# Entropy in Probability Kinematics

Stefan Lukits

## Contents

## 1 Abstract

If an agent's belief state is representable by a probability distribution (or density), then there may be guidelines or norms for changing it in the light of new evidence. Richard Jeffrey calls the investigation of these guidelines or norms 'probability kinematics' (see Jeffrey, 1965). Standard conditioning is the best-known (though not uncontested) updating procedure: the | operator (probabilistic conditioning) provides a posterior probability distribution (or density) which obeys both basic probability axioms and intuitions about desired properties of updated probabilities.

Although almost all the claims, examples, and formal methods presented in the following pertain to discrete probability distributions, there are usually equivalent things to be said for continuous probability densities, sometimes (as is the case for the Shannon entropy) requiring some mathematical manoeuvering, sometimes (more commonly) requiring no more than substituting an integral sign for a sum sign. I will exclusively refer to probability distributions, but many of the claims can be extended to probability densities.

Bayesian probability kinematics (1) requires a prior probability distribution (or, in softened versions, a more imprecise credal state) to precede any meaningful evaluation of evidence and (2) considers standard conditioning to be mandatory for a rational agent when she forms her posterior probabilities, just in case her evidence is expressible in a form which makes standard conditioning an option. There are various situations, such as the *Judy Benjamin* problem or the *Brandeis Dice* problem, in which standard conditioning does not appear to be an option (although some make the case that it is, so this point needs to be established independently), and therefore the question arises whether there is room for a more general updating method, whose justification will entail a justification of standard conditioning, but not be entailed by it.

E.T. Jaynes has suggested a unified updating procedure which generalizes standard conditioning called the principle of maximum entropy, or PME for short. The PME additionally to Bayesian probability kinematics (3) uses information theory to develop updating methods which keep the entropy of probability distributions high (in the synchronic case) and their cross-entropy low (in the diachronic case). The PME is based on a subjective interpretation of probabilities as representing the degree of uncertainty, or the lack of information, of the agent holding these probabilities. In this interpre-

tation, probabilities do not represent frequencies or objective probabilities, although there are models of how probabilities relate to them.

Most Bayesians reject the notion that the PME enjoys the same level of justification as standard conditioning and maintain that there are cases in which it delivers results which a rational agent should not, or is not required to, accept. Note that we distinguish between separate problems within the Bayesian camp: on the one hand, there may be objective methods of determining probabilities prior to any evidence or observation (call these 'absolutely' prior probabilities), for example from some type of principle of indifference; on the other hand, there may be objective methods of determining posterior probabilities from given prior probabilities (which themselves could be posterior probabilities in a previous instance of updating, call these 'relatively' prior probabilities) in case standard conditioning does not apply. My work is concerned with the latter problem, although the PME can also be used to defend objectivism about the former problem. Absolutely prior probabilities, however, have little relation to relatively prior probabilities.

Formal epistemologists widely concur that the PME is beset with too many conceptual problems and counterexamples to yield a generally valid objective updating procedure. Against the tide of this agreement, my work establishes that considering the tests that have been applied to it the PME as the only candidate for a generally valid, objective updating procedure is also a successful candidate. Both the conceptual problems and the counterexamples are surmountable, as we will show in great detail.

Many of the portrayals of the PME's failings are flawed and motivated by a desire to demonstrate that the labour of the epistemologist in interpreting probability kinematics on a case-by-case basis is indispensable. This 'full employment theorem' of probability kinematics (which has a formally proven equivalent in computer science) is widely promulgated by textbooks and passed on to students as expert consensus. By contrast, the PME combines a powerful and simple idea (update your probabilities in accordance with constraints revealed by the evidence without gaining more information than necessary) with a sophisticated formal theory which confirms that the powerful and simple idea consistently works.

Although the role of the PME in probability kinematics as a whole will be the scope of my work, I will pay particular attention to a problem which has stymied its acceptance by epistemologists: updating or conditioning on conditionals. Two counterexamples, Bas van Fraasen's *Judy Benjamin* and

Carl Wagner's *Linguist* problem, are specifically based on updating given an observation expressed as a conditional and on the PME's alleged failure to update on such observations in keeping with strong intuitions.

The PME also faces much of its conceptual criticism on the same issue with respect to 'epistemic entrenchment.' Epistemic entrenchment updates on conditionals by assuming a second tier of commitment to propositions beneath the primary tier of quantitative degrees of uncertainty of belief such as probabilities (or ranks). For example, if I am confident that a coin is fair my epistemic entrenchment that the probability of heads on the next toss is 1/2 is much more pronounced (and resilient to countervailing evidence) than the epistemic entrenchment in the same belief if I have no information or confidence about the bias of the coin. The PME conceptualizes this second tier differently and is consequently at odds with the voluminous recent literature on epistemic entrenchment. A large part of my task is to address and defend the PME's performance with respect to conditionals, both conceptually and with a view to threatening counterexamples.

## 2 Literature Review

### 2.1 Inductive Logic

David Hume posed one of the fundamental questions for the philosophy of science, the problem of induction. There is no deductive justification that induction works, as the observations which serve as a basis for inductive inference are not sufficient to make an argument for the inductive conclusion deductively valid. An inductive justification would beg the question. The late 19th and the 20th century brought us two responses to the problem of induction relevant to our project: (i) Bayesian epistemology and the subjective interpretation of probability focused attention on uncertainty and beliefs of agents, rather than measuring frequencies or hypothesizing about objective probabilities in the world, and on decision problems (John Maynard Keynes, Harold Jeffreys, Bruno de Finetti, Frank Ramsey; against, for example, R.A. Fisher and Karl Popper). (ii) Philosophers of science argued that some difficult-to-nail-down principle (indifference, simplicity, laziness, symmetry, entropy) justified entertaining certain hypotheses more seriously than others, even though more than one of them may be compatible with experience (Ernst Mach, Rudolf Carnap).

Two pioneers of Bayesian epistemology and subjectivism were Harold Jeffreys (see Jeffreys, 1931, and Jeffreys, 1939) and Bruno de Finetti (see de Finetti, 1931 and 1937). They personified a divide in the camp of subjectivists about probabilities. While de Finetti insisted that any probability distribution could be rational for an agent to hold as long as it obeyed the axioms of probability theory, Jeffreys considered probability theory to be an inductive logic with rules, resembling the rules of deductive logic, about the choice of prior and posterior probabilities. While both agreed on subjectivism in the sense that probabilities reflect an agent's uncertainty (or, in Jeffreys' case, more properly a lack of information), they disagreed on the subjectivist versus objectivist interpretation of how these probabilities are chosen by a rational agent (or, in Jeffreys' case, more properly by the rules of an inductive logic—as even maximally rational agents may not be able to implement them). The logical interpretation of probabilities began with John Maynard Keynes (see Keynes, 1921), but soon turned into a fringe position with Harold Jeffreys (for example Jeffreys, 1931) and E.T. Jaynes (for example Jaynes and Bretthorst, 2003) as advocates who were standardly invoked for refutation.

The problem in part was that the logical interpretation could not get off the ground with plausible rules about how to choose absolutely prior probabilities. No one was able to overcome the problem of transformation invariance for the principle of indifference (consider Bertrand's paradox, see Paris, 2006, 71f), not even E.T. Jaynes (for his attempts see Jaynes, 1973; for a critical response see Howson and Urbach, 2006, 285 and Gillies, 2000, 48).

One especially intractable problem for the principle of indifference was Ludwig von Mises' water/wine paradox. According to van Fraassen, it showed why we should "regard it as clearly settled now that probability is not uniquely assignable on the basis of a principle of indifference" (van Fraassen, 1989, 292). Van Fraassen went on to claim that the paradox signals the ultimate defeat of the principle of indifference, nullifying the Pyrrhic victory won by Poincaré and Jaynes in solving other Bertrand paradoxes (see Mikkelson, 2004, 137). Donald Gillies called von Mises' paradox a "severe, perhaps in itself fatal, blow" to Keynes' logical theory of probability (see Gillies, 2000, 43). De Finetti's subjectivism was an elegant solution to this problem and marginalized the logical theory.

While Jaynes threw up his hands over von Mises' paradox, despite the success he had landed addressing Bertrand's paradox (see Jaynes, 1973), Jeffrey Mikkelson recently suggested a promising solution to von Mises' paradox (see

Mikkelson, 2004). There may still be hope for an objectivist approach to absolutely prior probabilities. Nevertheless, my thesis remains agnostic about this problem. The domain of my project is probability kinematics. Relatively prior probabilities are assumed, and their priority only refers to the fact that they are prior to the posterior probabilities (one may call these distributions or densities antedata and postdata rather than prior and posterior, to avoid confusion) and not necessarily prior to earlier evidence.

This raises a conceptual problem: why would anybody be interested in a defence of objectivism in probability kinematics when the sense is that objectivism has failed about absolutely prior probabilities? My intuition is in line with Keynes, who maintains that all probabilities are conditional: "No proposition is in itself either probable or improbable, just as no place can be intrinsically distant; and the probability of the same statement varies with the evidence presented, which is, as it were, its origin of reference" (see Keynes, 1909, chapter 1).

The problem of absolutely prior probabilities is therefore moot, and it becomes clear that 'objectivism' is not really what we are advocating. Jaynes, who was initially interested in objectivism about absolutely prior probabilities as well, seemed to have come around to this position when in his last work *Probability Theory: The Logic of Science* he formally introduced probabilities as conditional probabilities (and later asserted that "one man's prior probability is another man's posterior probability," see Jaynes and Bretthorst, 2003, 89).

A logic of induction only provides rules for how to proceed from one probability distribution to another, given certain evidence. It does not claim that everybody should arrive at the same distribution (inasmuch as they claim to be rational), because there is no initial point at which two rational agents must agree. Just as in deductive logic, we may come to a tentative and voluntary agreement on a set of rules and presuppositions and then go part of the way together.

Another paradigm case for this kind of objectivity is Carnap's conventionalism in geometry. A subjectivist interpretation of probability, which both the more strictly subjectivist probability theorists (such as de Finetti) endorse as well as those who advocate the logical interpretation of probability—call both of these schools together the Bayesian school—sideline the frequentist's question about which probability/frequency corresponds to the real world as the conventionalist sidelines the metaphysician's question about which geometry corresponds to the real world. The logical interpretation goes further along with the conventionalist in attending to what is reasonable to believe given certain formal rules that we accept even if we have no Archimedean leverage point to start.

Probability kinematics rests on the idea that there are not only static norms about the probabilities of a rational agent, but also dynamic norms. The rational agent is not only constrained by the probability axioms, but also by standard conditioning as she adapts her probabilities to incoming evidence. Paul Teller gave a diachronic Dutch-book argument for standard conditioning (see Teller, 1973; Teller, 1976), akin to de Finetti's more widely accepted synchronic Dutch-book argument (for detractors of the synchronic Dutch-book argument see Seidenfeld et al., 1990; Foley, 1993, §4.4; and more recently Rowbottom, 2007; for a defence see Skyrms, 1987a). Brad Armendt expanded Teller's argument for Jeffrey conditioning (see Armendt, 1980). In contrast to the synchronic argument, however, there was considerable opposition to the diachronic Dutch-book argument (see Hacking, 1967; Levi, 1987; and Maher, 1992). Colin Howson and Peter Urbach even made the argument that standard conditioning itself as a diachronic norm of updating was inconsistent (see Howson and Urbach, 2006, 81f).

An alternate route to justify subjective degrees of belief and their probabilistic nature is to use Ramsey's approach of providing a representation theorem. Representation theorems make rationality assumptions for preferences such as transitivity (the standard reference work is still Sen, 1971) and derive from them a probability and a utility function which are unique up to acceptable transformations. Ramsey only furnished a sketch of how this could be done. The first fully formed representation theorem was given by Leonard Savage (see Savage, 1954); but soon Jeffrey noted that its assumptions were too strong. Based on mathematical work by Ethan Bolker (see Bolker, 1966; and a summary for philosophers in Bolker, 1967), Jeffrey provided a representation theorem with more acceptable assumptions (in Jeffrey, 1978). Since then, representation theorems have proliferated (there is, for example, a representation theorem for an acceptance-based belief function in Maher, 1993, and one for decision theory in Joyce, 1999). They are formally more complex than Dutch-book arguments, but well worth the effort because they make less controversial assumptions.

In summary, despite the success among epistemologists of de Finetti's more strictly subjectivist viewpoint, which is suspicious towards the claims of objectivity on part of the logical interpretation (with which, confusingly, de Finetti shares an overall subjectivist interpretation of probability as a measure of an agent's uncertainty, lack of information, or partial belief), the logical interpretation still commands intuitive appeal, internal consistency, and formal substance.

## 2.2 Information Theory and the Principle of Maximum Entropy

When Jaynes introduced the PME (see Jaynes, 1957a; 1957b), he was less indebted to the philosophy of science project of giving an account of semantic information (as in Carnap and Bar-Hillel, 1952; 1953) than to Claude Shannon's mathematical theory of information and communication. Shannon identified information entropy with a numerical measure of a probability distribution fulfilling certain requirements (for example, that the measure is additive over independent sources of uncertainty). The focus is not on what information is but how we can formalize an axiomatized measure. Entropy stands for the uncertainty that is still contained in information (certainty is characterized by zero entropy).

Shannon introduced information entropy in 1948 (see Shannon, 2001), based on work done by Norbert Wiener connecting probability theory to information theory (see Wiener, 1939). Jaynes also traced his work back to Ludwig Boltzmann and Josiah Gibbs, who built the mathematical foundation of information entropy by investigating entropy in statistical mechanics (see Boltzmann, 1877; Gibbs, 1902).

For the further development of the PME in probability kinematics it is important to refer to the work of Richard Jeffrey, who established the discipline (see Jeffrey, 1965), and Solomon Kullback, who provided the mathematical foundations of minimum cross-entropy (see Kullback, 1959). In probability kinematics, contrasted with standard conditioning, evidence is uncertain (for example, the ball drawn from an urn may have been observed only briefly and under poor lighting conditions).

Jeffrey addressed many of the conceptual problems attending probability kinematics by providing a much improved representation theorem, thereby creating a tight connection between preference theory and its relatively plausible axiomatic foundation and a probabilistic view of 'beliefs.' Jeffrey and Isaac Levi's (see Levi, 1967) debates on partial belief and acceptance, which Jeffrey considered to be as opposed to each other as Dracula and Wolfman (see Jeffrey, 1970), set the stage for two 'epistemological dimensions' (Henry Kyburg's term, see Kyburg, 1995, 343), which will occupy us in detail and towards which I will take a more conciliatory approach, as far as their opposition or mutual exclusion is concerned.

Kullback's divergence relationship between probability distributions made possible a smooth transition from synchronic arguments about absolutely prior probabilities to diachronic argument about probability kinematics (this transition was much more troublesome from the synchronic Dutch-book argument to the diachronic Dutch-book argument; for the information-theoretic virtues of the Kullback-Leibler divergence see Kullback and Leibler, 1951; Seidenfeld, 1986, 262ff; Guiaşu, 1977, 308ff).

Jaynes' project of probability as a logic of science was orginally conceived to provide objective absolutely prior probabilities by using the PME, rather than to provide objective posterior probabilities, given relatively prior probabilities. It was, however, easy to turn the PME into a method of probability kinematics using the Kullback-Leibler divergence. Jaynes presented this method in 1978 at an MIT conference under the title "Where Do We Stand on Maximum Entropy?" (see Jaynes, 1978), where he explained the *Brandeis* problem and demonstrated the use of Lagrange multipliers in probability kinematics.

Ariel Caticha and Adom Giffin recently demonstrated, using Lagrange multipliers, that the PME seamlessly generalizes standard conditioning (see Caticha and Giffin, 2006). Many others, however, thought that in one way or another the PME was inconsistent with standard conditioning, to the detriment of the PME (see Seidenfeld, 1979, 432f; Shimony, 1985; van Fraassen, 1993, 288ff; Uffink, 1995, 14; and Howson and Urbach, 2006, 278); Jon Williamson believed so, too, but to the detriment of standard conditioning (see Williamson, 2011).

Arnold Zellner, however, proved that standard conditioning as a diachronic updating rule (Bayes' theorem) is the "optimal information processing rule" (Zellner, 1988, 278), also using Lagrange multipliers. Standard conditioning is neither inefficient (using a suitable information metric), diminishing the output information compared to the input information, nor does it add extraneous information. This is just the simple conceptual idea behind the PME, although the PME only requires optimality, not full efficiency. Full efficiency implies optimality, therefore standard conditioning fulfills the PME.

Once the PME was formally well-defined and its scope established (for the latter, Imre Csiszár's work on affine constraints was important, see Csiszár, 1967), its virtues came to the foreground. While Richard Cox (see Cox, 1946) and E.T. Jaynes defended the idea of probability as a formal system of logic, John Shore and Rodney Johnson provided the necessary detail to establish the uniqueness of the PME in meeting intuitively compelling axioms (see Shore and Johnson, 1980).

## 2.3 Early Criticism

By then, however, an avalanche of criticism against the PME as an objective updating method had been launched. Papers by Abner Shimony (see Friedman and Shimony, 1971; Dias and Shimony, 1981; Shimony, 1993) convinced Brian Skyrms that the PME and its objectivism were not tenable (see Skyrms, 1985, 1986, and 1987b). Bas van Fraassen's *Judy Benjamin* problem (see van Fraassen, 1981) dealt another blow to the PME in the literature, motivating Joseph Halpern (who already had reservations against Cox's theorem, see Halpern, 1999) to reject it in his textbook on uncertainty (see Halpern, 2003).

Teddy Seidenfeld ran his own campaign against objective updating methods in articles such as "Why I Am Not an Objective Bayesian" (see Seidenfeld, 1979; 1986), while Jos Uffink took issue with Shore and Johnson, casting doubt on the uniqueness claims of the PME (see Uffink, 1995; 1996). Carl Wagner introduced a counterexample to the PME (see Wagner, 1992), again, as in the *Judy Benjamin* counterexample (but in much greater generality), involving conditioning on conditionals.

## 2.4 Late Criticism

In 2003, Halpern renewed his attack against the PME with the help of Peter Grünwald and the concept of 'coarsening at random,' which according to the authors demonstrated that the PME "essentially never gives the right results" (see Grünwald and Halpern, 2003, 243).

In 2009, Igor Douven and Jan-Willem Romeijn wrote an article on the *Judy Benjamin* problem (see Douven and Romeijn, 2009) in which they asked probing questions about the compatibility of objective updating methods with epistemic entrenchment.

Malcolm Forster and Elliott Sober's attack on Bayesian epistemology using

Akaike's Information Criterion was articulated in the 1990s (see Forster and Sober, 1994) but reverberated well into the next decade (Howson and Urbach call the authors the 'scourges of Bayesianism'). Because the attack concerns Bayesian methodology as a whole, it is not within our purview to defend the PME against it (for a defence of Bayesianism see Howson and Urbach, 2006, 292ff), but it deserves mention for its direct reference to information as a criterion for inference and provides an interesting point of comparison for maximum entropy.

Another criticism which affected both the weaker Bayesian claim for standard conditioning and the stronger PME was its purported excessive apriorism, i.e. the concern that the agent can never really move away from beliefs once formed—and that those beliefs always need to be fully formed all the time! It can be found as early as 1945 in Carl Hempel (see Hempel and Oppenheim, 1945, 107) and is vigorously raised again as late as 2005 by James Joyce (see Joyce, 2005, 170f). In *Bayes or Bust?,* excessive apriorism (among other things) led John Earman to his famous position of being a Bayesian only on Mondays, Wednesdays, and Fridays (see Earman, 1992, 1; for the detailed criticism Earman, 1992, 139f).

Gillies had similar reservations (see Gillies, 2000, 81; 84 and also for a pertinent quote by de Finetti see Gillies, 2000, 57), while Seidenfeld militated against objective Bayesianism in 1979 using excessive apriorism (we owe the term to him, see Seidenfeld, 1979, 414). Again, because the charge was directed at Bayesians more generally, we do not need to address it, but mention it because the PME may have resources at its disposal that the more general Bayesian position lacks (for this position, see Williamson, 2011).

## 2.5 Acceptance versus Probabilistic Belief

Epistemic entrenchment figures prominently in the AGM literature on belief revision (for one of its founding documents see Alchourrón et al., 1985) and is based on two levels of uncertainty about a proposition: its static inclusion in belief sets on the one hand, and its dynamic behaviour under belief revision on the other hand. It is one thing, for example, to think that the probability of a coin landing heads is 1/2 and consider it fair because you have observed one hundred tosses of it, or to think that the probability of a coin landing heads is 1/2 because you know nothing about it. In the former scenario, your belief that $P(X = H) = 0.5$ is more entrenched.

Wolfgang Spohn provided an excellent overview of the interplay between Bayesian probability theory, AGM belief revision, and ranking functions (see Spohn, 2012). The extent to which the PME is compatible with epistemic entrenchment and a distinction between the static and the dynamic level will be a major topic of my investigation. At first glance, the PME and epistemic entrenchment are at odds, because the PME operates without recourse to a second epistemic layer behind probabilities expressing uncertainty. Our conclusion will be that the content of this layer is expressible in terms of evidence and is not epistemic.

For a long time, there was unease between defenders of partial belief (such as Richard Jeffrey) and defenders of full (and usually defeasible or fallible) belief (such as Isaac Levi). This issue was viewed more pragmatically beginning in the 1990s with Patrick Maher's *Betting on Theories* (see Maher, 1993) and Wolfgang Spohn's work in several articles (later summarized in Spohn, 2012). Both authors sought to downplay the contradictory nature of these two approaches and emphasized how both were necessary and able to inform each other.

Maher argued that representation theorems were superior to Dutch-book arguments in justifying Bayesian methodology, but then distinguished between practical utility and cognitive utility. Whereas probabilism is appropriate in the arena of acting, based on practical utility, acceptance is appropriate in the arena of asserting, based on cognitive utility. Maher then provided his own representation theorem with respect to cognitive utility, underlining the resemblance in structure between the probabilistic and the acceptance-based approach.

In a similar vein, Spohn demonstrated the structural similarities between the two approaches using ranking theory for the acceptance-based approach. Together with the formal methods of the AGM paradigm, ranking theory delivered results that were astonishingly analogous to the already well-formulated results of Bayesian epistemology. Maher and Spohn put us on the right track of reconciliation between the two epistemological dimensions, and I hope to contribute to it by showing that the PME can be coherently coordinated with this reconciliation. This will only be possible if we clarify the relation that the PME has to epistemic entrenchment and how it conditions on conditionals, because on a surface level the two are difficult to accommodate to each other.

## 3 Proposal

My thesis is that the principle of maximum entropy (PME) is defensible against all counterexamples and conceptual issues raised so far as a generally valid objective updating method in probability kinematics. Subjectivists need to work harder to undermine the validity of the PME, and a fortiori the validity of objectivism.

The PME operates on the basis of an astonishingly simple principle: when updating your probabilities, waste no useful information and do not gain information unless the evidence compels you to gain it (see Jaynes, 1988, 280, Van Fraassen et al., 1986, 376, and Zellner, 1988, 278). The astonishingly simple principle comes with its own formal apparatus (not unlike probability theory itself): Shannon's information entropy, the Kullback-Leibler divergence, the use of Lagrange multipliers, and the sometimes intricate, sometimes straightforward relationship between information and probability.

My interpretation of the PME is an intermediate position between what I would call Jaynes' Laplacean idealism, where evidence logically prescribes unique and determinate probability distributions to be held by rational agents; and a softened version of Bayesianism exemplified by, for example, Richard Jeffrey and James Joyce (for the latter see Joyce, 2005).

I side with Jaynes in so far as I am committed to determinate prior probabilities, whether they are absolute or relative. Once a rational agent considers a well-defined event space, the agent is able to assign fixed numerical probabilities to it (this ability is logical, not practical—in practice, the assignment may not be computationally feasible). Because I consider this process to be contingent on previous commitments (there is no objectivity in choosing absolutely prior probabilities, if there is such a thing as an absolutely prior probability) and the interpretation of evidence, both of which introduce elements of subjectivity, I part ways with Jaynes about objectivity.

I part ways with the 'humanly faced' Bayesians because of my commitment to determinate probabilities. 'Humanly faced' Bayesians claim that rational agents typically lack determinate subjective probabilities and that their opinions are characterized by imprecise credal states in response to unspecific and equivocal evidence. There is a difference, however, between (1) appreciating the imprecision in interpreting observations and, in the context of probability updating, casting them into appropriate mathematical constraints for updated probability distributions, and (2) bringing to bear formal methods to probability updating which require numerically precise priors. The same is true for using calculus with imprecise measurements. The inevitable imprecision in our measurements does not attenuate the logic of using real analysis to come to conclusions about the volume of a barrel or the area of a circular flower bed. My project will seek to articulate these distinctions more explicitly.

One particularly strong advocate of imprecise credal states is James Joyce (see Joyce, 2005, 156f), with the unfortunate consequence that the updating strategies that Joyce proposes for these credal states are impotent. No amount of evidence can modify the imprecise credal state, because each member of the set of credal states that an agent accepts has a successor with respect to updating that is also a member of these credal states and that is consistent with its predecessor and the evidence. Although the feeling is that the imprecise credal state is narrowed by evidence towards more precision, set theory indicates that the credal state remains static, no matter what the evidence is, unless we introduce a higher-level distribution over these sets—but then the same problems arise on the higher level.

My project therefore promotes what I would call Laplacean realism, contrasting it with Jaynes' Laplacean idealism (Sandy Zabell uses the less complimentary term "right-wing totalitarianism" for Jaynes' position, see Zabell, 2005, 28), but also distinguishing it from contemporary softened versions of Bayesianism such as Joyce's or Jeffrey's (Zabell's corresponding term is "left-wing dadaists," although he does not apply it to Bayesians). What is distinctive about my approach to Bayesianism is the high value I assign to the role that information theory plays within it. My contention is that information theory, much like Jaynes intended it, provides a logic for belief revision. Almost all epistemologists, who are Bayesians, currently have severe doubts that information theory can deliver on this promise, not to mention their doubts about the logical nature of belief revision (see for example Zabell's repeated charge that advocates of logical probability have never successfully addressed Ramsey's "simple criticism" about how to apply observations to the logical relations of probabilities, see for example Zabell, 2005, 25).

One way in which these doubts can be addressed is by referring them to the more general debate about the relationship between mathematics and the world. The relationship between probabilities and the events to which they are assigned is not unlike the relationship between the real numbers we assign to the things we measure and calculate and their properties in the physical world. As unsatisfying as our understanding of the relationship between formal apparatus and physical reality may be, the power, elegance, and internal consistency of the formal methods is rarely in dispute. Information theory is one such apparatus, probability theory is another. In contemporary epistemology, their relationship is held to be at best informative of each other. Whenever there are conceptual problems or counterintuitive examples, the two come apart. I consider the relationship to be more substantial than currently assumed.

There have been promising and mathematically sophisticated attempts to define probability theory in terms of information theory (see for example Ingarden and Urbanik, 1962; Kolmogorov, 1968; Kampé de Fériet and Forte, 1967—for a detractor who calls information theory a "chapter of the general theory of probability" see Khinchin, 1957). While interesting, however, making information theory or probability theory derivative of the other is not my project. What is at the core of my project is the idea that information theory delivers the unique and across the board successful candidate for an objective updating mechanism in probability kinematics. This idea is unpopular in the literature, but as evidenced in the chapter outline the arguments on which the literature relies are not robust, neither in quantity nor in quality.

Carnap advises pragmatic flexibility with respect to inductive methods, although he presents only a one-dimensional parameter system of inductive methods which curbs the flexibility. On the one hand, Dias and Shimony report that Carnap's $\lambda$-continuum of inductive methods is consistent with the PME only if $\lambda = \infty$ (see Dias and Shimony, 1981). This choice of inductive method is unacceptable even to Carnap, albeit allowed by the parameter system, because it gives no weight to experience (see Carnap, 1952, 37ff). On the other hand, Jaynes makes the case that the PME entails Laplace's Rule of Succession ($\lambda = 2$) and thus occupies a comfortable middle position between giving all weight to experience ($\lambda = 0$, for the problems of this position see Carnap, 1952, 40ff) or none at all ($\lambda = \infty$). While Carnap's parameter system of inductive methods rests on problematic assumptions, we will show why Dias and Shimony's assignment of $\lambda$, given the PME, is erroneous, and why Jaynes' assignment is better justified.

Jeffrey, van Fraassen, Halpern, Skyrms, Persi Diaconis and Sandy Zabell (see Diaconis and Zabell, 1982), Colin Howson and Allan Franklin (see Howson and Franklin, 1994), Douven and Romeijn, however, all follow Carnap in giving weight not only to a logical and an empirical component in induction, but also to pragmatic considerations. The pragmatic considerations often turn into a requirement that an expert epistemologist study each situation calling for belief revision on its own and address it with recourse to a toolkit of updating procedures, each being context-appropriate under different circumstances.

In order to defend my position, I need to address three important counterexamples which at first glance discredit the PME: Shimony's Lagrange multiplier problem, van Fraassen's *Judy Benjamin* case, and Wagner's *Linguist*. For the latter two, I am confident that we can make a persuasive case for the PME, based on formal features of these problems which favour the PME upon closer examination. For the former (Shimony), Jaynes has written a spirited rebuttal: "This brings us, obviously, to the matter of Shimony. I am not a participant, but, like other readers, only a bewildered onlooker in the spectacle of his epic struggles with himself" (Jaynes, 1985, 134). According to Jaynes, errors in Shimony's argument have been pointed out five times (see Hobson, 1972; Tribus and Motroni, 1972; Gage and Hestenes, 1973; Jaynes, 1978; Cyranski, 1979). This does not, however, keep Brian Skyrms, Jos Uffink, and Teddy Seidenfeld from referring again to Shimony's argument in rejecting the PME in the 1980s, so the matter must be sorted out, and we promise to do so.

Once the counterexamples are out of the way, the more serious conceptual issues loom. There is good news and bad news for advocates of the PME. On the one hand, there are powerful conceptual arguments affirming the special status of the PME. Shore and Johnston, who use the axiomatic strategy of Cox's theorem in probability kinematics, show that relatively intuitive axioms only leave us with the PME to the exclusion of all other objective updating methods. Van Fraassen, R.I.G. Hughes, and Gilbert Harman's MUD method, for example, or their maximum transition probability method from quantum mechanics both fulfill their five requirements (see van Fraassen et al., 1986), but do not fulfill Shore and Johnston's axioms. Neither does Uffink's more general class of inference rules, which maximize the so-called Rényi entropies, but Uffink argues that Shore and Johnston's axioms rest on unreasonably strong assumptions (see Uffink, 1995). Caticha and Giffin counter that Skilling's method of induction (see Skilling, 1988) and Jaynes' empirical results in statistical mechanics and thermodynamics imply the uniqueness of Shannon's information entropy over rival entropies.

To continue with the good news, the PME seamlessly generalizes standard conditioning and Jeffrey's rule where they are applicable (see Caticha and Giffin, 2006). It underlies the entropy concentration phenomenon described in Jaynes' standard work *Probability Theory: the Logic of Science*, which also contains a sustained conceptual defence of the PME and its underlying logical interpretation of probabilities. Entropy concentration refers to the unique property of the PME solution to have other distributions which obey the affine constraint cluster around it. When used to make predictions whose quality is measured by a logarithmic score function, posterior probabilities provided by the PME result in minimax optimal decisions (see Topsøe, 1979; Walley, 1991; Grünwald, 2000) so that by a logarithmic scoring rule these posterior probabilities are in some sense optimal.

Jeff Paris has investigated different belief functions (probabilities, Dempster-Shafer, and truth-functional, see Paris, 2006) from a mathematical perspective and come to the conclusion that given certain assumptions about the constraints that experience normally imposes (we will have to examine their relationship to the affine constraints assumed by the PME), if a belief function is a probability function, only minimum cross entropy belief revision satisfies a host of desiderata (continuity, equivalence, irrelevant information, open-mindedness, renaming, obstinacy, relativization, and independence) while competitors fail on multiple counts (see Paris and Vencovská, 1990).

On the other hand, now turning to the bad news, the belief revision literature has in the last twenty years mostly turned its attention to the AGM paradigm (named after Carlos Alchourrón, Peter Gärdenfors, and David Makinson), which operates on the basis of fallible beliefs and their logical relationships. These are really at this point two different epistemic dimensions (to use Henry Kyburg's expression): the one where doxastic states are cashed out in terms of fallible beliefs which move in and out of belief sets; the other the dimension of probabilities where 'beliefs' are vague labels for a more deeply rooted, graded notion of uncertainty.

Jeffrey with his radical probabilism pursues a project of epistemological monism (see Jeffrey, 1965) which would reduce beliefs to probabilities, while Spohn and Maher seek reconciliation between the two dimensions, showing how fallible full beliefs are epistemologically necessary and how the formal structure of the two dimensions reveals many shared features so that in the end they have more in common than what separates them (see Spohn, 2012, 201 and Maher, 1993).

In the end, our project is not about the semantics of doxastic states. We do not argue the eliminativism of beliefs in favour of probabilities; on the contrary, the belief revision literature has opened an important door for inquiry in the Bayesian dimension with its concept of epistemic entrenchment. This is a good example for the kind of cross-fertilization between the two different dimensions that Spohn had in mind, mostly in terms of formal analogies and with little worry about semantics, important as they may be. Maher has given similar parallels between the two dimensions, also with an emphasis on formal relationships, in terms of representation theorems. Pioneering papers in probabilistic versions of epistemic entrenchment are recent (see Bradley, 2005; Douven and Romeijn, 2009).

The guiding idea behind epistemic entrenchment is that once an agent is apprised of a conditional (indicative or material), she has a choice of either adjusting her credence in the antecedent or the consequent (or both). Often, the credence in the antecedent remains constant and only the credence in the consequent is adjusted (Bradley calls this 'Adams conditioning'). Douven and Romeijn give an example where the opposite is plausible and only the credence in the consequent is left constant (see Douven and Romeijn, 2009, 12). Douven and Romeijn speculate that an agent could theoretically take any position in between, and they use Hellinger's distance to represent these intermediary positions formally (see Douven and Romeijn, 2009, 14).

Even though Bradley, Douven, and Romeijn are in the dimension of probabilities, they are using a notion frequently used and introduced by the AGM literature to capture formally analogous structures in probability theory. The question is how compatible the use of epistemic entrenchment in probabilistic belief revision (probability kinematics) is with the PME. The PME appears to assign probability distributions to events without any heed to epistemic entrenchment. The *Judy Benjamin* problem is a case in point. PME's posterior probabilities are somewhere in between the possible epistemic entrenchments, as though mediating between them, but they affix themselves to a determinate position (which in some quarters raises worries analogous to excessive apriorism).

My claim is that the PME does not accept two levels of epistemic commitment: the static and the dynamic. AGM belief revision theory, Spohn's ranking functions, and epistemic entrenchments according to Bradley, Douven, and Romeijn suppose that beneath our credences (static probabilities), believers entertain a second set of dynamic probabilities which are determinative of the kinematics once doxastic states are subject to change. This view is inconsistent with the PME, and so I hold against it that the PME can only understand this second dynamic set of graded commitments as information. In other words, entrenchments are evidential, not epistemic. To refer to the earlier example, the fact that a coin has been tossed a hundred times, so that now we consider it to be a fair coin (rather than assigning a 50:50 probability to heads and tails because we do not know any better), is information and part of our evidence; it is not part of the epistemic state of a rational agent, such as a belief or a probability function would be.

Despite the potholes in the historical development of the PME, on account of its unifying features, its simple and intuitive foundations, and its formal success it deserves more attention in the field of belief revision and probability kinematics, definitely more attention than the many competing ad hoc methods (such as Carl Wagner's) which patch one problem while raising many more somewhere else. The PME is the single principle which can hold things together over vast stretches of epistemological terrain (intuitions, formal consistency, axiomatization, case management) and calls into question the scholarly consensus that such a principle is not needed.

Epistemologists are generally agreed that such a principle is not needed because they expect pragmatic latitude in addressing questions of belief revision. Similar to how scientists subjectively choose absolutely prior probabilities, the full employment theorem gives epistemologists access to a wide array of updating methods. As there is already widespread consensus that objectivism has died on the operating table of absolutely prior probabilities, nobody sees any reason to prolong the dead patient's life on the sickbed of probability kinematics. The problem with this position is that the wide array of updating methods systematically leads to solutions which contradict the principle that one should use relevant information and not gain unwarranted information.

Often, independence assumptions sneak in through the back door where they are indefensible. Well-posed problems are dismissed as ambiguous (e.g. the *Judy Benjamin* problem), while problems that are over-determined may be treated as open to a whole host of solutions because they are deemed underdetermined (e.g. von Mises' water and wine paradox). Ad hoc updating methods proliferate which can often be subsumed under the PME with little mathematical effort (e.g. Wagner's *Linguist* problem).

Worst of all, epistemologists suffer from the mistaken perception that they will always be indispensable to the scientist's and the quotidien reasoner's quest for proper belief revision in the face of new evidence. I call this perception ‘full employment,’ so named after a theorem in computer science, where it can be formally proven that no computer program can write itself all necessary computer programs without the assistance of a computer scientist. Formal epistemologists would like to be in the same position with respect to probability updating. I maintain that information theory provides all the necessary tools to update probabilities effectively, universally, and consistently.

## 4 Chapter Outline

### 4.1 Introduction

The introduction provides a first framework for the problem of probability kinematics, underlines the philosophical relevance, and embeds the discussion in a wider epistemological context. It also introduces the problem of how we can identify evidence with affine constraints and presents the principle of maximum entropy both in its synchronic form (MAXENT) and in its diachronic form (*Infomin*). There is a view which holds MAXENT and *Infomin* to be inconsistent with each other, which needs to be addressed (see Wagner, 2002).

Information itself is notoriously difficult to define and comes in different varieties (Shannon information, Solomonoff complexity, quantum information, semantic information) that have not been successfully reduced to a single concept. We are primarily interested in Shannon information, i.e. the information associated with probability distributions or densities. It is an open question what the relationship between information theory and probability theory is, whether one is derivative of the other or whether they are independent accounts of uncertainty informing each other in interesting ways.

An affine constraint restricts evidence to subsets which are closed and convex in a suitable information topology of probability distributions or densities. For example, when standard conditioning is applicable (‘a die was rolled, and the result is an even number’) only those probability distributions which assign 1 to the observed event are eligible as posterior probabilities. I have not been able to find a systematic justification for the restrictions of affine constraints (the subset must be closed and convex), except that once they are in place they provide just the right formal assumptions to warrant the existence and uniqueness of a maximum entropy solution (Paris is mathematically most explicit, but also most unabashed about citing mathematical convenience as uppermost in accepting these restrictions, see Paris, 2006. Csiszár, 1967 is another place to consider and often referred to for moral support by advocates of the PME. It would constitute great progress if someone could bring more clarity to this question, although due to the mathematical challenges this may not be the task of a philosophy student (both Paris and Csiszár are mathematicians).

To my knowledge there are no counterexamples in which evidence is not an affine constraint, although it is probably possible to construct one. Paris, for example, refers to non-linear constraints of belief functions as too complicated to handle (undermining certain uniqueness claims, see Paris, 2006, 70) and in any case unlikely to be encountered in real-world problems (see Paris, 2006, 7). Affine constraints are equivalent to expectations provided by evidence (I have so far not been able to locate a formal proof for this result, but see Hobson, 1971), and those in turn cover all the classic examples of evidence: standard conditioning, Jeffrey conditioning, and partial information.

Affine constraints come in three forms: observation with certainty of an event (as above, 'a die was rolled, and the result is an even number'); complete re-partitioning of the event space ('a die was rolled, and the probability that the result is an even number rather than an odd number is 60:40'); and reassessment of expectation ('a die was rolled, and the expected value of the result is 4.5,' the *Brandeis* problem). This chapter shows how these forms are related. The third form is a generalization of the second form, and the second form is a generalization of the first form. Consequently, a method for solving the problem of probability kinematics for the third form automatically solves this problem for the other forms.

This chapter introduces standard conditioning and reviews the arguments why standard conditioning is widely accepted as a solution for affine constraints of the first form. This chapter introduces Jeffrey conditioning and reviews the arguments why Jeffrey conditioning is less widely accepted as a solution for affine constraints of the second form. The thesis of this work is conditional on the acceptance of Jeffrey conditioning as a solution to the problem of probability kinematics for the second form, although I will provide a brief overview of counterarguments. Then this chapter introduces affine constraints which are not covered by Jeffrey conditioning. Let us call these strictly affine constraints.

Strictly affine constraints exist, and they complete the list that we need to consider for the problem of probability kinematics. There is a formal basis for these two claims. My work provides counterarguments to the claim that Jeffrey conditioning and the PME can be subsumed under standard conditioning (for example in Domotor, 1985; Skyrms, 1985; or retrospective conditioning, summary and problems in Diaconis and Zabell, 1982, 822; applied to the *Judy Benjamin* problem see Grove and Halpern, 1997). Especially (but not exclusively) in statistical physics there are applications where standard conditioning is not an effective method, while Jeffrey conditioning and the PME are. As mentioned above, the claim that the three forms exhaust the list is formally challenging, and I am uncertain to what extent I can address the question in my project.

## 4.2 The Principle of Maximum Entropy: Virtues and Vices

This chapter introduces the PME and highlights its strengths and vulnerabilities. In terms of strengths, the PME provides a unique solution for all affine constraints and is as well-immunized against transformation variance as one can reasonably expect. Jaynes has identified the most important transformation groups (for example scale, translation, rotation, and location; for language invariance see Paris, 2006, 76), with respect to which the PME is transformation invariant (see Jaynes, 1973; and Jaynes and Bretthorst, 2003, 378). Howson and Urbach mention transformation groups (for example, arbitrary variations in space-time curvature and arbitrary coordinate transformations), with respect to which the PME is not transformation invariant (see Howson and Urbach, 2006, 285).

The PME solution uniquely fulfills several desiderata, and thus we are able to provide an axiomatic basis for its use (see Shore and Johnson, 1980; Tikochinsky et al., 1984; Skilling, 1988 for details; Uffink, 1996 for criticism). The PME accords with standard conditioning if the affine constraint is of the first form and with Jeffrey conditioning if the affine constraint is of the second form. While it is highly contested as an objective updating method for strictly affine constraints, it exhibits important virtues as such a method. It underlies the entropy concentration phenomenon.

A large majority of philosophers of science and epistemologists, however, rejects the idea that the PME is normative as an objective updating procedure given strictly affine constraints. Instead, they adhere to a type of 'full employment theorem,' according to which affine constraints must be submitted to the individualized attention of an expert before the problem can be considered properly investigated. I will call this camp 'opponents,' not because I necessarily believe that their claims are false, but because I believe that the arguments on which their claims rest are either faulty or incomplete.

They have not made a convincing case that the PME lacks generality. Their case rests on the vulnerabilities of the PME, which we will try to address in a comprehensive manner. This chapter portrays these vulnerabilities in their strongest form: the *Judy Benjamin* problem; the Shimony objection; the Seidenfeld objection; the Wagner objection; and Grünwald and Halpern's 'Coarsening at Random.' All of these objections derive counterintuitive results from the PME and claim that advocates of the PME find themselves in the awkward position of having to assent to something to which they would intuitively, and even after reflection, withhold assent.

### 4.2.1 Judy Benjamin

Opponents prominently cite van Fraassen's *Judy Benjamin* case to undermine the generality of the PME. This chapter shows that an intuitive approach to the *Judy Benjamin* case supports the PME. This is surprising because based on independence assumptions the anticipated result is that it would support the opponents. The chapter also demonstrates that opponents improperly apply independence assumptions to the problem. Not dissimilar to the (1/2)er camp and the (1/3)er camp in the *Sleeping Beauty* case, there is a (1/2)er camp and a (1/3)er camp in the *Judy Benjamin* case. This chapter gives rigorous arguments for the (1/3)er camp in the *Judy Benjamin* case (the analogy with Sleeping Beauty goes deeper than the labels, see Bovens, 2010).

### 4.2.2 The Shimony Objection

In a series of papers, Abner Shimony has highlighted cases in which the PME results in the counterintuitive claim that being informed of a Lagrange multiplier requires that one has expected it with certainty. This chapter introduces Shimony's objection and mounts a defence of PME against it. Attempts at a defence are available (see Hobson, 1972; Tribus and Motroni, 1972; Gage and Hestenes, 1973; Cyranski, 1979; Jaynes, 1985), but there is also strong support for Shimony's objection, coupled with various attempts at strengthening his results (see Seidenfeld, 1986; Skyrms, 1987b). I am not yet in a position to articulate a solution or a route towards one. It will probably require careful distinctions of what it looks like to gain information legitimately, especially with reference to using mathematical tools.

(I do not know if this is a helpful analogy, but it is alleged that the number $\pi$ is normal, i.e. there are no patterns in its digits, independent of the base. A formal proof that *pi* is normal is elusive. Consequently, its Shannon entropy is high, whereas its Kolmogorov complexity is low, as *pi* is expressible by a relatively simple formula. Therefore, a certain amount of care is needed in describing what it means to be informed that $x = \pi$. To be informed of the value of a Lagrange multiplier may reveal similar complications.)

One particular problem for the PME is that Shimony's argument identifies it with $\lambda = \infty$ in Carnap's parameter system of inductive logic. This is the case in which observation has no weight for induction, a state of affairs universally held to be undesirable. The game of deprecating an inductive method by showing that it implies $\lambda = \infty$ is played already by Carnap himself, his targets being C.S. Peirce, Ludwig Wittgenstein, and John Maynard Keynes (see Carnap, 1952, 40). Jaynes seeks to fend off the attack by showing that the PME entails $\lambda = 2$ (Laplace's Rule of Succession) and that Shimony's Lagrange multiplier argument is an instance of hysteron proteron.

### 4.2.3 The Seidenfeld Objection

Teddy Seidenfeld claims that the PME as a rule to update on partial information (an affine constraint of the third form) is unacceptable. It leads to precise probabilities that are excessively aprioristic, containing more information than the evidence generating them allows. This being a common objection to PME among philosophers, we will show how it cannot be coherently raised against the PME without being raised against Bayesian epistemology as a whole. This section also provides a brief overview of reasons why within appropriate contexts it is advisable to accept the normative claims of Bayesian epistemology. Therefore, the normative claims of the PME are immune to Seidenfeld's objection to the degree to which we have already accepted the normative claims of Bayesian epistemology.

Besides this more general objection, Teddy Seidenfeld develops several formal lines of argument against the PME, for example that under certain circumstances involving noise factors the PME will inappropriately provide more information based on less evidence. Another example involves an alleged incompatibility between Bayes' Theorem and the PME. Seidenfeld also explains and expands the Shimony objection in a new light. This section will look in detail at available lines of defence for the PME against Seidenfeld's objections.

### 4.2.4 The Wagner Objection

Carl Wagner is a typical representative of the full employment line of argument, who in his own words values the hard work of judging by warring against the temptations of mechanical updating (see Wagner, 1992, 255f). Other representatives of full employment with similar sentiments are Peter Grünwald, Joseph Halpern, Richard Bradley, Persi Diaconis, and Sandy Zabell. E.T. Jaynes, the most prominent proponent of the PME, sometimes speaks derisively about the relationship of the statistician (or formal epistemologist) and the client as one between a doctor and his patient, which I have characterized as full employment in allusion to the full employment theorem in computer science.

Wagner's formal attack against the PME concerns constraints that are a generalization of Jeffrey conditioning, for which Wagner suggests an intuitively plausible revision procedure. Wagner's procedure is based on what I call Jeffrey's principle: Leave the ratio of probabilities alone if they are not affected by your observation or your evidence. Let Mr. A, Mr. B, Ms. X, and Ms. Y be the exclusive group of people suspected of a crime and the ratio of probabilities which the detective assigns to them (as having committed the crime) be 2:3:1:4 (out of ten). Jeffrey's principle states that if the detective finds out that the culprit is male, the probabilities update to 4:6:0:0, all else being equal.

Wagner constructs a case, the *Linguist* problem, where Jeffrey conditioning is not applicable. He generalizes Jeffrey conditioning, using Jeffrey's principle, and finds that his generalization violates the PME. There are now two plausible intuitions inconsistent with each other: Jeffrey's principle and the principle of maximum entropy. Wagner concludes that, given all the other conceptual problems of the PME and the counterexamples, we ought to accept Jeffrey's principle and feel that the case against the PME has been corroborated once again.

Wagner's application of the PME which violates Jeffrey's principle, however, is incorrect. A correct application agrees with both with his procedure and with Jeffrey's principle. I present a formal proof that the PME seamlessly and elegantly generalizes Wagner's generalization of Jeffrey conditioning, with the added benefit that it is not merely an ad hoc rule such as Wagner's, but integrated in a unified approach to probability kinematics. This section conducts a careful analysis of Wagner's procedure, which is based on Dempster's Rule, and argues that the PME offers an elegant generalization of everything that Wagner wants except his rejection of determinate prior probabilities. We will make a case for them within a Bayesian framework.

It is generally not difficult to find alternatives to the PME (consider for example van Fraassen's Maximum Transition Probability and his less delicately named MUD method, both of which beat the PME on select performance criteria). My claim is that based on the virtues of the PME and based on the fact that all such alternatives violate the principle that belief revision should not result in unwarranted information gain, we should reject such alternatives, no matter how appealing some of their qualities are. For van Fraassen's MUD and MTP, for example, on the first of two performance criteria MUD places first, the PME second, and the MTP third. On van Fraassen's other performance criterion, the ranking is just the reverse. I count it as a virtue of the PME to do relatively well on both criteria and would not be surprised if the PME in some sense optimizes its performance across criteria rather than doing particularly well on smaller subsets.

### 4.2.5 Coarsening at Random

Coarsening at random (CAR) involves using more naive (or coarse) event spaces in order to arrive at solutions to probability updating problems. Grünwald and Halpern show that updating on the naive space rather than the sophisticated space is legitimate for event type observations when the set of observations is pairwise disjoint or when the CAR condition (as defined by them) holds. For Jeffrey type observations, there is a generalized CAR condition which applies likewise. For strictly affine constraints, the PME essentially never gives the right results, according to the authors.

Grünwald and Halpern mention the *Judy Benjamin* problem as their prime example for a case in which the PME delivers the wrong result. It does so in analogy to the evidently wrong result in the *Monty Hall* or the *Three Prisoners* problem by naive conditioning. This chapter shows how the analogy is misguided and the authors' claim misleading that the PME essentially always gives the wrong results.

## 4.3 Conceptual Problems

This chapter addresses conceptual problems that have been raised with respect to the PME rather than specific counterexamples to the procedure. First, Jos Uffink targets especially Shore and Johnson's assumptions when they identify the PME as the unique method for determining updated probability distributions, given certain types of rationality constraints. Uffink shows how a more reasonable restatement of Shore and Johnson's assumptions results in a whole class of updating procedures, the so-called Rényi entropies.

Van Fraassen, with collaborators R.I.G. Hughes and Gilbert Harman, also concludes that there is a family of candidates which fulfills five requirements or principles that he establishes with the *Judy Benjamin* problem in view. If the five requirements are complemented by two performance criteria, the PME is not the best updating method with respect to any of the performance criteria. The objection that instead of a unique rational updating method there is a family of such methods is again common among philosophers, especially because it accords with Rudolf Carnap's continuum of inductive methods in a different context. This chapter presents and expands on Ariel Caticha and Adom Giffin's arguments against families of updating methods.

Second, there is a large branch of belief revision literature which assumes that a person can only have meaningful updating methods, for ranks, probabilities, or other measures of uncertainty, if behind a first (static) layer there is a second (dynamic) layer which determines how the first layer behaves when changes occur. The principle of maximum entropy, by contrast, only operates on one level and must incorporate the second level as information. Degrees of belief about degrees of belief may be coherent and useful (see Jeffrey, 1974; Skyrms, 1980; Domotor 1980; 1981). Our contention is that they can be subsumed under evidence-handling and are thus of a completely different source and nature than first-order degrees of belief.

Third, and closely related to the previous point, upon learning the truth of a conditional, degree of belief in the antecedent and the consequent may be reevaluated. AGM belief revision theory has at its heart the idea of epistemic entrenchment, the degree to which an agent wants to maintain belief in a proposition when she is informed of its antecedent role in conditionals. Igor

Douven and Jan-Willem Romeijn have applied epistemic entrenchment to the *Judy Benjamin* problem and evaluate it in terms of 'Adams conditioning,' the kind of conditioning where the degree of belief in the antecedent remains unaltered. This solution, which receives support in a number of other papers (for example, by Joseph Halpern), in contrast to the PME contradicts van Fraassen, Hughes, and Harman's five requirements.

Epistemic entrenchment and the PME are incompatible at first glance. Wolfgang Spohn's work in ranking theory illustrates a strong commitment to the idea that a rational agent needs both a quantitative assessment (either in terms of rankings or in terms of probability) as well as parameters of epistemic entrenchment in order to make sense out of updating in light of new observation or evidence. The PME, however, updates by operating on probability distributions alone, only in conjunction with an objective principle of minimum information gain. This chapter will seek reconciliation between the strong intuitive appeal of both the PME and epistemic entrenchment.

Fourth, in the spirit of this reconciliation we will try to establish that the epistemic dimensions of probabilistic belief and acceptance are complementary, i.e. irreducible with respect to each other on the one hand, while on the other hand rich in cross-fertilization. They do not exist in parallel universes with unrelated sets of formal relations and strictly separate domains. On the contrary, they inform each other and are best kept in view simultaneously, both formally and in contexts of application. It will be the task of this chapter to clarify the nature of the relationship.

## 4.4 Epistemological Implications and Conclusion

In the concluding chapter, I want to look at the wider epistemological implications of a more affirmative stance towards the PME as an objective updating method. Importantly, information becomes a basic notion to which philosophers will increasingly pay attention, possibly at the expense of probability. There may be a parallel to the salience that preference gained in epistemology following Ramsey's work and the numerous representation theorems in its wake. It is plausible that information will fill a similar need for explicating the notion of probability, this time, however, not in the more subjective terms of preference, but in the more objective terms of information.

Probability theory is easily expressible by information theory, whereas the reverse is not true. For example, there is no analogue to Chaitin's incompleteness theorem in probability theory—in general, information theory is easily linked to algorithmic compressibility (Shannon information to Solomonoff/-Kolmogorov complexity), which may very well have substantial epistemological implications once belief revision is viewed in information-theoretic rather than probability-theoretic terms.

Questions in epistemology, in the theory of causation, in the philosophy of science, and perhaps also in the philosophy of cognition may be rearticulated and answered in different ways if information is more widely used as a common currency between these fields. In physics, information already plays a vital role, whereas the philosophy of information in many respects is in its infancy.

## 5 Bibliography

### References

Alchourrón, Carlos, Peter Gärdenfors, and David Makinson. "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions." *The Journal of Symbolic Logic* 50, 2: (1985) 510–530.

Armendt, Brad. "Is There a Dutch Book Argument for Probability Kinematics?" *Philosophy of Science* 583–588.

Bar-Hillel, Yehoshua, and Rudolf Carnap. "Semantic information." *The British Journal for the Philosophy of Science* 4, 14: (1953) 147–157.

Bolker, Ethan. "Functions Resembling Quotients of Measures." *Transactions of the American Mathematical Society* 124, 2: (1966) 292–312.

———. "A Simultaneous Axiomatization of Utility and Subjective Probability." *Philosophy of Science* 34, 4: (1967) 333–340.

Boltzmann, Ludwig. *On the Nature of Gas Molecules.* Taylor and Francis, 1877.

Bovens, Luc. "Judy Benjamin is a Sleeping Beauty." *Analysis* 70, 1: (2010) 23–26.

Bradley, Richard. "Radical Probabilism and Bayesian Conditioning." *Philosophy of Science* 72, 2: (2005) 342–364.

Carnap, Rudolf. *The Continuum of Inductive Methods.* University of Chicago, 1952.

Carnap, Rudolf, and Yehoshua Bar-Hillel. *An Outline of a Theory of Semantic Information.* Cambridge, MA: MIT, 1952.

Caticha, Ariel, and Adom Giffin. “Updating Probabilities.” In *MaxEnt 2006, the 26th International Workshop on Bayesian Inference and Maximum Entropy Methods.* 2006.

Cox, Richard. “Probability, Frequency and Reasonable Expectation.” *American Journal of Physics* 14: (1946) 1.

Csiszár, Imre. “Information-Type Measures of Difference of Probability Distributions and Indirect Observations.” *Studia Scientiarum Mathematicarum Hungarica* 2: (1967) 299–318.

Cyranski, John F. “Measurement, Theory, and Information.” *Information and Control* 41, 3: (1979) 275–304.

De Finetti, B. “Sul Significato Soggettivo della Probabilità.” *Fundamenta mathematicae* 17, 298-329: (1931) 197.

De Finetti, Bruno. “La prévision: ses lois logiques, ses sources subjectives.” In *Annales de l’institut Henri Poincaré.* Presses universitaires de France, 1937, volume 7, 1–68.

Diaconis, Persi, and Sandy Zabell. “Updating Subjective Probability.” *Journal of the American Statistical Association* 77, 380: (1982) 822–830.

Dias, P., and A. Shimony. “A Critique of Jaynes’ Maximum Entropy Principle.” *Advances in Applied Mathematics* 2: (1981) 172–211.

Domotor, Zoltan. “Probability Kinematics and Representation of Belief Change.” *Philosophy of Science* 47, 3: (1980) 384–403.

———. “Higher Order Probabilities.” *Philosophical Studies* 40, 1: (1981) 31–46.

———. “Probability Kinematics, Conditionals, and Entropy Principles.” *Synthese* 63, 1: (1985) 75–114.

Douven, I., and J.W. Romeijn. “A New Resolution of the Judy Benjamin Problem.” *CPNSS Working Paper* 5, 7: (2009) 1–22.

Earman, John. *Bayes or Bust?* Cambridge, MA: MIT, 1992.

Kampé de Fériet, J., and B. Forte. "Information et probabilité." *Comptes rendus de l'Académie des sciences* A 265: (1967) 110–114.

Foley, Richard. *Working Without a Net: A Study of Egocentric Epistemology.* New York, NY: Oxford University Press, 1993.

Forster, Malcolm, and Elliott Sober. "How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions." *The British Journal for the Philosophy of Science* 45, 1: (1994) 1–35.

Friedman, Kenneth, and Abner Shimony. "Jaynes's Maximum Entropy Prescription and Probability Theory." *Journal of Statistical Physics* 3, 4: (1971) 381–384.

Gage, Douglas W, and David Hestenes. "Comment on the Paper Jaynes's Maximum Entropy Prescription and Probability Theory." *Journal of Statistical Physics* 7, 1: (1973) 89–90.

Gibbs, Josiah Willard. *Elementary Principles in Statistical Physics.* New Haven, CT: Yale University, 1902.

Gillies, Donald. *Philosophical Theories of Probability.* London, UK: Routledge, 2000.

Grove, A., and J. Halpern. "Probability Update: Conditioning Vs. Cross-Entropy." In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence.* Providence, Rhode Island: Citeseer, 1997.

Grünwald, Peter. "Maximum Entropy and the Glasses You Are Looking Through." In *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence.* Morgan Kaufmann, 2000, 238–246.

Grünwald, Peter, and Joseph Halpern. "Updating Probabilities." *Journal of Artificial Intelligence Research* 19: (2003) 243–278.

Guiaşu, Silviu. *Information Theory with Application.* New York, NY: McGraw-Hill, 1977.

Hacking, Ian. "Slightly More Realistic Personal Probability." *Philosophy of Science* 311–325.

Halpern, Joseph. "A Counterexample to Theorems of Cox and Fine." *Journal of Artifcial Intelligence Research* 10: (1999) 67–85.

———. *Reasoning About Uncertainty.* Cambridge, MA: MIT, 2003.

Hempel, Carl G., and Paul Oppenheim. "A Definition of Degree of Confirmation." *Philosophy of science* 12, 2: (1945) 98–115.

Hobson, A. *Concepts in Statistical Mechanics.* New York, NY: Gordon and Beach, 1971.

Hobson, Arthur. "The Interpretation of Inductive Probabilities." *Journal of Statistical Physics* 6, 2: (1972) 189–193.

Howson, Colin, and Allan Franklin. "Bayesian Conditionalization and Probability Kinematics." *The British Journal for the Philosophy of Science* 45, 2: (1994) 451–466.

Howson, Colin, and Peter Urbach. *Scientific Reasoning: The Bayesian Approach, Third Edition.* Chicago: Open Court, 2006.

Ingarden, R. S., and K. Urbanik. "Information Without Probability." *Colloquium Mathematicum* 9: (1962) 131–150.

Jaynes, E.T. "Information Theory and Statistical Mechanics." *Physical Review* 106, 4: (1957a) 620–630.

———. "Information Theory and Statistical Mechanics II." *Physical Review* 108, 2: (1957b) 171.

———. "The Well-Posed Problem." *Foundations of Physics* 3, 4: (1973) 477–492.

———. "Where Do We Stand on Maximum Entropy." In *The Maximum Entropy Formalism*, edited by R.D. Levine, and M. Tribus, Cambridge, MA: MIT, 1978, 15–118.

———. "Some Random Observations." *Synthese* 63, 1: (1985) pp. 115–138.

———. "Optimal Information Processing and Bayes's Theorem: Comment." *The American Statistician* 42, 4: (1988) 280–281.

Jaynes, E.T., and G.L. Bretthorst. *Probability Theory: the Logic of Science.* Cambridge, UK: Cambridge University, 2003.

Jeffrey, Richard. *The Logic of Decision.* New York, NY: McGraw-Hill, 1965.

———. "Dracula Meets Wolfman: Acceptance Versus Partial Belief." In *Induction, Acceptance and Rational Belief*, edited by Marshall Swain, Springer, 1970, 157–185.

———. "Preference Among Preferences." *The Journal of Philosophy* 71, 13: (1974) 377–391.

———. "Axiomatizing the Logic of Decision." In *Foundations and Applications of Decision Theory*, Springer, 1978, 227–231.

Jeffreys, Harold. *Scientific Inference.* Cambridge University, 1931.

———. *The Theory of Probability.* Cambridge University, 1939.

Joyce, James. *The Foundations of Causal Decision Theory.* Cambridge University, 1999.

———. "How Probabilities Reflect Evidence." *Philosophical Perspectives* 19, 1: (2005) 153–178.

Keynes, John Maynard. *A Treatise on Probability.* Diamond, 1909.

———. *A Treatise on Probability.* London, UK: Macmillan, 1921.

Khinchin, A.I. *Mathematical Foundations of Information Theory.* New York, NY: Dover, 1957.

Kolmogorov, A.N. "Logical Basis for Information Theory and Probability Theory." *IEEE Transactions on Information Theory* 14, 5: (1968) 662–664.

Kullback, Solomon. *Information Theory and Statistics.* Dover Publications, 1959.

Kullback, Solomon, and Richard Leibler. "On Information and Sufficiency." *The Annals of Mathematical Statistics* 22, 1: (1951) 79–86.

Kyburg, Henry. "Book Review: Betting on Theories by Patrick Maher." *Philosophy of Science* 62, 2: (1995) 343.

Levi, Isaac. "Probability Kinematics." *The British Journal for the Philosophy of Science* 18, 3: (1967) 197–209.

———. "The Demons of Decision." *The Monist* 70, 2: (1987) 193–211.

Maher, Patrick. “Diachronic Rationality.” *Philosophy of Science* 120–141.

———. *Betting on Theories.* Cambridge University, 1993.

Mikkelson, Jeffrey. “Dissolving the Wine/Water Paradox.” *The British Journal for the Philosophy of Science* 55, 1: (2004) 137–145.

Paris, Jeff. *The Uncertain Reasoner’s Companion: A Mathematical Perspective*, volume 39. Cambridge, UK: Cambridge University, 2006.

Paris, Jeff, and Alena Vencovská. “A Note on the Inevitability of Maximum Entropy.” *International Journal of Approximate Reasoning* 4, 3: (1990) 183–223.

Rowbottom, Darrell. “The Insufficiency of the Dutch Book Argument.” *Studia Logica* 87, 1: (2007) 65–71.

Savage, Leonard. *The Foundations of Statistics.* New York, NY: Wiley, 1954.

Seidenfeld, Teddy. “Why I Am Not an Objective Bayesian; Some Reflections Prompted by Rosenkrantz.” *Theory and Decision* 11, 4: (1979) 413–440.

———. “Entropy and Uncertainty.” In *Advances in the Statistical Sciences: Foundations of Statistical Inference*, Springer, 1986, 259–287.

Seidenfeld, Teddy, Mark Schervish, and Joseph Kadane. “When Fair Betting Odds Are Not Degrees of Belief.” In *Proceedings of the Biennial Meeting of the Philosophy of Science Association.* JSTOR, 1990, 517–524.

Sen, Amartya. “Choice Functions and Revealed Preference.” *The Review of Economic Studies* 38, 3: (1971) 307–317.

Shannon, Claude Elwood. “A Mathematical Theory of Communication.” *ACM SIGMOBILE Mobile Computing and Communications Review* 5, 1: (2001) 3–55.

Shimony, Abner. “The Status of the Principle of Maximum Entropy.” *Synthese* 63, 1: (1985) 35–53.

———. *The Search for a Naturalistic World View.* Cambridge University, 1993.

Shore, J., and R.W. Johnson. “Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy.” *IEEE Transactions on Information Theory* 26, 1: (1980) 26–37.

Skilling, John. “The Axioms of Maximum Entropy.” In *Maximum-Entropy and Bayesian Methods in Science and Engineering*, edited by G.J. Erickson, and C.R. Smith, Dordrecht, Holland: Springer, 1988, 173–187.

Skyrms, Brian. “Higher Order Degrees of Belief.” In *Prospects for Pragmatism: Essays in Memory of F.P. Ramsey*, edited by David Mellor, Cambridge University, 1980, 109–137.

———. “Maximum Entropy Inference as a Special Case of Conditionalization.” *Synthese* 63, 1: (1985) 55–74.

———. “Dynamic Coherence.” In *Advances in the Statistical Sciences: Foundations of Statistical Inference*, Springer, 1986, 233–243.

———. “Coherence.” In *Scientific Inquiry in Philosophical Perspective*, edited by N. Rescher, Lanham, MD: University Press of America, 1987a, 225–242.

———. “Updating, Supposing, and Maxent.” *Theory and Decision* 22, 3: (1987b) 225–246.

Spohn, Wolfgang. *The Laws of Belief: Ranking Theory and Its Philosophical Applications.* New York, NY: Oxford University, 2012.

Teller, Paul. “Conditionalization and Observation.” *Synthese* 26, 2: (1973) 218–258.

———. “Conditionalization, Observation, and Change of Preference.” In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Dordrecht: Reidel, 1976.

Tikochinsky, Y, NZ Tishby, and RD Levine. “Alternative Approach to Maximum-Entropy Inference.” *Physical Review A* 30, 5: (1984) 2638.

Topsøe, F. “Information-Theoretical Optimization Techniques.” *Kybernetika* 15, 1: (1979) 8–27.

Tribus, Myron, and Hector Motroni. “Comments on the Paper Jaynes’s Maximum Entropy Prescription and Probability Theory.” *Journal of Statistical Physics* 4, 2: (1972) 227–228.

Uffink, Jos. “Can the Maximum Entropy Principle Be Explained as a Consistency Requirement?” *Studies in History and Philosophy of Science* 26, 3: (1995) 223–261.

———. “The Constraint Rule of the Maximum Entropy Principle.” *Studies in History and Philosophy of Science* 27, 1: (1996) 47–79.

Van Fraassen, Bas. “A Problem for Relative Information Minimizers in Probability Kinematics.” *The British Journal for the Philosophy of Science* 32, 4: (1981) 375–379.

Van Fraassen, Bas, R.I.G. Hughes, and Gilbert Harman. “A Problem for Relative Information Minimizers, Continued.” *The British Journal for the Philosophy of Science* 37, 4: (1986) 453–463.

Wagner, Carl. “Generalized Probability Kinematics.” *Erkenntnis* 36, 2: (1992) 245–257.

———. “Probability Kinematics and Commutativity.” *Philosophy of Science* 69, 2: (2002) 266–278.

Walley, P. *Statistical Reasoning with Imprecise Probabilities.* London, UK: Chapman and Hall, 1991.

Wiener, Norbert. “The Ergodic Theorem.” *Duke Mathematical Journal* 5, 1: (1939) 1–18.

Williamson, Jon. “Objective Bayesianism, Bayesian Conditionalisation and Voluntarism.” *Synthese* 178, 1: (2011) 67–85.

Zabell, Sandy. *Symmetry and Its Discontents: Essays on the History of Inductive Probability.* Cambridge University Press, 2005.

Zellner, Arnold. “Optimal Information Processing and Bayes’s Theorem.” *The American Statistician* 42, 4: (1988) 278–280.