

A Mathematical Theory of Evidence turns 40

Glenn Shafer
gshafer@business.rutgers.edu
www.glennshafer.com

June 7, 2016

Abstract

The book that launched the Dempster-Shafer theory of belief functions appeared 40 years ago. This intellectual autobiography looks back on how I came to write the book and how its ideas played out in my later work.

Contents

1	What does probability mean?	2
2	Dempster's upper and lower probabilities	4
3	A dissertation about allocating belief	5
4	A book that succeeded in an unexpected way	7
5	Unfinished business	10
6	Constructive probability	12
7	Independence and conditional probability	15
8	Parametric statistics	18
9	Artificial intelligence	20
10	Causal conjecture	25
11	Game-theoretic probability	27
12	Dempster-Shafer redux	30
13	Conclusion	32

This year, 2016, is the 40th anniversary of the publication of *A Mathematical Theory of Evidence*. Thierry Denoeux, the editor of *International Journal of Approximate Reasoning*, has asked me to use the occasion to reflect on how I came to write the book and how I experienced its subsequent reception. I am honored and pleased by this invitation.

The occasion has reminded me how many individuals have worked on the Dempster-Shafer theory of belief functions and how little I know about many of their contributions. I offer these reflections as an intellectual autobiography, not as a comprehensive review of this vast body of work nor as an attempt to assess which contributions are most important and lasting.

As I have retraced my steps, I have also been reminded of how many friends I made and how many other lives touched mine, often only briefly but usually with good will and good cheer. Few of these friends and acquaintances are mentioned here, but memories of their generosity infuses these pages for me.

1 What does probability mean?

Early in my doctoral studies, I found I was interested most of all in the philosophical and historical aspects of probability and statistics. What does probability mean? How can statistical inferences be justified? What do the debates of centuries past teach us about these questions?

My interest in the history and philosophy of mathematics began in Marion Tyler's algebra class in my small home town, Caney, Kansas.¹ Mrs. Tyler had a small mathematics library in the corner of her classroom, and I was fascinated by James R. Newman's four-volume *The World of Mathematics* [73]. At the end of my sophomore year at Princeton, I hesitated between majoring in philosophy and majoring in mathematics, switching to mathematics at the last minute.

When I decided to attend graduate school, after a Peace Corps stint in Afghanistan, I was not thinking about probability and statistics. But when I showed up at Berkeley in July 1969, the faculty members who advised applicants to the Mathematics Department were on vacation. Lucien LeCam, the graduate advisor in the Statistics Department, was in his office with the door open, and he enthusiastically assured me, "We do mathematics here." Indeed they did, brilliantly. But it was mathematics with a philosophy about its applications that I found puzzling. In almost every meeting of his class, I would ask Peter Bickel, "How big does n need to be?" Peter could not answer the question, though he never lost his good-natured unflappability.

Someone suggested that statistics at Harvard would be more to my taste, and I spent a very enjoyable year there in 1970–1971. Then, leaving Harvard for personal reasons, I finished my doctoral work back at Princeton in 1971–1973.

¹I was born on 21 November 1946 at my grandmother's farmhouse west of Caney and spent my childhood at my parents' farmhouse east of Caney. Because many of its residents worked as professionals for the oil companies headquartered twenty miles south in Bartlesville, Oklahoma, Caney had a better school system than most Kansas towns of its size. Mrs. Tyler, who had graduated first in her class from Grinnell College, was married to an engineer who worked for Phillips Petroleum in Bartlesville.

This unusually peripatetic doctoral program worked out better than one might expect. I benefited especially from the exposure it gave me to the history of my subject. At Berkeley in the spring of 1970, I attended F. N. David's course on the history of probability.² At Harvard in the spring of 1971, I was fascinated by Art Dempster's course on statistical inference, with its heavily historical flavor. At Princeton in the fall of 1972, I was captivated by Ivo Schneider's seminar on the history of probability.³ Princeton also offered the wonderful historical resources of Fine Library and a stream of stimulating visitors. Those who became important contributors to the history of probability and statistics included Ian Hacking, Eugene Seneta, and Steve Stigler.

Today's student of the history of probability and statistics does not lack for secondary sources; there are dozens of books, hundreds of articles, even a journal (*Electronic Journal for History of Probability and Statistics*, www.jephs.net). The picture was quite different in 1970. *Biometrika* was publishing a series of historical articles, David had published her book on the early history of probability in 1962 [19], and of course we still had Isaac Todhunter's methodical treatise from 1865 [136], but that was about it. I am delighted by what my contemporaries have accomplished since then, but the paucity of historical work forty years ago may actually have been a blessing to me. It gave me license to believe that I could make a contribution to history, and that I might find in the work of historical figures insights that had been lost to my contemporaries.

I was hardly alone in the early 1970s in being dissatisfied with the objectivist Neyman-Pearson philosophy of probability and statistical inference that I had encountered at Berkeley. The pendulum was already swinging back to a subjectivist view of probability. Bruno de Finetti and Jimmie Savage, still active, were attracting more and more attention. De Finetti's *Teoria Delle Probabilità* [23] appeared in 1970 and would appear in English a few years later. Savage died in November 1971, but his celebrated article on elicitation appeared in the *Journal of the American Statistical Association* [80] the following month.

There were always many statisticians who sought a balance between objectivist and subjectivist methods. Maurice G. Kendall, one of the great systematizers and organizers of mathematical statistics, published an eloquent plea for reconciliation in 1949 [58]. The best known application of Bayesian methods in the 1960s was the study of the authorship of the Federalist papers by two very mainstream statisticians, Frederick Mosteller at Harvard and David Wallace at Chicago [72]. But the leading proponents of subjectivism in the 1970s, including de Finetti, Savage, I. J. Good, and Denis Lindley, wanted to push the pendulum

²This was when the Berkeley campus exploded over the bombing of Cambodia. Florence Nightingale David (1909–1993) had worked for decades at University College London, beginning as an assistant to Karl Pearson in 1933. There she had experienced R. A. Fisher's misogyny and developed a lasting friendship with Jerzy Neyman. Long encouraged by Neyman to move to the United States, she finally took a position at Riverside in 1967, usually spending three days a week at Berkeley ([66], p. 243). She looked out our classroom window askance at the demonstrators fleeing tear gas, anticipating political repercussions for higher education in California and the statistics department she was building at Riverside.

³Ivo Schneider (born 1938) visited Princeton's History Department in 1972–1973. He spent most of his later career as a distinguished historian of science at the University of Munich.

farther. They contended that mathematical probability, properly understood, is subjective always and subjective only. Rationality and coherence somehow demand, they contended, that a person have numerical beliefs that obey the probability calculus.

I found the subjectivist arguments about rationality and coherence as unpersuasive as the Neyman-Pearson methodology. I found older writers much more persuasive. For example, Augustin Cournot's *Exposition de la théorie des chances et des probabilités*, published in 1843 [16], seemed to me fresh and full of common sense. For Cournot, it was evident that probabilities were sometimes objective and sometimes subjective, and that some propositions might not have numerical probabilities of any sort.

2 Dempster's upper and lower probabilities

In the summer of 1971, after leaving Harvard, I began an intensive study of Art Dempster's articles on upper and lower probabilities. Art had discussed them briefly in his course on statistical inference. Others had spoken of them as important, philosophically deep, but hard to understand. What was he talking about?

In the mid-1950s, when Dempster was completing his own doctoral study at Princeton, R. A. Fisher's fiducial probabilities were being widely debated. According to Fisher, the probabilities involved in a statistical model can sometimes be used, after observations from the model are obtained, to give probabilities for a parameter in the model, without the use of prior probabilities as required by the inverse method introduced in the eighteenth century by Bayes and Laplace; the probabilities thus obtained for the parameter constituted its *fiducial distribution*. Fisher had first offered examples of his fiducial argument in 1930 [46], and he was still stubbornly defending it in *Statistical Methods and Scientific Inference* in 1956 [47]. Fisher's stature assured that he continued to have a hearing, but most statisticians were more persuaded by Jerzy Neyman's confidence intervals, which at least had a clear definition [68, 82]. Donald Fraser, a statistician who became a friend of Dempster's when he visited Princeton during the first year of Dempster's doctoral study, summed up the status of the fiducial argument with these words in 1961:

The method has been frequently criticized adversely in the literature, particularly in recent years. Some of the grounds for this criticism are: conflict with the confidence method in certain problems; non-uniqueness in certain problems; disagreement with 'repeated sampling' frequency interpretations; and a lack of a seemingly-proper relationship with *a priori* distributions.

John Tukey, a leading statistician at Princeton, devoted his 1958 Wald Lectures at the Massachusetts Institute of Technology to an attempt to systematize and interpret Fisher's various fiducial arguments. In his notes, eventually published in his collected works [137], he concluded, "I see no justified general conclusions."

Dempster, who joined the Harvard faculty in 1958 after a brief stint at Bell Labs, ventured into the fiducial controversy in the early 1960s, first giving additional examples of inconsistencies [25, 26] and then offering some examples he found more convincing and discussing how the probabilities involved might be better understood [27, 28]. Then he attacked the problem of discontinuous data. The fiducial argument required continuous data, but in 1957 Fisher had suggested that quasi-probabilistic conclusions, expressed by intervals rather than point probabilities, might be appropriate in discontinuous cases such as the binomial [48]. In the late 1960s, Dempster developed this idea in a series of articles on upper and lower probabilities [29, 30, 31, 33, 32, 34].⁴

Of the several models in Dempster’s articles of the 1960s, I was most intrigued by his “structures of the second kind” in [29]. These models used a uniform probability distribution on a $k - 1$ dimensional simplex to define random sampling from a multinomial distribution with k categories and then continued to believe this uniform distribution in the fiducial spirit after data were observed. In principle, this provided posterior upper and lower probabilities for the parameters of any parametric statistical model; you could condition the distribution on the simplex to define random sampling for a model that allowed fewer probability distributions on the k categories, and you could take limits to obtain models in which the data take infinitely many values, even values in a continuous range ([29], p. 373). My first contribution to these ideas came in the summer and fall of 1971, when I devised postulates that implied this simplicial model.

3 A dissertation about allocating belief

Perhaps I was always more interested in the meaning of probability than in statistical inference. In any case, by the fall of 1971 my focus had shifted from Dempster’s methods for statistical inference to the calculus of belief embedded in them. In addition to defining upper and lower probabilities for statistical hypotheses based on a single observation, Dempster had proposed a rule for combining the upper and lower probabilities obtained from independent observations. There was also a rule of conditioning that generalized Bayesian conditioning and could be understood as a special case of the rule of combination.

Before Cournot had distinguished between objective and subjective probabilities, Poisson had distinguished between chances (objective) and probabilities (subjective). I found the rules of the probability calculus (or Kolmogorov’s axioms, as we say more often nowadays) convincing for frequencies and objective chances but not convincing for degrees of belief or subjective probabilities. Why

⁴Dempster discussed Fisher’s and Tukey’s influence on his thinking in [39]. He does not mention Fisher’s 1957 article there or in his articles of the 1960s. Fisher’s lack of clarity in the 1957 article makes this unsurprising. Dempster has told me that he did try to read it several times and remembers hearing Fisher talk about it in a crowded auditorium at Columbia University around 1961. It may have triggered Dempster’s thinking, but it evidently provided little guidance for his thought.

not adopt Dempster’s calculus for degrees of belief based on non-statistical evidence?

Jacob Bernoulli famously declared that probability differs from certainty as a part differs from the whole.⁵ The British logician William Fishburn Donkin wrote that the theory of probabilities is concerned with how we should distribute our belief [44].⁶ This mode of expression probably sounded naive or merely metaphorical to most twentieth-century statisticians, but I took it seriously. Why not take the notion of distributing belief as basic and allow the distribution to be more or less specific? Confronted with a finite number of exhaustive and mutually exclusive hypotheses, we might distribute some portions of our belief to individual hypotheses while assigning other portions only to groups of the hypotheses.

Dempster had interpreted his upper and lower probabilities as “bounds on degrees of knowledge” ([31], p. 515). He had characterized his lower probability $P_*(A)$ as the “probability that necessarily relates to an outcome in A ” and his upper probability $P^*(A)$ as “the largest amount of probability that may be related to outcomes in A ” ([33], p. 957). My thought was to surrender the word *probability* to the objective concept and to build a new subjective theory using mainly the word *belief*. To give it the same mathematical autonomy as Kolmogorov’s axioms gave the usual theory of probability, we could adopt the inclusion-exclusion properties Dempster had noticed ([30], p. 333) as axioms.

Kolmogorov’s axioms are general and abstract. A probability measure can be defined on any space, finite or infinite. I aimed for the same generality and abstractness. In my doctoral dissertation [83], I renamed Dempster’s lower probabilities *degrees of belief*. I gave axioms for a function Bel on a Boolean algebra \mathcal{A} , which could be an algebra of propositions or a σ -algebra of subsets as in Kolmogorov’s theory. I called functions satisfying the axioms *belief functions*; probability measures were a special case. I showed that any belief function can be represented by a mapping from a probability algebra. I called the mapping an *allocation of probability*; each element of the probability algebra (thought of as a portion of belief) was mapped or *allocated* to the most specific element of \mathcal{A} to which it was committed. This was the structure that Dempster had represented as a multivalued mapping from a measure space [30] or as a random closed interval [33], now made general and abstract. It was also Donkin’s distribution of belief, except that belief committed to a proposition was not necessarily distributed among more specific propositions. Within this abstract picture, I showed how two belief functions on the same Boolean algebra could be combined and called the method *Dempster’s rule of combination*. Conditioning was a special case. The rule was appropriate when the bodies of evidence on which the belief functions were based were independent—i.e., their uncertainties were distinct.

⁵In the original Latin, p. 211 of [8]: “*Probabilitas* enim est gradus certitudinis, & ab hac differt ut pars à toto.” In Edith Sylla’s translation, p. 315 of [9]: “*Probability*, indeed, is degree of certainty, and differs from the latter as a part differs from the whole.”

⁶Augustus De Morgan also wrote about amounts of belief ([24], pp. 179–182). See Zabell [145].

While working in Princeton, I kept in close touch with Dempster, corresponding and occasionally visiting Cambridge. He was always very supportive, then and in the decades since. Geoff Watson, the chair of the Statistics Department at Princeton, who also became an enthusiastic supporter, chaired my dissertation. Dempster was on the committee, along with Gil Hunt, a well known probabilist in the Mathematics Department at Princeton. Probably the only person who actually read the dissertation at the time, Gil certified that my mathematics was sound. Since he shared the Princeton mathematicians' skepticism about the whole enterprise of mathematical statistics (Fisher's work had been nonsense, and the field had not seen any interesting ideas since Wald), his encouragement made me feel like a big deal. I was also encouraged by the philosophers I knew at Princeton, especially Dick Jeffrey, a personal friend. I submitted the dissertation in May 1973 and defended it in September.

The dissertation had been more or less finished by the summer of 1972. Watson wanted to keep me at Princeton, and I was more anxious to work on my ideas than to travel the country giving job talks. So I did not bother to apply elsewhere and agreed to become an assistant professor at Princeton starting in the fall of 1973. Applying also to Harvard and Stanford might have been a better career move.⁷

I spent 1972–1973 planning a larger treatise. I now think of this as the first of my several attempts to write the great American probability book. I explained my ambition in the preface to the dissertation. I would begin with “a historical and critical account of the Bayesian paradigm”. Then would come the material in the dissertation. Then an “exposition and justification of some of A. P. Dempster's methods of statistical inference”. This project soon fell by the wayside, but I did write an overlong article that summarized the dissertation and stated the postulates that implied Dempster's structures of the second kind. I presented it at a meeting of philosophers and statisticians organized by the philosophers at the University of Western Ontario in May 1973, and it appeared in the meeting's proceedings [85]. This article is not very accessible nowadays, but it brought my ideas to the attention of philosophers and philosophically minded statisticians, including Terry Fine, I. J. Good, Henry Kyburg, Denis Lindley, and Pat Suppes.

4 A book that succeeded in an unexpected way

When we go beyond statistical modeling and undertake to assess evidence that arises naturally in the world, we must settle on some level of detail. We cannot formalize every aspect of our experience. This is especially clear in the theory of belief functions, because in addition to the set of possibilities that defines the propositions for which we assess degrees of belief (the *frame of discernment*, I called it), we must have less formal bodies of experience (*bodies of evidence* I called them) that provide the basis for the assessments. But we must make our

⁷Affirmative action had not yet taken hold at Princeton. By the late 1970s no assistant professor could have been hired so casually from the inside.

formal list of possibilities detailed enough to discern what the evidence is telling us about the questions that interest us. If the evidence suggests an inside job, then Sherlock Holmes’s frame of discernment must distinguish between whether the burglar was an insider or an outsider. The importance of this point is clear in both the Bayesian and belief-function formalisms. When we implement a Bayesian conditioning design, in which we construct subjective probabilities over a frame and then condition on one or more subsets representing additional evidence, the frame must be detailed enough to describe the evidence in a way that captures its pointing to an inside job. When we implement a belief-function design in which we use Dempster’s rule to combine the evidence for an inside job with other evidence, say evidence the burglar was tall, the frame must be detailed enough to separate the uncertainties involved in the two bodies of evidence. If a stool might have been available to an insider, then the frame must consider whether or not this was the case.

My teaching duties as an assistant professor not being onerous (Watson saw to that), my research continued apace, and I was soon preoccupied with issues that arise when we look at an evidential problem in more and more detail, thus refining the frame. Perhaps if we carry on far enough, decomposing our evidence as we go, we will arrive at individual items of evidence all as simple as the testimony of one of the witnesses imagined by Bishop George Hooper in the seventeenth century, justifying belief p in a particular proposition and leaving the remaining $1 - p$ uncommitted.⁸ I called a belief function of this form a *simple support function*, and I called a belief function obtained by combining simple support functions a *separable support function*. When a simple support function gives degree of belief p to A , I called $-\log(1 - p)$ the *weight of evidence* for A . When we combine two simple support functions for A we multiply the uncommitted beliefs $(1 - p_1)$ and $(1 - p_2)$ and hence add the weights of evidence. In general, a separable support function is characterized by the total weights of evidence against the various propositions it discerns. How are these total weights of evidence affected by refinement? If a separable support function on one frame turns out to be the coarsening of a separable support function on a finer frame, is there some consistency? Will the total weight of evidence against A , say, be at least as great in the finer frame as in the coarser one? I conjectured that it will be but had no proof. The conjecture seemed very important, and I threw myself into writing a paper to explain it.

There was a lot to explain. Since the conjecture involved only finite frames, the relatively sophisticated mathematics in my dissertation was not needed. Since the conjecture had nothing to do with statistical inference, the statistical ideas in Dempster’s articles were not needed. But I still needed to explain

⁸Hooper’s rules for concurrent and successive testimony [55] are particularly simple examples of Dempster’s rule. The rule for concurrent testimony says that if two independent witnesses testify to the same event, and we give them probabilities p_1 and p_2 of being individually reliable, then the probability that neither is reliable is $(1 - p_1)(1 - p_2)$, and so we have a degree of belief $1 - (1 - p_1)(1 - p_2)$ that the event happened. His rule for successive testimony says that if we give witness A probability p_1 of being reliable, we give witness B probability p_2 of being reliable, and B says that A said that an event happened, then we have a degree of belief p_1p_2 that the event happened.

the picture that motivated my conjecture: the axioms for belief functions, the representation in terms of portions of belief, Dempster's rule of combination, the rule of conditioning, the relation to Bayesian theory, simple support functions and weights of evidence, and the notion of refinement. The article got longer and longer. Soon it was a book. To round it off, I added a simple method for statistical inference, translating likelihoods into what I called *consonant belief functions*⁹ rather than using Dempster's structures of the second kind. I entitled the book *A Mathematical Theory of Evidence* [84].

At Watson's suggestion, I talked with Bea Shube, the longtime editor of the Wiley Series in Probability and Statistics. The series had deep ties with Princeton, having been founded by Sam Wilks of Princeton and Walter Shewhart of Bell Labs.¹⁰ Shube was eager to publish my book, and Wiley would have marketed it well, but I felt the price Wiley would charge, around \$20, would be too high for many potential readers. I also believed that because my research for the book had been supported by a government grant, it should have a non-profit publisher. The Princeton University Press was willing to issue the book in both hardcover and paperback and to charge only about \$10 for the paperback. I went with Princeton and insisted on a contract with no royalties. Surely another bad career move by an insufferable young idealist. The masses did not buy the book even at \$10, and Princeton's marketing capacity was even weaker than I had expected. In the 1980s, unable to manage a book that had been issued in both hardcover and paperback, they routinely told bookstores that tried to order it for their customers that it was out of print, even though only the paperback stock was exhausted.

By the time the book came out, I had made another bad career move, abandoning Princeton's Department of Statistics and my east-coast mentors to move to the Department of Mathematics at the University of Kansas. "Mathematicians don't know how to treat statisticians," Watson warned me. "I can get along with anyone," I said. Kansas was home.

But *Blow a kiss, take a bow! Honey, everything's coming up roses!* Within a few years, *A Mathematical Theory of Evidence* was an immense success, giving me a scholarly reputation across disciplines hardly matched by any statistician of my generation and enabling me to keep following my own drummer indefinitely. Everyone interested in statistics and reasoning with uncertainty had heard about the *Dempster-Shafer theory*. At Kansas, the university promoted me against the

⁹In *Statistical Methods and Scientific Inference* [47], which I read and reread in those years, Fisher had contended that values of the likelihood are better fitted than Neyman's confidence intervals "to analyze, summarize, and communicate statistical evidence too weak to supply true probability statements" (p. 75). In Chapter 11 of *A Mathematical Theory of Evidence*, where I showed how to translate a likelihood function into a consonant belief function, I also acknowledged that a consonant belief function had the same mathematical structure that the economist G. L. S. Shackle had used to express "potential surprise." To my surprise, the philosopher Isaac Levi later took me to task for distorting Shackle's ideas [69]. For a recent review of the use of likelihoods to construct consonant belief functions, see [43].

¹⁰The series has long carried the subtitle "Established by Walter A. Shewhart and Samuel S. Wilks." Shewhart was already editing Wiley's series in statistics in the late 1940s; Wilks joined him in the 1950s.

recommendation of the faculty in Mathematics, and a few years later I moved to the Business School, where statisticians were full citizens and where I was soon appointed to an endowed chair.

What had I done right? I had written a book with ideas that many people, mostly outside probability and statistics, could build upon. I remember Richard Hamming telling a doctoral class at Princeton that in order to become widely cited, you should not write a paper that exhausts its topic; instead, you should leave things for others to do. I had done just that, inadvertently.

The conjecture that had inspired the book was little noticed, even when its main point was proven by Nevin Zhang [146]. What attracted attention were the introductory chapters of the book: the challenge to Bayesian orthodoxy, the straightforward vocabulary for talking about partial belief and ignorance, the convincing representation of ignorance by zero beliefs on both sides of a question, the straightforward method of discounting, and the simple examples of Dempster’s rule. For many people in engineering, artificial intelligence, and other applied areas, the mathematical level was just right—abstract but elementary for anyone with strong undergraduate mathematics—and there was obvious work to do in implementation and elaboration. I had not told all I knew about belief functions, and what I did tell was self-contained, leaving readers with a sense of freedom. My rules seemed newly made up; others could make up some more.

5 Unfinished business

By the time I had finished the book, I had a substantial backlog of related projects to complete. My list of unfinished papers was always at the top left of my blackboard, whether in my sunny second floor office in Fine Hall at Princeton or in the basement of Strong Hall, which housed the University of Kansas’s Mathematics Department at the time. Most of them eventually found their way into print.

The first to appear was my account of non-additive probabilities in the work of Bernoulli and Lambert, which had been gestating since I had participated in Schneider’s seminar as a graduate student.¹¹ It appeared in *Archive for History of Exact Sciences* in 1978 [86].¹²

¹¹I first encountered Bernoulli’s non-additive probabilities in Bing Sung’s translation of Part IV of Bernoulli’s *Ars Conjectandi*. Sung made the translation while working with Dempster at Harvard, but I was not aware of it until I attended Schneider’s seminar. I was so enthusiastic about the importance of understanding Bernoulli’s entire viewpoint that in the early 1980s I raised a modest amount of money from the National Endowment for the Humanities to translate all of *Ars Conjectandi* into English. I recruited Edith Sylla, an historian accomplished in late medieval Latin, to help, and in the end she completed the translation without my participation [9]. Her translation and commentary constitutes a very important contribution to the history of probability, but it has not, sadly, dented the established tradition of praising the theorem Bernoulli proved in Part IV (now often called the weak law of large numbers) while ignoring his broader concept of probability judgement and his insistence that probability judgements cannot always be additive.

¹²This article also discussed George Hooper’s 1699 article in the *Philosophical Transactions*

I also had things to say about some of my friends' ideas. Dick Jeffrey had introduced a generalization of Bayesian conditioning that he called *probability kinematics*; my article showing that his rule is a special case of Dempster's appeared in a philosophy journal, *Synthese*, in 1981 [89].¹³ The following year, the *Journal of the American Statistical Association* published my belief-function treatment of the possible discrepancy between Bayesian and frequentist assessments of near-matches in forensic testing, discussed earlier by Harold Jeffreys and my friend Denis Lindley [93].

The article I most enjoyed writing concerned the house where my family and I lived at the time, in the country north of Lawrence, Kansas. Like many houses, it had been built in stages. I wanted to remove the wall between the kitchen and an adjoining room. One of these two rooms must have been the original building from which the house had grown. Which one? I gathered some evidence: hints from the ceiling above, from the crawl space below, from talking with neighbors, and from my experience with how things were done in Kansas. Why not use my theory to combine this evidence? I carried out a belief-function analysis and a Bayesian analysis. This was a real example, with ordinary evidence, without a statistical model in sight. Ian Hacking had invited me to talk about my work at the 1978 Biennial Meeting of the Philosophy of Science Association in San Francisco. I told about my Kansas house there and published my analysis in the meeting's proceedings [90]. This is another relatively inaccessible publication, but I remain very fond of it, for it demonstrates how you can convincingly handle everyday evidence with belief functions.

And then there was my dissertation. The book that was supposed to incorporate it was not happening. I was in a Mathematics Department, and my dissertation was my most mathematical work. I should publish it in article form, as mathematicians do. There was a wrinkle. A couple years after defending the dissertation, I had discovered that André Revuz, a student of Gustave Choquet, had long before published the bit I had thought most important—the representation theorem for functions satisfying my axioms [78]. So, again as mathematicians do, I generalized the theorem and emphasized some auxiliary ideas. I submitted the resulting paper to the *Annals of Probability* at the end of 1976; it appeared in 1979 [87]. That was essentially the first half of my dissertation. I then duly drafted the second half, which treated Dempster's rule of combination in the same generality and submitted it to the same journal. No, the editor said. Enough of this. The mathematics is probably fine, but not enough of our readers would be interested.

of the Royal Society of London [55], but at that time I and my fellow historians of statistics did not know that Hooper was its author; see [95].

¹³Born in 1926, Jeffrey died of lung cancer in 2002. His great American probability book appeared posthumously in 2004 [57].

6 Constructive probability

After my first two years at Kansas, I was awarded a Guggenheim fellowship and used it to spend most of the 1978–1979 academic year at the Department of Statistics at Stanford, then as now arguably the leading department of statistics in the United States. This was a stimulating time for me. The chair of the department, Brad Efron, was a generous and congenial host. Aside from statisticians, Stanford’s own and a flock of visitors, I met many philosophers and psychologists who were interested in probability judgement and decision theory. Pat Suppes hosted a lively seminar where I met Lotfi Zadeh. The distinguished mathematical psychologist David Krantz, known for his work on measurement and color, showed up at my office one day to tell me how important he thought my book was; he had learned about it from Amos Tversky. The Belgian scholar Philippe Smets also came by to introduce himself. After having earned a medical degree in Brussels in 1963 and a master’s degree in statistics at North Carolina State, he had just completed a doctoral dissertation applying belief functions to medical diagnosis. All these individuals were a decade or a generation older than me, but I also met Akimichi Takemura, a young student visiting from Japan, who would become an important collaborator decades later.

The new friend who influenced me most was Tversky, a gentle and brilliant man loved by all who knew him.¹⁴ Krantz, one of Tversky’s collaborators on the theory of measurement, introduced me to him. Having just moved to Stanford from Hebrew University, Tversky was already famous for his work with Daniel Kahneman on how untutored reasoning violates the standard rules of probability, and he was looking for nonstandard rules that might fit practice better. Did I think belief functions described people’s behavior better than Bayesian probabilities? Were my rules meant to be descriptive or normative? Neither, I responded. My theory could not account for the “biases” that Tversky and Kahneman had identified, but neither it nor the Bayesian theory were norms in my eyes. Each was a calculus for assessing and combining evidence that might or might not produce persuasive and useful results in any particular problem. The proof was always in the pudding.

Krantz and Tversky thought my answer needed more depth. They were experts on measurement theory. Along with Suppes and R. Duncan Luce, they had published the first volume of a treatise on the topic, *Foundations of Measurement*, in 1971 [63], and they were still working on Volume II.¹⁵ Measurement theory was concerned with creating and using scales against which psychological magnitudes could be measured, and Krantz, Tversky, Suppes, and Luce had made it very mathematical, adducing delicate axioms and complicated consequences. If my theory measured the strength of evidence, then I must be creating and using some scale that could be analyzed in their terms.

The notion that subjective probability judgement as practiced by Bayesians involves comparing one’s evidence or belief to a scale of canonical examples is

¹⁴Born in 1937, Tversky died of skin cancer in 1996. He is still missed, personally and professionally, by a legion of admirers.

¹⁵Volumes II and III finally appeared in 1989 and 1990.

very old; we can find it in the work of Joseph Bertrand in 1889 ([11], p. 27) and Emile Borel in 1924 ([13], pp. 332–333). The examples that form the scale are provided by chance devices: urns with known composition, coins with known biases, etc. What scale did belief functions use? Hooper’s sometimes reliable witnesses come immediately to mind in the case of simple support functions. We can compare the strength of evidence pointing to A by comparing it to a scale of such witnesses, who are reliable with different probabilities p . We can generalize this idea to an arbitrary belief function by positing a scale of examples involving random codes. In each example a code is drawn at random, with code c_i drawn with probability p_i for $i = 1, \dots, n$, where $p_1 + \dots + p_n = 1$. We do not see which code is drawn, but we see a coded message, which means A_i if c_i was used. In the fiducial spirit, we continue to believe the probabilities and thus commit p_i of our belief to A_i and nothing more specific, for $i = 1, \dots, n$. This defines our belief function.

Back in Kansas, I explained this idea in an lengthy article entitled “Constructive probability”, published in *Synthese* in 1981 [88]. This article also responded to several reviews and commentaries on *A Mathematical Theory of Evidence*, especially Peter Williams’s review in *The British Journal for the Philosophy of Science* [142]. Williams and others often wanted to interpret belief functions as one-sided betting rates, more conservative than the two-sided betting rates of Bayesian theory. This is more or less equivalent to thinking of belief-function degrees of belief as lower bounds on a set of possible probabilities. One can indeed build a theory of probability judgement from this idea; in this theory each example on the scale to which you fit your evidence involves a set of probability measures, lower bounds over which you can call lower probabilities and upper bounds over which you can call upper probabilities. Such a theory was developed by Williams himself and later much further developed by Peter Walley, who called it a theory of *imprecise probabilities* [141]. But it is a different theory, with a different scale of canonical examples, from belief functions.

Tversky felt that we could develop these ideas further, and in 1985 he and I published together an article in which we argued that the Bayesian and belief-function formalisms should be seen as different languages of probability judgement and pointed out that these languages were not complete guides to how a problem should be analyzed. For any particular problem, we also need a design for collecting, decomposing, assessing, and combining judgements [120].

Dempster readily accepted the notion of a variety of theories of probability judgement, each with its own scale of canonical examples,¹⁶ and the notion also found resonance among other authors interested in belief functions. It made little headway, however, among proponents of Bayesian subjective probability. I recall discussing their normative claims with Tversky and Howard Raiffa at a conference at the Harvard Business School in 1983.¹⁷ Raiffa, along with Ralph Keeney, had long emphasized the constructive nature of successful Bayesian decision analysis, and Tversky thought that the word *prescriptive*, which Raiffa

¹⁶See, for example, [39].

¹⁷*Conference on Decision Making: Descriptive, Normative, and Prescriptive Interactions, 16–18 June 1983*. This conference led to an edited volume of papers on its theme, [6].

sometimes preferred, might address my unhappiness with *normative*. As Raiffa and Tversky and David Bell explained ([7], pp. 17–18), we might distinguish three kinds of models:

- Descriptive models are evaluated by their *empirical validity*, that is, the extent to which they correspond to observed choices.
- Normative models are evaluated by their *theoretical adequacy*, that is the degree to which they provide acceptable idealizations of rational choice.
- Prescriptive models are evaluated by their *pragmatic value*, that is, by their ability to help people make better decisions.

Did this open a space for alternatives to Bayes? Would it be helpful to call belief-function assessments of evidence *prescriptive*? To me, it seemed more helpful to say merely that they are legitimate arguments, which may or may not be persuasive. Perhaps the belief-function calculus is like another tool in the toolbox or another medicine in the pharmacy. A decision analyst might prescribe it for a particular problem, but it might still fail.

For Raiffa also, the word *prescriptive* did not really open the way to alternatives to Bayes. For him, being in a prescriptive mode merely meant being realistic about how close to the normative Bayesian ideal you can bring your decision maker. The Bayesian ideal was still uniquely normative. As Raiffa’s collaborator John Pratt wrote in 1986, in response to my explanation of why I did not find Savage’s Bayesian axioms convincing,

If your procedures or decisions or feelings are intransitive or otherwise discordant with subjective expected utility, they are incoherent, “irrational,” or whatever you want to call it, and trying to justify them as coherent or find other rationalities is a waste of time. ([97], p. 498)

In our personal conversations, Tversky agreed with me that such rhetoric was over the top but confessed that when he talked with Raiffa, he also agreed with him.

Pratt’s comments struck the earnest young Glenn Shafer as egregiously unfair, as did many other critiques of *A Mathematical Theory of Evidence*. Always the wise counselor, Tversky advised me to take them in stride. In challenging Bayes, I had criticized an established tradition of thought, with passionate followers, and I had to expect to be criticized in return. Would that I had heeded this advice more often. Instead I sometimes made my own tone more strident, especially in book reviews. Perhaps this is why E. T. Jaynes labeled me a “fanatical anti-Bayesian” in his posthumous great American probability book ([56], p. 717).

My work with Tversky had still not dealt with the question Dave Krantz had raised. Can we put the measurement of belief in the framework of the theory of measurement, and will this justify rules for belief functions? David and I discussed this for more than a decade, and he eventually obtained some interesting

results with his students; see for example [77]. We talked about collaborating on the theme, but I never really contributed. As Krantz observed at the time, I had become more interested in relating belief functions to traditional concepts of probability. In the last analysis, throughout my career as at the beginning, I was always most interested in the meaning of probability.

The normative claims of my Bayesian critics were tied up, of course, with choice and decision making. *A Mathematical Theory of Evidence* had been concerned only with probability judgement. But if we do construct degrees of belief using the book’s theory, can we use them for decision making? I addressed this question in a working paper entitled “Constructive decision theory,” which I completed in December 1982. In this paper I argued that value, like belief, usually has limited specificity and cannot be distributed over ever more finely described states or consequences. The goals we construct and adopt do not have infinite detail. This limited specificity of goals matches the limited specificity of Dempster-Shafer degrees of belief and allows us to score proposed actions without worrying, as we used to say, about the price of beans in China. This idea addresses Jimmie Savage’s *problem of small worlds*, which I discussed in an appendix to the paper. I submitted the paper to *Statistical Science*, and the appendix evolved into a discussion article in that journal [97], but the paper itself, with its proposal for Dempster-Shafer decision making, was never published.¹⁸

Although I had an opportunity to teach decision analysis after moving to the business school at Kansas, and enjoyed doing so, I never returned to the topic of Dempster-Shafer decision making. Other authors took it up, including Ron Yager [143] and Philippe Smets [127]. Jean-Yves Jaffray’s approach, developed starting in the late 1980s, was especially sophisticated [140].¹⁹

7 Independence and conditional probability

Dempster’s rule of combination, I explained in *A Mathematical Theory of Evidence*, requires that the belief functions to be combined be “based on entirely distinct bodies of evidence and that the frame of discernment discerns the relevant interaction of those bodies of evidence” (p. 57). I elaborated on this in my paper for the 1978 meeting of the Philosophy of Science Association:

Probability judgments are, of course, always difficult to make. They are, in the last analysis, a product of intuition. But the job is often easier if we can focus on a well-defined and relatively simple item of evidence, such as my neighbor’s testimony, and make our judgments

¹⁸It appears finally in this issue of *International Journal of Approximate Reasoning*. A similar proposal about combining goals and values was developed very persuasively, without reference to Dempster-Shafer theory, by Krantz and Howard Kunreuther in 2007 [62].

¹⁹This article, by Peter P. Wakker, appeared in a special issue of *Theory and Decision* dedicated to Jaffray’s memory. Student of and successor to Jean Ville at the University of Paris, Jaffray was a very generous individual. I was his guest for a sabbatical year in Paris in 1996–1997, and in later years I spoke with him many times as I collected material for a biography of Ville [113]. Born in 1939, he died of pancreatic cancer in 2009.

on the basis of that evidence alone. The theory of belief functions suggests that we try to break our evidence down into such relatively simple components, that we make probability judgments separately on the basis of each of these components, and that we then think about how to combine these judgments. This makes sense intuitively if the different components of evidence seem to involve (depend for their assessment on) intuitively independent small worlds of experience. And in this case the theory offers a formal rule for combining the probability judgments—Dempster’s rule of combination. ([90], p. 444)

Many readers found these explanations insufficient. I was proposing to replace the Bayesian rule of conditioning, which is built into probability theory, with a complicated rule that lacked clear justification and guidance about its application. The standard probability calculus used by Bayesians had a definition of independence. I had only intuition. As Teddy Seidenfeld put it ([81], p. 486), “At least the Bayesian theory provides the machinery for deciding whether the data are mutually independent.”

In my view, these criticisms were based on a misunderstanding of Bayesian probability arguments. Bayesian conditioning requires the same kind of intuitive judgement of independence as Dempster’s rule requires. In both cases, the judgement of independence is constitutive of the probability argument. We make a judgement whether to use the machinery. The machinery does not tell us whether to make the judgement.

I first arrived at this view of the matter in the fall of 1972, when I discussed Bayes’s posthumous essay on probability [5] in Schneider’s seminar. Most discussion of Bayes’s essay focuses on how he represented ignorance about an unknown probability. I was interested instead in his argument for updating: if at first we have probabilities $P(A)$, $P(B)$, and $P(A\&B)$, and we subsequently learn that B has happened, then we should change our probability for A from $P(A)$ to $P(A\&B)/P(B)$. Nowadays we lead students to think that this is true almost by definition; we call $P(A\&B)/P(B)$ the “conditional probability of A given B ” and write $P(A|B)$ for it. But our notion of conditional probability was invented only in the late 19th and early 20th centuries. Bayes needed an argument for changing from $P(A)$ to $P(A\&B)/P(B)$, and the argument he gave hardly has the force of a mathematical theorem. It comes down to a fiducial-type judgement of independence: we judge that learning B was independent of the uncertainties involved in a bet on A that was to be called off if B did not happen, and so we continue to believe these latter probabilities.

Writing up my critique of Bayes had always been on the to-do list on my blackboard. It finally appeared in the *Annals of Statistics* in 1982 [91]. It was taken seriously by Andrew Dale, who mentioned it several times in his comprehensive history of inverse probability [17], relating it to early sequels to Bayes. My friends Steve Stigler and Lorraine Daston also cited it in their masterful books on the history of probability and statistics [131, 18], but without commenting on the point I was making. I could hardly hold this against them;

they were wise to keep their historical work out of the line of fire from our impassioned contemporary Bayesians.

For Bayes, as for other eighteenth-century writers, probability was situated in time. He took into account what happened first. If B is to be decided first in the game we are playing, and the game defines the probability for A after B has happened or failed, then the relation

$$P(A\&B) = P(B)P(A \text{ after } B \text{ has happened})$$

follows from arguments made by Bayes's predecessors, especially Abraham De Moivre, and repeated by Bayes in the proof of the third proposition in his essay. The argument to which I objected was the one Bayes gave for his fifth proposition, which concerns the case where A is to be decided first and we learn of B 's happening without knowing whether A has happened or failed. It is here that we need a constitutive judgement about whether our learning about B 's happening or failing is independent of the uncertainties involved in A 's happening or failing. If we do not face up to this need, we open the way to paradoxes, as Joseph Bertrand was explaining already in the 19th century.²⁰

The arrow of time was effectively excluded from the foundations of probability theory by Kolmogorov's axioms. Stochastic processes were central to the work of Kolmogorov and the other mathematicians who developed measure-theoretic probability in the twentieth century, but when they based abstract probability on functional analysis, Fréchet and Kolmogorov made the time structure auxiliary to the basic structure and therefore apparently auxiliary to the meaning of probability [61, 122, 123]. My analysis of Bayes's two arguments convinced me that time should be brought back into the foundations. The fundamental object of probability theory should not be a flat space with a probability measure; it should be a probability tree—i.e., a tree with branching probabilities. We should begin abstract probability theory by proving the rule of iterated expectation and the law of large numbers in this setting. I developed this theme in "Conditional probability" in 1985 [94] and in a stream of subsequent articles.

By the mid-1980s, even as the Dempster-Shafer theory gained steam, putting probability trees into the foundation of probability had become the intellectual project closest to my heart. I believed it could unify the disparate and competing philosophies of probability: frequentist, Bayesian, logical, etc.—and even make a place for belief functions. As I saw it, the different meanings and applications of probability derived from an ideal picture involving a sequential betting game. Basic theorems such as the law of large numbers gained their applied meaning from the principle of the excluded gambling strategy, which said that no strategy for betting at the odds given in the game would result in a large gain. The different interpretations of probability were tied together in a circle of implication, from knowledge of the long run to fair price to warranted belief and back to knowledge of the long run.

²⁰See pp. 2–3 of Bertrand's 1889 book [11]. Later writers who discussed such paradoxes include Ball in 1911 ([2], p. 32) and many contemporary authors (see for example [3], pp. 262–264 of [94], pp. 3–15 of [122], and [138]).

In 1988–1989, I spent a sabbatical year at the Center for Advanced Study in the Behavioral Sciences in Palo Alto, where I proposed to begin a book entitled *The Unity and Diversity of Probability*. Like the previous versions of my great American probability book, it was never completed. “The unity and diversity of probability” became instead merely the title of a discussion article in *Statistical Science* [102].²¹ But probability trees led me to new intellectual vistas: causal conjecture and game-theoretic probability.

The year at Palo Alto also produced something more important: my marriage to the distinguished Princeton historian Nell Irvin Painter. Nell and I made history at the Center, being the first couple to come single and leave engaged. The marriage launched me on a period of happiness and productivity that I had not known since my earlier years in Princeton, and it brought me back to Princeton for many years.²²

8 Parametric statistics

“You have a great spiel about probability, but what happened to statistics?” This was Geoff Watson’s question when I dropped by to see him in Princeton in the early 1980s. Dempster’s papers had been about mathematical statistics. My first great success at Princeton had been to devise postulates that led to Dempster’s most general model. My first article had been entitled “A Theory of Statistical Evidence”. I had been one of the most promising young statisticians of my generation. What happened to statistics?

I never published the proof that my postulates implied Dempster’s structure of the second kind. The postulates singled out some nice properties, but the structure also had less appealing properties. John Aitchison had pointed to one in his discussion of Dempster’s 1968 paper at the Royal Statistical Society: even when two hypotheses give an observation the same probability, the observation can shift their upper and lower probabilities, not favoring the one over the other but indicating less uncertainty. Denis Lindley brought Aitchison’s example up in the discussion of my talk at the University of Western Ontario in May 1973, and in response I conceded the point and suggested that the simpler likelihood method that I later used in Chapter 11 of *A Mathematical Theory of Evidence* might be preferable.²³ This simpler method also had its paradoxes, which critics did not overlook, but at least it was simple.

The problem that Dempster had addressed with his upper and lower probabilities begins with the assumption that x , some not-yet observed element of a known set \mathcal{X} , is somehow governed by a probability distribution that we do not know. We assume that this distribution is one of a set of distributions that we do know, and this set is indexed by a parameter, say θ . We may call the

²¹See also [103, 106, 107].

²²Nell and I were married in the chapel on the Princeton campus in October 1989. Dick Jeffrey’s wife Edie was the witness for our New Jersey marriage license.

²³Aitchison’s comments and Dempster’s response are on pp. 234–237 and 245–246 of [32]. My exchange with Lindley is on pp. 435–436 of [85].

indexed set of probability distributions, say $\{P_\theta\}_{\theta \in \Theta}$, our *parametric model* or *specification*. Now we observe x . What does this observation tell us about the value of θ ? Does it give us degrees of belief or probabilities for the value? If we begin with probabilities for θ , then $P_\theta(x)$ can be interpreted as the probability of x after θ is decided, and Bayes's third proposition gives updated probabilities for θ ; this Bayesian solution is uncontroversial if the initial probabilities for θ are agreed to. But if we have no initial probabilities for θ , then we have a problem, the problem Fisher proposed to solve with his fiducial argument and Dempster proposed to solve with his upper and lower probabilities.

After completing my dissertation, I was happy with the theory of evidence that I had extracted from Dempster's work but not equally happy with any solution it provided to the question he had been addressing. Perhaps the fault was with the question. Perhaps it was not well posed. How could you know that x was governed by P_θ without knowing θ ? What kind of evidence could provide such knowledge? In the final chapter of *A Mathematical Theory of Evidence*, I hinted at this disenchantment by emphasizing that a parametric model or specification is always a provisional and unsupported assumption. Dempster, with whom I talked frequently, went a step farther in the foreword he wrote for the book; its last sentence reads, "As one thought for the future, I wish to suggest that, just as traditional Bayesian reasoning has been shaken loose from its moorings, perhaps there will appear a comparable weakening of the strong form of information implied by a typical specification."

The list on my blackboard always included returning to this topic, and I finally did so with a paper that I presented to the Royal Statistical Society in May 1982 [92]. Here I noted that a given parametric model can be represented by many different belief functions. For a given model $\{P_\theta\}_{\theta \in \Theta}$, that is to say, there are many different belief functions Bel on $\Theta \times \mathcal{X}$ that produce P_θ when conditioned on θ and can be conditioned on x to produce posterior degrees of belief about θ . The choice among these should, I argued, depend on our evidence for the parametric model. If this evidence tells us only that x is random, with no restriction on its possible probability distributions, then perhaps we should use Dempster's simplicial multinomial model. If the different values of θ are different diseases and we have extensive experience with each disease, perhaps we should use a model developed by Smets in his dissertation. If our evidence actually takes the form of a single frequency distribution, such as a distribution for errors, and we have constructed the model by anchoring the frequency distribution to different values of θ , then we should surely use what amounts to a fiducial argument that continues to believe the frequency distribution. If our evidence is scanty, perhaps it merits no more than the likelihood method in Chapter 11 of *A Mathematical Theory of Evidence*, for all its shortcomings.

My presentation was followed by a lengthy and friendly discussion. Afterwards Peter Williams whisked me off to visit Bristol and the village of Tunbridge Wells, famous to statisticians as the home turf of Bayes. When the paper appeared in the society's journal, the discussion included contributions from good friends, others I had known only by reputation (such as George Barnard), and a brilliant newcomer, Terry Fine's student Peter Walley. The contribution that

impresses me most now is Dave Krantz's; he argued, convincingly, that in addition to my randomly coded messages we can use other canonical examples to construct belief functions, including the likelihood examples he had developed with John Miyamoto [64].

I drifted away from mathematical statistics in the 1980s, and in retrospect the paper for the Royal Statistical Society feels almost like a farewell. Its underlying message is that the problem of parametric inference, central as it has been to mathematical statistics, is artificial; we never really have evidence that justifies parametric models, so let's look instead at the evidence we do have and model it. Let's do evidence instead of mathematical statistics. But this is not the way I felt at the time. I was a proud statistician, proud to have been invited to read a paper at the Royal Statistical Society. It was only after a campaign to convince the administration at Kansas to launch a statistics department that I moved to the business school,²⁴ and my attention drifted away from statistics as a discipline only because other disciplines came insistently calling.

9 Artificial intelligence

At the beginning of the 1980s, *A Mathematical Theory of Evidence* had been reviewed in several statistics and philosophy journals, and I was happily busy working through the list on my blackboard (and busy with family; my first son arrived in 1978, the second in 1981). But hardly anyone else had taken up belief functions. Aside from Dempster and myself, the only person who published a journal article on the topic in the 1970s was Hung Nguyen, who had related belief functions to random sets while working with Lotfi Zadeh [74].²⁵ This changed abruptly in 1981, when belief functions were discovered by the field of artificial intelligence.

At the Seventh International Joint Conference on Artificial Intelligence in Vancouver, British Columbia, in August 1981, three papers by California researchers concluded that *A Mathematical Theory of Evidence* provided the means to reason under uncertainty. Jeff Barnett, at Northrop in Los Angeles [4], christened the theory the *Dempster-Shafer theory*, a name that stuck.²⁶ Tom Garvey and his group at SRI in Palo Alto called it *evidential reasoning* [50]. Leonard Friedman, at the Jet Propulsion Laboratory in Pasadena, called

²⁴All the statisticians on campus supported the proposal for a new department. Fred Mosteller of Harvard and Ingram Olkin of Stanford, two of the country's most respected statistical statesmen, came to campus to support it, and we wrote a persuasive report on how the university would benefit. The subsequent success of statistics nationwide confirms the wisdom of that report, even from a purely financial point of view. But moving Glenn Shafer to the Business School was obviously more feasible than finding money for a new department in the College of Liberal Arts and Sciences.

²⁵In [75], Nguyen recalls that Zadeh asked him to read my dissertation when he arrived at Berkeley in the late winter of 1975. After doing so, Nguyen explained the connection to random sets to Zadeh. Shortly afterwards, after hearing Suppes say something similar about my dissertation, Zadeh asked Nguyen to put his insights into writing.

²⁶For years afterwards, Barnett would apologize, every time we talked, that he had not called it the *Shafer-Dempster theory*.

it *plausible inference* and suggested using it in systems such as MYCIN and PROSPECTOR [49].

The authors of the three papers had not discovered *A Mathematical Theory of Evidence* independently. They had all learned about it from John Lowrance, who was using belief functions in a doctoral dissertation at the University of Massachusetts at Amherst and had toured the California labs with a job talk in 1980, finally taking a position in Garvey's group at SRI.²⁷ John had first seen the book in 1978 or 1979, when Nils Nilsson, then visiting Amherst from Stanford, gave him a copy he had received from Dick Duda. The book was evidently circulating in the Stanford artificial intelligence community, probably as a result of the influence of Suppes, Tversky, Zadeh, and my own presence at Stanford in 1978–1979.

Artificial intelligence was resurgent in the early 1980s, with new applications, new excitement and expanded funding. Everyone was talking about expert systems. Numerous researchers followed up on the discovery of belief functions at the Vancouver meeting, and I was soon being invited across the country and to Europe to talk about my work at corporate research centers, university computer science and engineering departments, and artificial intelligence conferences. I was hardly a computer scientist, but people were looking to me to tell them how to use degrees of belief in computer systems.

During the 1960s and 1970s, artificial intelligence had developed its own culture, distinct even within computer science and very different from the culture of statistics. It seemed to think in formal logic. I spent a month with Tom Garvey's group at SRI in June 1985, talking about the construction of belief-function arguments and wondering whether I could contribute more if I learned to express my thoughts in Lisp.²⁸ In 1984 I had submitted a paper to the annual meeting of the American Association for Artificial Intelligence, showing by example how dependent items of evidences can be resorted into items that are independent in the sense of Dempster's rule. Not surprisingly, it had been rejected.²⁹ Aside from its brevity, it hardly resembled a computer-science conference paper. I had more success telling statisticians about probability in artificial intelligence [100].

But you did not need to become a hard-core AI researcher to join the AI bandwagon. All sorts of engineers and applied mathematicians were jumping on. Zadeh was promoting his fuzzy sets and possibility measures with considerable success. Jim Bezdek invited me to the First International Conference on Fuzzy

²⁷Lowrance's Structured Evidential Argumentation System, eventually deployed widely in federal intelligence agencies, used a Likert scale to obtain inputs from analysts but combined arguments using Dempster's rule and belief-function discounting; see [70]. For an explicit discussion of the use of belief functions in intelligence, see [134, 135].

²⁸My visit at SRI was funded by Enrique Ruspini's grant from the Air Force. Others in Garvey's group included Lowrance and Tom Strat. I encountered e-mail for the first time there; it was fun and sometimes even useful, even when my correspondents were down the hall.

²⁹I distributed and cited it as Kansas School of Business Working Paper No. 164, "The problem of dependent evidence", October 1984. It is published for this first time in this issue of *International Journal of Approximate Reasoning*.

Information Processing in Hawaii in 1984, where I met a whole new cast of characters, including the dynamic duo Didier Dubois and Henri Prade, who were making the whole world fuzzy with infectious good nature.³⁰ People were often surprised that I looked so young. Famous for a classic book, I should have been ancient. I contributed a long paper to Bezdek's proceedings, trying to place Sugeno measures and Zadeh's possibility measures alongside Bayes and belief functions as languages for managing uncertainty [98]. I also contributed to the first Uncertainty in Artificial Intelligence meeting, in Los Angeles in 1985 [99]. In those years the UAI meetings brought together everyone with opinions about the management of uncertainty, including Zadeh and the usual quota of impassioned Bayesians.

Although I enjoyed the out-of-town attention, my fondest academic memories from the mid-1980s are of the seminar I led in the Business School at Kansas after moving there from Mathematics in 1984. Several of my fellow statisticians joined in, but my strongest collaborations were with two colleagues who were not statisticians, Prakash Shenoy, who had worked in game theory and decision theory, and Raj Srivastava, a physicist turned accountant. Prakash and I focused on belief-function computation, and Raj and I worked on the applications of belief functions to financial auditing. Prakash and Raj were generous, thoughtful, and capable colleagues, and we had funding for students.³¹

In 1985, Jean Gordon and Ted Shortliffe at Stanford published an approximate implementation of Dempster's rule for simple support functions that could be arranged hierarchically in a tree [53]. In 1984, when they sent me a preprint of the article, I realized that the approximation was unneeded, and I asked a graduate student in Mathematics at Kansas, Roger Logan, to work out the details. Since Roger was new to belief functions, this took a while, but by the middle of 1985 we completed a working paper and submitted it to *Artificial Intelligence*. It appeared in 1987 [116]. In the meantime, I discussed the problem with my colleagues in the seminar in the Business School, and it became clear that the modular computation made possible by Gordon and Shortliffe's hierarchical structure was much more general. Judea Pearl was just beginning, at that time, to discuss modular Bayesian computation in the context of graphical models, and we saw that our picture was also more general than his. So we published our own results as soon as we could. In a conference paper at a meeting in Miami Beach at the end of 1985, I pointed out that Gordon and Shortliffe's hierarchical picture made modular computation possible even if the individual

³⁰Dubois and Prade also had plenty to say about belief functions; see for example [45].

³¹I could now attract federal and corporate funding well beyond the summer-salary grants I had been receiving since I became an assistant professor at Princeton. When Ron Yager became a program officer at the National Science Foundation, he telephoned me out of the blue to urge me to apply for a larger grant. Additional funding came from the business school's Ronald G. Harper Chair in artificial intelligence, to which I was appointed in 1988. In 1971, Harper had launched a business using an algorithm he learned as an MBA student at Kansas to locate retail outlets. His methodology became steadily more complex, and by the 1980s he was advertising it as artificial intelligence. When he took his business public, he donated funds for the school to endow a chair in artificial intelligence. A national search led to the conclusion that I was better qualified to fill it than anyone else the school could recruit.

belief functions were more complicated than simple support functions. In an article in *IEEE Expert* in 1986, Shenoy and I explained the general structure of modular belief-function computation; we also explained that Pearl's Bayesian scheme is a special case, because Bayesian conditioning is a special case of Dempster's rule [125]. This theme was further developed in a doctoral dissertation by Khaled Mellouli (1987) and articles that he co-authored, including [118]. The seminar also produced doctoral dissertations by Pierre Ndilikilikisha (1992) and Ali Jenzarli (1995).

The most influential product of my collaboration with Shenoy was the article in which we put the basic ideas of probability propagation into axiomatic form; it appeared in the proceedings of the 4th Uncertainty in AI meeting in 1988 and in a more widely circulated form in 1990 [126]. Shenoy developed the idea into a general methodology for expert systems [124]. In 1991, realizing that very similar schemes for modular computation arise in many other areas of applied mathematics, including database management, dynamic programming, and the solution of linear equations, I wrote a lengthy working paper trying to identify axioms that would bring out the common features of all these schemes [104].³² I ultimately left this project aside as too ambitious, but others, especially Jürg Kohlas and his collaborators [60], have continued to develop it.

My collaboration with Srivastava was also very rewarding, because Dempster-Shafer theory is well adapted to assessing the evidence that arises in financial auditing. Although statistical sampling has played some role there, most of the evidence that an auditor must consider does not involve statistical models, and combination of different types of evidence is of paramount importance.³³ Moreover, standard advice on combining the different risks in an audit has long been formulated in ways that echo Hooper's rules. Srivastava, Shenoy, and I attracted grants from the foundation funded by the accounting firm KPMG, developed expert systems that used belief-function algorithms,³⁴ and published articles in leading accounting journals [119, 130]. The collaboration wound down after I moved to Rutgers in 1992, but Srivastava has continued to develop the use of Dempster-Shafer theory in accounting, often in collaboration with his students and with other leading accounting researchers. Several of Raj's students, including Saurav Dutta and Peter Gillett, were later my colleagues in the accounting department at Rutgers. Gillett had built an expert system based on belief functions for an accounting firm before he pursued his doctorate [51], and

³²This working paper profited from advice from Steffen Lauritzen, who hosted me for a month at Aalborg in the spring of 1991.

³³The combination of evidence in legal settings has these same features. I did not have the opportunity to collaborate with legal professionals as I did with accounting professionals during the 1980s, but my work on the construction of arguments did bring me into contact with the legal tradition in this area. I especially enjoyed my interactions with Peter Tillers, who was a thoughtful leader in studying the role of probability in the law until his death from Lou Gehrig's disease in 2015. My work on the use of belief functions arguments in legal settings appeared in the *Boston University Law Review* in 1986 [96].

³⁴Raj wanted to patent one of our algorithms, but I did not want to go down that path. Raj later had commercial success with algorithms for searching financial statements, including one for which he and the University of Kansas obtained a patent.

at Rutgers he worked with me on causality.

The computational issues that arose in the 1980s also brought Art Dempster back into active work on belief functions. Three of Dempster's doctoral students wrote dissertations on belief-function computation in the 1980s: Augustine Kong (1986), Mathis Thoma (1989), and Russell Almond (1989).³⁵ Kong and Dempster published together on computational issues [42], and Almond's work appeared in book form [1]. Other writing on belief functions by Dempster in this period included [35, 36, 37]. He and I continued to keep in touch; I continued to visit Harvard, and he also visited Kansas. We never actually wrote together, but the degree to which we remained on the same page about basic issues is remarkable.

The work at Kansas and Harvard was only a fraction of the work on belief functions in the second half of the 1980s and the early 1990s. Then as now, the greatest part of this work was in Europe. From his base in Brussels, Philippe Smets played the most significant leadership role. He was known not only for his own innovations but also for a distinctive vocabulary; instead of talking about *committing belief* as I had done in my dissertation and in *A Mathematical Theory of Evidence*, he preferred to talk about *transferring belief*, and he shied away from the normalization step in Dempster's rule [128]. At least equally important was his role in organizing exchanges among others working on belief functions.³⁶ Other leaders included Jean-Yves Jaffray and Bernadette Bouchon-Meunier in Paris and Jürg Kohlas in Switzerland. Kohlas and his collaborators added yet another dialect to the language of belief functions; for them, it was a *theory of hints* [59].

For my own part, I did not keep up with the explosion of work on belief functions in the late 1980s. I was turning more and more to probability trees and the unity of probability, and I felt I had said what I had to say about belief functions. I was repeating myself, and the repetition was no longer productive. In the first decade of my work on belief functions, in the 1970s and into the 1980s, I was working in a relatively small community of statisticians and philosophers. We had no internet in those days; you could not so easily bring yourself up to date on someone else's work. But those interested in the foundations of probability and statistics saw and discussed each other's work by correspondence and in a relatively small number of journals and meetings. The late 1980s felt different. We still had no internet, and the interest in belief functions had outgrown any manageable community. Many who read *A Mathematical Theory of Evidence* in the late 1980s and then proceeded to use, extend, and criticize it were unaware of what I and others had said five or ten years earlier about questions they were raising. Indeed, the very features that made the book so successful—its self-contained appearance and sense of innovation—encouraged reinvention³⁷ and disinterest in what might have come before.

³⁵In the early 1970s, two other students of Dempster's had completed dissertations on upper and lower probabilities: Herbert Weisberg (1970) and Sandra West (1972).

³⁶Smets's contributions were cut tragically short when he died of a brain tumor in 2005, at the age of 66 [10].

³⁷Tom Strat's invention of continuous belief functions in his paper for the 1984 meeting

International Journal of Approximate Reasoning, launched by Jim Bezdek in 1987, became recognized over time as the leading outlet for work on belief functions, providing a center that helped make the topic less fragmented than in the 1980s. In 1990, in a special issue of the journal devoted to belief functions, I once again gave an overview and responded to critics [101], and I expanded on the response two years later [105]. I also joined with Judea Pearl, with whom I had enjoyed a friendship of some years, to publish a book of readings that situated Bayes, belief functions, and other approaches to uncertainty management in artificial intelligence in the larger traditions of probability and logic [117]. This was published in 1990 by Morgan Kaufman, a well regarded publisher in artificial intelligence, and it served a need at the time. These contributions marked the end of a chapter for me. In the 1990s, I would focus on causality and game-theoretic probability, not on belief functions, and I would drift away from artificial intelligence as I had earlier drifted away from statistics.

I recall another revealing interaction with *International Journal of Approximate Reasoning*. My paper on Dempster’s rule had been setting on the shelf since its summary rejection by *Annals of Probability* and another journal a decade earlier. When Bezdek recruited me as one of the initial associate editors of his new journal and told me about his interest in articles on belief functions, it occurred to me that this old paper might finally find a home. He encouraged the submission, but the paper was badly out of date, making no reference to newer work on Sugeno and Choquet integrals. One referee made the point with exasperation: “Shafer only knows what Shafer has done.” Ouch. A direct hit. But as I realized, my disinterest extended to what I had been trying to do in my dissertation. The point of constructing Dempster’s rule abstractly in infinite spaces had been to give belief functions an interpretation and a mathematics independent of probability theory. Now, as Dave Krantz had told me, my philosophy of constructive probability and canonical examples had taken me back to a view closer to where Dempster had started. A belief function was obtained by “continuing to believe” probabilities. Since the combination involved in Dempster’s rule could be accomplished in the probability domain, there was no need for a mathematics of combination autonomous from probability. For this reason, I was not interested enough to find the time to update the paper to satisfy the referees. It remained on the shelf.³⁸

10 Causal conjecture

In 1970, when I began studying mathematical statistics, statisticians often avoided talk about causation. The word appeared most often in a negative

of the American Association for Artificial Intelligence [132] was emblematic for me. Tom was an important contributor, and the paper did more than most to advance the use of belief functions in artificial intelligence, but neither Tom nor his reviewers were aware that aside from *A Mathematical Theory of Evidence*, most previous work on belief functions had emphasized the continuous case.

³⁸It appears for the first time in this issue of *International Journal of Approximate Reasoning*.

statement: correlation does not imply causation. The goal of most statistical investigations is to understand causal relations, but because our concept of probability was clearly objectivist, we could express causal conclusions without using causal language. Once you had said that the probability of B is high whenever property A is present, it was superfluous and perhaps unnecessarily provocative to say that A causes B . This changed as Bayes gained ground, especially among those who believed that all probabilities are subjective. If probabilities are subjective, then we need causal language to describe relationships in the world. Influential calls for explicitly causal probabilistic language came from the philosopher Pat Suppes in 1970 [133] and the statistician Don Rubin in 1974 [79].

Probability theory is about events and variables. So when you marry it with causal talk, you will probably call these formal objects causes. So it was in Suppes's and Rubin's work, and so it was in the late 1980s, when researchers in artificial intelligence and statistics began calling networks of events and variables causal, as in the classic 1988 articles by Pearl [76] and by Lauritzen and Spiegelhalter [67].³⁹

Believing as I did that probability trees provide a more fundamental foundation for probability, I saw this as superficial. Events in the probabilistic sense are merely reports on how things come out at the end of a process. Causes, if we want to use this word, are more specific events localized in time—steps in nature's probability tree, if you will. Variables may be the cumulative result of many small steps in nature's probability tree. When I explained this to my friend Judea Pearl, he argued that causal relations must be law-like relations, not isolated happenings. I agreed, but causal laws should be rules governing nature's probabilities for what happens next on various steps across and through nature's tree, not laws governing the relations between the superficial and global variables that we are able to measure. Relations between such global variables can provide grounds for conjectures about this largely hidden causality, but that does not make the variables causes.

The distinction between probabilistic relations among variables and causal laws in the underlying probability tree can be illustrated with the concept of independence. When we consider two variables X and Y in a probability space, independence has a purely formal definition: X and Y are probabilistically independent if the probability of a pair of joint values, say x and y , can always be obtained by multiplication: $P\{X = x \text{ and } Y = y\} = P\{X = x\}P\{Y = y\}$. This is not a causal relation, but it is implied by a causal relation that might (or might not) hold in the underlying probability tree. The probabilities for X and Y change as time moves through the tree, and it is reasonable to say that X and Y are *causally independent* if there is no step in the tree that changes the probabilities for both X and Y . (The steps in the tree are the causes, and X and Y are causally independent if they have no common causes.) As it turns out, this causal independence implies probabilistic independence. So when we

³⁹These were the same "Bayesian" networks for which we were all studying schemes for local computation. I tried to explain these schemes in the simplest possible terms in a series of lectures I gave in North Dakota in 1992, which I finally published in 1996 [110].

observe probabilistic independence, we may conjecture that it is due to causal independence.

I became fascinated with this idea in early 1992, as I was preparing to move from Kansas to Rutgers in New Jersey. My wife Nell taught at Princeton, and after commuting between Kansas and New Jersey for three years, we wanted to live in the same state. I spent much of my time in Princeton for the first couple of years, and in the fall of 1992, I persuaded some of my friends in the philosophy department there⁴⁰ to join me in a seminar devoted to probabilistic causality, especially the work by Peter Spirtes, Clark Glymour, and Richard Scheines at Carnegie-Mellon [129]. During this seminar I explained my little theory of causal independence and tried to generalize it from independence to conditional independence, which characterized the networks that had been occupying the attention of statisticians and our colleagues in artificial intelligence.

Success. I did find a generalization to conditional independence. I started talking about it at the Fourth International Workshop on Artificial Intelligence and Statistics in Fort Lauderdale in January 1993. For the next two years, I did not want to talk about anything else. The result was my second book, *The Art of Causal Conjecture* [109], which appeared in 1996, twenty years after *A Mathematical Theory of Evidence*. I thought it equally important. It provided a probabilistic account of causality, and it demonstrated the power of my probability-tree foundation for probability. Of course, it did not have the unique explosive success that *A Mathematical Theory of Evidence* had enjoyed. My colleagues in statistics, philosophy, and artificial intelligence were still enthusiastically developing their own ways of thinking about causal networks of variables, and the book suffered from a defect I had only inadvertently avoided in *A Mathematical Theory of Evidence*: it did too much. I had discovered not just a single causal interpretation of a single probabilistic concept of conditional independence but a plethora of different probabilistic concepts with different causal interpretations. *The Art of Causal Conjecture* catalogs them all.⁴¹ I still believe that its picture will prove influential in the long run, but I did not do as much as I might have to make this picture more widely understood after the book appeared, because I was quickly drawn into a related and deeper enterprise: game-theoretic probability.⁴²

11 Game-theoretic probability

Early in 1992, I received a letter from Moscow. Dated 17 January, it read,

Dear Professor Shafer,

⁴⁰Dick Jeffrey, Gil Harman, and Bas van Fraassen. We were joined by Paul Holland from ETS and Adam Grove from NEC.

⁴¹The best introduction to the simplest example is in my article in the proceedings of the Paris meeting organized by Bernadette Bouchon-Meunier in 1994 [108].

⁴²I did return to causality in the late 1990s and early 2000s. I touched on the theme of causality and responsibility [111], and with Peter Gillett and Richard Scherl, I studied a logic of events derived from the idea of indefinitely refining a probability tree [115, 52]. This logic of events merits additional study.

In your 1990 paper in *Statist. Science* I found excellent exposition of ideas that guided me in my research during the past several years. It seems that we are moving in the same direction. Please find enclosed one of my papers. With great interest I am looking forward to reading your forthcoming book *The Unity and Diversity of Probability*.

Sincerely yours,

Volodya Vovk

Very flattering, all the more so that the enclosed paper was obviously brilliant and totally mystifying. I responded positively, and within a couple months Vovk and I were corresponding about possible collaboration.

Vovk was also in contact with Phil Dawid in London, and Dawid and I found multiple occasions to discuss his work. In September 1992, Vovk gave a discussion paper at the Royal Statistical Society [139] that built on Dawid's notion of prequential probability [20, 21, 22]; I could not attend but participated in the discussion. Dawid's prequential (sequential prediction) picture was related to my picture of a probability tree and had influenced my thinking as well as Vovk's. Central to Dawid's picture was his *prequential principle*, according to which the performance of a forecaster who gives successive probabilities should be assessed only by comparing his probabilities to the outcomes, ignoring any strategy the forecaster might have followed and any probabilities such a strategy might have recommended had events taken a different path.

Even though I was occupied with my book on causality from fall 1992 to spring 1995, Vovk and I continued to correspond about possible collaborations, including grant applications. By February 1994 we were discussing a possible book that would use probability trees to develop the theories of classical probability, filtering and control, and statistics. In May 1994 I traveled to Moscow to spend a week talking with him about our joint interests.

In June 1995, Steffen Lauritzen invited Vovk, Dawid, myself and others to a week-long seminar at Aalborg at which I presented the main results of *The Art of Casual Conjecture* and Vovk presented some of his results. Vovk's picture had evolved; instead of probability trees, he was now using formal games. These games are fully described by a probability tree when a probability distribution for outcomes is fixed in advanced, but the probabilities might instead be given in the course of the game, and this greater generality permits a more natural treatment of many applications of probability. During the meeting, with Lauritzen's encouragement, Vovk returned to the notion that he and I should write a book together. I was hesitant, because I was still eager to complete that great American probability book I had started at the Center in Palo Alto. But Vovk was insistent, and I agreed to his proposal a year later, in the summer of 1996, when we met again at my home in Princeton. He and his family were then arriving in the United States for a year at the Center in Palo Alto, and Nell and I were about to leave for a year's sabbatical in Paris. The book appeared five years later: *Probability and Finance: It's Only a Game!* [121]. In the 15

years since then, Vovk and I and others, especially Kei Takeuchi and Akimichi Takemura in Tokyo, have continued to develop its ideas.

Since the late 1930s, it has been a commonplace that Kolmogorov's axioms can be interpreted and used in different ways. They can be interpreted as axioms for objective probabilities, as axioms for degrees of belief, or as axioms for logical degrees of support. Mathematically, the game-theoretic foundation for probability Vovk and I proposed is a generalization of Kolmogorov's picture. Philosophically, it allows us to deepen our understanding of the diversity of interpretation. We begin with a flexible game that typically has three players: (1) Forecaster, who prices gambles, (2) Skeptic, who decides how to gamble, and (3) Reality, who decides the outcomes. We can use games of this form in many ways. For example:

Classical probability. Classical theorems such as the law of large numbers and the central limit theorem become more general, because the probabilities may be given by Forecaster freely as the game proceeds, rather than by a probability distribution for the whole sequence of Reality's moves, which can be thought of as a strategy for Forecaster.

Statistical testing. If the role of Forecaster is played by a theory (quantum mechanics, for example), we can take the role of Skeptic and test the theory by trying to multiply the capital we risk by a large factor; it turns out that all statistical tests can be represented in this way.

Forecasting. To make probabilistic forecasts, we can put ourselves in the role of Forecaster and play to beat Skeptic's tests.

Subjective probability. We can take the role of Forecaster to state prices that express our belief. Instead of suggesting, as de Finetti did, that we are offering these prices to all comers (this may be dangerous, because surely there are others who know more), we may say that we do not think we could beat them if we took the role of Skeptic.

Finance. We can put an actual financial market in the roles of Forecaster (the day's opening prices being Forecaster's move) and Reality (the day's closing prices being Reality's move), so that the hypothesis that Skeptic, the investor, cannot multiply his capital by a large factor becomes a testable version of the efficient-market hypothesis.

Causality. We can imagine that the probabilistic predictions are successful (inasmuch as Skeptic cannot beat them), but that the game is played out of our sight, so that we see only some aspects of Reality's moves. This leads to the theory of causality I had developed in *The Art of Casual Conjecture*.

This is a picture far richer and more powerful than what I imagined when I began writing, 25 years ago, about the unity and diversity of probability.

The greatest historical and philosophical lesson I have drawn from 20 years of work on game-theoretic probability is the importance of what I called the

principle of the excluded gambling strategy in the 1980s. The game-theoretic picture gives this principle a more precise form; it says that no simple strategy will allow Skeptic to multiply the capital he risks by a large factor. When used in finance, this is a game-theoretic form of the efficient-market hypothesis. It is fundamental to most applications of game-theoretic probability, just as the corresponding principle that no event of small probability singled out in advance will happen is fundamental to most applications of measure-theoretic probability.

I call the principle that Skeptic will not multiply the capital he risks by a large factor the *game-theoretic Cournot principle*. As I learned in the historical studies I undertook in order to understand game-theoretic probability, it was Cournot who first argued that probability connects with the world only by means of the principle that an event of small probability will not happen. This principle was adopted explicitly by Cournot's greatest continental successors, including Borel and Kolmogorov [122, 112, 12]. But the principle was never as prominent in the English-language literature, and it became less salient as the center of gravity of probability and statistics shifted to English after World War II. Its game-theoretic reformulation reminds us of its central role.

The game-theoretic Cournot principle also enriches the interpretation of belief functions. As I have explained, Dempster-Shafer judgements measure evidence by comparing it to canonical examples where we continue to believe certain probabilities. What exactly are we continuing to believe about these probabilities? We are continuing to believe that they cannot be beat—that a betting opponent will not, using a simple strategy, multiply the capital he risks by a large factor betting at the implied odds. I developed this idea in *International Journal of Approximate Reasoning* in 2011 [114].

12 Dempster-Shafer redux

Although I have done little work directly related to belief functions since the 1980s, I have been pleased to see that research on the topic has continued apace and is, in some respects, resurgent. We see a steady stream of applications of belief functions in meetings such as FUSION (International Conference on Information Fusion) and ISIPTA (International Symposium on Imprecise Probability: Theories and Applications), and other applications arise in a bewildering variety of fields.

The belief-function community in Europe suffered a heavy blow with Philippe Smets's death in 2006, but it has clearly recovered, with the organization of the Belief Functions and Applications Society in 2010 and its subsequent growth. I was honored to participate in the Society's founding meeting, along with Art Dempster, in Brest in April 2010. Thierry Denoeux, the Society's current president, has outstripped Smets in the breadth of belief-function applications that he has explored with his students at Compiègne.

In 2009, Dempster and I accepted honorary doctorates from the University of Economics in Prague, where Radim Jiroušek has led a strong group in belief

functions. After focusing on applied problems, including climate change, in the 1990s, Dempster returned to working and publishing on belief functions. In 2001, he explained the belief-function interpretation and computation of Kalman filters [38]. He reviewed the history of belief functions in 2003 [39], published on the use of belief functions for object recognition in 2006 [41], and published a general exposition for statisticians in 2008 [40]. Two additional students have completed dissertations on belief functions at Harvard under his supervision: Wai Fung Chiu (2004) and Paul Edlefsen (2009).

Within mathematical statistics, which plays a relatively large role in the United States, belief functions have been largely dormant since the 1980s. The Bayesian paradigm has mushroomed in importance during this period, as statisticians have realized that Neyman-Pearson methods cannot cope with the field's increasingly complex parametric models, and as computational methods for Bayesian solutions have been developed. One might even say that the dominant research paradigm has become the development of Bayesian computational methods. Discussions of philosophical foundations, even in forums where they were once popular, such as the International Society for Bayesian Analysis, have been displaced by discussions of computation. A similar shift has occurred in artificial intelligence, for example in the Uncertainty in AI meetings.

In very recent years, however, puzzles encountered by complex Bayesian models are leading to a reconsideration of fiducial and Dempster-Shafer methods. Even when they use Bayesian methods to obtain posterior probabilities for parameters, most statisticians retain an objectivist conception of the parametric model, and they have been most comfortable with Bayesian methods when they can identify prior probability distributions for the parameters that seem to represent ignorance, inasmuch as the posterior probabilities come close to having frequency properties like those required for Neyman-Pearson confidence coefficients. As parametric models have become increasingly complex, with huge numbers of parameters, it has become clear that this desideratum cannot be satisfied for all parameters. A prior probability distribution that seems to represent ignorance about one group of parameters may clearly express very strong opinions about another. In response to this conundrum, a number of authors have been studying fiducial and Dempster-Shafer style arguments that have attractive properties for targeted parameters. See the recent book by Ryan Martin and Chuanhai Liu [71] and the recent review article by Jan Hannig and his collaborators [54].

In the spirit of my earlier discussion of parametric models (and Dempster's remark in the foreword to *A Mathematical Theory of Evidence*), we might address this issue by asking what evidence could possibly justify such complex parametric models. To construct a fiducial and or Dempster-Shafer analysis, we must decide what aspects of a parametric model we can continue to believe after observations are made. Perhaps we should first ask what aspects we can believe in the first place. What bets do we really believe can be safely offered to an opponent trying to multiply their capital by a large factor?

13 Conclusion

After studying probability and partial belief for 45 years, my sturdiest belief about the enterprise is that the most enduring advances will draw on history. Intellectual developments, like so many other aspects of life, are path-dependent. We cannot understand where we are, nor how to get out, without understanding how we got there.

The Dempster-Shafer theory of belief functions was launched 40 years ago in the context of long-standing puzzles concerning the representation of ignorance and the interplay between subjective and objective aspects of probability. My intellectual path since then has involved looking back at how these puzzles arose: how probability theory grew out of the picture of betting, how we acquired the concept of conditional probability, how probabilities became additive, how time was pushed out of the foundations. I have found a few answers to these questions, and I believe these answers can help us move forward.

References

- [1] Russell G. Almond. *Graphical Belief Modeling*. Chapman and Hall, New York, 1995. 24
- [2] Walter William Rouse Ball. *Mathematical Recreations and Essays*. Macmillan, London, 5th edition, 1911. 17
- [3] Maya Bar-Hillel and Ruma Falk. Some teasers concerning conditional probabilities. *Cognition*, 11:109–122, 1982. 17
- [4] Jeffrey A. Barnett. Computational methods for a mathematical theory of evidence. In *IJCAI-81: Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 868–875, 1981. Reprinted as Chapter 8 of [144]. 20
- [5] Thomas Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London for 1763*, 53:370–418, 1764. 16
- [6] David E. Bell, Howard Raiffa, and Amos Tversky, editors. *Decision Making: Descriptive, Normative, and Prescriptive Interactions*. Cambridge University Press, Cambridge, 1988. 13, 32, 34, 39, 41
- [7] David E. Bell, Howard Raiffa, and Amos Tversky. Descriptive, normative, and prescriptive interactions in decision making. In *Decision Making: Descriptive, Normative, and Prescriptive Interactions* [6], pages 9–30. 14
- [8] Jacob Bernoulli. *Ars Conjectandi*. Thurnisius, Basel, 1713. 6, 32
- [9] Jacob Bernoulli. *The Art of Conjecturing, together with Letter to a Friend on Sets in Court Tennis*. Johns Hopkins University Press, Baltimore, 2006. Translation of [8] with commentary, by Edith Sylla. 6, 10

- [10] Hughes Bersini et al. In memoriam: Philippe Smets (1938–2005). *International Journal of Approximate Reasoning*, 41:iii–viii, 2006. 24
- [11] Joseph Bertrand. *Calcul des Probabilités*. Gauthier-Villars, Paris, 1889. Second edition 1907. 13, 17
- [12] Laurent Bienvenu, Glenn Shafer, and Alexander Shen. On the history of martingales in the study of randomness. *Electronic Journal for History of Probability and Statistics*, 5(1), 2009. jehps.net. 30
- [13] Émile Borel. A propos d’un traité de probabilités. *Revue philosophique de la France et de l’étranger*, 98:321–336, 1924. Reprinted in [15], Volume 4, pp. 2169–2184, and in [14], pp. 134–146. An English translation by Howard E. Smokler was published in the first edition (only) of [65]. 13
- [14] Émile Borel. *Valeur pratique et philosophie des probabilités*. Gauthier-Villars, Paris, 1939. 33
- [15] Émile Borel. *Œuvres de Émile Borel*. Éditions du Centre National de la Recherche Scientifique, Paris, 1972. Four volumes. 33
- [16] Antoine-Augustin Cournot. *Exposition de la théorie des chances et des probabilités*. Hachette, Paris, 1843. 4
- [17] Andrew I. Dale. *A History of Inverse Probability From Thomsas Bayes to Karl Pearson*. Springer, New York, second edition, 1999. 16
- [18] Lorraine Daston. *Classical Probability in the Enlightenment*. Princeton University Press, Princeton, NJ, 1988. 16
- [19] Florence Nightingale David. *Games, Gods, and Gambling*. Griffin, London, 1962. 3
- [20] A. Philip Dawid. Fisherian inference in likelihood and prequential frames of reference (with discussion). *Journal of the Royal Statistical Society. Series B*, 53:79–109, 1991. 28
- [21] A. Philip Dawid. Prequential data analysis. In Malay Ghosh and Pramod K. Pathak, editors, *Current Issues in Statistical Inference: Essays in Honor of D. Basu*, IMS Lecture Notes–Monographs Series 17, pages 113–126. Institute of Mathematical Statistics, Hayward, CA, 1992. 28
- [22] A. Philip Dawid and Vladimir G. Vovk. Prequential probability: Principles and properties. *Bernoulli*, 5:125–162, 1999. 28
- [23] Bruno de Finetti. *Teoria Delle Probabilità*. Einaudi, Turin, 1970. An English translation, by Antonio Machi and Adrian Smith, was published as *Theory of Probability* by Wiley in two volumes in 1974 and 1975. 3

- [24] Augustus De Morgan. *Formal Logic: or, The Calculus of Inference, Necessary and Probable*. Taylor and Walton, London, 1847. 6
- [25] Arthur P. Dempster. Further examples of inconsistencies in the fiducial argument. *Annals of Mathematical Statistics*, 33:884–891, 1963. 5
- [26] Arthur P. Dempster. On a paradox concerning inference about a covariance matrix. *Annals of Mathematical Statistics*, 33:1414–1418, 1963. 5
- [27] Arthur P. Dempster. On direct probabilities. *Journal of the Royal Statistical Society, Series B*, pages 100–110, 1963. 5
- [28] Arthur P. Dempster. On the difficulties inherent in Fisher’s fiducial argument. *Journal of the American Statistical Association*, 59:56–66, 1964. 5
- [29] Arthur P. Dempster. New methods for reasoning towards posterior distributions based on sample data. *Annals of Mathematical Statistics*, 37:355–374, 1966. 5
- [30] Arthur P. Dempster. Upper and lower probabilities induced by a multi-valued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967. 5, 6
- [31] Arthur P. Dempster. Upper and lower probability inferences based on a sample from a finite univariate population. *Biometrika*, 38:512–528, 1967. 5, 6
- [32] Arthur P. Dempster. A generalization of Bayesian inference (with discussion). *Journal of the Royal Statistical Society, Series B*, 30:205–247, 1968. 5, 18
- [33] Arthur P. Dempster. Upper and lower probabilities generated by a random closed interval. *Annals of Mathematical Statistics*, 39:957–966, 1968. 5, 6
- [34] Arthur P. Dempster. Upper and lower probability inferences for families of hypotheses with monotone density ratio. *Annals of Mathematical Statistics*, 40:953–969, 1969. 5
- [35] Arthur P. Dempster. Probability, evidence, and judgment (with discussion). In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics 2*, pages 119–131. North-Holland, Amsterdam, 1985. Reprinted in [6], pp. 284–298. 24
- [36] Arthur P. Dempster. Bayes, Fisher, and belief functions. In Seymour Geisser et al., editor, *Bayesian and Likelihood Methods in Statistics and Econometrics. Essays in Honor of George Barnard*, pages 35–47. North-Holland, 1990. 24

- [37] Arthur P. Dempster. Construction and local computation aspects of belief functions. In R. M. Oliver and J. Q. Smith, editors, *Influence Diagrams, Belief Nets, and Decision Analysis*, pages 121–141. Wiley, New York, 1990. 24
- [38] Arthur P. Dempster. Normal belief functions and the Kalman filter. In A. K. Mohammed E. Saleh, editor, *Data Analysis from Statistical Foundations: A Festschrift in honour of the 75th birthday of D. A. S. Fraser*, pages 65–84. Nova, Huntington, New York, 2001. 31
- [39] Arthur P. Dempster. Theory of belief functions: History and prospects. In Bernadette Bouchon-Meunier et al., editor, *Intelligent Systems for Information Processing: From Representation to Applications*, pages 213–222. Elsevier, Amsterdam, 2003. 5, 13, 31
- [40] Arthur P. Dempster. The Dempster-Shafer calculus for statisticians. *International Journal of Approximate Reasoning*, 48:365–377, 2008. 31
- [41] Arthur P. Dempster and Wai Fung Chiu. Dempster–Shafer models for object recognition and classification. *International journal of Intelligent Systems*, 21(3):283–297, 2006. 31
- [42] Arthur P. Dempster and Augustine Kong. Uncertain evidence and artificial analysis. *Journal of Statistical Planning and Inference*, 20:355–368, 1988. 24
- [43] Thierry Dencœux. Likelihood-based belief function: Justification and some extensions to low-quality data. *International Journal of Approximate Reasoning*, 55(7):1535–1547, 2014. 9
- [44] William Fishburn Donkin. On certain questions relating to the theory of probabilities. *Philosophical Magazine*, 1:353–368, 1851. Continued in Volume 2, pp. 55–60. 6
- [45] Didier Dubois and Henri Prade. A set-theoretic view of belief functions. *International Journal of General Systems*, 12:193–226, 1986. 22
- [46] Ronald A. Fisher. Inverse probability. *Proceedings of the Cambridge Philosophical Society*, 26(4):528–535, 1930. 4
- [47] Ronald A. Fisher. *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh, first edition, 1956. 4, 9
- [48] Ronald A. Fisher. The underworld of probability. *Sankhya*, 18:201–210, 1957. 5
- [49] Leonard Friedman. Extended plausible inference. In *IJCAI-81: Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 487–495, 1981. 21

- [50] T. D. Garvey, J. D. Lowrance, and M. A. Fischler. An inference technique for integrating knowledge from disparate sources. In *IJCAI-81: Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 319–325, 1981. 20
- [51] Peter Gillett. Automated dynamic audit programme tailoring: An expert system approach. *Auditing: A Journal of Practice & Theory*, 12 Supplement:173–189, 1993. 23
- [52] Peter R. Gillett, Richard B. Scherl, and Glenn Shafer. A probabilistic logic based on the acceptability of gambles. *International Journal of Approximate Reasoning*, 44(3):281–300, 2007. 27
- [53] Jean Gordon and Edward H. Shortliffe. A method for managing evidential reasoning in a hierarchical hypothesis space. *Artificial Intelligence*, 26:323–357, 1985. Reprinted as Chapter 12 of [144]. 22
- [54] Jan Hannig, Hari Iyer, Randy C. S. Lai, and Thomas C. M. Lee. Generalized fiducial inference: A review. *Journal of the American Statistical Association*, 2016. To appear, doi:10.1080/01621459.2016.1165102. 31
- [55] George Hooper. A calculation of the credibility of human testimony. *Philosophical Transactions of the Royal Society of London*, 21:259–365, 1699. 8, 11
- [56] Edwin T. Jaynes. *Probability theory: The Logic of Science*. Cambridge University Press, 2003. 14
- [57] Richard Jeffrey. *Subjective Probability (The Real Thing)*. Cambridge University Press, Cambridge, 2004. 11
- [58] Maurice G. Kendall. On the reconciliation of theories of probability. *Biometrika*, 36(1/2):101–116, 1949. 3
- [59] Jürg Kohlas and Paul-André Monney. *A Mathematical Theory of Hints: An Approach to the Dempster-Shafer Theory of Evidence*. Springer, Berlin, 1995. 24
- [60] Jürg Kohlas et al. Generic local computation. *Journal of Computer and System Sciences*, 78(1):348–369, 2012. 23
- [61] Andrei Nikolaevich Kolmogorov. Grundbegriffe der Wahrscheinlichkeitsrechnung. In *Ergebnisse der Mathematik*, volume 2, pages 195–262. Springer (for *Zentralblatt für Mathematik*), Berlin, 1933. An English translation by Nathan Morrison appeared under the title *Foundations of the Theory of Probability* (Chelsea, New York) in 1950, with a second edition in 1956. 17
- [62] David H. Krantz and Howard C. Kunreuther. Goals and plans in decision making. *Judgment and Decision Making*, 2:137–168, 2007. 15

- [63] David H. Krantz, R. Duncan Luce, Patrick Suppes, and Amos Tversky. *Foundations of Measurement, Vol. I: Additive and polynomial representations*. Academic Press, New York, 1971. Volume II, Geometrical, threshold, and probabilistic representations, appeared in 1989. Volume III, Representation, axiomatization, and invariance, appeared in 1990. 12
- [64] David H. Krantz and John Miyamoto. Priors and likelihood ratios as evidence. *Journal of the American Statistical Association*, 78:418–423, 1983. 20
- [65] Henry E. Kyburg Jr. and Howard E. Smokler, editors. *Studies in Subjective Probability*. Wiley, New York, 1964. A second edition, with a slightly different selection of readings, was published by Krieger, New York, in 1980. 33
- [66] Nan Laird. A conversation with F. N. David. *Statistical Science*, 4(3):235–246, 1989. 3
- [67] Steffen L. Lauritzen and David J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 50(2):157–224, 1988. 26
- [68] Erich L. Lehmann. *Fisher, Neyman, and the Creation of Classical Statistics*. Springer, New York, 2011. 4
- [69] Isaac Levi. Dissonance and consistency according to Shackle and Shafer. In Peter D. Asquith and Ian Hacking, editors, *PSA 1978, Volume 2*, pages 466–477, East Lansing, 1981. Philosophy of Science Association. Proceedings of the Biennial Meeting of the Philosophy of Science Association in San Francisco, 27–29 October 1978. 9
- [70] John Lowrance et al. Template-based structured argumentation. In Alexandra Okada et al., editor, *Knowledge Cartography: Software Tools and Mapping Techniques*, pages 443–470. Springer, London, 2008. 21
- [71] Ryan Martin and Chuanhai Liu. *Inferential models: Reasoning with uncertainty*. CRC Press, Boca Raton, 2016. 31
- [72] Frederick Mosteller and David L. Wallace. *Inference and Disputed Authorship: The Federalist*. Addison-Wesley, Reading, Massachusetts, 1964. Reprinted in 2006 by CSLI Publications. 3
- [73] James R. Newman. *The World of Mathematics*. Simon and Schuster, New York, 1956. Four volumes. 2
- [74] Hung T. Nguyen. On random sets and belief functions. *Journal of Mathematical Analysis and Applications*, 65:531–542, 1978. Reprinted as Chapter 5 of [144]. 20

- [75] Hung T. Nguyen. On belief functions and random sets. In Thierry Denoeux and Marie-Hélène Masson, editors, *Belief Functions: Theory and Applications. Proceedings of the 2nd International Conference on Belief Functions, Compiègne, France, 9–11 May 2012*, pages 1–19. Springer, Berlin, 2012. 20
- [76] Judea Pearl. Embracing causality in formal reasoning. *Artificial Intelligence*, 35(2):173–215, 1988. 26
- [77] Bonnie K. Ray and David H. Krantz. Foundations of the theory of evidence: Resolving conflict among schemata. *Theory and Decision*, 40(3):215–234, 1996. 15
- [78] André Revuz. Fonctions croissantes et mesures sur les espaces topologiques ordonnés. *Annales de l’Institut Fourier*, 6:187–269, 1955. 11
- [79] Donald Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974. 26
- [80] Leonard J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66:783–801, 1971. 3
- [81] Teddy Seidenfeld. Evidence and belief functions. In Peter D. Asquith and Ian Hacking, editors, *PSA 1978, Volume 2*, pages 478–489, East Lansing, 1981. Philosophy of Science Association. Proceedings of the Biennial Meeting of the Philosophy of Science Association in San Francisco, 27–29 October 1978. 16
- [82] Teddy Seidenfeld. R. A. Fisher’s fiducial argument and Bayes’ theorem. *Statistical Science*, 7(3):358–368, 1992. 4
- [83] Glenn Shafer. *Allocations of Probability*. PhD thesis, Department of Statistics, Princeton University, 1973. 6
- [84] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976. 9, 42
- [85] Glenn Shafer. A theory of statistical evidence (with discussion). In William L. Harper and Clifford A. Hooker, editors, *Foundations and philosophy of statistical inference: Volume II of the proceedings of an international research colloquium held at the University of Western Ontario, London, Canada, 10–13 May 1973*, pages 365–436, Boston, 1976. Dordrecht. 7, 18
- [86] Glenn Shafer. Non-additive probabilities in the work of Bernoulli and Lambert. *Archive for History of Exact Sciences*, 19:309–370, 1978. Reprinted as Chapter 6 of [144]. 10

- [87] Glenn Shafer. Allocations of probability. *Annals of Probability*, 7:827–839, 1979. Reprinted as Chapter 7 of [144]. 11
- [88] Glenn Shafer. Constructive probability. *Synthese*, 48:1–60, 1981. Reprinted as Chapter 9 of [144]. 13
- [89] Glenn Shafer. Jeffrey’s rule of conditioning. *Philosophy of Science*, 48:337–363, 1981. 11
- [90] Glenn Shafer. Two theories of probability. In Peter D. Asquith and Ian Hacking, editors, *PSA 1978, Volume 2*, pages 441–466, East Lansing, 1981. Philosophy of Science Association. Proceedings of the Biennial Meeting of the Philosophy of Science Association in San Francisco, 27–29 October 1978. 11, 16
- [91] Glenn Shafer. Bayes’s two arguments for the rule of conditioning. *Annals of Statistics*, 10:1075–1089, 1982. 16
- [92] Glenn Shafer. Belief functions and parametric models (with discussion). *Journal of the Royal Statistical Society, Series B*, 44:322–352, 1982. Reprinted as Chapter 10 of [144]. 19
- [93] Glenn Shafer. Lindley’s paradox (with discussion). *Journal of the American Statistical Association*, 77:325–351, 1982. 11
- [94] Glenn Shafer. Conditional probability (with discussion). *International Statistical Review*, 53:261–277, 1985. 17
- [95] Glenn Shafer. The combination of evidence. *International Journal of Intelligent Systems*, 1:155–179, 1986. 11
- [96] Glenn Shafer. The construction of probability arguments (with discussion). *Boston University Law Review*, 66:799–823, 1986. Reprinted in [6], pages 193–234. 23
- [97] Glenn Shafer. Savage revisited (with discussion). *Statistical Science*, 1:463–501, 1986. Reprinted in [6], pages 193–234. 14, 15
- [98] Glenn Shafer. Belief functions and possibility measures. In James C. Bezdek, editor, *The Analysis of Fuzzy Information, Vol. 1: Mathematics and Logic*, pages 51–84. CRC Press, Boca Raton, 1987. 22
- [99] Glenn Shafer. Probability judgment in artificial intelligence. In J. F. Lemmer and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence*, pages 127–135. North-Holland, Boca Raton, 1987. 22
- [100] Glenn Shafer. Probability judgment in artificial intelligence and expert systems (with related articles and discussion). *Statistical Science*, 2:3–44, 1987. 21

- [101] Glenn Shafer. Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning*, 4:323–362, 1990. 25
- [102] Glenn Shafer. The unity and diversity of probability (with discussion). *Statistical Science*, 5(4):435–462, 1990. 18
- [103] Glenn Shafer. The unity of probability. In George von Furstenberg, editor, *Acting Under Uncertainty: Multidisciplinary Conceptions*, pages 95–126. Kluwer, Boston, 1990. 18
- [104] Glenn Shafer. An axiomatic study of computation in hypertrees. Technical Report 232, School of Business, University of Kansas, October 1991. www.glennshafer.com. 23
- [105] Glenn Shafer. Rejoinder to comments on “Perspectives on the theory and practice of belief functions”. *International Journal of Approximate Reasoning*, 6:445–480, 1992. 25
- [106] Glenn Shafer. What is probability? In David C. Hoaglin and David S. Moore, editors, *Perspectives on Contemporary Statistics*, MAA Notes 21, pages 93–105. Mathematical Association of America, 1992. 18
- [107] Glenn Shafer. Can the various meanings of probability be reconciled? In Gideon Keren and Charles Lewis, editors, *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues*, pages 165–196. Lawrence Erlbaum, Hillsdale, NJ, 1993. 18
- [108] Glenn Shafer. The subjective aspect of probability. In George Wright and Peter Ayton, editors, *Subjective Probability*, pages 53–73. Wiley, New York, 1994. 27
- [109] Glenn Shafer. *The Art of Causal Conjecture*. MIT Press, Cambridge, MA, 1996. 27
- [110] Glenn Shafer. *Probabilistic Expert Systems*. SIAM, Philadelphia, 1996. 26
- [111] Glenn Shafer. Causality and responsibility. *Cardozo Law Review*, 22:101–123, 2001. 27
- [112] Glenn Shafer. From Cournot’s principle to market efficiency. In Jean-Philippe Touffut, editor, *Augustin Cournot: Modelling Economics*, pages 55–95. Edward Elgar, Cheltenham, 2007. 30
- [113] Glenn Shafer. The education of Jean André Ville. *Electronic Journal for History of Probability and Statistics*, 5(1), 2009. jehps.net. 15
- [114] Glenn Shafer. A betting interpretation for probabilities and Dempster-Shafer degrees of belief. *International Journal of Approximate Reasoning*, 52:127–136, 2011. 30

- [115] Glenn Shafer, Peter R. Gillett, and Richard B. Scherl. Probabilistic logic. *Annals of Mathematics and Artificial Intelligence*, 2000. 27
- [116] Glenn Shafer and Roger Logan. Implementing Dempster’s rule for hierarchical evidence. *Artificial Intelligence*, 33:271–298, 1987. Reprinted as Chapter 19 of [144]. 22
- [117] Glenn Shafer and Judea Pearl. *Readings in Uncertain Reasoning*. Morgan Kaufmann, San Mateo, 1990. 25
- [118] Glenn Shafer, Prakash Shenoy, and Khaled Mellouli. Propagating belief functions in qualitative Markov trees. *International Journal of Approximate Reasoning*, 1:349–400, 1987. 23
- [119] Glenn Shafer and Rajendra Srivastava. The Bayesian and belief-function formalisms: A general perspective for auditing (with discussion). *Auditing: A Journal of Practice and Theory*, 9 Supplement:110–148, 1990. 23
- [120] Glenn Shafer and Amos Tversky. Languages and designs for probability judgment. *Cognitive Science*, 9:309–339, 1985. Reprinted on pages 237–265 of [6] and as Chapter 13 of [144]. 13
- [121] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It’s Only a Game!* Wiley, New York, 2001. 28
- [122] Glenn Shafer and Vladimir Vovk. The origins and legacy of Kolmogorov’s *Grundbegriffe*, 2004. Working Paper 4, <http://www.probabilityandfinance.com>. 17, 30
- [123] Glenn Shafer and Vladimir Vovk. The sources of Kolmogorov’s *Grundbegriffe*. *Statistical Science*, 21:70–98, 2006. 17
- [124] Prakash Shenoy. A valuation-based language for expert systems. *International Journal of Approximate Reasoning*, 3(5):383–411, 1989. 23
- [125] Prakash Shenoy and Glenn Shafer. Propagating belief functions with local computations. *IEEE Expert*, 1(3):43–52, 1986. 23
- [126] Prakash Shenoy and Glenn Shafer. Axioms for probability and belief-function propagation. In Ross D. Shachter et al., editor, *Uncertainty in Artificial Intelligence 4*, pages 169–198. North-Holland, Amsterdam, 1990. Reprinted as Chapter 20 of [144]. 23
- [127] Philippe Smets. Decision making in the TBM: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38:133–147, 2005. 15
- [128] Philippe Smets and Robert Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994. Reprinted as Chapter 28 of [144]. 24

- [129] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. Springer, New York, 1993. Lecture Notes in Statistics No. 81. 27
- [130] Rajendra P. Srivastava and Glenn Shafer. Belief-function formulas for audit risk. *Accounting Review*, 67:249–283, 1992. Reprinted as Chapter 23 of [144]. 23
- [131] Stephen M. Stigler. *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard University Press, Cambridge, MA, 1986. 16
- [132] Tom Strat. Continuous belief functions for evidential reasoning. In Ronald Brachman, editor, *AAAI-84: The Fourth National Conference on Artificial Intelligence*, pages 308–313, 1984. 25
- [133] Patrick Suppes. *A Probabilistic Theory of Causality*. North-Holland, Amsterdam, 1970. 26
- [134] Gheorghe Tecuci, Dorin Marcu, Mihai Boicu, and David A. Schum. *Knowledge Engineering: Building Cognitive Assistants for Evidence-Based Reasoning*. Cambridge, New York, 2016. 21
- [135] Gheorghe Tecuci, David A. Schum, Dorin Marcu, and Mihai Boicu. *Intelligence Analysis as Discovery of Evidence, Hypotheses, and Arguments: Connecting the Dots*. Cambridge, New York, 2016. 21
- [136] Isaac Todhunter. *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace*. Macmillan, London, 1865. 3
- [137] John W. Tukey. Handouts for the Wald lectures 1958. In *The Collected Works of John W. Tukey, Volume VI: More Mathematical, 1938–1984*, pages 119–148. Wadworth, Pacific Grove, California, 1990. 4
- [138] Marilyn vos Savant. *The Power of Logical Thinking*. St. Martin’s Press, New York, 1996. 17
- [139] Vladimir G. Vovk. A logic of probability, with applications to the foundations of statistics (with discussion). *Journal of the Royal Statistical Society. Series B*, 55(2):317–351, 1993. 28
- [140] Peter P. Wakker. Jaffray’s ideas on ambiguity. *Theory and Decision*, 71:11–22, 2011. 15
- [141] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991. 13
- [142] Peter M. Williams. On a new theory of epistemic probability (review of [84]). *The British Journal for the Philosophy of Science*, 29:375–387, 1978. 13

- [143] Roland R. Yager. Decision making under Dempster-Shafer uncertainties. *International Journal of General Systems*, 20:233–245, 1992. Reprinted as Chapter 24 of [144]. 15
- [144] Roland R. Yager and Liping Liu, editors. *Classic Works of the Dempster-Shafer Theory of Belief Functions*. Springer, Berlin, 2008. 32, 36, 37, 38, 39, 41, 42, 43
- [145] Sandy Zabell. *Symmetry and Its Discontents: Essays on the History of Inductive Probability*. Cambridge, New York, 2005. 6
- [146] Nevin L. Zhang. Weights of evidence and internal conflict for support functions. *Information Science*, 38:205–212, 1986. Reprinted as Chapter 15 of [144]. 10