

DON'T FOOL YOURSELF!

How to avoid pitfalls in logical thinking and systematic trading

Saurabh Singal

TRADE TECH ASIA

19 OCTOBER, 2016

The Very First Clinical Trial

- King Gustav III of Sweden viewed coffee consumption as being dangerous.
- The king ordered the experiment to be conducted using two identical twins who had been condemned to death.
- One twin was made to drink three pots of coffee every day, and the other, tea.
- Two physicians supervised this procedure.
- Both doctors died & Gustav III was assassinated before seeing the final results.
- Of the twins, the tea drinker was the first to die, at age 83!
- *Sweden has one of the highest coffee consumption per capita despite several attempts to ban coffee*

Surely You're Joking, Mr. Feynman!

- The inimitable Richard Feynman wrote a famous article called *Cargo Cult Science*, parts of which are also summarized in his famous autobiographical book.
- He says that the first principle is that you must not fool yourself - and you are the easiest person to fool. You must be willing to question and doubt your own theories and results, and investigate possible flaws in a theory or an experiment, even when you use others' published results.
- After you've not fooled yourself, it's easy not to fool other scientists. You just have to be honest in a conventional way after that...

That is our theme today....

Three Common Types of Dangerous Mistakes

- Unwitting misuses of Statistics
 - *Simpson Paradox, Law of Small numbers as just two examples*
- Use bad computational tools or sloppy methods in your own work.
 - *Reinhart and Rogoff; Number of Chromosome Pairs(23, not 24)*
compared to
Careful work and physical explanation in Weldon Dice study
- Rely on others' research without verifying it yourself – especially when the title of the research is flashy or the conclusion appears dazzlingly counter-intuitive.
 - *Rogue economist Steven Levitt (Freakonomics, SuperFreakonomics)*
compared to
the brilliant insights of Abraham Wald

How Sweet!

Mankind outsmarts itself

- We are programmed to like sweet tasting things because Nature wanted us to eat vitamins, which occur in fruits, and fruits are sweet- but we learnt to make sugar !
- In 1975, Indian chemist Shashikant Phadnis was trying to make a new insecticide in Queen Elizabeth College, London.
- He started by adding sugar to sulphuryl chloride, a highly toxic chemical, which has many chlorine atoms, just like DDT.
- His supervisor asked him to **test** the new chemical, but the young misunderstood and thought he had been asked to **taste** it.
- The new compound -Sucralose or Splenda – is about 600 times sweeter than sugar and is the best selling artificial sweetener in America.
- *Sugar is probably the most dangerous thing we consume today- but most Europeans before Columbus had never tasted cane sugar! WHAT IF Gustav III had studied the harmful effects of sugar?*

How We Fool Ourselves on the Road

- The advent of disc brakes, seatbelts, and airbags has led to virtually no measurable decrease in fatalities...
- Thus demonstrating a constant risk tolerance - drivers have increased their speeds and recklessness to max out the advantages in safer design !

What is Simpson's Paradox?

Simpson's paradox: a reversal of the direction of an association when data from several groups are combined to form a single group.

- For example, When the data are examined as one group, the association between X and Y is positive, but when the data are split into groups based on some other characteristic, say W, the association between X and Y is negative.
- This is a type of **Omitted-Variable Bias (OVB)** : occurs when a model incorrectly leaves out one or more important causal factors. The model compensates for the missing factor by over- or underestimating the effect of one of the other factors.

Simpson's Paradox Example

(courtesy Craig Dibble)

- Hooters has 2 jobs, cooks and service staff.
 - For the job of cooks, 20 men and 5 women apply.
 - For the job of and service staff, 5 men and 20 women apply.
- Hooters hires 4 men and 1 woman as cooks.
- Hooter also hires all 5 men and 20 women as service staff.
- Over all, Hooters hired 9 men (out of 25 male applicants) and 21 women (out of the 25 female applicants).

Simpson's Paradox Example - 2

- Over all, Hooters hired 9 men (out of 25 male applicants) and 21 women (out of the 25 female applicants).
- Total analysis: A candidate has 36% chance of being hired if male; and 84% if woman. *Apparent evidence of discrimination.*



Simpson's Paradox Example -3

- Granular analysis: A candidate applying for the job of a cook has a 20% chance to be hired, **and this is the same for men and women.**
- A candidate applying for the job of service staff has a 100% chance to be hired, **same for men and women.**
- *There is **NO** evidence of discrimination.*
- The scary thing is every time we do this sort of thing and get a result it does not mean anything if there is another confounding variable. For example the cooks might really be top chef and sous chef. They only hire men for top chef, only hire women for sous chef. You would miss the discrimination then.
- How far do you go down? Far enough to get the answers you want!!!

Law of Small Numbers

- The Bill and Melinda Gates Foundation has funded the movement that supports smaller schools.
- *It is claimed that outstanding schools are small schools.*
- *But weak schools tended to be smaller than average as well!*
- ***Evidence that Smaller Schools do not improve Student Achievement.*** This is the title of an article by *Howard Wainer & Harris L. Zwierling*.
- ***Which sample is likely to have the highest average? The answer is the sample with the smallest number of observations (Wainer & Zwierling 2006). The reason is that the variability of an average declines with the number of observations.***

Law of Small Numbers - 2

- The counties in which the incidence of kidney cancer is ***lowest*** are mostly rural, sparsely populated, and located in traditionally Republican states in the Midwest, the South, and the West.
- The incidence of kidney cancer is ***highest*** in counties which tend to be mostly rural, sparsely populated, and located in traditionally Republican states in the Midwest, the South, and the West.

Law of Small Numbers - 3

- The very small population in the counties under question makes it drawing inferences prone to error.
- A county with small population has a much greater chance to be healthier, cleaner, dirtier, richer, ..., than average. But it also has a much greater chance to be less healthy, dirtier, poorer, ..., than average.
- If a Nobel laureate decided to move to a village with a small population, of say 49 people, suddenly this village would rightfully boast that 2% of its residents are Nobel prize winners. (But if you or I decide to subsequently decide to reside in the same village, our chances of getting a Nobel are not going to be 2%!)

Law of Small Numbers - 4

- Daniel Kahneman in *Thinking Fast and Slow* illustrates this further. One person tosses 4 fair coins repeatedly. Second person tosses 7 fair coins repeatedly.
- Now note the proportion of times each person gets a homogeneous result (either all heads or all tails). The respective probabilities are $0.5^4 \times 2$ and $0.5^7 \times 2$. Or 12.5% and 1.56% respectively, the larger number being 8 times the smaller.
- If a coin toss is replaced by probability of cancer and number of coins (4 or 7) is replaced by population of a county, then the smaller counties will dominate the top and bottom rankings.
- Imagine adding two more people that toss a single coin each. And a third who tosses a million coins
- ***This is just a statistical fact, and we must be cognizant of this.***

Oops-onomics: Quirky is not always Correct

- Well known statistician Andrew Gelman & Kaiser Fung wrote in *Freakonomics: What Went Wrong?* “... we noted a discouraging tendency in the *Freakonomics* body of work to present speculative or even erroneous claims with an air of certainty.”
- In *Super Freakonomics* - *driving drunk is safer than walking drunk!*
- In *Freakonomics* and in his 2001 paper co-authored by Donahue, Levitt claims *legalized abortion reduces crime rate*. Several other researchers disputed this finding. In particular, Christopher L. Foote and Christopher F. Goetz found a “*programming error*” in Levitt’s work, which causes the “*abortion effect*” to vanish. Levitt acknowledged his error but claimed that re-doing the calculations after fixing the errors supports the original results.

Armour on Airplanes & Survivorship Bias

- In World War II, bullet holes on fighter planes returning from combat were analysed to determine where more armour should be put.
- Too much armour makes the plane heavier and difficult to maneuver and also unable to fly longer distances. Too little armour makes it more vulnerable to enemy fire.
- Initial suggestion by researchers: maximum damage was on fuselage (wings, nose and tail), so put more armour there. Less damage around engines, so reduce armour there.
- **Abraham Wald's** suggestion: put more armour on engines! A hit there causes no survivors. A hit to the fuselage might still lead to survival. The initial reaction was to reinforce where damage had been found but in fact it is best to reinforce other areas because those were the planes that didn't return!
- A person shot in the hand or feet might survive; a shot to the head is likely to be fatal.

Sadly Abraham Wald and his wife died in a plane crash in Nilgiri mountains.

Another Quirky and Logical Finding

We saw example of quirky findings - some good (Wald) and some bad(Levitt). Here is another from road safety- *cyclists with helmets were at higher risk than cyclists without helmets!*

Behavioral reason: Car Drivers were more careful when passing cyclists without helmets and drove farther from them, putting more space between car and cycle and reducing the chance of crash.

Physical reason: Without a helmet, a big crash would fracture your skull, which allows your brain to expand and is survivable. These days the helmet stops your skull from shattering but does nothing to prevent bruising to the brain, which still expands, but with life-ending consequences in an unbroken skull.

How not to Excel at Economics: Reinhart & Rogoff

- Carmen Reinhart and Kenneth Rogoff (authors of *This Time is Different*) wrote a 2010 research paper, “*Growth in a Time of Debt*”, concluding that higher government debt is associated with slower economic growth.
- On average annual growth was **-0.1%** in countries with gross government debt $\geq 90\%$ of GDP.
- Thomas Herndon, a grad student at UMass, re-ran the numbers and found the “corrected” number is **+2.2 %**. Herndon pointed out some other mistakes too.
- R& R admitted accidentally excluding five rows while computing an “AVG” in their Microsoft Excel spreadsheet.
- In a “science lab”, experimenters would be required to record every detail of their work, including experimental design, procedures, equipment, raw results, data processing, statistical methods
- But in many analytic “Big Data” frameworks , relatively few take such care in their work.

There are none so blind as do not wish to see!

- In 1921, a Texas scientist, Theophilus S. Painter published a finding that human cells have 24 pairs of chromosomes, the same number as in the great apes (chimpanzees, gorillas and orangutans).
- In 1955, an Indonesian-American scientist Joe-Hin Tijo found that the correct number of chromosome pairs in humans is 23.
- At that point, some other scientists reported that in their experimental work, they could also count only 23 pairs, not 24 - and admitted that lacked the self-belief to challenge the existing wisdom.
- In fact, photographs printed in several textbooks published between 1921 and 1955 were examined and in all of them, only 23 pairs could be seen.
- *There are none so blind as do not wish to see!*

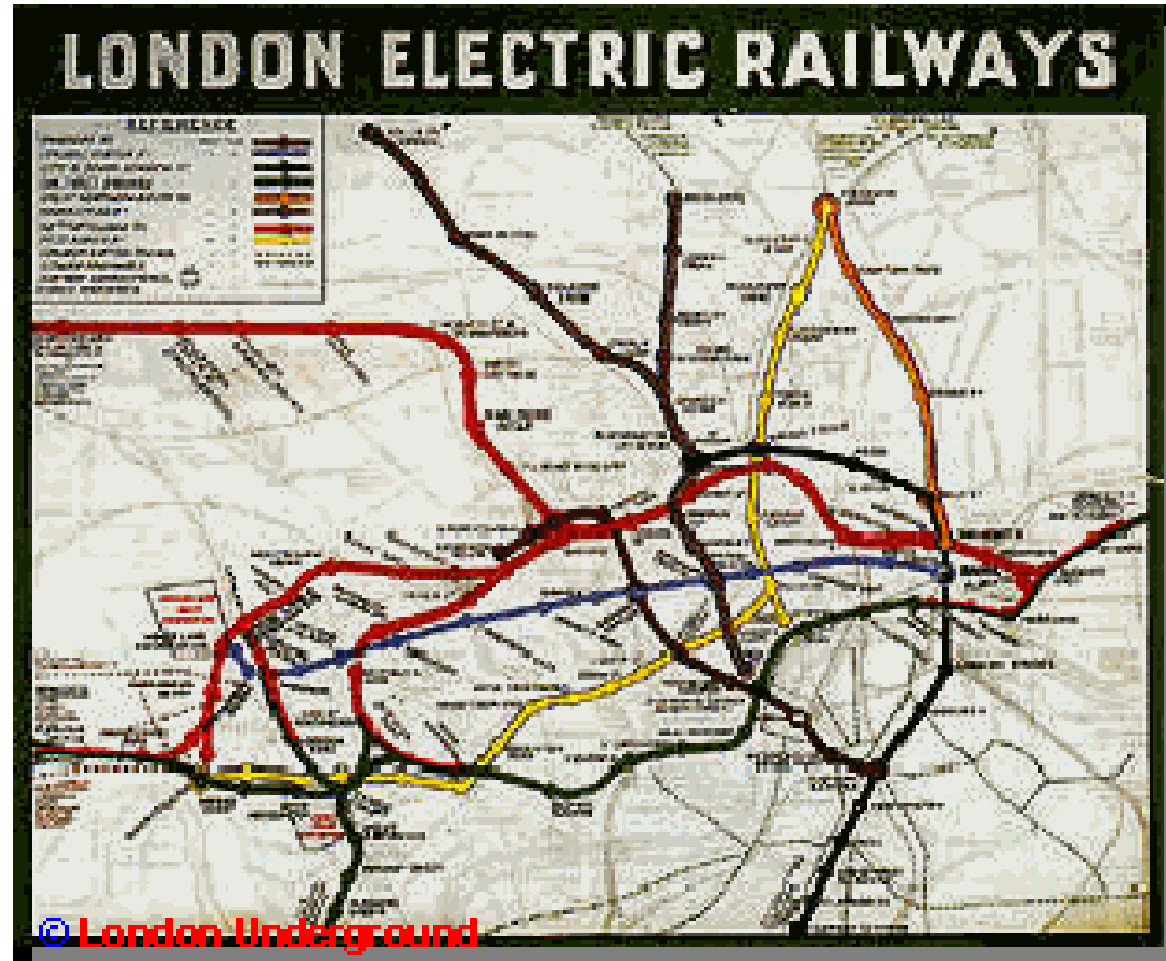
Weldon's Dice

- When we roll a die, do all numbers 1,2,...,6 show up $1/6^{\text{th}}$ of the time?
- Weldon reported to Galton after tossing 12 dice more than 26,306 times.
- He found a small but clear bias towards getting 5 or 6; p -value of 0.0003

Explanation

- Most inexpensive dice have hollowed-out pips; the face with 6 pips is lighter than its opposing face, which has only 1 pip. Therefore 6 shows up more often than 1;
- Like wise 5 shows up more often than 2.
- However 3 and 4 pips have less difference in weight and the effect is not observed for 3 vs 4.

Less Accurate, but more Appropriate



Less Accurate, but more Appropriate - 2



Less Accurate, but more Appropriate - 3

- Before the map made by Harry Beck in 1931, the various lines were shown superimposed on an actual map of the city.
- The central stations were shown too close together, the further ones were too naturally too far.
- Beck realised that passengers were not concerned with geographical accuracy, but how to get from one place to another and where to change trains.
- While drawing an electrical circuit layout, Beck got the inspiration for the tube map- a schematic where most stations were roughly equally spaced.
- Originally there was opposition to it. He was denied credit even after his idea was adopted.
- *This diagram is an evolution of the original design conceived in 1931 by Harry Beck is printed on every London Underground map.*

Conclusion

In particular, watch out for

- 1) Omitted variables that are confounded - If they were obvious they wouldn't have been omitted.
- 2) Datasets that are too small - check the significance numbers
- 3) Interesting results get reported, non interesting ones buried - ask to the analyses that your staff didn't show you.
- 4) Watch out for sloppy methods - think of your grumpiest professor. Would he be convinced?
- 5) Finally "Think!"

Challenge

In your own business, find examples of

1. Survivor analysis
2. Example of Simpson's paradox
3. Law of small numbers
4. Where simplifying justifies the loss of accurate representation
5. Where the underlying reason justifies a seemingly strange anomaly
6. And finally, any research paper or results that you might have come across which did not "test out" - providing you with your own *oops!* moment