Deduction, Induction, Statistics, and Topology

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

Amsterdam 2016

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.

Taxonomy of Inference

inference

- deductive
  - truth preserving, monotonic.
- inductive
  - Everything else

Taxonomy of Inference

inference

deductive

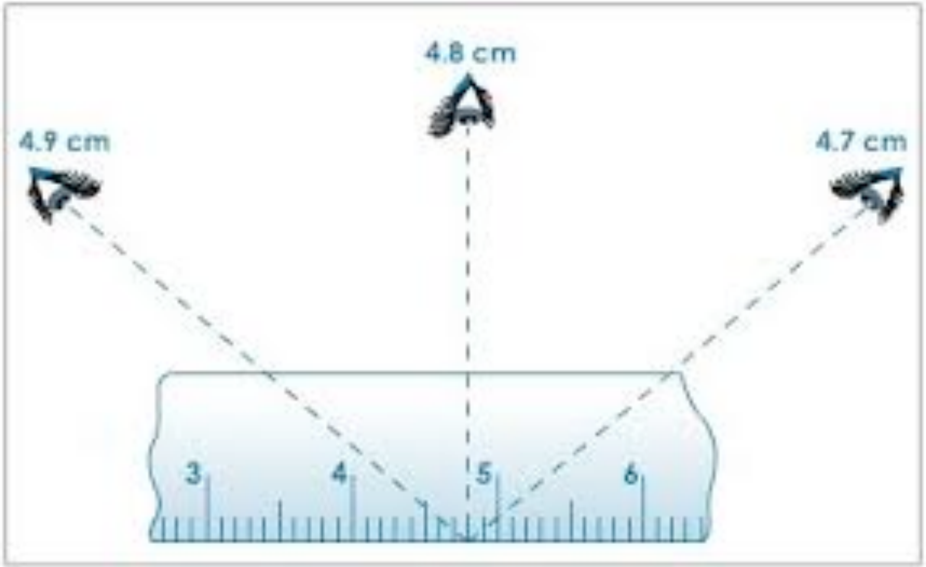
- Calculation
- Refuting universal *H*
- Verifying existential *H*
- Deciding between universal *H*, *H'*
- Predicting *E* from *H*
- Hypotheses compatible with *E*

inductive

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

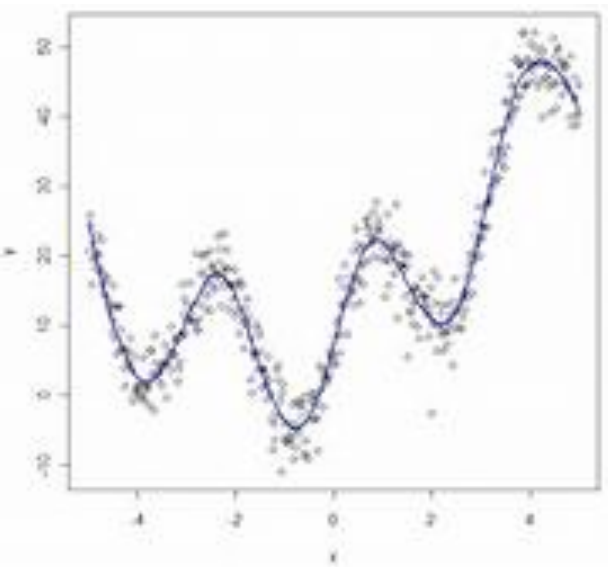
Real Data

- All real measurements are subject to probable error.
  - It can be reduced through redundancy (sample size).



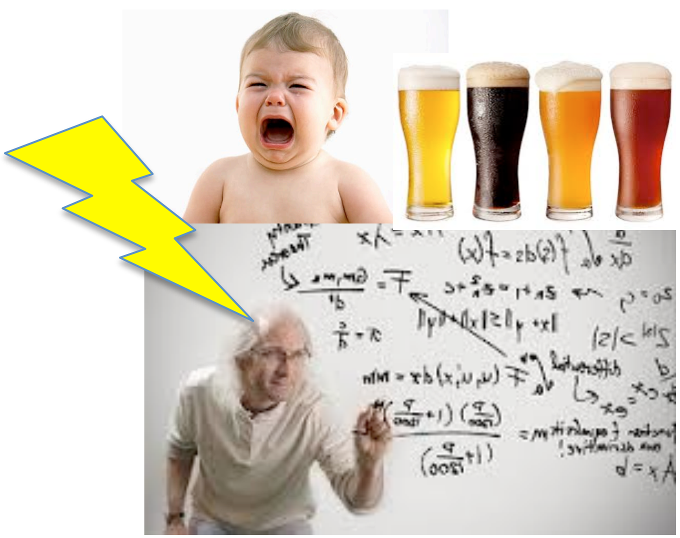
Real Predictions

- The predictions of probabilistic theories are subject to probable error.
  - It can be reduced through repeated sampling.



Real Calculations

- All real calculations are subject to probable error.
  - It can be reduced by redundant codes, circuits, and refereeing.



Stochastic 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.

Taxonomy of Inference

inference

inductive

strictly deductive

stochastically deductive

everything else

Taxonomy of Inference

inference

inductive

strictly 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

stochastically 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

Bad Taxonomy

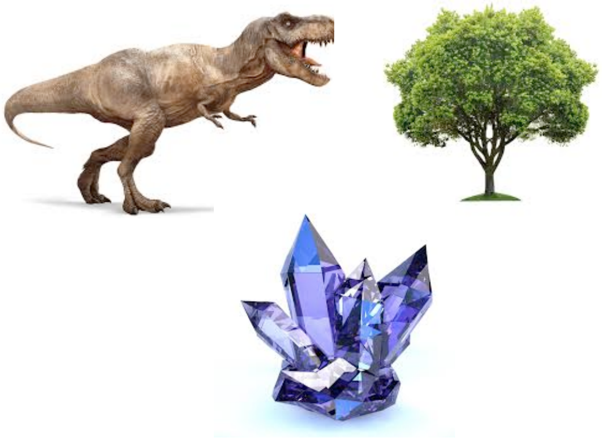
things

everything else

white roses

non-white roses

everything else



Improved Taxonomy of Inference

inference

deductive

strict

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

stochastic

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

strict

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

stochastic

1. Inferring simple H_0
2. Model selection

Improved Taxonomy of Inference

inference

deductive

strict

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

stochastic

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

strict

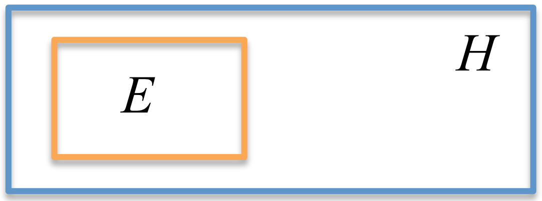
1. Inferring universal H_0
2. Choosing between universal H_0, H_1, H_1, ...

stochastic

1. Inferring simple H_0
2. Model selection

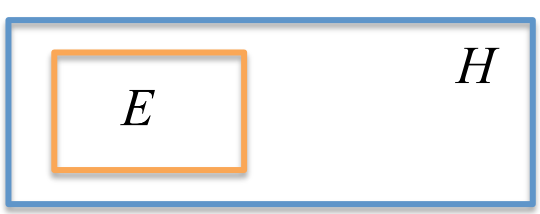
Question

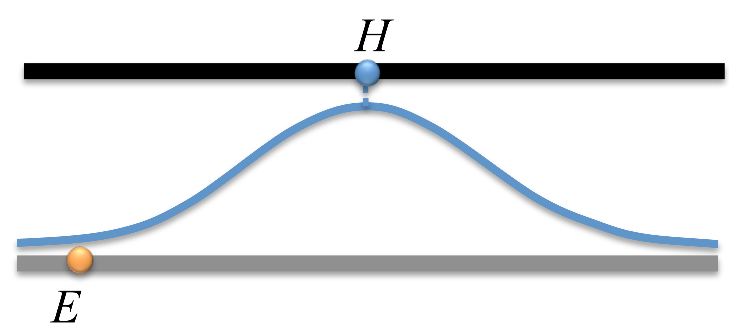
- In strict deduction, the evidence rules out possibilities.



Question

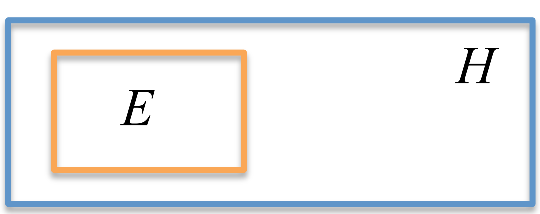
- In strict deduction, the evidence rules out possibilities.
- In statistical deduction, the sample is logically compatible with every possibility.

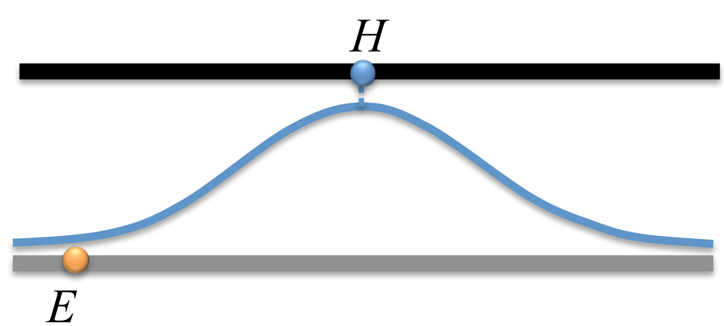




Question

- In strict deduction, the evidence rules out possibilities.
- In statistical deduction, the sample is logically compatible with every possibility.
- Is there a common, underlying sense of **empirical information**?



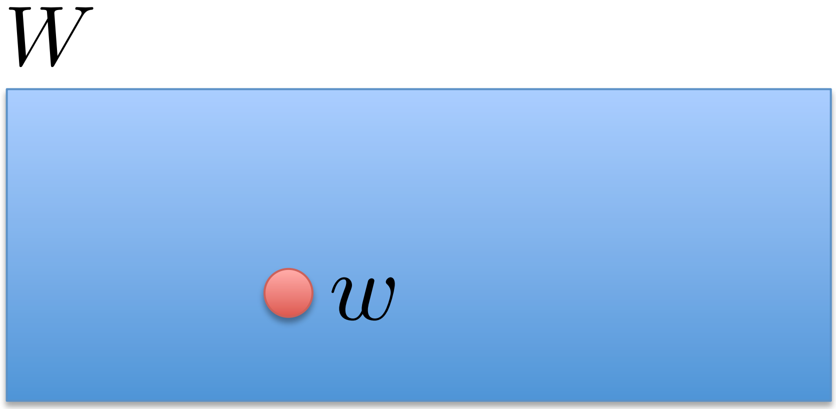


The Topology of Information

CMU ILLC

Worlds

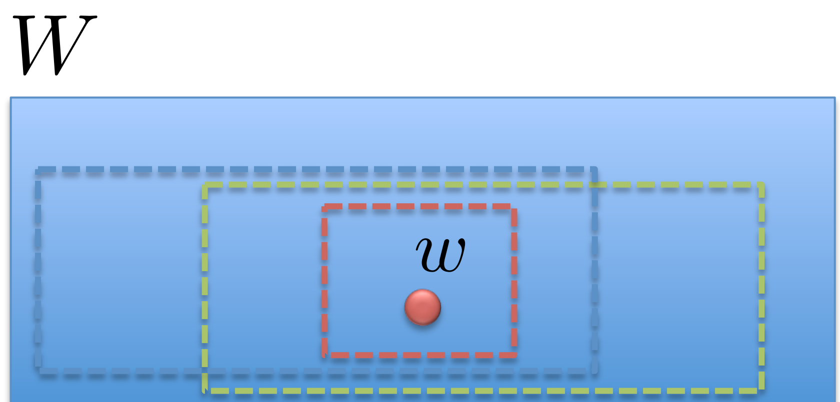
- The points in W are **possible worlds**.



The Structure of Information

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

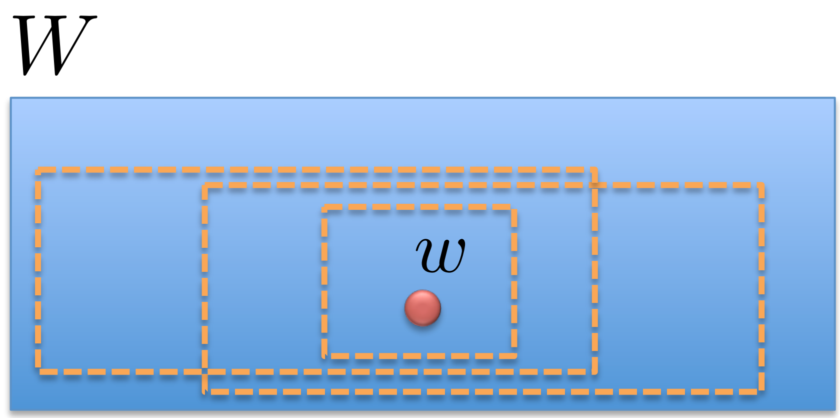
1. some information state true;
2. each true pair of information states is entailed by a true information state.



The Structure of Information

Local information basis at w:

$$\mathcal{I}(w) \;:=\; \{E \in \mathcal{I} : w \in E\}.$$



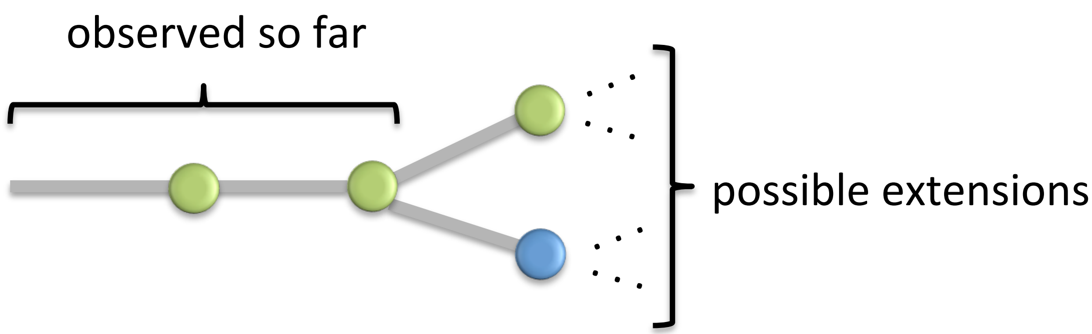
Sleeping Beauty Theorist

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

Example: Sequential Binary Experiment

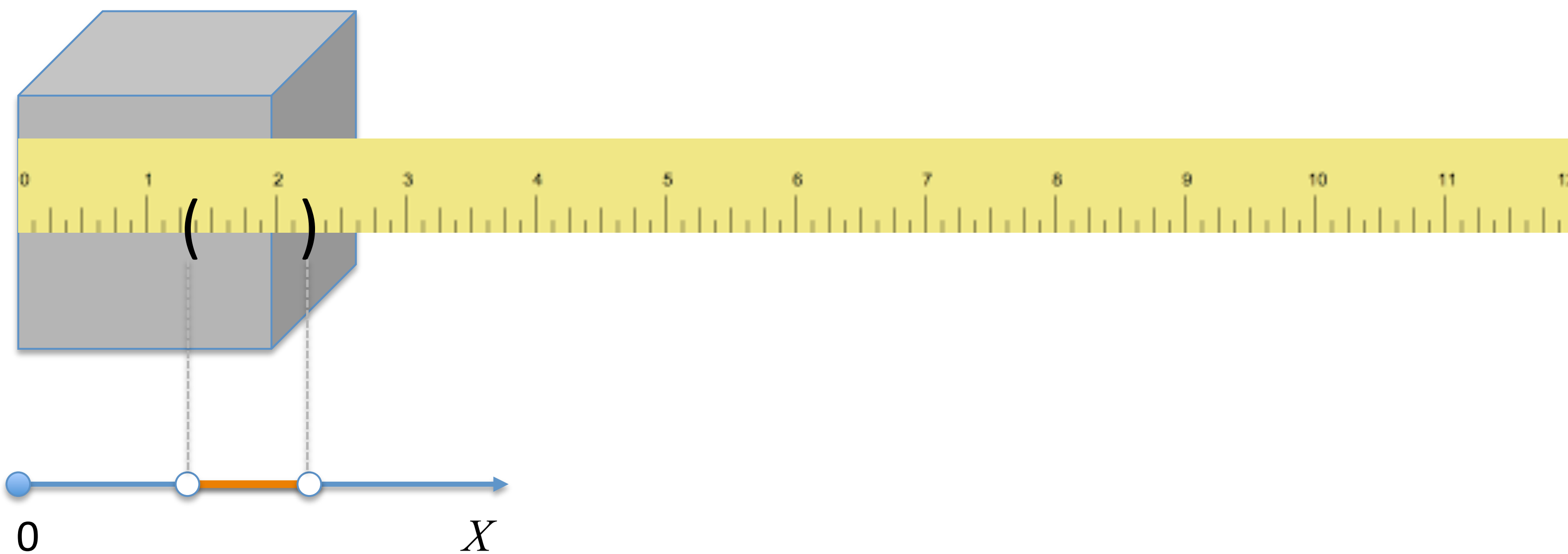
Worlds = infinite discrete sequences of outcomes.

Information states = cones of possible extensions:



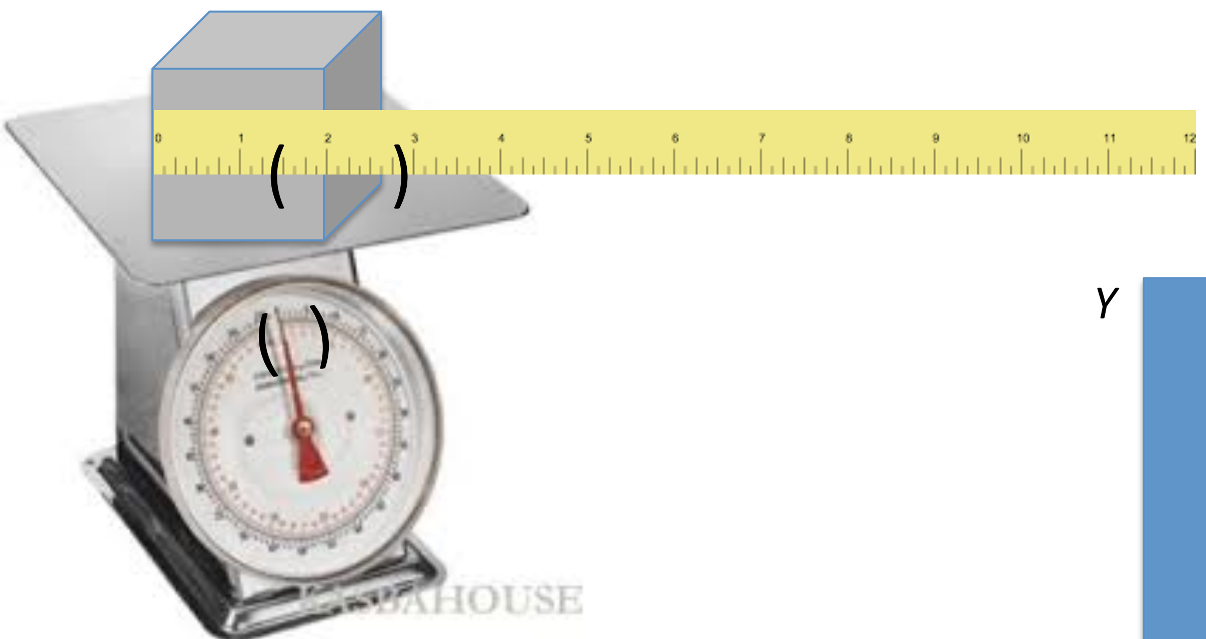
Example: Measurement of X

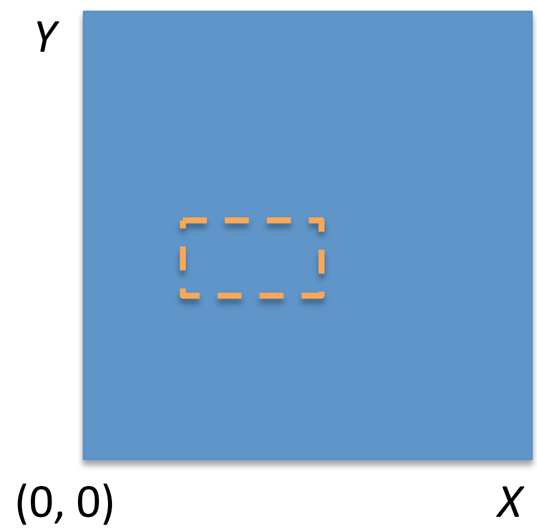
- **Worlds** = real numbers.
- **Information states** = open intervals.



Example: Joint Measurement

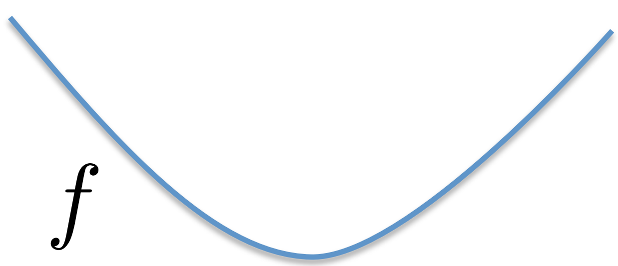
- **Worlds** = points in real plane.
- **Information states** = open rectangles.





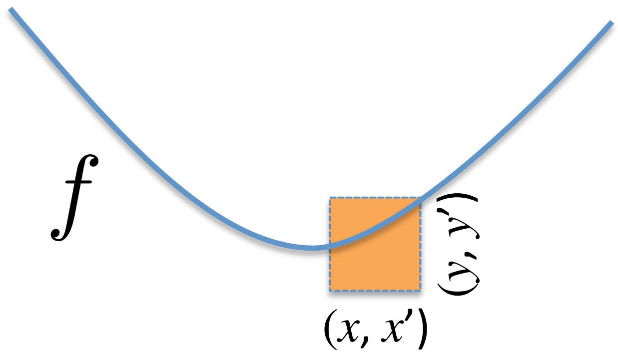
Example: Equations

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



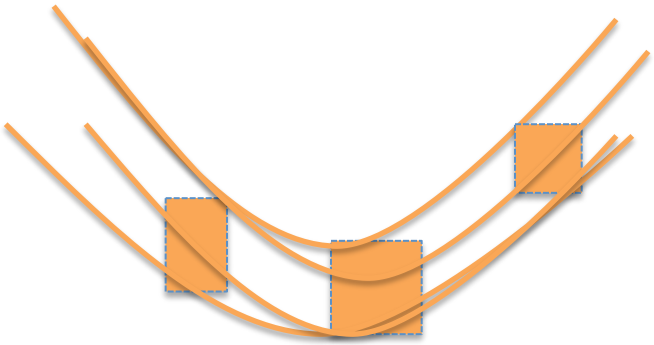
Example: Laws

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



Example: Laws

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

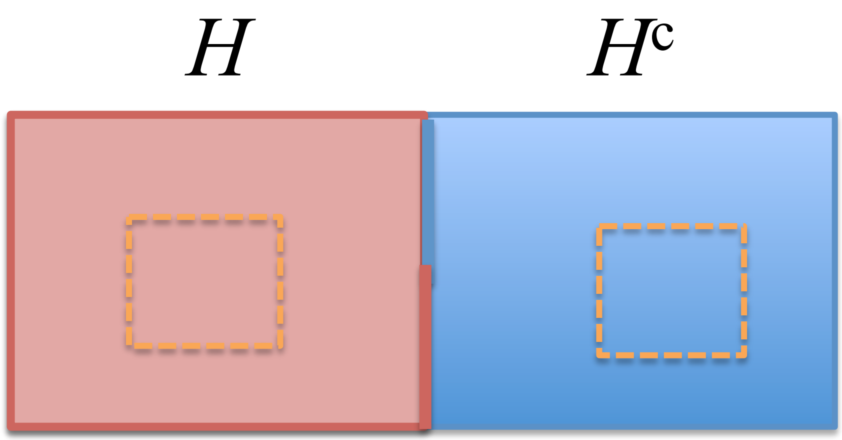


Deductive Verification and Refutation

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.

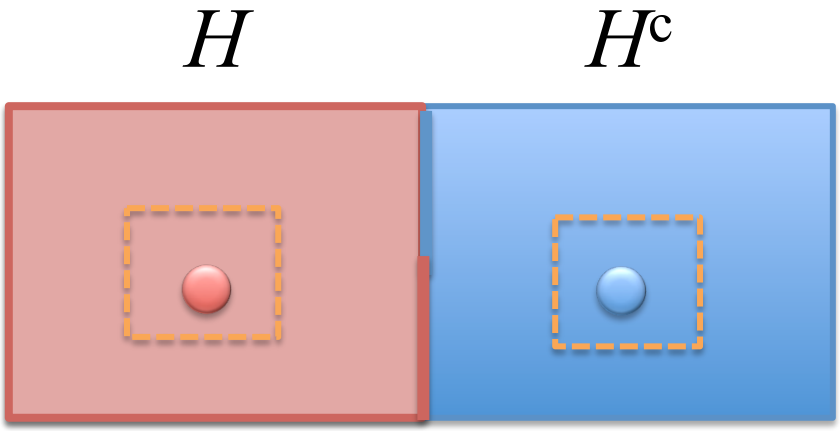


Will be Verified

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

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

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

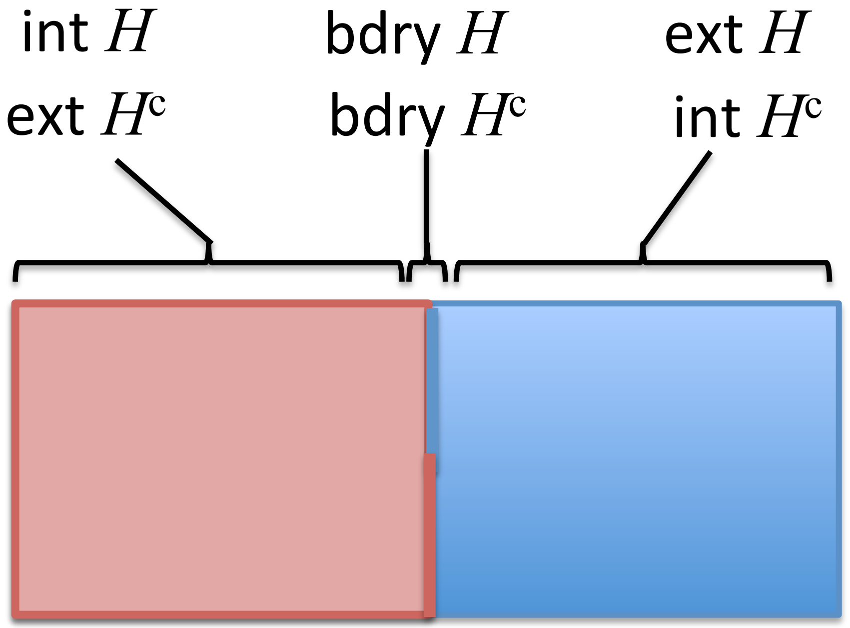


Will be Verified

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

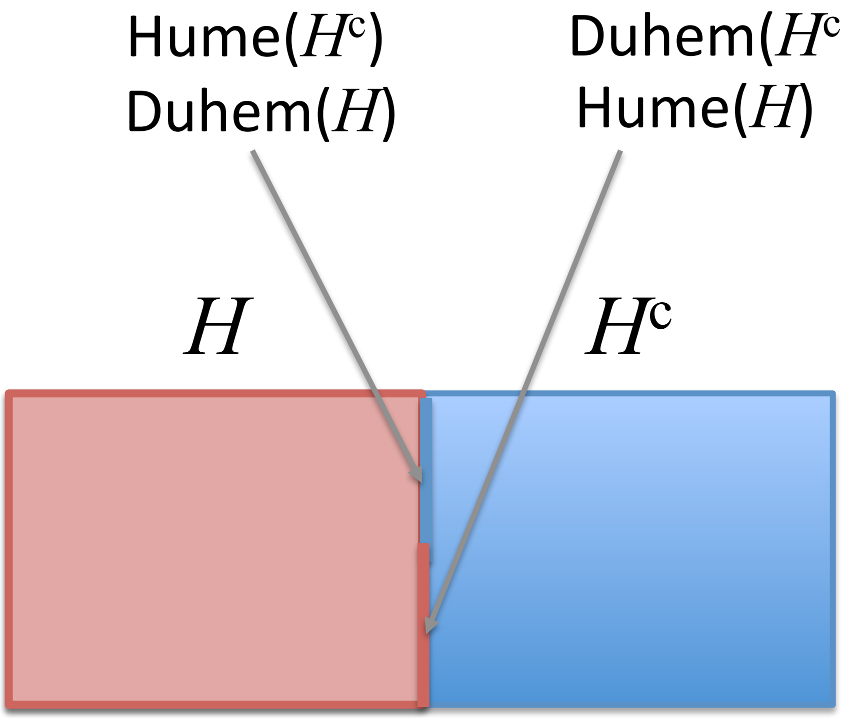
ext H := the proposition that H **will be refuted.**

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



Will be Verified

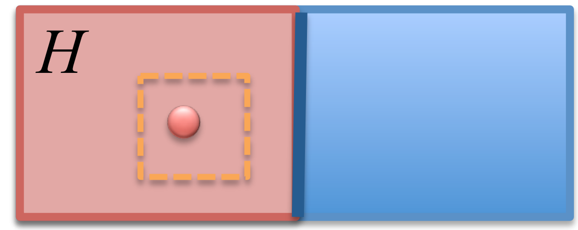
- bdry(H) ∩ H = "you face **Hume's problem** w.r.t. H";
- bdry(H) ∩ H^c = "you face **Duhem's problem** w.r.t. H"



Verifiability, Refutability, Decidability

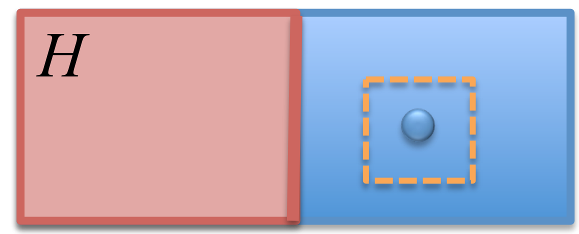
H is **verifiable** iff $H \subseteq \text{int}(H)$.

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

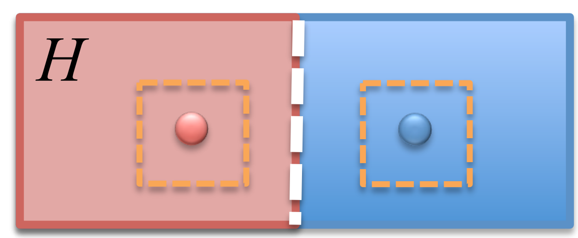


H is **refutable** iff $\text{cl}(H) \subseteq H$.

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



H is **decidable** iff *H* is both verifiable and refutable.



Methods

- A **verification method** for H is an inference rule $V(E) = A$ such that in every world w:
  1. $w \in H$: V converges to H without error.
  2. $w \in H^c$: V always concludes W.

Methods

- A **verification method** for H is an inference rule $V(E) = A$ such that in every world w:
  1. $w \in H$: V converges to H without error.
  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.

Methods

- A **limiting verification method** for H is an inference rule $V(E) = A$ such that in every world w:

  $w \in H$ iff V 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.

Methods

- A **verification method** for H is an inference rule $V(E) = A$ such that in every world w:
  1. $w \in H$: V converges to H without error.
  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.
- H is **methodologically verifiable [refutable, decidable, etc.]** iff H has a method of the corresponding kind.

Verification, Refutation, and Decision are Deductive

Proposition (truth preservation).

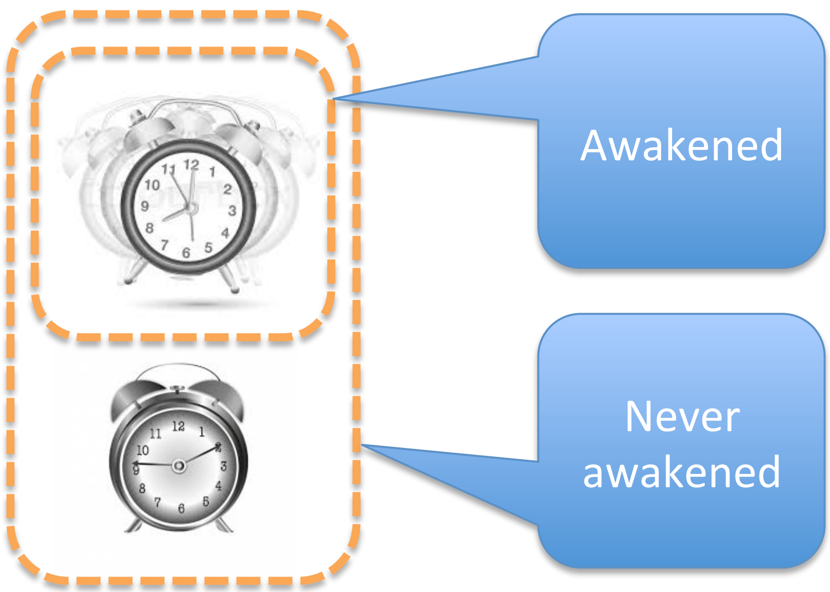
If V is a verifier, refuter or decider for H and $V(E) = A$, then $E \subseteq A$.

Proposition (monotonicity).

If there is a verifier, refuter or decider for H, then there is a monotonic one that never drops H or H^c after having concluded it.

Limiting Verification, Refutation, and Decision are Inductive

Proposition. No limiting verifier of "never awakened" is truth preserving or monotonic.



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.

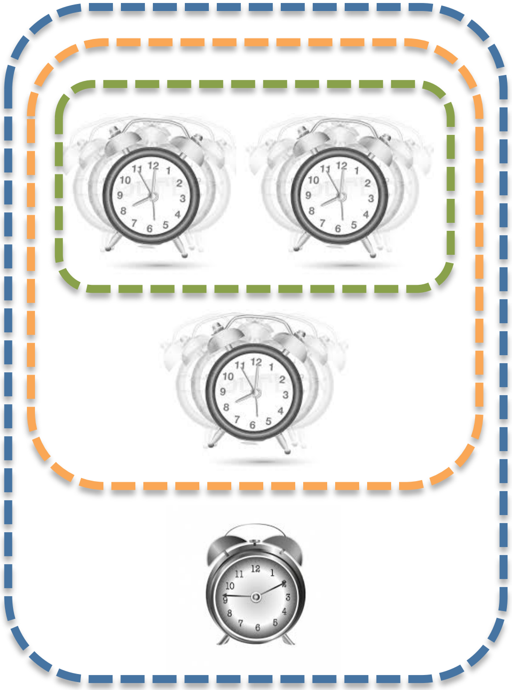
- H is **locally closed** iff H is a difference of open sets.

Sleeping Theorist Example

H_2 = "Awakened twice" is open.

H_1 = "Awakened once" is locally closed.

H_0 = "Never awakened" is closed.

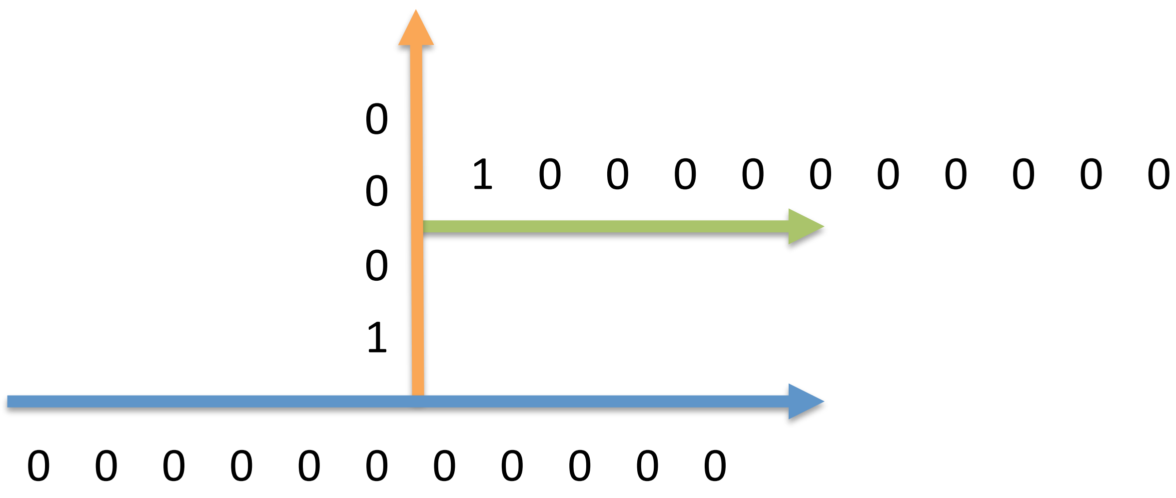


Sequential Example

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.

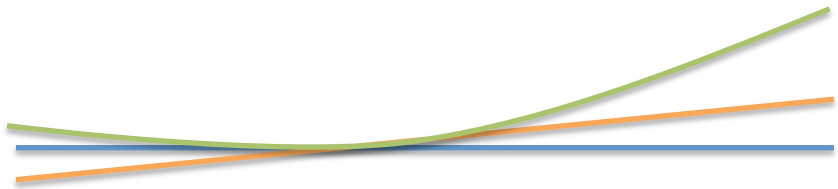


Equation Example

H_2 = "quadratic" is open.

H_1 = "linear" is locally closed.

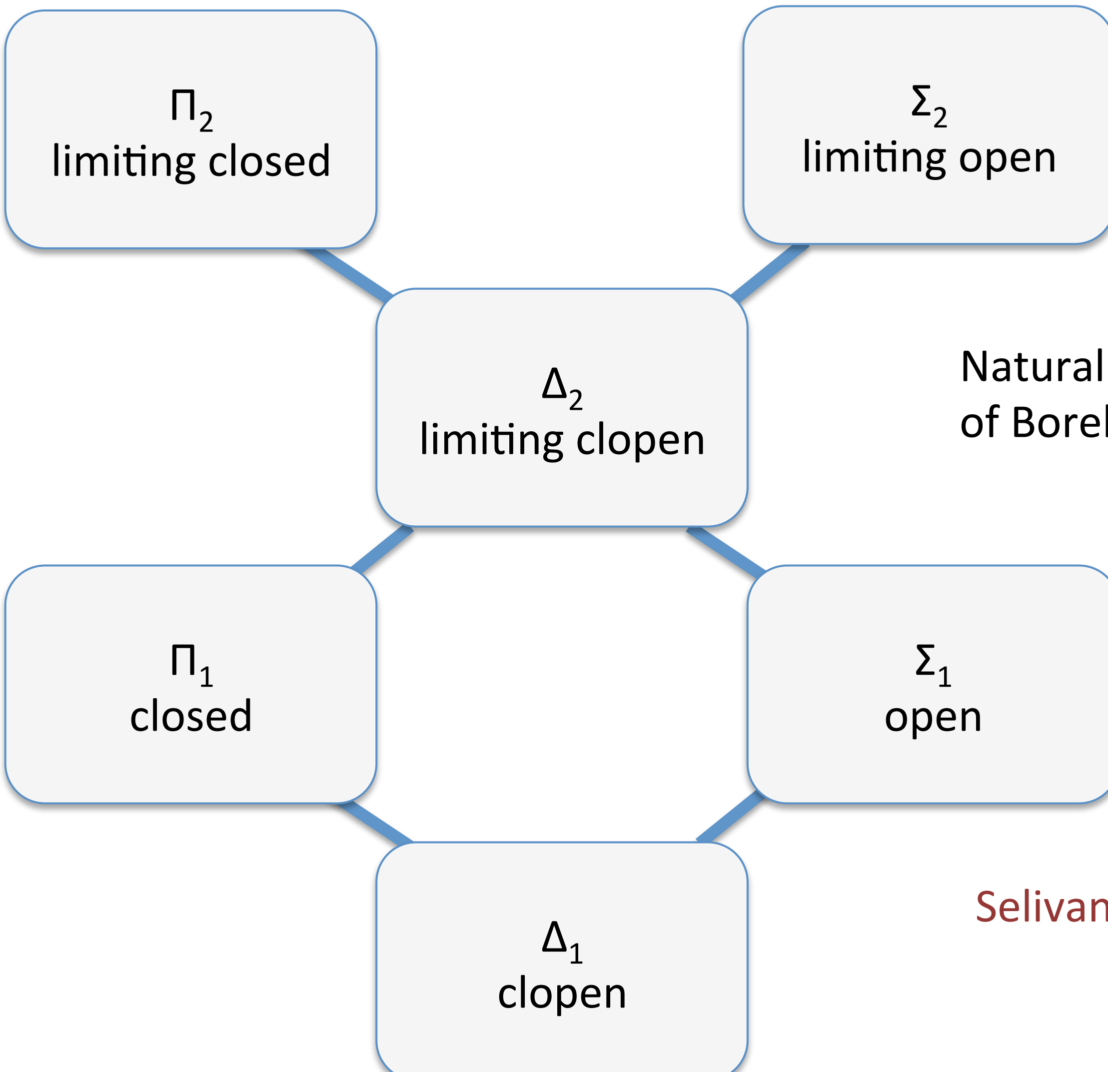
H_0 = "constant" is closed.



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.

deBrecht Hierarchy

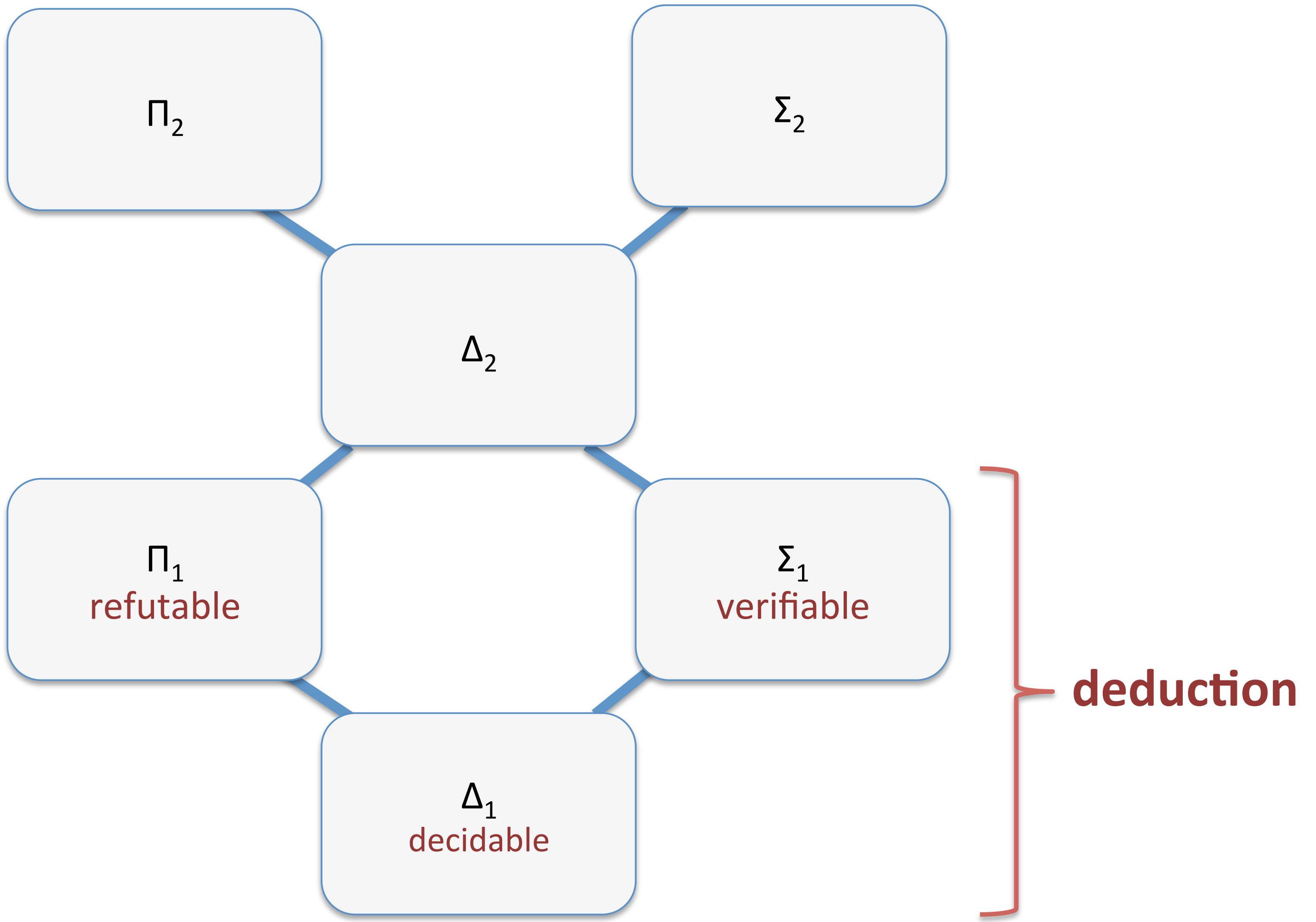


Natural generalization of Borel hierarchy.

Selivanov 2006 (?)

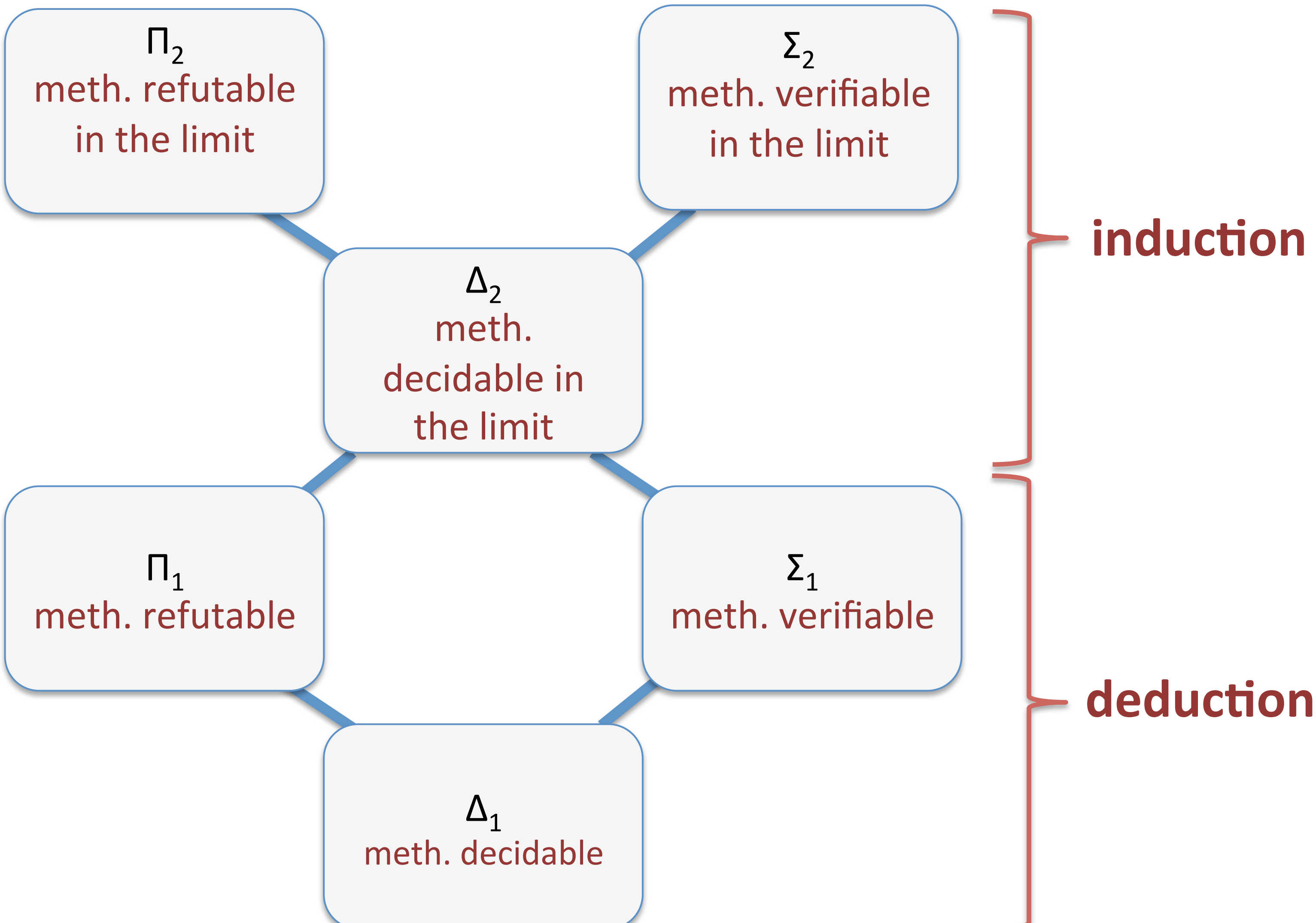
Topology and Pragmatics

Prop.



Topology and Methodology

Prop.



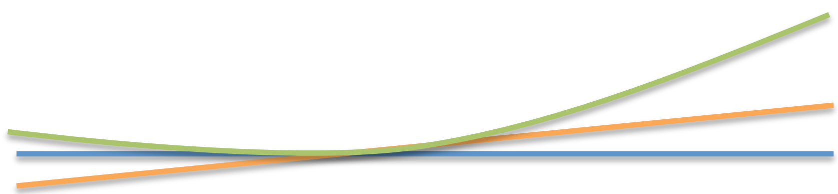
OCKHAM'S TOPOLOGICAL RAZOR

Popper Was Doing Topology!

Popper's simplicity relation:

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

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

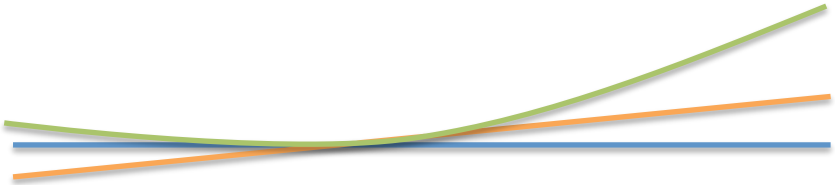


A Slight Revision

Our simplicity relation:

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

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



Ockham's Razor

- A **question** partitions W into possible answers.
- A **relevant response** is a disjunction of answers.

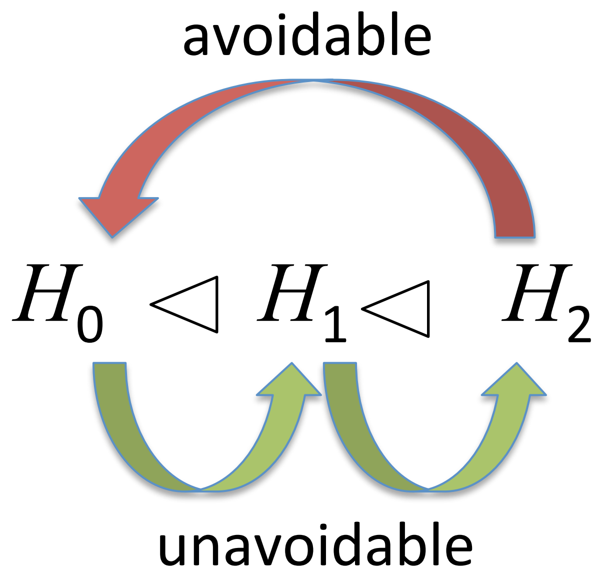
Proposition. The following principles are **equivalent**.

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

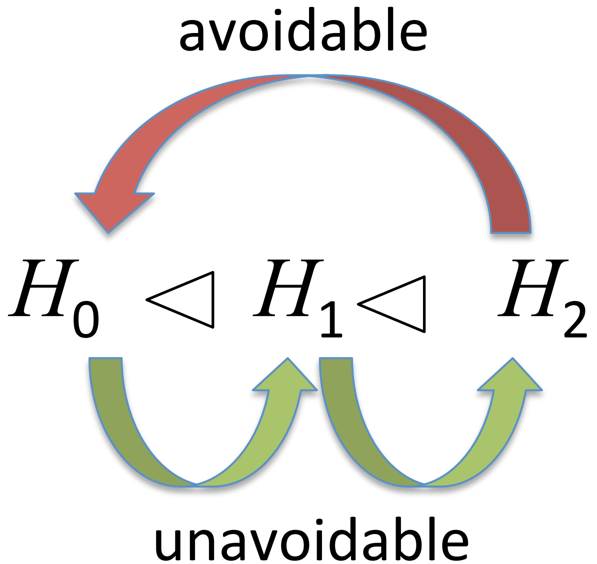
If you **violate Ockham's razor** then

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

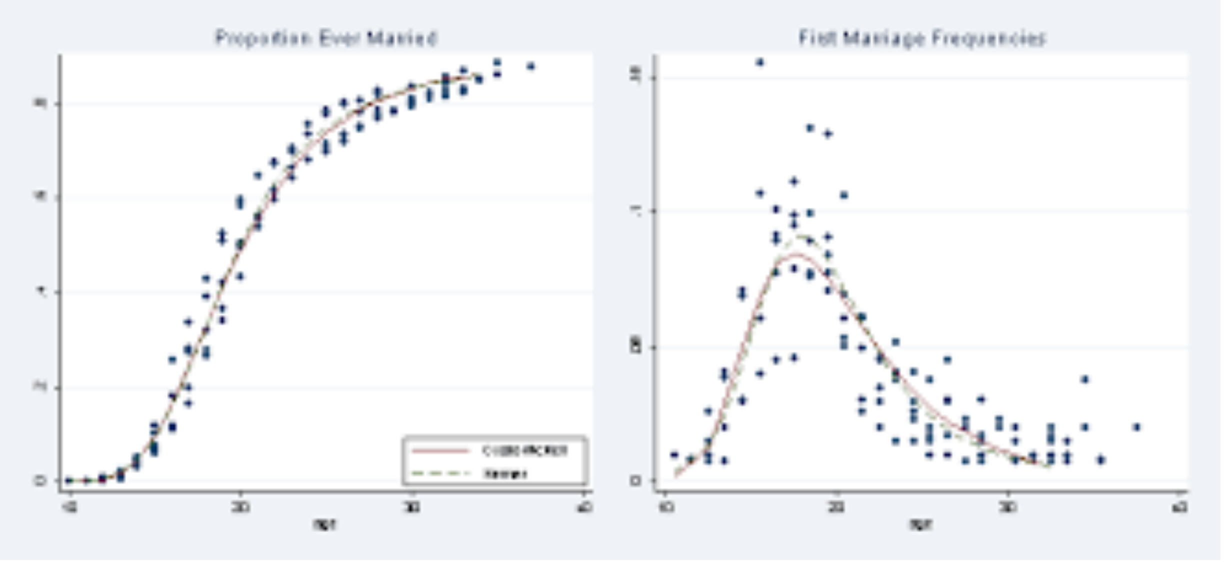
Indeed, by **favoring** a **complex** hypothesis, you incur the avoidable cycle in a **complex** world!



STATISTICAL INFORMATION TOPOLOGY

Statistical Information Topology

= the topology that lifts the preceding results to statistical inference.



Skepticism

The above account...

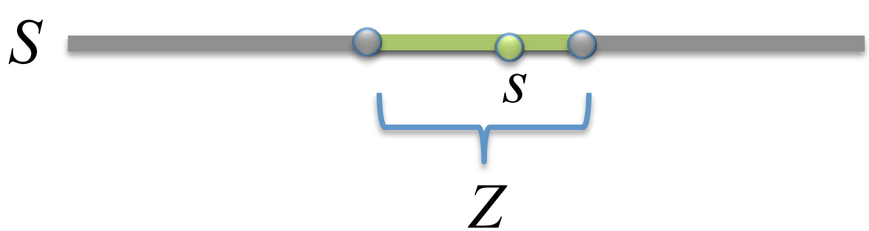
"may be okay if the candidate theories are **deductively** related to observations, but when the relationship is **probabilistic**, I am **skeptical** ...".

Eliott Sober, *Ockham's Razors*, 2015

Epistemology of the Sample

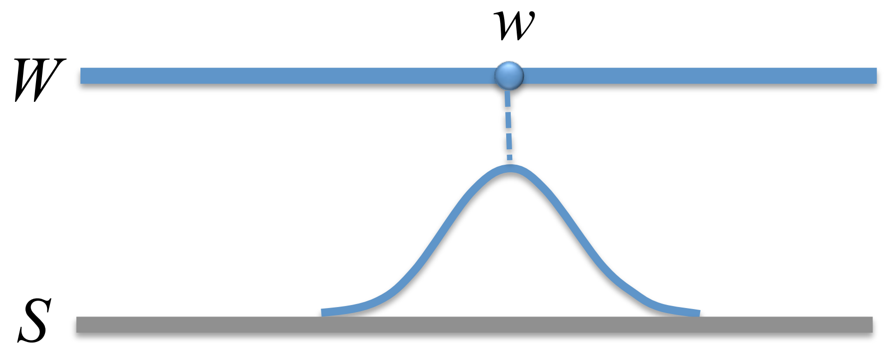
- The sample space S always comes with its own topology $\mathcal{T}$.
- $\mathcal{T}$ reflects what is verifiable about the **sample** itself.

s definitely falls within open interval Z .



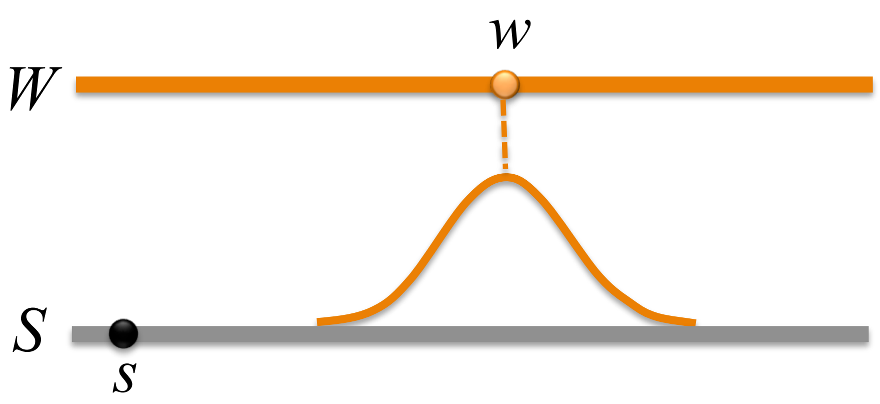
Statistics

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



The Difficulty

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



Response

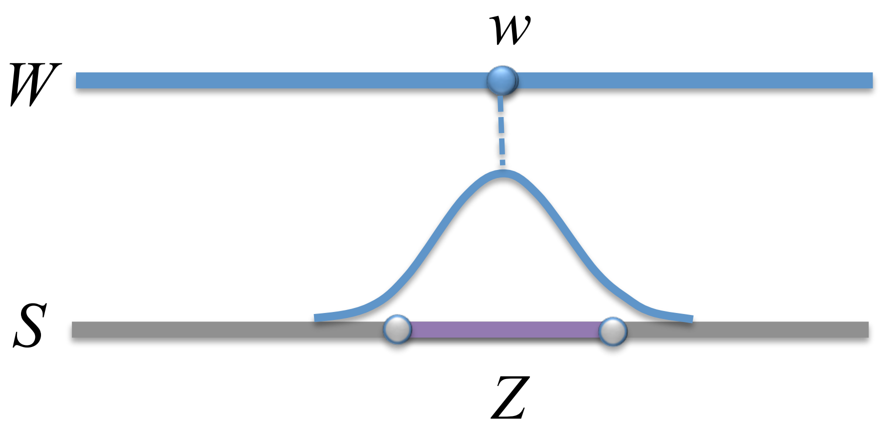
- Solve for the unique topology such that:

statistically verifiable = **open.**

Topology  Statistics

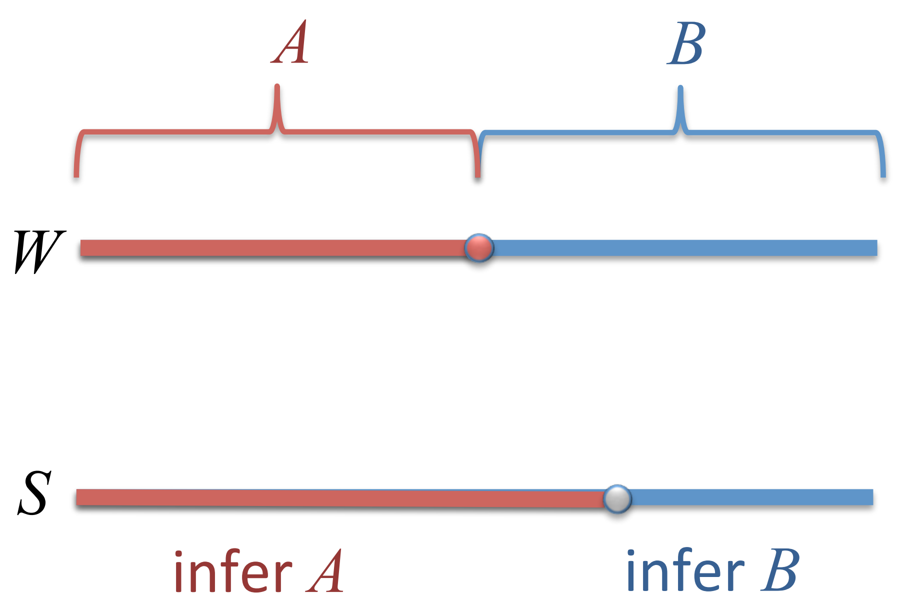
Feasible Sample Events

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



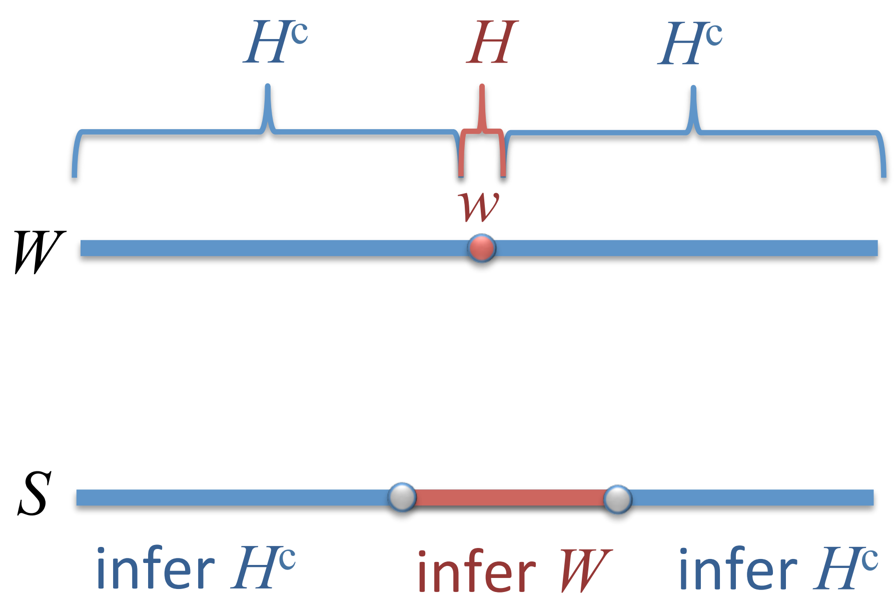
Feasible Method

A feasible method M is a measurable function from samples to propositions over W such that $M^{-1}(A)$ is feasible, for all A.



Feasible Tests

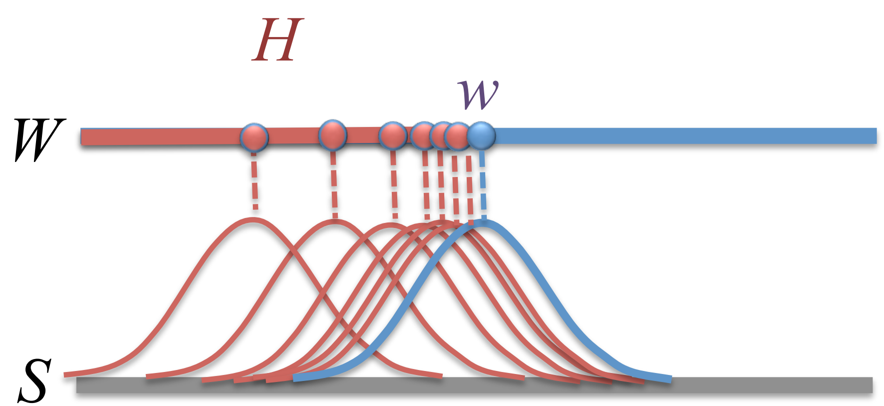
A feasible test of H is a feasible method that outputs H^c or W.



Statistical Information Topology

$w \in \text{cl}\, H$ 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 $1\text{-}\alpha$.

- 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.

Methods

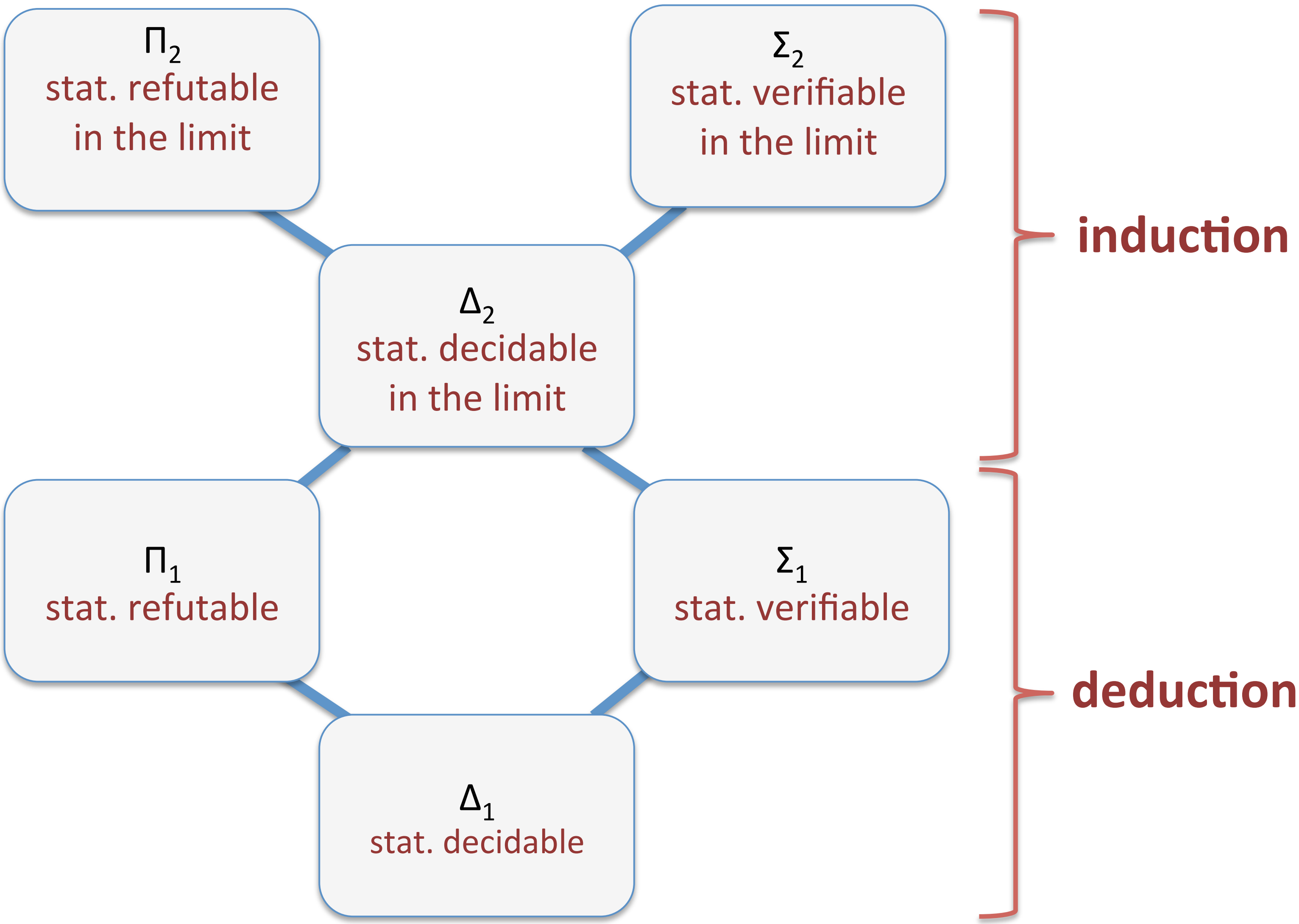
- A **limiting statistical verification method** for H is a sequence (M_n) of feasible methods such that:

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

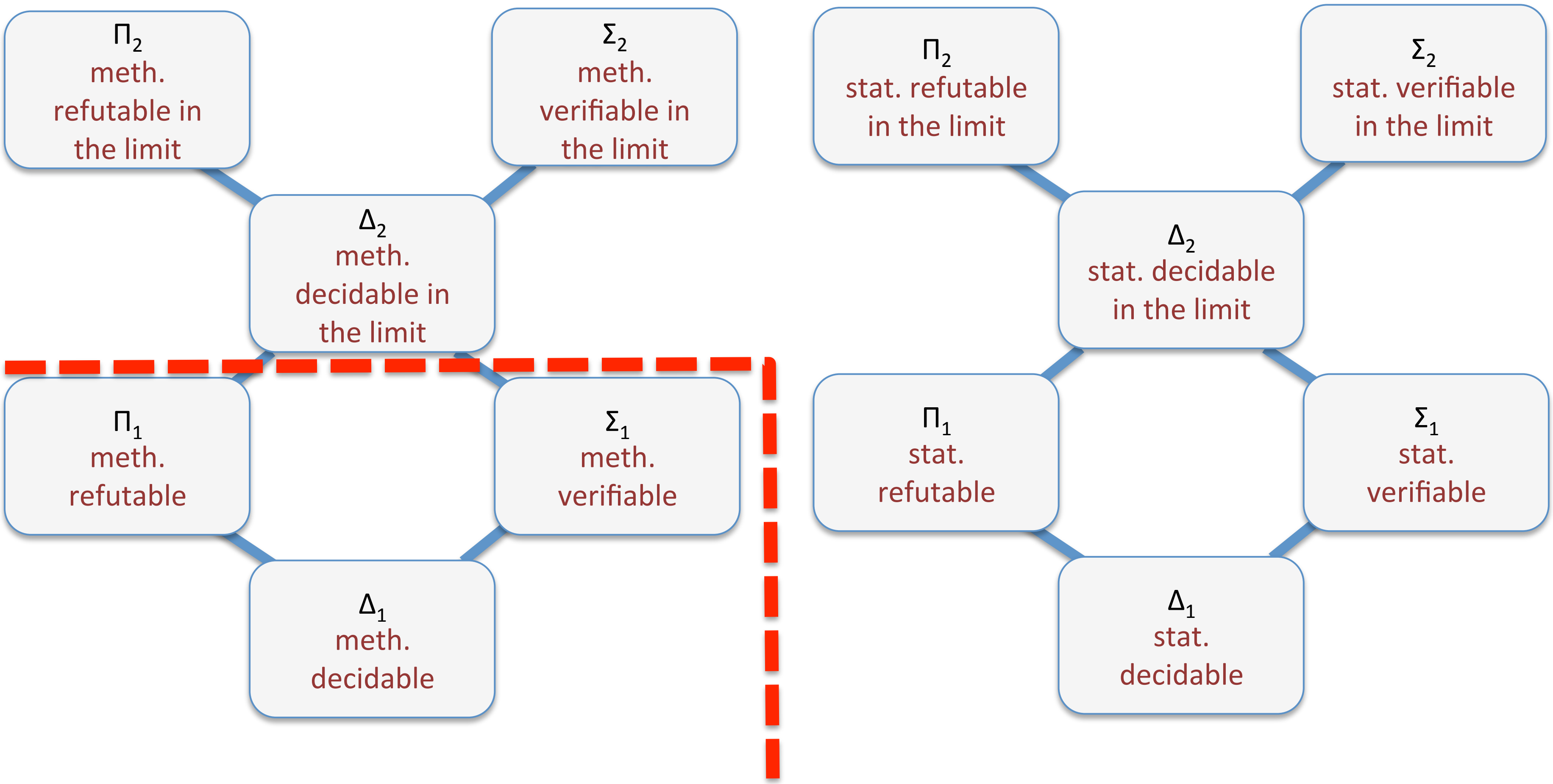
- A **limiting statistical refutation method** for H is a limiting verification method for H^c.
- A **limiting statistical decision method** for H is a limiting verification method and a limiting refutation for H.

Topology and Statistical Methodology

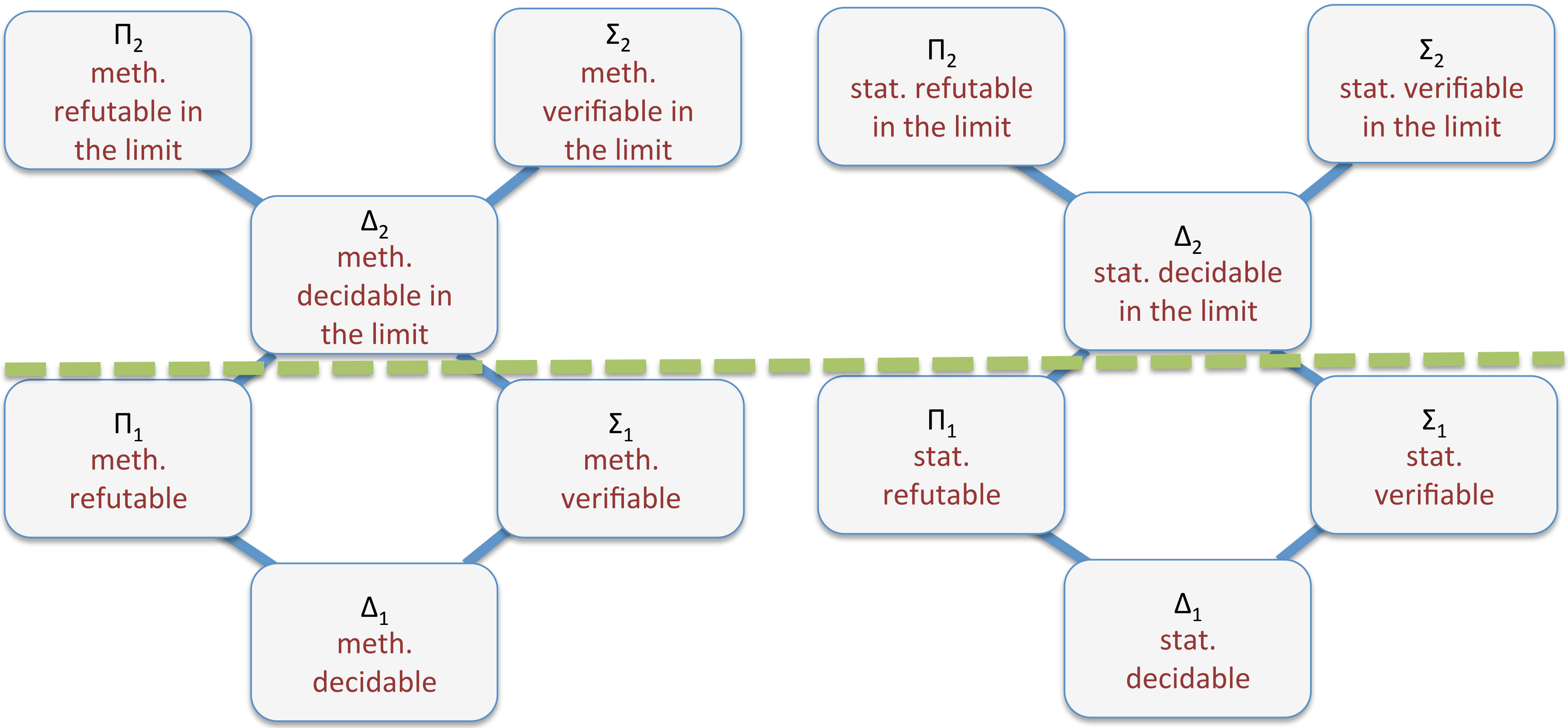
Prop.



Deduction vs. Induction: Wrong



Deduction vs. Induction: Right



Monotonicity

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

$$p_w^{n_2}(M_{n_2} = H) - p_w^{n_1}(M_{n_1} = H) < \alpha,$$

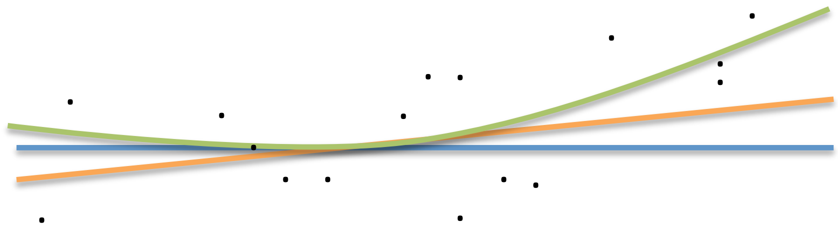
for $n_2 > n_1$.

Topological Simplicity

It still makes sense in terms of statistical information topology!

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

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



Ockham's Statistical Razor

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

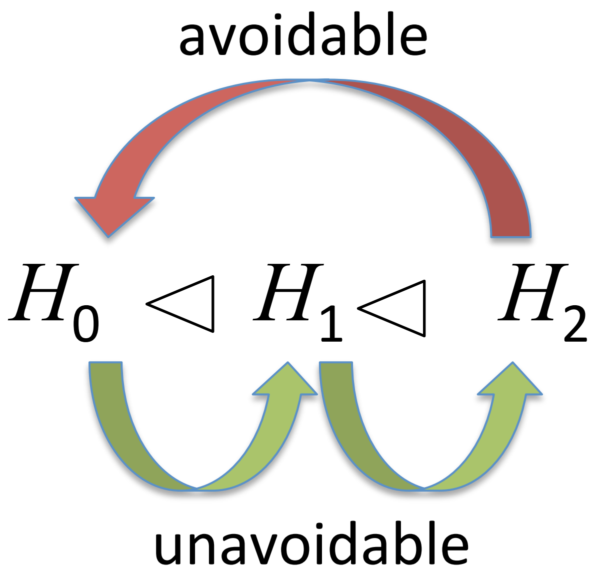
Response: the third formulation of O.R. does not mention compatibility with experience!

3. Infer a relevant response that is more complex than the true answer **with chance < α**.

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.



A New Objective Bayesianism

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

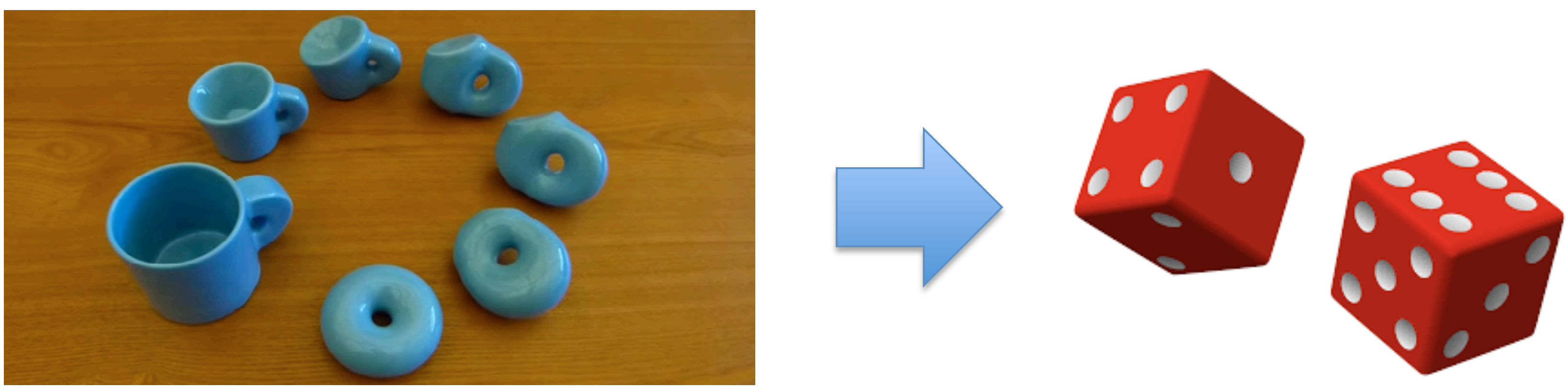
✗ Indifference = ignorance.

✓ truth-conduciveness.

CONCLUSION

A Method for Methodology

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



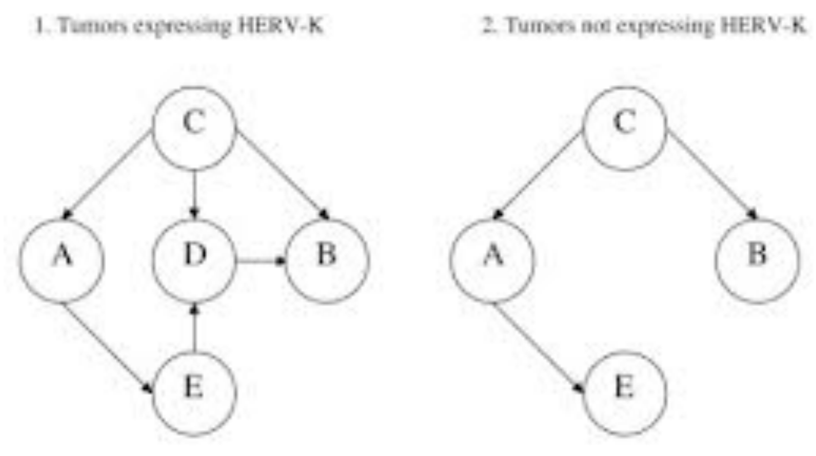
Some Concluding Remarks

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** in a direct and uniform fashion 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.



Symbols: "A" = cause (YFV); "B" = outcome (cancer); "C" = confounders (recreational solar exposure and high social class); "D" and "E" = mediators (HERV-K antigen and immune response).

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.