 REVIEW BY JAMES GUSZCZA AND THOMAS N. HERZOG

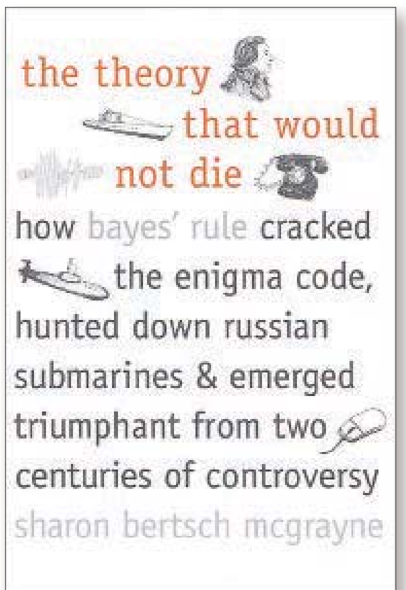
$p(theory) \xrightarrow[data]{} q(theory)$

The Theory That Would Not Die

by Sharon Bertsch McGrayne (Yale University Press, New Haven, 2011)

DURING A ROUTINE MIDAIR REFUELING in January 1966, a U.S. B-52 bomber armed with four hydrogen bombs brushed against the nozzle of a fuel tanker. Forty thousand gallons of fuel burst into flames, killing seven of the crew members and raining tons of aircraft parts—and four hydrogen bombs—on the isolated Spanish town of Palomares. None of the bombs exploded, and three of the four bombs were located within 24 hours. But the fourth bomb remained elusive. Strategic Air Command needed to use all available information to locate the wayward nuclear bomb as soon as possible.

Thus begins one of the many gripping, and occasionally startling, stories that grace Sharon Bertsch McGrayne's highly enjoyable new history of Bayesian inference, *The Theory That Would Not Die*. McGrayne's account is capacious enough to include vignettes as disparate as the search for lost nuclear weapons, the freeing of Alfred Dreyfus from Devil's Island, the successful effort to crack the Nazis' Enigma cipher machine, determining the authorship of the Federalist Papers, and the founding of the Casualty Actuarial Society (CAS).



Bayesian Logic

Bayes' "inverse probability" rule concerns the updating of probabilities about uncertain statements in light of empirical data. In modern terms, it takes the form of a simple conditional probability calculation:

$$p(theory) \xrightarrow[data]{} q(theory) = p(theory \mid data) = \frac{p(data \mid theory)p(theory)}{p(data)}.$$

Here $p(theory)$ and $q(theory)$ represent one's "prior" and "posterior" probabilities of the truth of an uncertain statement before and after the receipt of a body of relevant data. The second equality in the above expression is an elementary fact of probability theory known as Bayes' theorem. But the real substance of Bayesian inference is encapsulated in the first equality: $q(theory) = p(theory \mid data)$. This is a philosophical assumption, not a mathematical fact. And it encapsulates the central Bayesian principle that conditional probability is the mechanism by which new data should be combined with background beliefs or knowledge to make predictions and inferences. (See our article, "Enhanced Credibility," in the September-October 2010 issue of *Contingencies* for more details.)

This is the logic that the Navy's top scientist, John Craven, developed to guide the search for the lost nuclear bomb near Palomares and later used to find the nuclear submarine USS Scorpion, lost at sea in May 1968. The area of search is divided into a grid, each cell of which is assigned a probability of containing the lost object. The search is prioritized in the areas of highest probability, and Bayes' theorem is used to update these probabilities in the light of failing to find the object in the areas that have been searched. This procedure incorporates the probability of finding the object in a location X if it actually is in that location. Bayesian updating reflects that fact that failing to find the lost object in, say, a deep water area, conveys different information than failing to find it in an open field. McGrayne (in a private conversation with us) and others have speculated reasonably that Bayesian search probably was used in the detection of Osama bin Laden in Abbottabad, Pakistan.

Theoretical Origins

A striking aspect of the history of Bayesian inference is the large number of times it independently has been rediscovered. The rule is of course named after the Rev. Thomas Bayes, who first outlined the logic of inverse probability in a paper that was posthumously presented to the British Royal Society in 1763 by his friend Richard Price.

Bayes' work relates interestingly to both philosophy and actuarial science. McGrayne astutely links Bayes' work with the philosophical discussions of cause and effect that were "in the air" subsequent to the publication of David Hume's landmark *Treatise on Human Nature*. And indeed Price used Bayes' rule to dispute Hume's skeptical thesis (today echoed in Nassim Taleb's *Black Swan*) that no amount of experience is sufficient to establish a cause-and-effect relationship. Price, interestingly, also was a founding father of the actuarial profession—he consulted for the Equitable Life Assurance Society of

$$= p(theory \mid data) = \frac{p(data \mid theory)p(theory)}{p(data)}$$

London, and his nephew, William Morgan, was the first person whose professional designation contained the word "actuary."

Like the calculus, Bayes' rule was independently discovered at essentially the same historical moment in two countries. Pierre Simon Laplace, one of history's greatest scientists, independently discovered Bayes' rule, formulated it in more modern terms than did Bayes, and used it in applications ranging from estimating the mass of planets to assessing the probability of a defendant's guilt given the evidence.

While Laplace is remembered primarily for his contributions to the physical sciences and discovery of the central limit theorem, he was also an Enlightenment thinker. His probabilistic analysis of court trials was of a piece with his opposition to capital punishment. McGrayne quotes Laplace's comment, "The possibility of atoning for these errors is the strongest argument of philosophers who have wanted to abolish the death penalty."

Many decades later, the great mathematician Henri Poincaré used Bayesian inference to help free Alfred Dreyfus, a French officer falsely accused of spying for the Germans, from life imprisonment on Devil's Island. It is ironic that many contemporary lawyers view the Dreyfus Affair as an example of why probabilistic arguments should be curbed in criminal cases.

Expanding Development

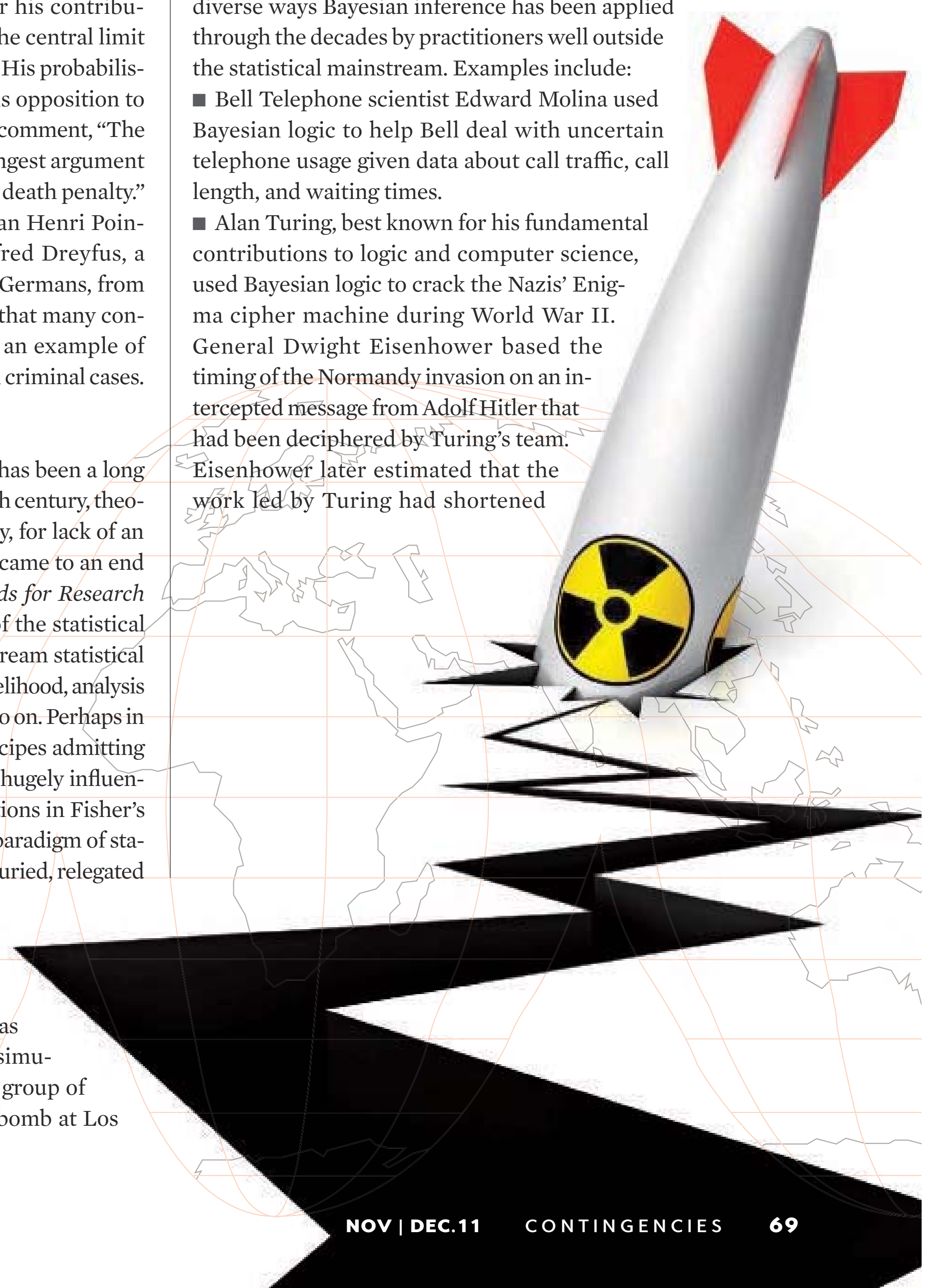
Widespread acceptance of Bayes' updating rule has been a long time coming. From Laplace's time to the early 20th century, theorists tended to accept Bayes' rule only grudgingly, for lack of an alternative. A protracted period of ambivalence came to an end when R. A. Fisher published *Statistical Methods for Research Workers* in 1925. Fisher's book outlined many of the statistical procedures that have become codified in mainstream statistical education and computing packages: maximum likelihood, analysis of variance, significance tests using *p*-values, and so on. Perhaps in no small part because of Fisher's emphasis on recipes admitting of automation, his approach to statistics became hugely influential. *Statistical Methods* went through seven editions in Fisher's lifetime, and his methods became the dominant paradigm of statistical inference. Bayesian inference was all but buried, relegated to a generally reviled minority status.

The year 1990, however, ushered in a remarkable reversal of fortune and marked the beginning of a Bayesian revolution that even today continues to gain force. The key insight was that Markov Chain Monte Carlo (MCMC)—a simulation technique that emanated from the same group of research scientists that worked on the atomic bomb at Los Alamos—could be used to bypass the intractable integrations needed to compute Bayesian posterior probabilities. By providing working statisticians with a powerful computational tool (if not a collection of Fisher-style recipes), MCMC has enabled Bayesian practice to flourish in a wide and expanding array of domains.

Hidden Histories

There is more to the story, of course, and McGrayne's book is particularly fascinating in the light it sheds on the diverse ways Bayesian inference has been applied through the decades by practitioners well outside the statistical mainstream. Examples include:

- Bell Telephone scientist Edward Molina used Bayesian logic to help Bell deal with uncertain telephone usage given data about call traffic, call length, and waiting times.
- Alan Turing, best known for his fundamental contributions to logic and computer science, used Bayesian logic to crack the Nazis' Enigma cipher machine during World War II. General Dwight Eisenhower based the timing of the Normandy invasion on an intercepted message from Adolf Hitler that had been deciphered by Turing's team. Eisenhower later estimated that the work led by Turing had shortened



PAVELIS / ISTOCK

the war in Europe by at least two years. Turing's work remained classified and little known outside the intelligence community for many decades.

- Jerome Cornfield, an employee of the National Institutes of Health and a Bayesian autodidact, pioneered the use of case-control studies as part of his effort to link smoking with cancer. Cornfield went on to become the most influential biostatistician of his time. It is interesting that R. A. Fisher, who consulted for the tobacco industry, vigorously disputed Cornfield's analysis.
- Albert Madansky of the RAND Corporation used Bayesian logic to estimate the probability of an accidental nuclear detonation—a probabilistic analysis of an unprecedented event. In a similar manner, the actuary L. H. Longley-Cook estimated the probability of two planes crashing in midair before such an event had ever happened. Longley-Cook's prescient Bayesian analysis anticipated two such crashes in the subsequent years.
- The Princeton/Bell Labs statistician and Exploratory Data Analysis pioneer John Tukey used Bayesian "shrinkage" to forecast presidential elections for NBC. Tukey's work was a forerunner of Nate Silver's hierarchical Bayes election forecasting at FiveThirtyEight.com.

Actuarial Influences

Actuaries play a particularly notable role in McGrayne's hidden history of 20th-century Bayes. Albert Wurts Whitney, one of the earliest members of the CAS, led a committee focused on providing a methodology for pricing workers' compensation insurance using the meager data available at the time. Though versed in statistics and the Bayes-Laplace theory, Whitney was also a pragmatist. He appreciated the need for a simple formula that working actuaries and underwriters could use to combine relevant pieces of information with subjective judgments, background knowledge, and collateral data sets. Whitney's famous credibility expression—$Z=P/(P+K)$—anticipated the shrinkage estimator that Charles Stein discovered in the 1950s.

Also remarkable in the development of Bayesian statistics is the role played by the midcentury actuary Arthur Bailey. Bailey, who merits an entire chapter of McGrayne's book, studied actuarial science at the University of Michigan and, as was typical, started his career as a frequentist in the Fisherian mold. At first dismayed by the apparent lack of rigor behind Whitney's credibility theory, Bailey spent much of the 1940s studying the issue. His library included a 1940 reprint of Bayes' paper with a preface by Bell Telephone's Edward Molina. In 1950, Bailey, by then a vice president at Kemper, presented the fruit of his investigations—a paper entitled "Credibility Procedures: Laplace's Generalization of Bayes' Rule and the Combination of Collateral Knowledge with Observed Data" at a CAS event. Bailey's paper anticipated the work of such core 20th-century Bayesians as Jimmy Savage and Bruno de Finetti. One particular passage in Bailey's paper vividly captures the actuarial profession's early embrace of the Bayesian philosophy:

> At present, practically all methods of statistical estimation appearing in textbooks… are based on an equivalent to the assumption that any and all collateral information or a priori knowledge is worthless. There have been rare instances of rebellion against this philosophy by practical statisticians who have insisted that they actually had a considerable store of knowledge apart from the specific observations being analyzed… However it appears to be only in the actuarial field that there has been an organized revolt against discarding all prior knowledge when an estimate is to be made using newly acquired data.

Bailey's paper was positively reviewed by the Harvard statistician Richard von Mises, who stated that he hoped it would help "the unjustified and unreasonable attacks on the Bayes theory, initiated by R.A. Fisher, fade out."

The paper attracted the right audience. Jimmy Savage, then a recent convert to the Bayesian paradigm, learned of Bailey's work from University of Michigan actuarial science professor Allen Mayerson. He subsequently reported Bailey's work to Bruno de Finetti, himself a former actuary. The two attended a conference together in Trieste, where they spread the word about Bailey and the Bayesian roots of actuarial credibility theory. Among the conference attendees were Dennis Lindley, one of the leading Bayesians of the second half of the 20th century, and Hans Bühlmann, who went on to develop Bailey's work into a full theory of Bayesian credibility.

As Bayesian methodology continues its march into the statistical mainstream, *The Theory that Would Not Die* serves as a helpful—and highly entertaining—lesson in the extraordinary, and often hidden, contributions that Bayesian methods historically have made in scholarship, government, military intelligence, jurisprudence, medicine, actuarial science, and beyond. □

JAMES GUSZCZA, a fellow of the Casualty Actuarial Society and member of the Academy, is an assistant professor of actuarial science at the University of Wisconsin: Madison business school.

THOMAS N. HERZOG, a fellow of the American Statistical Association, an associate of the Society of Actuaries, and a member of the Academy, was most recently distinguished scholar in insurance regulation at the National Association of Insurance Commissioners in Washington.