“Putting the Brakes on the Breakthrough, or ‘How I used simple logic to uncover a flaw in a controversial 60-year old ‘theorem’ in statistical foundations’”

- **Example 1**: Trying and Trying Again: Optional stopping
- **Example 2**: Two instruments with different precisions
(you shouldn’t get credit/blame for something you didn’t do)
- **The Breakthrough**: Birnbaumization
- **Imaginary dialogue** with Allan Birnbaum

My reluctance….

1. It differs from the presentations on The Stat Wars and their Casualties—even though the casualties of this battle have been very large

2. Understanding the full arguments requires more than could be given in a talk like this, and careful distinctions with symbols of the sort I develop (2014) but won't use here.

3. It's something I haven't been immersed in, and 6 years plus a pandemic later, there are some fine points I may not be up on.

But we can always continue the discussion later on the phil-stat-wars.com site or in future forums

Casualties…

The likelihood principle is incompatible with the main body of modern statistical theory and practice, notably the Neyman–Pearson theory of hypothesis testing and of confidence intervals, and incompatible in general even with such well-known concepts as standard error of an estimate and significance level. [Birnbaum (1968), page 300.]

"[I]t is clear that reporting significance levels violates the LP [SLP], since significance levels involve averaging over sample points other than just the observed **x**." [Berger and Wolpert (1988), page 105.]

My reluctance….

1. It differs from the presentations on The Stat Wars and their consequences—even though the consequences of this battle have been very large

2. Understanding the full arguments requires more than could be given in a talk like this, and careful distinctions with symbols of the sort I develop (2014) but won't use here.

3. It's something I haven't been immersed in, and 6 years plus a pandemic later, there are some fine points I'm surely not up on.

But we can always continue the discussion later on the phil-stat-wars.com site or in future forums

Example 1: Trying and Trying Again: Optional stopping

To show "a startling and practically serious example of conflict with the LP", Berger and Wolpert, 104-5:

Keep sampling until H_0 is rejected at the .05 level

E_i Random sample from Normal distribution

$Y_i \sim N(\mu,\sigma)$ and testing H_0: $\mu=0$, vs. H_1: $\mu\neq 0$.

i.e., keep sampling until |relative frequency of success| ≥ 2 standard deviation units (*nominal* p-value of .05)

After 10 trials, failing to rack up enough successes, go to 20 trials, failing again go to 30 trials, and then 10 more, until finally say on trial 169 they attained a statistically significant result.

Stopping rule: rule for when to stop sampling

Optional stopping vs fixed sample size (example throughout)

y*: outcome in E_2 optional stopping (stops with n = 169)
x*: outcome in fixed sample size experiment E_1 (n = 169)

Comparing ***y**** and ***x**** the evidential assessment seems different

For a legitimate .05 assessment (fixed sample size):

Pr (P-value < .05; H_0)= .05

But with optional stopping, the probability of erroneously rejecting the null accumulates

Pr (reach a P-value < .05 in optional stopping n = 169; H_0) = ~.55

[This is a proper stopping rule, it will stop in finitely many trials, and is assured of rejecting the null, even if the null is true]

Both experiments seek to make inferences about the same parameter θ, (with the same value), the statistical model is given.

$Y_i \sim N(\mu,\sigma)$ and testing H_0: $\mu=0$, vs. H_1: $\mu \neq 0$.

- These two outcomes $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ have <u>proportional likelihoods</u>:
- The data can be viewed as a string < z_1 z_2, z_3,…. z_{169} >
- Regardless of whether it came from E_1 *or* E_2

  $$\Pr(\boldsymbol{y}^*; H_0) = c\Pr(\boldsymbol{x}^*; H_0), \text{ c some constant}$$

- Their likelihoods are proportional over values of parameter.

Likelihood Principle (LP):

If likelihoods from different experiments E_1 and E_2 are proportional Pr($\boldsymbol{y}$*; H_0) = cPr($\boldsymbol{x}$*;H_0), then inferences from the two should be the same.

Because there are so many "givens" in the LP, I adopt a shorthand:

LP-pair: Whenever ***x**** and **y*** from different experiments E_1 and E_2 have proportional likelihoods over the parameter, I call them LP pairs (The * indicates they are LP pairs)

Infr[*E*,*z*]: The inference from outcome ***z*** from experiment *E*

In the paper: $\text{Infr}_E[\boldsymbol{z}]$

Likelihood Principle (LP). If ***x**** and ***y**** are LP pairs, then
Infr[E_1,$\boldsymbol{x}$*] = Infr[E_2,$\boldsymbol{y}$*]

i.e., $\text{Infr}_{E1}[\boldsymbol{x}^*] = \text{Infr}_{E2}[\boldsymbol{y}^*]$

- The LP is the controversial principle that I'm discussing.
- The controversial discussion stems from statistician Allan Birnbaum (1962).
- He wanted to consider general principles of "informative" inference (Infr), over all methods and schools, inferences about a parameter θ within a given statistical model.

Example 1: Trying and Trying Again: Optional stopping

The likelihood principle emphasized in Bayesian statistics implies, … that the rules governing when data collection stops are irrelevant to data interpretation.
(Edwards, Lindman, Savage, 1963, p. 239).

Savage: The LP restores "simplicity and freedom that had been lost" with frequentist methods: "optional stopping is no sin"

The LP follows if inference is by way of Bayes's theory
Bayesians call this the *Stopping Rule Principle SRP*.

In general, suppose that you collect data of any kind whatsoever – not necessarily Bernoullian, nor identically distributed, nor independent of each other. . . – stopping only when the data thus far collected satisfy some criterion of a sort that is sure to be satisfied sooner or later, then the import of the sequence of *n* data actually observed will be exactly the same as it would be had you planned to take exactly *n* observations in the first place.
(Edwards, Lindman, and Savage 1963, 238-9)

In calculating [the posterior], our inference about μ, the only contribution of the data is through the likelihood function….In particular*, if we have two pieces of data* ***x****** *and* ***y***** with [proportional] likelihood function ….the inferences about μ from the two data sets should be the same. This is not usually true in the orthodox [frequentist] theory and its falsity in that theory is an example of its incoherence.* (Lindley 1976, p. 36).

Violates Weak Repeated Sampling

For sampling theorists, by contrast, this example “taken in the context of examining consistency with $\theta = 0$, is enough to refute the strong likelihood principle” [Cox (1978), page 54], since, with probability 1, it will stop with a ‘nominally’ significant result even though $\theta = 0$.

It contradicts what Cox and Hinkley call “the weak repeated sampling principle” [Cox and Hinkley (1974), page 51].

Within sampling theory (frequentist error statistics) inference from $\boldsymbol{x}^*$ from E_1 differs from the inference from $\boldsymbol{y}^*$ from E_2

While p_1 would be .05, p_2 would be ~.55.

LP violation: $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ form an LP pair, but
$\text{Infr}[E_1,\boldsymbol{x}^*] \neq \text{Infr}[E_2,\boldsymbol{y}^*]$

Note: Birnbaum's argument is to hold for any form of inference, p-values, confidence levels, posteriors, likelihood ratios. I'm using p-values for simplicity and since they are used in the initial argument.

Someone could say they don’t care about error probabilities; anyway.

But I’m not here to convince you of the error statistical account that violates the LP.

Birnbaum seemed to show the frequentist couldn’t consistently allow these LP violations.

"Within the context of what can be called classical frequency-based statistical inference, Birnbaum (1962) argued that the conditionality and sufficiency principles imply the [strong] likelihood principle" [Evans, Fraser and Monette (1986), page 182]. (Mayo 2014, p. 228)

The Birnbaum result heralded as a breakthrough in statistics!

Savage**:**
Without any intent to speak with exaggeration it seems to me that this is really a historic occasion. This paper is a landmark in statistics … I myself, like other Bayesian statisticians, have been convinced of the truth of the likelihood principle for a long time. Its consequences for statistics are very great.

… I can't stop without saying once more that this paper is really momentous in the history of statistics. It would be hard to point to even a handful of comparable events. (Savage, 1962).

All error statistical notions, p-values, significance levels, … all violate the likelihood principle (ibid.)

(Savage doubted people would stop at the halfway house, but that they'd go on to full blown Bayesianism; Birnbaum never did.)

I have no doubt that revealing the flaw in the alleged proof will not be greeted with anything like the same recognition (Mayo, 2010).

Sufficiency or Weak LP

Note: If there were only one experiment, then outcomes with proportional likelihoods are evidentially equivalent:

If $\boldsymbol{y_1}^*$ and $\boldsymbol{y_2}^*$ are LP pairs within a single experiment E, then

Infr[E,$\boldsymbol{y_1}^*$] = Infr[E,$\boldsymbol{y_2}^*$]

Also called **sufficiency principle or weak LP**.

(Birnbaum will want to make use of this…)

Aside on *sufficiency*:

If a random variable Y, in a given experiment E, arises from f(y; θ), and the assumptions of the model are valid, then all the information about θ contained in the data is obtained from consideration of its sufficient statistic S and its *sampling distribution* $f_S(s; \theta)$.

Using k Bernouilli trials to make inferences about θ, the probability of success on each trial, S = K, the number of successes (S has a Binomial sampling distribution)

In using abbreviation Infr[*E*,***z***], or preferably, Infr_E[***z***], we assume E includes the probability model, parameters and sampling distribution corresponding to the inference.
(Cox and Mayo 2010, p. 287)

All that was Example 1.

Example 2: ***Mixture Experiment:*** *Cox (1958) Two instruments of different precisions (you shouldn't get blamed for something you didn't do)*

- We flip a fair coin to decide whether to use a very precise or imprecise tool: E_1 or E_2.
- The experiment has 2 steps, first flip the coin to see whether E_1 *or* E_2 is performed, then perform it and record outcome

(Weak) Conditionality Principle CP

- *CP: Once it is known which E_i produced **z**, the p-value or other inferential assessment should be made conditioning on the experiment actually run*.

Note: The randomizer (the flipping) must be irrelevant to the parameter θ: a θ irrelevant randomizer.

Example: in observing a normally distributed **Z** in testing a null hypothesis $\theta = 0$, E_1 has variance of 1, while that of E_2 is 10^6.

The same ***z*** measurement would correspond to a much smaller p- value if from E_1 rather than E_2: The p-value from E_1 (p_1) is much less than from E_2 (p_2) $p_1 \ll p_2$

Unconditional or convex combination

What if someone reported the precision of the experiment as the average of the two:

$$[p_1 + p_2]/2$$

This *unconditional* assessment is referred to as the *convex combination* of the p-values averaged over the two experiments.

Cox identifies the CP to make this point: Suppose we know we observed E_2 with its much larger variance:

> The unconditional test says that we can assign this a higher level of significance than we ordinarily do, because if we were to repeat the experiment, we might sample some quite different distribution. But this fact seems irrelevant to the interpretation of an observation which we know came from a distribution [with the larger variance] (Cox 1958, 361).

[The measurement using the imprecise tool shouldn't get credit because the measurer might have used a precise tool; nor should the precise guy be blamed because he could have used the lousy imprecise tool.]

Conditionality Principle

CP: Once it is known which E_i has produced **z,** condition on the E_i producing the result.

Do not use the unconditional formulation.

It seems obvious

Note: CP requires the experiment and its outcome to be given or known: If it is given only that **z** came from E_1 or E_2, and not which, then CP does not authorize [conditioning].

.

If it’s so obvious why was Cox’s paper considered one of the most important criticisms of (certain) frequentist methods?

CP is not a mathematical principle, but one thought to reflect intuitions about evidence.

There are contexts where the frequentist would report the *unconditional* assessment e.g., if there were gong to be many applications and you were reporting average precision or the like.

> "If the object of the analysis is to make statements by a rule with certain specified long-run properties, the unconditional test just given is in order..If, however, our object is to say ‘what we can learn from the data that we have’, the unconditional test is surely no good.” (Cox 1958, 361).

The surprising upshot of the argument I will consider is it claims to show if you condition on the randomizer you have to go all the way to the data point, which means no error probabilities (the LP).

That's what Birnbaum's result purports: **(SP and CP) entails LP!**

> It is not uncommon to see statistics texts argue that in frequentist theory one is faced with the following dilemma: *either to deny the appropriateness of conditioning on the precision of the tool chosen by the toss of a coin, or else to embrace the strong likelihood principle,* which entails that frequentist sampling distributions are irrelevant to inference once the data are obtained. This is a false dilemma. . . . The 'dilemma' argument is therefore an illusion. (Cox and Mayo 2010, 298)

But the illusion is not so easy to dispel; thus this paper.

NOW FOR THE BREAKTHROUGH

Shorthand: x^* = (E$_1$,x^*); y^* = (E$_2$,y^*)

LP: If x^* and y^* are LP pairs, then Infr[E_1,x^*] = Infr[E_2,y^*]

Whenever ***x****** and ***y****** form an LP pair, but

Infr[E$_1$,$\boldsymbol{x}^*$] ≠ Infr[E$_2$,$\boldsymbol{y}^*$]

we have a *violation of the Likelihood Principle* (LP).

We saw in Example 1:

y^* (a 2 s.d. result) in E$_2$ optional stopping (stops with n = 169) had a p-value ~.55
x^* (a 2 s.d. result) in E$_1$, fixed n = 169, had a p-value =.05

The two outcomes form an LP pair, so we have an *LP violation*.

Birnbaum Experiment: Birnbaumization

Step 1: Birnbaum Experiment $\boldsymbol{E}_{\mathbf{B}}$:
Birnbaum will describe a funny kind of 'mixture' experiment based on an LP pair;

You observed $\boldsymbol{y}^*$ from experiment E_2 (optional stopping)

- *Enlarge it to an imaginary mixture*: I am to imagine it resulted from getting tails on the toss of a fair coin, where heads would have meant performing the fixed sample size experiment with n = 169 from the start.

- *Erasure*: Next, erase the fact that $\boldsymbol{y}^*$ came from E_2 and report it as if it came from from (E_1, $\boldsymbol{x}^*$)

Birnbaumization

We are to imagine that performing E_2 was the result of flipping a fair coin (or some other randomizer given as irrelevant to θ) to decide whether to run E_1 or E_2. Cox terms this the "enlarged experiment" [Cox (1978), page 54], E_B. We are then to define a statistic T_B that stipulates that if $(E_2, \mathbf{y}^*)$ is observed, its SLP pair $\mathbf{x}^*$ in the unperformed experiment is reported;

$$T_B(E_i, \mathbf{Z}_i) = \begin{cases} (E_1, \mathbf{x}^*), & \text{if } (E_1, \mathbf{x}^*) \text{ or } (E_2, \mathbf{y}^*), \\ (E_i, \mathbf{z}_i), & \text{otherwise.} \end{cases}$$

Birnbaum's argument focuses on the first case and ours will as well.

T_B is a sufficient statistic within the Birnbaumized experiment

Inference within Birnbaum's Experiment E_B

If making an inference within E_B: Report the convex combination

$[p_1 + p_2]/2$

p_1 = .05 (E_1 fixed sample size n = 169)
p_2 = .55 (E_2 optional stopping, stops at n = 169.

So the average is ~ .3

(the ½ comes from the imagined fair coin)

I said it was a funny kind of mixture, two reasons:

- It didn't happen, you only observed **y*** from E_2
- You are to report an outcome as ***x**** from E_1 even though you actually observed **y*** from E_2
- The inference is based on the unconditional mixture

We may call it *Birnbaumizing* the result you got.

You are to do it for any LP pair, we're focusing on just these two.

Within E_B, it appears that you must treat ***x**** as evidentially equivalent to its LP pair, ***y**** (after all: you report y* as x*)

Premise 1: If use Birnbaumization, the inference *within* E_B

Infr[***x****] = Infr[***y****]   both are equal to $(p_1 + p_2)/2$)

Premise 2 (a): If use conditionality CP:

Infr[***y****] = p_2

Premise 2 (b): Infr[***x****] = p_1

He wants to infer: $p_1 = p_2$

But we know $p_1 \neq p_2$

Construed this way, it's an invalid argument (true premises, false conclusion)

Formulated as a valid argument, it will not be sound. You can uphold SP and CP, but not apply them in such a way as to violate self-consistency.

Instead of continuing with the formal paper, let me sketch an exchange I post on my blog each New Year's eve, one I imagine "Midnight With Birnbaum" akin to that *Midnight in Paris* movie.

Just as that writer gets to go back in time to get appraisals from famous past writers, I imagine checking my criticism of Birnbaum with him.

Just to zero in on the main parts…

ERROR STATISTICAL PHILOSOPHER: It's wonderful to meet you Professor Birnbaum….I happen to be writing on your famous argument about the likelihood principle (LP).

BIRNBAUM: Ultimately you know I rejected the LP as failing to control error probabilities (except in the trivial case of testing point against point hypotheses, provided they are predesignated.).

ES PHILOSOHER: I suspected you found the flaw in your argument …? It doesn't seem to work…

BIRNBAUM: Well, I shall happily invite you to take any case that violates the LP and allow me to demonstrate that the frequentist is led to inconsistency, provided she also wishes to adhere to Sufficiency and Conditionality Principle CP

Whoever wins this little argument pays for this whole bottle of vintage champagne.

ES PHILOSOPHER: I really don't mind paying for the bottle.

BIRNBAUM: Good, you will have to. Take any LP violation.

In two-sided Normal testing, ***y***** from optional stopping E_2 (stops with a 2 s.d. difference at *n* = 169).

I will show you that ***y***** must be inferentially equivalent to ***x***** from corresponding fixed sample size E_1 (n = 169).

i.e. Infr[***x*****] = Infr[***y*****]

(We assume both experiments seek to make inferences about the same parameter θ, and the model is given.)

Do you agree that: For a frequentist, ***y***** is not evidentially equivalent to ***x*****?

ES PHILOSOPHER: Yes, that's a clear case where we reject the LP, and it makes perfect sense to distinguish their corresponding p-values: the optional stopping experiment makes the p-value quite a bit higher than with the fixed sample size.

For optional stopping $p_2 = .55$

For fixed $p_1 = .05$

I don't see how you can make them equal.

BIRNBAUM: Well, I'll show you'll have to renounce another principle you hold, the CP (I assume you hold SP: sufficiency)

ES PHILOSOPHER: Laughs.

BIRNBAUM: Suppose you've observed $\boldsymbol{y}^*$ from E_2.You admit, do you not, that this outcome could have occurred as a result of a different experiment? It could have been that a fair coin was flipped where it is agreed that heads instructs you to perform E_1 (fixed sample size) and tails instructs you to perform the optional stopping test E_2, and you happened to get tails and managed to stop at 169. A mixture experiment.

ES PHILOSOPHER: Well, that is not how I got $\boldsymbol{y}^*$, but ok, it could have occurred that way.

BIRNBAUM: Good. Then you must grant further that your result could have come from a special variation on this mixture of mine, call it a Birnbaum experiment E_B.

In an E_B, if the outcome from the experiment you actually performed has an outcome with a proportional likelihood to one in some other experiment not performed, E_1, then we say your result has an "LP pair".

In that case, E_B stipulates that you are to report ***y***** as if you had determined whether to run E_1 or E_2 by flipping a fair coin (given as irrelevant to the parameter of interest).

ES PHILOSOPHER: So let's see if I understand a Birnbaum experiment E_B. I run E_2 and when I stop I report my observed 2 s.d. difference came from E_1, and as a result of this strange type of a mixture experiment.

BIRNBAUM: Yes, or equivalently you could just report

y**: my result is a 2 s.d. difference and it could have come from either E_1 (fixed sampling, *n*= 169) or E_2 (optional stopping, which happens to stop at the 169th trial). The proof is often put this way, it makes no difference.

ES PHILOSOPHER: You're saying if my result has an LP pair in the experiment not performed, I should act as if I accept the LP and so if the likelihoods are proportional in the two experiments (both testing the same mean), the outcomes yield the same inference.

BIRNBAUM: Well, wait, experiment $E_{B,}$ as an imagined mixture it is a single experiment, so really you only need to apply the *sufficiency principle* (the weak LP) which everyone accepts. Yes?

ES PHILOSOPHER: So Birnbaumization has turned the two experiments into one. But how do I calculate the p-value within your *Birnbaumized* experiment?

BIRNBAUM: I don't think anyone has ever called it that.

ES PHILOSOPHER: I just wanted to have a shorthand for the operation you are describing, … So how do I calculate the p- value within E_B?

BIRNBAUM: You would report the overall (*unconditional*) p-value, which would be the average over the sampling distributions:
$(p_1 + p_2)/2$.

(The convex combination.) Say p_1 is .05 and p_2 is ~.55; anyway, we know they are different, that's what makes this a violation of the LP.

ES PHILOSOPHER: So you're saying that if I observe a 2-s.d. difference from E_2, I do not report the associated p-value .55, but instead I am to report the average p-value, averaging over some other experiment E_1 that could have given rise to an outcome with a proportional likelihood to the one I observed, even though I didn't obtain it this way?

BIRNBAUM: I'm saying that you have to grant that ***y****** from an optional stopping experiment E_2 could have been generated through Birnbaum experiment E_B.

If you follow the rules of E_B, then ***y****** is evidentially equivalent to ***x****** This is premise 1.

Premise 1: Within E_B

The inference from ***y****** = the inference from ***x****** both are equal to $(p_1 + p_2)/2)$

(defn of E_B and sufficiency)

ES PHILOSOPHER: But this is just a matter of your definition of inference within your Birnbaumized experiment E_B. We still have p_1 different from p_2 so I don't see the contradiction with my rejecting the LP.

BIRNBAUM: Hold on; I'm about to get to the second premise, premise 2. So far all of this was premise 1.

ES PHILOSOPHER: Oy, what is premise 2?

BIRNBAUM: Premise 2 is this: Surely, you agree (following CP), that once you know from which experiment the observed 2 s.d. difference actually came, you ought to report the p-value corresponding to that experiment. You ought not to report the average $(p_1 + p_2)/2$ as you were instructed in E_B

This gives us:

Premise 2 (a): From CP:

The inference from *y is p_2**

ES PHILOSOPHER: So, having first insisted I imagine myself in a Birnbaumized, E_B and report an average p-value, I'm to return to my senses and "condition" to get back to where I was to begin with?

BIRNBAUM: Yes, at least if you hold to the conditionality principle CP (of D. R. Cox)

Likewise, if you knew it came from E_1,then the p-value is

Premise 2 (b): From CP:

The inference from *x is p_1**

So you arrive at (2a) and (2b), yes? Surely you accept Cox's CP.

ES PHILOSOPHER: Well, the CP is defined for mixtures, where one flips a coin to determine if E_1 or E_2 is performed. One cannot perform the following: Toss a fair coin. If it lands heads, perform an experiment E_1 that yields a member of an LP pair ***x****, if tails observe E_2 that yields ***y****.

We do not know what would have resulted from the unperformed experiment, much less that it would form an an LP pair with the observed **y***. There is a single experiment, and CP stipulates that we know which was performed and what its outcome was.

BIRNBAUM: Wasn't I clear that I'm dealing with hypothetical or mathematical mixtures (rather than actual ones)? A lot of people—including some brilliant statisticians—have tried to block my argument because it deals in mathematical mixtures.

Mathematical or Hypothetical Mixtures

ES PHILOSOPHER: I want to give your argument maximal mileage. I will allow CP is definable for hypothetical θ-irrelevant mixtures out there, and that we can Birnbaumize an experimental result.

Envision the mathematical universe of LP pairs, each imagined to have been generated from a θ-irrelevant mixture (for the inference context at hand). When we observe ***y**** we pluck the ***x**** companion needed for the argument. If the outcome doesn't have an LP pair, just proceed as normal.

You're task is to show that

> If ***x**** and ***y**** are LP pairs, then the inference from ***x**** = the inference from ***y****.

That's the LP.

BIRNBAUM: Exactly, given y*, pluck down x* and get premise (1). Together with premises (2a) and (2b) we get the LP!

ES PHILOSOPHER: Clever, but your “proof” is obviously unsound: invalid, or valid with false premises.

Invalid

Premise 1: If use Birnbaumization (T_B is sufficient in E_B)

Infr[***x****] = Infr[***y****] both are equal to $(p_1 + p_2)/2$)

Premise 2 (a): If use conditionality CP:

Infr[***y****] = p_2

Premise 2 (b): If use conditionality CP:

Infr[***x****] = p_1

Conclusion $p_1 = p_2$

Putting the premises as “if thens”, they can both be true but can still have $p_1 \neq p_2$

Valid but premises conflict

Premise 1: **In** $\boldsymbol{E}_B$, the inference from $\boldsymbol{y}^*$ = inference from $\boldsymbol{x}^*$ [both are equal to $(p_1 + p_2)/2$)]

Premise 2 (a): from CP (conditionality)

In $\boldsymbol{E}_B$ (given $\boldsymbol{y}^*$), the inference from $\boldsymbol{y}^*$ = p_2

Premise 2 (b): from CP:

In $\boldsymbol{E}_B$, (given $\boldsymbol{x}^*$), the inference from $\boldsymbol{x}^*$ = p_1

Conclusion: The inference from $\boldsymbol{y}^*$ = the inference from $\boldsymbol{x}^*$. ($p_1 = p_2$)

Premise 1 conflicts with the premises about $\boldsymbol{E}_B$ so it's valid but unsound

If Premise 1 is true, 2a and b are false.

- *Keeping to this valid formulation, the premises can hold true only in a special case: when there is no LP violation.*

- With LP violations there are three different distributions E_1, E_2, E_B

- Else the premises contradict each other. Of course, anything follows from a contradiction (X and ~X), and the argument may be playing on that fact.

- Alternatively, I can get the premises to come out true, but then the conclusion is false (for any LP violation)—so it is *invalid*.

Every LP violation will be a case where the argument is *unsound* (either invalid or premises conflict).

BIRNBAUM: Of course Bayesians and Likelihoodists have much simpler ways to derive the LP, e.g., it follows directly from inference by way of Bayes Theorem. My argument was to hold for "approaches which are independent of this [Bayes'] principle" [Birnbaum (1962), page 283].

ES PHILOSOPHER: There's no contradiction for a frequentist (by dint of accepting SP and CP, and denying the LP.)

[Aside: Contemporary *nonsubjective Bayesians* concede they "have to live with some violations of the likelihood and stopping rule principles" (Ghosh, Delampady, and Sumanta 2006, 148).]

:

BIRNBAUM: Yet some people still think it is a breakthrough.

ES PHILOSOPHER: I've come to see that clarifying the entire argument turns on defining the CP.

BIRNBUM: Yes, the "monster of the LP" arises from viewing CP as an equivalence, instead of going in one-sided form (from mixtures to the known result).

ERROR STATISTICAL PHILOSOPHER: In my 2014 paper I too construe CP as giving an "equivalence" but there is an equivocation that invalidates the purported move to the LP.

Equivalently,

Given ***y**** is known to have come from E_2, it doesn't matter if E_2 was the result of an irrelevant randomizer to choose between E_1 and E_2, or E_2 was fixed all along.

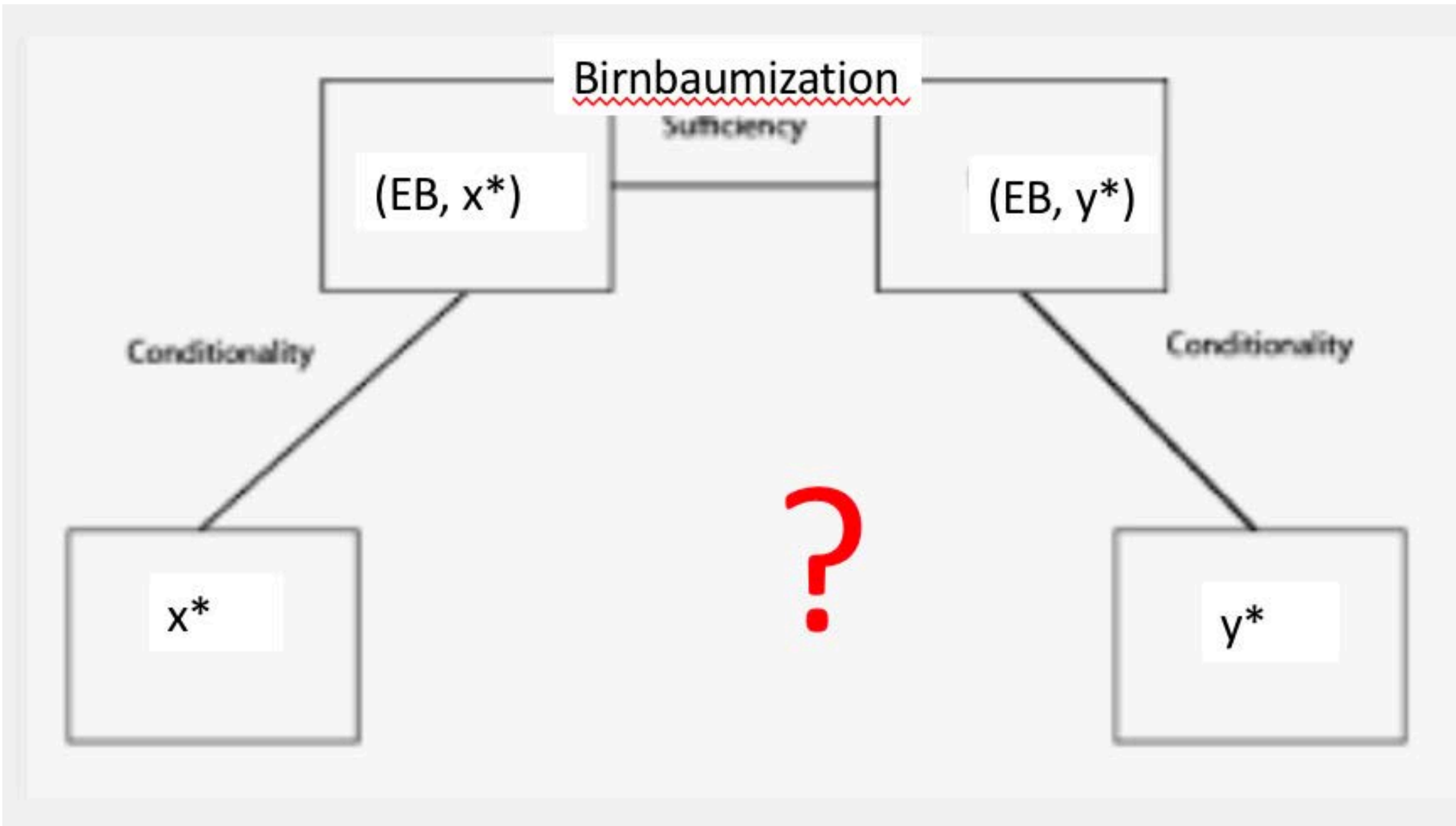
(because knowing it came from E_2, I have all I need to compute the inference)

It does not follow that:

Given ***y**** is known to have come from E_2, I don't have to know it came from E_2, (so it's equal to just reporting the Birnbaumized result)

The problem stems from mistaking CP as the equivalence

Infr[E_B, z] = Infr[E_i,z] i = 1,2


Birnbaumization
Sufficiency
(EB, x*)
(EB, y*)
Conditionality
Conditionality
x*
?
y*

Fallacy of Four Terms

(old)
Every computer has a mouse.
Every mouse is an animal.
Therefore, every computer has an animal.

Here is another:
A ham sandwich is better than nothing.
Nothing is better than eternal happiness.
A ham sandwich is better than eternal happiness.

- **Example 1**: Trying and Trying Again: Optional stopping

- **Example 2**: Two instruments with different precisions
  (you shouldn’t get credit/blame for something you didn’t do)

- **The Breakthrough**: Birnbaumization

- **Imaginary dialogue** with Allan Birnbaum

END

Msc

Contemporary *nonsubjective Bayesians* concede they "have to live with some violations of the likelihood and stopping rule principles" (Ghosh, Delampady, and Sumanta 2006, 148), since their prior probability distributions are influenced by the sampling distribution. ..[O]bjectivity can only be defined relative to a frame of reference, and this frame needs to include the goal of the analysis." (J. Berger 2006, 394).

Msc

Nothing precludes the Neyman–Pearson theory from choosing the procedure “which is best on the average over both experiments” in E_{mix} [Lehmann and Romano (2005), page 394]. They ask the following: “for a given test or confidence procedure, should probabilities such as level, power, and confidence coefficient be calculated conditionally, given the experiment that has been selected, or unconditionally?” They suggest that “[t]he answer cannot be found within the model but depends on the context” (Lehmann & Romano 2005, p. 394).

- Msc
- The evidential meaning of any outcome of any mixture experiment is the same as that of the corresponding outcome of the corresponding component experiment, ignoring the over-all structure of the mixture experiment.
- Dawid's definition is a portion of the one found in Birnbaum (1962), page 271. It assumes, of course, all of the other stipulations, for example, we are making "informative" inferences about θ, it is a θ-irrelevant mixture, the outcome is given, and all the rest.

I don't say:

"the problem stems from mistaking CP as the equivalence" simpliciter, but rather it stems from the incorrect equivalence!

"The problem stems from mistaking CP as the equivalence Infr[E-mix, z] = Infr[E_i,z] (whether the mixture is hypothetical or actual)."

In the notation used in this presentation:

The problem stems from mistaking CP as the equivalence

Infr[E_B, z] = Infr[E_i,z] i = 1,2

Msc

"Mayo's comments here are true of the operational sufficiency principle…Birnbaum's sufficiency principle does not say anything about how inference is to be performed" (Gandenberger 2015).

Yes, it's a universal generalization: For any case of informative inference (about a parameter in a given model), from any school, if SP and CP, then LP

So all I need to do is supply a single counterexample, any LP violation in sampling theory will do.

He seems to think there's a principle distinct from operationalizing it! It must be operationalized to "turn 2 into 1" and get x* = y* (evidentially equivalency) in Birnbaumization.

Birnbaum's argument was to give an argument that is relevant for a sampling theorist and for "approaches which are independent of this [Bayes'] principle"[Birnbaum (1962), page 283].

Its assumed implications for sampling theory is why it was dubbed "a landmark in statistics" [Savage (1962b), page 307].

Bayesians and Likelihoodists have much simpler ways to derive the LP, e.g., it follows directly from inference by way of Bayes Theorem.