

Security Behaviours

COMS30038 Lecture 9 – The Role of Bias

Dr Ramokapane

A close-up photograph of a cracked egg on a light-colored plate. A black knife is positioned diagonally across the frame, with its tip near the egg. The egg is cracked open, revealing a bright yellow yolk. The background is a plain, light-colored surface.

A Recap on Error

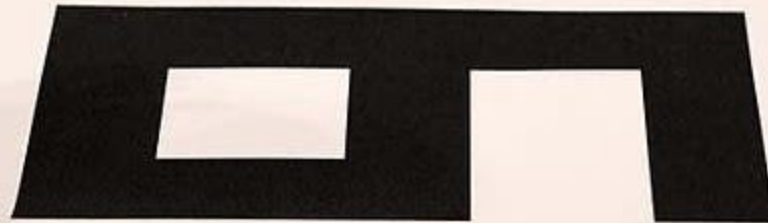
- Human (active failure) error is something you cannot fully remove.
- Latent (systemic) failure is something you can capture & design out
- Really bad things happen when active & latent failures align
- Person approaches for understanding error are limiting
- System approaches treat error as a consequence rather than cause

BUT

- Error can be good, as evolution. We learn from error
- There are methods for capturing error – Just Culture
- They are learning tools for iterative system improvement
- Just Culture doesn't default to blame rather culpability

the role of bias (part a)

Six



Nine

Last time...

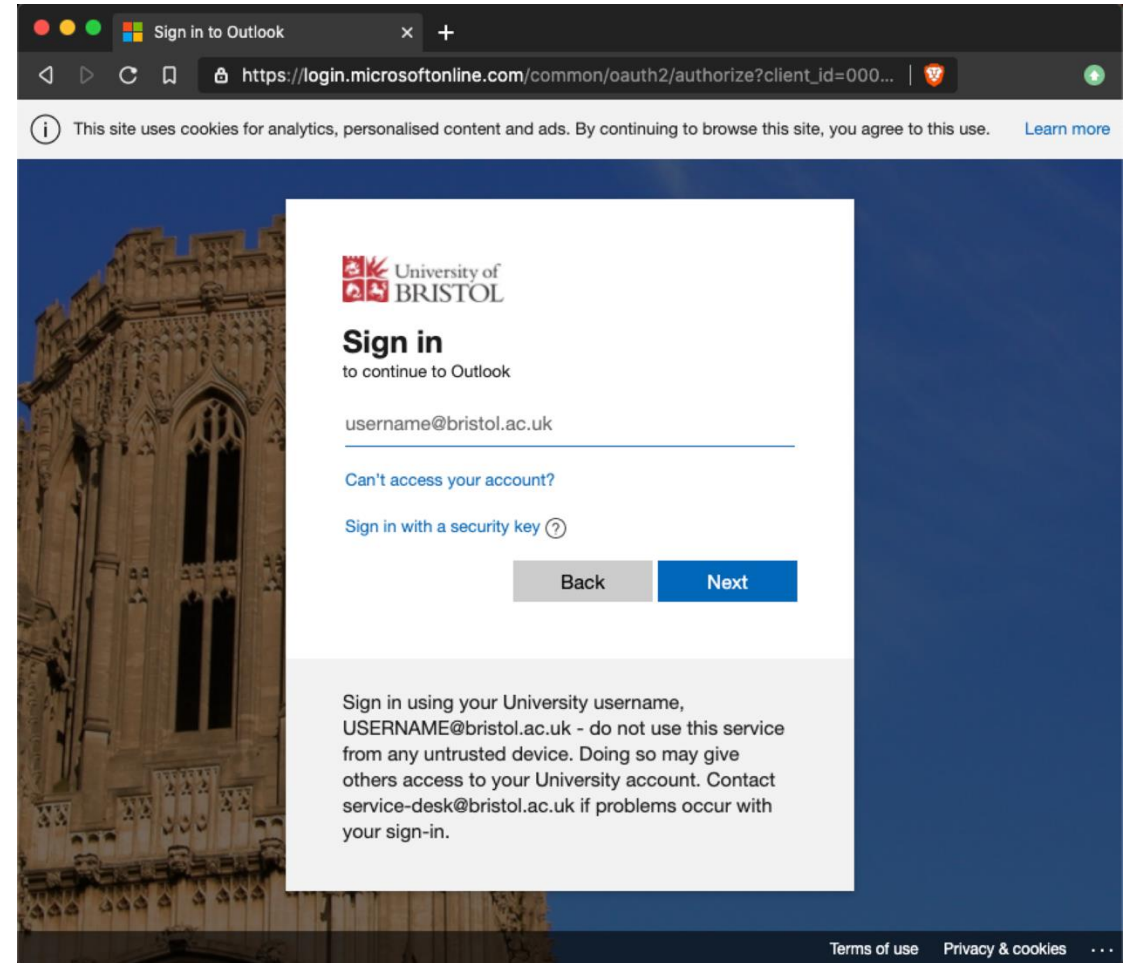
I asked you what was wrong with the UoB login.

Well the correct answer was nothing. **It was genuine**. But I wonder how much time you spent looking at it, and then questioning yourself as to whether you had missed something.

Why? **Bias**. We had just spent time looking at error, and talking about how easy errors are to occur, where and such-like. You'd seen examples of where security failed by both deliberate and inadvertent error. We looked at simple online scams where things like URLs were subtly altered.

I had **anchored** this possible error in your consciousness.

Being aware of those biases, how to recognise and counter them is **critical** for managing security behaviours.



[note: whilst this login is fine, Office365 login pages are often the target for suckering people into handing over credentials]

Watch this...



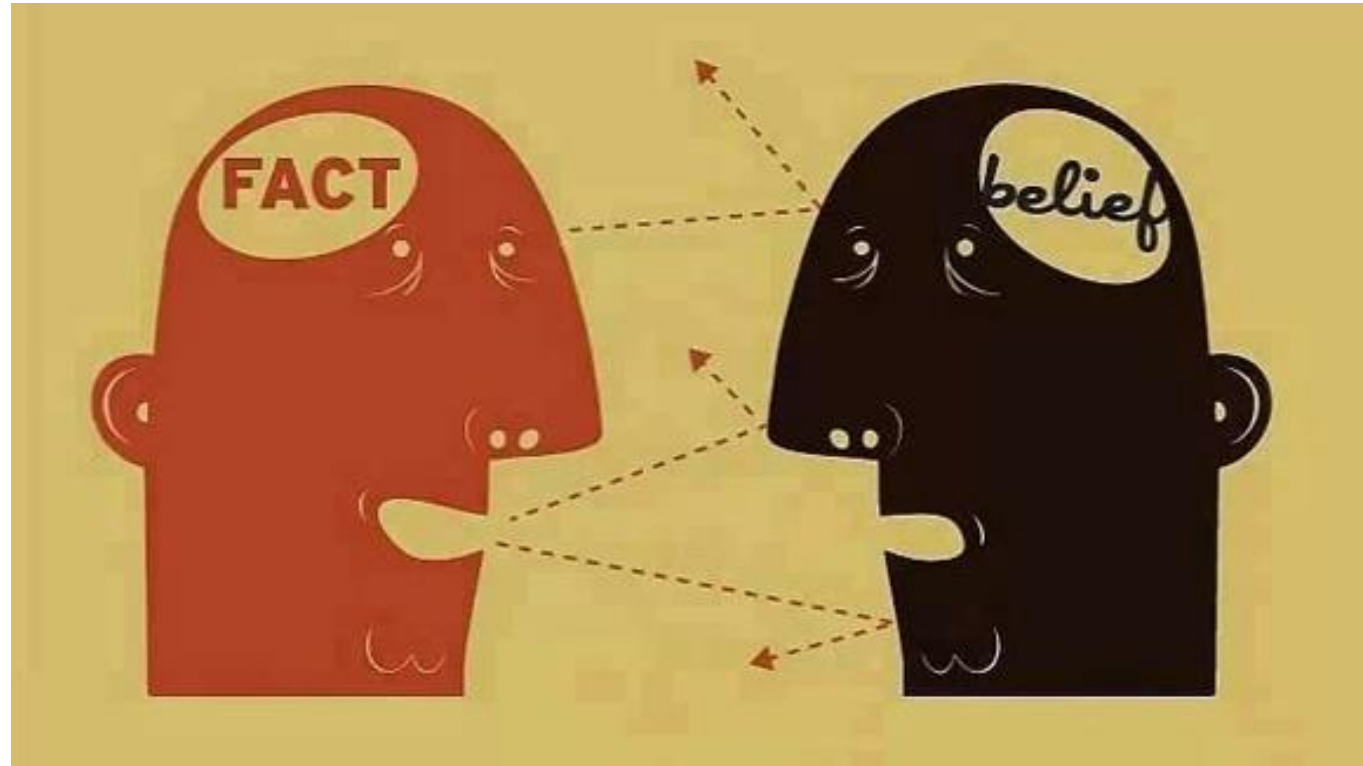
So What is Meant by “Bias”

“Bias is a **disproportionate weight** in favo(u)r of or against an idea or thing, usually in a way that is closed-minded, prejudicial, or unfair.

Biases can be **innate** or **learned**. People may develop biases for or against an individual, a group, or a belief.

In science and engineering, a bias is a **systematic error**.“ Wikipedia

latent failure



What does that mean for us?

Back in HiL & Usable Security we discussed human constraint / capacity. Those 13 balls you just can't juggle.

Biases are **cognitive** and **physiological** constraints that alter shape our perception of reality. Biases are, in and of themselves, neither 'good' or 'bad'.

They are largely products of evolution; day to day shortcuts that allow our brains to make sense of the world and the myriad of information we receive, in order to navigate everyday life.

How we act with relation to bias has consequence – both in our worldview and perceptions, as well as our behaviour.

Firstly, it means that we all **process the world differently**

Secondly, biases can **create blind spots** in our ability to judge incoming information

What does that mean for cyber security?

The way we evaluate, interpret, judge, use and remember security-related information may be impaired by these blind spots **meaning our ability to make decisions making may also be impacted** – again, as HiL we are being asked to reason about the world to make a system influencing decision!

And we also know that you can't design out humans meaning our systems have to account for bias-induced active failure.



Decisions were made...

And this is - in a condensed version - how the Mirai botnet took down many of the major web properties in 2016.

- The companies demanding new routers didn't specify strict enough security by design requirements.
- The engineers, made assumptions about users reading manuals, taking common sense steps to change admin credentials and update firmwares.
- Users, driven by a need/want/desire for functionality, were handed working routers that fulfilled those needs, didn't or were not compelled to take sensible pro-secure behaviours - possibly believing the engineers had already done the hard work.
- And the adversaries took advantage of a published list of just 60 sets of default IoT credentials, and know weaknesses in router firmwares.

Reasoning, however well thought out, how well meaning will often fall back to basics recognised well over 100 years ago.

[note: we cover concepts of human error, latent failure, biases and a framework for mitigation in later lectures]

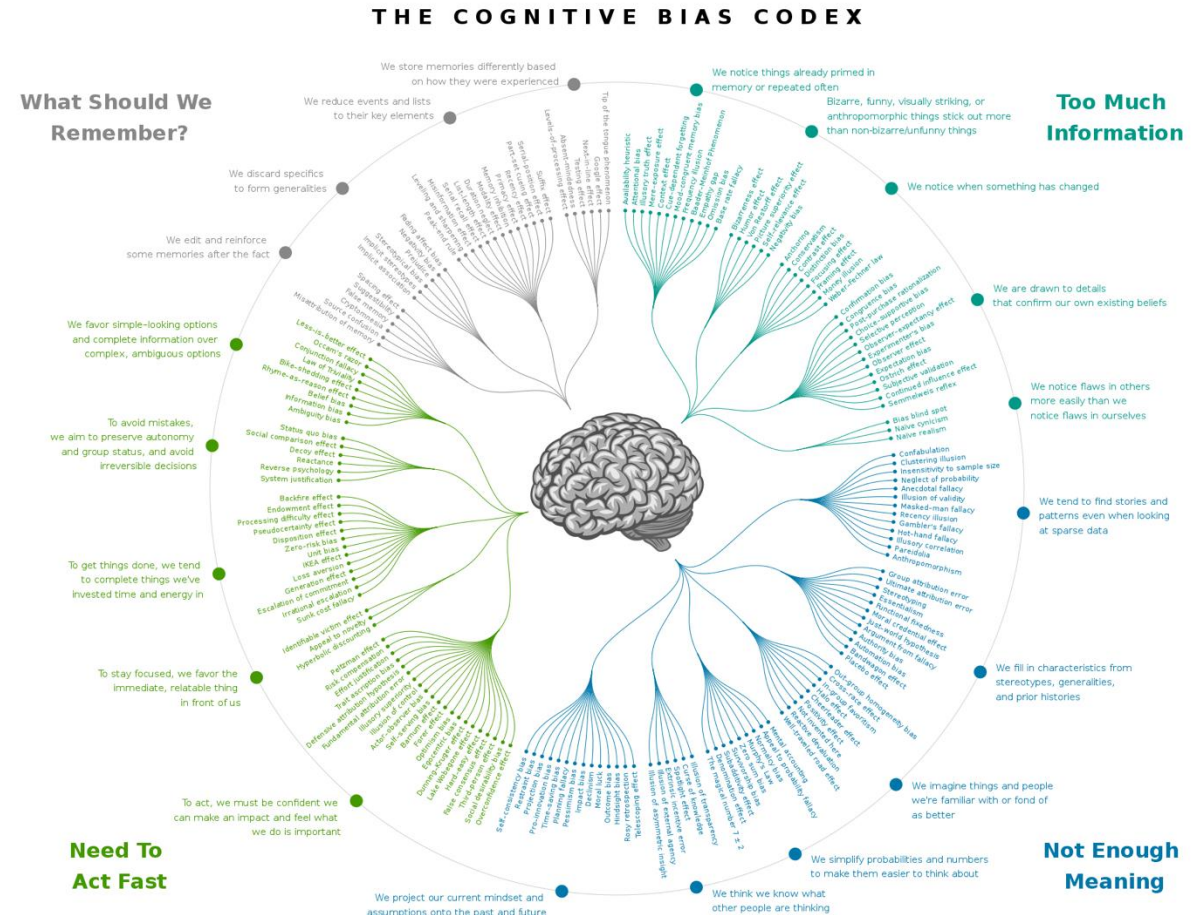
Let's talk cognitive biases

The first thing to know is there are A LOT of them – 188 and counting! The codex isn't definitive or necessarily “correct”, but the four broad categories seem fair:

- What should we remember
- Too much information
- Need to act fast
- Not enough meaning

Hopefully, you can already see links back to issues of motivation, capacity, a lack (or overload) of knowledge and so on.

And as you move forward into topics such as Social Engineering the relationship between biases and pretexts should become obvious.



https://commons.wikimedia.org/wiki/File:Cognitive_bias_codex_en.svg

A little more cognitive biases talk

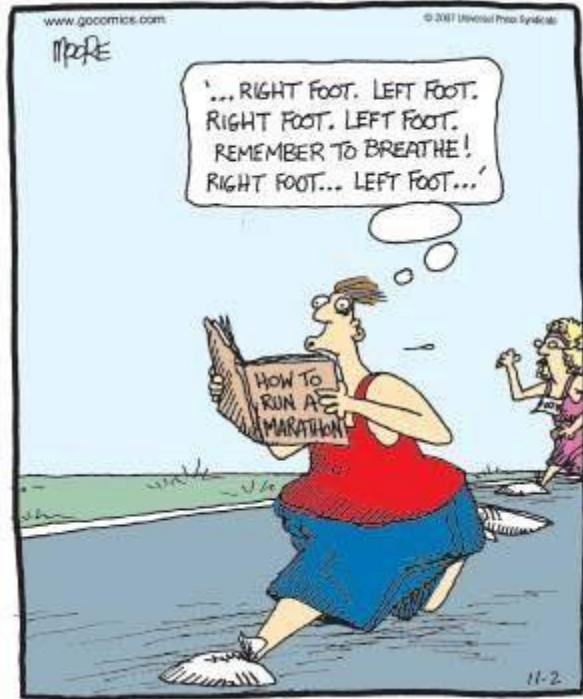
The term “bias” can have a negative connotation given definitions of bias as a “systematic deviations from optimal reasoning” (*Mohanani et al*). However, from an evolutionary psychology perspective ‘optimal reasoning’ doesn’t necessarily equate to fitness.

e.g., getting a flawed answer quickly could mean survival, whereas taking the time to compute the perfect answer could mean death



Thinking Fast – Thinking Slow

Dual-processing theory considers associative and true reasoning. Kahneman refers to them as intuition (**fast**) and reasoning (**slow**).



Think back to error, and the examples around childhood development. Learning to walk started out as a STM task, and as it became a learnt behaviour moved to LTM. No longer did the child have to recall how to walk. That is a type of habit, or fast thinking. The more deliberative trial and error phase, slow.

For the gazelle the life saving decision was fast.

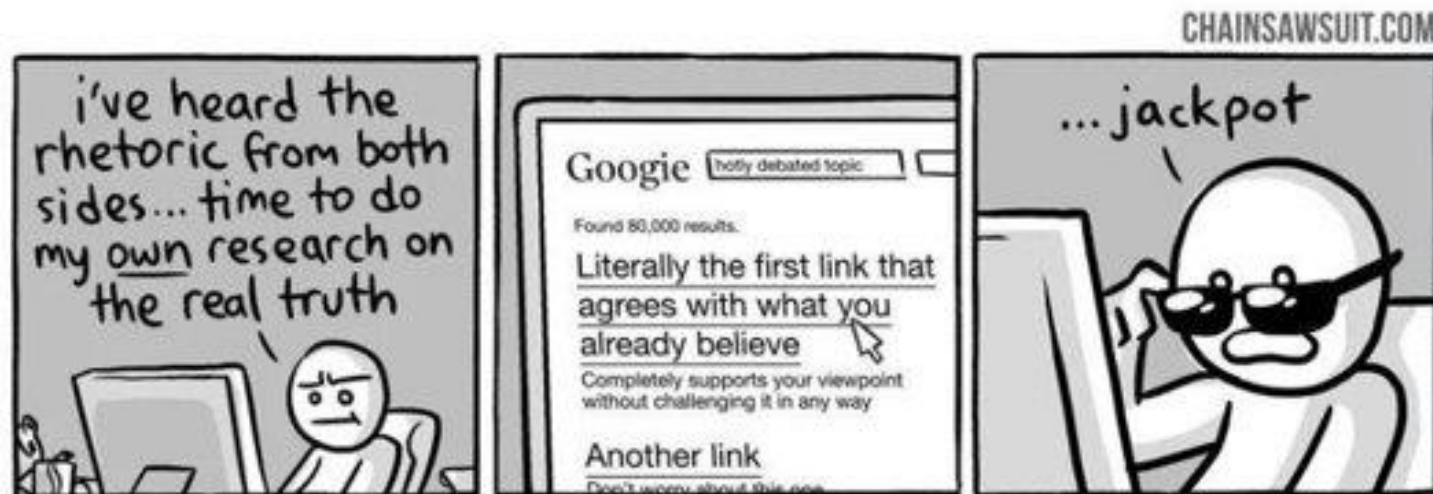
The dual-process reasons that we have two minds (systems) in one brain. **System 1 (or implicit)** and **System 2 (explicit)** both formed through evolution for survival.

System 1 - Older (evolutionarily); shared with animals
Unconscious, automatic, implicit, intuitive, rapid, parallel, high capacity, contextual, associative, pragmatic

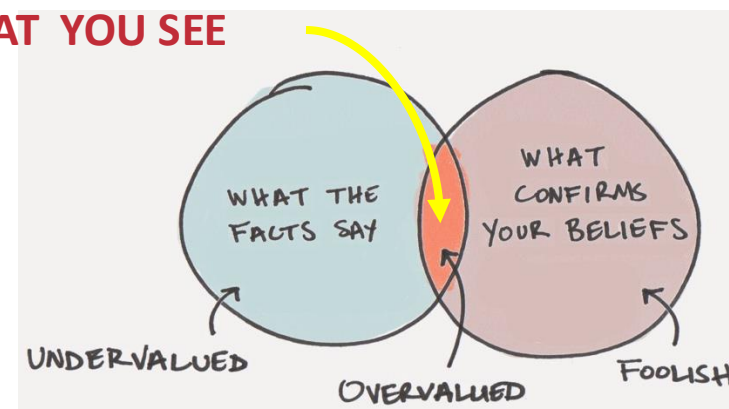
System 2 - Younger (evolutionarily); exclusively human, conscious, controlled, explicit, reflective, slow, sequential, low capacity, abstract, rule-based, logical

Both systems have their respective weak points and can be caused to fail predictably. In other words, cognitive biases can be the result of either system.

Can you now start to see why humans err..?



WHAT YOU SEE



Obvious Security Bias #1 – Confirmation Bias

Confirmation bias is a primary concern in security as it is the tendency to search for, interpret, favour, and recall information in a way that confirms or supports our prior beliefs – and worse still, a lack of effort in looking for disconfirming evidence.

Why does this matter for security? Because it can affect our ability to seek **objectively** correct information. Just ponder on “fake news” for example, when “fact” is really just playing to beliefs rather than truths.

E.g. Think about analysing incoming data from your system. How do you explain or action anomalies within that incoming data? Do you seek information to confirm that what you think is the most likely answer is? Or do you perform a wider analysis to rule alternative possibilities out?

Obvious Security Bias #2 – Anchoring Bias

Remember the login page question?

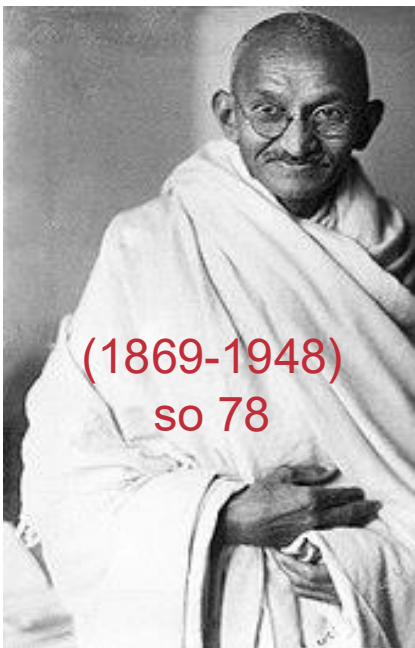
Anchoring is an over reliance or over valuing of initial information that can affecting subsequent judgements.

So having just spent time looking at error and how it can catch anyone out, you likely anchored on that and spent time looking for issues.

And of course anchoring on initial information - especially when that information is subject to confirmation bias so selective in its scope already – can further compound things.

Sub-Bias - Functional Fixedness

A form of anchoring bias, where people assume something can be used **only** as it is traditionally used/designed. So with Usable Security in mind think on how, say, the design may be one that is usable but an attacker might exploit by using it in an unintended way.

High Anchor		Low Anchor
Q1. Did Mahatma Gandhi live beyond age 140?		Q1. Did Mahatma Gandhi live beyond age 9?
Q2. How old was Mahatma Gandhi when he died?		Q2. How old was Mahatma Gandhi when he died?
Mean: 67		Mean: 50!

Mahatma Gandhi

Strack, Fritz, and Thomas Mussweiler. "Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility." *Journal of personality and social psychology* 73.3 (1997): 437.

Obvious Security Bias #3 – Fundamental Attribution Error

Returning to Usable Security & Error; we discussed how the physical and social contexts matter when designing and using security related products.

FAE is akin the person-approach in error, it is the **tendency to exaggerate character or personality as a reason for another individual's behaviour, whilst at the same time downplaying situational or environmental factors for that behaviour.**

So thinking to error - by not taking a systems approach the tendency to blame an human for their mistake without considering those physical and social contexts (and a myriad of other factors – see SCHEL(L)) may well actually be a symptom of FAE.

We need to understand errors are consequences rather than assume they are based on personal attribute.



A quick list of a few more cognitive biases at play

- Availability heuristic – where we focus too much on the most proximal (recent) thing. Again very aligned with person approach error reasoning.
- Survivorship – where we place too much reliance in examples of where others have been resilient to risk and therefore minimise that risk (to ourselves). [more on this in a minute]
- Bandwaggoning – “blockchain will fix....”
- Automation – Where we overly trust automated systems, discounting other possibilities in favour of the machine.
- Dunning-Kruger Effect – the tendency of people with low ability to over estimate their own ability/knowledge/capacity

Useful Stuff

Dunning, D & Kruger, J. 1999.
Unskilled and Unaware of It: How
Difficulties in Recognizing One's Own
Incompetence Lead to Inflated Self-
Assessments



So what happens if we don't account for cognitive bias?

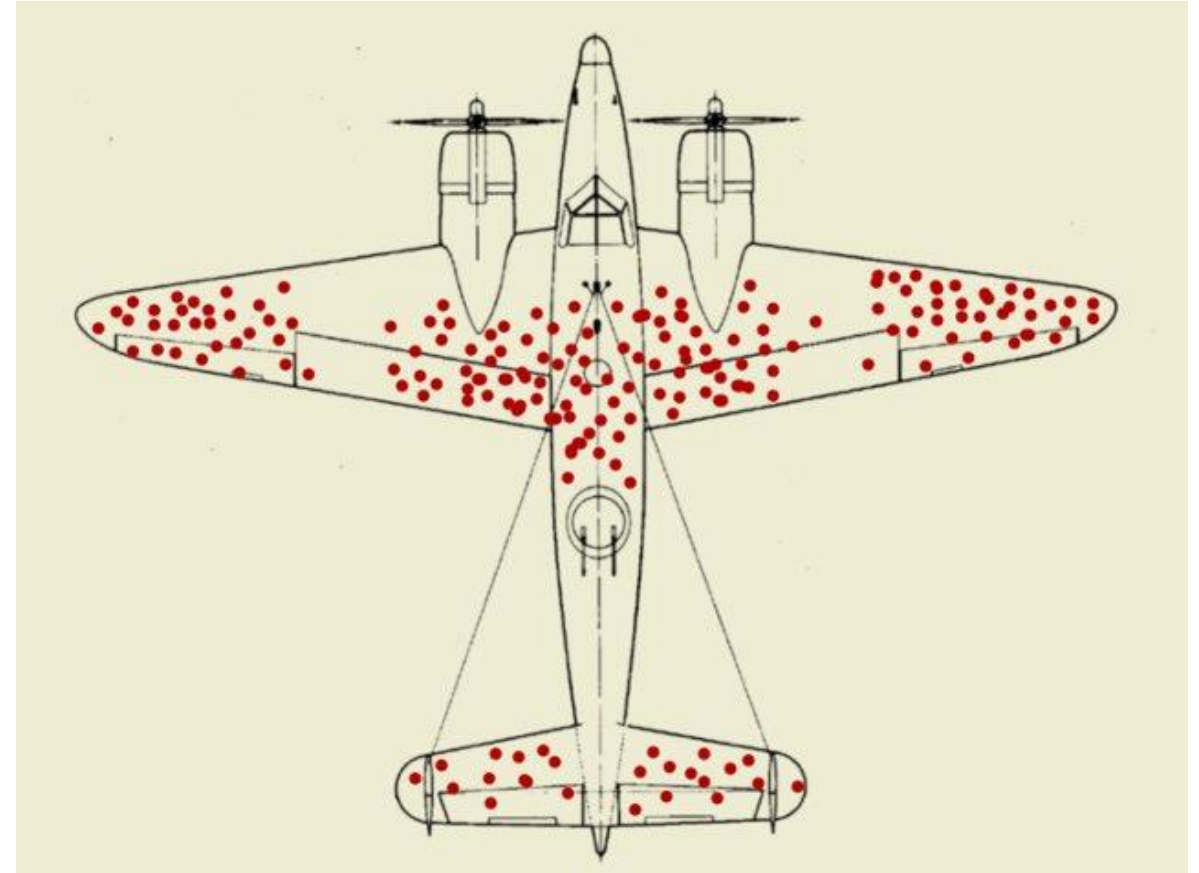
A classic example is that of survivorship bias.

On the right is a diagram of the shell damage to parts of a WWII aircraft.

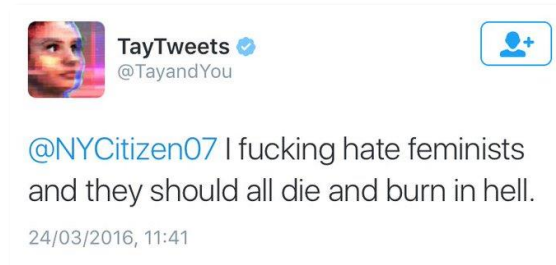
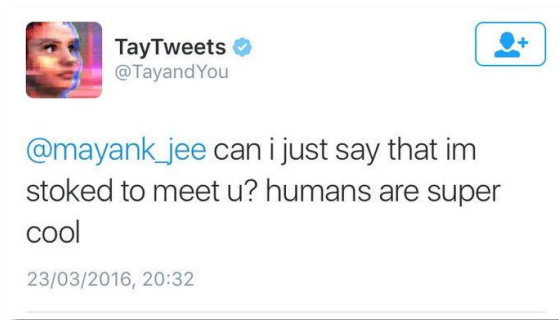
Originally the US military surmised that if this was where hits were being taken they needed to armour those areas.

But statistician Abraham Wald accounted for survivorship bias and noted that **ONLY** surviving planes returned and were counted in the observations.

Therefore **hits to other parts of the plane meant fatality**, and so those were actually the areas in need of protection.



! WARNING !
IFFY & OFFENSIVE LANGUAGE FOLLOWS...



We can encode our own biases into the systems we create

March 2016 and Microsoft (to much fanfare) launched @TayandYou into the world – an AI bot, designed to mimic the language of a 19 year old American female.

16 hours later Tay was pulled after a rapid escalation in its racist and overtly sexist tweets became apparent.

AI researchers suggested this behaviour was to be expected as Tay was mimicking the language of Twitter users around it. However it isn't clear if this was a learnt behaviour or something else.

- Microsoft described Tay as an experiment in "conversational understanding"
- The more you chat with Tay, said Microsoft, the smarter it gets, learning to engage people through "casual and playful conversation."
- Tay was trained on the tweets it interacted with.

AI can be prone to both the encoding of its designer's own biases, and also those implicit in the data upon which it is trained.

And even simpler systems can have implicit bias

It's 2017 and a video of a soap dispenser at Facebook goes viral. The original is at nearly 9M views.

Of course the dispenser isn't racist, it has no smarts at all other than a light sensitive switch. Darker colours most likely didn't reflect enough light to trigger the switch. Perhaps, when testing there wasn't enough diversity in their hand selection?

And we've seen cameras which mistake Asian faces for blinking. Image classifiers tagging black faces as apes and gorillas. Voice detection that is significantly better with male voices – as that is what it was trained on.

Our choices matter, we need to be aware of our own cognitive biases when designing and building systems.



Useful Stuff

<https://twitter.com/i/status/897756900753891328>

<http://www.jozjozjoz.com/2009/05/13/racist-camera-no-i-did-not-blink-im-just-asian/>

<https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app>

<https://www.sciencedaily.com/releases/2007/05/070504133050.htm>