Commentary on Jan Sprenger's "A Confirmation-Theoretic Guide to Explanation"

Justin C. Fisher
University of British Columbia

What makes a proposed explanation good?

- Truth
  - Hard to achieve
  - Hard to know when you've achieved it.
- Relevance
  - Are more relevant explanations *more useful*?
  - Are they *more strongly confirmable*?
  - Or are we just trying to fit intuitions about relevance?
- Other good-making features?

What makes a proposed explanation good?

- Truth
  - Hard to achieve
  - Hard to know when you’ve achieved it.
- Relevance
  - Are more relevant explanations *more useful*?
  - Are they *more strongly confirmable*?
  - Or are we just trying to fit intuitions about relevance?
- Other good-making features?

Jan's Proposal

"relative to the actual background knowledge [K] of the subject, knowing the explanans [C] boosts the likelihood of the explanandum [E]."

- I.e., a proposed explanation is more explanatorily relevant, the more it boosts the probability of its explanandum.

$$r_d = P(E \mid C \,.\, K) - P(E \mid K)$$

- – Violates objectivity desideratum [r_d varies with $P(C|K)$.]

$$r_s = P(E \mid C \,.\, K) - P(E \mid \neg C \,.\, K)$$

- – Jan notes this also violates objectivity! So why favor it?

- If $P(C|K)$ is small, as it often will be, then $r_d \approx r_s$.
- Even if we grant that r_s is the best game currently in town, that doesn't mean we have to play it.

Why is the sky blue?

E = “Daytime clear skies on earth are blue”

K = Our background knowledge

C = *Any* proposed explanation such that
C and $\neg$ C are each compatible with K.

$$P(E \mid K) = 1$$

Thus…

$$P(E \mid C \,.\, K) = 1$$

$$P(E \mid \neg C \,.\, K) = 1$$

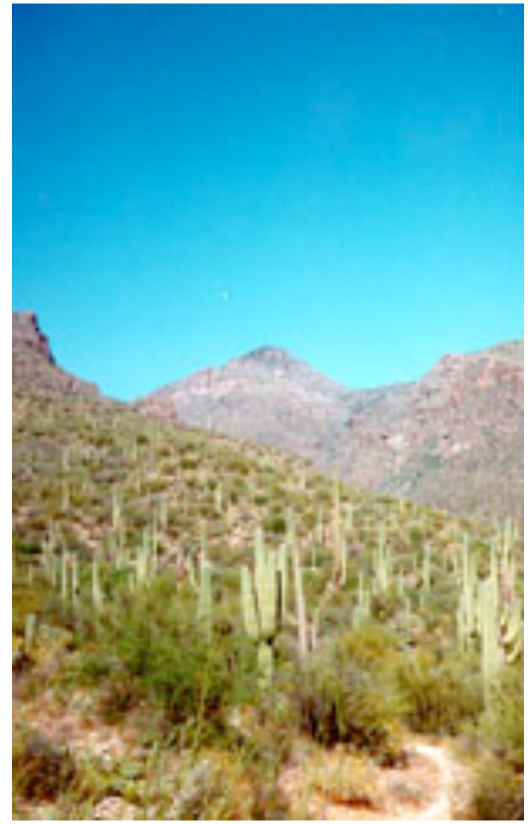
And thus…

$$r_s = 1 - 1 = 0$$

If you were already sure E was true, then there’s no room for a proposed explanation to boost E’s probability, so none will count as “relevant”!

Non-Dogmatic Version

E = “Daytime clear skies on earth are blue”
K = Our background knowledge
C = A great explanation of E (given K).



$$P(E \mid K) = 0.99$$
$$P(E \mid C \,.\, K) = 1.00$$
$$P(E \mid \neg C \,.\, K) = 0.98$$

Thus…

$$r_s = 1 - 0.98 = 0.02$$

If you were already **quite** sure E was going to occur, then there’s **little** room for a proposed explanation **with low prior probability** to boost E’s probability, so none will count as “relevant”!

Why did this atom decay?

E = “This atom decays”
K = Our background knowledge
C = The correct physical theory of radioactivity.

$$P(E \mid K) = 0.10$$
$$P(E \mid C \,.\, K) = 0.01$$

It follows that…

$$r_s < 0.01 - 0.10 = \mathbf{-\ 0.09}$$

Compare: C* = “My favorite color is yellow.”

- This has *zero* explanatory relevance to E.
- **Zero** is significantly higher than **-0.09**.
- But C* *can’t* be more relevant to E than the correct physical theory of radioactivity!

Further Problems…

E = "He will apparently saw a lady in half"
K = Our knowledge going into the show.
C = He brandishes a saw conspicuously.

$$P(E \mid K) = 0.10$$
$$P(E \mid C \,.\, K) = 0.90$$

It follows that…

$$r_s > 0.90 – 0.10 = 0.80$$

Mere correlates shouldn't be so highly relevant!

Compare: C* = Full contents of his notebook.

– This *should be* explanatorily relevant to E.
– But it may not even boost E's probability at all.

"Probabilification Value"

$$r_s = P(E \mid C \,.\, K) - P(E \mid \neg\, C \,.\, K)$$

- I would call Jan's proposed notion something like "*probabilification value*"
  - It is a measure of how much learning the actual truth value of a particular claim C would impact upon the probability of E, relative to background knowledge K.

- *Probabilification value* is interesting, and something that scientists often hope to achieve.

But Probabilification Value is not closely linked to Explanatory Relevance

Sometimes K already confers enough probability on E that there's little probabilification value left to be had, even by highly relevant explanations.

Sometimes K confers *too much* probability on E, such that good, relevant, explanations will end up being *anti*-probabilifying.

Sometimes K is sparse enough that unexplanatory correlates of E will have high probabilification values, whereas quite relevant supplements to K will themselves have low probabilification values.