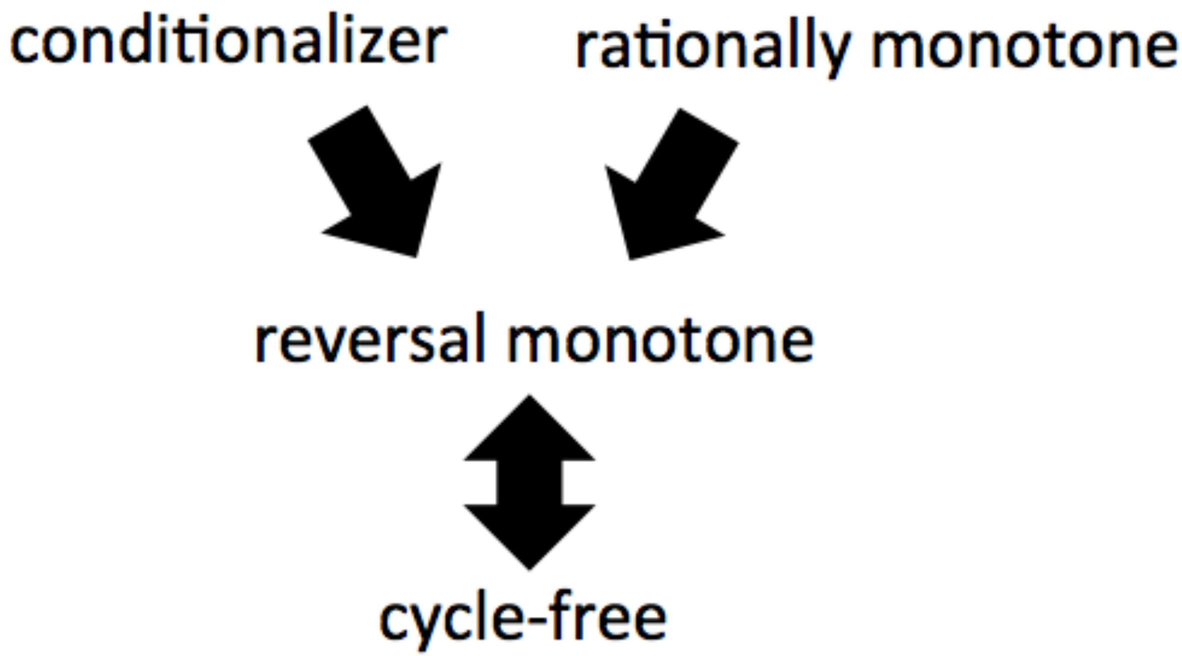


The Proof Idea



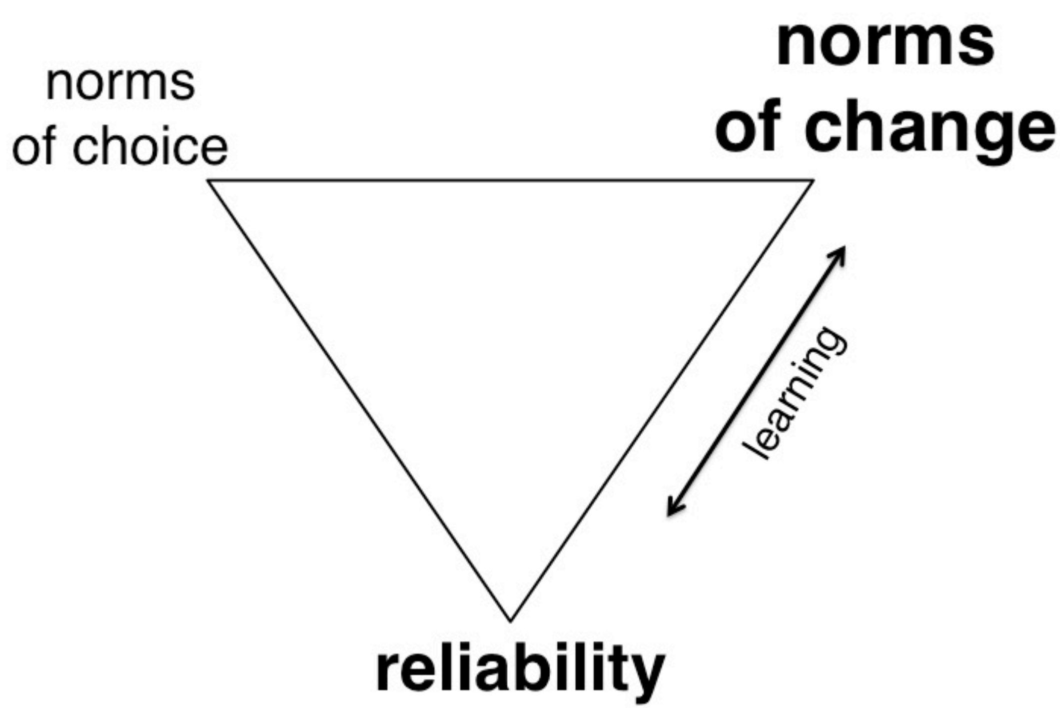
Theorem

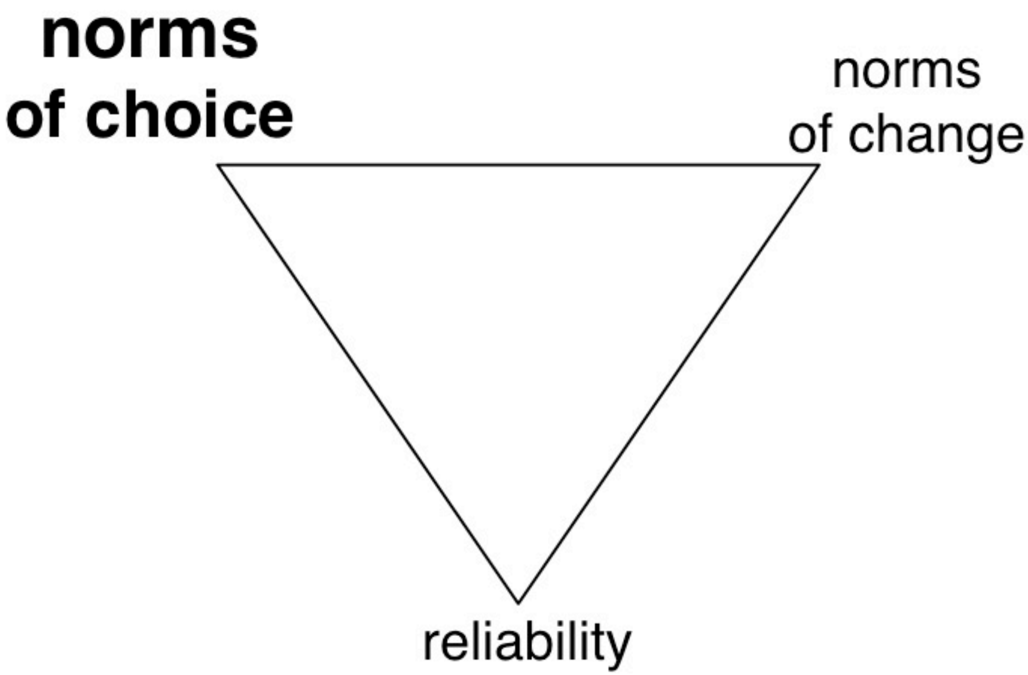
If M is a consistent solution to $\mathfrak{P}$, then M is cycle-free iff M is reversal monotone.



Theorem

If M is a consistent solution to $\mathfrak{P}$, then M is cycle-free iff M is reversal monotone.





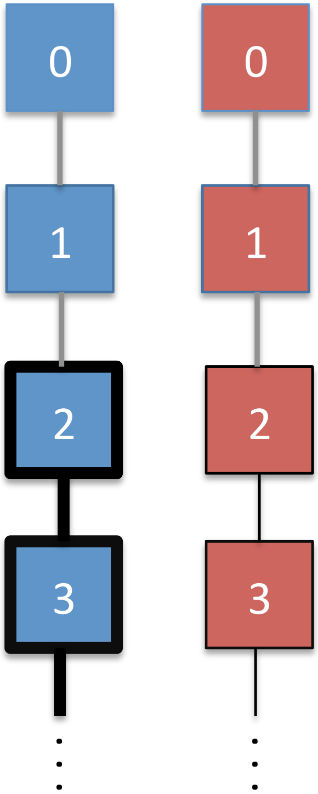
7. OCKHAM'S RAZOR

Simplest Relevant Responses Given E

B is a simplest relevant response given E iff any relevant response simpler than B is incompatible with E.

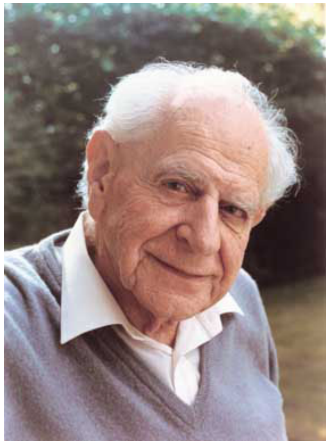
Ockham's Razor

- Output a simplest relevant response given E.
  - Allows for suspension of judgment.
  - Rules out "even" in co-finite even/odd problem.
  - Makes sense for infinite descending chains.



Popper's Razor

- Output a relevant response that is refutable (closed) in the problem restricted to E.



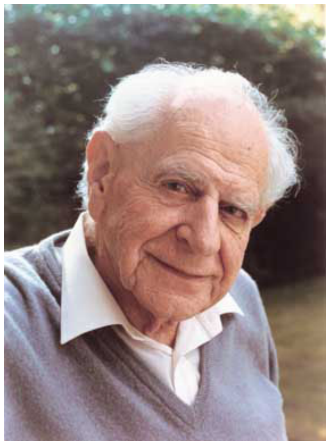
Error Razor

- "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.

Theorem

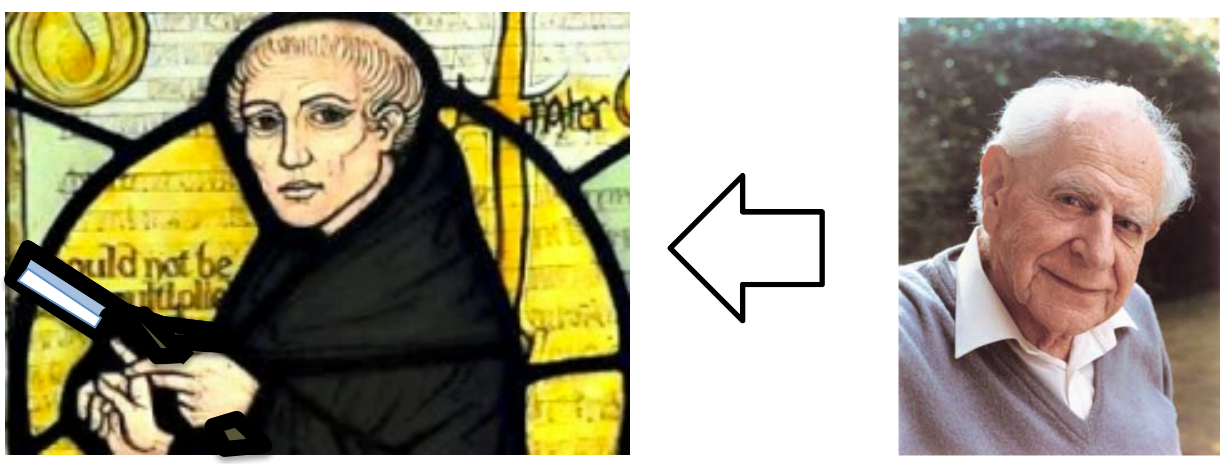
For natural problems, Ockham's Razor = Popper's Razor.

=



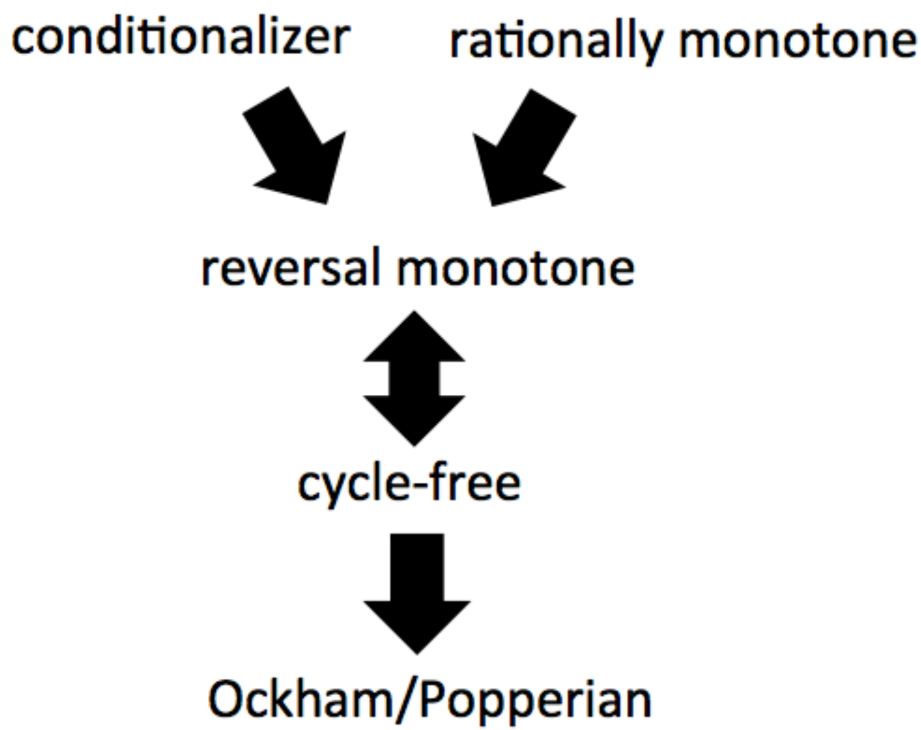
Theorem

In general, Popper's razor is stronger.



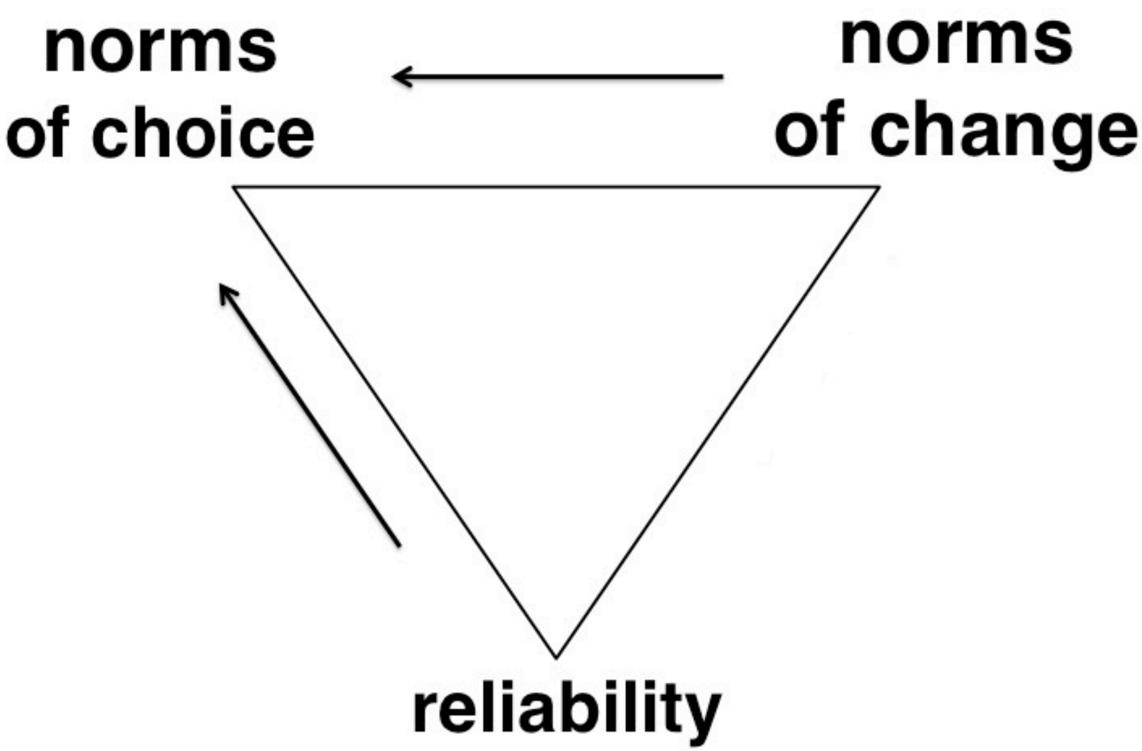
Theorem

If *M* is a cycle-free solution, then *M* is Popperian.

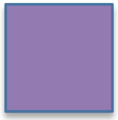


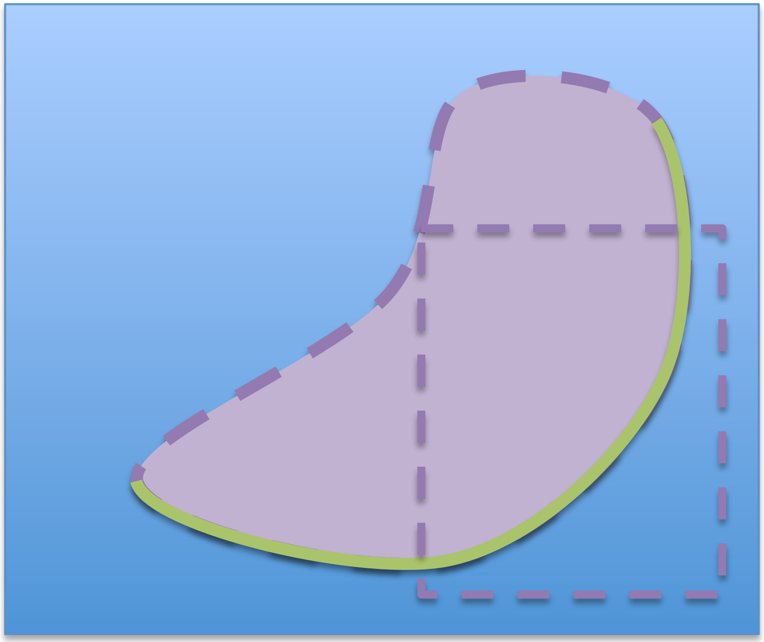
Theorem

If M is a cycle-free solution, then M is Popperian.

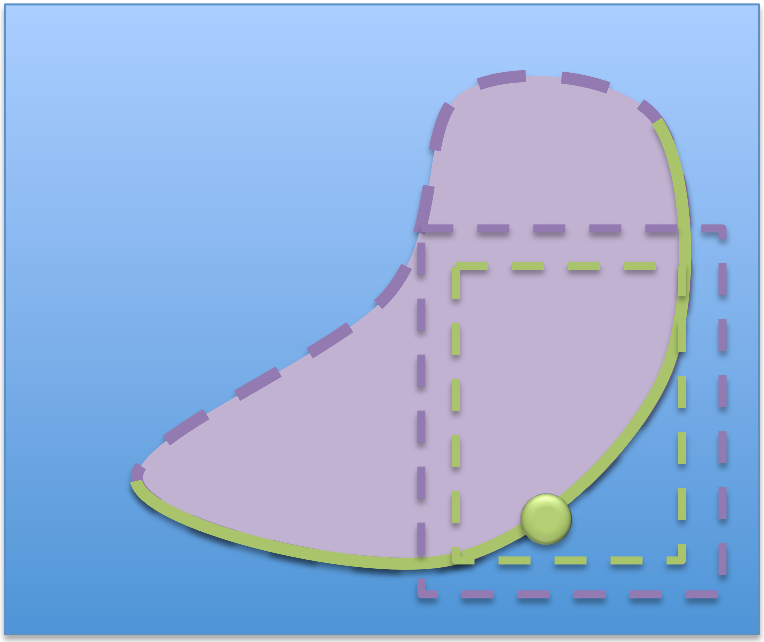


The Proof Idea

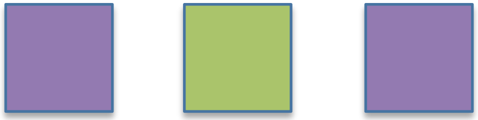


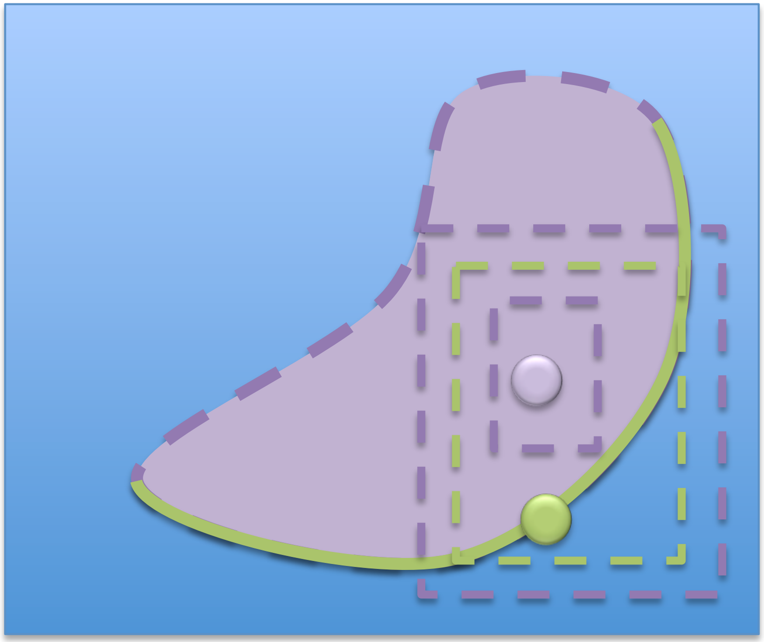


The Proof Idea



The Proof Idea





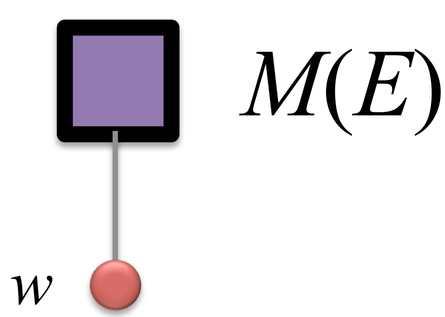
Theorem

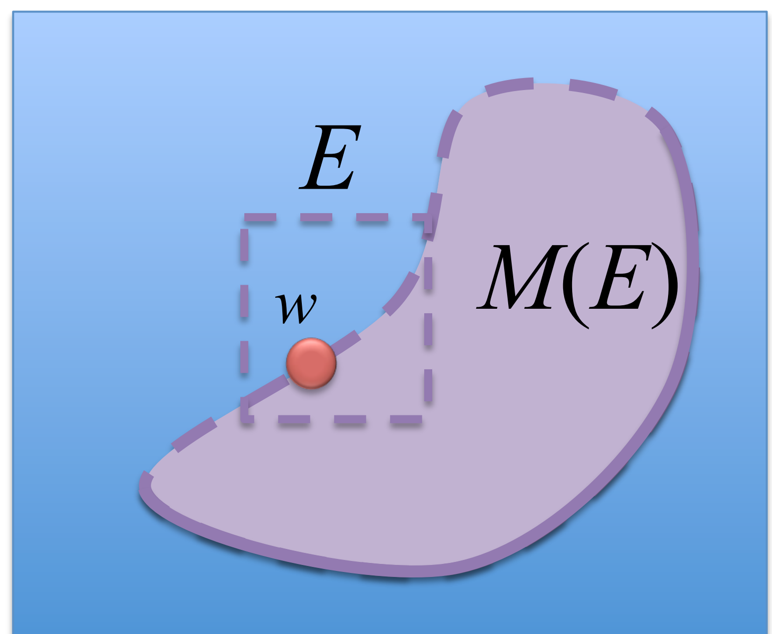
Every natural problem has a cycle-free solution (and every such solution is Ockham).

Equivalence

Lemma 1: Error razor implies Popper's razor.

Suppose that M violates Popper's razor on E.
So $M(E)$ isn't closed given E.
Let w be a missing boundary point.
So $\{w\}$ is simpler than $M(E)$.

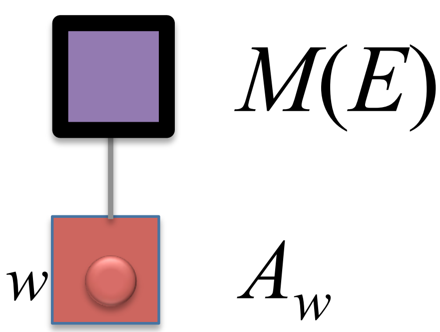


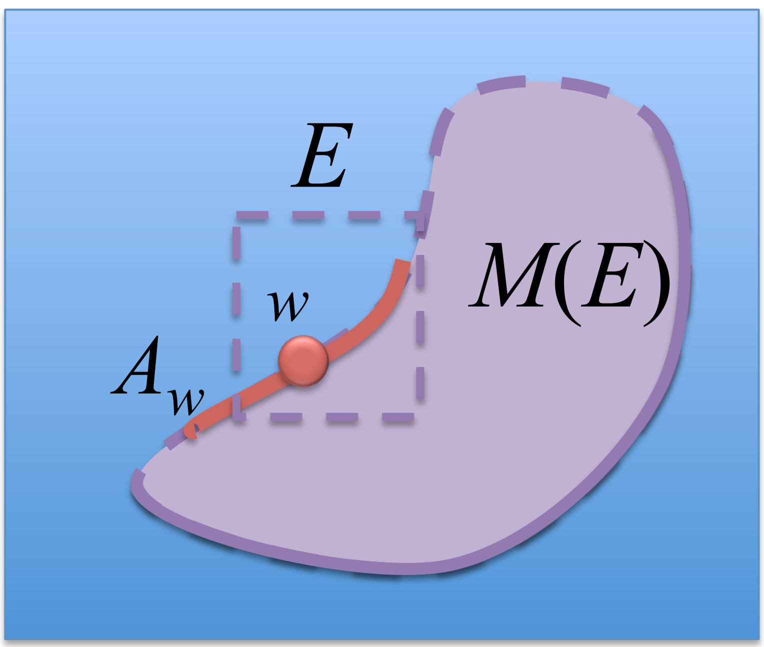


Equivalence

Lemma 1: Error razor implies Popper's razor.

Suppose that M violates Popper's razor on E.
So $M(E)$ isn't closed given E.
Let w be a missing boundary point.
So $\{w\}$ is simpler than $M(E)$.
Apply homogeneity.
So A_w is simpler than $M(E)$.

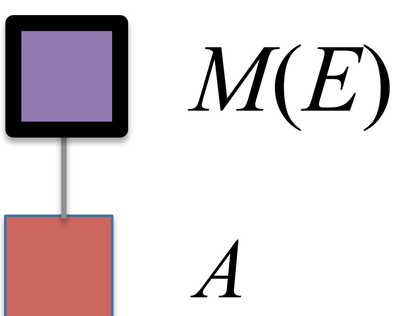


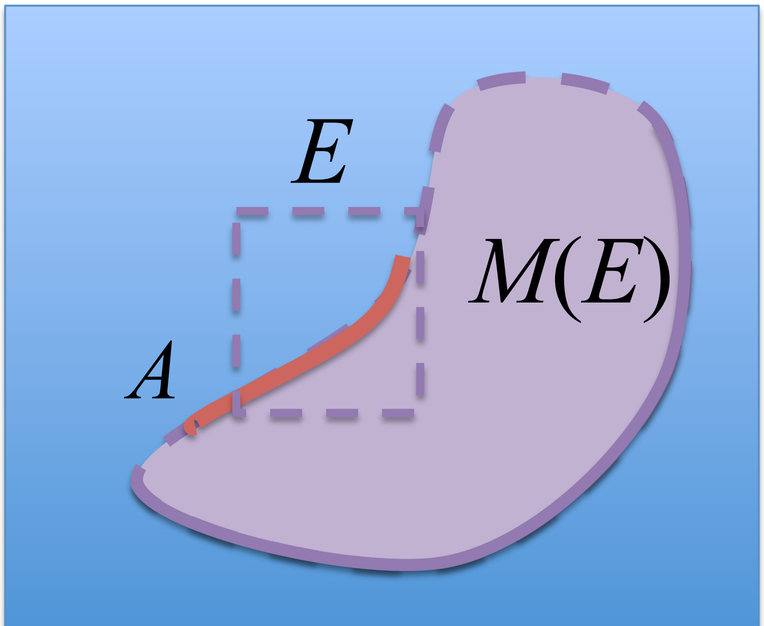


Equivalence

Lemma 2: Popper's razor implies Ockham's razor.

Suppose that M violates Ockham's razor on E.
So some A compatible with E is simpler than $M(E)$.

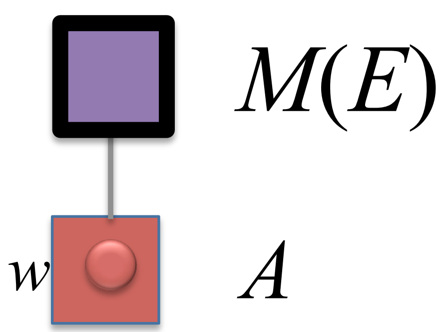


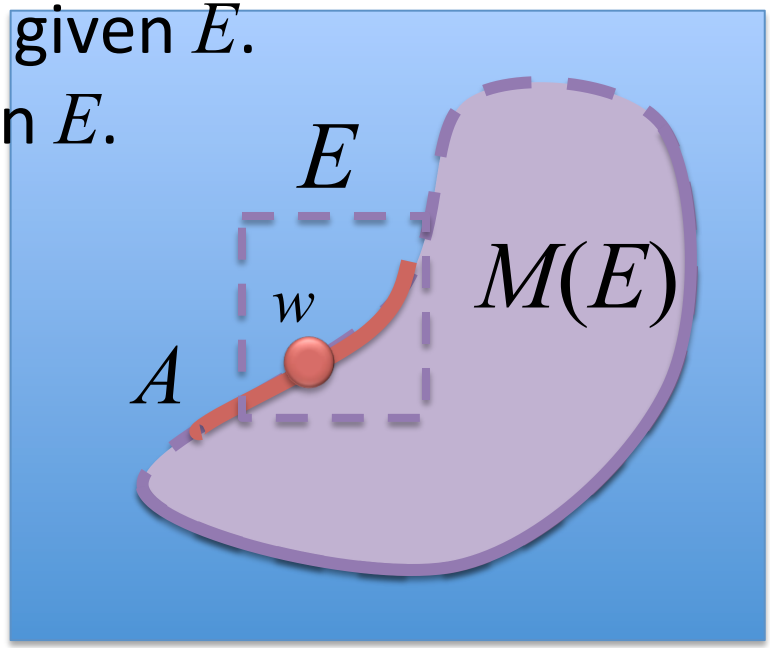


Equivalence

Lemma 2: Popper's razor implies Ockham's razor.

Suppose that M violates Ockham's razor on E.
So some A compatible with E is simpler than $M(E)$.
Choose w to witness compatibility.
w witnesses that is not closed given E.
So M violates Popper's razor on E.

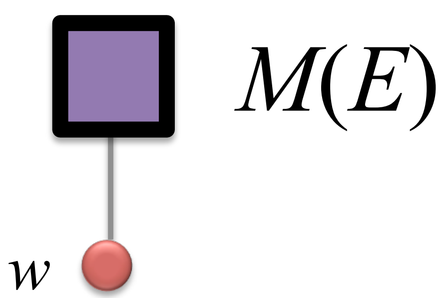


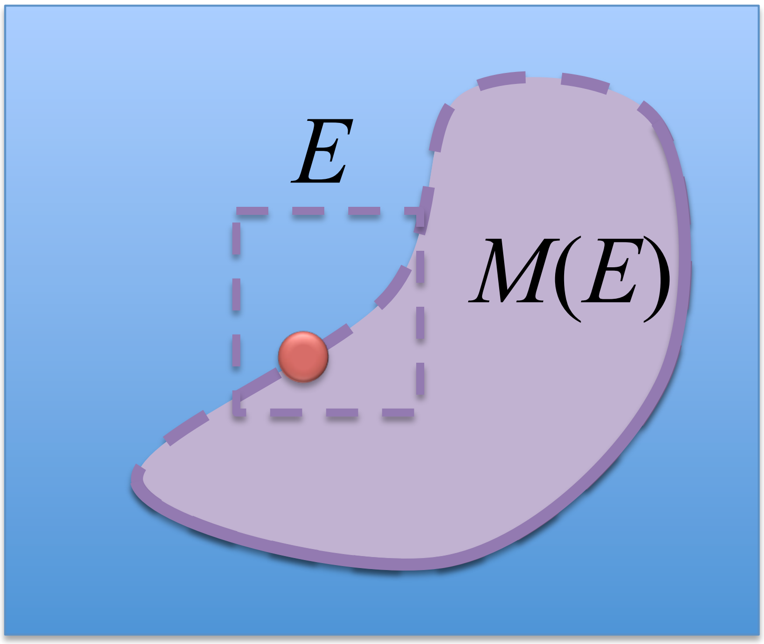


Equivalence

Lemma 3: Ockham's razor implies Error razor.

Suppose that M violates the error razor in w.
So w presents E such that A_w is simpler than $M(E)$.

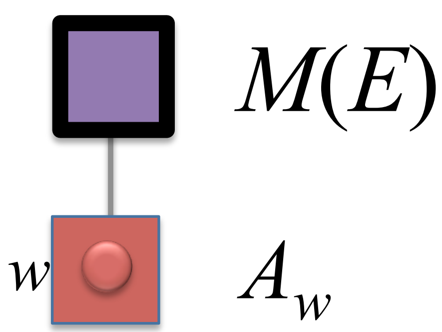


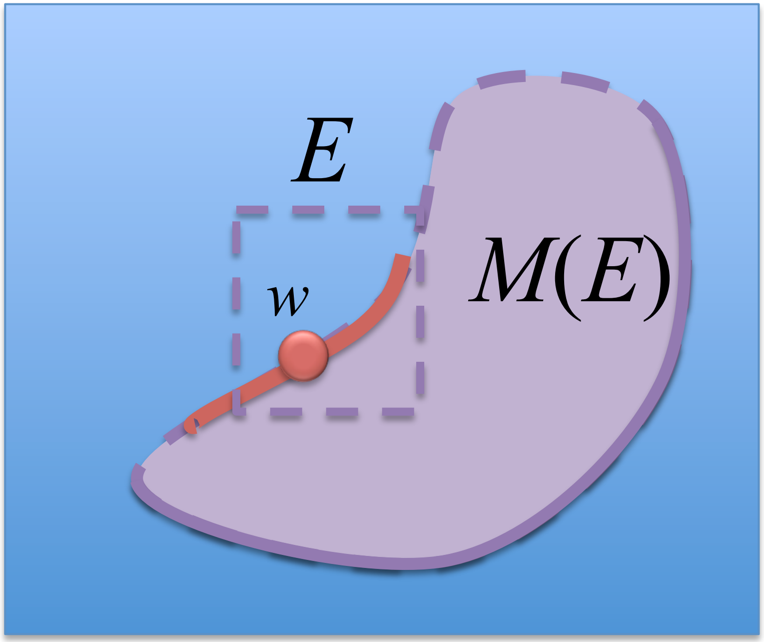


Equivalence

Lemma 3: Ockham's razor implies Error razor.

Suppose that M violates the error razor in w.
So w presents E such that A_w is simpler than $M(E)$.
Apply homogeneity.
So A_w is simpler than $M(E)$.
That is an Ockham violation.

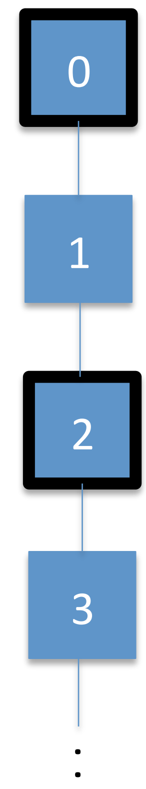




Patience

- 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



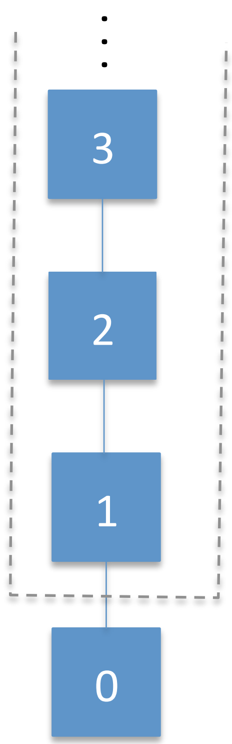
Patience

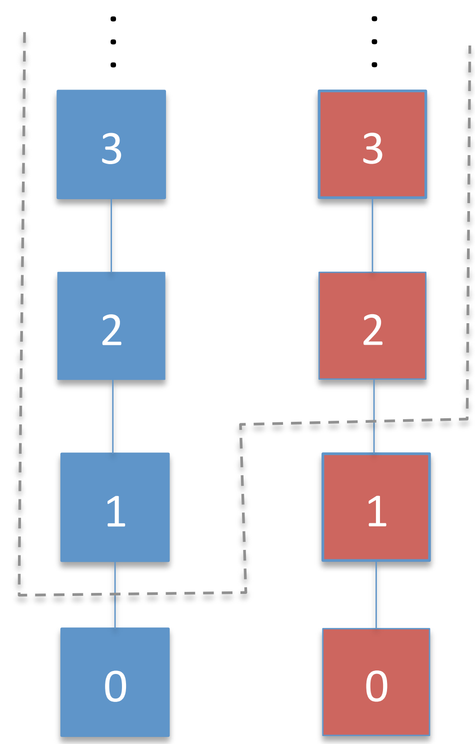
- Never rule out a simplest relevant response given E.
  - Says that Ockham's razor is the only reason for inductive leaps beyond experience.
  - Logically independent of Ockham's razor.

Ockham but
not patient

Normal Science vs. Revolutionary Science

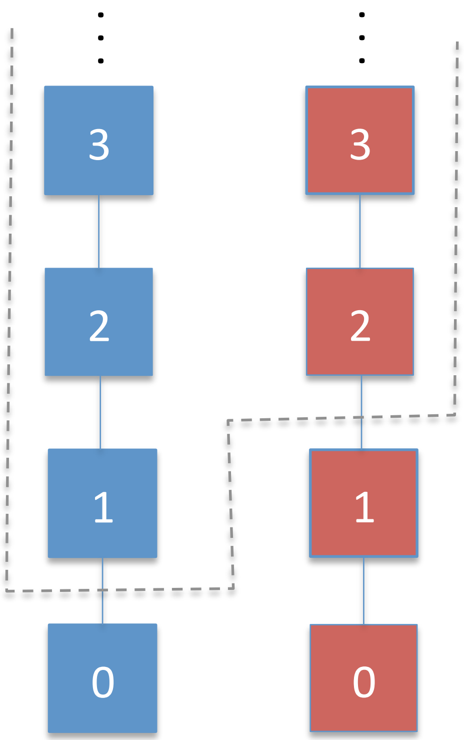
- A problem is normal iff the disjunction of every upward-closed set of answers in the simplicity order is verifiable.
- Else it is revolutionary.





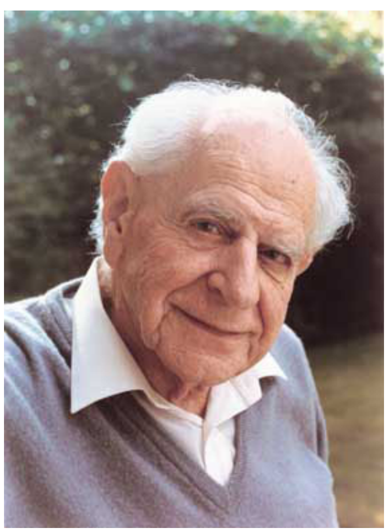
Patient Learnability

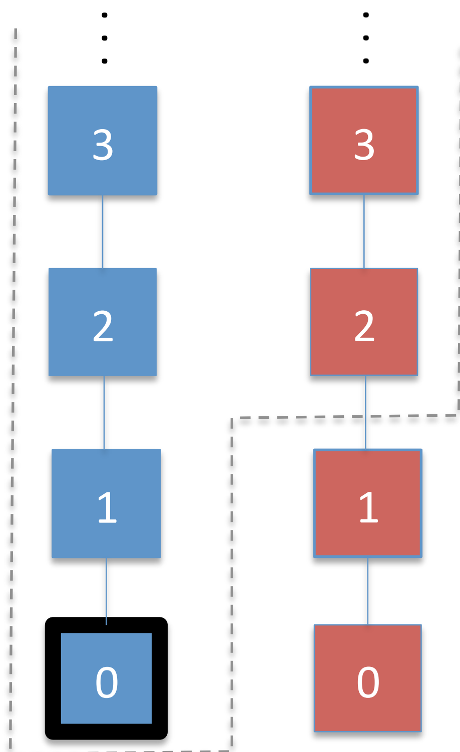
- **Proposition.** A natural problem has reversal-optimal solution iff it is normal.
- **Idea:** Just suspend judgment until some answer is uniquely simplest.



Revolutionary Ockham Solutions

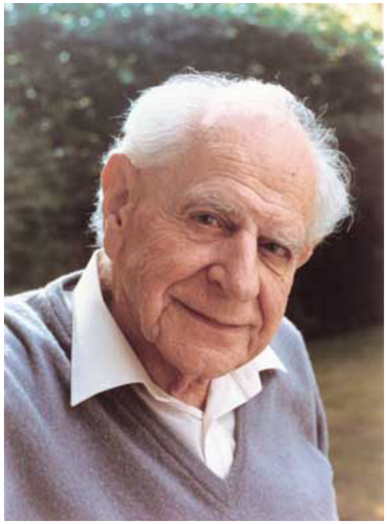
"Popper":
choose the paradigm with fewer free parameters.

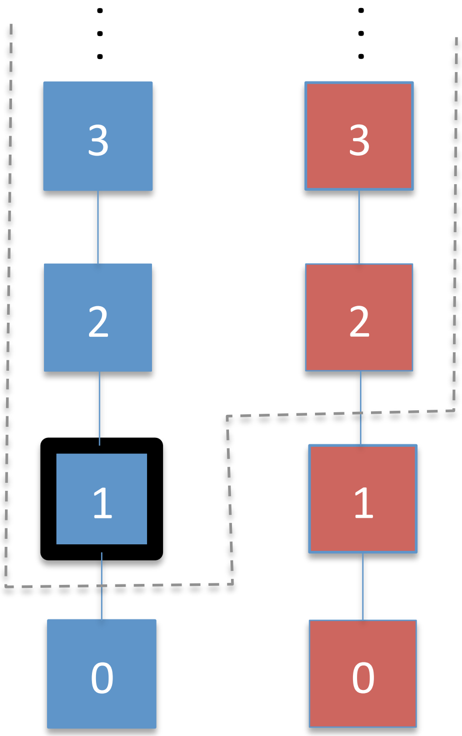




Revolutionary Ockham Solutions

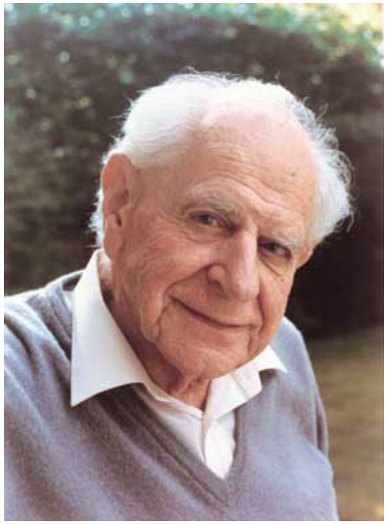
"Popper":
choose the paradigm with fewer free parameters.

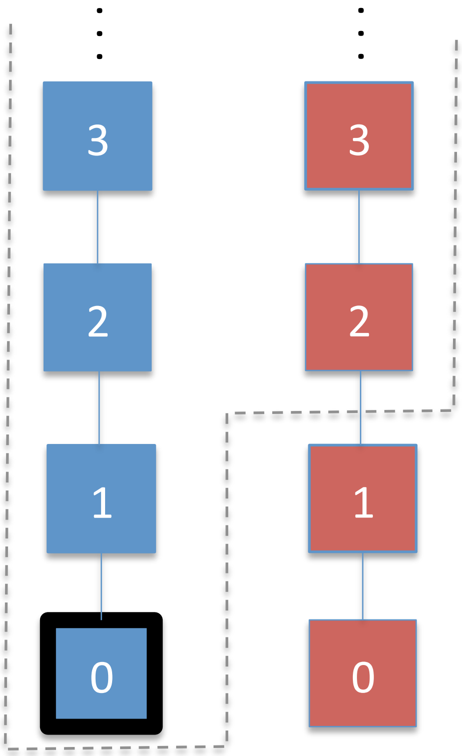




Revolutionary Ockham Solutions

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

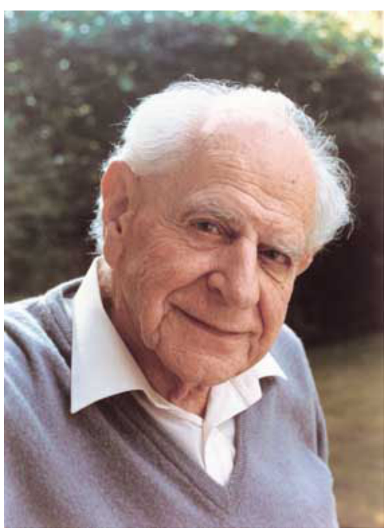


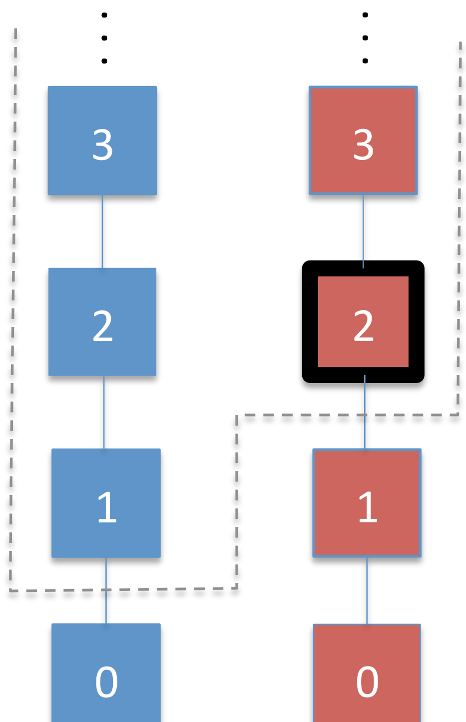


Lakatos:
choose the paradigm that was adjusted least recently.

Revolutionary Ockham Solutions

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





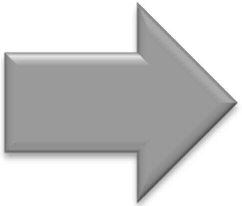
Lakatos:
choose the paradigm that was adjusted least recently.

Contextual Justification

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

Theorem

In normal problems, a solution is reversal optimal iff it is patient.



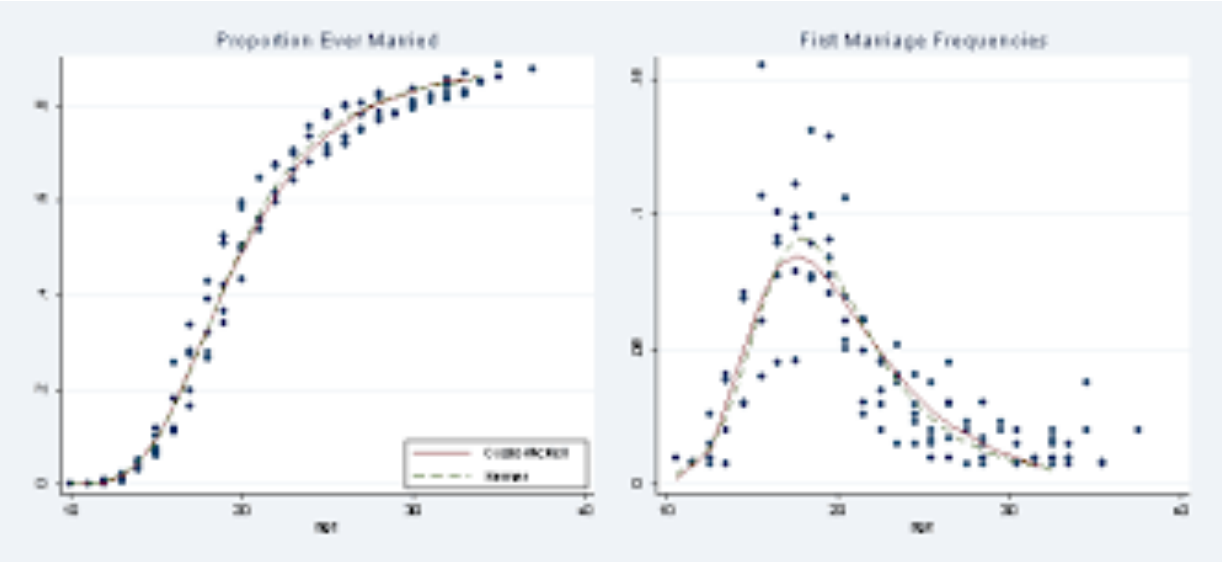
Summary and Discussion

- Simplicity is a topological feature of problems.
- Ockham's razor is necessary for cycle-optimal convergence to the true answer.
- Patience is necessary for reversal-optimal convergence to the true answer.
- Optimally straight convergence is weak, but its implications for scientific method are strong.

9. OCKHAM’S STATISTICAL RAZOR

Noisy Data

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



Stats Wars

Bayesianism

- (+) Induction
- (-) Unreliable

Frequentism

- (-) No induction
- (+) Reliable

Darth Bayeser   Luke Estimator

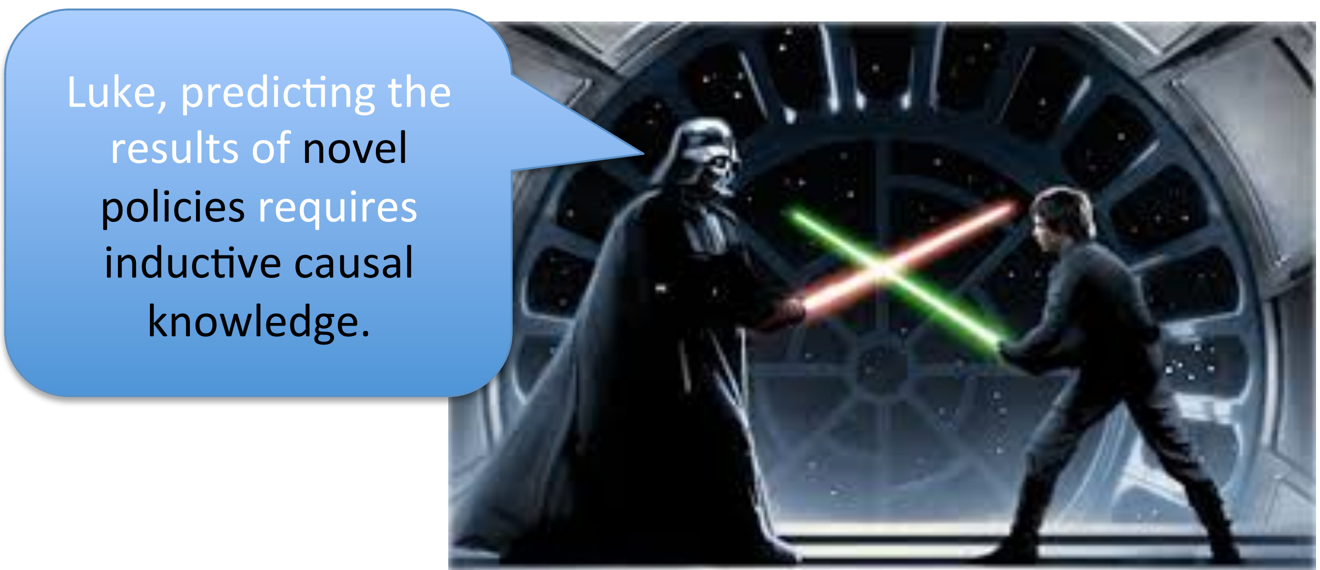
Stats Wars

Bayesianism

- (+) Induction
- (-) Unreliable

Frequentism

- (-) No induction
- (+) Reliable



Darth Bayeser

Luke Estimator

Stats Wars

Bayesianism

- (+) Induction
- (-) Unreliable

Frequentism

- (-) No induction
- (+) Reliable

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 (after 15 years)

Everything carries over very nicely, if you do it just right.

A glance at how it is done follows.

"In Chance" Translation

Topology

- *W*
- Input information
- Method
- Topology on *W*

- Simplicity

- Convergence
- Reversals
- Ockham's razor
- Patience

Statistics

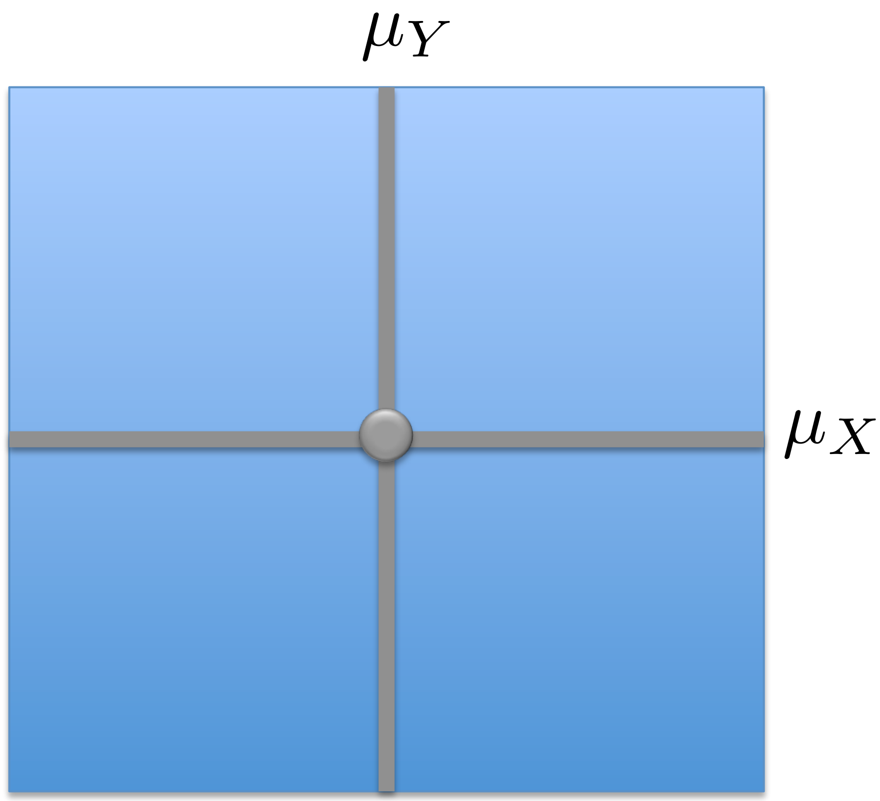
- A set of probability measures
- Random (iid) sample $\mathbf{X}_N$
- A measurable function M_N of $\mathbf{X}_N$
- $f(p) = p(\mathbf{X}_N \in S)$ is continuous

- Simplicity

- Convergence **in chance**
- Reversals **in chance**
- Error razor **in chance**
- Error patience **in chance**

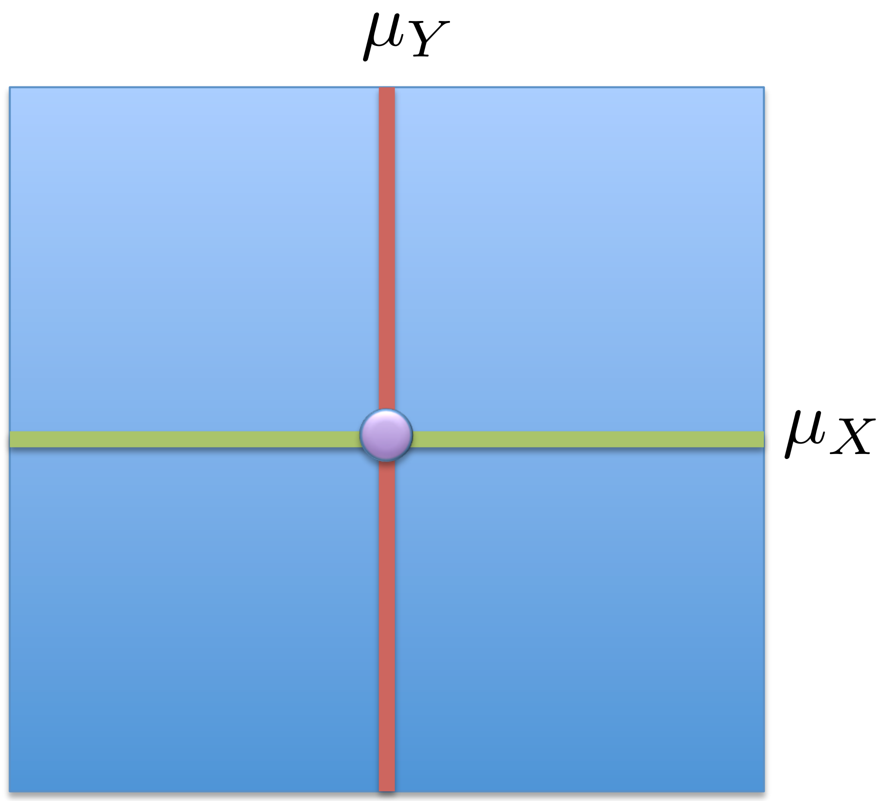
Example

- Worlds: bivariate independent normal distributions.



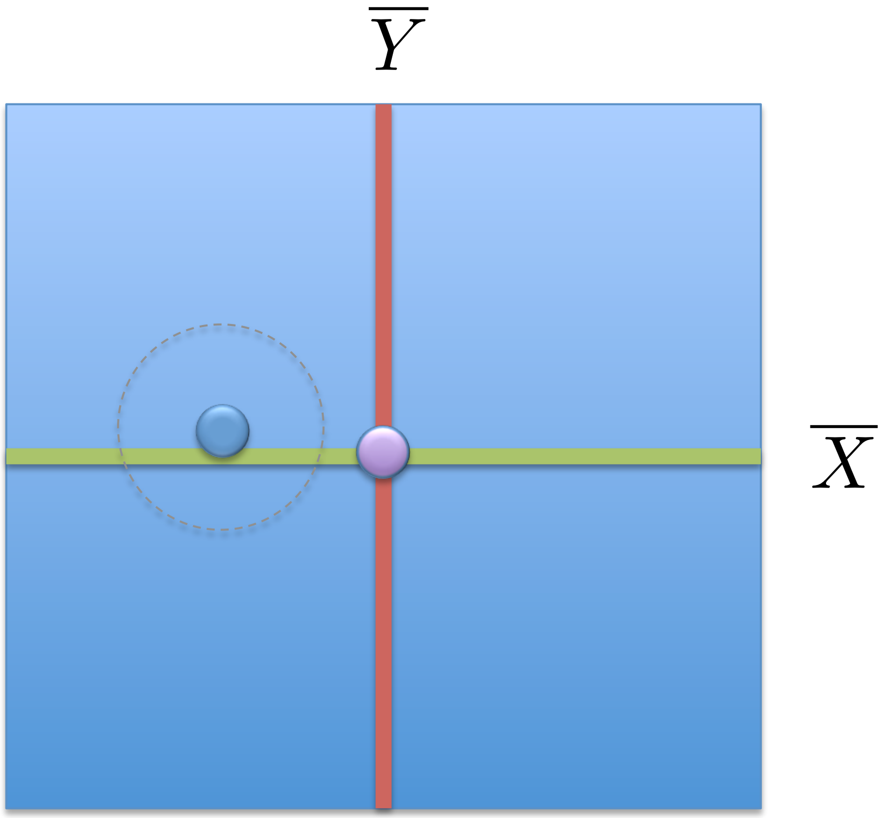
Example

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



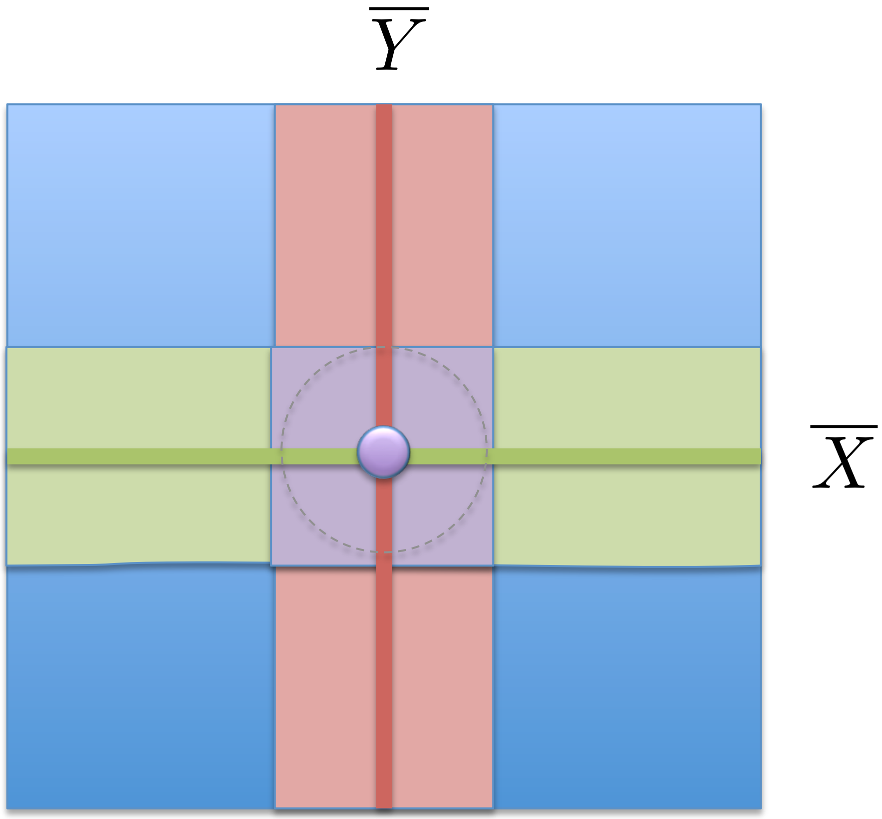
Example

- 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 answers.



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 N and Borel event S in R^N.

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

Reversals in Chance

Sampling distribution: p

Method: M

Reversal sequence: (A_0, A_1, A_2)

Ascending sample sizes: (N_0, N_1, N_2)

Output chances: $p^i_j \; := \; p_{N_j}(M = A_i).$

Pre-reversal odds: $a_i \; := \; p^{i+1}_i / p^i_i.$

Post-reversal odds: $b_i \; := \; p^{i+1}_{i+1} / p^i_{i+1}.$

Odds differences: $d_i \; := \; b_i \dot{-} a_i.$

Reversals in Chance

Reversal sequence: (A_0, A_1, A_2)

Ascending sample sizes: (N_0, N_1, N_2)

"How much of the drop in chance of producing A_i can be accounted for by the rise in chance of producing A_{i+1}?"

$$\begin{aligned} p^i_j &:= p_{N_j}(M = A_i). \\ d_i &:= \underbrace{(p^i_i \dot{-} p^i_{i+1})}_{\text{Drop for } A_i} \dot{-} \underbrace{(p^{i+1}_{i+1} \dot{-} p^{i+1}_i)}_{\text{Rise for } A_{i+1}}. \end{aligned}$$

Reversals in Chance

Reversal sequence: (A_0, A_1, A_2)

Ascending sample sizes: (N_0, N_1, N_2)

"How much of the drop in chance of producing A_i can be accounted for by the rise in chance of producing A_{i+1}?"

$$
\begin{aligned}
p_j^i &:= p_{N_j}(M = A_i). \\
d_i &:= (p_i^i \dot{-} p_{i+1}^i) \dot{-} (p_{i+1}^{i+1} \dot{-} p_i^{i+1}).
\end{aligned}
$$

Reversals in Chance

Reversal sequence: (A_0, A_1, A_2)

Ascending sample sizes: (N_0, N_1, N_2)

"How much of the drop in chance of producing A_i can be accounted for by the rise in chance of producing A_{i+1}?"

$$
\begin{aligned}
p^i_j &:= p_{N_j}(M = A_i).\\
d_i &:= (p^i_i \mathbin{\dot{-}} p^i_{i+1}) \mathbin{\dot{-}} (p^{i+1}_{i+1} \mathbin{\dot{-}} p^{i+1}_i).
\end{aligned}
$$

Reversals in Chance Compared

$$(A_0, \quad A_1, \quad A_2)$$
$$d_0 \qquad d_1$$

reverses as badly as

$$(A'_0, \quad A'_1, \quad A'_2)$$
$$d'_0 \qquad d'_1$$

iff $A_i \subseteq A'_i$, for $i \leq 2$;
$d_i \geq d'_i$, for $i < 2$.

Comparison with α Tolerance

$$\begin{array}{ccccc} (A_0, & & A_1, & & A_2) \\ & d_0 & & d_1 & \end{array}$$

reverses as
α -badly as

$$\begin{array}{ccccc} (A'_0, & & A'_1, & & A'_2) \\ & d'_0 & & d'_1 & \end{array}$$

iff $A_i \subseteq A'_i$, for $i \leq 2$;
$d_i \geq d'_i + \alpha$, for $i < 2$.

Method Comparison

- Solution M' reverses as α-badly as solution M iff:

  For each reversal sequence generated by solution M at some p and sample sizes,

  method M' generates a reversal sequence at some p' and sample sizes that reverses as α-badly.

- Similarly for cycles.
- Strict partial order, for $\alpha > 0$.

Optimal Truth Conduciveness

- Solution M is α reversal-optimal iff:

  Each solution in chance reverses as α-badly as M.

- Similarly for α cycle-optimality.

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

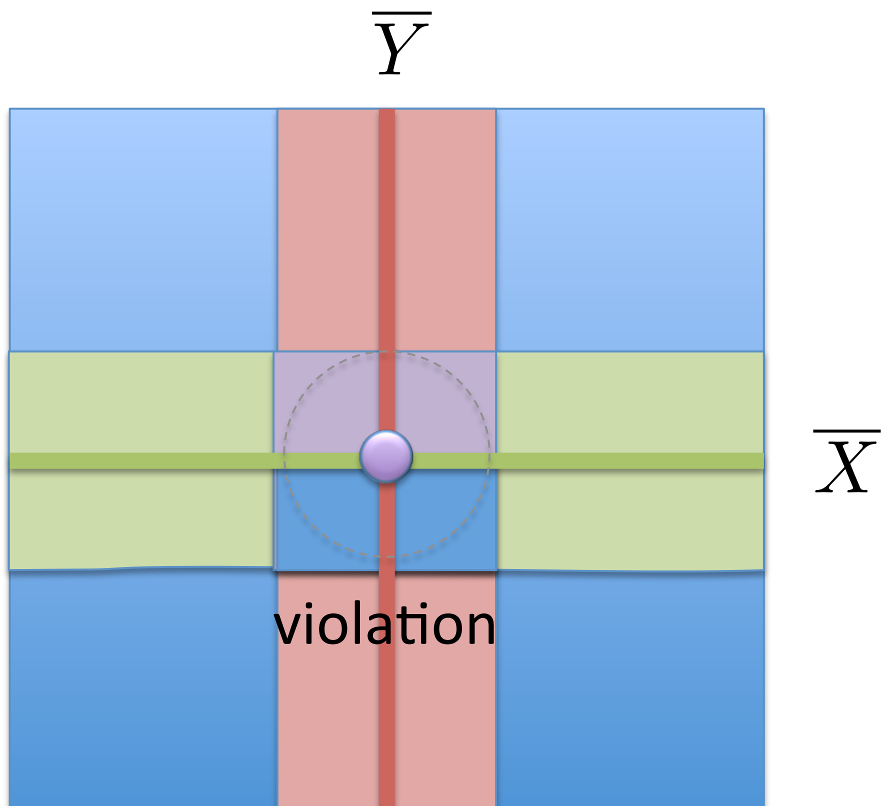
Proposition. In natural problems:

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

Ockham's α-Razor

Violate the error razor with at most chance α.

If A_p is properly simpler than B, then $p(M_N = B) < \alpha$.



Error patience

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

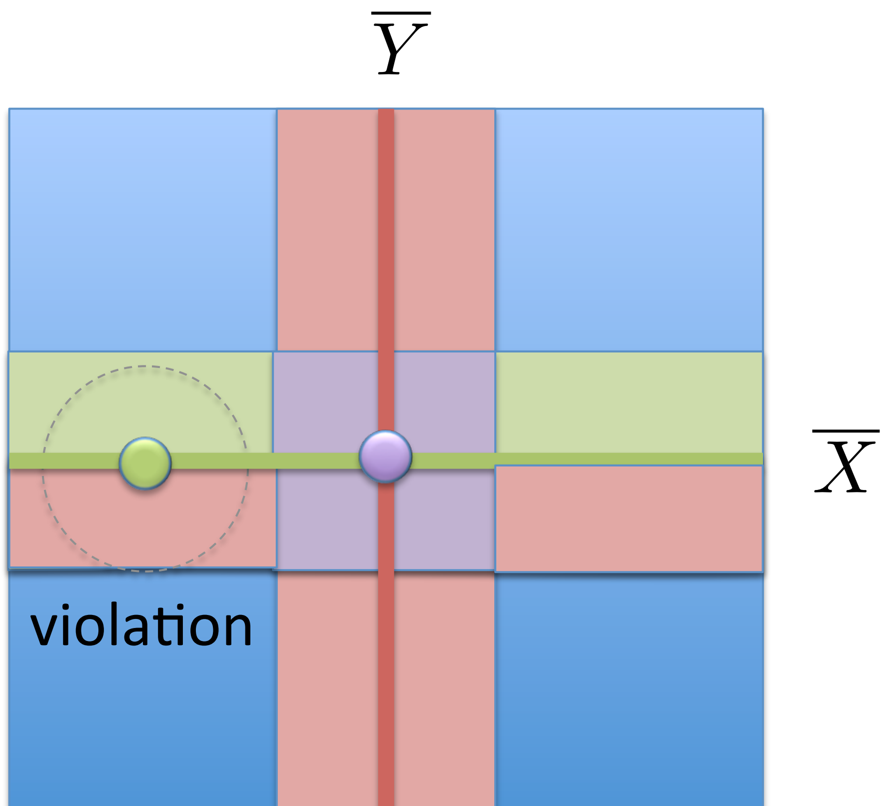
Equivalence

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

α-Patience

Violate error patience with at most chance α.

If B rules out every answer as simple as A_p, then $p(M_N = B) < \alpha$.



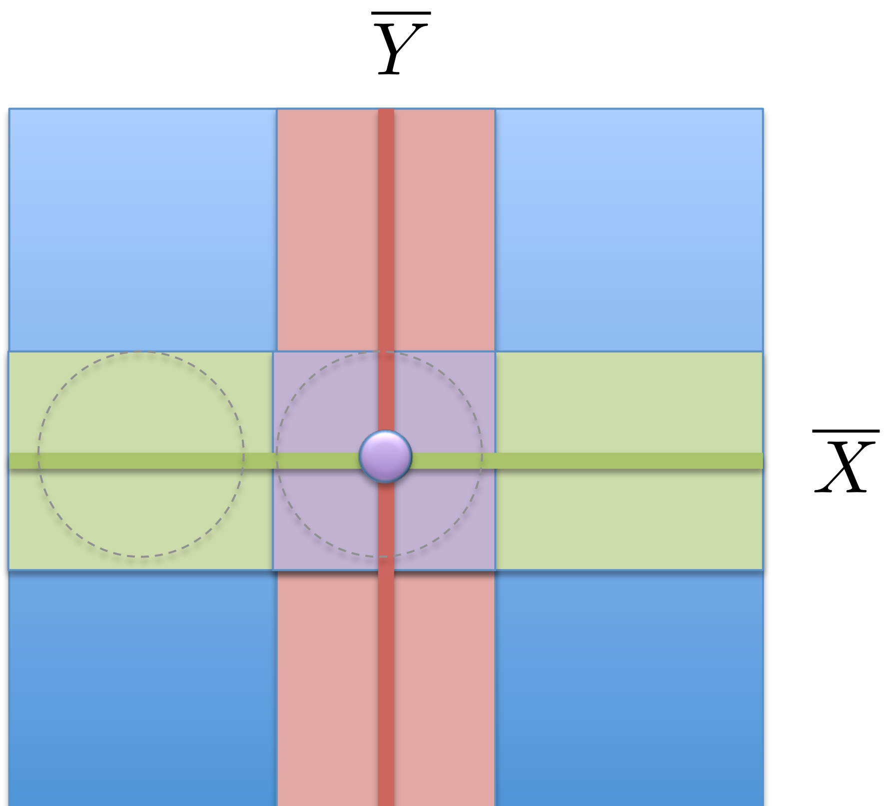
Theorem

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

Theorem

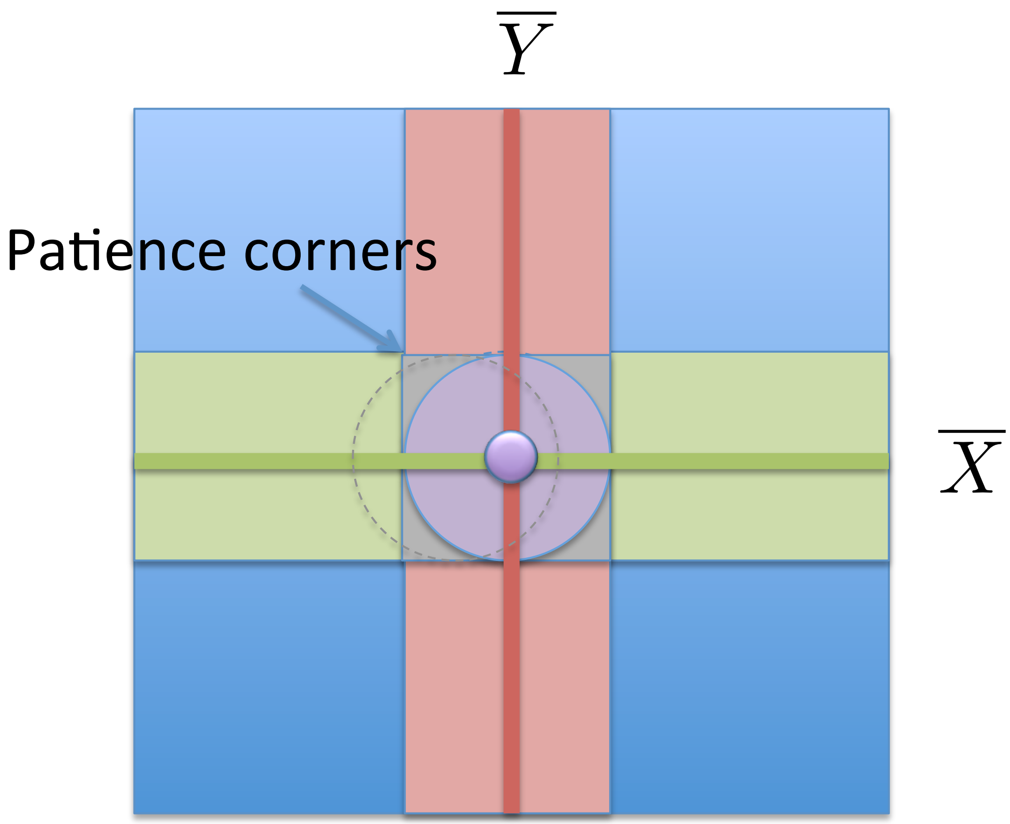
- **Proposition:** Every α-reversal optimal solution in chance to a statistical problem satisfies α-patience.

α-Patient, α-Ockham Solution



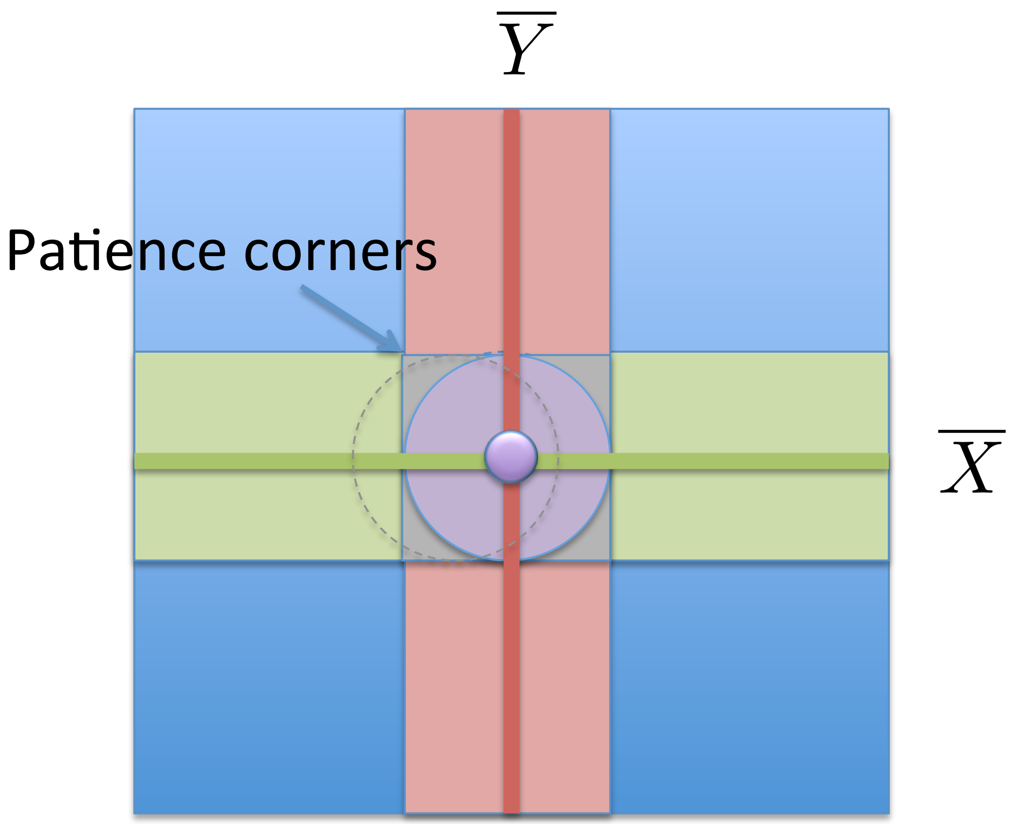
α-Patient, α-Ockham Solution

- Power-optimized version



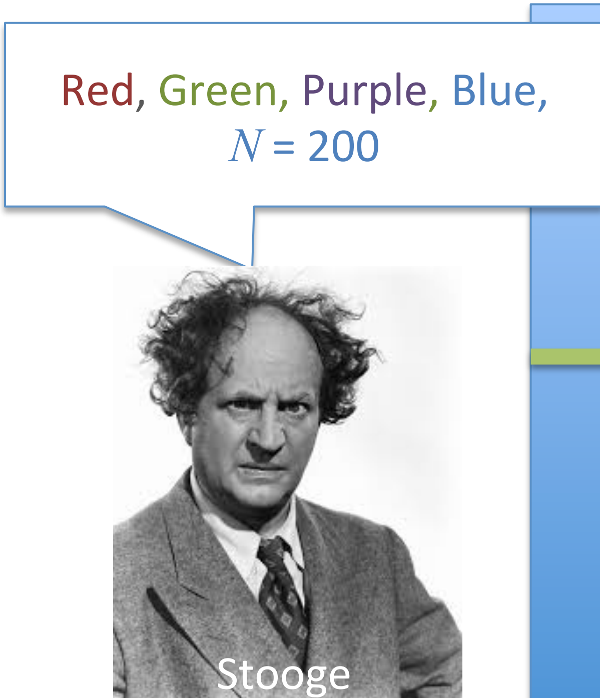
Fishing with Tests

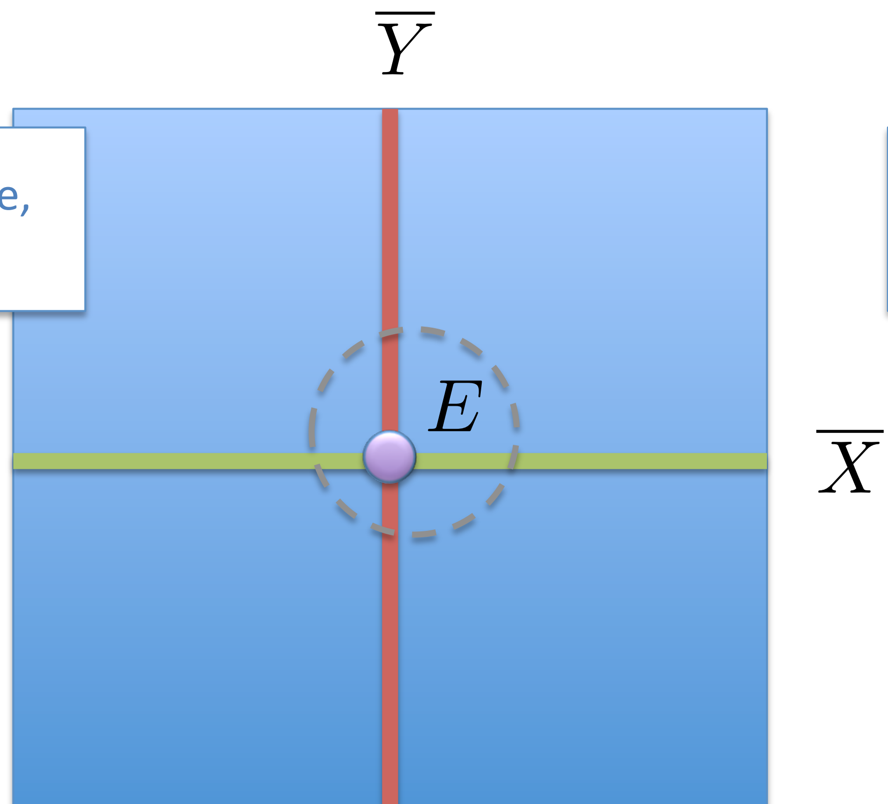
"Power" = get the unavoidable reversals over with as soon as possible for given α.



Confidence Intervals as Information

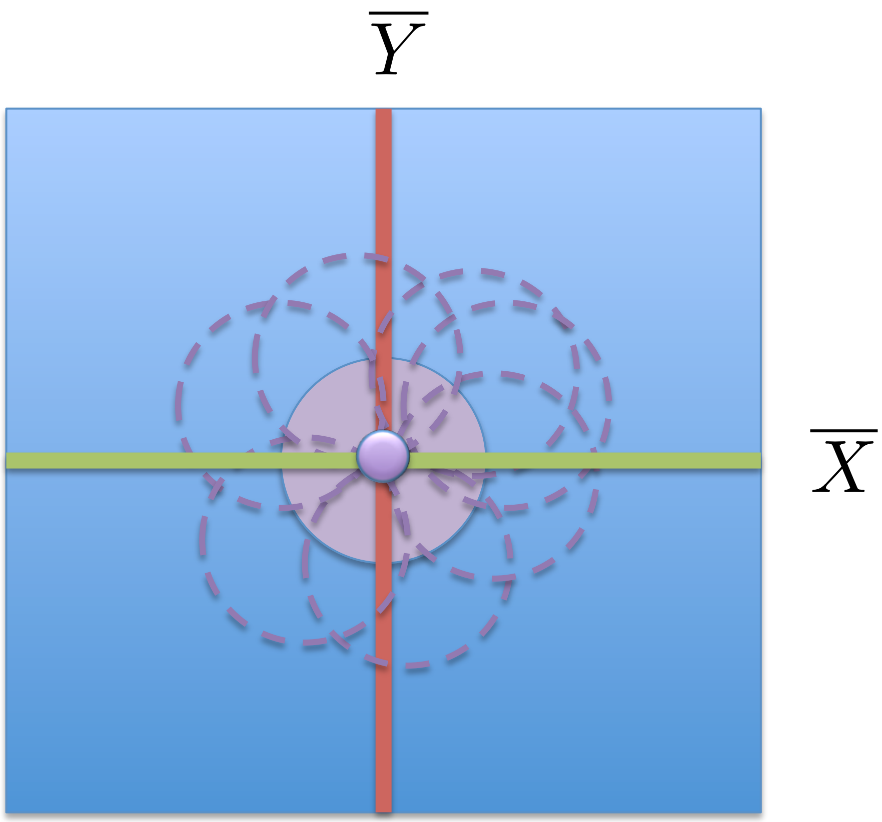
- Statistician's stooge constructs a confidence ball E and reports to the theorist the sample size N and which answers are logically compatible with E.
- Theorist applies the logical versions of Ockham's razor and patience.





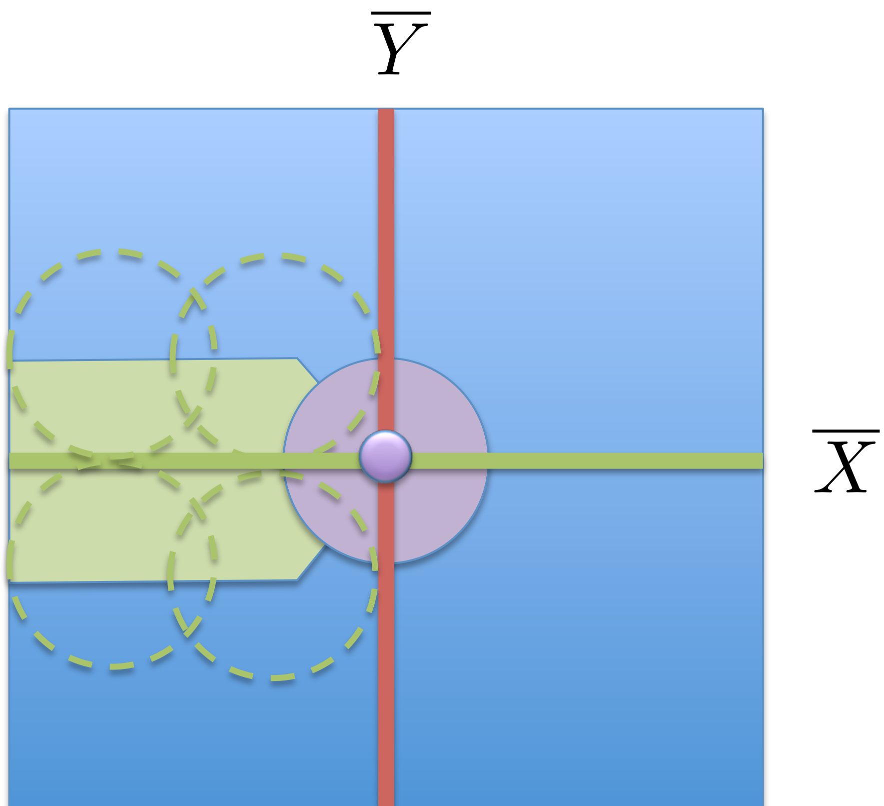
Confidence Intervals as Information

- Statistician's stooge constructs a confidence ball E and reports to the theorist which answers are logically compatible with E.
- Apply the logical versions of Ockham's razor and patience.



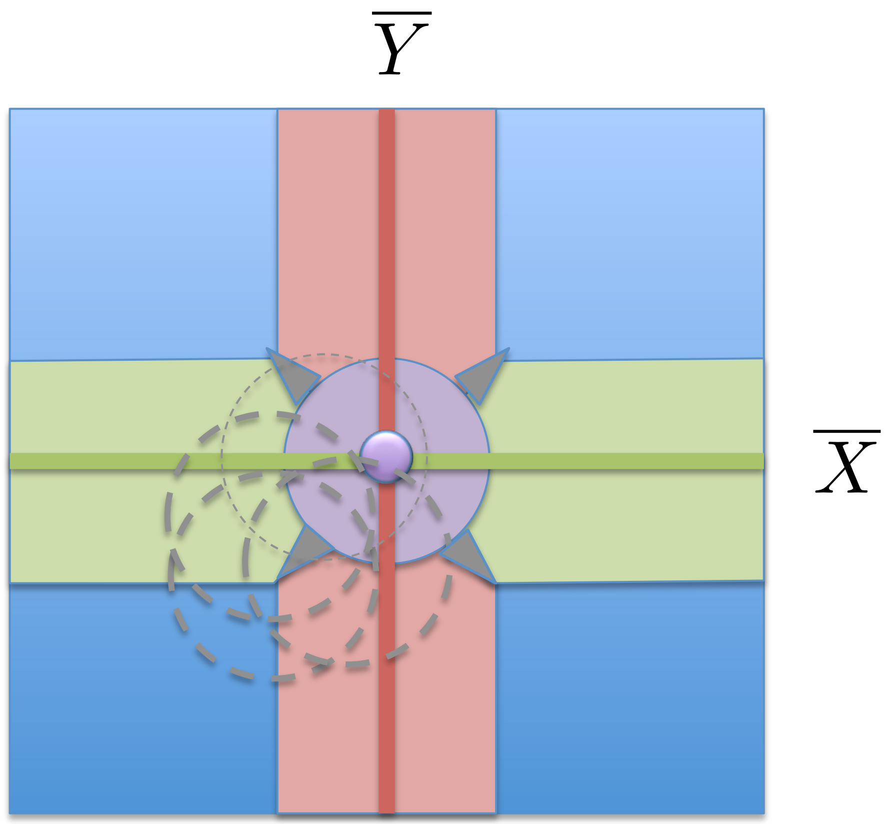
Confidence Intervals as Information

- Statistician's stooge constructs a confidence ball E and reports to the theorist which answers are logically compatible with E.
- Apply the logical versions of Ockham's razor and patience.



Confidence Intervals as Information

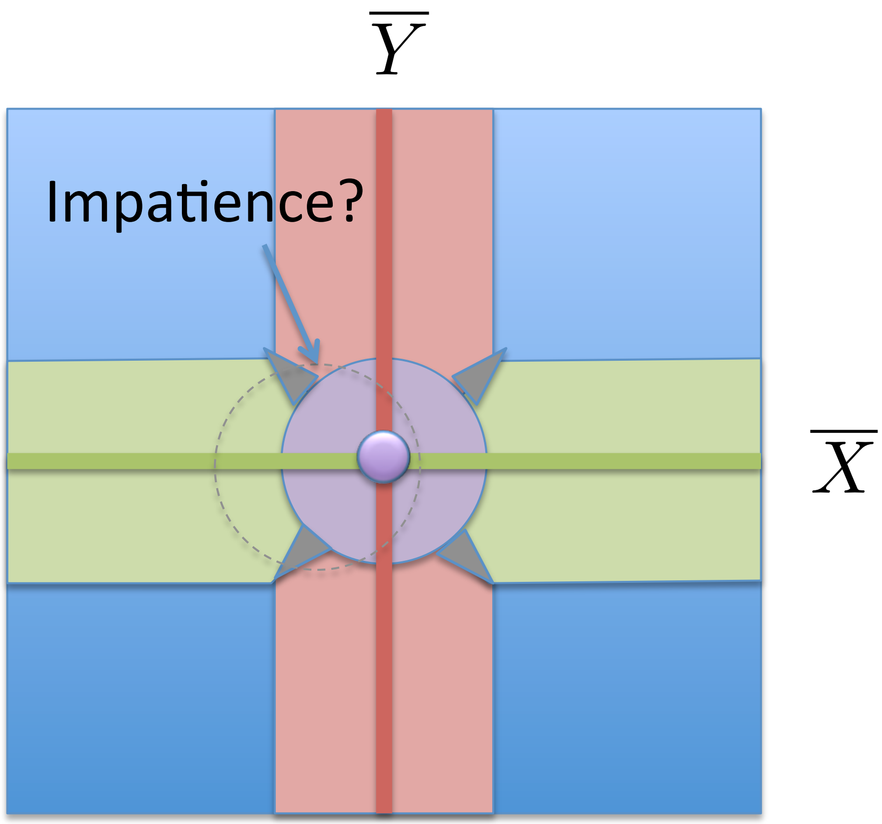
- Statistician's stooge constructs a confidence ball E and reports to the theorist which answers are logically compatible with E.
- Apply the logical versions of Ockham's razor and patience.



Confidence Intervals as Information

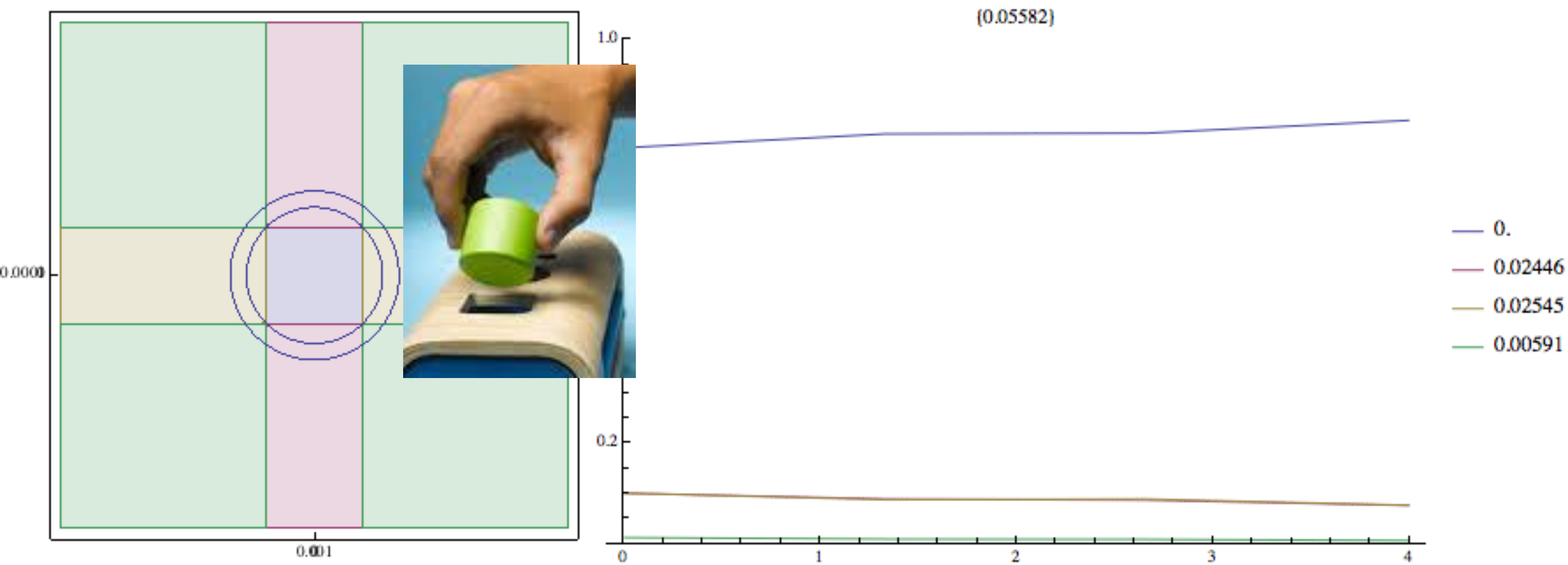
Reversal sub-optimal due to impatience?

Fishing with tests may be better.



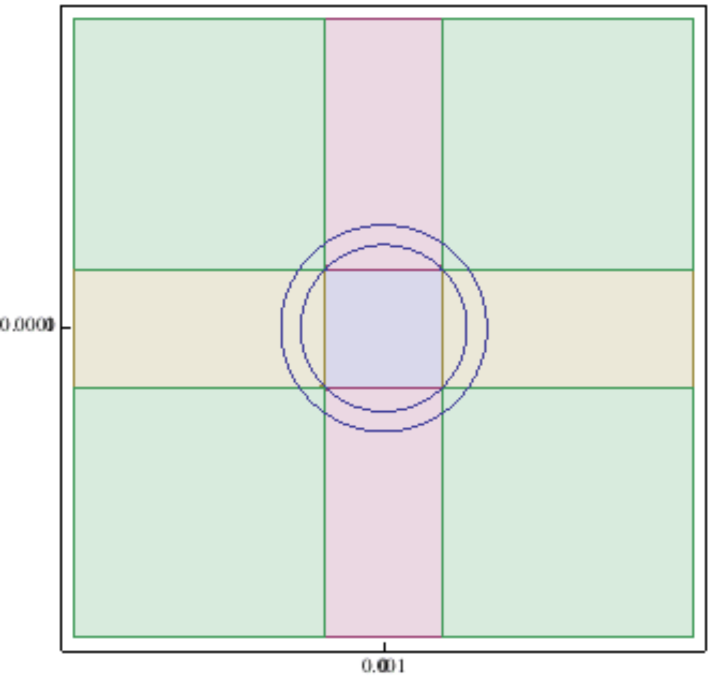
Simulations

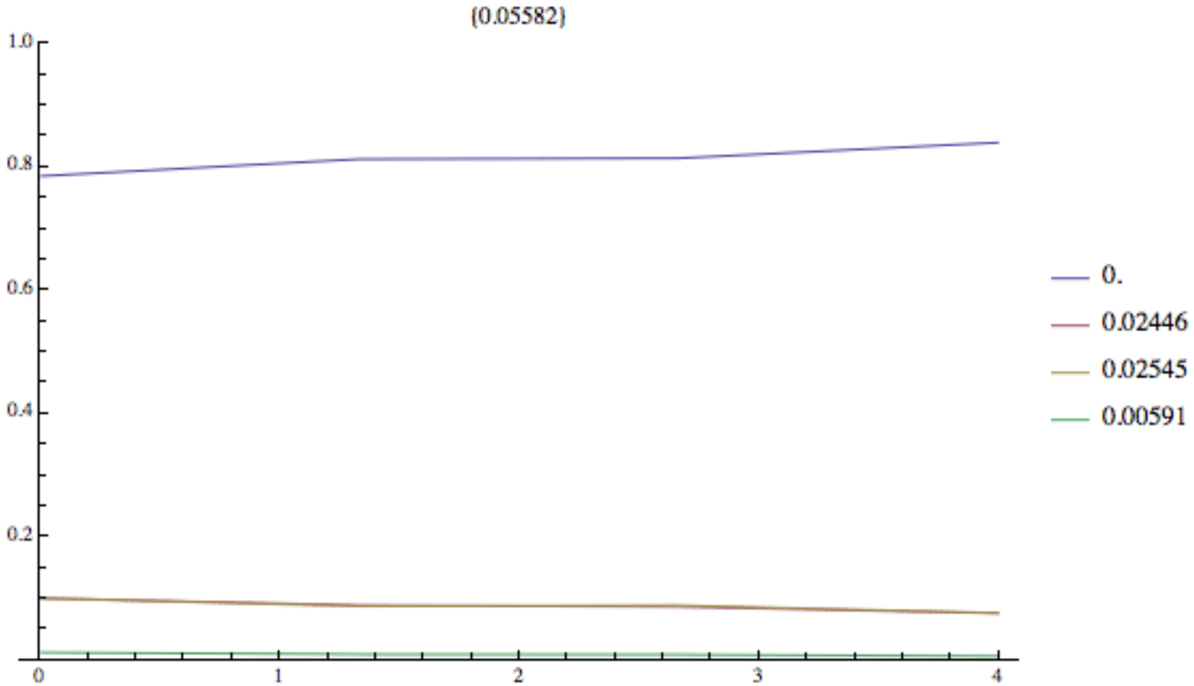
- Method: Minimize BIC score.
- 20% Ockham violation near 0,0.
- 10% impatience near 0,0.
- Blue zone is optimized against small errors in simplicity.



Simulations

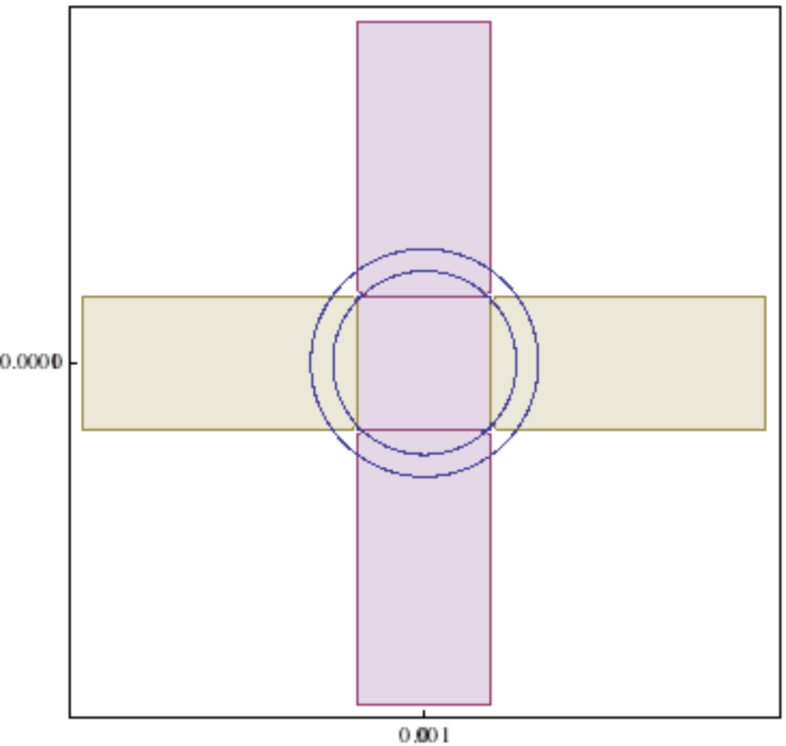
- Method: Minimize BIC score.
- 20% Ockham violation near 0,0.
- 10% impatience near 0,0.
- Blue zone is optimized against small errors in simplicity.

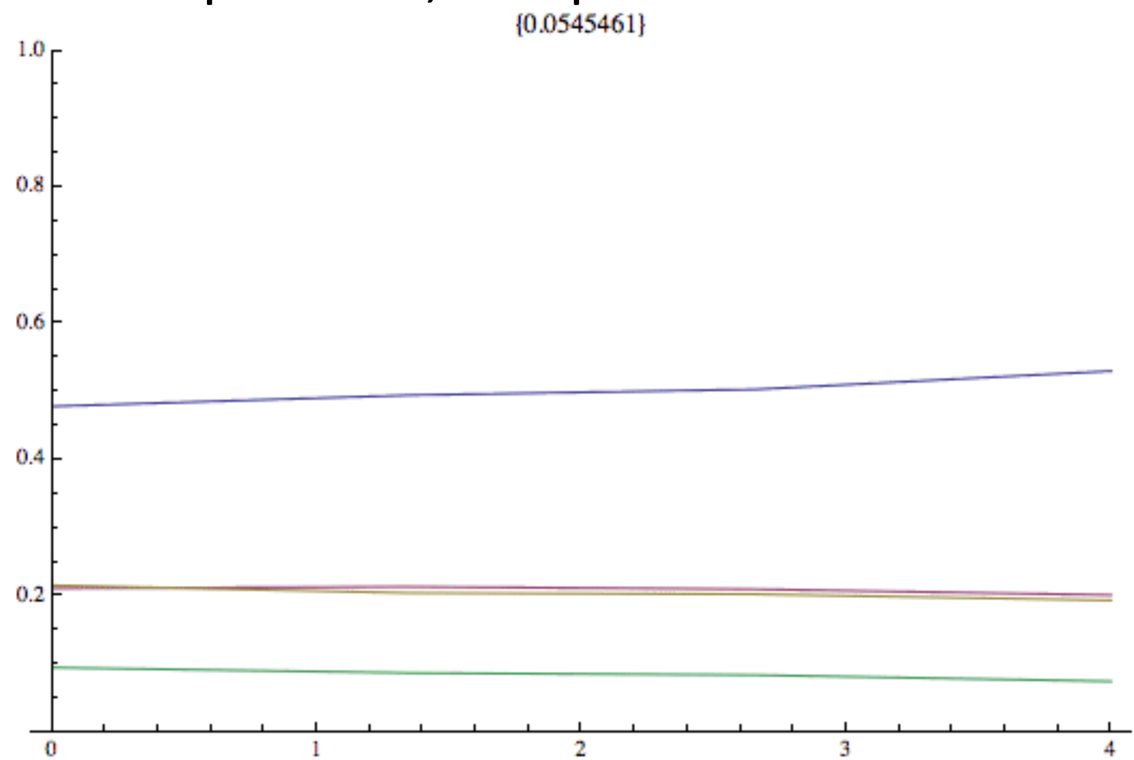




Bayes

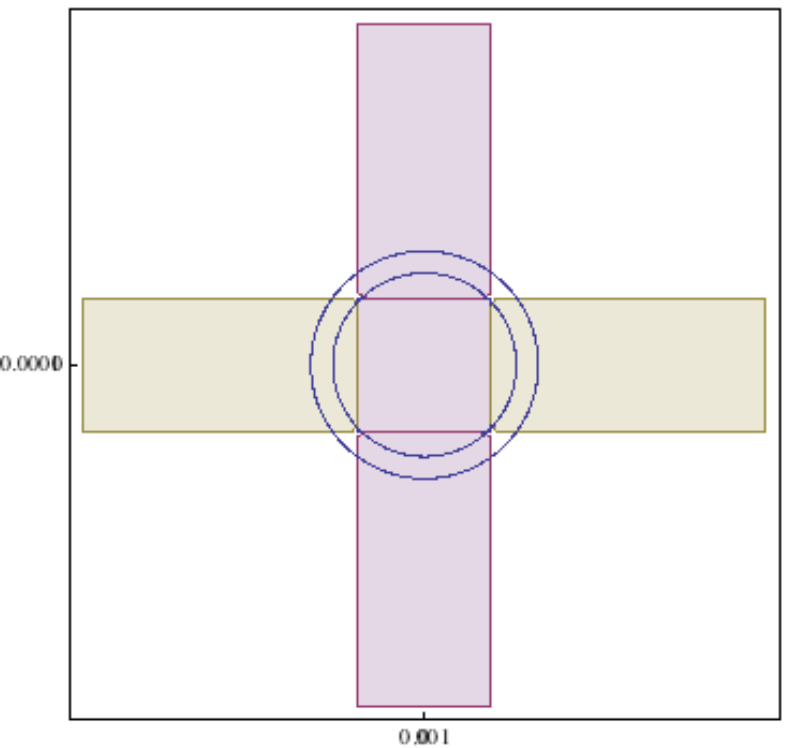
- Method: Maximize Bayesian credence,
- Even priors on models, Gaussian priors on parameters.
- 20% Ockham violation, 20% impatience, bad power.

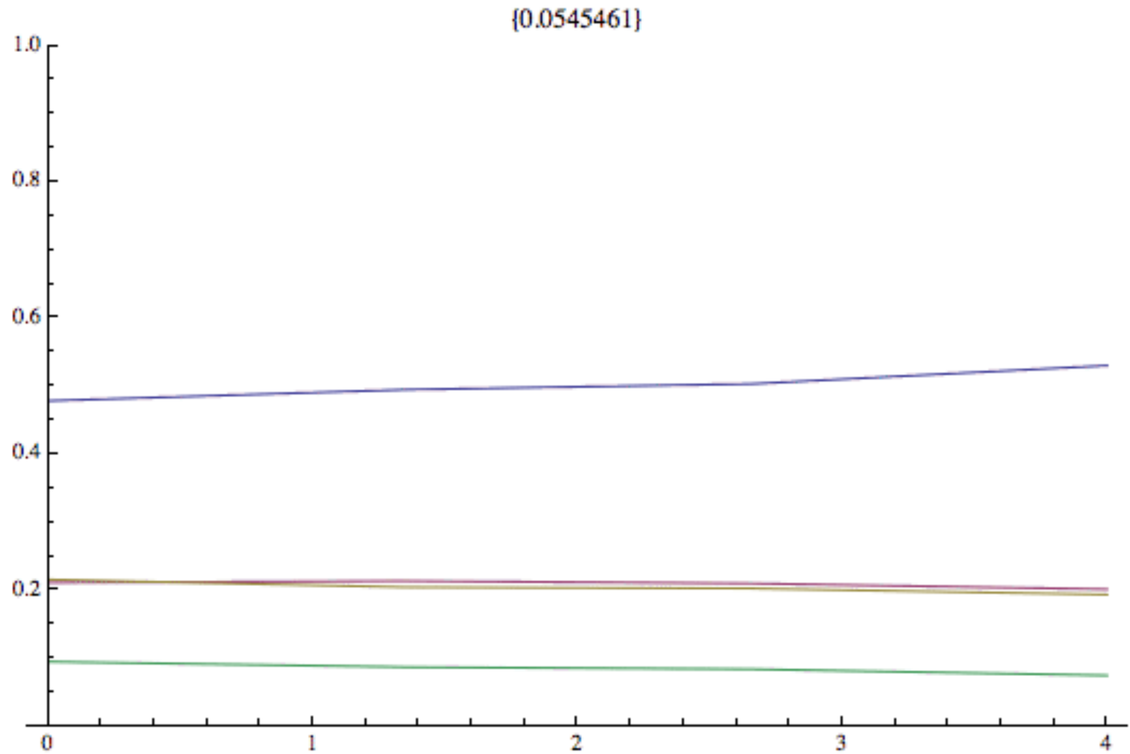




Bayes

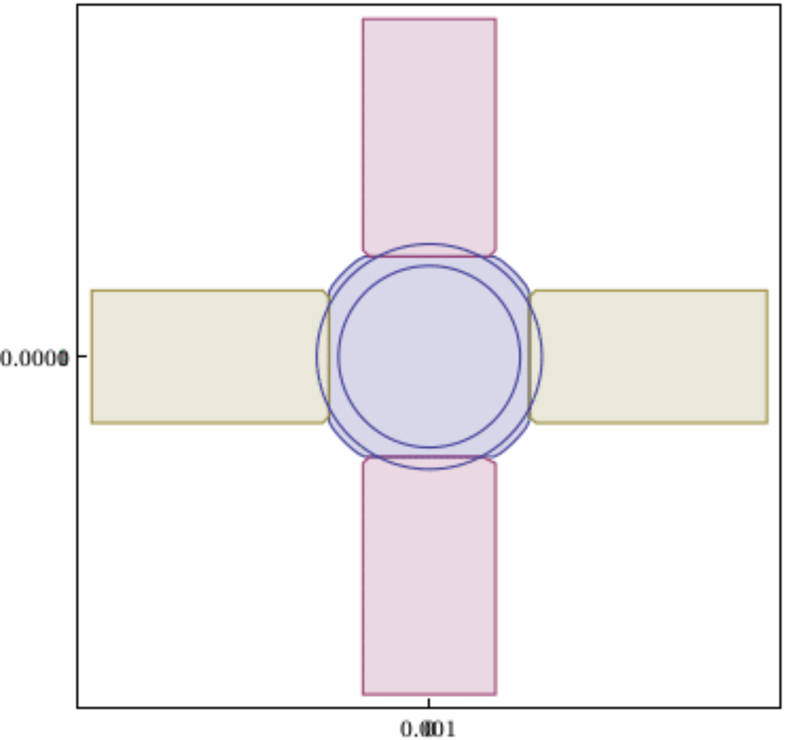
- Method: Maximize Bayesian credence,
- Even priors on models, Gaussian priors on parameters.
- 20% Ockham violation, 20% impatience, bad power.

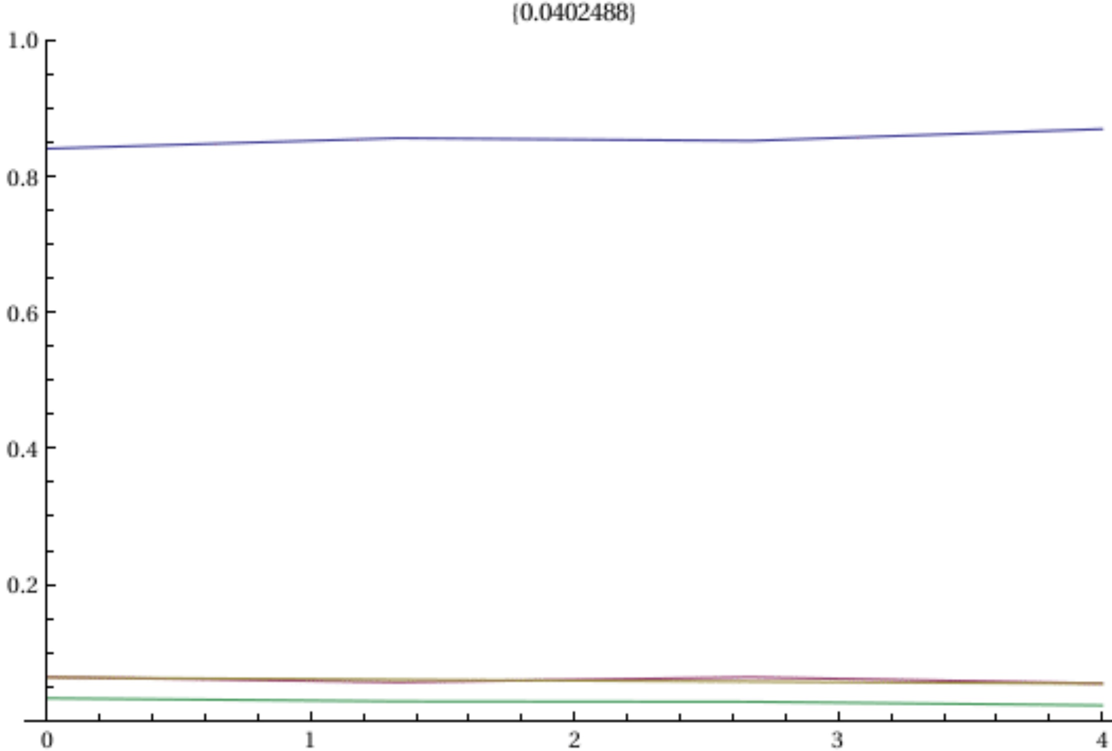




Bayes

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



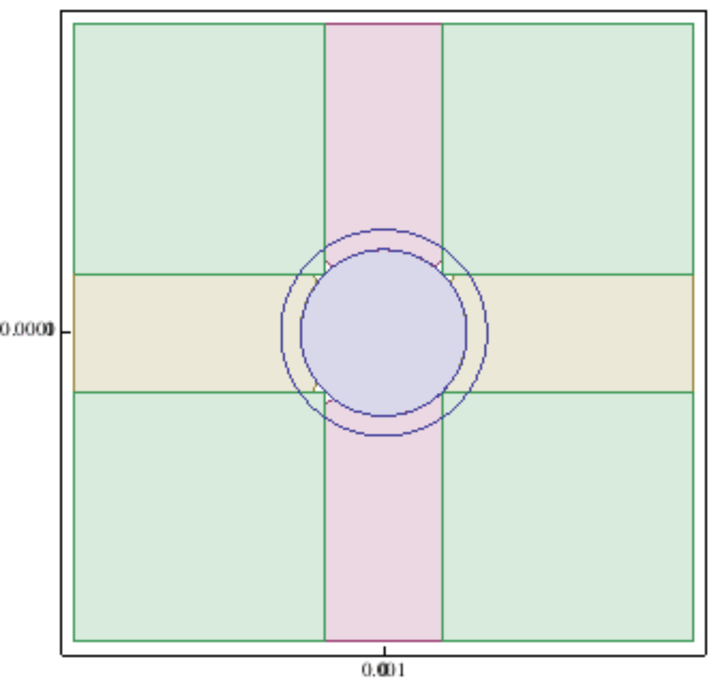


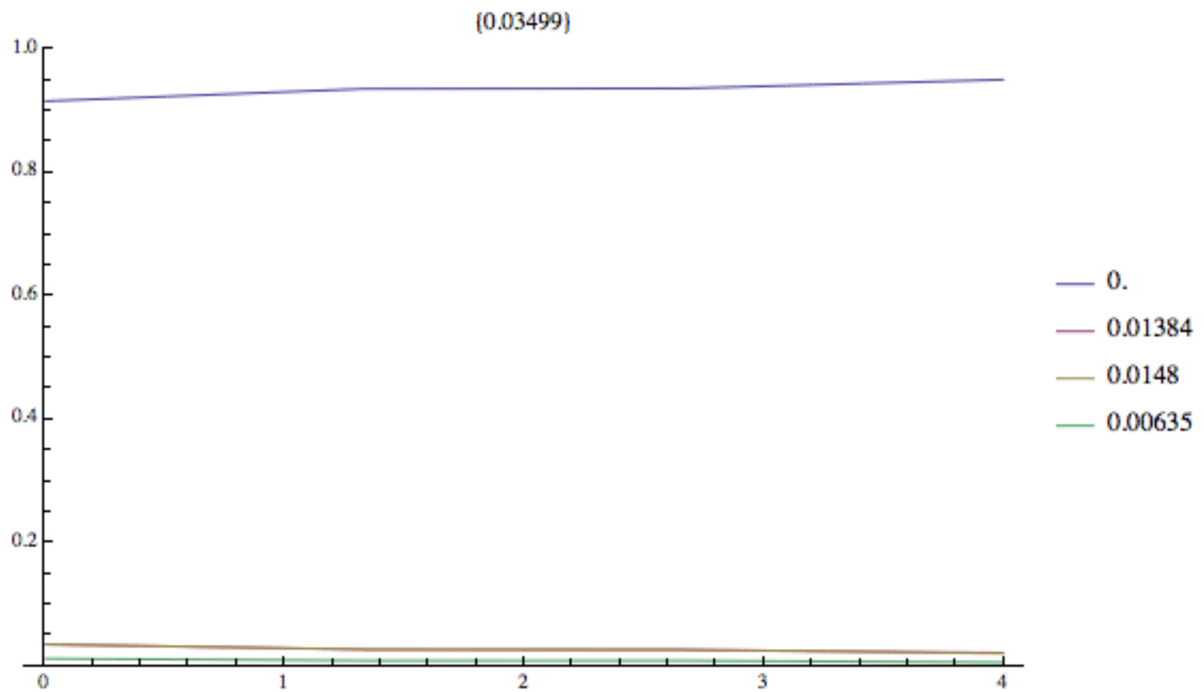
Plausible Advice

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

Simulations

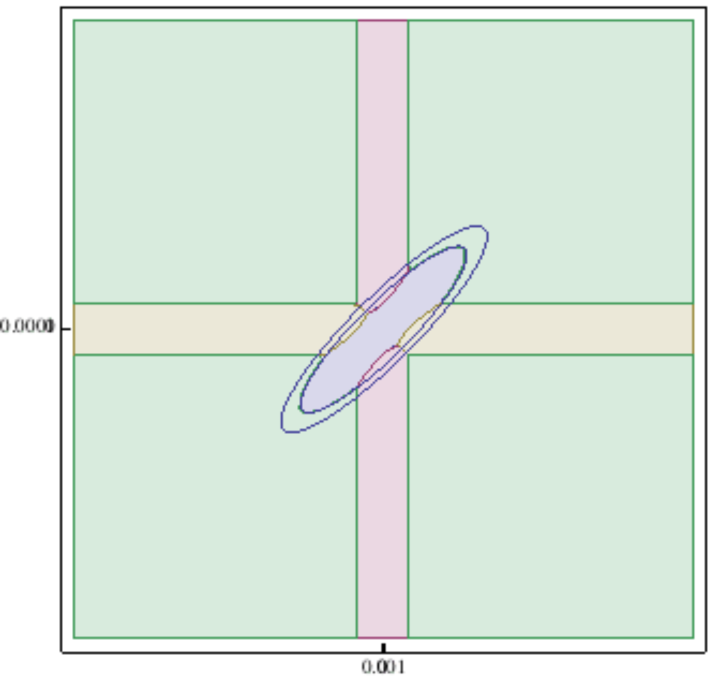
- Method: Nested BIC.

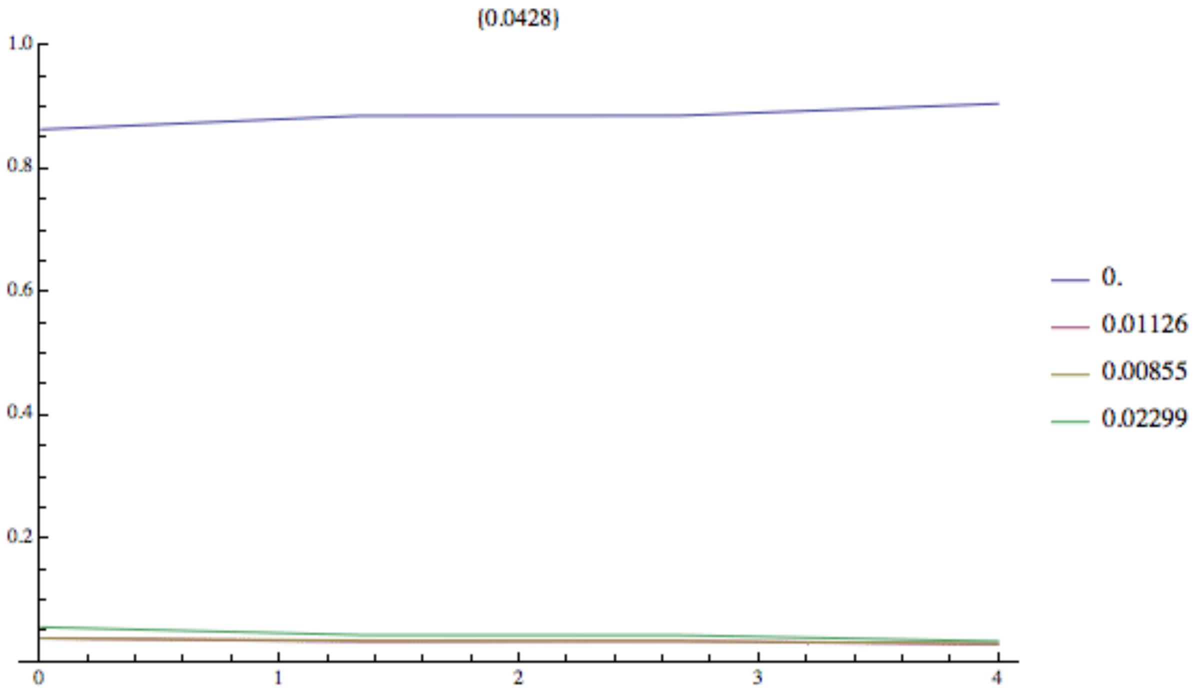




Simulations

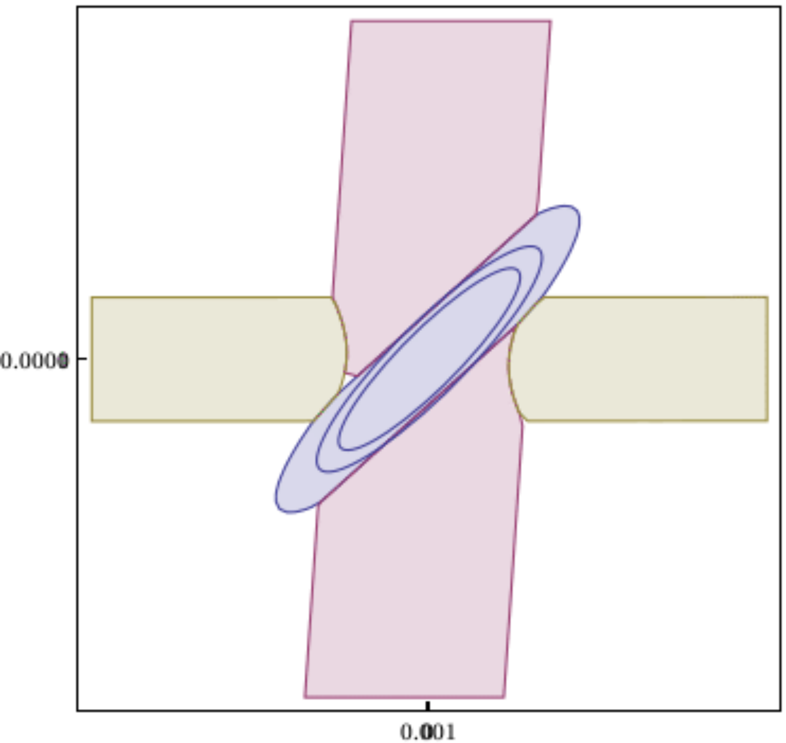
- Method: Minimize BIC.

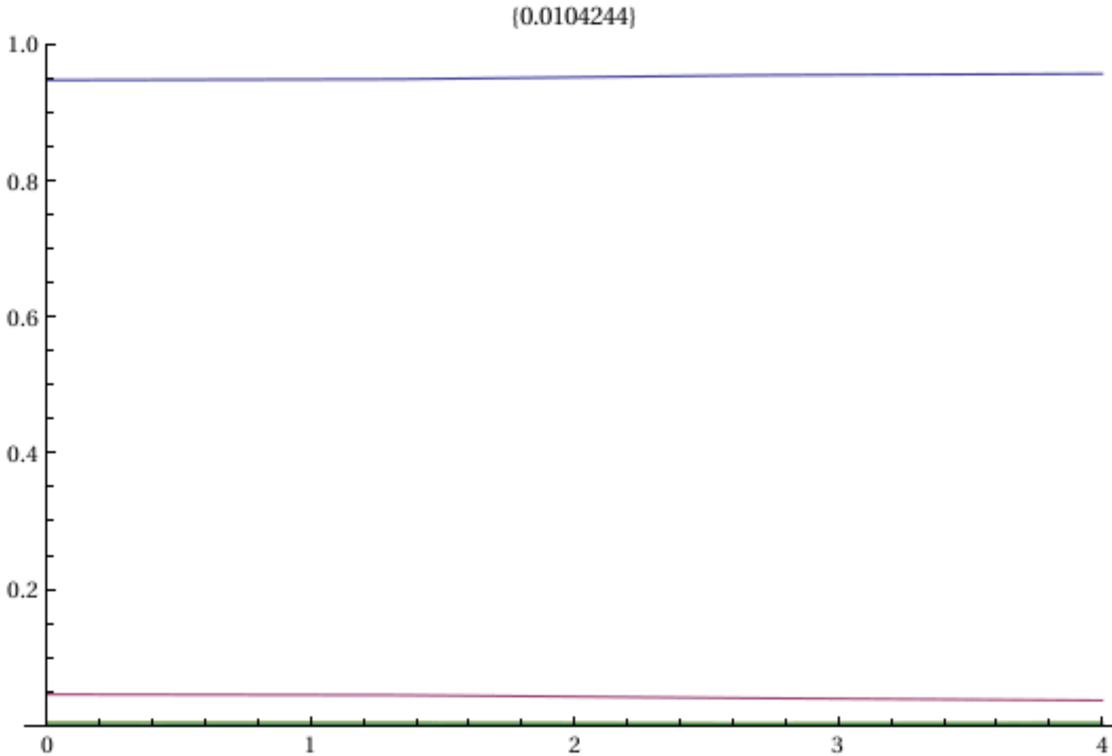




Bayes with Correlation

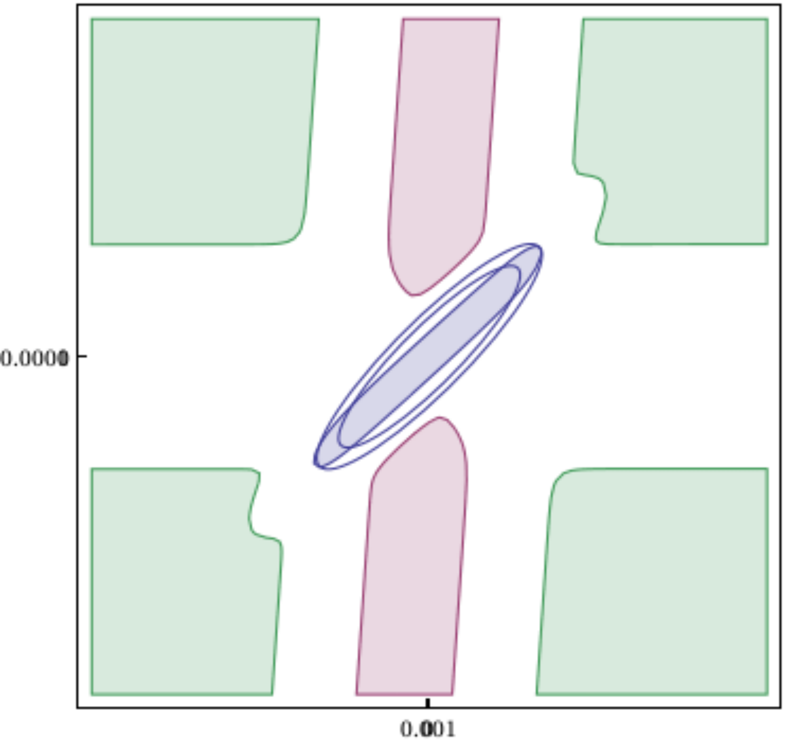
- Method: Maximize Bayesian credence,
- Prior bias toward simplicity, Gaussian priors on parameters.
- 40% impatience, bad power.

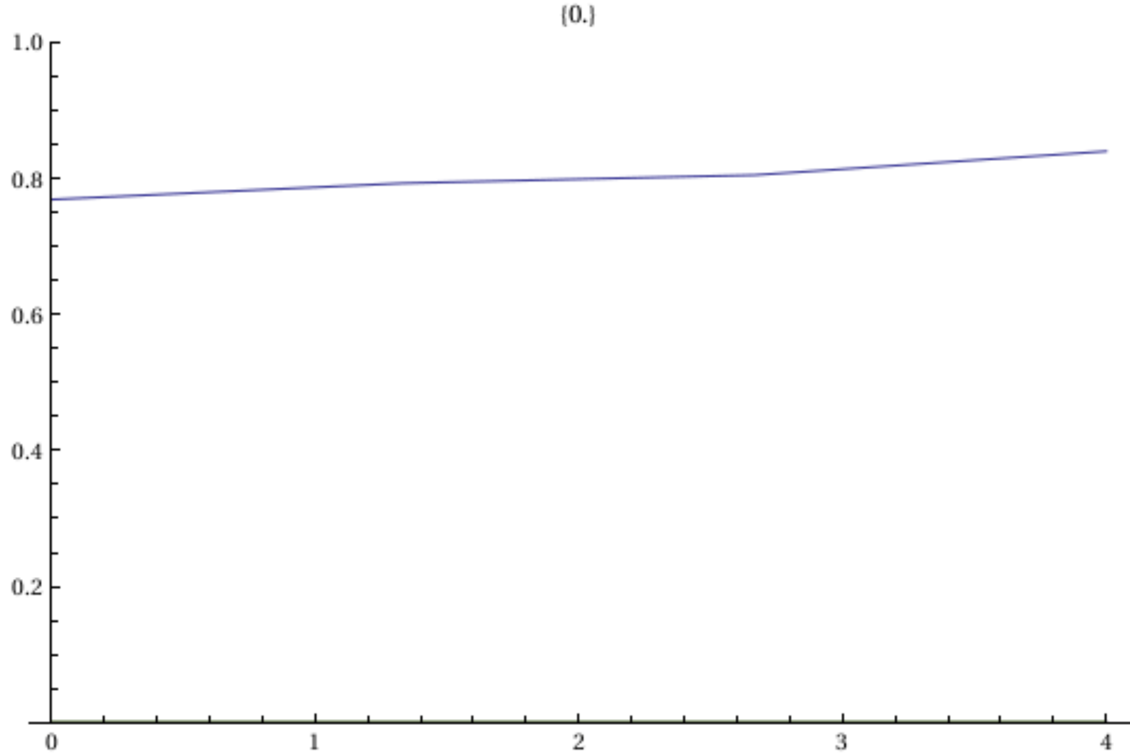




The Power of Modesty

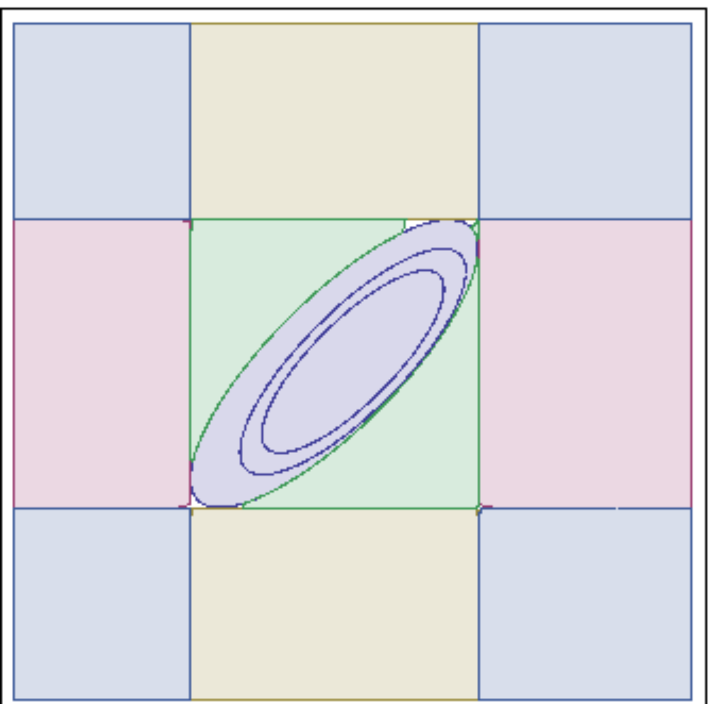
- Method: 95% threshold for Bayes Posterior
- Waiting for "confirming data" brings reversals in chance down to around 5%.

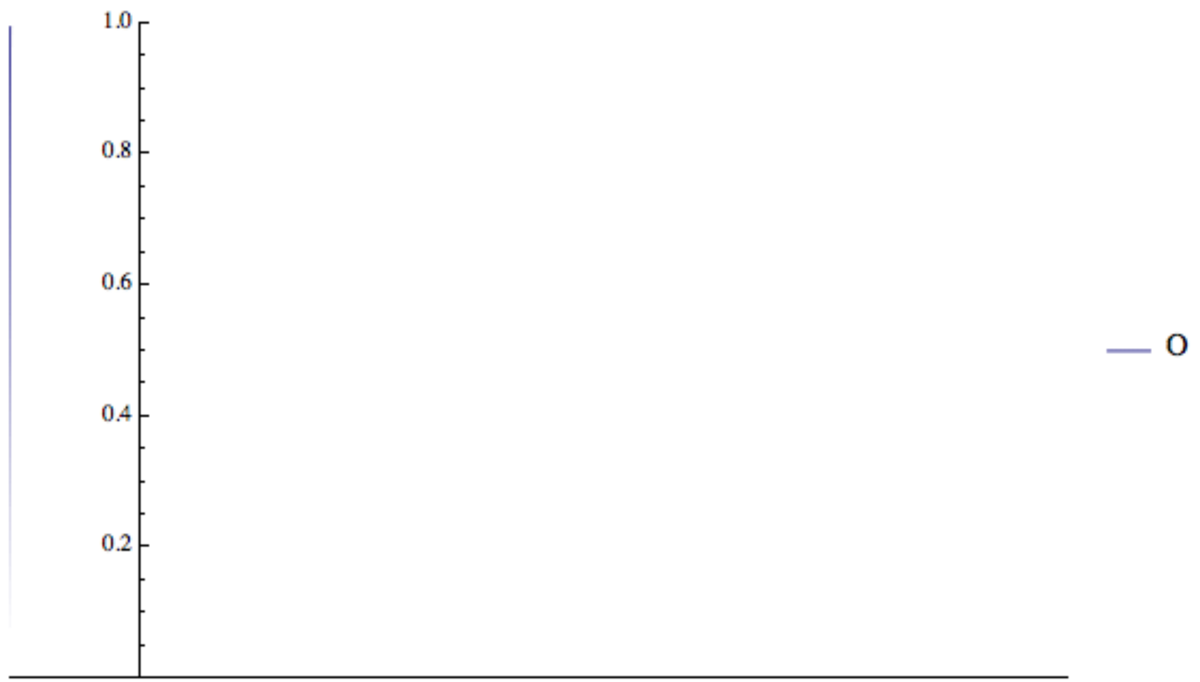




Ockham's Frequentist Razor

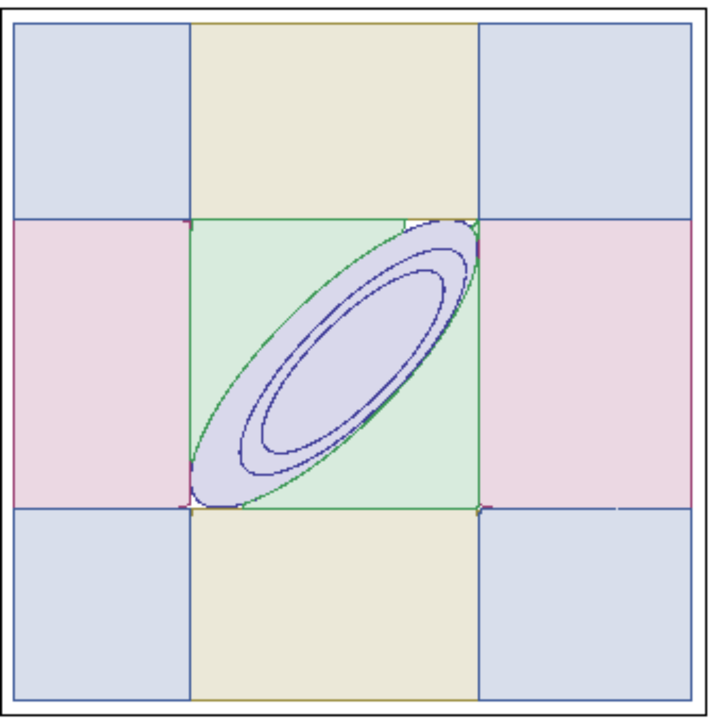
- Choose powerful nested tests at significance α .
- Disjoin the simplest models whose tests do not reject.

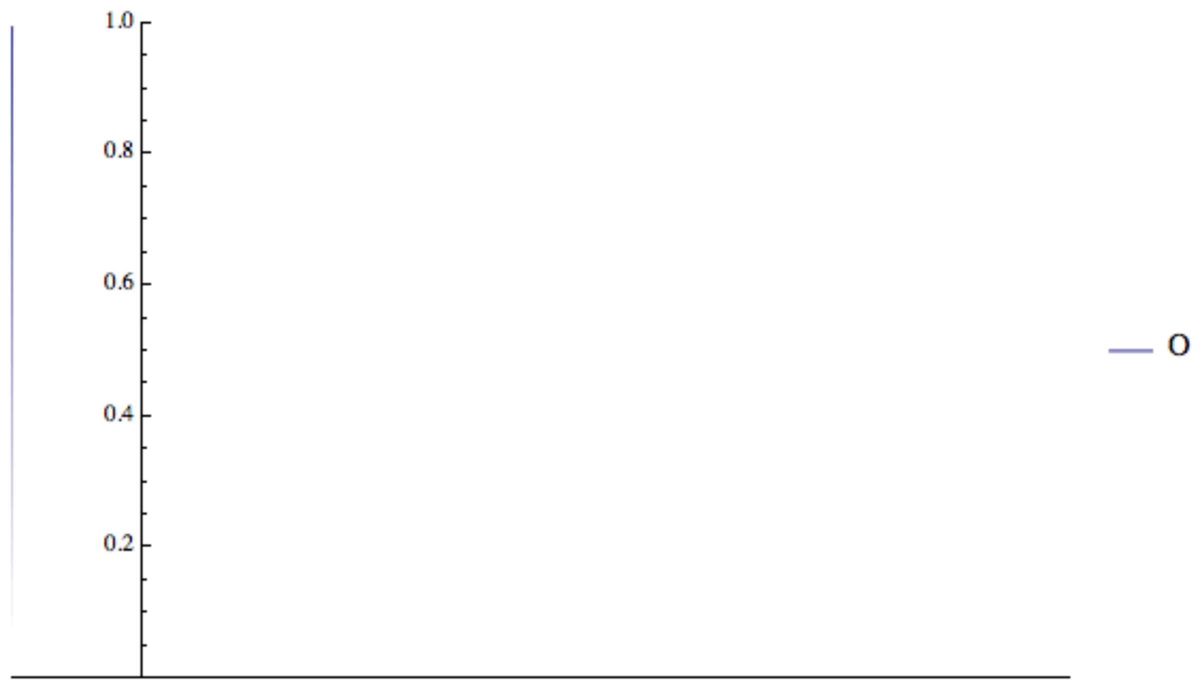




Significance and Power Reinterpreted

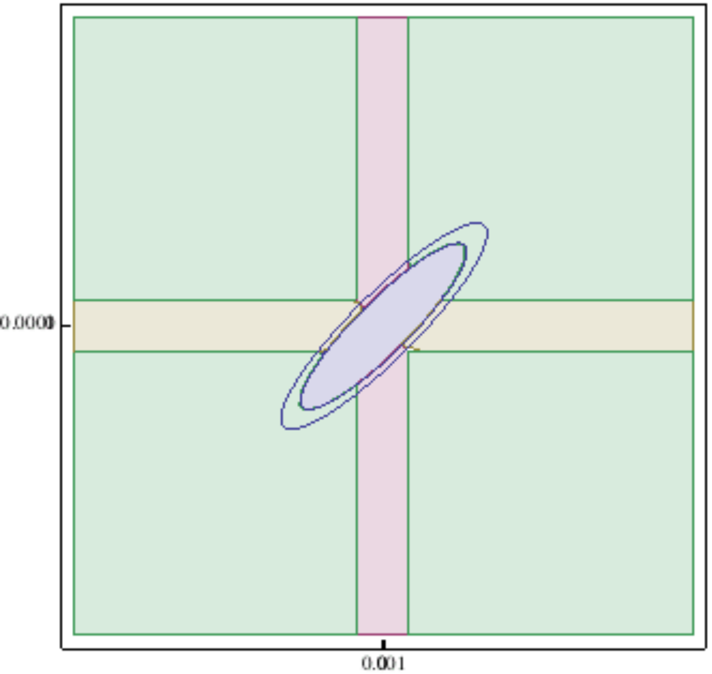
- “Significance” = tolerance on cycles and reversals in chance.
- “Power” = if you are destined to drop a model, get it over with a.s.a.p.

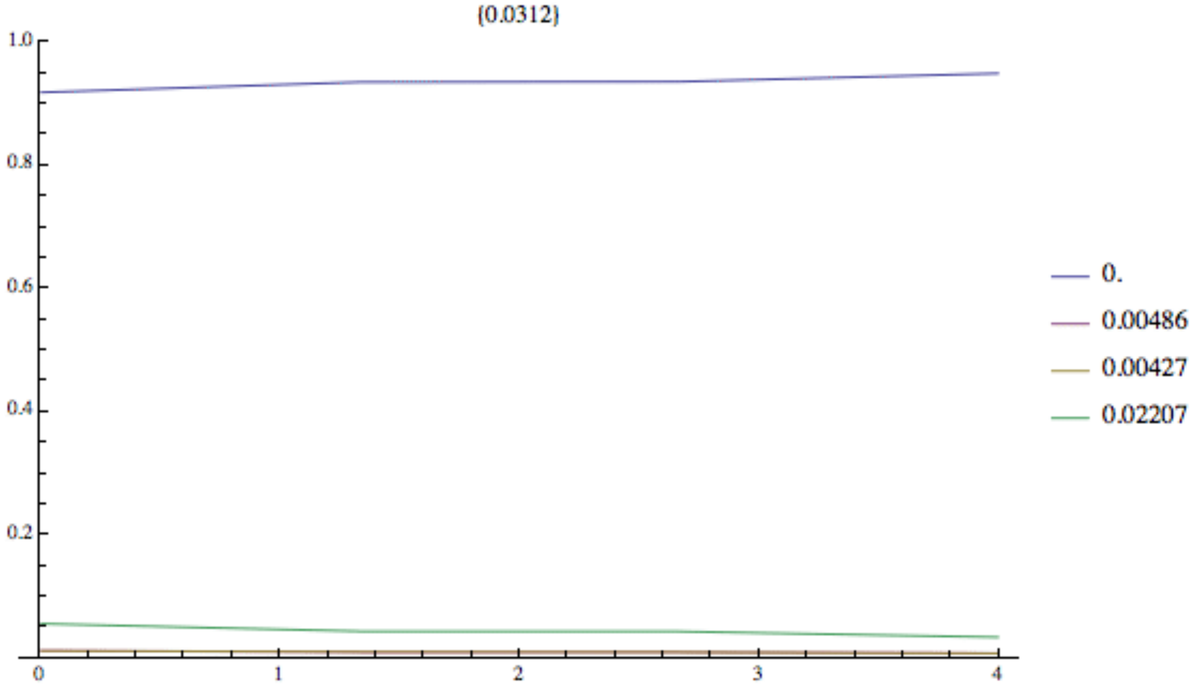




Simulations

- Method: Nested BIC





Thanks for *your* patience!