**Weighing the Fog of War**

Niall MacKay, Chris Price and Jamie Wood use a naval battle of the First World War to explain how Bayesian thinking can aid the study of history

What exactly is it that a historian does? Academic historians are uninterested in simple narrative, the “just one damned thing after another” trope, and strive to explain *why* things happened as they did. In peer review they must convince of the “importance” of their research to historical understanding. To do so inevitably admits the possibility that things might have happened differently, but this is to open a Pandora’s box of ramifications, of supposition piled on supposition. The difficulty of reasoning one’s way through these problems – and the ease with which intentionally misleading counterfactuals can be created by charlatans – is perhaps why, as renowned historian Michael Howard has said, “Grown-up historians don’t waste time on counterfactuals”. Implicitly, however, they must at least admit alternatives, or history is indeed reduced to narrative.

The term “counterfactual history” does not help matters, for it suggests the serious consideration of impossibilities and the generation of fantasies. The opposite conceit, however, is to assume that human history is a rational and ordered process. This itself a manifestation of intellectual weakness, a flinching from the frightening acknowledgment that human affairs are heavily contingent – a truth recognised by Harold Macmillan in his “events, dear boy, events”. There are indeed tides in the affairs of men and women, but these are generated, perturbed and deflected by actual occurrences. “Grown-up historians” do not believe in fate; rather they accept that no event is inevitable, simply more or less probable. Thinking about “what ifs?” or even “if onlys” (when emerging on the wrong side of the historical outcome) is a necessary part of deciding which historical events are important and which are not – which changed the course of history, and which are incidental. We need not trace everything back to the horse-shoe nail; but to the unexpected loss of the battle, at least, we can impute some future implications, especially where with hindsight we now understand how such a loss erroneously misled contemporary decisions.

More broadly, one could argue that the Bayesian mind-set is the natural one in which to conduct historical analysis, for the course of history develops within an envelope of probabilities surrounding the historical narrative. Like gamblers watching “in-running” or “live” odds, historians observe the perpetual game of history being played out, and alter their views in the light of events and newly-emergent historical facts. The same is true of the historical actors themselves, and it is then the historians’ job to understand these actors’ implicit prior estimates of chances, and how these changed as events unfolded. If we decide that a real event was unlikely, we alter completely our estimation of the reasonableness or otherwise of differing prior views, and of the correctness of subsequent analyses of the event.

This is a difficult task, and historians typically regard it as impossible. For example, Niall Ferguson, a rare advocate of “counterfactual” history, has nevertheless despondently asked “How exactly are we to distinguish probable unrealized alternatives from improbable ones?” Historians of all persuasions might reasonably ask its complement “How exactly are we to distinguish probable real events from improbable ones?”

In a recent paper for the journal *Historical Methods* we advocated Approximate Bayesian Computation as a partial answer.[[1]](#endnote-1) But it can only work in very special, uncommon circumstances. Essentially one needs a simple situation governed by a well-founded model with tightly estimated parameters. Then, when the actual historical outcome turns out to be extremely unlikely, one can argue that this must indeed be the case rather than that the models or parameters are wrong.

All of this was the case in the first all-big-gun naval battle of the dreadnought era. On 24th January 1915, the battlecruisers of the British and German navies lined up early in the morning for an engagement in which the fleets would steam in straight lines for several hours, exchanging fire at distance in what immediately became known as the battle of the Dogger Bank [see boxes]. Dogger Bank is an excellent case study for this type of analysis: the battle was a single act with no distinct phases or tactical development (a statistician might call it “stationary”) and with no light forces between the battle lines complicating matters.

**The model: re-creating the battle**

The simulation model is obvious and simple: each shell fired has a probability of hitting, independent of the number of ships or guns on either side. Thus, on average, each side causes damage in proportion to its numbers. All that is needed for this to hold is that each shell has a single ship as its target, and will not hit another by chance if it misses its original target. This was certainly the case at Dogger Bank. This model is usually attributed to the British engineer Frederick Lanchester, and has since been misapplied to many other military contexts in which it demonstrably does not hold. However, in its original context of big-gun naval war, it is surely correct, and its inevitable conclusions were derived independently in France, Russia, the USA (twice) and by Lanchester (but not, it appears, in Germany): marginal numerical superiority results in an accelerating loss ratio and a large margin of victory, scaling by individual guns’ hit-rates multiplied by the *square* of their numbers.

We must first decide what is to be the unit in the model: is it the gun, the turret or the ship? The unit which naturally stands or falls together is the turret – unless the entire ship is vulnerable to a single critical hit. This typically occurred via ammunition fires, when fire spread through the cordite propellant [see box]. As we have seen, *Seydlitz* sustained such a fire at Dogger Bank, and the same nearly happened in *Lion*. The Germans benefited greatly from the explosion on *Seydlitz*, subsequently tightening up their handling procedures and thereby preventing future such explosions, in addition to which their cordite was inherently less volatile than the British. But the British learned no such lessons, and at the 1916 battle of Jutland, three British battlecruisers blew up from this cause. So one might infer that the British were very lucky not to lose at least one ship at Dogger Bank, and this is one of the central hypotheses we tested.

We used the basic Lanchester model, with prior parameters grouped by class of ship (to enable a hierarchical approach) and including probabilities for various outcomes – a hit, a hit on a turret, a critical hit causing an ammunition fire. We made no attempt to simulate the spatial aspects, here a simple exchange between battle-lines. Instead each ship was given a time at which it came into action, its target being chosen randomly and then changed at each step with a certain probability. This reflects the vagaries of targeting and conditions – at Dogger Bank there were frequent target changes. For simplicity, we did not take account of the arcs of fire of the turrets. This creates a slight inaccuracy in the fighting strength of ships with wing turrets, but both sides faced similar problems.

The simulation works by computing a firing propensity for each ship based on its rate of fire and the number of remaining turrets. The next ship to fire is then computed using a standard stochastic simulation algorithm, with each turret firing a single shot treated as an individual event. (Historically warships at least attempted to fire in salvos, but coordination was often lost in combat, so that very few salvos resulted in multiple hits.)

The ship-to-fire then has a certain probability of hitting its target. If a hit is achieved the probability of damage is then the product of two factors, which quantify the effectiveness of the attacker’s shell and of the defender’s armour. Damage here means effect on the Lanchestrian unit, the turret. If a damaging hit is scored then we also check if this is catastrophic – we include small probabilities for flash explosions and disablements which result in the complete loss of the ship. If not, then a single turret is lost from the target. For our prior parameters we used the average of Dogger Bank and Jutland – effectively, we are asserting that the sum of what happened at the two battles is the best starting point from which to approach the distribution of likely possibilities at Dogger Bank.

This process is then iterated for a fixed time, and the number of shells fired from each ship, hits received and so forth recorded for each run, constituting one re-creation of the historical battle. It is important not to attach too much meaning to the precise outcomes of the individual runs: we are primarily interested in the summative results of the battle in terms of hits and effect on the integrity of the ships. Whilst the exact states of the individual ships might be superficially attractive to assess as historical examples, to do so would be misconceived: we can only hope for a successful fit at an aggregate level.

**Bayesian fitting: re-creating the battle a million times**

Bayesian fitting is something like exhaustive, multiple war-gaming. We begin with prior estimates for the model’s parameters, together with the natural distributions for these. We then perform many millions of “runs”, each one a simulation of the battle, to assess how well these priors predict the real outcome. It is this sheer scale that gives the advantage over traditional (table-based or computerised) war-gaming: the space of possible outcomes, and thereby the parameter space, is exhaustively explored.

The techniques of Approximate Bayesian Computation (ABC), developed over the last two decades, offer historians a systematic methodology for gaining some control over uncertainty and randomness. These appear at every stage. First, even with the best possible estimates for parameters, it is unlikely that these will reproduce events. Indeed, with a stochastic model, even with the same parameters we will almost never get the same result twice. The same is true of real events: just because one side is superior to the other does not always mean the outcome will be the expected (“mean”) result. History is a particularly demanding context: we only ever have a statistical sample of one. Further, our parameters could be wrong for many reasons: our data could be incorrect, the model parameter might not be accessible from the data, or, as we shall see, the event from which we take our results might be an unusual outcome and lie far from the expected result.

Each run draws parameters from the prior distributions, and computes an outcome for the battle in the form of a set of summary statistics. These are then compared with the battle itself to produce a measure of how good the simulation is. In the case of a traditional Bayesian MCMC analysis the likelihood is an exactly-computed mathematical function which gives a precise measure of the “goodness” of the simulated outcome relative to the real outcome. In the more complex simulations presented here no likelihood function can be constructed, and for these we need the “approximate” methods of ABC. Here we use a criterion developed in biology for dealing with gross summary statistics of models based on individual interactions: a simple maximum distance between the real and simulated data is required for each statistic in order for the simulation to be considered “good”. On the basis of these good simulations the prior estimates are then updated to give “posteriors” – essentially, better estimates of the parameters in the light of what actually happened. As the parameters are improved over many runs, with the priors becoming more finely-tuned towards the posteriors, the distance requirement for a good simulation is successively reduced. We use six summary statistics in this study, three for each side (hits received, turrets lost, ships lost), these being evaluated (crucially) just prior to the disablement of *Lion* and *Bluecher*.

**Conclusions**

At their simplest, our results are represented by the histograms in Figure 1. Most striking is that our model captures the number of shell hits accurately, with true results close to the medians. But there is significant deviation in losses of the key Lanchestrian unit, the turret. Most striking of all is the histogram of ship losses. The Germans had not, at this stage, lost a ship, and this is the median result, although they later lost *Bluecher*. But for the British the median result is the loss of a ship in the modelled, early stage of the battle. The simulation is attempting to drag the values of the shell hits away from the median, so as to produce a larger number of units lost, but is unable to do so because of the (appropriate) prior values.

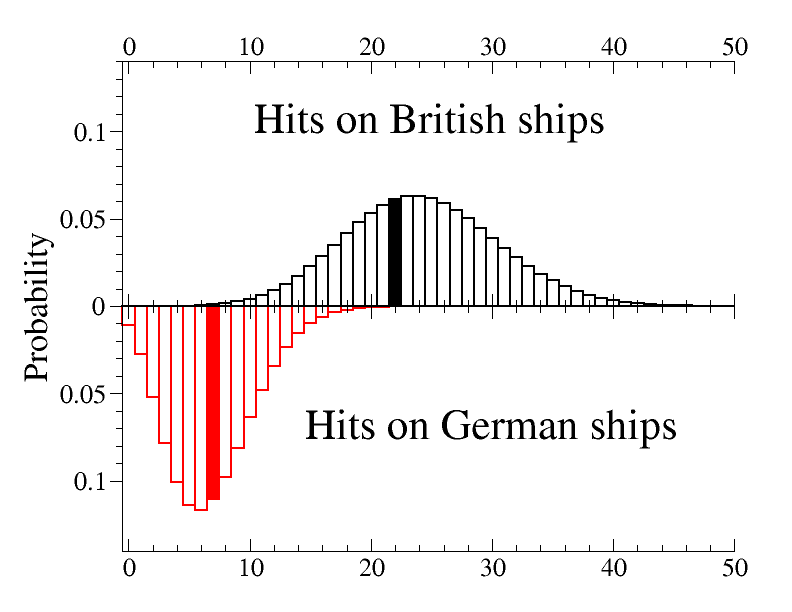
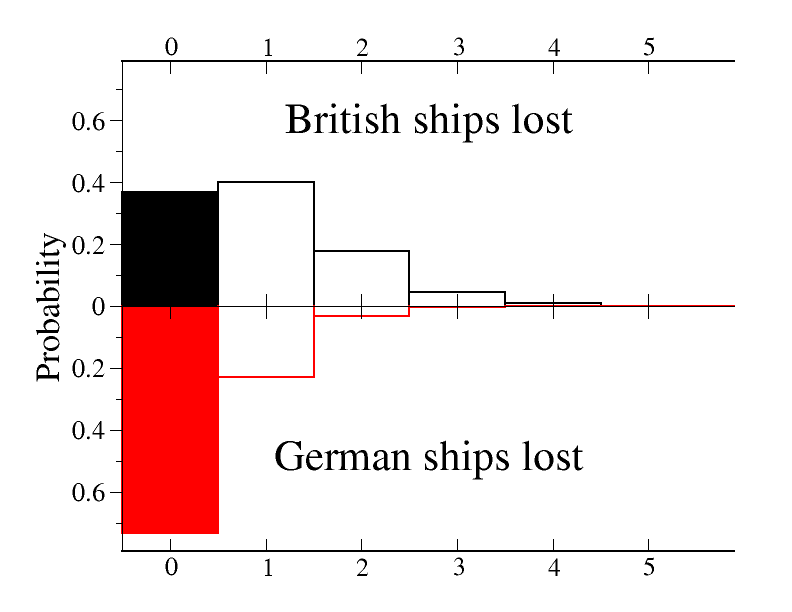
The essential point is that the priors are difficult to reconcile with the realized events, so that either the real outcome was highly improbable, or the prior estimates of the parameters were very wrong, or a mixture of the two. But the priors here are rather precise: for any parameters consistent with later events at Jutland, the Dogger Bank result was achieved with very low probability. The most probable unfolding of events would have diverged greatly from the actual trajectory of the battle, and the Bayesian analysis is a powerful quantitative statement of just how fortunate the British were at Dogger Bank. Even in the historical, abbreviated battle up to 11 a.m. the British should have lost at least one ship, while the expected result of a protracted battle might well have been the destruction of as many as three. The Germans were unlucky to lose *Bluecher*, but its relative vulnerability is still present in our data.

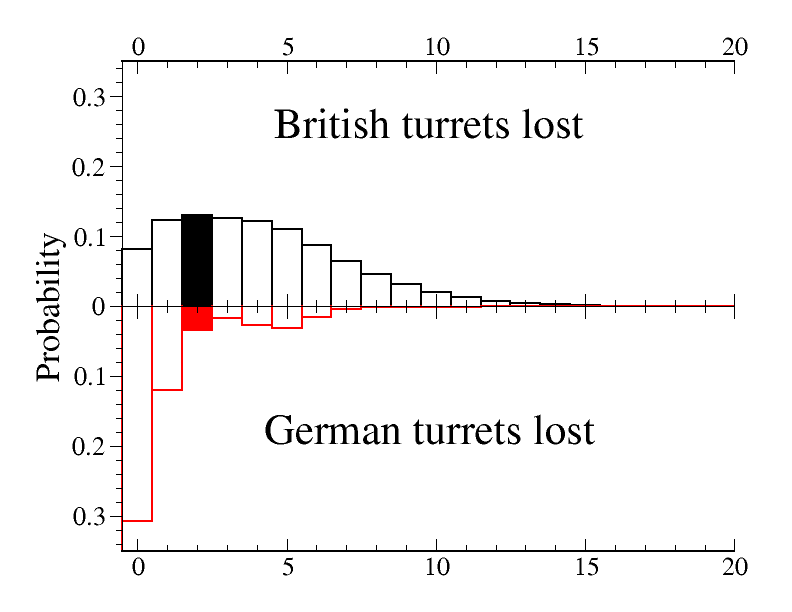
The British belief, later repeated by historians and analysts, was mostly that the battle of the Dogger Bank was a victory squandered. The British commander David Beatty took the view that “We had a great day. ... The [German 12"] projectile is no good, and I am sure we can stand a lot of it”. The Royal Navy’s “lessons learned” said that “German shell, for incendiary effect and damage to personnel, are far inferior to ours. Their only good quality lies in armour penetration and damage to material” – as if that was immaterial. The wider world took Dogger Bank as evidence of the superiority of the battlecruiser concept: articles in the US Naval Institute’s *Proceedings* opined that the Germans had narrowly avoided disaster and that “the battle-cruiser is the mistress of the sea.” Yet a young lieutenant, later to be an admiral, could write that “We were marvellously lucky to escape as we did as their shooting was damned good”, and the evidence supports him: our results show that the muddled British signalling that cut short the engagement between the battlecruisers almost certainly prevented British losses.

More broadly, we argue that Bayesian methods have much to offer to historical analysis. It is axiomatic that each historical event began as one of many possibilities and remained no more than a probability until it occurred. A much fuller historical understanding can be achieved when we have a tool for effective examination of the full range of prior possibilities.

Above all, improbable actual outcomes distort historical judgment. Historical actors’ behaviour which might seem inexplicable in the light of subsequent events may seem more reasonable when we know that the actual events were improbable. Similarly we may make poor judgments of actors’ subsequent behaviour if we falsely judge events to have been inevitable, or at least highly probable, when they were not.

Our view, demonstrated by the example of the battle of the Dogger Bank, is that the synergy between historical method and Bayesian simulation offers a step-change in the precision and rigour available to the historian, and a significant unvisited arena for scientific application. The power and utility of this approach will only increase as more large historical data sets become available – for example, Geographic Information System (GIS)-based recording of archaeological finds on battlefields. Even without simulation, the perspective of Bayesian ideas provides a yardstick by which historians can measure their views against new historical information, and a framework by which, as Niall Ferguson advocated, historians can “recapture the uncertainty of decision-makers in the past, to whom the future was merely a set of possibilities”.



**Figure 1**: Results. The three sets of summary statistics are presented here: the total number of hits received by each side, the total number of turrets lost by each side, and the total number of ships lost by each side. The British figures are on the upper panel and German on the lower. The actual result in each case is shown by a filled box. Note that *Bluecher* was not lost in the action modelled here – the simulation finishes prior to the disablement of *Bluecher* and its subsequent loss. In the turrets-lost column the small bulge in the German distribution at around 6 is caused by the relative vulnerability of *Bluecher* and its loss, which would result in the loss of all six of its turrets.

**BOX**: The battlecruiser concept

In the decade before the First World War, Britain and Germany engaged in one of the great arms races of history, building huge fleets of warships whose principal armament was 10-14 big guns of the same 11”-15” calibre. The first of these was HMS *Dreadnought*, and all subsequent ships were known as dreadnoughts.

Two types of dreadnought were built: the battleship, and the battlecruiser (at first known as the “armoured cruiser”). The essential calculation is that, within a given tonnage of ship, the designer has to combine guns, armour and engines. The battleship balanced these, so that the ship’s armour would provide a reasonable defence against the equivalent of its own shells.

Very much the brainchild of British Admiral Sir John (“Jacky”) Fisher, the battlecruiser was more unbalanced. It combined big engines (and hence high speed) with battleship-standard guns but thin armour. It could catch and destroy enemy merchant raiders, act as the eyes of the fleet, and outrun any ship it could not outfight. But if it slowed down and fought a battleship it would be vulnerable.

**BOX:** The battle of the Dogger Bank: the first clash of dreadnoughts

The German navy was smaller than the British, and could not fight it all at once. The German strategy was rather to draw out a portion of the British fleet, typically by bombarding British coastal towns, and thereby create a popular clamour for an active response. Then the German battlecruisers would destroy whichever British ships it met, and perhaps bring the overall fleets to something like equal numbers.

But the British were decrypting the German codes. The Germans did not realize this, and instead suspected that fishing boats were radioing information on sightings of German ships. It was with the intention of clearing the fishing ground of the Dogger Bank of trawlers and small warships that the German fleet sortied there on 23rd-24th January 1915. The British, forewarned by decrypts, were there to meet them with a larger force.

The British had five battlecruisers. The Germans had four, one of which, *Bluecher*, was hardly a battlecruiser at all. Slower and worse-armed than any of the others, it had been the Germans’ misconceived response to what they thought the British were building when the battlecruiser was introduced.

The fleets steamed south-east in lines. The faster British had closed to within range at about 9 a.m. At 0943 the German *Seydlitz* suffered a hit which caused a massive ammunition fire, destroying two turrets. Around 1030 the British flagship *Lion* was hit several times, and had to haul out of the line. Simultaneously *Bluecher* at the rear of the German line was badly hit and its speed greatly reduced.

The pivotal event, at 1102, was a signal from *Lion*, muddled in sending and misinterpreted when received, which caused the other British ships to abandon the chase for the German fleet and instead concentrate on *Bluecher*. Overwhelmed, it sank around 1 p.m.

**BOX**: Lanchester’s Laws

In its simplest, deterministic form, Lanchester’s model takes Red and Blue numbers R(t) and B(t) and assumes that these evolve according to

.

The trajectories are hyperbolae, so that is conserved. Initial imbalances are amplified, and with initial numbers and the eventual victor (Red, say) annihilates its enemy with units remaining, where is given by

.

This “square law” became “quite famous in the [British] Grand Fleet”, as its commander John Jellicoe wrote to Lanchester in July 1916. But no calculus is needed to reach its essential conclusions, which were convincingly demonstrated by Bradley Fiske in the USA: his 1905 “tables” – in modern terms, spreadsheets – made the effects of the square law quite clear.

**BOX:** Flash fire and ammunition handling

The rotating gun turrets on the decks of a dreadnought are simply the business end of a cylindrical structure which extends down to the magazines where shells and, at the deepest level, propellant are kept. These are then lifted to the turret in mechanical hoists. The propellant is cordite – low explosives extruded as a cord-like substance. Cordite burns, but only does so explosively when under pressure.

Gun crews on both sides knew the importance of loading and firing rapidly. The relative safety of cordite compared to the old gunpowder often led them to store it in inappropriate places between magazine and turret, and to circumvent the steel “scuttles”, metal traps for exchanging cordite charges which were always closed on one side or the other. The upshot was a severe but unappreciated risk of “flash fire”, spreading down from the turret and eventually causing a giant, explosive fire in the magazine.

Following Dogger Bank, tensions in British practice remained. The Admiralty issued a memorandum warning about the dangers of flash fire, but the battlecruiser squadrons continued to emphasise rapidity of shooting. For example, Ernle Chatfield, captain of *Lion*, wrote that “a mistake was made in firing too slowly during the earlier stages… …rapidity of fire is essential.” but that “Plunging fire is a great danger to ammunition anywhere between decks. ...lids of powder cases should not be removed faster than necessary." Records suggest that it is likely that British handling practices were much the same at Jutland as at Dogger Bank: an example of the military wisdom that it is often the loser who learns most from a battle about how to fight the next one.

1. MacKay, N., Price, C. and Wood, A. J. (2016) Weighing the fog of war: Illustrating the power of Bayesian methods for historical analysis through the Battle of the Dogger Bank, *Historical Methods,* **49**(2), 80-91. This article contains citations for all sources referred to above. [↑](#endnote-ref-1)