What is Statistical Deduction?

{Kevin T. Kelly, Konstantin Genin}
Carnegie Mellon University

June 2017

INDUCTIVE VS. DEDUCTIVE INFERENCE

Taxonomy of Inference

All the objects of human ... enquiry may naturally be divided into **two kinds**, to wit,

1. **Relations of Ideas**, and
2. **Matters of Fact**.

David Hume, *Enquiry*, Section IV, Part 1.

Taxonomy of Inference

- Any ... **inference** in science belongs to one of **two kinds**:
    1. either it yields **certainty** in the sense that the conclusion is **necessarily true**, provided that the premises are true,
    2. or it does not.
- The first kind is ... **deductive inference** ....
- The second kind will ... be called **'inductive inference'**.
- R. Carnap, *The Continuum of Inductive Methods*, 1952, p. 3 .

Taxonomy of Inference

- Explanatory arguments which ... account for a phenomenon by reference to **statistical** laws are not of the **strictly deductive** type.
- An account of this type will be called an ... **inductive** explanation.
- C. Hempel, “Aspects of Scientific Explanation”, 1965, p. 302.

Deductive Inference

Truth Preserving

- In each possible world:
  - if the premises are true,
  - then the conclusion is true.

Monotonic

- Conclusions are stable in light of further premises.

Logical Taxonomy of Inference

inference

deductive

truth preserving,
monotonic.

inductive

Everything else

Logical Taxonomy of Inference

inference

deductive

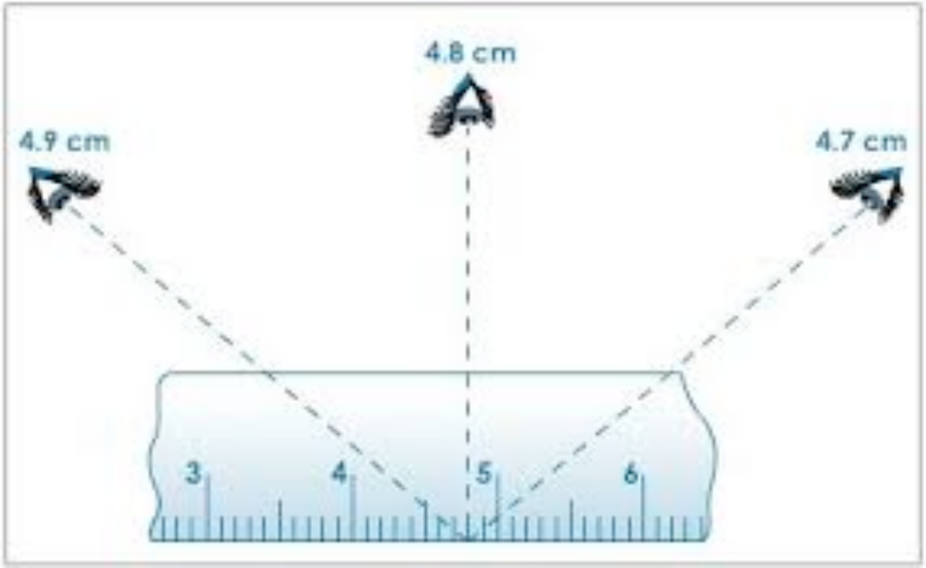
- Calculation
- Refuting universal *H*
- Verifying existential *H*
- Deciding between universal *H*, *H'*
- Predicting *E* from *H*
- Hypotheses compatible with *E*

inductive

- Inferring universal *H*
- Choosing between universal H_0 , H_1 , H_2 , ...

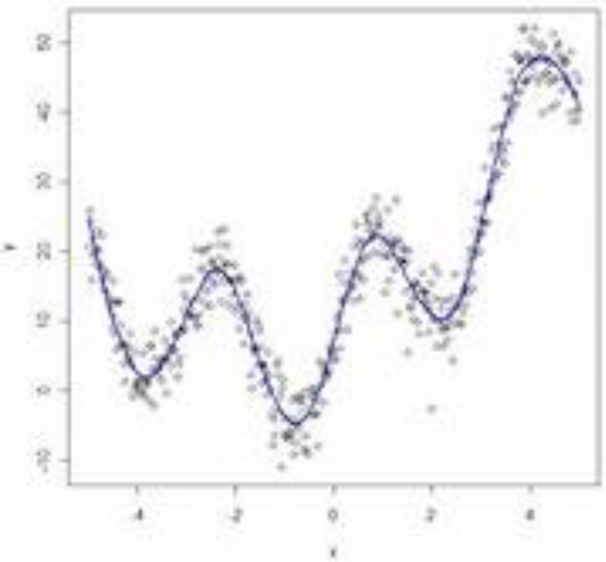
Real Data

- All real measurements are subject to probable error.
  - It can be reduced by averaging repeated samples.



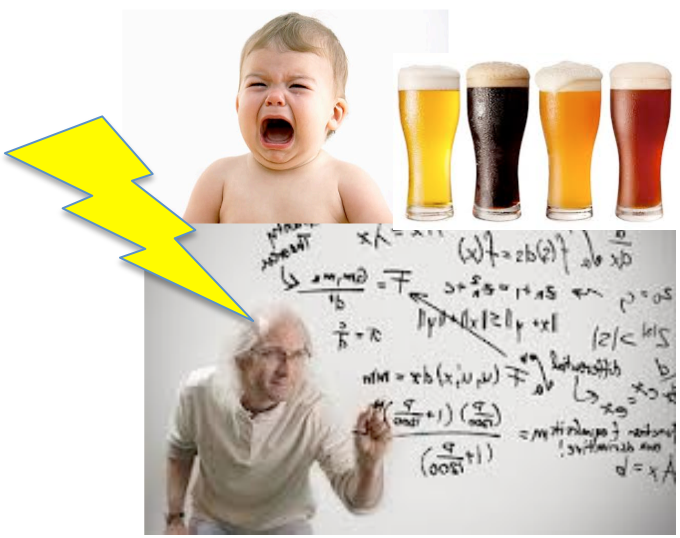
Real Predictions

- All real predictions are subject to probable error.
  - It can be reduced by predicting averages of repeated samples.



Real Calculations

- Even all real calculations are subject to probable error.
  - It can be reduced by comparing repeated calculations.



Real Deductive Inference

Truth preserving in chance

- In each possible world:
  - if the premises are true,
  - then the chance of drawing an erroneous conclusion is low.

Monotonic in chance

- The chance of producing a conclusion is guaranteed not to drop by much.

Taxonomies Can be **Bad**

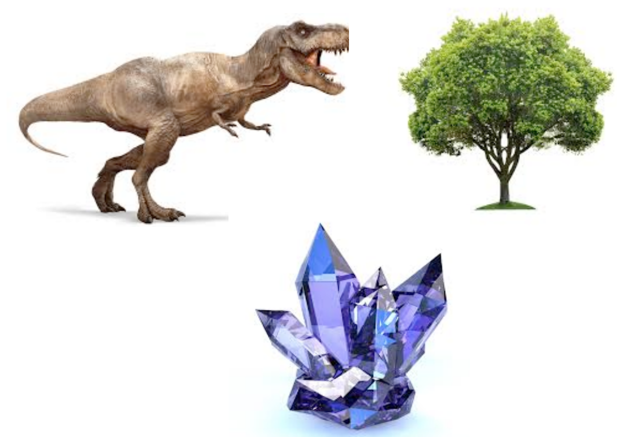
things

everything else

white roses

non-white roses

everything else



Traditional Taxonomy of Inference

inference

inductive

logically deductive

statistically deductive

everything else

Missed Opportunities for Philosophy

inference

inductive

logically deductive

1. Ideal calculation
2. Refuting universal H_0
3. Verifying existential H_1
4. Deciding between universal H_0, H_1
5. Predicting E from H
6. Hypotheses compatible with E

statistically deductive

1. Real calculation
2. Refuting point null H_0
3. Verifying composite H_1
4. Deciding between point hypotheses H_0, H_1
5. Direct inference of E from H
6. Non-rejection.

everything else

1. Inferring universal H_0
2. Choosing between universal H_0, H_1, H_1, ...

1. Inferring simple H_0
2. Model selection

Better Taxonomy of Inference

inference

deductive

logically

1. Refuting universal H_0
2. Verifying existential H_1
3. Deciding between universal H_0, H_1
4. Predicting E from H
5. Compatibility with E
6. Ideal calculation

statistically

1. Refuting point null H_0
2. Verifying composite H_1
3. Deciding between point hypotheses H_0, H_1
4. Direct inference of E from H
5. Non-rejection.
6. Real calculation

inductive

logically

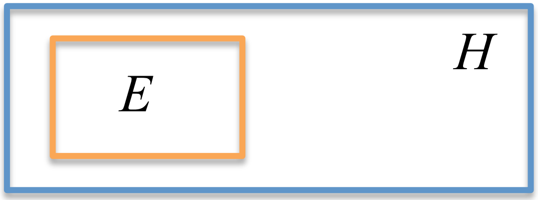
1. Inferring universal H_0
2. Choosing between universal H_0, H_1, H_1, ...

statistically

1. Inferring simple H_0
2. Model selection

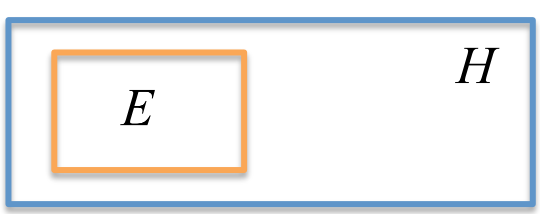
Main Objection

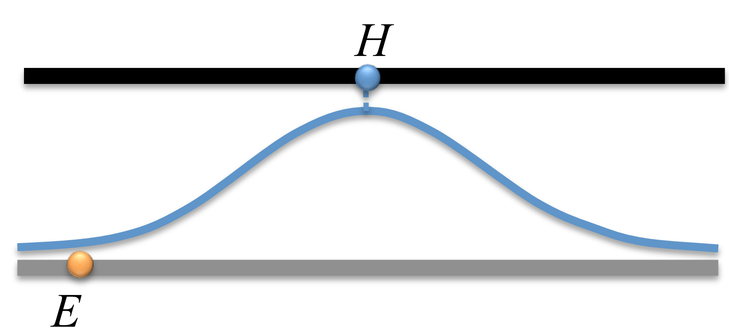
- In logical deduction, the evidence definitely rules out possibilities.



Main Objection

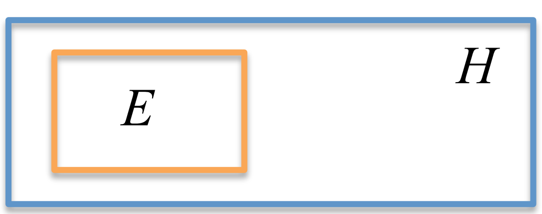
- In logical deduction, the evidence logically rules out possibilities.
- In statistical deduction, the sample is logically compatible with every possibility.

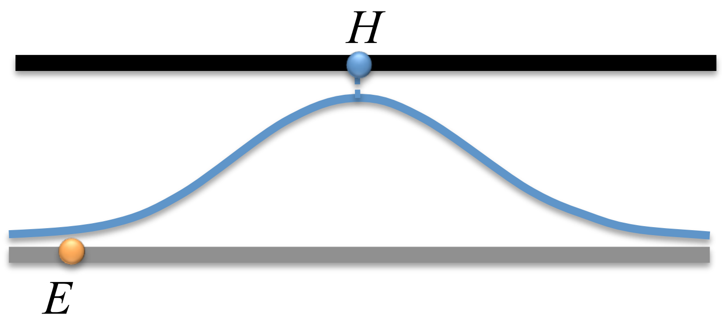




Main Objection

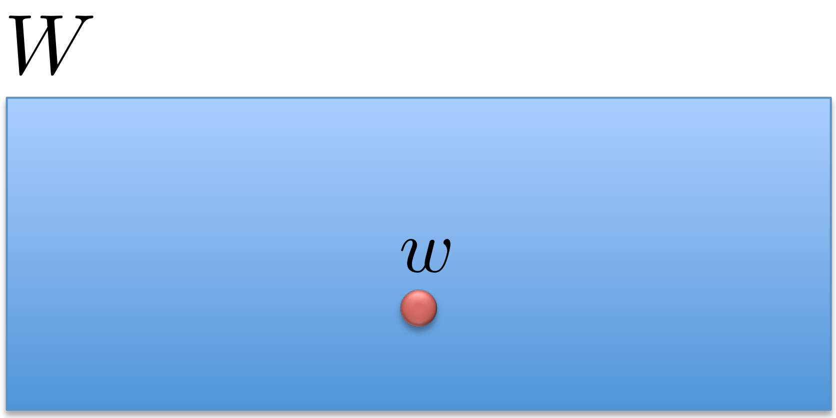
- In logical deduction, the evidence logically rules out possibilities.
- In statistical deduction, the sample is logically compatible with every possibility.
- The situations are not even similar.





THE LOGICAL SETTING

Possible Worlds

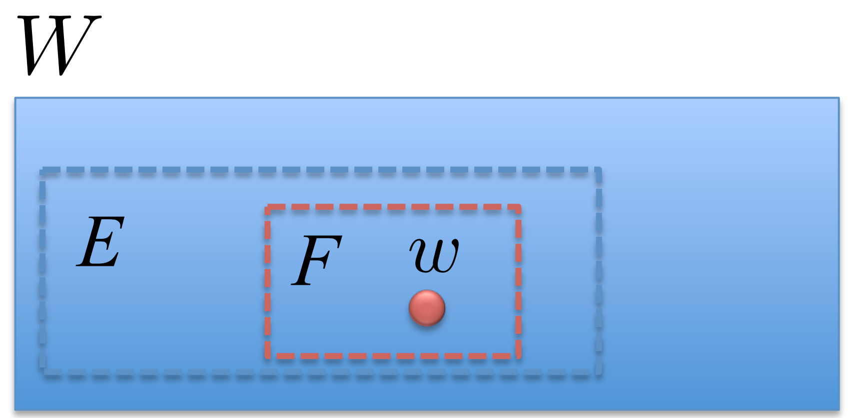


Propositional Information State

The logically strongest proposition you are informed of.

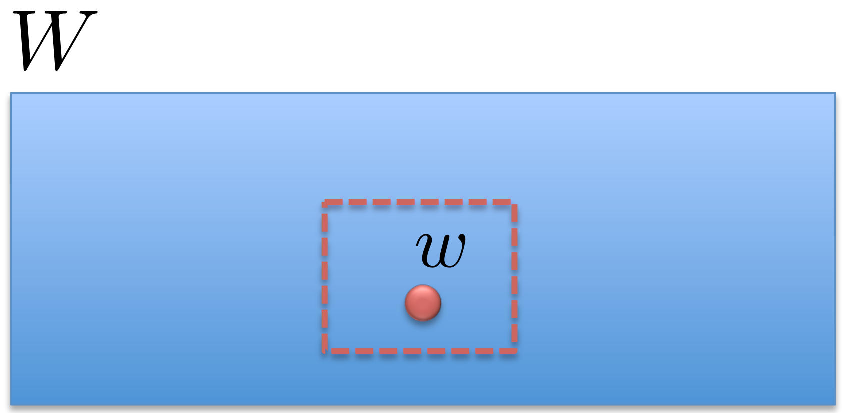
The Situation We are Modeling

In world w, a diligent inquirer eventually obtains true information F that deductively entails arbitrary information state E true in w.



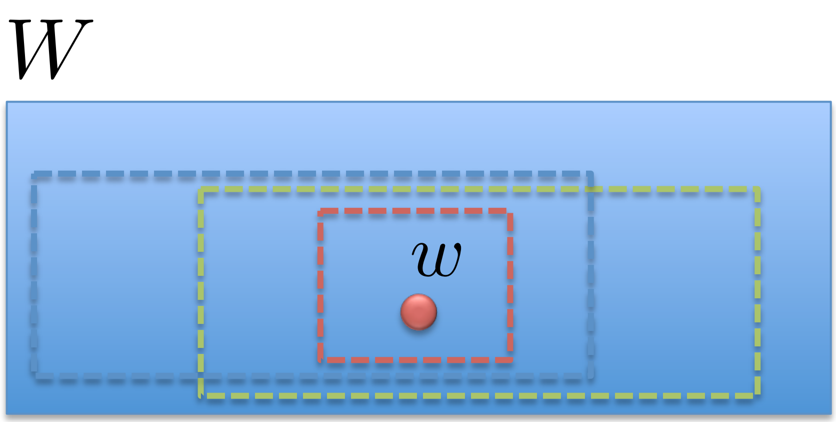
Three Axioms

1. Some information state true in w.



Three Axioms

1. Some information state true in w.
2. Each pair of information states true in w is entailed by a true information state true in w.

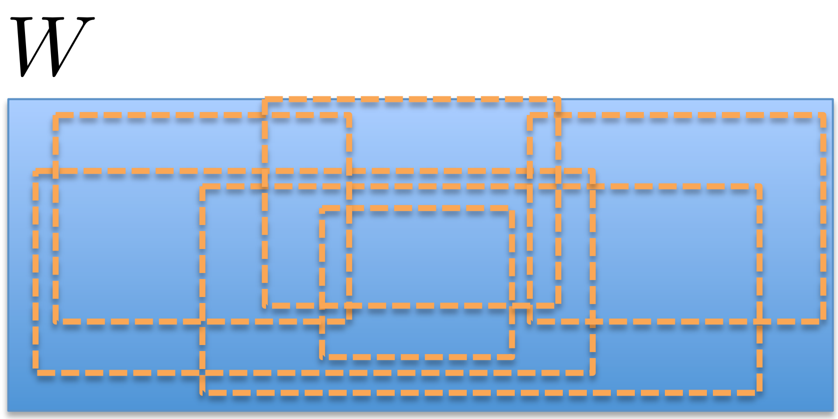


Three Axioms

1. Some information state true in w.
2. Each pair of information states true in w is entailed by a true information state true in w.
3. There are at most countably many information states.

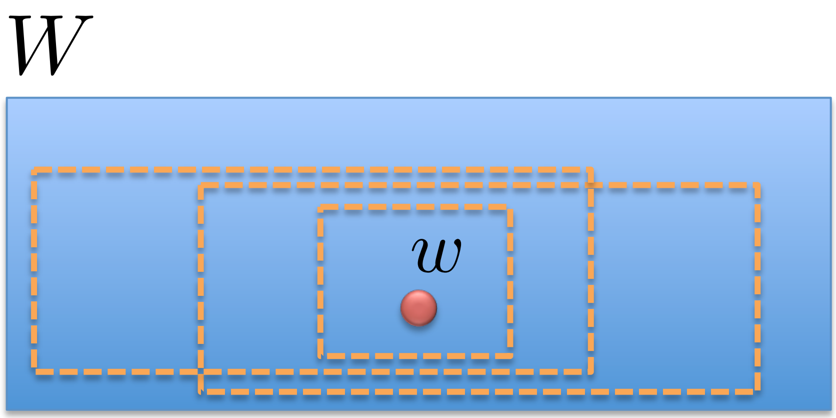
Information States

$\mathcal{I}$ = the set of all information states.



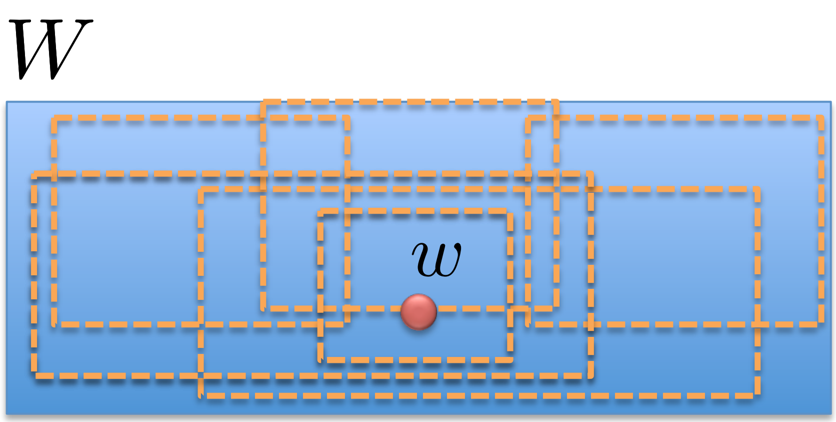
Information States

$$\mathcal{I} = \text{the set of all information states.}$$
$$\mathcal{I}(w) = \text{the set of all information states true in } w.$$



The Topology of Information

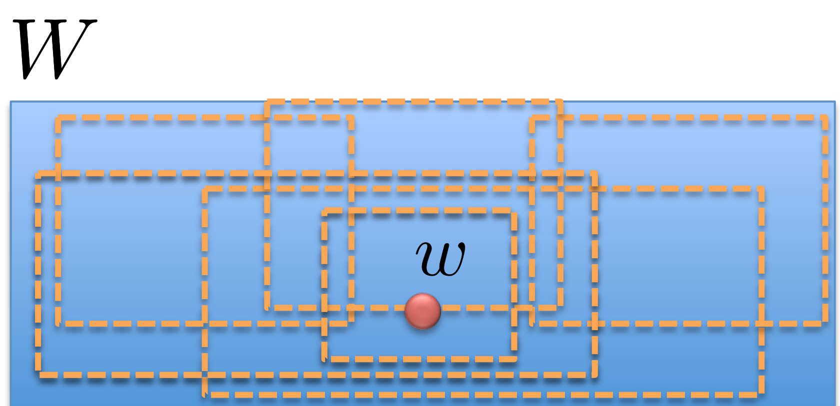
- $\mathcal{I}$ is a **topological basis** on W.
- Closing $\mathcal{I}$ under infinite disjunction yields a **topologial space** on W.



The Topology of Information

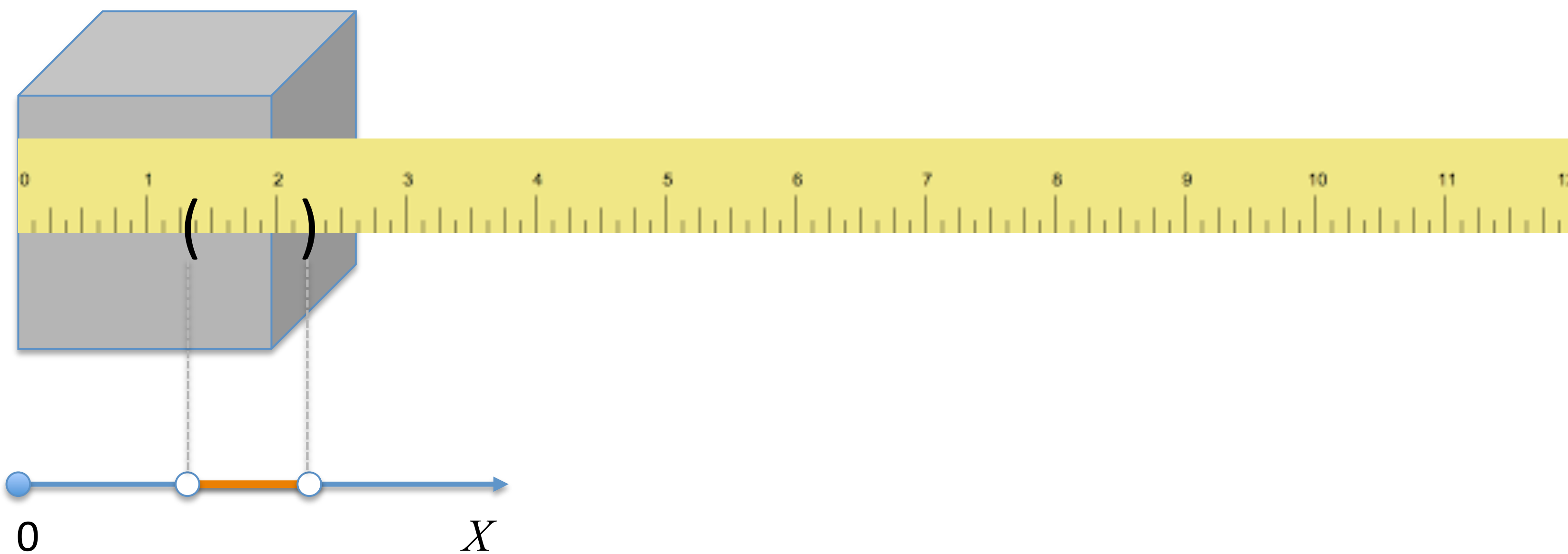
- $\mathcal{I}$ is a **topological basis** on W.
- Closing $\mathcal{I}$ under infinite disjunction yields a **topological space** on W.

Topological structure isn't imposed; it is already there.



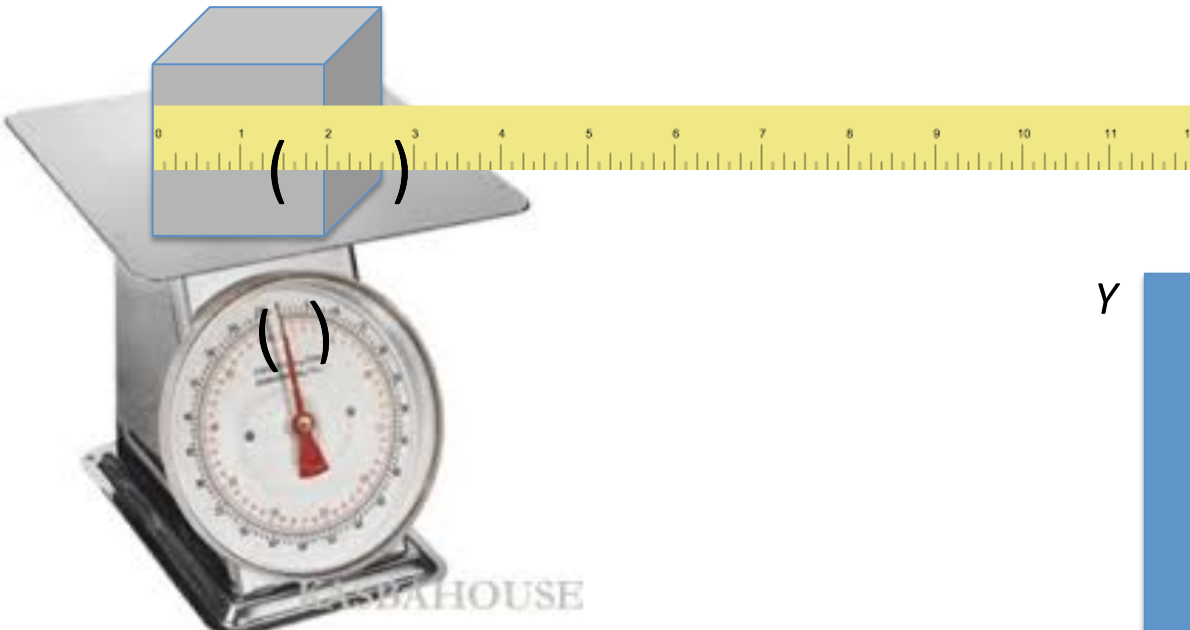
Example: Measurement of X

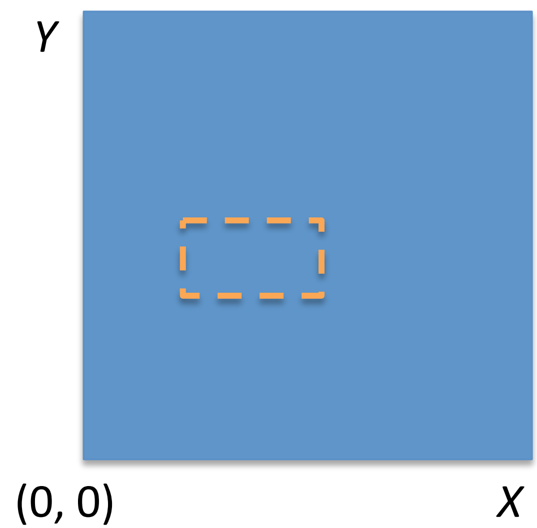
- **Worlds** = real numbers.
- **Information states** = open intervals.



Example: Joint Measurement

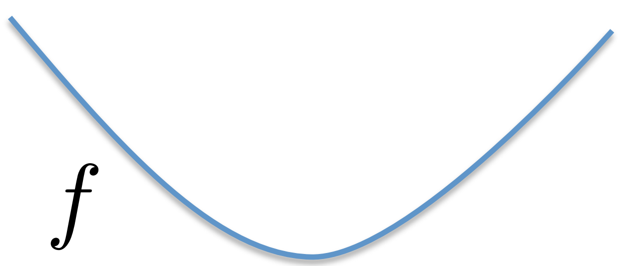
- **Worlds** = points in real plane.
- **Information states** = open rectangles.





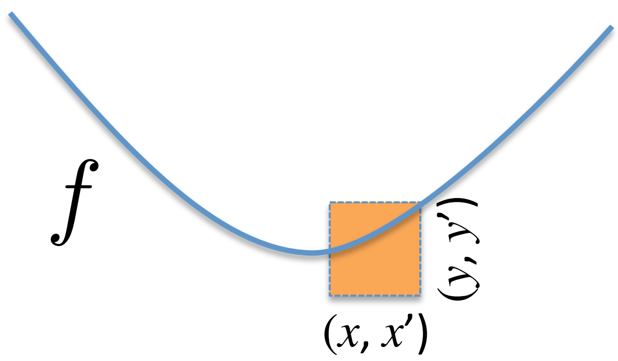
Example: Equations

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



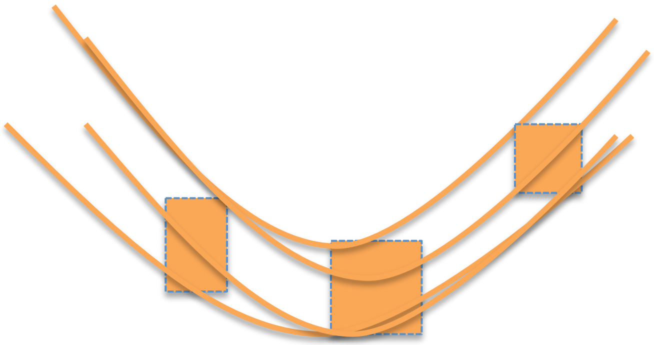
Example: Laws

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



Example: Laws

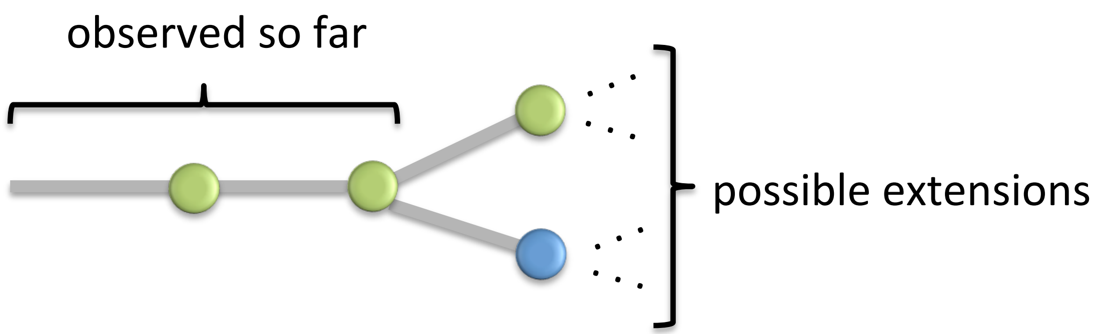
- The **information state** is the set of all worlds that touch each observation.



Example: Sequential Binary Experiment

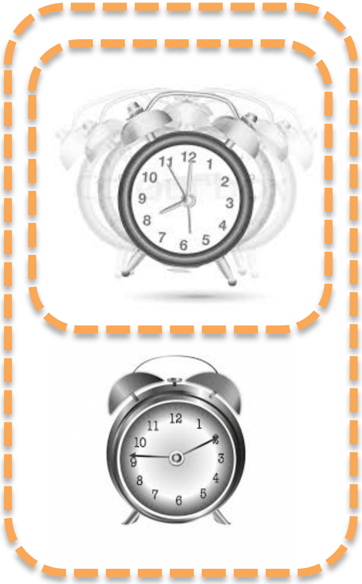
World = infinite discrete sequence of outcomes.

Information state = all extensions of a finite outcome sequence:



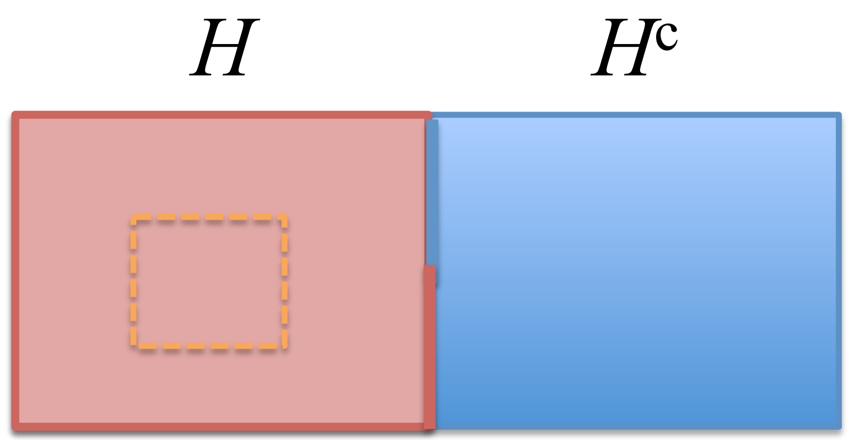
The Sleeping Scientist

- The theorist is awakened by her graduate students only when her theory is refuted.



Deductive Verification and Refutation

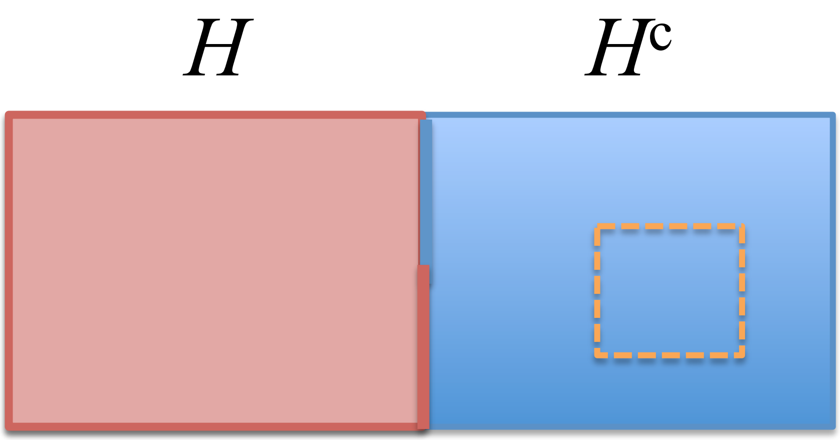
H is **verified** by E iff $E \subseteq H$.



Deductive Verification and Refutation

H is **verified** by E iff $E \subseteq H$.

H is **refuted** by E iff $E \subseteq H^c$.

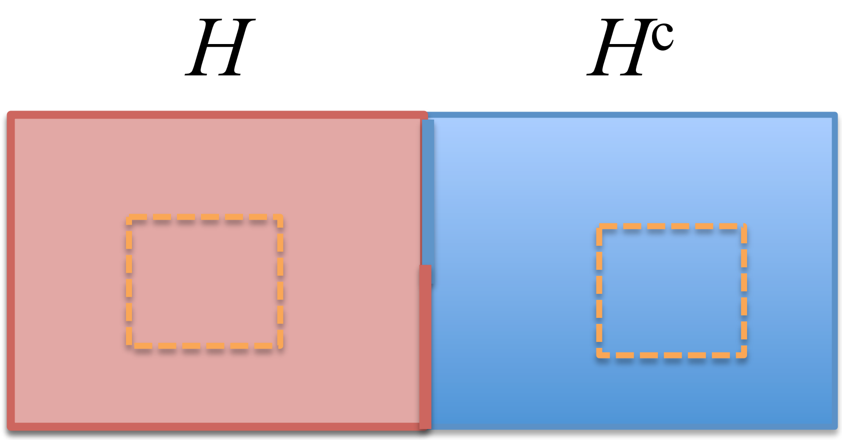


Deductive Verification and Refutation

H is **verified** by E iff $E \subseteq H$.

H is **refuted** by E iff $E \subseteq H^c$.

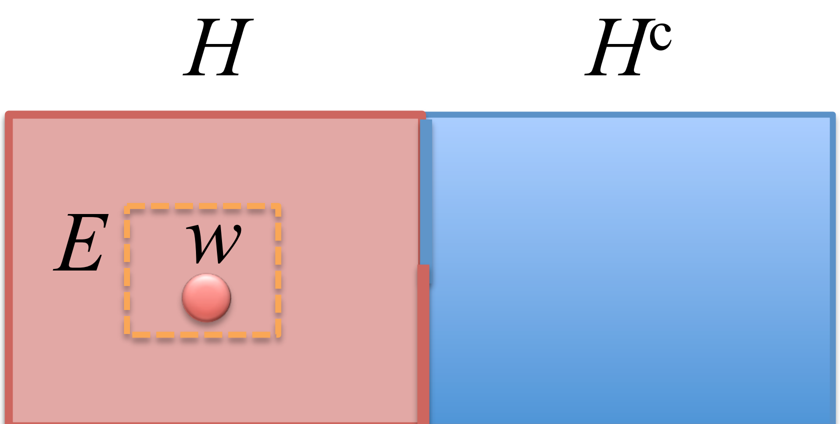
H is **decided** by E iff H is either verified or refuted by E.



H Will be Verified in w

w is an **interior point** of H iff

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



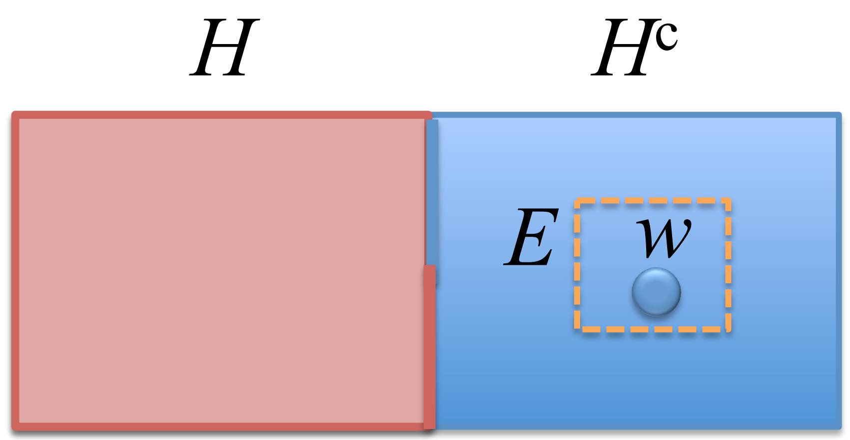
H Will be Refuted in w

w is an **interior point** of H iff

iff H **will be** verified in w

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

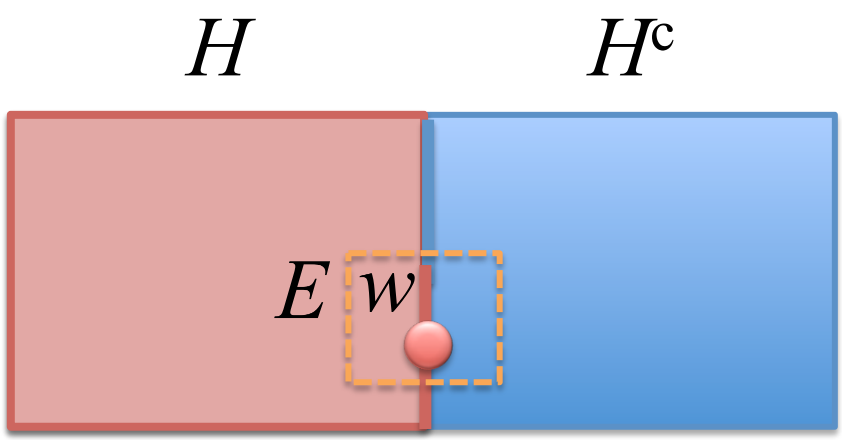
w is an **exterior point** of H iff w is an interior point of H^c.



Popper’s Problem of Metaphysics in w

w is a **frontier point** of H iff

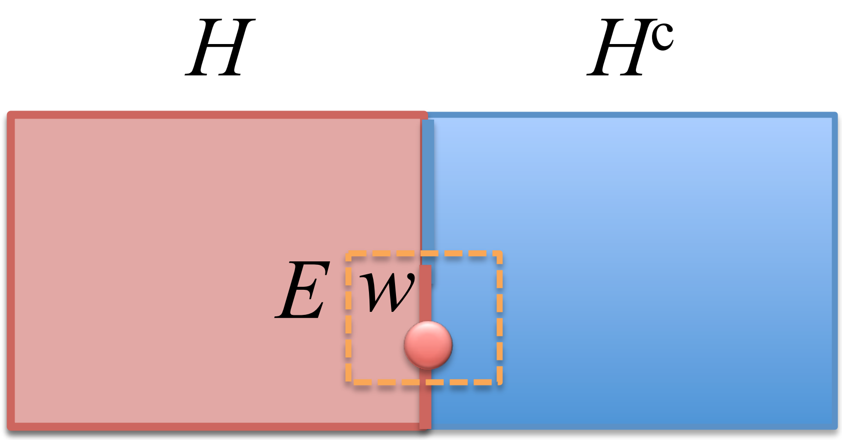
- H is false in w but will never be refutedin w.



Hume's Problem of Induction in w

w is a **frontier point** of H^c iff

- H is true in w but will never be verified in w.



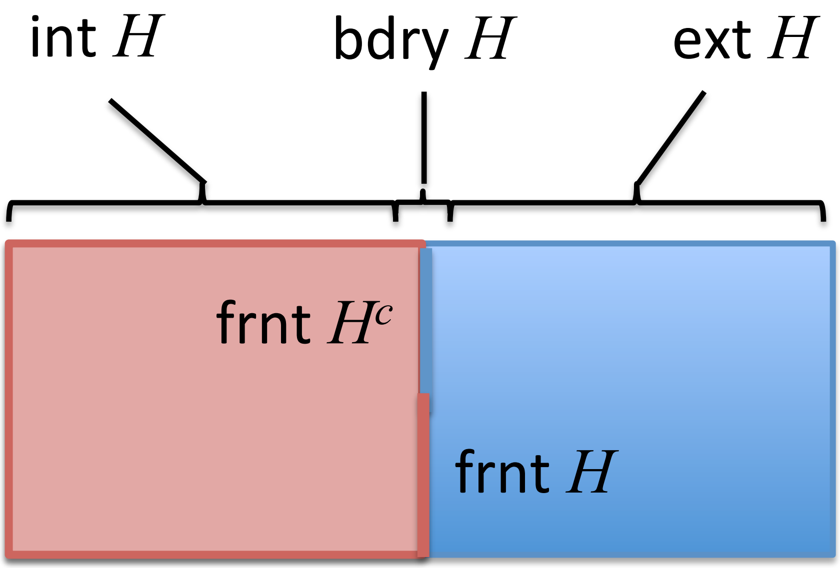
Topological Operations as Modal Operators

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

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

frnt H := the proposition that H is **false** but **will never be refuted.**

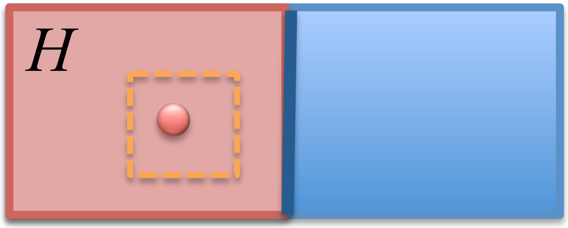
frnt H^c := the proposition that H is **true** but **will never be verified.**



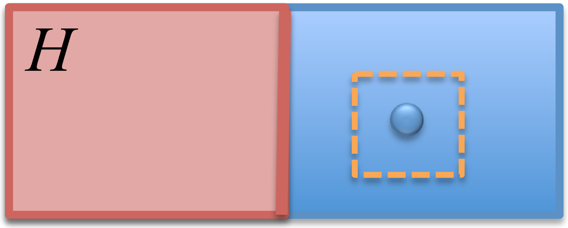
Verifiability, Refutability, Decidability

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

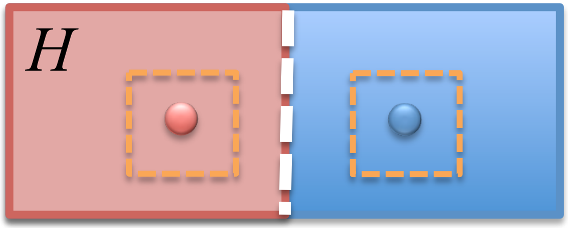
i.e., iff H will be verified however H is true.



H is **closed (refutable)** iff H^c is open.

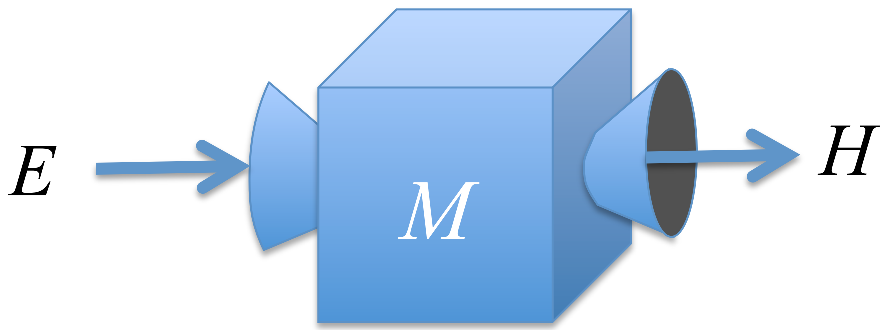


H is **clopen (decidable)** iff H is both open and closed.



Propositional Methods

- **Propositional methods** produce propositional conclusions in response to propositional information.



Deductive Success

- A **verification method** for H is an method M such that in every world w:
  1. $w \in H$: M converges infallibly to H;
  2. $w \in H^c$: V always concludes W.

Deductive Success

- A **verification method** for H is an method M such that in every world w:
  1. $w \in H$: M converges to H and never concludes H^c;
  2. $w \in H^c$: V always concludes W.
- A **refutation method** for H is just a verification method for H^c.

Deductive Success

- A **verification method** for H is an method M such that in every world w:
  1. $w \in H$: M converges to H and never concludes H^c;
  2. $w \in H^c$: V always concludes W.
- A **refutation method** for H is just a verification method for H^c.
- A **decision method** for H converges to H or to H^c without error.

Deductive Success

Proposition.

If M is a verifier, refuter, or decider for H,

then M produces only conclusions that are deductively entailed by the given information.

The Topology of Deductive Success

Proposition. H has a verifier, refuter, or decider iff H is open, closed, or clopen.

Inductive Success

- A **limiting verification method** for H is a method M such that in every world w:

$w \in H$ iff M converges to some true H' that entails H.

Inductive Success

- A **limiting verification method** for H is a method M such that in every world w:

  $w \in H$ iff M converges to some true H' that entails H.

- A **limiting refutation method** for H is a limiting verification method for H^c.

Inductive Success

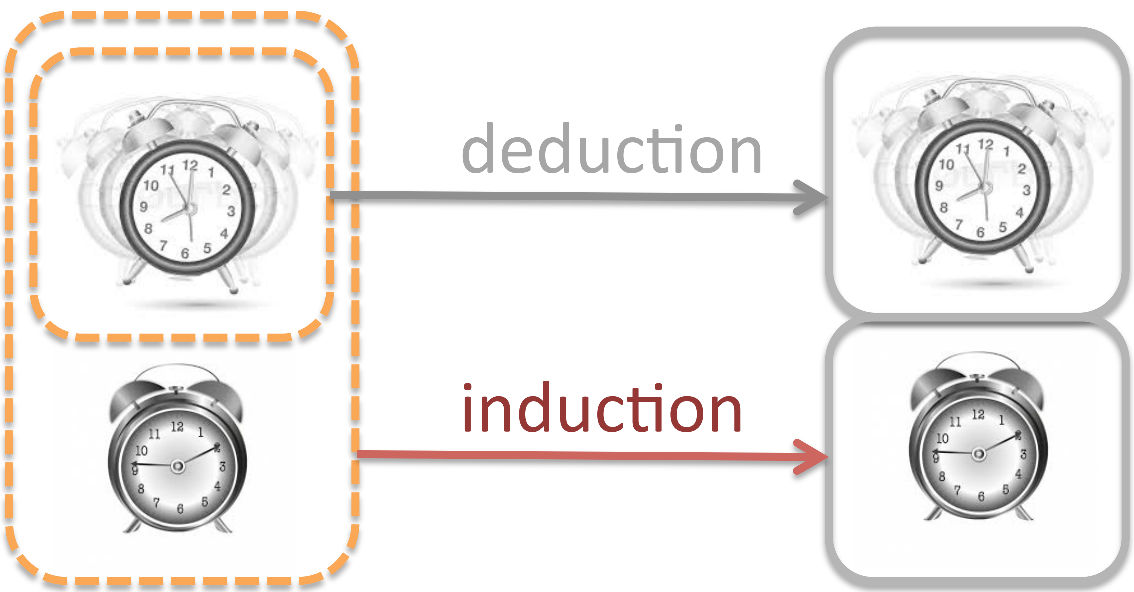
- A **limiting verification method** for H is a method M such that in every world w:

  $w \in H$ iff M converges to some true H' that entails H.

- A **limiting refutation method** for H is a limiting verification method for H^c.
- A **limiting decision method** for H is a limiting verification method and a limiting refutation for H.

Inductive Success

Proposition. No limiting verifier of "never awakened" is deductive.



Scientific Models

H is **locally closed** iff H can be expressed as a difference of open (verifiable) propositions.

Thesis: Scientific models are locally closed propositions.

Topology

Let $\mathcal{I}^*$ denote the closure of $\mathcal{I}$ under union.

Proposition:

If $(W, \mathcal{I})$ is an information basis

then $(W, \mathcal{I}^*)$ is a topological space.

Topology

- H is **open** iff $H \in \mathcal{I}^*$.
- H is **closed** iff H^{c} is open.
- H is **clopen** iff H is both closed and open.

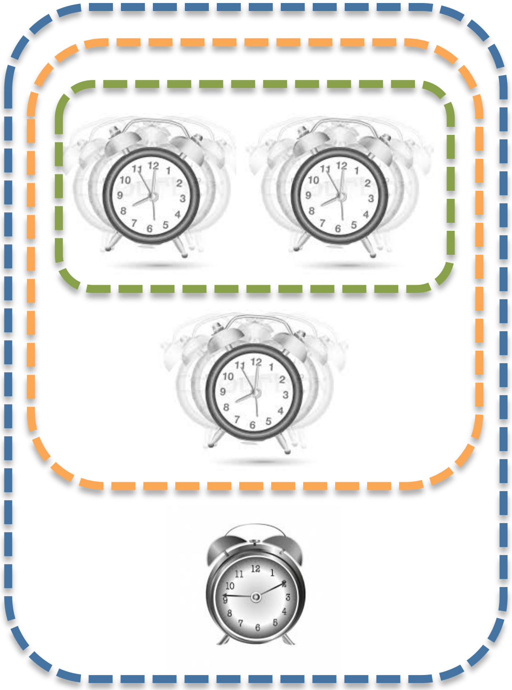
- H is **locally closed** iff H is a difference of open sets.

Sleeping Theorist Example

H_2 = "Awakened twice" is open.

H_1 = "Awakened once" is locally closed.

H_0 = "Never awakened" is closed.

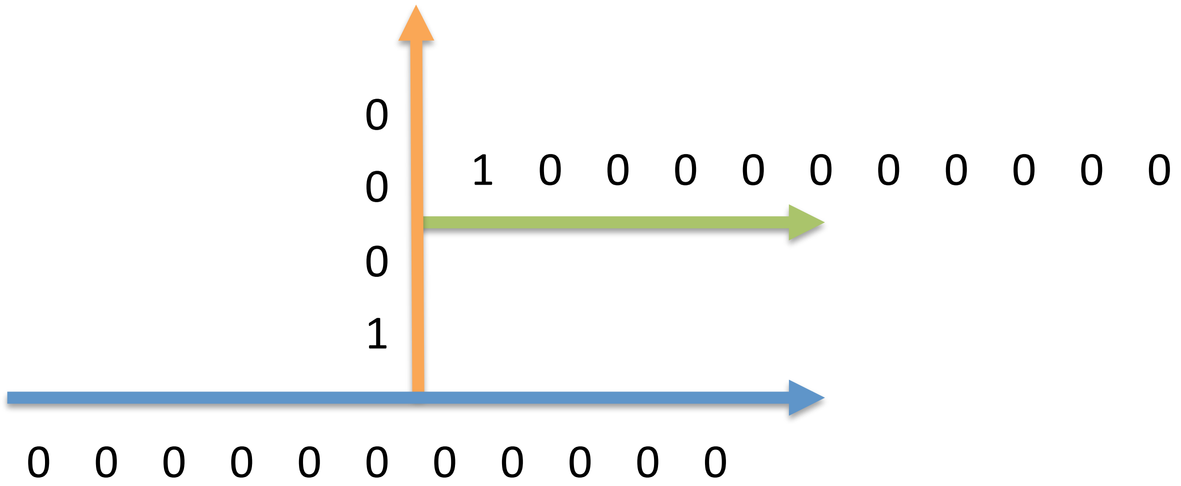


Sequential Example

H_2 = "You will see 1 exactly twice" is open.

H_1 = "You will see 1 exactly once" is locally closed.

H_0 = "You will never see 1" is closed.

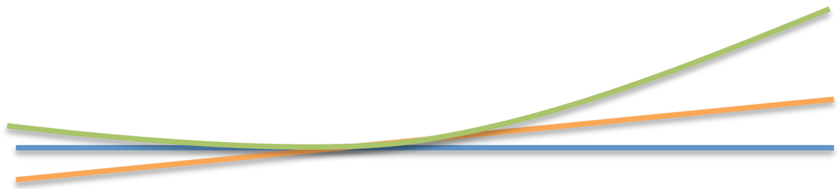


Equation Example

H_2 = "quadratic" is open.

H_1 = "linear" is locally closed.

H_0 = "constant" is closed.



Scientific Theories and Paradigms

H is **limiting open** iff H can be expressed as a countable union of locally closed propositions.

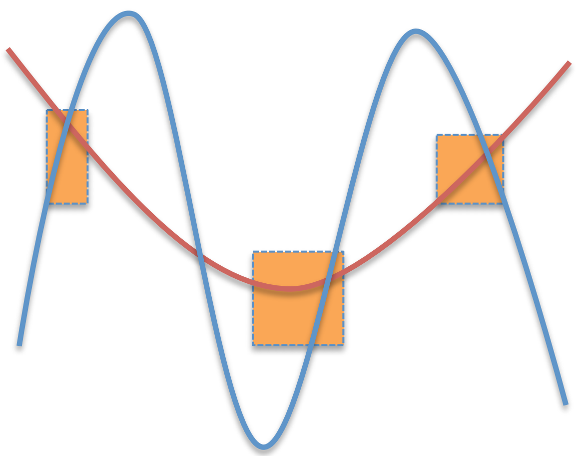
Theses:

1. Scientific theories are limiting open.
2. Each locally closed disjunct of a theory is a possible **articulation** of the theory.
3. **Duhem's problem:** a theory in trouble can always be re-articulated to accommodate the data.

Equation Example

H_0 = the true law is polynomial.

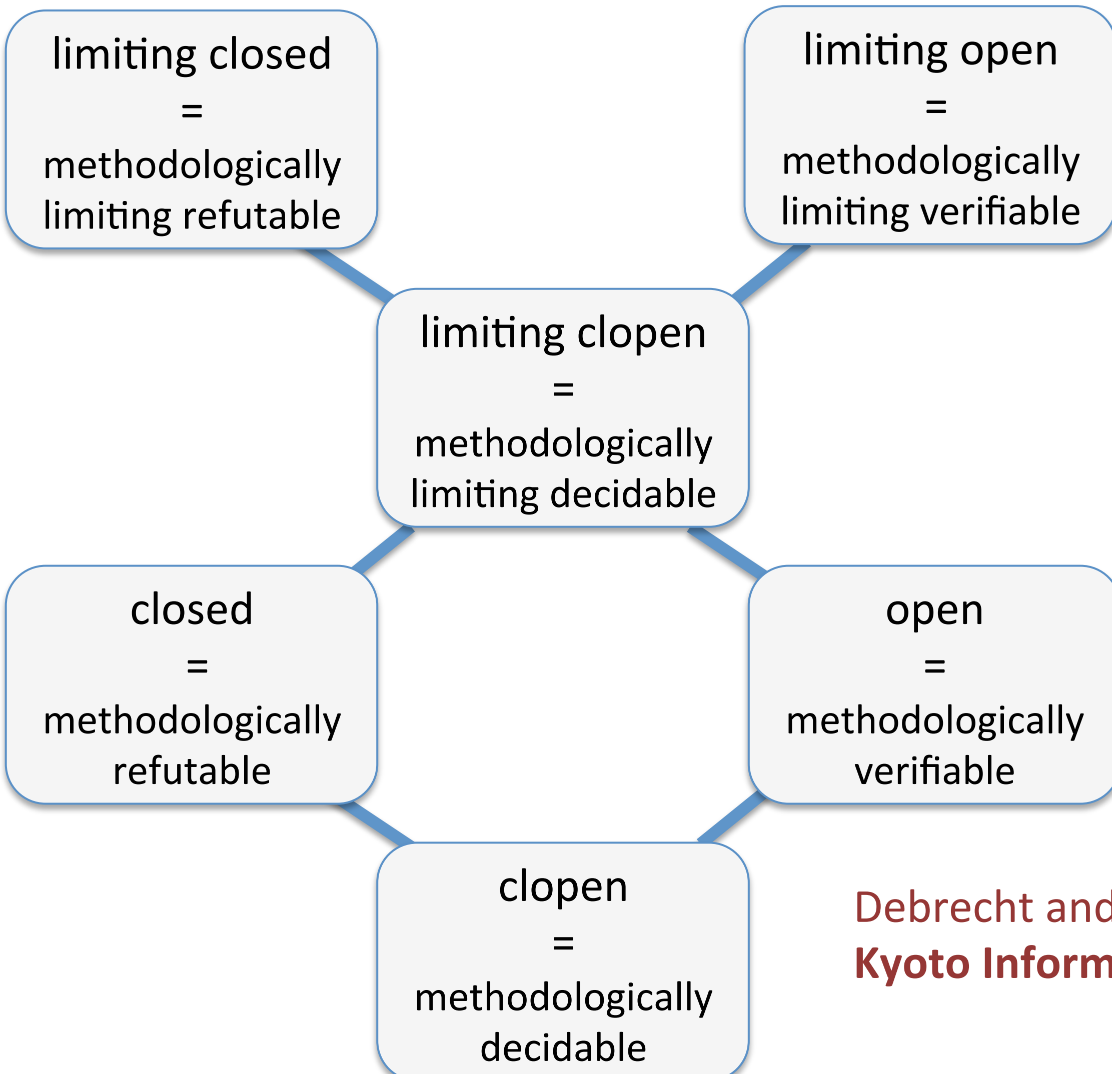
H_1 = the true law is a trigonometric polynomial.



Topology

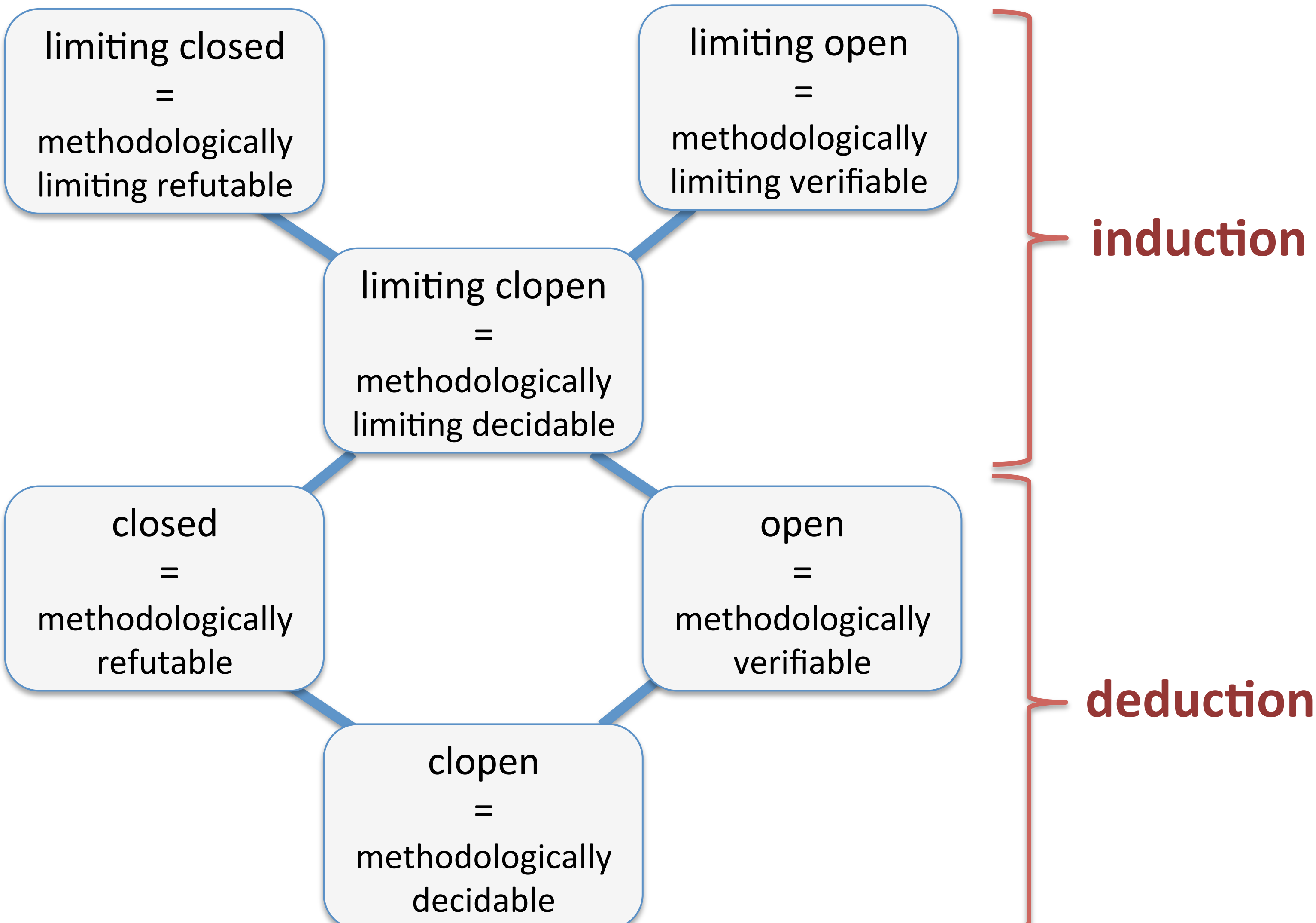
- H is **limiting open** iff H is a countable union of locally closed sets.
- H is **limiting closed** iff H^c is limiting open.
- H is **limiting clopen** iff H is both limiting open and limiting closed.

Theorem.



Debrecht and Yamamoto,
Kyoto Informatics

Theorem



limiting closed = methodologically limiting refutable
limiting open = methodologically limiting verifiable
limiting clopen = methodologically limiting decidable
closed = methodologically refutable
open = methodologically verifiable
clopen = methodologically decidable
induction
deduction

THE STATISTICAL SETTING

Can We Do the Same for Statistics?

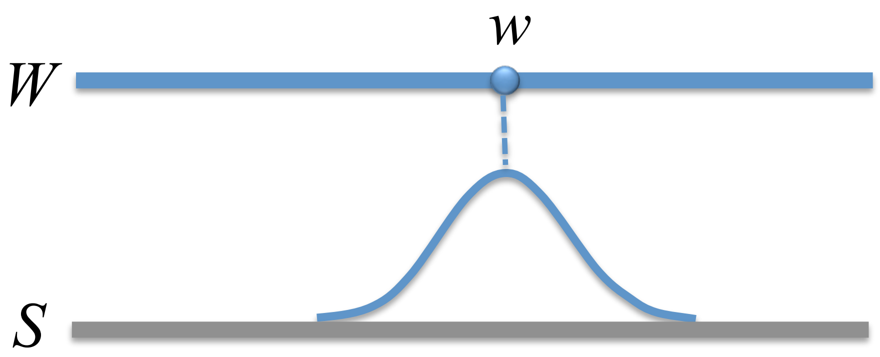
Kelly's topological approach...

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

Eliott Sober, *Ockham's Razors*, 2015

Statistics

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



Statistical Verification

- A **statistical verification method** for H at **significance level** $\alpha > 0$:
  1. converges in probability to conclusion H, if H is true.
  2. always concludes W with probability at least $1\text{-}\alpha$, if H is false.

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

Methods

- A **statistical verification method** for H at level $\alpha > 0$ is a sequence (M_n) of feasible tests of H^c such that for every world w and sample size *n*:

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

  for $\alpha_n \rightarrow 0$, and dominated by α.

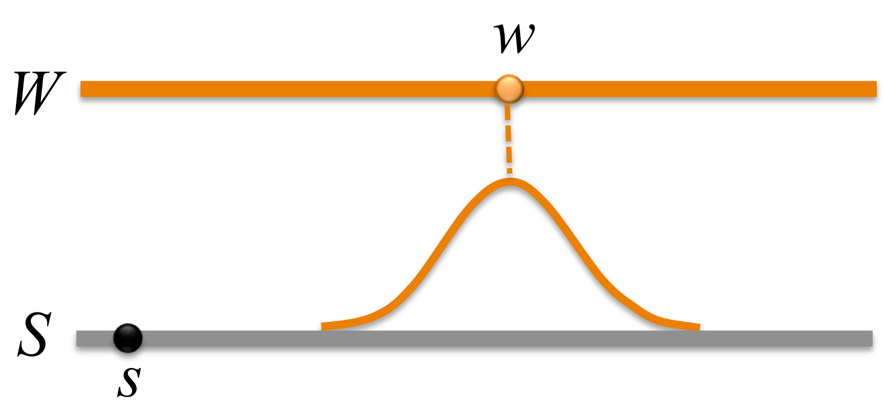
Statistical Verifiction in the Limit

- A **limiting statistical α-verification method** for H
  1. produces only conclusions H or W
  2. converges in probability to H iff H is true.

- H is **statistically verifiable in the limit** iff H has a limiting statistical α-verification method, for each $\alpha > 0$.

Recall the Fundamental Difficulty

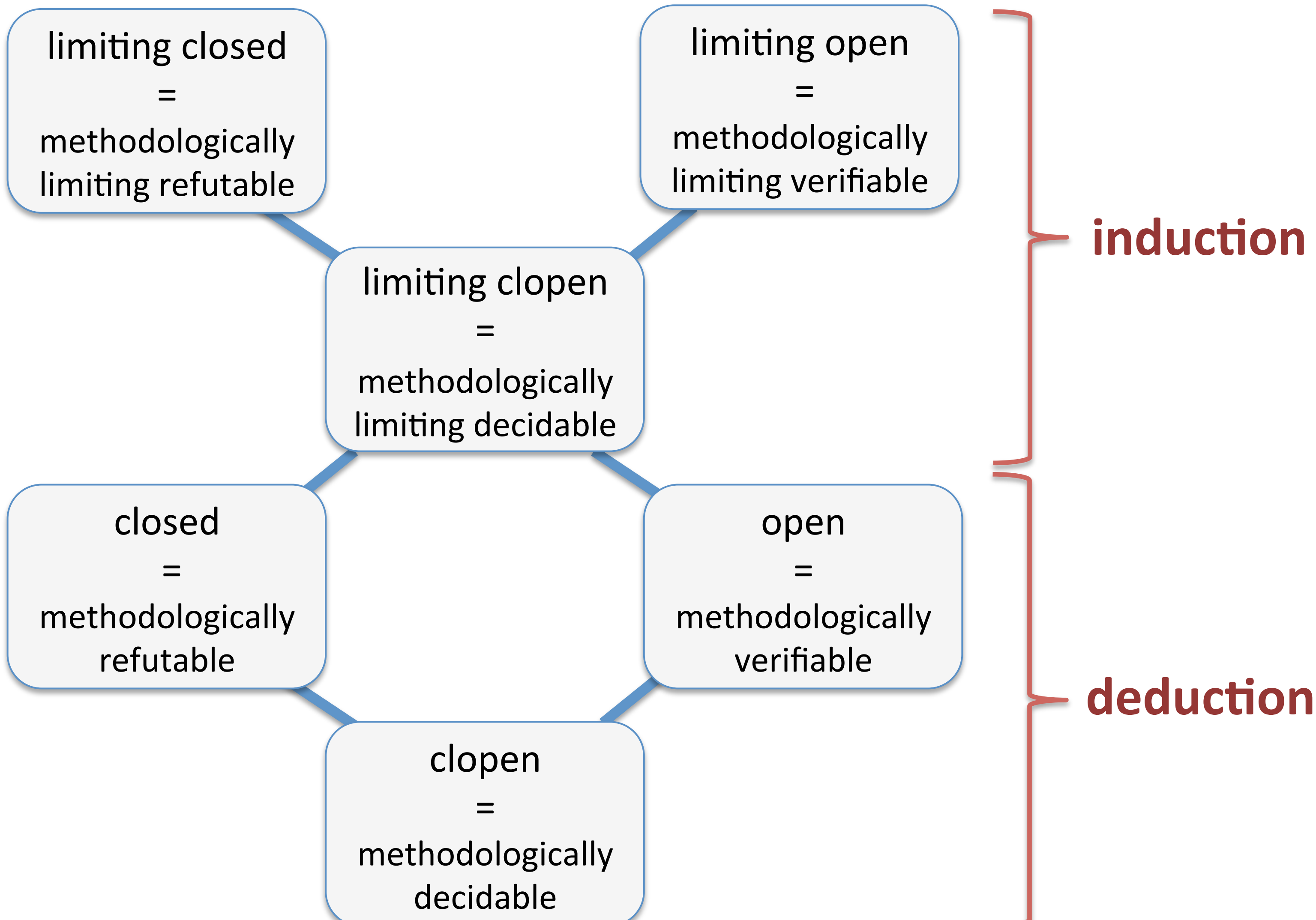
- Every sample is logically consistent with all worlds!
- So it seems that statistical information states are all trivial!



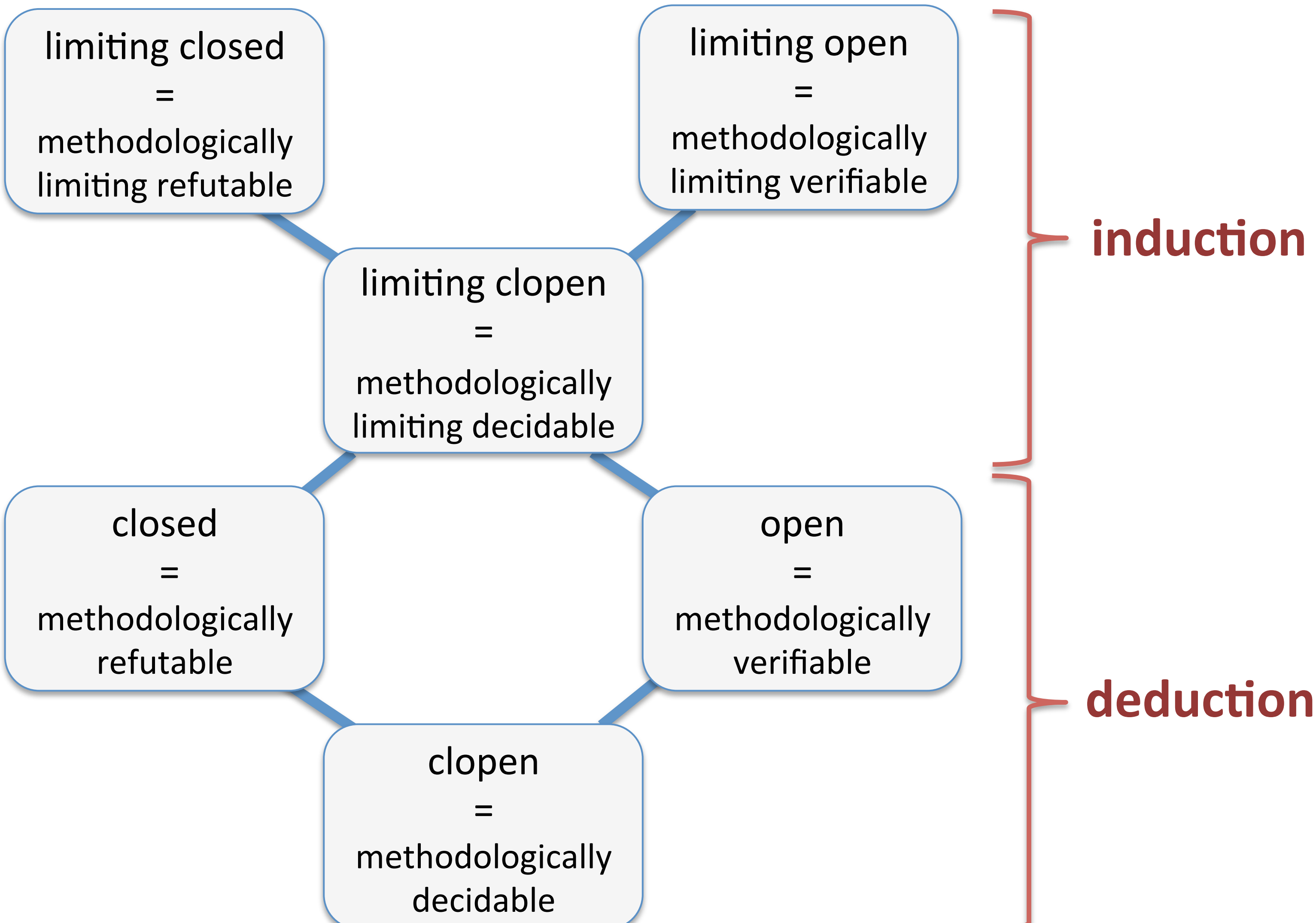
The Main Result

- Under mild and natural assumptions...
- there exists a unique and familiar topology on probability measures for which...

The Main Result



So in Both Logic and Statistics:

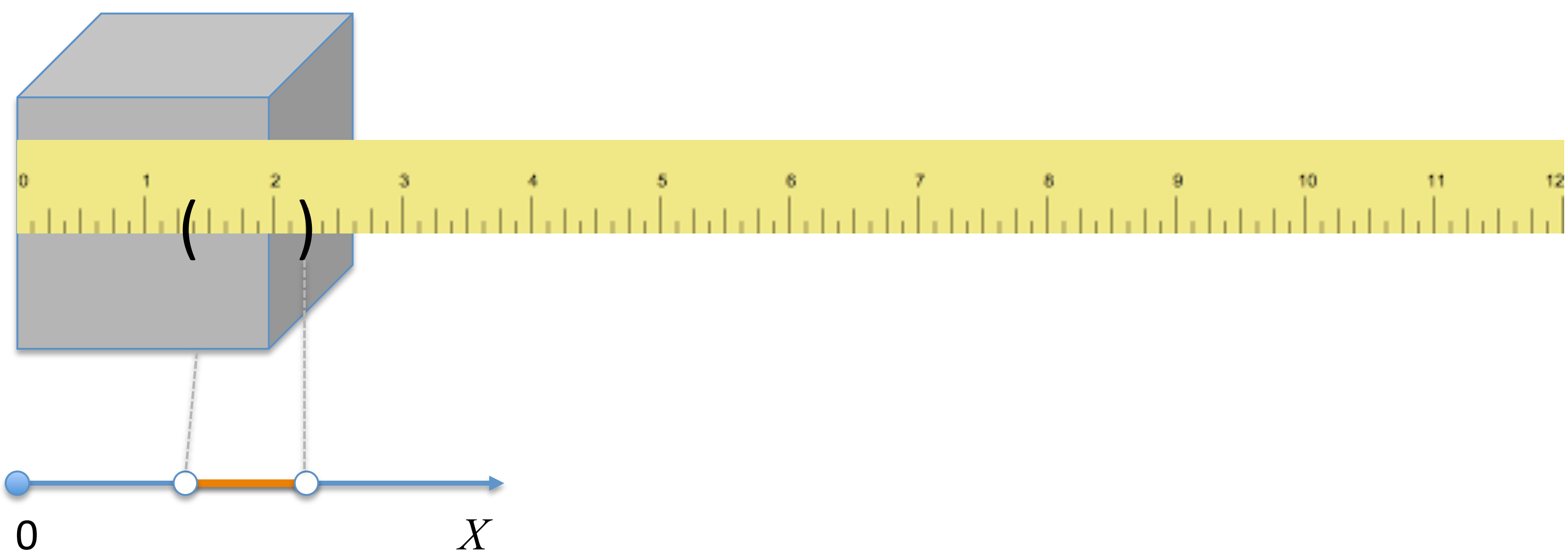


From Logic to Statistics

- Start with purely (topo)logical insights about scientific methodology.
- Transfer them to statistics via the preceding result.

The Key Idea

- Even with arbitrarily powerful magnification, it is infeasible to verify that a given cube is **exactly** 2 inches wide.



The Key Idea

- Similarly, it is awkward to say that a given attempt at measuring length yields **exactly** a given value.
- More decimal places of expansion might violate **exact** identity at any stage of approximation:
  - 2.3578000000000000000000000000**1.**

The Key Idea

- So if there were a non-zero chance of a sample hitting **exactly** on the boundary of the acceptance zone of a statistical test...
- one would have a non-zero chance of **implementing** the test incorrectly.
- I.e., the test would be **infeasible**.

- A sample event is **almost surely decidable** in W iff every possible probability measure in W assigns its boundary chance 0.

Almost Surely Decidable Sample Events

- A sample event is **almost surely decidable** in W iff there is zero chance that a sampled measurement hits exactly on its boundary.

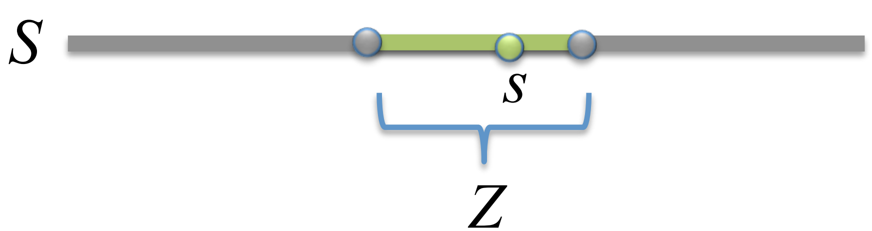
The Weak and Natural Assumptions

1. Entertain only **feasible methods** whose acceptance zones for various hypotheses are almost surely decidable.
2. The sample space has a **countable basis of almost surely decidable regions**.
   - True for discrete random variables.
   - True for continuous random variables.
3. Sampling is IID.

Epistemology of the Sample

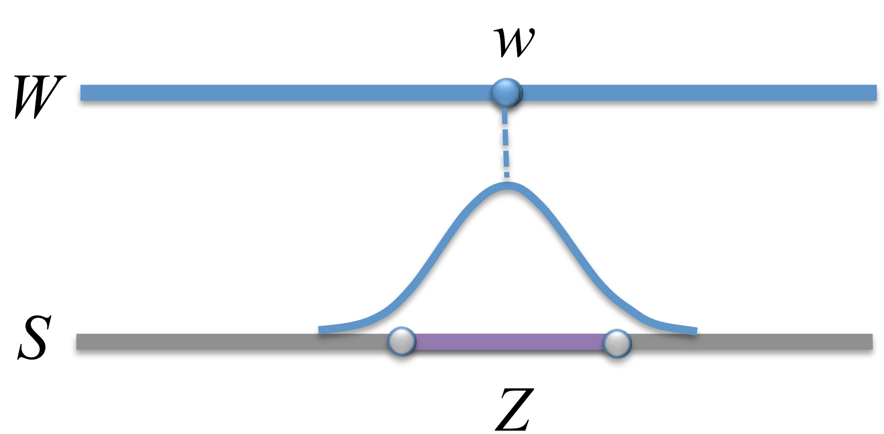
- The sample space S always comes with its own topology $\mathcal{T}$.
- $\mathcal{T}$ reflects what is verifiable about the **sample** itself.

s definitely falls within open interval Z .



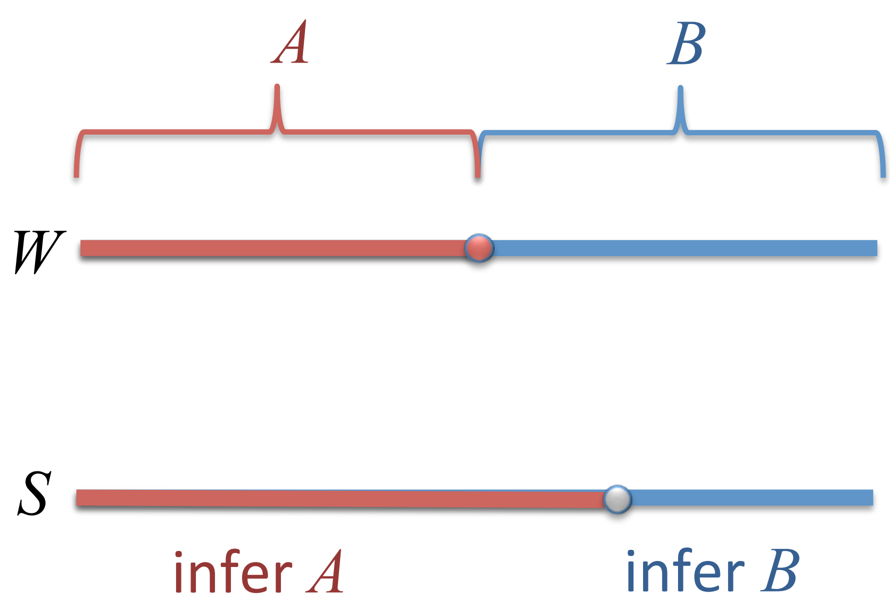
Feasible Sample Events

- It's impossible to decide whether a sample that lands right on the boundary of sample zone Z is really in or out of Z.
- Z is feasible iff the chance of its boundary is zero in every world, i.e. Z is *almost surely* decidable.



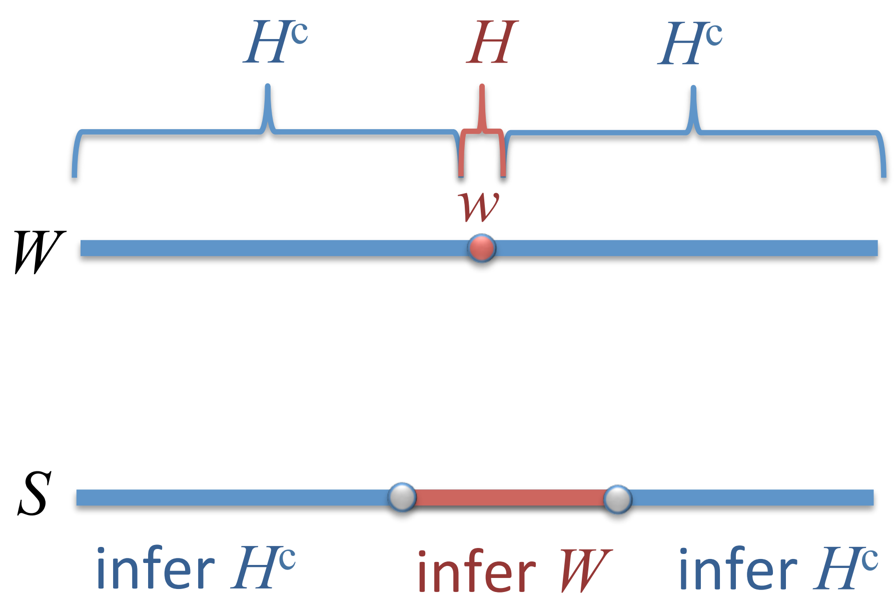
Feasible Method

A feasible method M is a statistical method whose acceptance zones for various conclusions are all feasible.



Feasible Tests

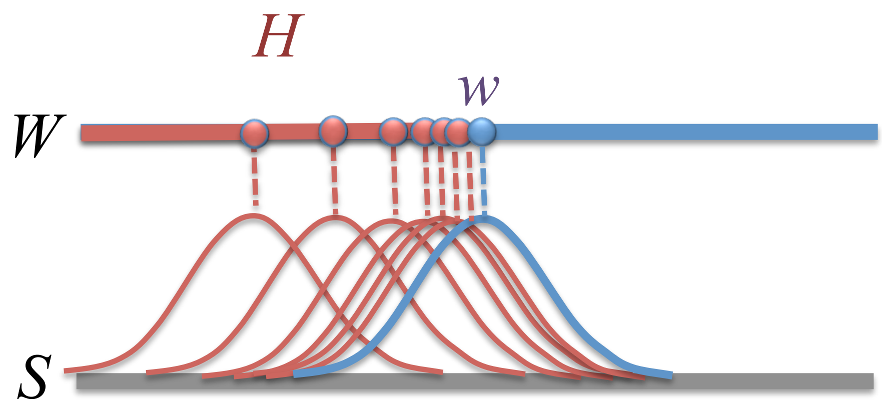
A feasible test of H is a feasible method that outputs H^c or W.



The Weak Topology

w ∈ cl H iff there exists sequence (w_n) in H, such that for all feasible tests M :

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



Weak Topology

Proposition: If $\mathcal{T}$ has a countable basis of feasible regions, then:

statistical information topology = weak topology.

Weak Topology

Proposition: If $\mathcal{T}$ is second-countable and metrizable, then the weak topology is second-countable and metrizable e.g., by the Prokhorov metric.

Methods

- A **statistical verification method** for H at level $\alpha > 0$ is a sequence (M_n) of feasible tests of H^c such that for every world w and sample size *n*:

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

  for $\alpha_n \rightarrow 0$, and dominated by α.

Monotonicity

Conjecture: For any open H and $\alpha > 0$, there exists (M_n) a verification method at level α such that if $w \in H$:

1. if $w \in H$: $p_w^{n_2}(M_{n_2} = H) + \alpha > p_w^{n_1}(M_{n_1} = H)$,
2. if $w \in H^c$: $p_w^{n_2}(M_{n_2} = W) > p_w^{n_1}(M_{n_1} = W)$,

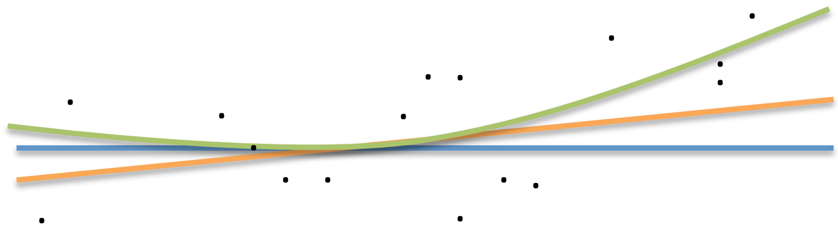
for all $n_2 > n_1$.

Topological Simplicity

It still makes sense in terms of statistical information topology!

$$A \lhd B \;\Leftrightarrow\; A \cap \mathsf{cl}(B) \setminus B \;\neq\; \varnothing.$$

$$H_1 \lhd H_2 \lhd H_3.$$



Ockham's Statistical Razor

Concern: "compatibility with E" is no longer meaningful.

Response: the third formulation of O.R. does not mention compatibility with experience!

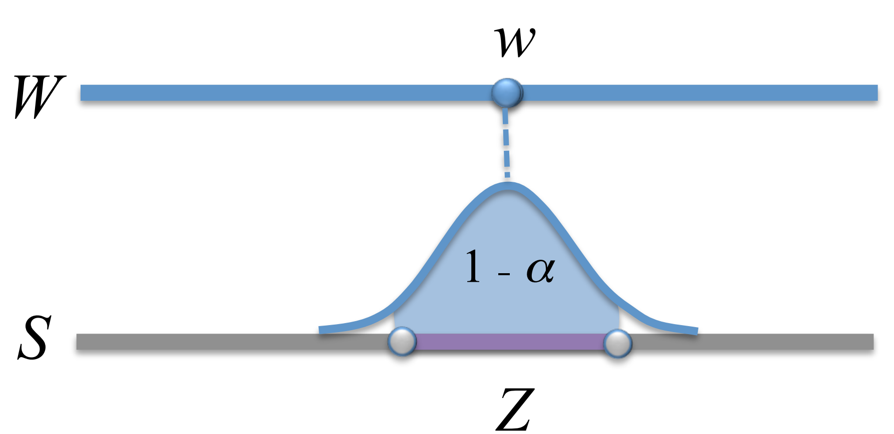
APPLICATION: OCKHAM'S STATISTICAL RAZOR (UNDER CONSTRUCTION)

Ockham's α-Razor

Statistical version of the error-razor:

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

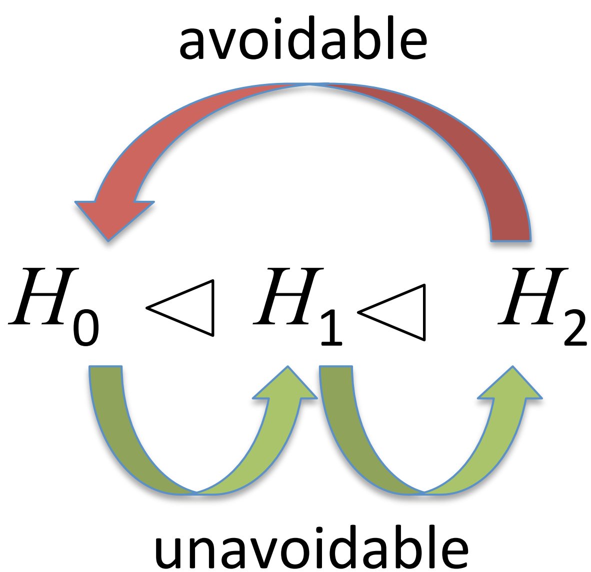
Agrees with significance for simple vs. complex binary questions!



Epistemic Mandate for Ockham's Razor

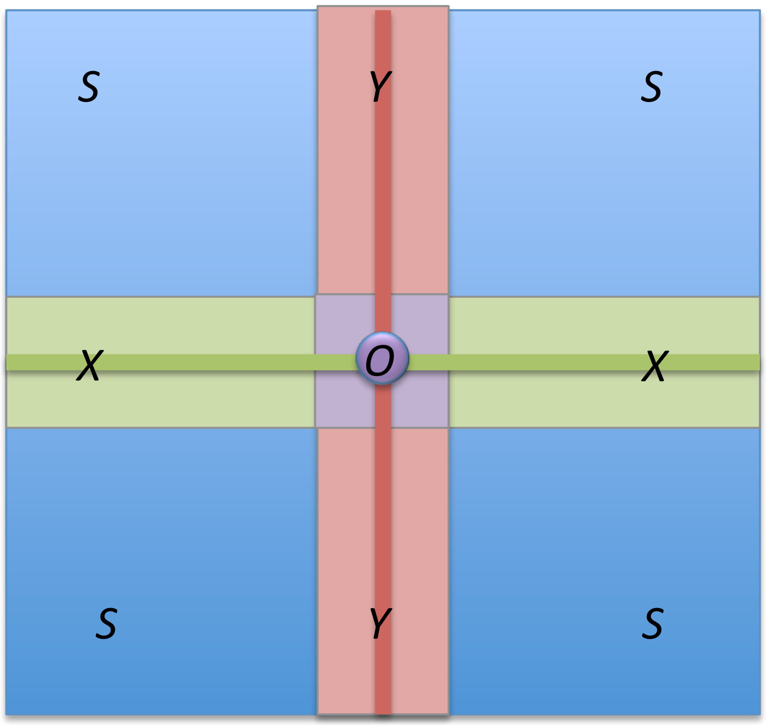
If you **violate Ockham's razor** with chance α, then

1. either you **fail to converge** to the truth in chance or
2. nature can force you into an α-**cycle of opinions** (complex-simple-complex), even though such cycles are avoidable.



O-Cycle Solution, Uniform Case

- Worlds: uniform distributions with unit square support
- Question: which mean components are non-zero?
- Method: output the simplest answer such that no sample point falls outside of its zone.



Progressive Methods

- Say that a solution is **progressive** iff the objective chance that it outputs the true answer is an increasing function of sample size.

- Say that a solution is α-**progressive** iff the chance that it outputs the true answer never decreases by more than α.

Result

- **Proposition:** If there is an enumeration of the answers $A_1, A_2, A_3, \ldots$ agreeing with the simplicity order, then there is an α-progressive solution for every α.

(Whenever α-monotonic verifiers exist for **ext** A_i)

Result

- **Proposition:** Every α-progressive solution is α-Ockham.

A New Objective Bayesianism

How much **prior bias toward simple models** is necessary to avoid α-cycles?

✗ Indifference = ignorance.

✓ truth-conduciveness.

CONCLUSION

A Method for Methodology

1. **Develop** basic methodological ideas in **topology**.
2. **Port** them to **statistics** via **statistical information topology**.

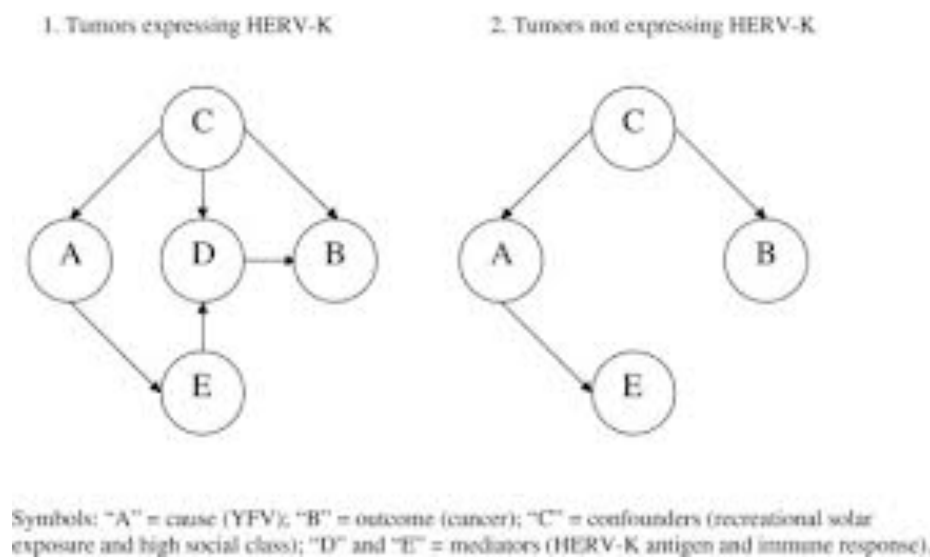
Some Concluding Remarks

1. **Information topology** is the structure of the scientist's problem context.
2. The apparent **analogy** between statistical and ideal methodology reflects **shared topological structure**.
3. Thereby, **ideal logical/topological ideas** can be **ported** directly to statistics.
4. The result is a new, systematic, **frequentist** foundation for **inductive inference** and **Ockham's razor**.

ETC.

Application: Causal Inference from Non-experimental Data

- **Causal network** inference from **retrospective** data.
- That is an **inductive** problem.
- The search is strongly guided by **Ockham's razor**.
- We have the **only non-Bayesian foundation** for it.



Application: Science

- All scientific conclusions are supposed to be counterfactual.
- Scientific inference is strongly simplicity biased.
- Standard ML accounts of Ockham's razor do not apply to such inferences (J. Pearl).
- Our account does.

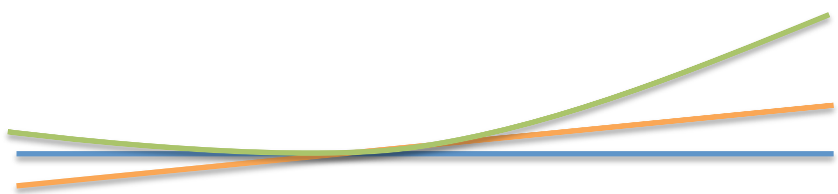
OCKHAM’S TOPOLOGICAL RAZOR

Popper Was Doing Topology

Popper's simplicity relation:

$$A \preceq B \;\Leftrightarrow\; A \subseteq \mathrm{cl}B.$$

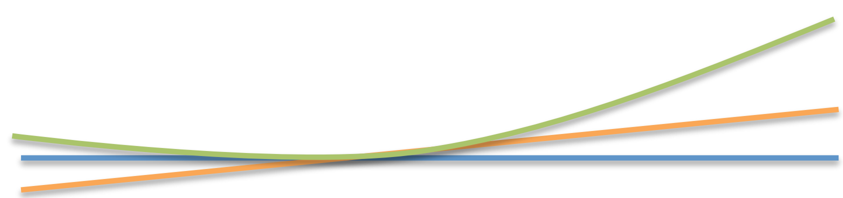
$$H_1 \preceq H_2 \preceq H_3.$$



An Improvement

$$A \lhd B \;\Leftrightarrow\; A \cap \mathsf{cl}(B) \setminus B \;\neq\; \varnothing.$$

$$H_1 \lhd H_2 \lhd H_3.$$



Topological Simplicity

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

Ockham's Razor

- A **question** partitions W into possible answers.
- A **relevant response** is a disjunction of answers.
- A **solution** is a method that converges to the true answer in every world in W.

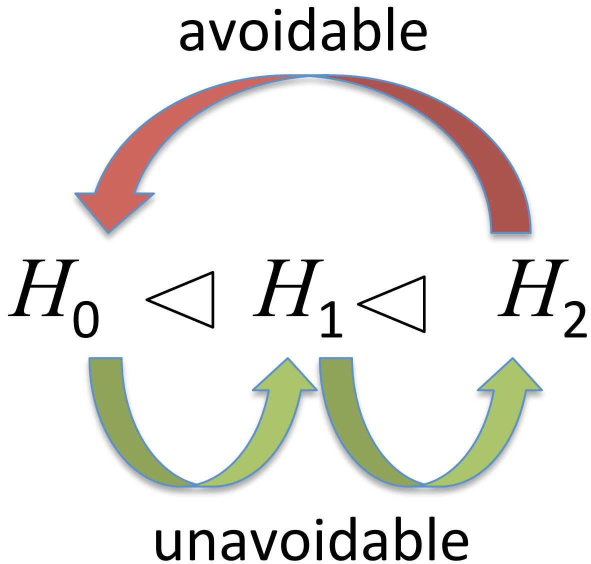
Proposition. The following principles are **equivalent**.

1. Infer a **simplest** relevant response in light of E.
2. Infer a **refutable** relevant response compatible with E.
3. Infer a relevant response that is **not more complex than the true answer**.

Epistemic Mandate for Ockham's Razor

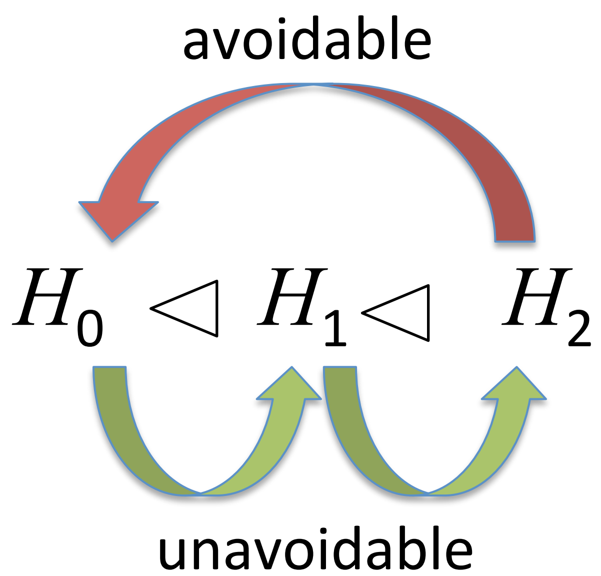
If you **violate Ockham's razor** then

1. either you **fail to converge** to the truth or
2. nature can force you into an **avoidable cycle of opinions**.



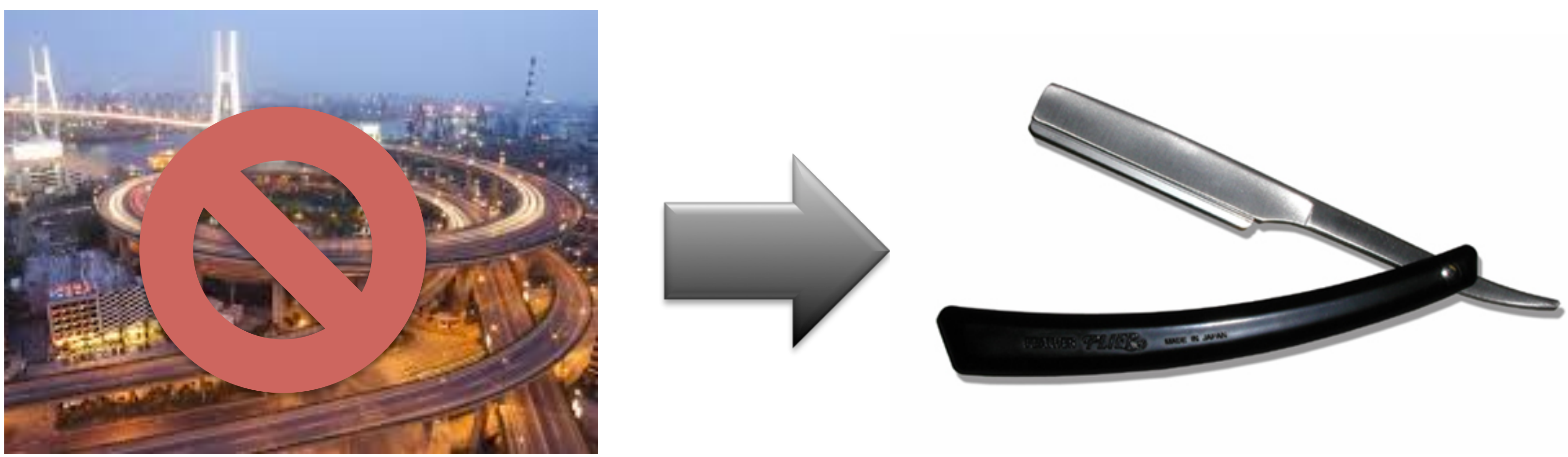
Does Not Presuppose Simplicity

Indeed, by **favoring** a **complex** hypothesis, you incur the avoidable cycle in a **complex** world!

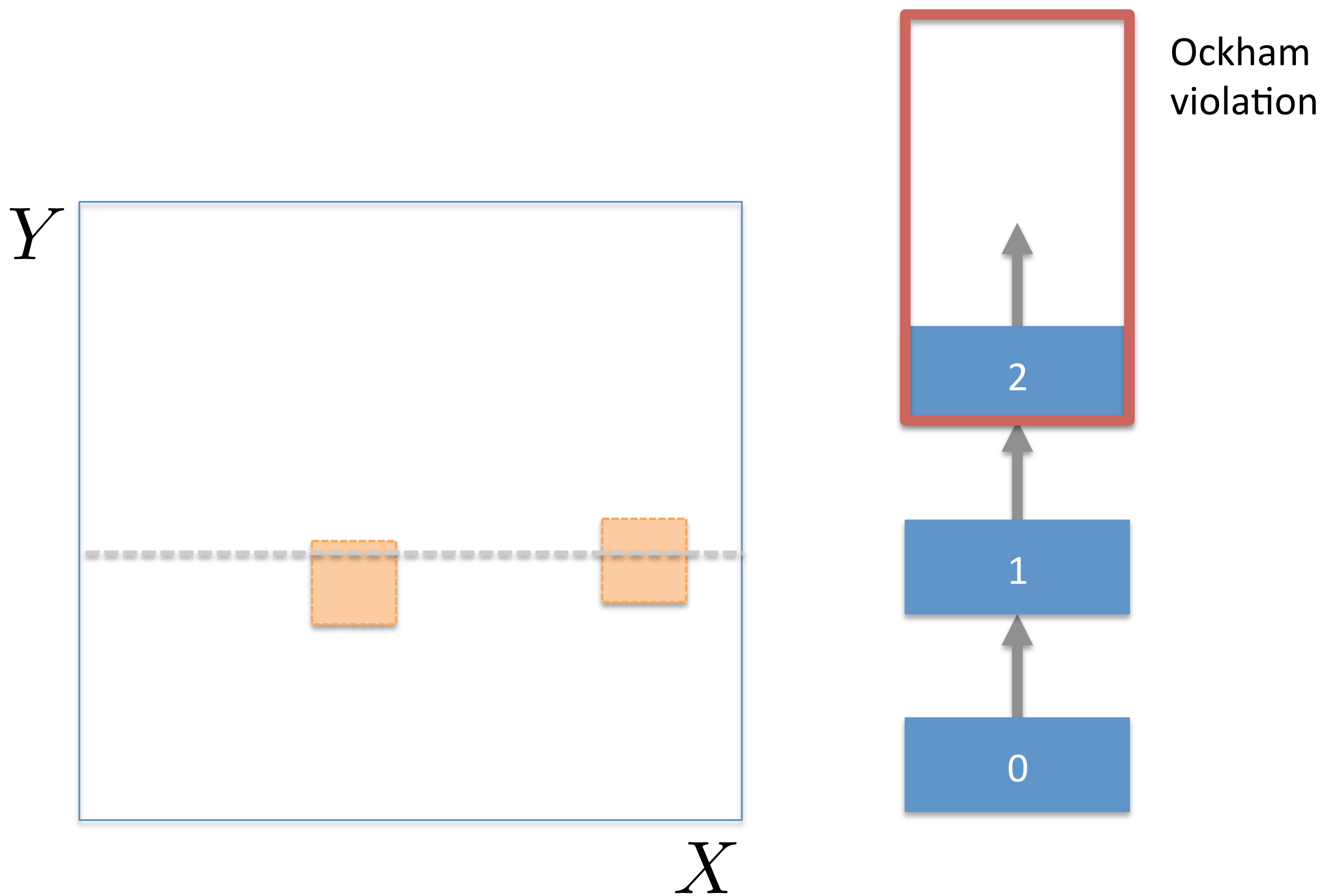


Result

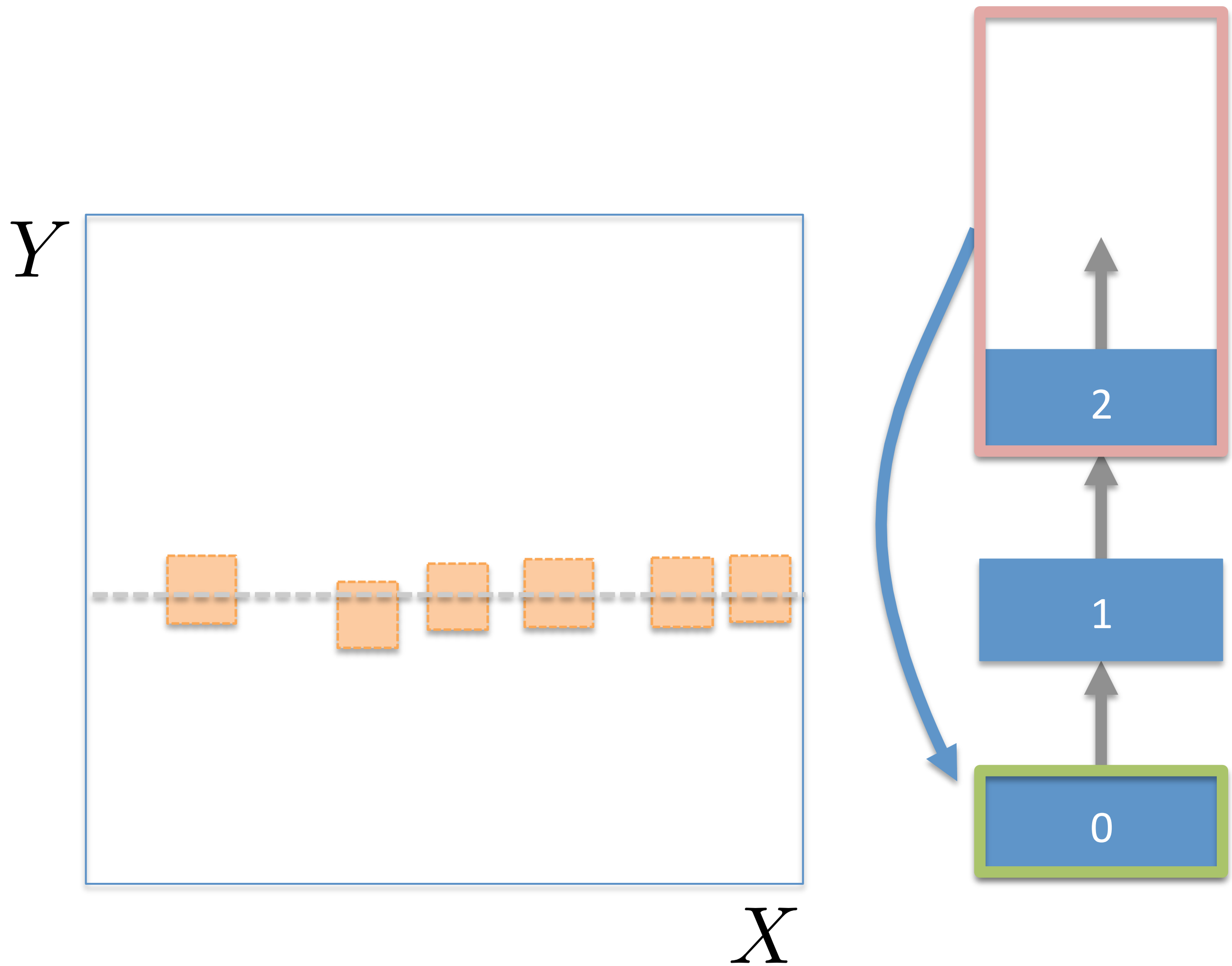
- **Proposition:** Every cycle-free solution satisfies Ockham's razor.



The Idea

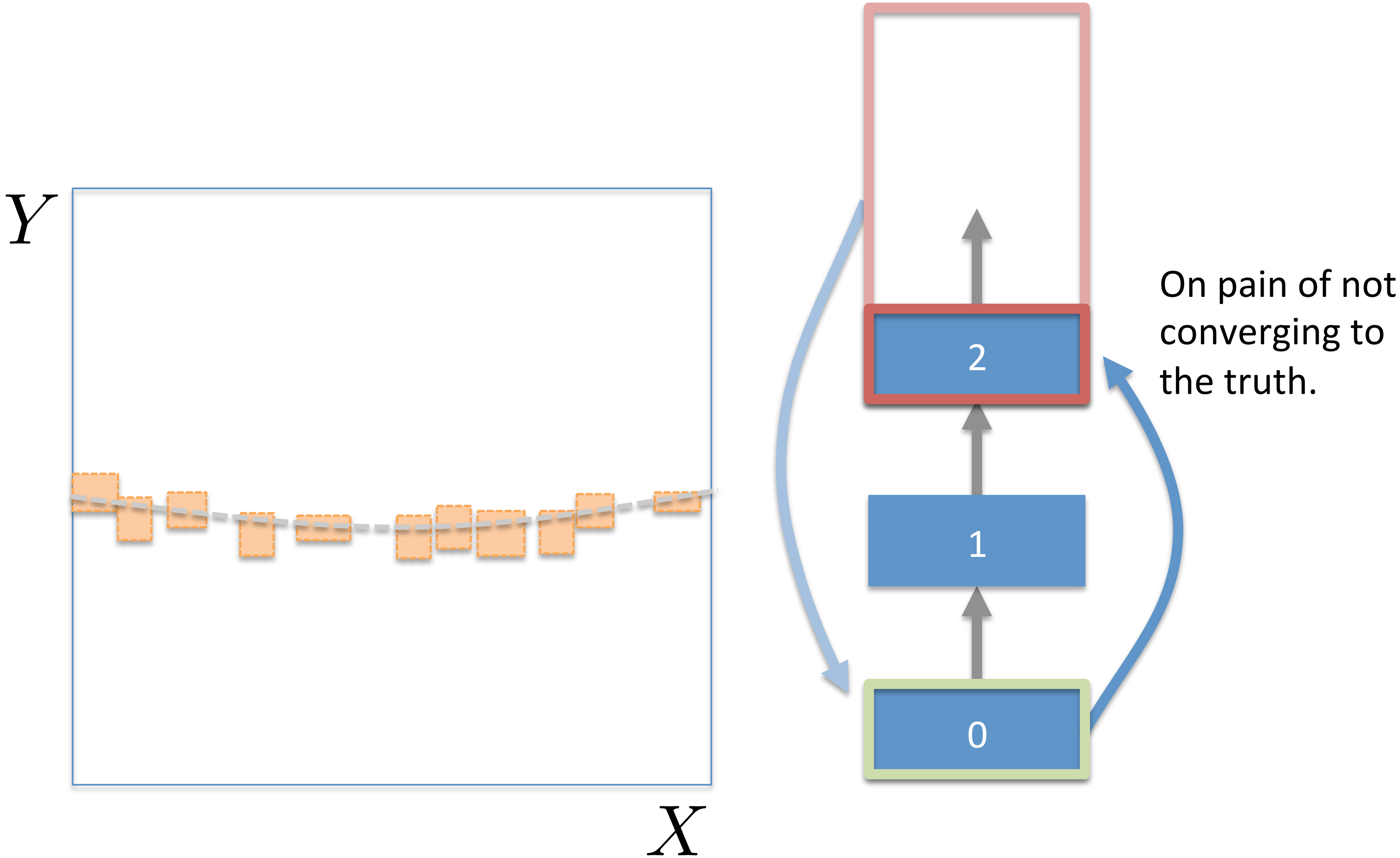


The Idea



On pain of not converging to the truth.

The Idea



Result

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