

Weighing evidence in an information war

Paul McKeigue

Usher Institute of Population Health Sciences and Informatics

23 January 2019

Fake news and disinformation

Government response to the House of Commons Digital, Culture, Media and Sport Committee (2018):

- *'fake news' is a poorly-defined and misleading term that conflates a variety of false information, from genuine error through to foreign interference in democratic processes*
- *the Government has sought to move away from 'fake news' and instead has sought to address 'disinformation' and wider online manipulation.*
- *we have defined disinformation as the **deliberate creation and sharing of false and/or manipulated information that is intended to deceive and mislead audiences**, either for the purposes of causing harm, or for political, personal or financial gain.*
- *we will consider options to **improve critical thinking skills and resilience to disinformation** in the context of political engagement.*

Disinformation, conspiracy theory or truth? – some disputed explanations

Year	Event	Official explanation	Alternative explanation
1981	Yellow rain in Laos and Cambodia	Communist mycotoxin warfare	Mass defecation flights of Asian honeybees
2001	WTC collapse	Hijacked aircraft	Planned demolition
2006	Litvinenko poisoning	Russian assassins	Accidental mishandling of polonium
2013-18	Alleged chemical attacks in Syria	Regime chemical warfare	Managed massacre of captives
2016	Brexit referendum result	Revolt against migration policy	Manipulation of voters by informatics companies
2017	Noise-induced illness of diplomats in Cuba	Communist sonic weapons	Calling song of Indies short-tailed cricket
2018	Salisbury poisonings	Russian assassins	Provocation intended to damage relations with Russia

How degrees of belief are updated by evidence

Wrinch and Jeffreys (1921) rewrote Bayes theorem in terms of the *odds* favouring one hypothesis \mathcal{H}_1 over another \mathcal{H}_2



Dorothy Wrinch

$$(\text{prior odds } \mathcal{H}_1 : \mathcal{H}_2) \times \frac{\text{likelihood of } \mathcal{H}_1}{\text{likelihood of } \mathcal{H}_2} = (\text{posterior odds } \mathcal{H}_1 : \mathcal{H}_2)$$

- **prior odds** encode your subjective degree of belief favouring \mathcal{H}_1 over \mathcal{H}_2 , before you have seen the observations.
- The **likelihood** of a hypothesis is the conditional probability of the observations given that hypothesis.
 - “Support” would be a better word than “likelihood”

Likelihoods and Bayes factor

- We can think of the likelihood as measuring how well the hypothesis would have predicted the observations.
 - to evaluate the likelihood, we have to envisage what would be expected to happen if the hypothesis were true.
- The ratio $\frac{\text{likelihood of } \mathcal{H}_1}{\text{likelihood of } \mathcal{H}_2}$ is the **Bayes factor** favouring \mathcal{H}_1 over \mathcal{H}_2 .
 - **likelihood principle**: all evidence for \mathcal{H}_1 versus \mathcal{H}_2 is contained in the Bayes factor.
 - If you make an assertion about the strength of evidence favouring one hypothesis over another, you are making an assertion about the magnitude of the Bayes factor.
 - Likelihoods on their own have no meaning - information is conveyed only through the ratio of likelihoods

Example: accumulation of evidence

Two alternative hypotheses about a coin that is to be tossed: \mathcal{H}_1 that the coin is fair, \mathcal{H}_2 that the coin is two-headed. In most situations your prior belief would favour \mathcal{H}_1 over \mathcal{H}_2 .

- Suppose the coin is tossed once and comes up heads
 - likelihood of a fair coin is 0.5, likelihood of a two-headed coin is 1. The Bayes factor favouring a two-headed coin over a fair coin is 2. Your posterior odds are now twice your prior odds.
- Suppose the coin is tossed another nine times, and comes up heads every time
 - Bayes factor is $2^{10} = 1024$.
 - Do you now suspect that someone has got hold of a two-headed coin?

Taking logarithms: weight of evidence

The logarithm of the Bayes factor is called the **weight of evidence** favouring \mathcal{H}_1 over \mathcal{H}_2 . As taking logarithms replaces multiplying by adding, we can rewrite Bayes theorem as

$$\log(\text{prior odds}) + \text{weight of evidence} = \log(\text{posterior odds})$$

- Weights of evidence can be added over independent observations, like physical weights

If we use logarithms to base 2, the units of measurement of weight are **bits** (binary digits).

- For the coin example, weight of evidence favouring \mathcal{H}_2 (two-headed) over \mathcal{H}_1 (fair) is
 - one bit ($2^1 = 2$) after the coin has been tossed once and come up heads
 - 10 bits after the coin has been tossed 10 times coming up heads each time.

Practical applications



Hut 8, Bletchley Park



Alan Turing



Jack Good

- Alan Turing (1941):
 - first stage of Enigma codebreaking procedure was based on accumulating weights of evidence
 - derived sampling properties of the weight of evidence
- Jack Good (1950 to 2003):
 - wrote up and extended Turing's unpublished results
 - showed relationship of weights of evidence to information theory
 - discussed wider applications for instance to jury trials

Subjectivist view of probability

- Classical probability theory - in situations like coin-tossing and throwing dice, probabilities are imposed by physical symmetries.
- Bruno de Finetti (1931) - quantitative coherence of betting odds implies that subjective degrees of belief must obey rules of probability theory.
 - can elicit your subjective probability of an event as the price you would offer or accept for a ticket that will pay out £1 if the event occurs, and nothing if the event does not occur.
 - if the prices you specify over various combinations of events are not consistent with probability theory, someone else can construct a “Dutch book” against you.
- Richard Cox (physicist, 1946) - degrees of belief must obey the rules of probability theory if they satisfy simple axioms of logical consistency.
- Edwin Jaynes (physicist, 2002) - probability calculus is an extension of the ordinary rules of logic to uncertainties about propositions.

Some general principles, in non-mathematical language

- You cannot evaluate the evidence for or against a single hypothesis, only the weight of evidence favouring one hypothesis over an alternative
- Weight of evidence favouring one hypothesis over another is based on comparing how well each hypothesis would have predicted the observations.
 - applies in many fields: medical diagnosis, historical investigation, intelligence analysis, criminal trials.
- Assessment of the weight of evidence favouring one hypothesis over another does not depend on your degree of belief in which of these hypotheses is true.
 - people with different prior beliefs about the plausibility of the hypotheses being compared should be able to reach consensus on the weight of evidence.

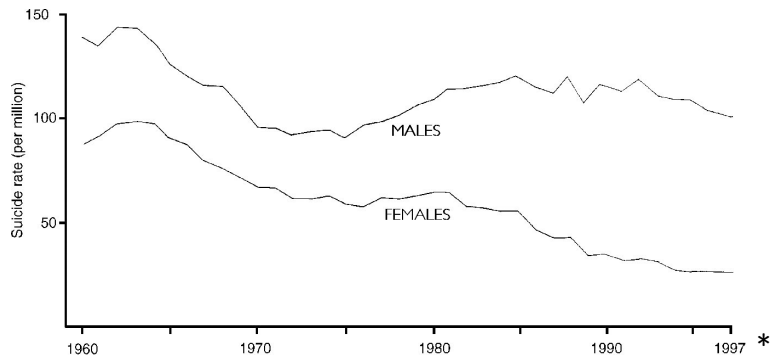
Weight of evidence is based on comparing hypotheses

To evaluate the evidence favouring one hypothesis over another, you have to assess, for each hypothesis in turn, the probability of the observations given that hypothesis:

- Mandy Rice-Davies (1963), when the barrister put to her that Lord Astor denied having an affair with her:

“Well he would, wouldn’t he.”

Weight of evidence is based on comparing hypotheses



Geoffrey Rose (epidemiologist, 1926-1993), considering the 35% fall in suicide rates in England & Wales between 1963 and 1975:

"You can't exclude the explanation you haven't considered."

Hempel's paradox of confirmation

"An observation that is consistent with a hypothesis is not necessarily evidence in favour of that hypothesis."

Good (1967): not a paradox but a corollary of Bayes theorem

- On island A, there are 32768 (2^{15}) birds of which 64 (2^6) are crows and all these crows are black.
- On island B, there are 32768 (2^{15}) birds of which 4096 (2^{12}) are crows and 512 (2^9) of these crows (that is, one eighth of all crows) are black.

You wake up on one of these islands and the first bird that you observe is a black crow. Is this evidence that you are on island A, where all crows are black?

The confirmation fallacy: interpreting an observation that is consistent with a hypothesis as support for that hypothesis

- On island A, there are 2^{15} birds of which 2^6 are crows and all these crows are black.
- On island B, there are 2^{15} birds of which 2^{12} are crows and 2^9 of these crows (that is, one eighth of all crows) are black. You observe a black crow. Is this evidence that you are on island A, where all crows are black?
 - Likelihood of island A, given this observation, is 2^{-9} .
 - Likelihood of island B, given this observation is 2^{-3} .
- Weight of evidence favouring island B over island A is $[-3 - (-9)] = 6$ bits.

Alleged chemical attack in Ghouta in 2013: regime attack or false flag chemical attack?

21 August 2013, opposition-held eastern suburbs of Damascus:

- ~500 bodies appeared at hospitals and improvised morgues in the middle of the night.
- Images of some victims still alive and being treated were uploaded by opposition media outlets.
- Chemical attack with rockets was reported.

Two alternative hypotheses were initially proposed:

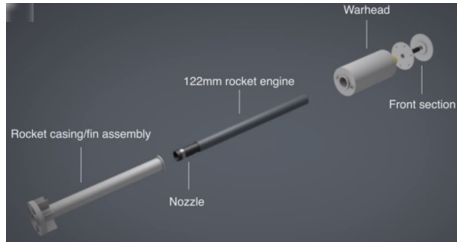
- \mathcal{H}_1 (UK Joint Intelligence Committee): “it is highly likely that the regime was responsible for the CW attacks on 21 August”
- \mathcal{H}_2 (Saar Wilf, Israeli data scientist/entrepreneur): a false-flag chemical attack was carried out by the Syrian opposition, with the object of drawing the US into an attack on Syria.

Munition allegedly used in Ghouta in 2013

“Volcano” rocket had previously been used by Syrian army as improvised siege weapon.



Alleged impact site



Rocket components

Is the type of munition evidence for the regime attack hypothesis?

Eliot Higgins (2013):

“There is no evidence of Syrian rebel forces ever using this type of munition — and only Syrian government forces have ever been shown using them.”

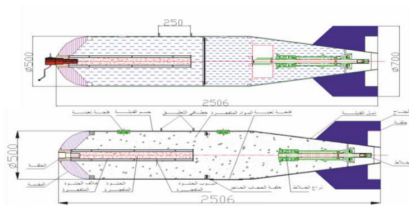
Patrick Radden Keefe, Daily Telegraph (2014)

“The proof that the Syrian regime had used chemical weapons came from an unemployed blogger working from his sofa”

Likelihood of regime attack hypothesis, given the observed type of munition

To evaluate the likelihood of \mathcal{H}_1 , put yourself in the shoes of a Syrian general ordered to carry out a chemical attack: you would expect the Syrian military to use its stocks of munitions designed to deliver nerve agent.

Syrian chemical bomb declared to OPCW in 2013: 2 compartments for binary precursors, mixed by turning crank to rupture the membrane.



Improbable that Syrian military would use Volcano rocket: no mechanism for mixing precursors, range of only ~2 km would necessitate launching from the front line.

Likelihood of \mathcal{H}_1 given the observation is < 0.1

Likelihood of opposition false flag, given the observed type of munition

To evaluate the likelihood of \mathcal{H}_2 , put yourself in the shoes of an opposition commander planning a false-flag chemical attack. Two possible ways to implicate the regime:

1. to fake an air strike, with fragments of air-delivered munitions matching something in the government arsenal, or
2. to use rockets or artillery shells matching something in the government arsenal. Volcano rockets, captured from Syrian army stocks or copied, would have been ideal.

If we assign equal probabilities to these two options under \mathcal{H}_2 , the likelihood of \mathcal{H}_2 given the observation is 0.5.

- Weight of evidence favouring \mathcal{H}_2 over \mathcal{H}_1 is at least $\log_2 \frac{0.5}{0.1}$:
2.3 bits

Quantitative approach identifies where analysts differ

Where two analysts evaluating the same observations disagree on the weight of evidence, working through the probability calculation will identify where their assumptions differ, and may suggest how to resolve these differences:

- seek more information on a key observation
- identify a reference class of events that can be used to assign probabilities

Evidence contributed by the absence of something expected

Sherlock Holmes: “curious incident of the dog in the night-time” - the dog did not bark when the horse was taken.

Some dogs that did not bark in the alleged chemical attacks in Syria:

- From Ghouta 2013 and Khan Sheikhoun 2017, no images of search and rescue operations appeared.
- No family photos with victims were produced by bereaved survivors
- No funerals: most bodies were buried in unmarked mass graves
- Relevant laboratory results not released by OPCW:
 - no quantitative results of tests for sarin exposure
 - no scientific papers from the labs
 - no multivariate chemical signature attribution studies

Alleged chemical attack in Ghouta in 2013: regime attack or managed massacre?

Of ~20 local media ops that uploaded videos, not one showed search and rescue of victims. This led to hypothesis \mathcal{H}_3 : managed massacre of captives, with rockets and sarin used to lay a trail of forensic evidence.

For each of the following observations, we can assess the probability given \mathcal{H}_1 and the probability given \mathcal{H}_3 :

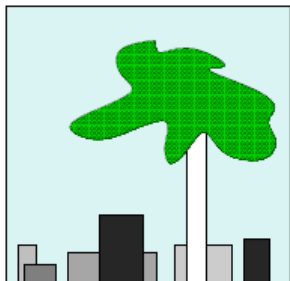
- The only video of bodies purportedly found *in situ* (Zamalka Ghost House) was an obvious fraud: family group executed beforehand in unfinished building.
- Only a few burials were shown: most victims disappeared into unmarked mass graves.
- O'Brien (2014) - reconstruction of images from Kafr Batna shows that one of the victims woke up in the morgue and had his throat cut by the "emergency responders".

How probable is a managed massacre for the purpose of a false flag?

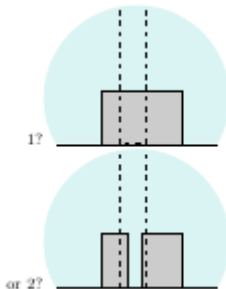
- “Laundering” of massacres to blame the other side is common in war.
- Massacres undertaken for the purpose of a false flag have been rare, at least until recent years
 - market attacks in Sarajevo (1992, 1994, 1995) alleged to be false flags
 - “humanitarian intervention” in Yugoslavia and Libya may have provided an incentive.
- Conditional and unconditional probabilities are often confused:
 - fallacious argument used by defence lawyer in OJ Simpson trial.
- given that a massacre has occurred and that there is a prospect of humanitarian intervention, the probability of a managed massacre is not negligible.

Hypothesis testing and Occam's razor (Mackay 2003)

Accept the simplest explanation that fits the data



A tree and some boxes

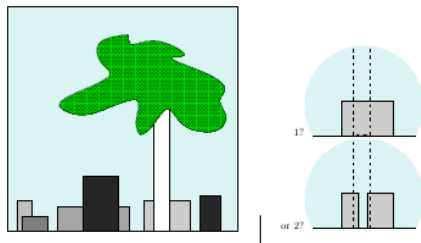


One box, or two?

Both hypotheses - single box behind tree, or two boxes - explain the data equally well.

The two-box hypothesis relies on the improbable coincidence that both boxes have same height, same colour, and each has one edge hidden behind the tree or behind the other box.

Occam's razor: a corollary of Bayes theorem



Suppose there are 16 equally probable box heights, 16 equally probable colours, and equal probabilities for an edge of a box to be visible or hidden

Under the two-box hypothesis, the probability that the second box has the same height and same colour as the first is $\frac{1}{16} \times \frac{1}{16} \times \frac{1}{2} \times \frac{1}{2}$.

- weight of evidence in favour of a single box is $-\log_2 \frac{1}{1024} = 10$ bits.

Hypotheses that adapt to the data to explain the observations are automatically penalized by Bayesian inference.

Collapse of the World Trade Center's Twin Towers

- \mathcal{H}_1 : planes crashed by hijackers
- \mathcal{H}_2 : planned demolitions set to coincide with the plane crashes

A nearby smaller tower (Building 7), collapsed several hours later.

- Likelihood of \mathcal{H}_1 given the observation
 - probability of collapse of a nearby smaller tower given \mathcal{H}_1 is low (unusual construction?).
- Likelihood of \mathcal{H}_2 given the observation
 - Proponents of \mathcal{H}_2 explain the observation by adapting an unknown variable - the intended target of the fourth plane - to be Building 7.
 - as the three buildings attacked were symbols of American military or economic power, it is probable that the fourth target was another such building, rather than an unremarkable building in the WTC complex.

Weight of evidence contributed by the observation favours \mathcal{H}_1 .

Khan Sheikhoun 2017: regime attack or managed massacre?

\mathcal{H}_1 : chemical attack by the Syrian air force using sarin.

\mathcal{H}_2 : deception operation involving a managed massacre of captives and small quantities of sarin used to lay a trail of forensic evidence.



A “mountain of evidence” (Monbiot) for \mathcal{H}_1 ?

- images of alleged impact site
- images showing victims in quarry, trucks, or morgues
- interviews with purported eyewitnesses
- environmental and blood samples tested positive for sarin.

Operation Mincemeat: a legendary deception operation

April 1943: badly decayed body dressed in Royal Marines uniform and chained to briefcase, found by fisherman off the coast of southern Spain.

A mountain of evidence for the identity of Major William Martin – or an echo chamber?



- documents included identity card, letters, receipts.
- listed among British casualties published in *The Times*.
- cables from London to the British vice-consul emphasized importance of recovering the briefcase.

Observations contribute evidence only if they are more probable under one hypothesis than under the other.

How would a managed massacre of captives be planned?

- A well-staffed media operation would be ready to edit the raw footage into clips and stills badged with the logos of opposition media organizations.
- Improvised explosive devices and possibly smoke generators could be placed at key locations in the town to make the civilian population believe they were under chemical attack.
- Low doses of sarin could be administered to volunteers so that they would test positive for exposure.
- Actors could be prepared to play the part of bereaved parents, and provided with photos of children who were to be killed.
- Captives (most likely religious minorities or families of government supporters) would be killed with a gas that leaves no residue
 - for maximal emotional impact, the victims should include a high proportion of children, and some would be left alive to be filmed before they were finished off.

What observations could not easily be generated from a managed massacre?

- Documentation that the victims seen dead in the images had lived in the locality from which they were supposedly rescued
- Interviews with bereaved survivors with documentation that the dead victims were their relatives, including family photos with adults or older children seen among the victims.
- Videos of search and rescue operations – hard to stage on a large scale without the cooperation of civilians.
- A multivariate chemical signature match between the environmental sarin samples and Syrian military stocks, with named authors and enough detail for peer review
- Quantitative results of blood tests on “survivors” showing exposure to sarin at levels high enough to have caused severe poisoning.
 - tests for sarin exposure can detect exposure at levels far lower than those required to cause symptoms.

What observations would be likely to give away a managed massacre (improbable in a real chemical attack, probable in a managed massacre)?

- Flight track inconsistent with alleged impact site (for an attack alleged to have been by air)
 - rebels might expect an airstrike but would not know the flight track in advance
- Alleged location of impact site and victims not consistent with wind direction
 - locations would have to be planned in advance
- Video evidence that scenes were staged or that victims were captives
 - dozens of video clips and still images that are meant to show rescue workers dealing with large numbers of victims have to be recorded, edited and uploaded in a few hours.
 - Timing of images and matching of identities of victims across clips and stills may reveal an incriminating story.

Evaluating the weight of evidence: five key observations from Khan Sheikhoun

- No images of search and rescue
 - White Helmets have helmet cams and are famous for documenting their operations on video
- Key witness had no photos of his deceased wife
 - reported that his house had been destroyed in an airstrike.
- Wind direction inferred from smoke plumes
 - alleged location of victims was upwind of alleged impact site
- Children with head injuries received after “rescue” by White Helmets.
 - at least two children shown in morgues with recent head injuries did not have these injuries when first “rescued”.
- Flight track of the Syrian jet
 - UN/OPCW Joint Investigative Mechanism reported that jet was no closer than 5 km from the town.

Laying the basis for disinformation operations: Strategic Communication

- National Security Council: StratCom is “the systematic and co-ordinated use of all means of communication to deliver UK national security objectives by influencing the attitudes and behaviours of individuals, groups and states”.
- UK government StratCom programmes are largely covert:
 - Syria programme (since 2012): “UK's largest of its kind since the Cold War” can be discerned from FCO expenditure records.
 - Russia programme (since 2015, now named Counter Disinformation and Media Development): UK is represented as being under attack, requiring a counter-offensive (“Integrity Initiative”) by military information warfare specialists.
- “Narrative development” is outsourced to NGOs and private companies (often established by ex-military officers): creates echo chamber of seemingly independent sources.
- Journalists and academics are recruited into “clusters”

What can universities do to “improve critical thinking skills and resilience to disinformation in the context of political engagement”?

