03. Induction and Confirmation

Topic: Relation between theory and evidence.

1. Induction

David Hume
(1711-1776)

Problem of Induction: What reason do we have for thinking the future will resemble the past?

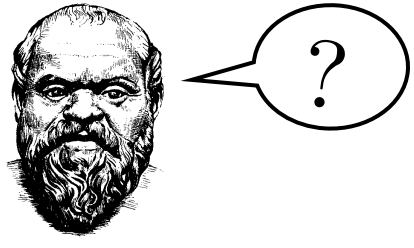
- *Initial Response*: In the past, the future has resembled the past. So shouldn't we expect it to continue to do this?
- *But*: This is circular! It infers future consequences from past consequences; and this is what is being questioned in the first place!

What does "induction" refer to?

- Two types of argument:

Deductive argument: If premises are true, conclusion must be true.

All men are mortal.
Socrates is a man.
∴ Socrates is mortal.



Inductive argument: Truth of premises does not guarantee truth of conclusion.

(a) *Enumerative Induction*. Inference from a finite number of observations to a generalization.

90% of smokers got lung cancer.

$\therefore$ Smoking causes lung cancer.

(b) *Projection*. Inference from finite number of observations to the next case.




Swan #1 observed at time t_1 is white.

Swan #2 observed at time t_2 is white.

$\vdots$

$\therefore$ The next swan observed will be white.










(c) *Explanatory inference* = inference from observations to a hypothesis that best explains them.

Dinosaur extinction event 65 million years ago.

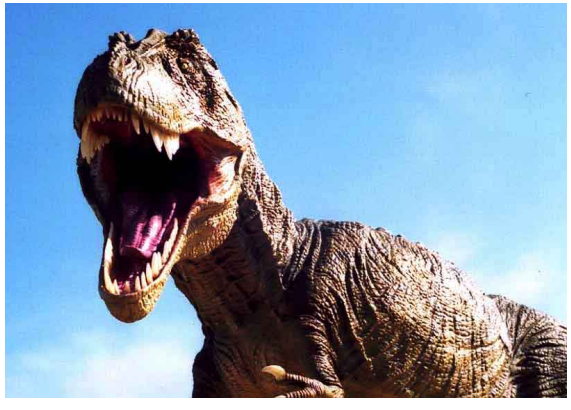
High levels of iridium in layers of Earth's crust from ~65 million years.

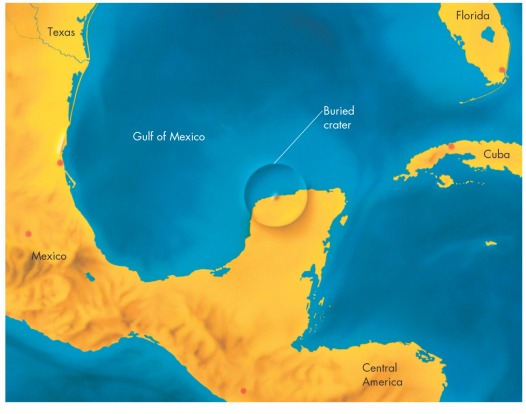
Iridium is commonly found in meterorites.

Impact crater off Yucatan penninsula dates to ~65 million years

---

$\therefore$ A giant meteor impacted the Earth 65 million years ago causing the extinction of the dinosaurs.





- Is one form of induction more fundamental than the others?
- Can a logic of induction be constructed?

2. Confirmation

(A) Hypothetico–Deductivism (HD)

Given a hypothesis H,

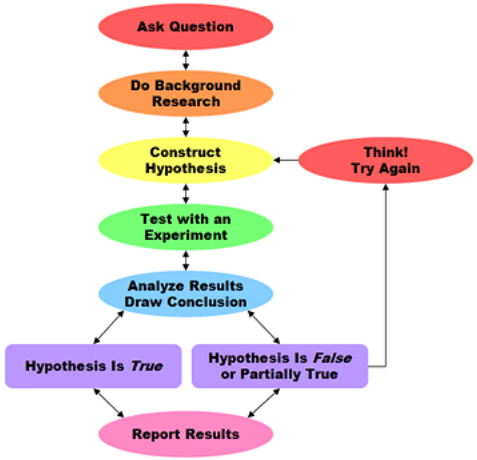
Step 1. Derive a prediction E from H. (*deductive inference*)

Step 2. Test the prediction.

Step 3. If E is true, then H is "confirmed".
If E is false, then H is "disconfirmed". } *inductive inference*

Basic idea: "E confirms H" means "H entails E, and E is true".
HD models confirmation on entailment.

- Is this the "Scientific Method"?



Two Problems with HD

Pierre Duhem (1861-1916)

Willard Quine (1908-2000)

Problem 1. Duhem-Quine Problem

General form of HD reasoning:

If H is true, then E is true.

E is true.

$\therefore$ H is confirmed.

- *But*: To derive a prediction E from H, we need additional assumptions.

Amended form of HD reasoning:

If (H & A_1 & A_2 & ...) are true, then E is true.

E is true.

$\therefore$ (H & A_1 & A_2 & ...) is confirmed.

- *Which* of H, A_1, A_2, ... does E confirm?

Problem 2. Confirmation as Entailment is too weak!

Claim 1: Any true observation S HD-confirms any hypothesis T.

- Let T be any hypothesis, and S be any sentence.
- *Then*: T entails T-*or*-S (whenever T is true, so is T-*or*-S).
- *And*: If S is true, then so is T-*or*-S.
- *So*: If S is true, then T entails T-*or*-S, and T-*or*-S is true.
- *Thus*: If S is true, then T-*or*-S HD-confirms T.
- *But:* This makes confirmation too easy! Let T = special relativity and S = "There are mice in my cupboard".

"E HD-confirms H" means "H entails E, and E is true".

Claim 2: If E HD-confirms T, then E HD-confirms the conjunction of T with any other hypothesis.

- If T entails E, then T & S entails E for any S.
- *So*: If E HD-confirms T, then E HD-confirms T & S.
- *But*: This makes confirmation too easy! Let T = Newtonian mechanics, S = Creationism, E = "The orbits of the planets are ellipses".

(B) Instance Confirmation

Basic idea: "E confirms H" means "E is an *instance* of H".

Ex: H = All ravens are black.
E = A black raven.

Notion of an "instance":
- Assume that all hypotheses in science are of the form "All Fs are Gs".
- An *instance* of a hypothesis is then an F that is also a G.

Problem. "Ravens Paradox"

- "All Fs are Gs" is *logically equivalent* to "All non-Gs are non-Fs".
  - *Whenever "All Fs are Gs" is true, so is "All non-Gs are non-Fs", and vice-versa.*
- *So*: A non-G that is a non-F instance-confirms "All non-Gs are non-Fs", and thus it instance-confirms "All Fs are Gs"!
- *Which means*: A white shoe instance-confirms "All ravens are black"!

Initial Response: Bite the bullet (Hempel's response)

Carl Hempel (1905-1997)

- "All ravens are black" means "If it's a raven, then it's black", which is a claim about *everything* in the universe.
- *So*: A white shoe *does* instance-confirm it, although very minutely.
- *But*: A white shoe is also a non-blue, non-aardvark, so it also instance-confirms "All aardvarks are blue".

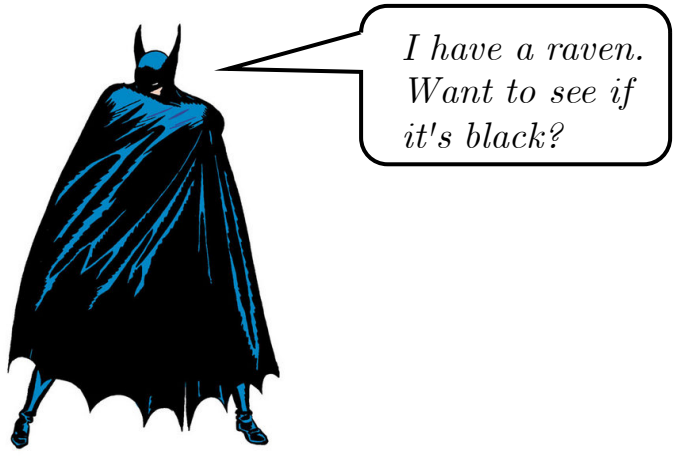
Two more responses:

1. Whether or not an instance confirms a hypothesis may depend on other factors.

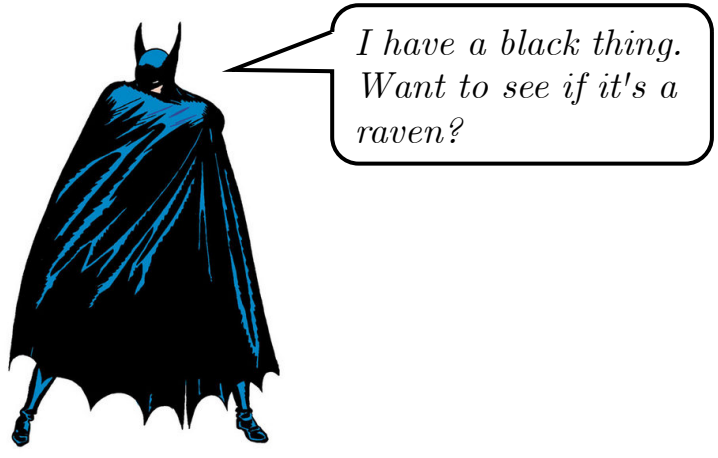
> (i) *All ravens are black and they are extremely rare.*
> (ii) *All ravens are very common, most are black, and a few are white.*

- A black raven observed outside your window (a common sighting) will confirm (ii) but not (i).

2. Whether or not an instance confirms a hypothesis depends on the *potential* for the instance to *refute* the hypothesis.



<u>*You*</u>: Yes! (Your hypothesis is "All ravens are black".)



<u>*You*</u>: No! (Your hypothesis is not "All black things are ravens".)

- What if it's a black raven in both cases?
- Whether it confirms your hypothesis depends on the order in which you discover its properties!

- What about that white shoe?

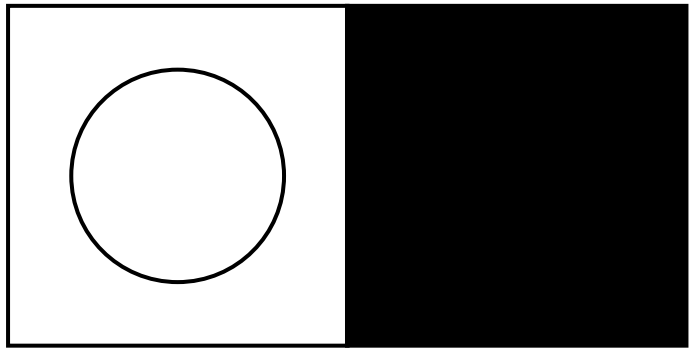
You: Yes! (If it's a raven, then your hypothesis is doomed!)

You: No! (Your hypothesis is not "All black things are ravens".)

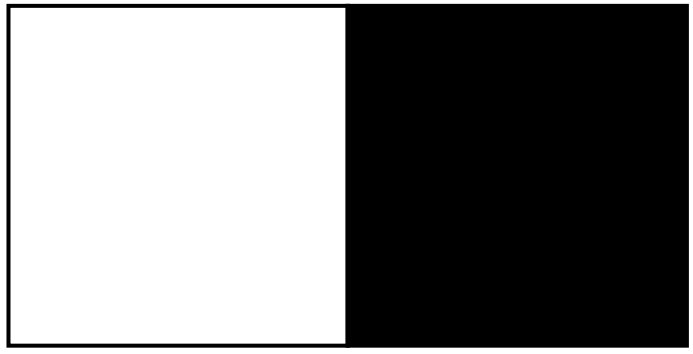
Moral: Some black raven observations confirm "All ravens are black". Some white shoe observations confirm "All ravens are black". Others, in both cases, don't!

General moral:
- Observations are not "automatically" relevant to hypotheses.
- Whether or not they are relevent may depend on their order and on other information.

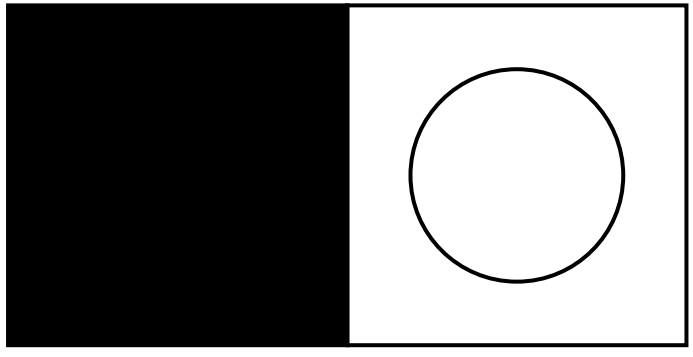
The Selection Task



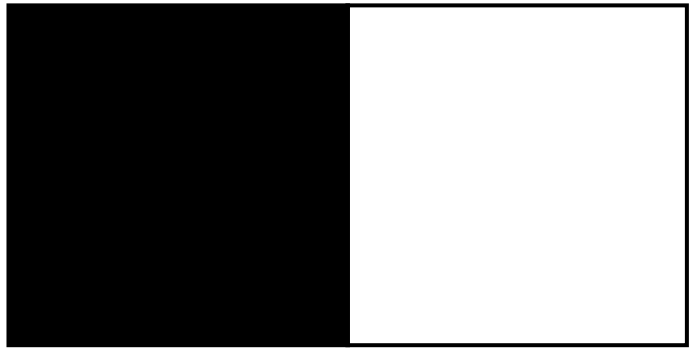
A



B



C



D

- Which masks need to be removed to test the truth of "If there's a circle on the left, then there's a circle on the right"?

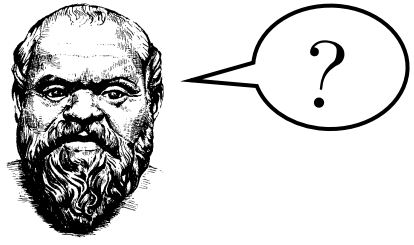
> *Hint*: "If there's a circle on the left, then there's a circle on the right" *is logically equivalent to* "If there isn't a circle on the right, then there isn't a circle on the left".

Goodman's New Riddle of Induction

Claim: There can be no *formal* theory of confirmation.

- *Idea*: Deductive logic is the logic of argument *forms*:

All men are mortal.
Socrates is a man.
Therefore Socrates is mortal.



All Fs are Gs.
a is an F.
Therefore a is a G.

Goodman's claim: Confirmation can't similarly be analyzed at the formal level.

Def. 1 *grue* = green if observed before 2018, or blue if not observed before 2018.

- Many things are grue:

- *Question*: Are *all* emeralds (those that have been observed before 2018 and those *yet* to be observed after 2018) grue?

(A) All observed emeralds prior to 2018 have been green.
Therefore all emeralds are green.

(B) All observed emeralds prior to 2018 have been grue.
Therefore all emeralds are grue.

- (A) and (B) have the same *form*:

  All observed emeralds prior to 2018 have been G.
  Therefore all emeralds are G.

- *But*: (A) seems like a strong inductive argument.
- (B) seems weak: Should we believe that emeralds we've not observed prior to 2018 will be blue if observed after 2018?
- What is wrong with (B)?

(1) *A good theory of induction shouldn't use time-indexed words like "grue".*

- *But*: Whether or not a word is time-indexed is language relative.

> *Def. 1* *grue* = green if observed before 2018, or blue if not observed before 2018.
>
> *Def. 2.* *bleen* = blue if observed before 2018, or green if not observed before 2018.

- In English, "grue" and "bleen" are time-indexed, and "green" and "blue" aren't.
- But why can't there be another language, Blinglish, in which "grue" and "bleen" are primitive, and "green" and "blue" are time-indexed?

> *Def. 3.* *green* = grue if observed before 2018, or bleen if not observed before 2018.
>
> *Def. 4.* *blue* = bleen if obseved before 2018, or grue if not observed before 2018.

- How could we know today whether we speak English or Blinglish?

(2) *Maybe the words we use aren't the problem. Maybe it's the properties they refer to.*

- Greenness is a natural property: it picks out a "natural kind" in nature.
- Grueness is unnatural: it doesn't pick out a natural kind.

- *But*: How do we come to have knowledge of natural kinds?
  - *elements? (hydrogen, helium, lithium, etc.)*
  - *elementary particles? (electron, neutrino, quark, etc.)*
  - *biological species?*
  - *economies with very high inflation?*
  - *mental disorders in the DSM IV?*

- Problem of identifying the right category for prediction and extrapolation.