Coincidences and How to Think about Them

Elliott Sober

Examples – Mere Coincidence or …?

- Fir trees are green and so are iguanas.
- The King James Bible was published in the year that Shakespeare turned 46. In Psalm 46, the 46th word is "shake" and the 46th word from the end is “spear”.
- Evelyn Marie Adams won the New Jersey lottery twice.
- Two of Sally Clark’s children died of symptoms indistinguishable from SIDS.

What is a coincidence?

- Diaconis and Mosteller (1989) give this working definition: it as “a surprising concurrence of events, perceived as meaningfully related, with no apparent causal connection.”
- This definition makes coincidence a subjective category. Is it possible to do better?

Let's distinguish coinciding observations and coincidence explanations.

- Two observations coincide when they are similar (e.g., the color of fir trees coincides with the color of iguanas, or two people arrive at the same street corner at the same time).
- The coinciding of two observations is a "mere coincidence" when there is no causal connection between them.

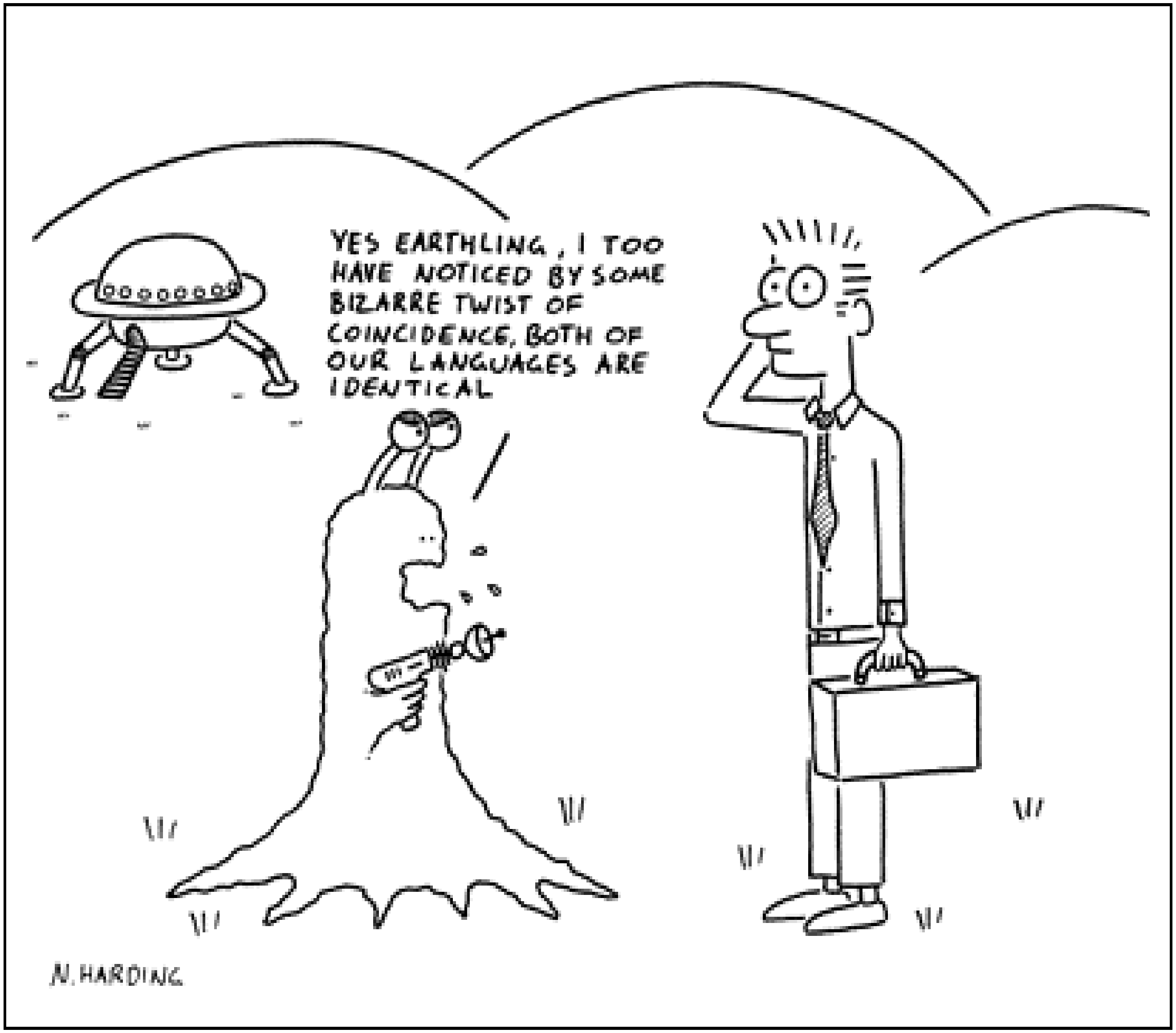
Two events are causally connected when one causes the other or they are effects of a common cause.

- "It is not a coincidence that offspring resemble their parents and it is not a coincidence that siblings resemble each other."
- This definition of coincidence explanation makes it objective, if causation is objective.

When should the coinciding of two observations not be dismissed as a mere coincidence?

- That is, when should you infer that they are causally connected?
- Suggestion #1: when their co-occurrence would be too improbable if they were causally unconnected – that is, when the hypothesis that “it’s just a coincidence” strains your credulity too much.

The principle at work



Suggestion #1 = Use <u>probabilistic modus tollens</u>

- The coinciding observations (the two languages are similar in numerous ways).
- Pr(the coinciding observations | the two languages are causally unrelated) = tiny

----

Reject the hypothesis that the two languages are causally unrelated.

Prob MT leads to paranoia -- to finding conspiracies everywhere

- Example: What is the probability that Adams would win the lottery last week and this week too, if this was just a coincidence? (1 in 17 million)
- Suggestion #2: Redescribe the observations -- for example, as someone wins two lotteries in the same state in the whole history of state lotteries. This is a logically weaker observation and so is more probable.

Another example of Suggestion #2

- My birthday (06061948) appears at the 16,769,633th position of 👌. The probability of this is very small. Is it small enough to reject the hypothesis that this is a coincidence?
- Suggestion #2 allows you to duck this question by weakening the description of the observation. What I should ask instead is: what is the probability of my birthday's appearing somewhere in the 1st 100 million digits? It's about 2/3.

The two suggestions summarized

(S1) When two or more observations coincide, use <u>probabilistic modus tollens</u> to see if you should reject the hypothesis that their similarity is "just a coincidence."

(S2) If you think the coinciding of the observations <u>is</u> "just a coincidence," but the coincidence hypothesis says that the observations are very improbable, weaken your description of the observations.

Both suggestions are wrong

(S1) Prob MT is incorrect. Example: forensic DNA. Suppose two individuals are examined at 12 loci and they have the same rare genotype at each. Pr(match | sibs) is tiny, but that doesn't show that you should reject the sib hypothesis.

(S2) Weakening the observations at will violates the Principle of Total Evidence. It also means that anything goes. Example: the Martians who speak English.

Another approach: the Law of Likelihood + the Principle of Total Evidence

- LL: The observations O favor H_1 over H_2 if and only $Pr(O \mid H_1) > Pr(O \mid H_2)$.
- Don't confuse the likelihood of H [Pr(O | H)] with the probability of H [Pr(H | O)] – think about gremlins.
- LL helps you evaluate what the evidence says; it doesn't tell you what to accept or reject. The embarrassing question that attaches to Prob-MT no longer arises.

How to apply the Law of Likelihood to the problem of interpreting observations that coincide?

- When two observations (O) coincide (e.g., Martians and we both speak English), consider two hypotheses: CC = the two are causally connected. Indep = they are causally independent of each other.
- Compare Pr(O | CC) and Pr(O | Indep).
- The example of forensic DNA …

Problem with the Likelihood Approach

Implausible “conspiracy theories” often have higher likelihoods than hypotheses that say that the coinciding of the observations is just a coincidence.

Example: Adams won the lottery twice.

A Possible Solution -- Bayesianism

- Consider, not just Pr(O | CC) and Pr(O | Indep), but the prior probabilities Pr(CC) and Pr(Indep) as well.
- Pr(CC | O) > Pr(Indep | O) iff

  Pr(O | CC)Pr(CC) > Pr(O | Indep)Pr(Indep).
- Example: the case of Sally Clark.

A Bayesian analysis of the case of Sally Clark (due in part to Phil Dawid)

- Two hypotheses: Indep = the two children had SIDS. CC = Clark smothered them. O = the observations yielded by the post mortems.
- Pr(O | Indep) ≈ Pr(O | CC).
- Pr(Indep) > Pr(CC).
- Therefore: Pr(Indep | O) > Pr(CC | O).

Is Bayesianism a complete solution to the problem of thinking about coincidences?

- When prior probabilities and likelihoods are objective, Bayesianism provides an objective solution.
- However, the hypotheses considered in science often do not have objective prior probabilities.

Example: the tides and the phases of the moon

- That the tides coincide with the phases of the moon has been noticed for centuries.
- Newton's theory of gravitation explains this by showing how the moon causes the tides.
- Pr(moon-tide association | NT) >>

  Pr(moon-tide association | INDEP).
- But what of Pr(NT) and Pr(INDEP)?

Continental Drift – Alfred Wegener observed that …

The wiggles in the East coast of South America and those in the West coast **of Africa coincide.**

The geological strata down the two coasts are correlated.

The distributions of extant and fossil species down the two coasts are similar.

Likelihood favors C-Drift over C-Stasis, but what about the Priors?

- Pr(data | C-Drift) >> Pr(data | C-Stasis).
- What about Pr(C-Drift) and Pr(C-Stasis)?
- Many physicists rejected C-Drift until a possible mechanism was described (plate tectonics). Many others thought the evidence was already overwhelming.
- Did the two groups have different priors? If so, who was right?

A New Approach: Model Selection and a Larger Data Set

- Consider a very big data set – the complete record of who bought which tickets, and which tickets won, on all the New Jersey lotteries to date. Adams' double win is in this set, but so are lots of other facts
- There are many models we might consider.

Some Models

(Fair) Pr(t wins | t was purchased in lottery i) = α_i

(PL) Pr(t wins | t was purchased in lottery i by player j) = β_{ij}

(PLT) Pr(t wins | t was the kth ticket purchased in lottery i by player j) = γ_{ijk}

(One) Pr(t wins | t was purchased in a lottery) = δ

Three facts about these models

- They are nested: One →Fair → PL → PLT. Stronger models are special cases of weaker models.
- Different parameters in a model can have different values, but need not.
- Models don’t make predictions; fitted models do.

Akaike's 2 Innovations

- The goal: find models that make accurate predictions when fitted to old data.
- The means: an unbiased estimate of the predictive accuracy of model M ≈

  Pr[data | L(M)] – k. This is A's theorem.

Consequences of applying AIC

- The most complex model M (the PLT model) is such that Pr[data | L(M)] = 1, but M also pays the largest penalty for complexity.
- Simpler models may well score better overall. Maybe Fair is the best model.

Three properties of conspiracy theories

- They usually are developed by focusing on a very narrow data set.
- They usually are constructed so as to confer on that data a probability of unity.
- When asked to address a larger data set, they seem to be part of a larger model that contains many parameters.

Summary – how to reason about coincidences?

- Prob MT is a bad idea, as is violating the Principle of Total Evidence.
- The Law of Likelihood is better, but it is subject to a problem – how can one identify what is wrong with absurd conspiracy theories?
- Bayesianism can solve this problem when prior probabilities are objective.
- When priors are not objective, AIC can sometimes help determine when coinciding observations should be dismissed as a "mere coincidence."