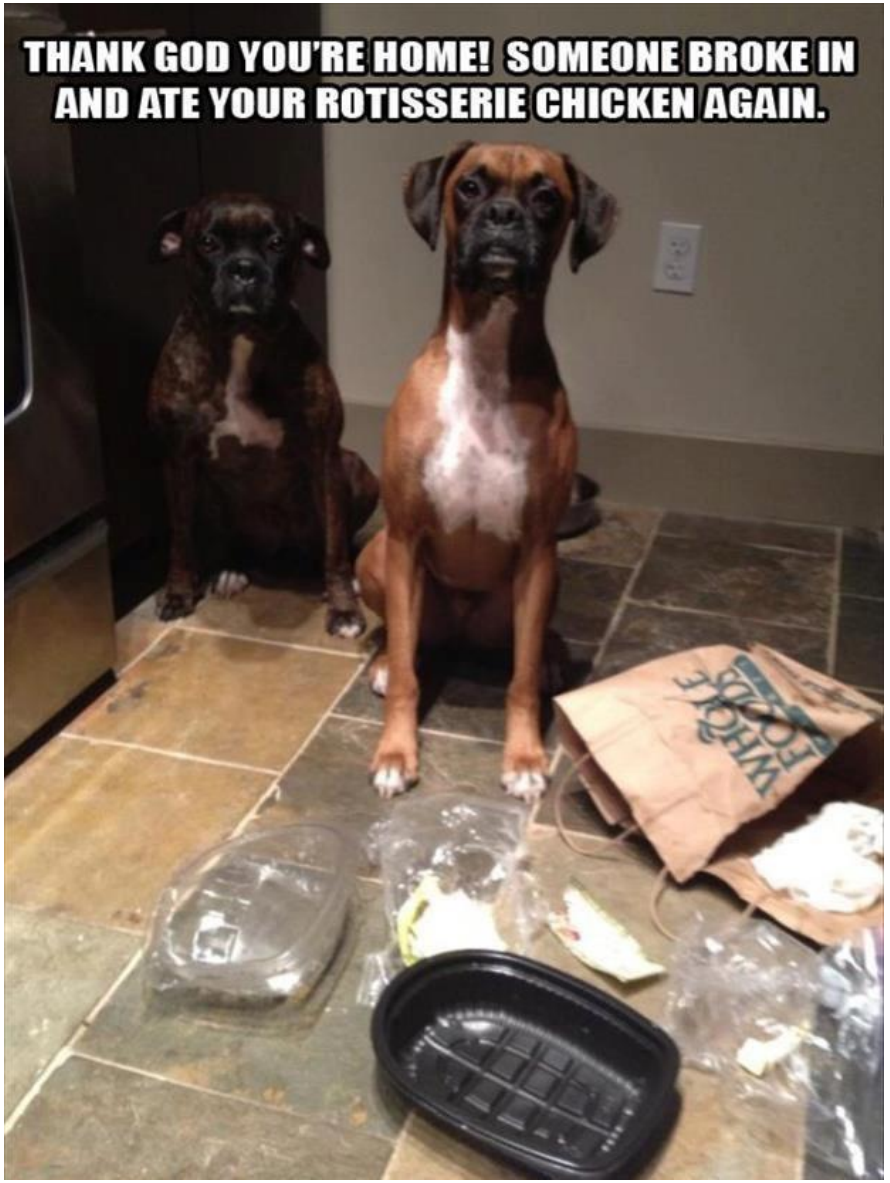
Inference to the Best Explanation

What most likely happened here?



Inference to the Best Explanation

- **Definition**

  An *explanation* is a hypothesis (or story) about what caused an object to exist, or an event to occur.

  An *inference to the best explanation* (IBE) means judging a hypothesis to be probably true, on the basis that it explains the available evidence *better* than any *competing* hypothesis.

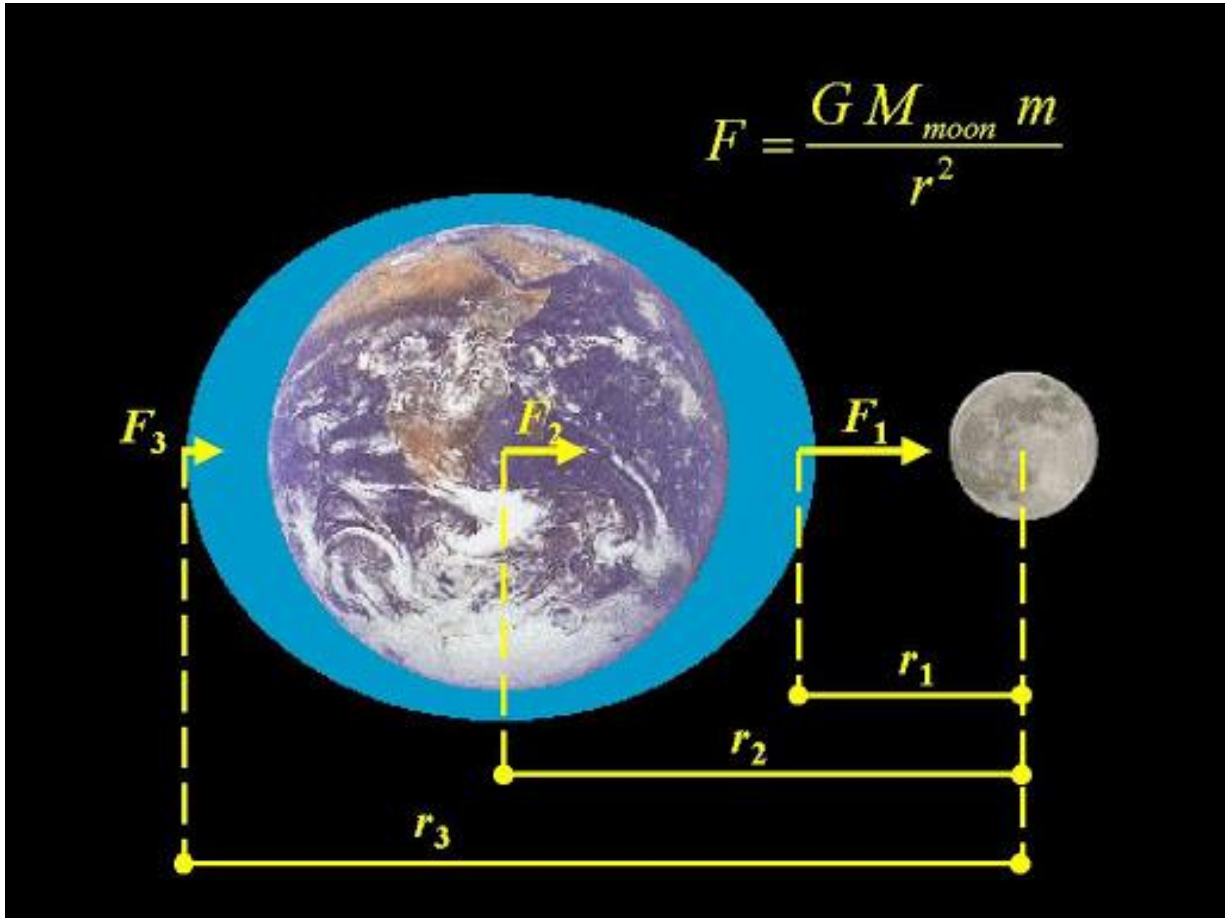
Explanation and Inference

- But there's a bit more to explanation than just pointing to a cause. For example, it was indeed Newton who first correctly explained the tides.
  - But others had previously *claimed* that the moon caused the tides. (E.g. Kepler)

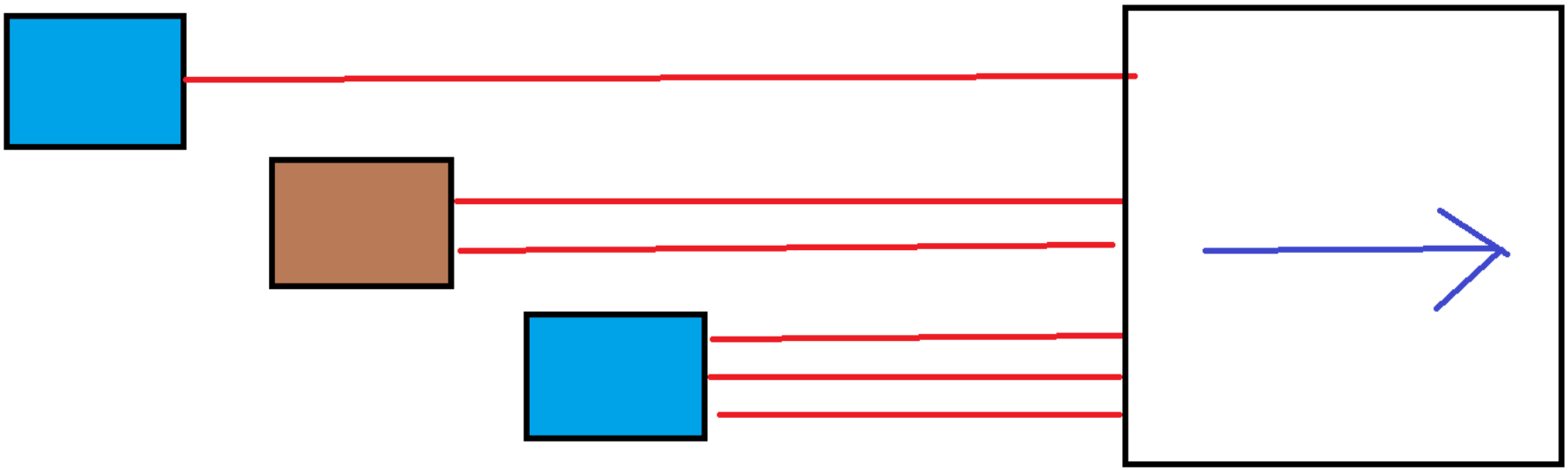
- So, when Newton explained the tides, he had to do more than just say "the moon is doing it".
- What more did he have to do?

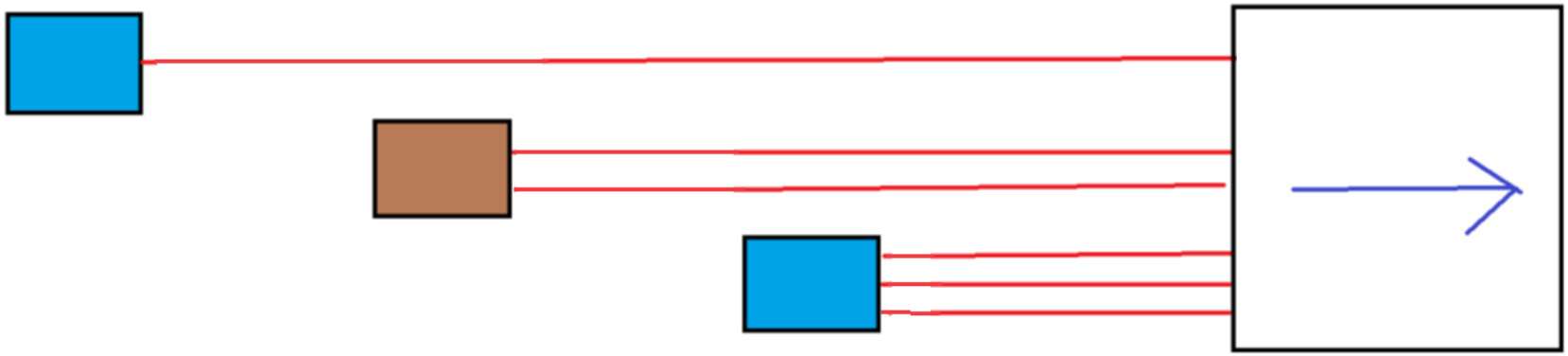
- He had to *show* that, assuming his theory of gravity, and certain features (e.g. mass) of the moon, *one would expect 2 tides per 24 hour period*. And one would expect the oceans to rise and fall so many feet.
- In short, he had to *infer* or *predict* the tides from his (supposed) cause, i.e. the moon's variable gravity.
  - A bunch of hard math!

- Why are there *two* high tides per 24 hours, not just one?



- What happens when a truck accelerates, pulling 3 trailers by bungee cords of different strengths?





E.g. “Monty Hall problem”

- On the TV show *Let’s Make a Deal*, the contestant had to pick one door from a choice of 3, keeping whatever prize lay behind that door.

“Monty Hall problem”

Clearly you just make a random choice, with a 1/3 chance of being right.

But the twist is that Monty Hall (who knows where the prize is, and won’t reveal it) opens one of the other doors (always a “zonk”) revealing (e.g.) a goat or llama.

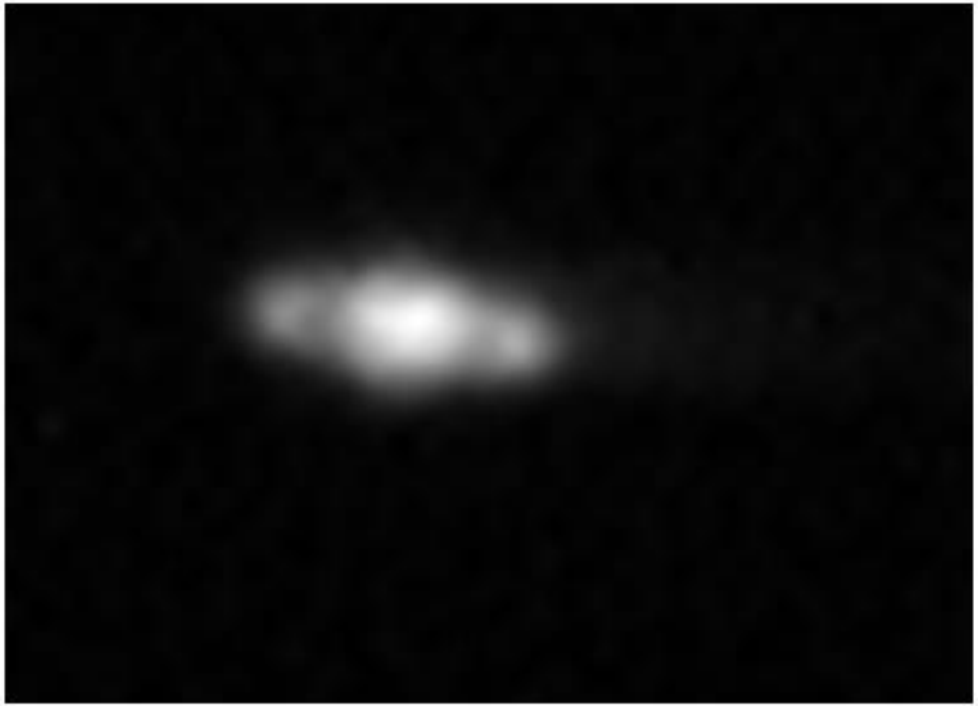
You then have one final chance to change your mind. Either keep the door you have, or switch to the one remaining closed door.

- Which is the rational choice? (Or are they rationally equivalent?)

- This problem is easily solved using the method of inference to the best explanation (IBE) and also by probability theory.

What's up with Saturn?

In 1610 Galileo looked at Saturn through his telescope and saw something like the image below. How do we best explain this data?



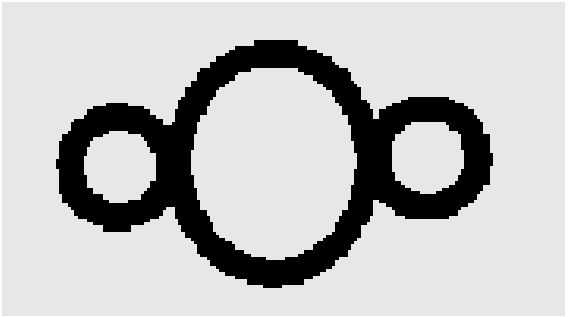
Competing Hypotheses

- H_1: Saturn is a composite of 3 planets, with two equal small planets flanking the main one.
- H_2: Saturn is a giant soup tureen, with handles.
- H_3: Saturn has a flat ring around its equator

1st Hypothesis: a triple planet

On 30 July 1610 Galileo he wrote to his Medici patron:

> "the star of Saturn is not a single star, but is a composite of three, which almost touch each other, never change or move relative to each other, and are arranged in a row along the zodiac, the middle one being three times larger than the lateral ones, and they are situated in this form:

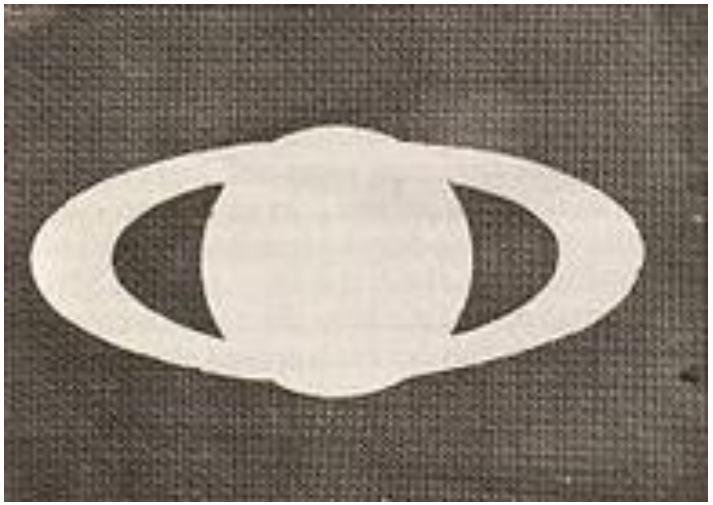


2nd Hypothesis: Giant Soup Tureen

(Galileo never *quite* proposed this theory. But he did say that Saturn appeared to have ‘handles’, or ‘ears’.)

3rd Hypothesis: A Ring

- In 1655, Huygens proposed that Saturn was surrounded by "a thin, flat ring, nowhere touching, and inclined to the ecliptic."



Best explanation?

- There could be more hypotheses, of course, perhaps thousands of them. (Limited only by one's imagination.)

- But suppose that these are the only three. Which of them is the best, and why?

- It would be nice to have some *criteria* for how good an explanation is.

Criteria for a Strong Explanation

1. Causation (or “Production”) Condition
   - A cause of the empirical data must be provided.

2. Prediction (Empirical Adequacy) Condition
   - The proposed cause must predict the empirical data.

3. Prior Plausibility Condition
   - The proposed cause must be plausible in itself.

When is *H* a good explanation of *E*?

1. *Causation* Condition (or "production")

   H makes a claim about something that *caused E*.

   (It may describe the nature of a known cause, or posit the existence of a previously unknown cause.)

- Which of the three hypotheses about Saturn provide a cause of its appearance?

- All of them.

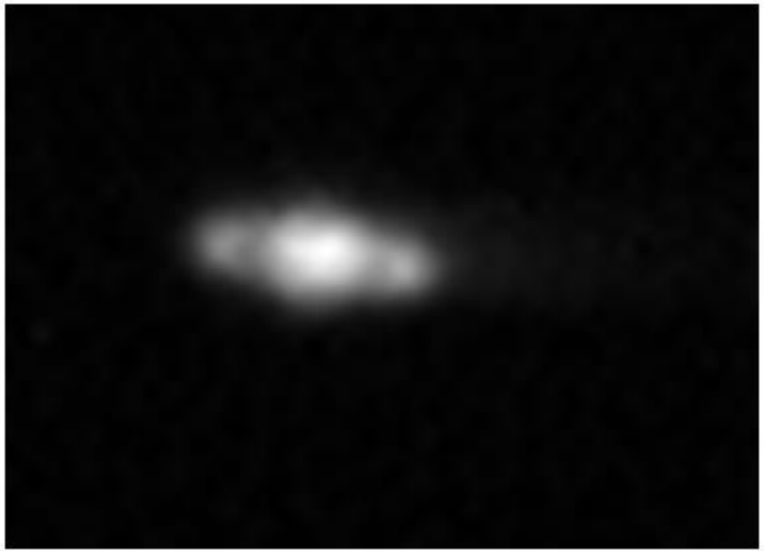
When is *H* a good explanation of *E*?

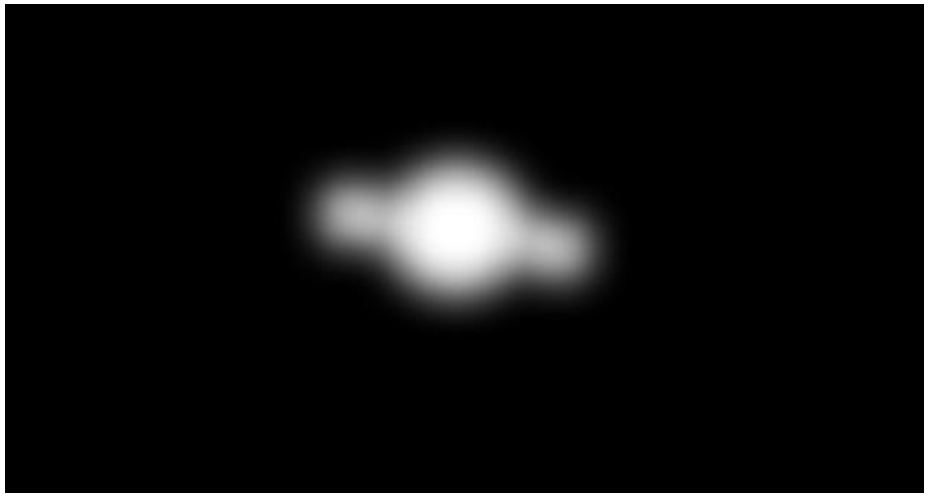
- *Prediction* (empirical adequacy) Condition

  E can be predicted (or inferred) from *H*, to a high degree. In other words, assuming *H* to be true, one would *expect* to observe *E*.

- A theory that predicts the data is *empirically adequate*, and "saves the phenomena".
  - (This is clearly a criterion that *empiricists* will insist on, but rationalists accept it as well.)

Does H_1 predict the data?



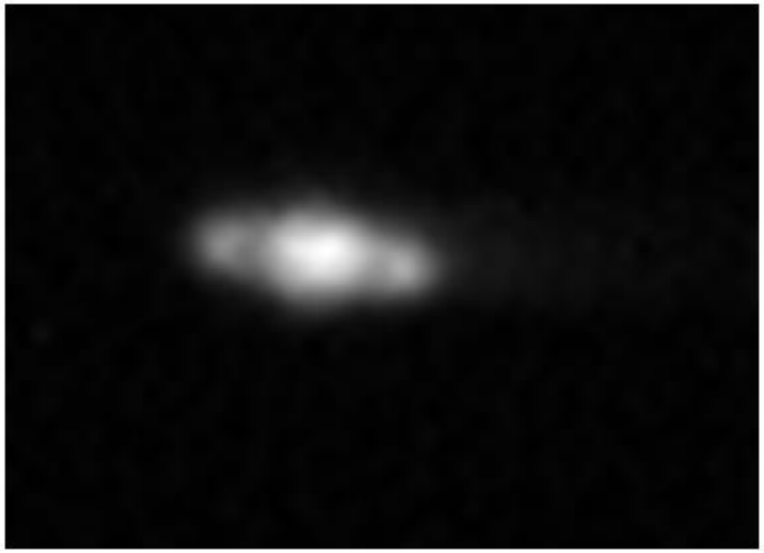


Data

1st theory prediction

Somewhat, but not too great.

Does H_2 predict the data?

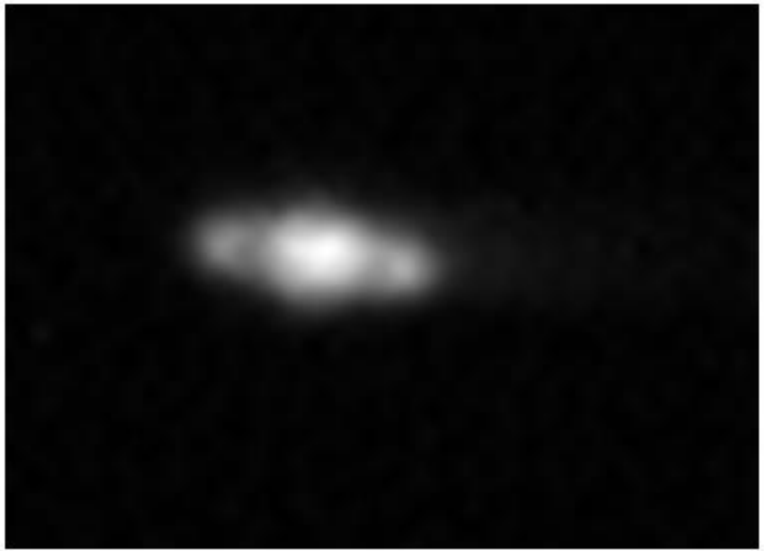


data

2nd theory prediction

A better fit.

Does H_3 predict the data?



data

3rd theory prediction

About as good as H_2.

When is *H* a good explanation of *E*?

- *Prior Plausibility* Condition

  The cause proposed by *H* is plausible, given our background knowledge.

  N.B. The judgement of plausibility has nothing to do with the evidence *E*. This is the *prior* plausibility, *before* we learn about *E*.

  (This criterion is one that *rationalists* will emphasize.)

How do we judge plausibility?

- A plausible hypothesis is one that agrees with what we already believe about the world. I.e. it fits our "background knowledge", or "paradigm". (This background knowledge may include empirical data other than *E*.)
- A *simple* hypothesis is more plausible than a complex one, other things being equal. (Ockham's Razor.)
- A hypothesis involving known causes is more plausible than one involving unknown causes.

- Is it plausible that Saturn is a triple planet?
  - This is a new idea. No other planet is thought to be triple. But Jupiter and the earth have surrounding planets. Prior plausibility ACCEPTABLE
- Is it plausible that Saturn is a soup tureen? Or has ears?
  - This is simply ridiculous, and complicated. Prior plausibility *FAIL*
- Is it plausible that Saturn has a ring?
  - This is also a new idea, and rather odd. But it isn't as ridiculous as handles or ears. JUST BARELY ACCEPTABLE

Overall, which is best?

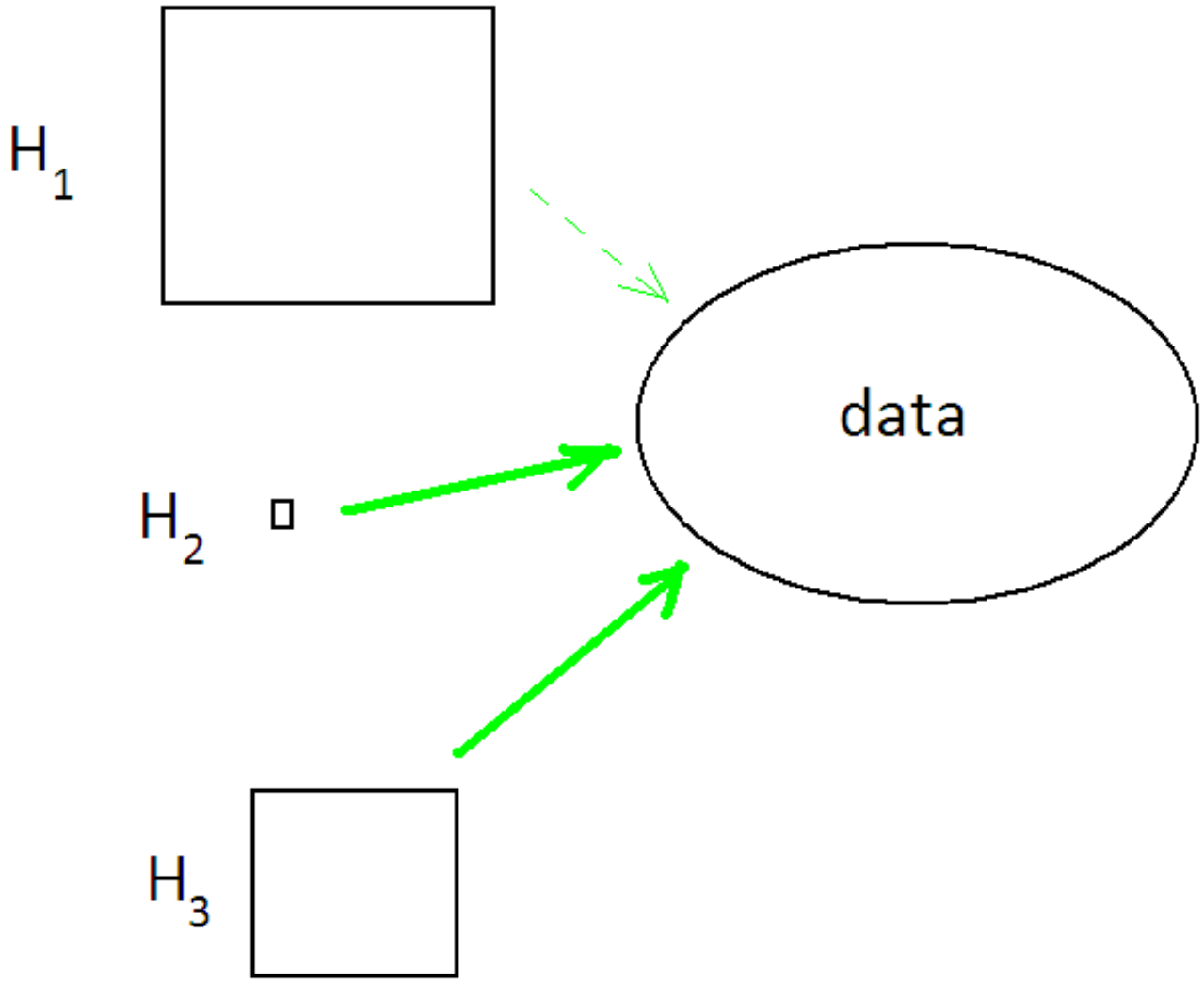
	Cause proposed?	Cause is plausible?	Cause predicts E?
H_1 (triple planet)	Yes	Somewhat	Poorly
H_2 (handles)	Yes	No	Well
H_3 (ring)	Yes	Barely	Well

H_1 is *weak* because it fails to predict the evidence.

H_2 is *weak* because it is implausible.

H_3 is *strongest* because it is barely plausible and predicts the evidence.

$\Rightarrow H_3$ is the best explanation.


H_1
data
H_2
H_3

	Cause proposed?	Cause is plausible?	Cause predicts E?
H_1 (triple)	Yes	Somewhat	Poorly
H_2 (handles)	Yes	No	Well

Which of H_1 and H_2 is stronger, do you think?

If they were the only options, which would *you* believe?

Stick with your priors, or be persuaded by evidence?

Or sit tight and wait for a better theory?

The Monty Hall problem

- Assume that you initially select Door A, and then Monty Hall opens Door B to reveal a Zonk.
- There are now two remaining possible hypotheses:
  - The prize is behind Door A
  - The prize is behind Door C.

- Which one is more probable?
  - The one that best explains why Monty Hall opened Door B.

Degrees of Plausibility

- In reality, the plausibility of a hypothesis isn't a yes/no matter. There are *degrees* of plausibility.

- The degree to which a hypothesis is plausible, prior to the (new) evidence, is called its *prior probability*.

- Relative to background knowledge *K*, we can write the prior probability of H as $P_K(H)$.

A scientist's sense of plausibility is fallible …

"In 1825, Mr. McEnery, of Torquay, discovered worked flints along with the remains of extinct animals in the celebrated Kent's Hole Cavern, but his account of his discoveries was simply laughed at. In 1840, one of our first geologists, Mr. Godwin Austin, brought this matter before the Geological Society, and Mr. Vivian, of Torquay, sent in a paper fully confirming Mr. McEnery's discoveries, but it was thought too improbable to be published. Fourteen years later, the Torquay Natural History Society made further observations, entirely confirming the previous ones, and sent an account of them to the Geological Society of London, but the paper was rejected as too improbable for publication. "

- (From A. R. Wallace, 1870)

Degrees of Prediction

- In a similar way, there are in reality degrees to which a hypothesis predicts a piece of evidence. (Some hypotheses predict the evidence more strongly than others.)

- The degree to which a hypothesis predicts the evidence is called the *likelihood* of the evidence, under that hypothesis.

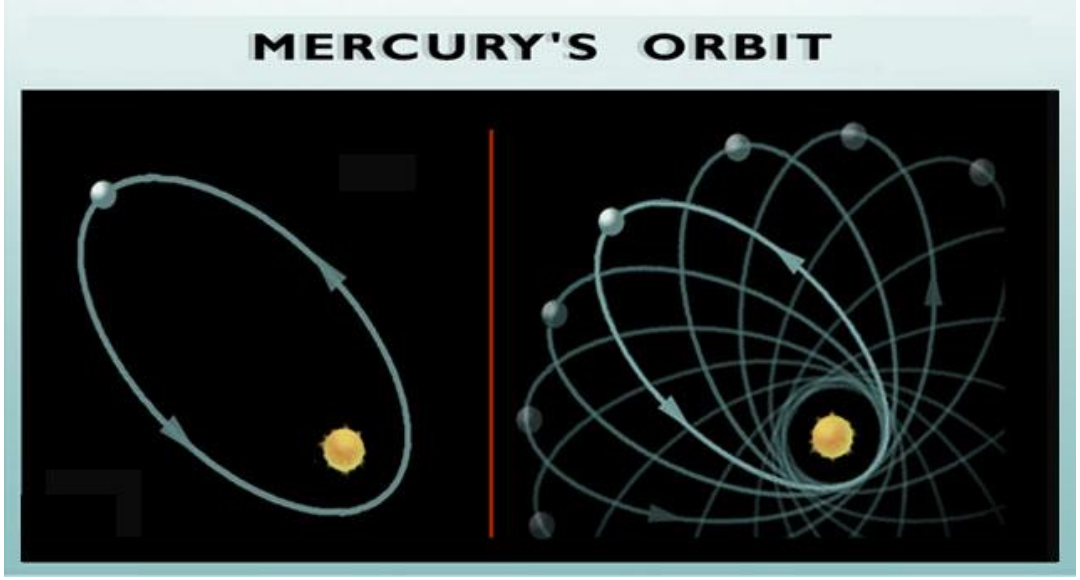
- The likelihood of E under H can be written $P_K(E \mid H)$.

“Predicting” old data?

- This technical sense of “H predicts E”, that $P_K(E \mid H)$ takes a high value, is a little different from ordinary prediction.
- For example, a theory H may “predict” (in this technical sense) *old data*, i.e. data that was already known before H was proposed.
- “H predicts E” just means that E may be logically inferred from H, together with background knowledge *K*.

E.g. Mercury's perihelion shift

- In 1915 Einstein used relativity theory (GTR) to correctly "predict" the perihelion shift of Mercury to be 574 arc-seconds per century (Newtonian mechanics predicts 531 arc-seconds). But the observed value was known since 1859! (Is GTR confirmed by this?)



Part 2

The Strength of an Explanation

Strength of an Explanation

- The overall strength of an a hypothesis H, as an explanation of E, relative to background knowledge K, is:

  $$\text{Strength}(H) = P_K(H) \times P_K(E \mid H).$$

  (Strength = plausibility $\times$ empirical adequacy)

  In other words, a strong hypothesis has to be plausible *and* predict the evidence.

Simple example

- Suppose a box contains two coins, A and B.
  - Coin A has *Chance*(heads) = 2/3
  - Coin B has *Chance*(heads) = 1/3

- A coin is drawn at random from the box, and tossed once. It landed tails.
- *Which coin was it?* Coin A or coin B?
- Coin B, most likely. Priors are the same for both hypotheses. The coin-B hypothesis *predicts the evidence twice as strongly*.

***Strength is absolute*, not relative to other hypotheses**

- The strength of a hypothesis H, as an explanation E, depends only on H, E, and your background information.

- The strength of an explanation is an absolute measure, not a relative one. (The “best” explanation is relative, of course.)

Avoid these fallacies

- Hence, one should never argue that:
  - H predicts E, since the only alternative, H', doesn't predict E.
  - H is plausible, since H' is implausible.
  - H does not predict E, since H' predicts E better.

(Etc.)

Strong explanation vs. true explanation

- A strong explanation need not be true, and a true explanation need not be strong.

- E.g. the Vulcan explanation for Mercury's anomalous perihelion shift was pretty good (at the time) yet false. (The planet Vulcan does not exist.)
- Also, a true explanation need not be good. Some observed patterns are due simply to chance. Yet chance is always a poor explanation, since it only very weakly predicts the evidence.

- - Lincoln was elected to Congress in 1846
  - Kennedy was elected to Congress in 1946

  - Lincoln was elected president in 1860
  - Kennedy was elected president in 1960

  - Kennedy had a secretary named Lincoln
  - Kennedy was shot in a car named Lincoln

  - Lincoln was succeeded, after assassination, by vice-president Johnson
  - Kennedy was succeeded, after assassination, by vice-president Johnson

  - Andrew Johnson was born in 1808
  - Lyndon Johnson was born in 1908

Year 1981

1. Prince Charles got married.
2. Liverpool won the UEFA Champions league.
3. The Pope Died.

Year 2005

1. Prince Charles got married.
2. Liverpool won the UEFA Champions league.
3. The Pope Died.

If Prince Charles decides to remarry and Liverpool make it to the final then someone should warn the Pope...!

The best explanation may be weak

- Imagine you flip a coin 100 times, and get the outcome:

  hhhthhttttththhhhhthttthtthhhhthtthhhttttthhthtttht
  hthhhhhhhttttthtthtthhthtthttthhthhthtthtthhthhtttt
  hhthth.

- Is there a good explanation for this *exact* sequence?

- No. It isn't predicted by *any* plausible theory.

- The coin-tossing case shows that some data are very difficult to explain. There is no good explanation of them.

- Yet there is a *true* explanation, and quite often a *best* explanation (chance in this case).

What if every possible explanation is weak?

"When you have eliminated the impossible, whatever remains, however improbable, must be the truth" (Sherlock Holmes)

Bayesian version:

"When you have eliminated the absurdly weak explanations then whatever remains, even if it's rather weak, is probably true."

Do you agree?

Bayes' theorem

$$P(H_1|E) = \frac{\text{strength}(H_1)}{\text{strength}(H_1) + \text{strength}(H_2) + \ldots}$$

- E.g. if strength(H_1) = 10^{-6}, and strength(H_2) = 10^{-9}, then

$$P(H_1|E) = \frac{10^{-6}}{10^{-6} + 10^{-9} + \ldots \text{ (even smaller)}} = 0.999$$

The Flaws in Bayesianism

1. There are cases in the history of science where (for a long period) no one even *thought* of the true explanation.

2. In other cases, the truth had been thought of, but dismissed as too implausible due to mistaken background ideas.

(N.B. Bayes's theorem is based on the assumption that we have thought of all the hypotheses, and that our background knowledge is correct.)

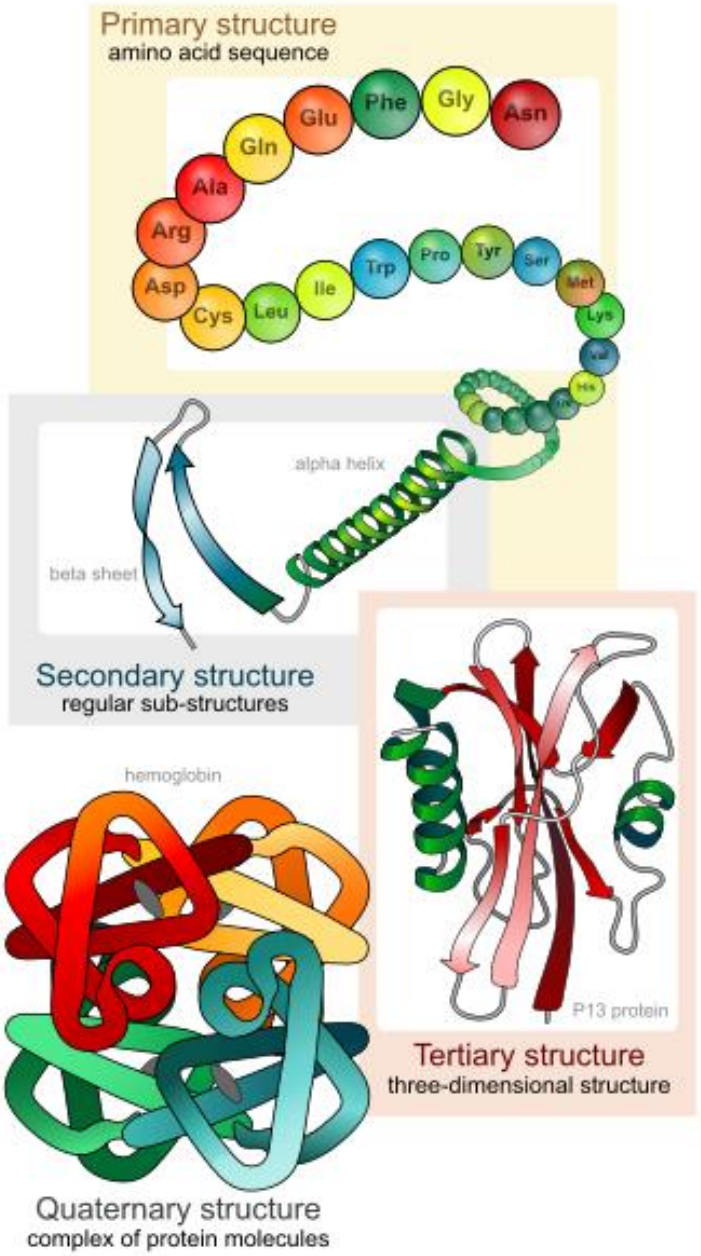
When the best explanation (we can think of) is very weak, should we regard it as probably true? Or should we guess that either (1) or (2) above applies?

Part 3

Case Study: The Origin of Life

The Structure of Life

- Living organisms are built out of proteins, which are *polymers*, i.e. long chain molecules made by connecting units together.
  - The units in these chains are amino acids, There are 20 different kinds of amino acid used in proteins.
- In order to function, a protein needs the right sequence of amino acids.
  - Problem: even if the right amino acids were available, how did they assemble themselves into the right sequence? Given the known laws of chemistry, it seems absurdly improbable.


Primary structure
amino acid sequence
Phe
Gly
Asn
Glu
Gln
Ala
Arg
Asp
Cys
Leu
Ile
Trp
Pro
Tyr
Ser
Met
Lys
alpha helix
beta sheet
Secondary structure
regular sub-structures
P13 protein
Tertiary structure
three-dimensional structure
hemoglobin
Quaternary structure
complex of protein molecules

DNA, RNA, etc.

- Well, how do proteins assemble in our bodies? They are built by machines, such as the ribosome, made from proteins (!)
- Also, the construction involves *other* polymers, such as DNA and RNA, that seem equally improbable.
- The question then is how proteins *first* formed without such aids.

- “Anyone who tells you that he or she knows how life started on earth some 3.4 billion years ago is a fool or a knave. Nobody knows.”

Stuart Kauffman, *At Home in the Universe* (New York: Oxford University Press, 1995), p. 31.

- “Although a biologist, I must confess that I do not understand how life came about. . . . I consider that life only starts at the level of a functional cell. The most primitive cell may require at least several hundred different specific biological macro-molecules. How such already quite complex structures may have come together, remains a mystery to me.”

  Werner Arber, microbiologist and Nobel laureate.

- “One has only to contemplate the magnitude of this task to conclude that the spontaneous generation of a living organism is impossible. Yet here we are -- as a result, I believe, of spontaneous generation.”

  George Wald, "The Origins of Life," in *The Physics and Chemistry of Life* (Simon & Schuster, 1955), p. 270.

  Isn’t it irrational to believe in an “impossible” theory?

Jim Tour criticises OOL research

The Possibilities

1. Abiogenesis is reasonably probable. We're presently unable to see this, since we are ignorant of certain physical/chemical facts.
2. Abiogenesis is fantastically improbable, yet it happened. Improbable events do occur every day! (Especially in a large universe.)
3. Abiogenesis didn't happen. Maybe God did it?

- What is it rational to believe in this case?
- And why?

The Battle: Priors vs. Likelihoods

- In some cases we have a choice between two weak explanations, but they're weak in different ways.

- One is implausible, but the other doesn't predict the evidence (or only very weakly)

- It's then a kind of tug-of-war.

Part 4

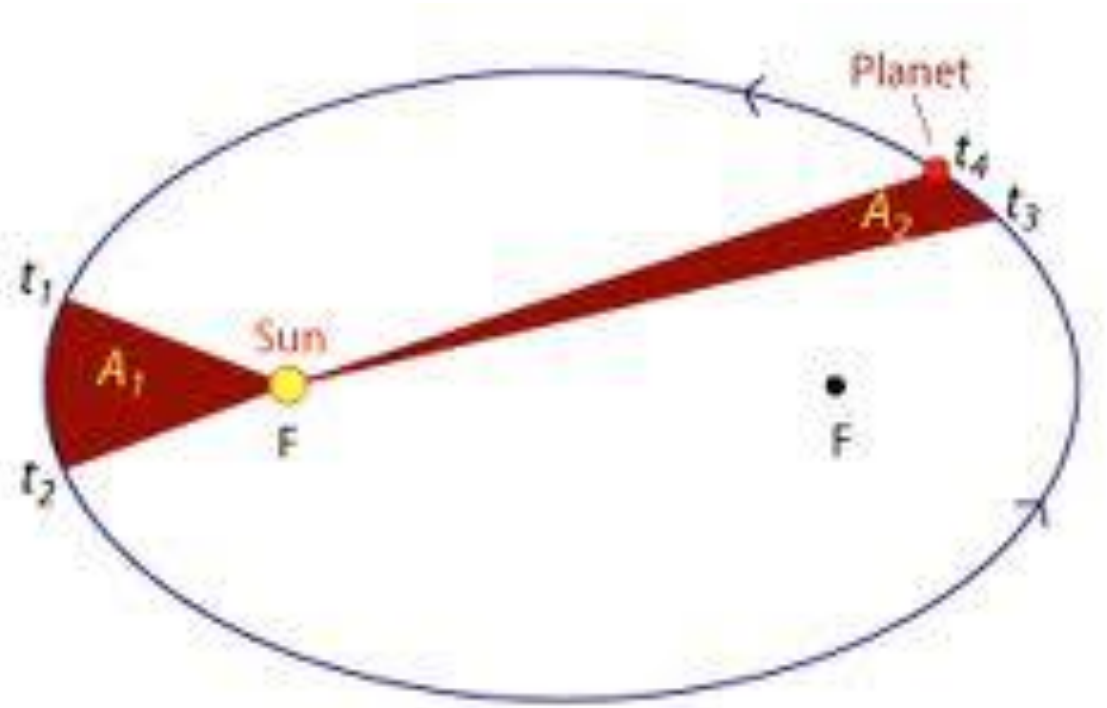
What if *we don't know* what a theory predicts?

- In many cases it is *not known* whether or not the hypothesis predicts a certain piece of evidence.
- What do we do then?

Example: Newton and the Moon

- After Newton formulated his laws of motion, and “inverse square” law of gravity, he tried to predict the motions of the planets from them.
  - He predicted that each planet would orbit the sun along an ellipse, with the sun at one focus.
  - He predicted Kepler’s Equal Areas Law.
- Predictive success!

Example: Newton and the Moon



Kepler's Second Law:
Law of Equal Areas
(greatly exaggerated)

- The lunar motion was more complex, however, since the moon is affected by both the earth and the sun.
- Even as late as 1740, certain features of the lunar motion could not be predicted from Newton's Laws.
  - (The problem was not the laws themselves, as it turned out, but the inability of mathematicians to solve the equations.)
- *In 1740*, were Newton's Laws a good explanation of the lunar orbit?

The need for *demonstration*

- How do you know, in a given case, whether H predicts E?

- What if, for example, Newton had confidently asserted that his theory predicts the lunar motion?
- Not good enough. We need a real mathematical demonstration.

Can “chance” be an explanation?

- N.B. Chance doesn’t mean no cause. It means a stochastic cause, i.e. a cause that doesn’t determine the effect. So the causation condition is OK.
- But the chance of the effect is often small (e.g. in origin of life). In such cases, the hypothesis is poor on the *prediction* condition, i.e. it has very low empirical adequacy.

- E.g. Richard Dawkins in *The Blind Watchmaker*.

- Dawkins says that Darwinian evolution *cannot* be an unguided random process, since in that case life would be a statistical miracle.
  - E.g. producing one part of the protein haemoglobin by a purely random process has a chance $\approx 10^{-190}$. (p. 45)

- A theory of evolution that appealed too much to chance would be very weak – it would fail the prediction condition.