**Weighing the Fog of War: Illustrating the power of Bayesian methods for historical analysis through the battle of the Dogger Bank**

N. J. MacKay[[1]](#footnote-1), C. Price[[2]](#footnote-2), A. J. Wood1,[[3]](#footnote-3)

ABSTRACT. The application of scientific methods to historical situations is restricted by the existence of a single outcome with no possibility of repetition. However, new computational methods make quantitative historical analysis nevertheless possible. We apply methods of approximate Bayesian computation to simulate a naval engagement of the First World War, the Battle of the Dogger Bank. We demonstrate that the battle’s outcome was highly unlikely, with significant implications both for subsequent actions and for historical understanding. Dogger Bank exemplifies our view that Bayesian methods offer historians the tool they need to grapple with the evolving probabilities of historical events, giving a sound scientific basis for counterfactual history, and opening up a wealth of possibilities for analysis.

KEYWORDS: Approximate Bayesian Computation (ABC); battle of the Dogger Bank

Counterfactual history has provoked an unfortunate response in academia: ‘Grown-up historians don’t waste time on counterfactuals,’ according to renowned historian Michael Howard.[[4]](#endnote-1) In this perspective, historians should confine themselves to what actually happened. Yet if their fundamental task is to explain why the past developed as it did, they must contend with other possible outcomes: the course of history is complex, contingent, and develops within an envelope of probabilities surrounding the historical narrative. Hindsight promotes the *post hoc* fallacy, leading to a temptation to view events as inevitable, even when demonstrably unlikely, and to arrive at misleading rationalizations about historical actors. An alternative methodology is to better understand and quantify the probabilities of historical outcomes, a problem which might appear intractable to historians. But today science can help. New techniques of Bayesian simulation (Marjoram *et al.* 2003) enable us to determine whether a particular outcome was especially unlikely or improbable – whether someone got lucky, and someone else did not. This can then alter our judgments about what was done and believed both before and after the event. Further, we argue that the mind-set required by the historian in such cases is precisely that of Bayesian probability: to understand historical actors' implicit prior estimates of chances, and how these changed as events unfolded. Where the evolution of probabilities can be quantified with scientific methods, it is the responsibility of historians to include this in their thinking. Niall Ferguson, a supporter of counterfactual history, has nevertheless asked: ‘How exactly are we to distinguish probable unrealized alternatives from improbable ones?’ (Ferguson 2000). In some cases, at least, Bayesian methods provide the answer.

As an example of the power of this approach we study a case from the First World War. On 24th January 1915 the battle cruiser fleets of the British and German Navies fought an action in the south of the North Sea, known as the Battle of the Dogger Bank. This was the first symmetric clash between the all-big-gun capital ships of the Dreadnought era, an eagerly-awaited confrontation that was expected to test the soundness of new doctrines and technology developed in the uncertainty of peacetime (Bennett 1974; Marder 1965; Philbin 2014).

Two powerful fleets met each other early in the day in conditions of good visibility. For the Germans, Admiral Hipper commanded the modern battlecruisers SMS *Seydlitz*, *Moltke* and *Derfflinger,* as well as the armored cruiser *Blücher*, a fast pre-dreadnought, and light forces. For the British, Admiral Beatty commanded battlecruisers *Lion*, *Tiger*, *Princess Royal* (together the 1st Battlecruiser Squadron, hereafter 1BCS) *New Zealand* and *Indomitable* (2BCS), and light forces. The outcome of Dogger Bank was inconclusive. It led to the sinking of *Blücher*, and the heavy damaging of another German ship, *Seydlitz*, through an internal explosion. *Moltke* was completely unhit, and *Derfflinger* had only minor damage. On the British side, *Lion* was disabled, while *Tiger* was hit numerous times without significant damage. The battle was truncated by confusing signals on the British side from *Lion*, as a result of which Beatty’s signalling officer was subsequently blamed for “losing” the certainty of a decisive victory for the British admiral.Disablement of the flagship, leading to its loss of electrical power and visual signals, were also major factors, however.

Dogger Bank is an excellent case study for this type of analysis, for both practical and historical reasons. The battle itself was fought in a single act: there were no distinct phases, and the long range at which the engagement was fought (especially in its early stages) and the positions of the light forces (outside the battle lines) make the big-gun battle a reasonable first approximation, at least until *Blücher* and *Lion* are unable to remain in the lines. During this section of the battle the spatial arrangement of the main vessels is largely unchanged.

Historically, Dogger Bank is of crucial importance, although its significance is sometimes undervalued by historians and analysts alike. After the clash, the British did not know how little damage they had inflicted on the surviving major ships of their opponent. There were rumours that the Germans had been ‘bashed about’, that one ship had sunk on the way home and another had been heavily damaged.[[5]](#endnote-2) Similar rumours, but of British ships sunk, circulated in the German fleet (Lützow 1931, 44-45). The result was that Dogger Bank was perceived as a battle that had got away, one that the British would undoubtedly have won conclusively had it not been for an errant signal,[[6]](#endnote-3) and this view has permeated to the modern day (Hughes 1986a, p75). Secondly, Dogger Bank was very much a precursor to the crucial opening ‘Run to the South’ phase of the Battle of Jutland. Beatty himself perceived the outcome of the battle to be unsatisfactory, and was less inclined to blame poor signals than the lack of initiative he perceived in subordinates who obeyed them instead of discounting their patent absurdity and engaging the enemy anyway (Philbin 2014, 152). His extensive report to the Admiralty of 9 February 1915 recommended changes, but not in gunnery training or ammunition handling, which did not appear to be causes for concern in terms of the perceived results of the battle (Philbin 2014, 144-145). The poor gunnery of the newly formed Battlecrusier Fleet [BCF] was also a result of its basing at Rosyth, where opportunities for practice were limited, and attempts to remedy this were far from complete by the time of Jutland (Goldrick 2015, 294).

Many historical issues are raised by the battle. For example, why did the British overestimate the effectiveness of their fire so markedly? Why were concerns amongst the sailors (for example, that ‘we were marvellously lucky to escape as we did’[[7]](#endnote-4)) not investigated more rigorously? Even if they had been, could anything effective realistically have been done? Such debates, however, rely on a perception of the facts of the encounter, and it is difficult to be critical of British leadership until the extent of their perceived errors can be estimated. The likely outcome of the battle if it had developed according to British expectations has previously been a matter of conjecture, but now probabilities can be calculated, and must be if a solid foundation for historical analysis is to be provided.

In the Bayesian approach, one begins by acknowledging that one always has some prejudices, some ‘prior’ beliefs, in the form of estimates of parameters, based on the best available information. These are then challenged by new information and systematically updated to give better, ‘posterior’ beliefs in the light of the new data. To illustrate the potential of Bayesian methods in history we conduct an analysis of the battle of the Dogger Bank using ‘approximate Bayesian computation’ (ABC) which, technically, comprises Markov Chain Monte Carlo (MCMC) techniques, adapted to cope with the absence of a mathematical likelihood function (Marjoram *et al.* 2003).[[8]](#endnote-5) The analysis begins by parametrizing the ships and their guns, and with prior estimates of these parameters. The parameters are connected with the outcomes via a simulation model – effectively, a simple war-game. The outcome of the simulation is captured by a set of summary statistics, which can then be compared with the same statistics for the historical outcome of the battle. The prior values of the parameters are then updated, to give posteriors, by comparing the outcome of the simulation with the historical outcome. The power of the method comes from running the simulation – playing the war-game – millions of times, thus enabling exploration of all regions of the parameter space.

**Simulation method: a stochastic Lanchester model**

First we need a reliable model of the naval encounter itself to form the basis of our simulation. We use a simple model based on the work of F. W. Lanchester (1913-1916), the underlying ideas of which pervaded writing on big-gun naval tactics in the decade before the First World War and were arrived at independently by authors of various nationalities (Chase 1902; Fiske 1905; Baudry 1914; Osipov 1915). The simple, central assumption of such models is that each side causes damage at a rate in proportion to its numbers. This assumption, and the models it results in, are largely discredited as aggregate models of land and air warfare, although they may still have value in understanding the tactical implications of local, temporary conditions. However, in naval warfare this assumption certainly captures the essence of big-gun combat, and may persist for much longer and therefore be more appropriate than in any of the many other military arenas to which the models have been applied (Hughes 1986b). Lanchester models' influence, both on the tactics of the era and on the 1922 Washington Treaty (Kahn 2004), was considerable, superseded only after the advent of aircraft carriers by an understanding of the more pulsed nature of their firepower.

Lanchester’s ‘square law’ was assumed to operate in the 1914-18 naval war, particularly in the context of Dreadnought battle lines, and formed an integral component of concentration as Fuller understood it when providing his seminal codification of the Principles of War (Fuller, 1916). For this reason we believe that the decisive phase of the Dogger Bank clash is an ideal proving ground for this form of modelling. The Lanchester equations, simplistic in most military scenarios, nevertheless capture the essence of the dynamics here, at least for the crucial period. We use a stochastic[[9]](#endnote-6) version of Lanchester’s model (Morse and Kimball 1951; Dolansky 1964) so that different outcomes can follow from otherwise-identical simulations.

A degree of simplification might be assumed from our focus on Dreadnought exchanges, bearing in mind that both Dogger Bank and Jutland were intricate affairs involving the integrated movements of scores of smaller vessels, including various types of cruisers, destroyers, torpedo boats and submarines, and aircraft. The smaller units had the capacity to inflict fatal injury on big ships and at Dogger Bank neither Beatty nor Hipper relished the danger of underwater damage from the many torpedoes carried by the light forces of both sides (Philbin 2015, 130) or their ability to drop mines in the path of the battle line (though no German vessels were carrying mines at Dogger Bank).

Such factors affected the thinking of both admirals and Beatty twice manoeuvred to avoid this type of attack. Nevertheless, the role of the light forces at Dogger Bank, after making the initial contact, was largely to cancel each other out and prevent interference in the movements of each side’s capital ships. German light forces moved ahead of the retreating Hipper, shielded by his battle cruisers but ready to intervene, while British light forces consciously avoided coming between the two lines and thus blocking the British battle cruisers’ lines of sight to their opponents (Goldrick 2015, 262). They were also kept at a distance by *Blücher* (Philbin 2014, 125).

Thus, light forces were prominent at Dogger Bank, but not in the way they were at Jutland, there being no equivalent of Scheer’s massed torpedo assault on the Grand Fleet, for example. The decisive clash came down to an exchange between two lines of Dreadnought battle cruisers and the hapless *Blücher*. This being so, we initially apply our techniques to Dogger Bank, aspiring to tackle in subsequent work the greater complexity of Jutland. There, not only are light forces more prominent, but the big-gun exchanges are effectively a series of Lanchestrian sub-battles determined by geometry and fluid spatial interactions.

The first decision to be made is the choice of unit, the relevant discrete element of fighting power (and the unit of mass in Fuller’s principle of concentration). In naval conflict this has been a surprisingly contentious choice. There are at least three possibilities: the gun, the turret or the ship. Influenced partly by clear statements made in the early works of Fiske (1905) we regard the turret as the natural choice. We do however include the possibility of a single hit causing loss of a ship. This happened three times to Dreadnought battlecruisers and twice to armoured cruisers on the British side during the battle of Jutland. Also at Jutland, the German pre-dreadnought *Pommern* exploded after a single hit by a torpedo. However, three of these ships were certainly obsolete, and *Indefatigable* and *Invincible* were arguably obsolescent, the former’s loss provoking Chatfield’s famously cold observation that she was ‘not a serious tactical loss’.[[10]](#endnote-7) Each ship has a number of parameters associated with it, listed in Table 1. We group ships by class in order that we may make use of a hierarchical Bayesian approach, as described below.

We make no attempt to simulate the spatial aspects, which here are well approximated by a simple exchange between battle-lines. Instead we give each ship an engagement time at which it becomes active and available in combat. This simulation starts at the point where the first ship is able to target all of the opposing side’s ships. The target is chosen randomly at commencement of action and then may be changed at each step with a certain probability. This is chosen to reflect the vagaries of targeting and conditions – at Dogger Bank there were frequent target changes, whereas during the Jutland ‘Run to the South’ phase there were not. For simplicity, we do not take any account of the arcs of fire of the turrets on the ships. This creates a slight bias in the fighting strength of ships with the P,Q wing-turret configuration, but we do not believe this is significant, as 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 (Gillespie 1976, Gibson & Bruck 2000). Note that we treat each turret firing a single shot as an individual event. This is historically inaccurate in the sense that we know that warships at least attempted to fire in salvos, typically with each turret contributing a single shot to each broadside. However, this coordination was often lost in combat and there is little evidence that the interaction between parallel firing contributes significantly to the damage caused or the likelihood of hitting. Indeed it is surprising how few salvos fired by each side resulted in multiple hits per salvo. For this reason we think our modelling choice is justified: it makes computation simpler, and at worst will mean that we are not completely capturing the fine detail of the noise in the simulation.

[Table 1]

The ship to fire then has certain probability of hitting its target (Table 1); if a hit is achieved the probability of damage is then computed. This probability is computed by multiplying two factors, which quantify respectively the effectiveness of the shell of the attacker and that of the armour of the defender (1 - resilience) to arrive at a probability. Damage here means probability of affecting the Lanchestrian unit, the turret. Whilst it is well known that German shells penetrated the British ships more frequently, British shell hits at Jutland were still able to deplete German turret numbers by causing extensive damage to secondary systems. Whilst this was not immediately clear after the Dogger Bank action, the state of German ships after Jutland makes this historically obvious. The German Battle cruisers, whilst still all afloat in the immediate aftermath of the exchanges, had so few operational turrets between them as to make them irrelevant in the combat model (Campbell 1986). 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 – and, if not, then a single turret is lost from the target.

The information in our prior parameters is the average of Dogger Bank and Jutland: 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. It is therefore essential to understand whether such probabilities changed greatly between Dogger Bank and Jutland 16 months later, as they certainly did for the Germans, who learned lessons from the near-loss of *Seydlitz* at Dogger Bank. The British Grand Fleet came increasingly to emphasize rapidity of fire during the intervening period (for a wide-ranging analysis see Lambert (1998)), but the evidence on battlecruiser force flash discipline is equivocal, and it is not clear that the danger at Jutland was greater than that at Dogger Bank.[[11]](#endnote-8)

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, which constitutes one re-creation of the historical event. 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 inflicted and received as well as the average 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 to gain some degree of successful fit at an aggregate level.) The implementation is written in JAVA and we use the stochastic simulation method of Gibson and Bruck (2000).

**Fitting the model: approximate Bayesian computation**

Bayesian fitting is something like exhaustive, multiple wargaming. 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, and which places a Bayesian analysis in contrast to the more limited approach of Connors *et al.* (2014), which uses a stochastic model to analyse a land battle but, crucially, without Bayesian fitting.

The techniques of approximate Bayesian computation (ABC), only developed during the last two decades, give the second main advantage of our approach: 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 also true of real events: just because one side is superior to the other does not always mean the outcome will achieve the expected (‘mean’) result. History is a particularly demanding example: 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 mean 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 statistics are then compared with those of 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 the ‘approximate’ methods of ABC have only recently been developed to handle them (Marjoram *et al.*, 2003). 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 of the statistics in order for the simulation to be considered ‘good’ (Toni & Stumpf, 2010).

On the basis of the ‘good’ simulations the prior estimates of the parameters 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.

Of course it may be that the priors are difficult to reconcile with the realized events, and on the basis of such difficulties one may conclude that either the real outcome was highly unlikely, or the prior estimates of the parameters were very wrong, or a mixture of the two. This is precisely what happened here: for any priors consistent with later events during Jutland, the Dogger Bank result was achieved with very low probability.

Was the Dogger Bank outcome unlikely, or were the parameters so changed by the time of the battle of Jutland as to invalidate their use in simulating the earlier battle? The ships, men and tactics were largely the same, and, as discussed above, we find no evidence of radical changes in loading practice. The BCF, based on the Firth of Forth, had poor opportunities to improve its shooting, and probably did not do so except for the 3rd battlecruiser squadron, which was detached to Scapa for gunnery practice shortly before Jutland and achieved much-improved accuracy.[[12]](#endnote-9) Our estimates for the priors of the model, based on a combination of accurate estimates from Dogger Bank and on the exhaustive literature from Jutland, definitively catalogued by Campbell (1986), are summarized in Table 1. Campbell assembled his tables for capital-ship ammunition expenditure and hits from the after-action reports of both navies, and, although these are subject to some *caveats* (concerning for example expenditure of shell of different types and assignment of hits, slightly differing British and German recording methods, and the peculiar visibility conditions pertaining at Jutland) the data provided are richly substantiated and it is difficult to see how they can be improved upon. The most salient possibility of error in Campbell’s data is of imperfect assignment of hits (as between the ship which fires the shell and the target), but our model treats all damage-causing hits as equivalent (as suggested by Okun (2001)) and so is not affected by this.

An important technical requirement is that we choose priors with ‘support’ across the full range of possible values, to enable complete exploration of the parameter space – that is, we disallow impossible values but admit all others as possibilities, however improbable. It is precisely the quality of information made available by an exhaustive search of the parameter space that permits, and justifies, detailed assessments of probabilities. For this reason we choose what are known as ‘log-normal’ distributions for positive parameters with unbounded values (that is, their logarithms are Gaussian ‘bell’ curves), and ‘beta’ distributions for parameters which are used as probabilities. Means and standard deviations are given in Table 2. We also use a hierarchical scheme whereby ships of the same class share the same distributions. This gives us the ability to better assess these distributions whilst still taking account of ship-to-ship variability due to details of combat (such as better gunnery officers, vagaries of visibility, loss of rangefinders or other firing components). The result of the Bayesian analysis is a posterior distribution for each parameter, resulting from our prior information after it has been challenged by the simulation data.

[Table 2]

As noted above, our acceptance criterion for a ‘good’ simulation uses a simple maximum distance between the real and simulated data. We base this on the number of hits received and given by each side, as well as their effect; we use six summary statistics in this study, three for each side (hits received, turrets lost, ships lost). Because the space of possible results is so wide (in common with other without-likelihood studies implemented on IBMs (Toni & Stumpf, 2010) we need a strategy to refine our parameters. We define acceptance as being within some vicinity of the true outcome within the summary statistic space. We use a distance between the true results and the simulation, denoted by ‘ε’ in ABC parlance; initially this is that each statistic must each independently be within a range of the true value. True values were taken to be: British hits received - 22, German hits received - 7, British Turrets lost – 2 (2 on Lion), German turrets lost – 2 (2 Supra-firing turrets on Seydlitz), no British or German ships lost. We emphasise that this evaluation is taken at the point prior to the disablement of Lion and Blucher. To achieve this refinement we implement a series of MCMC chains[[13]](#endnote-10) with a large value of ‘ε’ to get a refined number of intermediate posterior distributions. It has been calculated that the chain should have an optimal acceptance ratio of 0.07 (Sherlock *et al.*, 2014); however, due to the nature of our data it was found to be impossible to achieve both this and chain convergence, so a value an order of a magnitude smaller was chosen. The chain was warmed-in with high values of ‘ε’ and the value of the Gaussian transfer kernel was reduced as necessary to achieve a chain at the desired value of ‘ε’ at the chosen acceptance rate. Once found, the chain was then run for 100,000 iterations and thinned by a factor of 100 to give 1000 samples which constitute one chain. Ten such chains were run in parallel to enable convergence statistics to be computed.

In line with standard Bayesian practice (Liu 2008) we test convergence of the chain by combining visual inspection with calculation of the R statistic of Gelman and Rubin (1992). An example of three chains can be seen in Figure 1. After some preliminary inspection, and owing to the tight prior information necessarily imposed, the best chain behaviour was accomplished in the manner described above.

[Figure 1]

The convergence behaviour for most parameters was excellent, with R values very close to one for all parameters associated with the main German capital ships and the British 1BCS. Parameters for the other ships were less well represented owing to the very weak selection acting upon them. It may be argued that *New Zealand* and *Indomitable* played no role at all in the key stage of the Dogger Bank battle, but as we wish to explore the space of more protracted involvement we have elected to keep them in and more selectively filter the parameter chains associated with them. On the German side, because *Blücher* is only a single ship the chain converges slowly and the burn-in is insufficient for the chain to converge for some values of the parameters. This may be seen both in the high R parameters and visually.

A final, crucial point is that we end our fitting at the time at which *Blücher* and *Lion* drop out of their respective lines. This is, effectively, the end of the realized spatially-simple big-gun battle between opposing lines, and any information from after this point is unrepresentative of such an encounter.

**Conclusions**

At their simplest, our results are represented by the histograms in Figure 2, where the real outcomes are demonstrably some distance from the expected values. In particular the number of turrets − the natural Lanchestrian unit (Fiske 1905) − that the British would have been expected to lose is apparently much higher than actually occurred at Dogger Bank.

The central issue is then: is this truly the case, or is it that the model or its prior parameters are at fault? The latter we may discount, for the priors are based on the sum of Dogger Bank and the much larger Jutland action, and the Bayesian method explores all of the parameter space. We have already argued that, in matters such as gunnery accuracy and flash discipline, the fleet at Jutland did not differ significantly from 16 months previously. For the Lanchestrian assumption we can do no more than note its broad contemporary acceptance[[14]](#endnote-11) and its later resilience, both after the war in the Washington Treaty negotiations (see Kahn, 2004) and in war-gaming, and in later analysis (Hughes 1986a, 1986b), as the correct model for battleship gunnery, superseded only after the advent of aircraft carriers by an understanding of the more pulsed nature of their firepower projection (van der Tol 1997). In sum, the model is probably not very far wrong, and it would have to be greatly so indeed to warrant a conclusion other than that the British were very lucky at Dogger Bank.

Most striking is that our model captures the number shell hits accurately – with both true results close to the median – but there is significant deviation in losses of the key Lanchestrian unit, the turret. This has as a consequence that the median result is the loss of a British ship in the modelled, early stage of the battle. The simulation is then 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 most probable unfolding of events would have led to significant divergence 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. The expected result would have been the destruction of at least one British ship, and possibly as many as three. The Germans were unlucky to lose *Blücher*, but its relative vulnerability is still present in our data.

[Figure 2]

Narrowly concerning the Battle of the Dogger Bank, the obvious first conclusion is that the British belief in a victory squandered[[15]](#endnote-12), and the later repetition of this belief by historians and analysts (Marder 1965, vol.2; Hughes 1986a), must be considered false. The evidence, rather, is that the muddled British signalling that cut short the engagement between the battlecruisers almost certainly prevented British losses. But although 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’ (Temple Patterson 1966-68, vol.1, item 107), the British had ample evidence, both from the hits on *Lion* and from the earlier battle of the Falklands, of the potential vulnerability of their battlecruisers.

It is difficult to see what could have been done tactically at Dogger Bank to mitigate the thin armour on the first two classes of British battlecruiser, and indeed the addition of the *Queen Elizabeth* class of battleships to the BCF should have reduced their vulnerability at Jutland by reducing the concentration of fire against them and overwhelming the German battlecruisers in the Lanchestrian sense. Nevertheless, weaknesses in the direction of fire apparent at Dogger Bank were repeated at Jutland, when again a German battlecruiser was left unmolested and Beatty’s failure to concentrate his line (with the *Queen Elizabeths* effectively left behind) and signal correctly were again in evidence. It has been argued in the context of learning the lessons of Dogger Bank that: ‘What was lacking, particularly in tactical development, were organisations that could conduct dispassionate analysis of the evidence from first principles, rather than as immediate responses to individual, but potentially linked, problems. These were not to emerge until much later.’ (Goldrick 2015, 294). The tendency to sanguine conclusions was not confined to the British: as a result of Dogger Bank, two articles immediately appeared in the US Navy’s *Proceedings* (Bullard 1915; Stirling 1915) which opined that the Germans had narrowly avoided disaster and that ‘the battle-cruiser is the mistress of the sea.’

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. When inevitability is discounted, as it must be, a much fuller historical understanding may 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 it is unfair to judge actors for their subsequent behaviour if we know that their information about events was incomplete. Finally, the attempt to quantify probabilities gives us greater understanding of analytically-inclined historical actors, such as the commander of the British Grand (combined) Fleet, Admiral Jellicoe, whose natural inclination was to make judgments of finely-balanced chances even when under great pressure. Bayesian analysis and results can provide the historian with a rigorous basis for understanding decisions made by such commanders.

In sum, historians need a balanced perspective based on the contingencies facing historical actors. War-gaming (the *Kriegsspiel* and its more dignified modern incarnation as ‘military simulation’) has, for over two centuries, been the means by which this is achieved in military education. 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 which war-gaming can make 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.[[16]](#endnote-13) 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 historians can ‘recapture the uncertainty of decision-makers in the past, to whom the future was merely a set of possibilities’ (Ferguson 2000).

**References**

Baudry, Lt A. (French Navy). 1914. *The Naval Battle: studies of tactical factors.* London: Hughes Rees.

Bayliss, A., C. Bronk Ramsey, J. van der Plicht and A.Whittle. 2007. Bradshaw and Bayes: towards a timetable for the Neolithic. *Cambridge Archaeological Journal* 17:1-28.

Bennett, G. 1974. *Naval Battles of the First World War.* London: Macmillan.

Bullard, R. S. 1915. A plea for the battlecruiser. *Proceedings of the United States Naval Institute* 41:1897-1917.

Campbell, J. 1986. *Jutland: an analysis of the fighting*. London: Conway Maritime Press.

Chase, J. V. (US Navy). 1902. *A mathematical investigation of the effect of superiority of force in combats upon the sea*, unpublished restricted paper. Newport RI: Naval War College Archives.

Connors, D., M. J. Armstrong, J. Bonnett, A counterfactual study of the Charge of the Light Brigade. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* (in press).

Cresswell, Capt. J. (RN). 1932. Royal Naval Staff College lecture, Churchill Archives Centre CRES 3/2.

Csillery, K., M. G. B. Blum, O. E. Gaggiotti and O. Francois. 2010. Approximate Bayesian Computation (ABC) in practice, *Trends in Ecology and Evolution* 25:410-418.

Dolansky, L. 1964. Present state of the Lanchester theory of combat, *Operations Research* 12:344-358.

Doob, J. L. 1945. Markoff chains – Denumerable case. *Transactions of the American Mathematical Society.* 58:455–473.

Ferguson, N. 2000. *Virtual History: Alternatives and Counterfactuals* New York: Basic.

Fiske, Cdr B. A. (US Navy). 1905. American naval policy. USNI Prize Essay, *Proceedings of the United States Naval Institute* 31:1-80.

Fuller, J. F. C. (anon.).1916 The Principles of War, with Reference to the Campaigns of 1914-15, *Journal of the Royal United Services Institute* 61: 1-40.

Gelman, A. and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences, *Statistical Science* 7:457–472.

Gibson, M. A. and J. Bruck. 2000. Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry A* 104 (9):1876–1889.

Gillespie, D. T. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* 22:403–434.

Goldrick, J. 2015. *Before Jutland: The Naval War in Northern European Waters, August 1914- February 1915*. Annapolis: Naval Insitute Press.

Hughes, W. P. 1986a. *Fleet Tactics.* Annapolis: US Naval Institute Press.

Hughes, W. P. 1986b. Naval tactics and their influence on strategy, *Naval War College Review* 39 no.1:2-17

Kahn, D. 2004. *The Reader of Gentlemen's Mail.* New Haven: Yale University Press.

Kennedy, M. C. and A. O'Hagan. 2001. Bayesian calibration of computer models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63:425–464.

Lambert, N. A. 1998. ‘Our Bloody Ships’ or ‘Our Bloody System’? Jutland and the Loss of the Battle Cruisers, 1916. *Journal of Military History* 62:29-55

Lanchester, F. W. 1913-1916. Articles in *Engineering* 1913-14, reprinted in *Aircraft in Warfare: the Dawn of the Fourth Arm,* London: Constable & Co, 1916.

Liu, J. S. 2008. *Monte Carlo Strategies in Scientific Computing*. Berlin: Springer.

Lützow, A. D. 1931. *Der nordseekrieg: Doggerbank-Skagerrak.* Oldenburg i. O.: Stalling.

Marder, A. J. 1965. *From the Dreadnought to Scapa Flow: The Royal Navy in the Fisher era, 1904-1919*, 5 vols. Oxford: University Press.

Marjoram, P., J. Molitor, V. Plagnol and S. Tavaré. 2003. Markov chain Monte Carlo without likelihoods, *Proceedings of the National Academy of Sciences* 100:15324–15328

Morse, P. M. and G. E. Kimball. 1951. *Methods of Operations Research*. Cambridge, Mass.: MIT Press.

Okun, N. 2001. Major historical naval armor penetration formulae, at <http://www.navweaps.com/index_nathan/Hstfrmla.htm>, last accessed 16th October 2014.

Osipov, M. 1915. The influence of the numerical strength of engaged forces on their casualties. Trans. from the Russian by R. Helmbold and A. S. Rahm, *Naval Research Logistics* 42: 435-490 (1995).

Philbin, T. R. 2014. *The Battle of Dogger Bank: The First Dreadnought Engagement, January 1915*. Bloomington & Indianapolis: Indiana University Press.

Ranft, B. ed. 1989. The Beatty Papers: selections from the private and official correspondence of Admiral of the Fleet Earl Beatty. Vol. 1. Navy Records Society, vol. 128.

Sherlock, C., A. Thiery, G. O. Roberts and J. S. Rosenthal. 2013. On the efficiency of pseudo-marginal random walk Metropolis algorithms; to appear in the *Annals of Statistics*. [arXiv:1309.7209]

Stirling, Cdr Y. (US Navy). 1915. The arrival of the battlecruiser, *Proceedings of the United States Naval Institute* 41:1919-1924.

Temple Patterson, A., Ed. 1966-1968. *The Jellicoe Papers: selections from the private and official correspondence of Admiral of the Fleet Earl Jellicoe of Scapa,* vols 108 and 111. Navy Records Society.

Toni, T., and M. P. H. Stumpf. 2010. Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* 26.1: 104-110.

Van der Tol, J. M. 1997. Military Innovation and Carrier Aviation: The Relevant History. *Joint Force Quarterly* Summer 1997:77-87

**Appendix: a stochastic Lanchester model**

The Lanchester aimed-fire model is simply the assumption that each side causes damage in proportion to its numbers and independent of its opponent’s numbers. This typically holds where targets are straightforwardly and symmetrically (as between the two sides) acquired. The general view among military analysts, based on the combat data, is that this assumption does not hold for aggregate land, air or all-arms warfare. However, as noted in the text and by analysts such as Hughes (1986b), it is peculiarly appropriate to *Dreadnought*-era battleship gunnery, albeit perhaps at the level of simple sub-battles rather than general actions among mixed forces. The stochastic version of the model simply assigns to each shot a probability of a hit, so that the expected total rate of hitting is proportional to the number of ships firing. This translates to a master equation for the time evolution of the current state of the battle, , where refers to the configuration of the system – the number of ships remaining and the number of their turrets. The master equation may be written in general form as

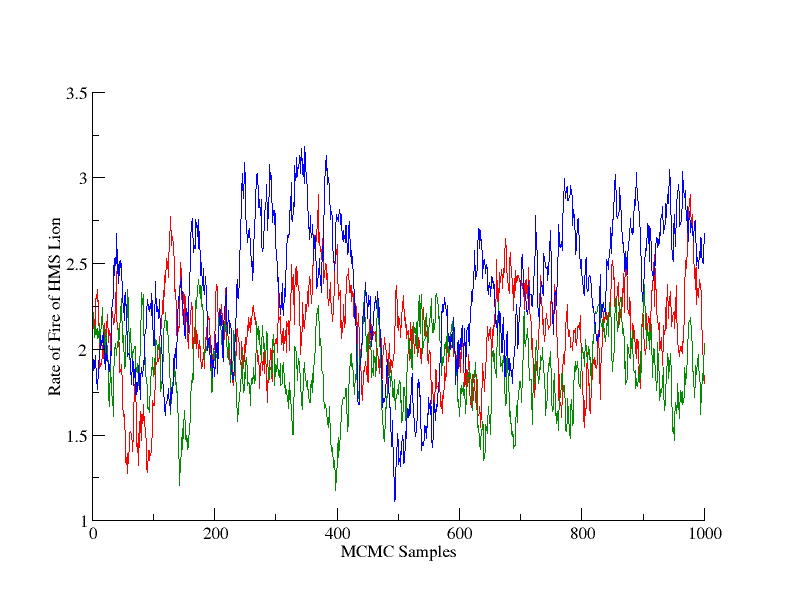
where the two sums represent respectively.all the ways in which the configuration can be achieved, and all the ways that the configuration can be exited. The transfer function, , consists of all the possible events in the system which change the configuration. For example a successful firing event of ship *i* which destroys a single turret of ship *j* contributes a term of the form

to the transfer function where *i* is the firing ship and *j* is the target. Equations of this type can be numerically solved with high accuracy by estabished standard stochastic simulation algorithms (SSAs) which are well-developed (Doob 1945, Gillespie 1976, Gibson & Bruck 2000), despite the apparent complexity of the master equation.

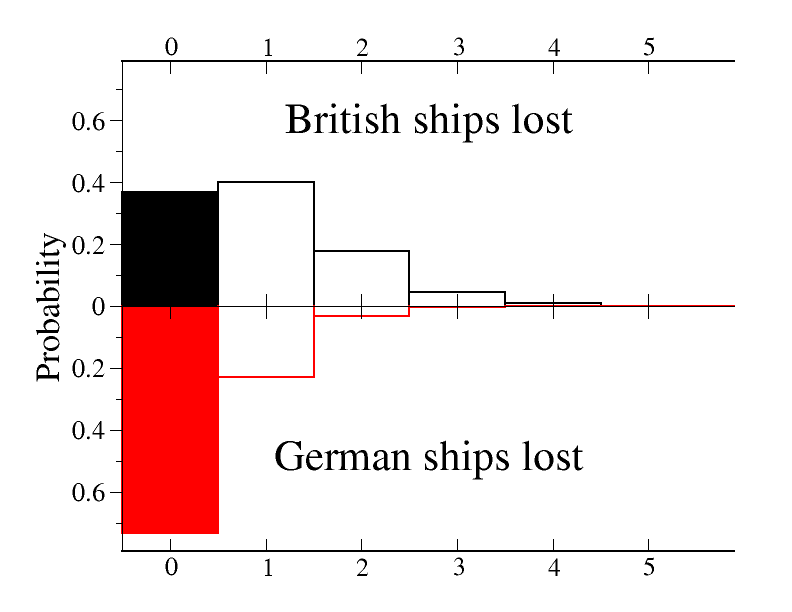
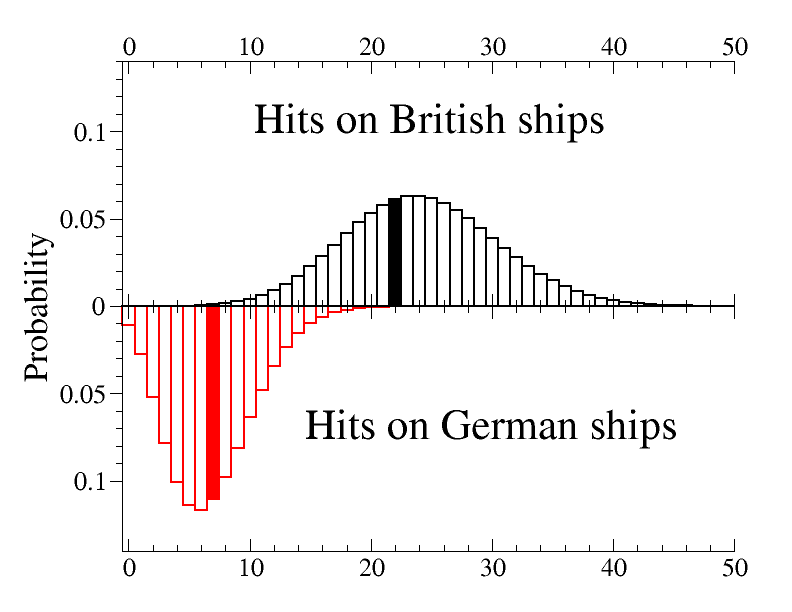
**Acknowledgment**

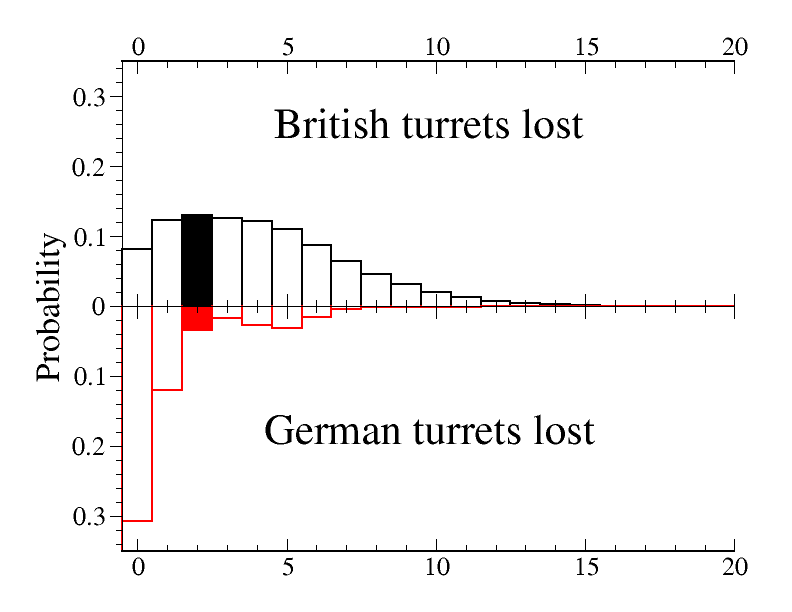
We should like to thank Wayne Hughes, James Goldrick and Matthew Seligmann for discussions.

1. Department of Mathematics, University of York, York YO10 5DD, UK. niall.mackay@york.ac.uk [↑](#footnote-ref-1)
2. Programme of History and American Studies, York St John University, York YO31 7EX. c.price@yorksj.ac.uk [↑](#footnote-ref-2)
3. Department of Biology, University of York, York YO10 5DD, UK. jamie.wood@york.ac.uk [↑](#footnote-ref-3)
4. Start the Week, BBC Radio 4, 11 June 2012 [↑](#endnote-ref-1)
5. Daniel, Lt C. (RN). Diary entry, Thursday 28th January 1915. Churchill Archives Centre, Churchill College, Cambridge, DANL. [↑](#endnote-ref-2)
6. Cresswell, Capt. J. (RN). Royal Naval Staff College lecture, 1932, Churchill Archives Centre CRES 3/2. See also Marder (1965). [↑](#endnote-ref-3)
7. Blagrove, Lt H E C. Letter to Lt Oswald Freeth, HMS Tiger, 6th March 1915, in Churchill Archives Centre, DRAX 1/47. [↑](#endnote-ref-4)
8. There is a fine technical distinction here: whereas ‘probability’ is used to refer to the chance of an outcome given the parameters, ‘likelihood’ refers to the chance of the parameters taking certain values given the outcome. Thus, for example, one might ask ‘what was the probability of *Hood* being sunk given that she was being fired upon by *Bismarck*?’ but ‘given that *Hood* was sunk, what is the likelihood that a shot from *Bismarck* penetrated a magazine?’. [↑](#endnote-ref-5)
9. That is, probabilistic, incorporating some randomness. [↑](#endnote-ref-6)
10. Quoted on p66 of Marder, *From the Dreadnought to Scapa Flow,* vol.3 of 2nd ed. (1978), p66. [↑](#endnote-ref-7)
11. Some directly-relevant comments are in the report of Capt. A. E. Chatfield (of *Lion*), item 122 of the Beatty Papers (Ranft 1989),, in which he says that ‘a mistake was made in firing too slowly during the earlier stages .... rapidity of fire is essential’. But this is balanced (in his ‘Recommendations’) by ‘Plunging fire is a great danger to ammunition anywhere between decks. ... Lids of powder cases should not be removed faster than necessary.’ The February 1915 revision of BCF orders (item 128) makes no mention of rapidity of fire, and there is nothing in its gunnery-training recommendations which might lead to worse flash discipline. An Admiralty memorandum of February 1915 urges better flash discipline, but was probably not widely acted upon (Lambert 1998). [↑](#endnote-ref-8)
12. The extent of this practice is made clear in *Indomitable*’s log, ADM 53/44832, National Archives, Kew. [↑](#endnote-ref-9)
13. This is a technical term for a means of moving around in – exploring – the parameter space. [↑](#endnote-ref-10)
14. For example, Jellicoe wrote to Lanchester on 15th June 1916 that ‘your N-square law has become famous in the Grand Fleet’. Letter held as B3/18, Lanchester archive, University of Coventry. [↑](#endnote-ref-11)
15. Cresswell, 1932 RNSC lecture. [↑](#endnote-ref-12)
16. Thanks to Hew Strachan for this suggestion.

    **Figure 1**: **Chain trace examples**. Shown here are three sample chain trajectories for a single parameter over multiple runs, in this case for the rate of fire of HMS *Lion* (in shells per minute). This parameter is drawn from a hierarchical distribution which also includes the rate of fire of HMS *Tiger*, both ships being of the same class.

    **Figure 2**: **Results.** The three sets of key results, or summary statistics, are presented here: the total number of hits by each side (A), the total number of turrets lost by each side (B), and the total number of ships lost by each side (C). The British figures (in a negative sense, *i.e.* hits received) are on the upper panel and German ones on the lower panel. The actual result in each case is shown by a filled box at the appropriate value. Note that *Blücher* was not lost in the action modelled here – the simulation finishes prior to the time of disablement of *Blücher* and prior to 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 *Blücher* and its loss resulting in the loss of all six of its turrets.

    **Table 1: Parameters of the model**

    |  |  |  |  |
    | --- | --- | --- | --- |
    | **Parameter** | **Type** | **British** | **German** |
    | Length of Combat | Global | At Dogger Bank the ships were engaged for approximately 2 hours | see left |
    | Target change | Global | 0.05 | see left |
    | Engagement time | Individual by Class | The British ships with larger guns engage first, so *Lion* engages as the simulation starts. The spread and poor order of ships is noticeable. | The Germans engage and begin firing approximately twenty minutes after the first British ship. |
    | Rate of fire | Individual by Class | Approximately 2 per minute | Approximately 3 per minute |
    | No of Turrets | Individual by Class | All have 4 | *Seydlitz*, *Moltke*:5*, Derfflinger*: 4 and *Blücher*: 6 |
    | Accuracy | Individual by Class | ≈0.03 | ≈0.05 |
    | Effectiveness | Individual by Class | 13.5” guns ≈0.36, 12” and 11” guns ≈0.3. See note in caption. | 11” and 12” guns ≈0.33, 8.2” guns ≈0.27. See note in caption |
    | Resilience | Individual by Class | 5.7” main belt armour corresponds to ≈0.57 for 1BCS, 3.8” corresponds to ≈0.38 for 2BCS. See note in caption. | 6.8” gives ≈0.68 for GBC and *Blücher* as a fraction of this gives ≈0.43. See note in caption. |
    | Flash | Individual by Class | With knowledge from Jutland this is a potential problem, this is present. | This is a potential hazard for the Germans at Dogger as many steps were taken between Jan 1915 and May 1916 to remedy after the contained flash explosion on *Seydlitz*. |
    | Disablement | Individual by Class | Small chance of disabling hit – becomes ‘sitting duck’. | as left |

    **Table 1**. Summary of the main parameters of the simulation, including indication of the values used. The actual values are drawn from a distribution around these values; see section on Bayesian treatment. Based on information from Campbell’s (1986) detailed description. The average probability of a hit causing loss of a turret at Jutland is approximately constant at 0.12, independently for each side. Our effectiveness and resilience must combine to give around this figure. We use main belt armour thickness divided by 10 as the base for resilience and assume British 12” guns have an effectiveness of 0.3. Other calibres are then computed relatively, based on Nathan Okun’s 2001 analysis of calibre, weight and muzzle velocity (British 13.5” 1.2, German Shells 1.1).

    **Table 2: Prior and posterior means and standard deviations**

    |  |  |  |  |  |  |
    | --- | --- | --- | --- | --- | --- |
    | **Parameter (probability, unless otherwise stated)** | **Prior mean** | **Prior stdev** | **Posterior mean** | **Posterior stdev** | **R value** |
    | Rate of Fire (British BC) (shells min-1) | 2.13 | 0.4 | 2.12 | 0.389 | 1.01 |
    | Accuracy (British BC) | 0.0104 | 0.00364 | 0.0102 | 0.0034 | 0.999 |
    | Effectiveness (13.5") | 0.363 | 0.0473 | 0.372 | 0.047 | 1 |
    | Resilience (Lion Class) | 0.569 | 0.0488 | 0.566 | 0.0481 | 1 |
    | Flash chance (Lion Class) | 0.0417 | 0.0399 | 0.0388 | 0.0358 | 1 |
    | Disablement chance (Lion Class) | 0.0833 | 0.0552 | 0.0875 | 0.056 | 1 |
    | Engagement time (Lion) (min) | 2 | 0.999 | 2.09 | 1.05 | 1.05 |
    | Rate of Fire (German Ships) (shells min-1) | 2.83 | 0.4 | 2.85 | 0.395 | 1.02 |
    | Accuracy (German Ships) | 0.0235 | 0.00484 | 0.0234 | 0.00429 | 0.999 |
    | Effectiveness (German BC) | 0.333 | 0.0464 | 0.332 | 0.0469 | 1 |
    | Resilience (German BC) | 0.676 | 0.046 | 0.674 | 0.0443 | 1 |
    | Flash chance (German ships) | 0.111 | 0.0993 | 0.0909 | 0.0747 | 1.01 |
    | Disablement chance (German BC) | 0.125 | 0.11 | 0.145 | 0.0944 | 1.03 |
    | Engagement time (Seydlitz) (min) | 21 | 0.999 | 21 | 0.997 | 1.02 |
    | Effectiveness (British 12") | 0.304 | 0.0453 | 0.305 | 0.0436 | 1 |
    | Resilience (NZ and Indomitable) | 0.382 | 0.0478 | 0.385 | 0.0481 | 1 |
    | Flash chance (NZ and Indomitable) | 0.25 | 0.193 | 0.207 | 0.149 | 1.05 |
    | Engagement time (NZ) | 54 | 4 | 54 | 3.95 | 1.13 |
    | Effectiveness (German AC) | 0.275 | 0.0439 | 0.268 | 0.042 | 1 |
    | Resilience (German AC) | 0.431 | 0.0488 | 0.419 | 0.0491 | 1 |
    | Disablement chance (German AC) | 0.333 | 0.178 | 0.352 | 0.0908 | 1.3 |
    | Engagement time (Tiger) (min) | 7 | 0.999 | 7.03 | 0.994 | 1.06 |
    | Engagement time (Princess Royal) (min) | 12 | 0.999 | 12 | 0.985 | 1.02 |
    | Engagement time (Indomitable) (min) | 110 | 4 | 112 | 5.26 | 1.39 |
    | Engagement time (Moltke) (min) | 20 | 0.999 | 20 | 0.98 | 1.02 |
    | Engagement time (Derfflinger) (min) | 19 | 0.999 | 19 | 0.989 | 1.01 |
    | Engagement time (Bluecher) (min) | 20 | 0.999 | 20.1 | 0.997 | 1.01 |

    **Table 2**. Prior and posterior means and standard deviations for the parameters used in the simulation, together with the Gelman-Rubin R statistic computed across 10 chains.

    [↑](#endnote-ref-13)