Bayesian Epistemology

Alan Hájek and Stephan Hartmann

1. Introduction

Bayesianism is our leading theory of *uncertainty*. Epistemology is defined as the theory of *knowledge*. So "Bayesian Epistemology" may sound like an oxymoron. Bayesianism, after all, studies the properties and dynamics of degrees of belief, understood to be probabilities. Traditional epistemology, on the other hand, places the singularly non-probabilistic notion of knowledge at centre stage, and to the extent that it traffics in belief, that notion does not come in degrees. So how can there be a Bayesian epistemology?

According to one view, there cannot: Bayesianism fails to do justice to essential aspects of knowledge and belief, and as such it cannot provide a genuine epistemology at all. According to another view, Bayesianism should supersede traditional epistemology: where the latter has been mired in endless debates over skepticism and Gettierology, Bayesianism offers the epistemologist a thriving research program. We will advocate a more moderate view: Bayesianism can illuminate various long-standing problems of epistemology, while not addressing all of them; and while Bayesianism opens up fascinating new areas of research, it by no means closes down the staple preoccupations of traditional epistemology.

The contrast between the two epistemologies can be traced back to the mid-17th century. Descartes regarded belief as an all-or-nothing matter, and he sought justifications for his claims to knowledge in the face of powerful skeptical arguments. No more than four years after his death, Pascal and Fermat inaugurated the probabilistic revolution, writ large in the *Port-Royale Logic*, in which the many shades of uncertainty are represented with probabilities, and rational decision-making is a matter of maximizing expected utilities (as we now call them). Correspondingly, the Cartesian concern for knowledge fades into the background, and a more nuanced representation of epistemic states has the limelight. Theistic belief provides a vivid example of the contrasting orientations. Descartes sought certainty in the existence of God grounded in apodeictic demonstrations. Pascal, by contrast, explicitly shunned such alleged 'proofs', arguing instead that our situation with respect to God is like a gamble, and that belief in God is the best bet—thus turning the question of theistic belief into a *decision* problem (which he, unlike Descartes, had the tools to solve). Bayesian epistemology owes its name to the Reverend Thomas Bayes, who, a century later, published an important theorem that underwrites certain calculations of conditional probability central to confirmation theory—more on this shortly. But really 'Bayesian epistemology' is something of a misnomer; 'Kolmogorovian epistemology' would be far more appropriate, as we will see.

Caveats: When we speak of 'traditional epistemology', we lump together a plethora of positions as if they form a monolithic whole. Other articles in this volume distinguish carefully among various positions that our broad banner conflates. For our purposes, they start out regarding knowledge and belief as the central concepts of epistemology, and then to go on to study the properties, grounds, and limits of these binary notions. We also speak of 'Bayesianism' as if it is a unified school of thought, when in fact there are numerous intra-mural disputes. I. J. Good (1971) calculates that there are (at least) 46,656 ways to be a Bayesian, while we will mostly pretend that there is just one. By and large, the various distinctions among Bayesians will not matter for our purposes. As a good (indeed, a Good) Bayesian might say, our conclusions will be robust under various precisifications of the position. Many traditional problems can be framed, and progress can be made on them, using the tools of probability theory. But Bayesian epistemology does not merely recreate traditional epistemology; thanks to its considerable expressive power, it also opens up new lines of enquiry.

2. What is Bayesian Epistemology?

Bayesian epistemology is the application of Bayesian methods to epistemological problems. Bayesianism models degrees of belief as *probabilities* along the lines of Kolmogorov's (1933) axiomatization. Let Ω be a non-empty set. A *field* (*algebra*) on Ω is a set $\mathcal{F}$ of subsets of Ω that has Ω as a member, and that is closed under complementation (with respect to Ω) and union. Let P be a function from $\mathcal{F}$ to the real numbers obeying:

1) $P(a) \geq 0$ for all $a \in \mathcal{F}$. (Non-negativity)
2) $P(\Omega) = 1$. (Normalization)
3) $P(a \cup b) = P(a) + P(b)$ for all $a, b \in \mathcal{F}$ such that $a \cap b = \varnothing$. (Finite additivity)

Call P a *probability function*, and $(\Omega, \mathcal{F}, P)$ a *probability space*.

One could instead attach probabilities to members of a collection of sentences of a formal language, closed under truth-functional combinations; this is more common in Bayesian confirmation theory. A lively area of current debate concerns just how finely grained such contents of probability attributions should be. For example, various problems of 'self-location' suggest that probabilities should attach to 'centred propositions', e.g. <possible world, individual, time> triples.

Kolmogorov extends his axiomatization to cover infinite probability spaces, requiring $\mathcal{F}$ to be closed under *countable* union, and strengthening 3) to *countable* additivity. He defines the *conditional probability of a given b* by the ratio of unconditional probabilities:

$$P(a \mid b) = \frac{P(a \cap b)}{P(b)}, \text{ provided } P(b) > 0.$$

If $P(a \mid b) = P(a)$, then a and b are said to be *independent* (relative to P).

Versions of *Bayes' theorem* can now be proven:

$$P(a \mid b) = \frac{P(b \mid a)P(a)}{P(b)}$$

$$= \frac{P(b \mid a)P(a)}{P(b \mid a)P(a) + P(b \mid \neg a)P(\neg a)}$$

More generally, suppose there is a partition of hypotheses $\{h_1, h_2, \ldots, h_n\}$, and evidence e. Then for each i,

$$P(h_i \mid e) = \frac{P(e \mid h_i)P(h_i)}{\sum_{j=1}^{n} P(e \mid h_j)P(h_j)}.$$

The $P(e \mid h_i)$ terms are called *likelihoods*, and the $P(h_i)$ terms are called *priors*. See Joyce (2008) for a more detailed discussion of varieties and uses of Bayes' theorem.

Bayesianism offers a natural analysis of the relation of *confirmation* between a piece of evidence e and a hypothesis h:

e confirms h (relative to P) iff $P(h \mid e) > P(h)$.

We may also define various probabilistic notions of *comparative* confirmation, and various *measures of evidential support* (see Eells and Fitelson 2000 and Fitelson 1999).

Bayesianism can be understood as combining a *synchronic* thesis about the degrees of belief or *credences* of a rational agent at a given time, and a *diachronic* thesis about how they evolve in response to evidence. Synchronically, the agent's credences are probabilities. Diachronically, her credences update according to the rule of *conditionalization.* Suppose that initially her credences are given by probability function $P_{initial}$, and that she becomes certain of *e* (where *e* is the strongest such proposition). Then her new credence function P_{new} is related to $P_{initial}$ *as* follows:

(Conditionalization) $P_{new}(x) = P_{initial}(x \mid e)$ (provided $P_{initial}(e) > 0$).

Jeffrey conditionalization allows for less decisive learning experiences in which her probabilities across a partition $\{e_1, e_2, \ldots\}$ change to $\{P_{new}(e_1), P_{new}(e_2), \ldots, \}$, where none of these values need be 0 or 1:

$$P_{new}(x) = \Sigma_i P_{initial}(x \mid e_i)\, P_{new}(e_i) \text{ (provided } P_{initial}(e_i) > 0).$$

(Jeffrey 1983). $P_{new}(x)$ is called the *posterior* probability function.

According to Bayesian orthodoxy, an agent begins with a 'prior' probability function and repeatedly updates by (Jeffrey) conditionalization as evidence comes in. This combines a striking permissiveness about the starting point of an agent's epistemological odyssey with considerable rigidity about how the agent should respond to evidence. But Bayesianism is a theme that admits of many variations—see Good (1971) and Hájek and Hall (2002) for discussion of some of them. Here, let us consider several further constraints on 'priors' that have been proposed.

A probability function is said to be *regular* if it assigns probability 1 only to tautologies, and probability 0 only to contradictions—to all other sentences it assigns intermediate values. It seems to be an epistemological desideratum that a prior be regular, reflecting an open-mindedness appropriate to an agent who is a *tabula rasa*.

A rationale is that to rule out (probabilistically speaking) a priori some genuine logical possibility would be to pretend that one's evidence was stronger than it really was. The *principle of indifference* also enjoins you to reflect the poverty of your evidence in your credences: you are required to give equal probability to all possibilities among which your evidence does not discriminate (and in a state of total evidential innocence, that is all of them). A sophisticated version of the principle of indifference, favoured by so-called *objective* Bayesians, has been explored by Jaynes (2003): maximize the probability function's *entropy*, which for an assignment of positive probabilities p_1, ..., p_n to n possibilities equals $-\Sigma_i\, p_i \log(p_i)$. See also Williamson (2005).

Then there are two principles that are meant to codify one's epistemic commitment to aligning one's credences to certain probabilistic hypotheses. In the first, Lewis's *Principal Principle* (here simplified), the hypotheses concern the *objective chance* of the relevant propositions (Lewis 1980):

$$C_0(a \mid ch_t(a) = x) = x \text{ (for all } a \text{ and } t \text{ for which this is defined).}$$

Here C_0 is some reasonable prior, a an arbitrary proposition, and $ch_t(a) = x$ the claim that the chance at time t of a is x. The idea is that one should align one's credences with what one takes the corresponding objective chances to be, where the latter are genuine probabilities *in the world*. In the second, van Fraassen's *Reflection Principle* (van Fraassen 1984), the hypotheses concern one's own future credences for the relevant propositions

$$C_t(a \mid C_{t'}(a) = x) = x \text{ (for all } t, t', a \text{ and } x \text{ for which this is defined).}$$

Here C_t is one's probability function at time t, and $C_{t'}$ one's function at later time t'.

The idea is that rationality requires a certain commitment to one's future opinions; when all is going well, one's future selves are better-informed versions of one's current self.

3. Contrasts between Traditional Epistemology and Bayesian Epistemology

We can now bring out several points of contrast between traditional and Bayesian epistemology. We have noted that 'knowledge' and 'belief' are binary notions, to be contrasted with the potentially infinitely many degrees of 'credence' (corresponding to all the real numbers in the [0, 1] interval). 'Knowledge' is not merely 'justified true belief', as Gettier has famously shown, but many epistemologists hope that some 'fourth condition' will complete the analysis—some kind of condition that rules out cases in which one has a justified true belief by luck, or for some anomalous reason. Some epistemologists (e.g. Sosa (1999)) advocate versions of *safety* as a condition on knowledge—roughly, at the closest worlds in which a given agent believes *p, p* is true. Others (e.g. Nozick (1981)) advocate *sensitivity*—roughly, at the closest worlds in which *p* is false, the agent does not believe *p*. (And some advocate both.) Note well: the *closest* worlds. Here we find another disjuncture between traditional and Bayesian epistemology: nothing in the standard Bayesian apparatus reflects the 'similarity' of worlds that has taken centre-stage in the analysis of knowledge.

Notice that truth, justification, and these anti-luck conditions may be characterized as at least partially *objective*, with 'belief' providing the only purely *subjective* component. This is in sharp contrast to orthodox Bayesianism, which refines and analyzes this doxastic notion, but which has no clear analogue of the 'objective' conditions. Most importantly, Bayesianism apparently has nothing that corresponds to the *factivity* of knowledge: that one can only know truths. And even when our beliefs fall short of knowledge, still it is a desideratum that they be true; but the Bayesian seems to have no corresponding desideratum for intermediate credences, which are its stock-in-trade. When you assign, for example, probability 0.3 to it raining tomorrow, what sense can be made of this assignment being *true?* It is also dubious whether Bayesianism can capture 'justification' (cf. Shogenji (2009) for an interesting proposal) or any 'anti-luck' condition on knowledge—more on this shortly.

Relatedly, all Bayesian claims must be relativized to a probability function, or more precisely, to a probability space—an entire probability model. We saw this above in the definitions of 'independence' and 'confirmation'—they came with parenthetical references to *P*. Many authors suppress these references, encouraging one to forget their inherent subjectivity (and even we secreted them away in slightly disingenuous parentheses!). Traditional epistemologists, by contrast, conduct much of their discussions in terms free of any such relativization—they speak of one proposition being *evidence* for another, of a process of belief acquisition being *reliable*, and so on without any qualification. And again, truth, justification, and 'anti-luck' conditions are typically supposed to hold or not independently of whether some agent *thinks* that they do, or whether some *model* says they do.

The synchronic requirement that an agent's credences obey the probability calculus may be regarded as generalizing the requirement familiar from traditional epistemology that one's beliefs should be consistent. The diachronic requirement of conditionalization is reminiscent of the Quinean principle of 'minimum mutilation' (change beliefs as little as the evidence allows)—a conservative recipe for belief revision. But nothing in traditional epistemology corresponds to Jeffrey conditionalization—it is *essentially* a probabilistic revision rule. The principle of indifference corresponds very roughly to a Cartesian admonition to suspend judgment when one's evidence is lacking, but it is far more specific. And there are no traditional analogues of the various additional constraints on priors. Going in the other direction, Bayesianism is silent about some of the cornerstones and more recent concerns of traditional epistemology—we will discuss this at greater length at the end.

Given the striking differences between traditional and Bayesian epistemology, are there reasons to prefer one to the other?

4. Thesis: Bayesian Epistemology Is Superior to Traditional Epistemology

Jeffrey, a famous Bayesian, suggests two main benefits accrued by the Bayesian framework in his (1992):

1. Subjective probabilities figure in *decision theory*, an account of how our opinions and our desires conspire to dictate what we should do. The desirability of each of our possible actions is measured by its *expected utility*, a probability-weighted sum of the utilities associated with that action. To complete Jeffrey's argument, we should add that traditional epistemology offers no decision theory (recall Descartes versus Pascal). The analysis of rational action surely needs to advert to more fine-grained mental states than binary belief and knowledge. (See Eriksson and Hájek 2007 for more discussion of why the intermediate credences that are necessary for that analysis cannot be reduced to these binary notions.)

2. Observations rarely deliver certainties—rather, their effect is typically to raise our

probabilities for certain propositions (and to drop our probabilities for others), without any reaching the extremes of 1 or 0. Traditional epistemology apparently has no way of accommodating such less-than-conclusive experiential inputs, whereas Jeffrey conditionalization is tailor-made to do so.

We may continue the list that Jeffrey has started of putative advantages of Bayesianism over traditional epistemology at some length:

3. Knowledge is unforgiving. Its standards are so high that they can rarely be met, at least in certain contexts. (This is related to the fact that knowledge does not come in degrees—near-knowledge is not knowledge at all.) This in turn plays into the hands of skeptics. But it is harder for skeptical arguments to get a toehold against the Bayesian. For example, the mere *possibility* of error regarding some proposition X undermines a claim of knowledge regarding X, but it is innocuous from a probabilistic point of view: an agent can simply assign X some suitable probability less than 1. Indeed, even an assignment of probability 1 is consistent with the possibility of error —under plausible assumptions, it can be shown that a dart thrown at random at a representation of the [0, 1] interval has probability 1 of hitting an irrational number, even though it might fail to do so.

4. Moreover, it is a platitude that doxastic states come in degrees, and the categories of 'belief' and 'knowledge' are too coarse-grained to do justice to this fact. You believe, among other things, that 2 + 2 = 4, that you have a hand, that London is in England, and (say) that Khartoum is in Sudan. But you do not have the same

confidence in all these propositions, as we can easily reveal in your betting behavior and other decision-making that you might engage in. The impoverished nature of ‘belief’ attributions is only exacerbated when we consider the wide range of propositions for which you have less confidence—that this coin will land heads, that it will rain tomorrow in Novosibirsk, and so on. We may conflate your attitudes to them all as ‘suspensions of belief’ (as Descartes would), but that belies their underlying structure. Such attitudes are better understood as subjective probabilities.

5. Relatedly, the conceptual apparatus of deductivism is impoverished, and comparatively little of our reasoning can be captured by it, either in science or in daily life (*pace* Popper and Hempel). After all, whether we like it or not, our epistemic practices constantly betray our commitment to relations of support that fall short of entailment (Oaksford and Chater 2007). We think that it would be irrational to deny that the sun will rise tomorrow, to project ‘grue’ rather than ‘green’ in our inductions, and to commit the gambler’s *fallacy*. Probability theory helps us to understand such relations.

6. Bayesianism has powerful mathematical underpinnings. It can help itself to a century of work in probability theory and statistics. Traditional epistemology may appeal to the occasional system of epistemic or doxastic logic, but nothing comparable to the formidable formal machinery that we find in the Bayesian’s tool kit.

7. Bayesian methods, in turn, have much wider application than any formal

systematization of ‘knowledge’ or ‘belief’. Look at the sciences, social sciences, engineering, and artificial intelligence if you need any convincing of this.

8. There are many arguments for Bayesianism, which collectively provide a kind of triangulation to it. For example, ‘Dutch Book arguments’ provide an important defense of the thesis that rational credences are probabilities. An agent’s credences are identified with her betting prices; it is then shown that she is susceptible to sure losses iff these prices do not conform to Kolmogorov’s axioms. There are also arguments from various decision-theoretic representation theorems (Ramsey 1931, Savage 1954, Joyce 1999), from calibration (van Fraassen 1984), from ‘gradational accuracy’ or minimization of discrepancy from truth (Joyce 1998), from qualitative constraints on reasonable opinion (see Earman 1996 for a discussion of results of Cox and others), and so on. Moreover, there are various arguments in support of conditionalization and Jeffrey conditionalization—e.g., Dutch book arguments (Armendt 1980, Lewis 1999) and arguments from minimal revision of one’s credences (Diaconis and Zabell 1982). Again, there is nothing comparable in traditional epistemology.

9. Finally, a pragmatic argument for Bayesianism comes from an evaluation of its fruits. As we show at greater length in section 7, Bayesianism is highly explanatory with minimal resources—a simple, fecund theory if ever there was one. Traditional epistemology is hard-pressed to offer the same rewards. For example, we will see how various important intuitions about confirmation can be vindicated by a Bayesian analysis, and some erroneous intuitions can be corrected. It seems that no analysis

couched purely in terms of 'knowledge' and 'belief' could pay such dividends.

So we see various advantages that Bayesianism apparently has over traditional epistemology. But this does not tell the whole story. For starters, the triumphs of Bayesian confirmation theory just touted are supposedly offset by the so-called *problem of old evidence* (Glymour 1980). If $P(e) = 1$, then e apparently cannot confirm anything by Bayesian lights: in that case, $P(h \mid e) = P(h \cap e)/P(e) = P(h)$. Yet we often think that such 'old evidence' can be confirmatory. Consider the evidence of the advance of the perihelion of Mercury, which was known to Einstein at the time that he formulated general relativity theory, and thus (we may assume) was assigned probability 1 by him. Nonetheless, he rightly regarded this evidence as strongly confirmatory of general relativity theory. The challenge for Bayesians is to account for this. (See Zynda (1995) for discussion.)

Bayesianism, then, is not without problems of its own. So let us revisit the contest between traditional epistemology and Bayesianism, this time looking at arguments that incline in favor of the former.

5. Antithesis: Bayesian Epistemology Is *Not* Superior to Traditional Epistemology

1. Bayesians introduce a new technical term, 'degree of belief', but they struggle to explicate it. To be sure, the literature is full of nods to betting interpretations, but these meet a fate similar to that of behaviorism—indeed, a particularly *localized* behaviorism that focuses solely on the rather peculiar kind of behavior that is mostly found at racetracks and casinos. Other characterizations of 'degree of belief' that fall out of decision-theoretic representation theorems are also problematic. (See Eriksson

and Hájek 2007.) 'Belief', by contrast, is so familiar to the folk that it needs no explication.

2. Recall the absence of any notion of *truth* of an intermediate degree of belief. Yet truth is the very aim of belief. It is usually thought to consist in *correspondence* to the way things are. Moreover, we want our methods for acquiring beliefs to be *reliable*, in the sense of being *truth-conducive*. What is the analogous aim, notion of correspondence, and notion of reliability for the Bayesian? The terms of her epistemology seem to lack the success-grammar of these italicized words. For example, one can assign very high probability to the period at the end of this sentence being the creator of the universe without incurring any Bayesian sanction: one can do so while assigning correspondingly low probability to the period *not* being the creator, and while dutifully conditionalizing on all the evidence that comes in. Traditional epistemology is not so tolerant, and rightly not.

3. Relatedly, the Bayesian does not answer the skeptic, but merely ignores him. Bayesianism doesn't make skeptical positions go away; it merely makes them harder to state.

4. The Bayesian similarly lacks a notion of 'justification'—or to the extent that she has one, it is too permissive. At least on what we have called Bayesian orthodoxy, any prior is a suitable starting point for a Bayesian odyssey, yet mere conformity to the probability calculus is scant justification.

Now, the Bayesian will be quick to answer this and the previous objections in a single stroke. She will appeal to various *convergence theorems*. For example:

> If observations are precise… then the form and properties of the prior distribution have negligible influence on the posterior distribution. From a practical point of view, then, the untrammeled subjectivity of opinion… ceases to apply as soon as much data becomes available. More generally, two people with widely divergent prior opinions but reasonably open minds will be forced into arbitrarily close agreement about future observations by a sufficient amount of data. (Edwards et al. 1963: 201)

Call this *convergence to intersubjective agreement;* such agreement, moreover, is often thought to be the mark of objectivity. The "forcing" here is a result of conditionalizing the people's priors on the data. Gaifman and Snir (1982) similarly show that for each suitably open-minded agent, there is a data set sufficiently rich to force her arbitrarily close to assigning probability 1 to the true member of a partition of hypotheses. Call this *convergence to the truth.* The Bayesian might even try to parlay these theorems into providing surrogates for that 'fourth condition' for knowledge, insisting that such convergences do not happen by luck, or for some anomalous reason, but are probabilistically guaranteed.

These are beautiful theorems, but one should not overstate their epistemological significance. They are 'glass half-full' theorems, but a simple reversal of the quantifiers turns them into 'glass half-empty' theorems. For each data set, there is a suitably open-minded agent whose prior is sufficiently perverse to thwart such convergence: after conditionalizing her prior on the data set, she is still nowhere near assigning probability 1 to the true hypothesis, and still nowhere near agreement with other people. And strong assumptions underlie the innocent-sounding phrases "suitably open-minded agent" and "sufficiently rich data set". No data set, however rich, will drive anywhere at all a dogmatic agent who concentrates all credence on a single world (maximally specific hypothesis). Worse, an agent with a wacky enough prior will be driven *away* from the truth. Consider someone who starts by giving low probability to being a brain in a vat, but whose prior regards all the evidence that she actually gets as confirming that she is. And we can always come up with rival hypotheses that no courses of evidence can discriminate between—think of the irresolvable conflict between an atheist and a creationist who sees God's handiwork in everything.

5. A proponent of Bayesianism may describe it as "fecund" (and we did); but an opponent may describe it as "empty". With so little constraint on priors, it is not surprising that Bayesianism accounts for so much. Indeed it is irrational to deny that the sun will rise tomorrow, to project 'grue' rather than 'green' in our inductions, and to commit the gambler's *fallacy*. The trouble is that Bayesianism condones all such inferential practices—for all are licensed by suitably perverse priors.

6. The traditional epistemologist may protest that Bayesians distance themselves from the world. Recall our discussion of the relativization of Bayesian claims to the subjective probability functions of agents. Rather than hooking up directly with the world, the terms of their epistemology are all *internal to probabilistic models* of the world. Moreover, the Bayesian apparently does not have much of a story about what makes a model good, or one model better than another. This is related to the concern that the Bayesian does not do justice to truth, justification, and the 'anti-luck' conditions.

What are we to make of these conflicting considerations for and against Bayesian epistemology? At this stage of the dialectic, any good Hegelian will insist that it's time for a:

6. Synthesis

Should we really prefer one approach to epistemology over the other? Should one of the two approaches be jettisoned? We will argue that we should not regard them as in competition. In fact, the two approaches complement each other in both subject matter and method.

Traditional epistemologists sometimes stress that philosophy differs from science and insist that philosophy has its own distinct method of enquiry, namely conceptual analysis. Bayesians, on the other hand, typically consider their work more in line with scientific theorizing. This is reflected in the many connections Bayesian epistemology has with Bayesian statistics, decision theory and the literature on causal discovery in artificial intelligence. It is also reflected in the importance Bayesians give to solving real problems. As we will see in the next section, Bayesianism is tremendously successful in this respect. Given these successes, Bayesians should hardly be expected to give up their framework just because it is not a complete panacea for all epistemological ills. By way of analogy, scientists rightly held onto Newtonian mechanics even in the face of some theoretical problems (for example, its commitment to action at a distance). And despite its being superseded by Relativity Theory, we understand precisely how Newtonian mechanics is still approximately true. We believe that the successor of Bayesianism will stand in a similar relation to Bayesianism as Relativity Theory stands to Newtonian mechanics.

So let us revisit the charges leveled at Bayesianism in the previous section, keeping in mind this view of it as a work-in-progress that nonetheless is clearly earning its theoretical keep—like any good scientific theory. We may happily take its fundamental concept, 'degree of belief' as a primitive in the absence of a successful analysis of it. It has earned *its* theoretical keep by its contribution to a virtuous total theory—like any primitive scientific concept. And it can live peacefully alongside traditional epistemology's primitive concept of *belief*, without any expectation of reduction of one to the other. Indeed, the prospects for such reduction strike us as unpromising. (See Eriksson and Hájek 2007.)

Against the charge of Bayesianism being empty, it can plead the good company of deductive logic. To be sure, crazy sets of belief can be consistent, and inferences from absurd premises can be valid—the slogan 'garbage in, garbage out' is as true in epistemology as it is in computer science. Bayesianism, like logic, can nevertheless play a salutary role in keeping our degrees of belief, like our beliefs, in harmony, and in policing our elicit inferences. After all, deductive logic is never regarded as a *complete* set of constraints on belief; similarly, the Bayesian constraints on degrees of belief should not be regarded as complete. Some additional constraints may well find their inspiration in traditional epistemology.

Against the charge of Bayesianism's verdicts being model-relative, it can plead the good company of science. After all, our best methods of enquiry in the physical and social sciences work like this. Arguably we should not expect epistemology to be different.

Nor need answering skeptical challenges be part of Bayesianism's job description, just as it is not part of traditional epistemology's job description to underwrite rational decision making, confirmation, and the use of probabilistic and statistical methods in the sciences. There is no harm in their labour being divided. They are two different ways to approach epistemology and they often answer different questions.

And where their questions are shared, their approaches can be complementary rather than mutually exclusive. We see that Bayesian epistemology helps to address some questions that we find in traditional epistemological debates. The Bayesian treatment of issues such as testimony and the coherence theory of justification (see section 7) are cases in point. It turns out that the formal machinery of Bayesianism is well suited to make certain questions more precise and to provide answers when our intuitions don't give clear verdicts.

Now let us see what one can do with the machinery of Bayesianism. It should be assessed by the problems it solves and how much it unifies, for example, the methodology of science.

7. Achievements of Bayesian Epistemology

Many of the Bayesian success stories are from confirmation theory. But Bayesianism has much more to offer as its domain of applicability also includes other parts of epistemology and philosophy of science. Here are five highlights.

1. Confirmation Theory. As we saw in section 2, Bayesians begin with the idea that confirmation is a matter of *probability-raising*. They then show how important intuitions about confirmation can be vindicated. Suppose that h entails e, so that $P(e \mid h) = 1$. Then the posterior probability of h is $P(h \mid e) = P(h)/P(e)$. Hence, for a fixed prior probability of h, the posterior probability of h increases if $P(e)$ decreases. From this we can immediately account for the methodological insight that *more surprising evidence confirms better*. Similarly, Bayesians have provided a rationale for the *variety-of-evidence thesis*—the more varied the evidence is, the better—and have provided illuminating discussions of the Duhem-Quine thesis (Earman 1996, ch. 3).

To address these issues, several model assumptions have to be made. In the case of the variety-of-evidence thesis, for example, "more varied" has to be explicated in probabilistic terms. This can be done in different ways. Many Bayesian discussions of the variety-of-evidence thesis assume that the evidence is certain. But we saw already that this is not always the case. It speaks in favor of the Bayesian framework that it provides the tools to model more complicated testing scenarios. See Bovens and Hartmann (2003) for more realistic Bayesian models of the variety-of-evidence thesis and the Duhem-Quine thesis.

Bayesian confirmation theory connects naturally with empirical psychology as a wealth of work in the psychology of reasoning under uncertainty demonstrates. See, for example, Chater and Oaksford (2008) and Oaksford and Chater (2007) for sophisticated Bayesian models that account for empirical findings. Crupi et al. (2008) show how the presence of the conjunction fallacy (i.e. that experimental subjects assign a higher probability to a conjunction than to one of the conjuncts), as famously demonstrated in psychological experiments by Tversky and Kahneman (1983), can be explained in confirmation-theoretical terms. Crupi et al. (2007) argue on normative and experimental grounds for a specific measure of evidential support (the so-called *Z*-measure). And Bayesian confirmation theory provides a flexible framework to rationally reconstruct specific episodes from the history of science. While traditional epistemology does not have the resources to study such episodes, the Bayesian framework is ideally suited for these purposes (Franklin 1990). It is also better suited than some system of epistemic or doxastic logic—imagine trying to illuminate some scientific episode solely with '*K*' and '*B*' operators!

2. Dynamics of Belief. Traditional epistemology, with its focus on the analysis of knowledge, is relatively silent about questions of belief dynamics. If there is talk about belief change, it is generally assumed that it takes place on the basis of learned evidence that is certain. Traditional epistemology shares this assumption with logical theories of belief revision such as the AGM theory (Gärdenfors and Rott 1995). However, Jeffrey taught us that learning often does not come in the form of certainties. To address these cases of learning and belief change, philosophers as well as researchers in artificial intelligence have formulated new updating rules (such as Jeffrey conditionalization) and developed powerful tools such as the theory of Bayesian networks (Neapolitan 2003).

3. Applications. Bayesianism has a symbiotic relationship with causation and powerful algorithms have been developed to learn causal relations from probabilistic data (Korb and Nicholson 2004, Pearl 2000, Spirtes et al. 2001). These algorithms use the theory of Bayesian networks. A Bayesian network organizes a set of variables into a *Directed Acyclic Graph* (DAG). A DAG is a set of nodes and a set of arrows between some of the nodes. The only constraint is that there are no closed paths formed by following the arrows. A *root node* is a node with outgoing arrows only, and a *parent* of a given node is a node from which an arrow points into the given node. Each node represents a propositional variable, which can take any number of mutually exclusive and exhaustive values. To make a DAG into a Bayesian network, one more step is required: we need to specify the prior probabilities for the variables in the root nodes and the conditional probabilities for the variables in all other nodes, given any combination of values of the variables in their respective parent nodes. The arrows in a Bayesian network carry information about the independence relations between the variables in the network. This information is expressed by the Parental Markov Condition: *A variable represented by a node in the Bayesian network is independent of all variables represented by its non-descendent nodes, conditional on all variables represented by its parent nodes.* In the causal modeling literature, this condition is called the Causal Markov Condition.

4. The Coherence Theory of Justification. Confronted with the Cartesian skeptic, coherentists point out that when our belief systems hang together well, with their different parts supporting each other, then this is an indication of the truth of the systems (BonJour 1985). However, the corresponding theory—the coherence theory of justification—suffers from several problems. Here are two. First, the theory is vague, as it is difficult to make precise what coherence is. Second, coherence is not necessarily truth-conducive. For example, fairy tales are made up, although the stories they tell may be highly coherent. Hence, the coherence of a set of propositions is at best truth-conducive *ceteris paribus*. But what goes in the ceteris paribus clause? This question is hard to address if we only have the toolbox of traditional epistemology. Bayesians can be of real help here. They have proposed and analyzed various measures of coherence and analyzed in detail under which conditions, if at all, coherence is truth-conducive (Bovens and Hartmann 2003, Douven and Meijs 2007, Olsson 2009).

5. Sources of Knowledge/Belief. Traditional epistemology examines sources of knowledge and belief such as our senses, memory and testimony. All three have inspired Bayesian model-building. First, the uncertainty of the evidence from our senses has prompted the development of a more realistic updating rule than strict conditionalization—Jeffrey conditionalization. Second, conditionalization represents an idealized version of the epistemological role of memory—one who updates only by conditionalization never forgets—while Bayesian models of bounded rationality allow for memory loss (Mehta et al. 2004). Third, Bayesians have the resources to model the effect of combining the testimony of several witnesses (Bovens and Hartmann 2003). There is also a growing literature on self-knowledge and self-location, as exemplified by the Sleeping Beauty problem (Elga 2000), and whether 'centred' information can rationally induce changes in opinions about 'uncentred' propositions concerning how the world is.

8. Avenues for Future Research

In this closing section, we briefly point to some topics that we would like to see addressed in future research. While some of them concern the relation of Bayesian epistemology to traditional epistemology and philosophy of science, others are internal to the Bayesian program.

1. More bridges between Bayesian epistemology and traditional epistemology. There is no harm in labour being divided between the two kinds of epistemology, as we have argued—but it would be all the better if they could become more cooperative enterprises. Think of some of the time-honored debates in traditional epistemology: skepticism, the analysis of knowledge, reliabilism, internalism vs. externalism. Think of some of the currently hot topics: contextualism, subject-sensitive invariantism, contrastivism, relativism, luminosity, 'knowledge how' (as opposed to 'knowledge that'), knowledge 'wh_' (who, where, when, which), … Where are the counterpart debates in Bayesian epistemology? Going in the other direction, think of some of the time-honored debates in Bayesian epistemology: constraints on priors, updating rules, the extension of subjective probabilities to infinite spaces. And think of some currently hot topics: credences about chances (as in the Principal Principle), credences about one's future credences (as in the Reflection Principle), updating credences on 'centred' or 'indexical' propositions, ... Where are the counterpart debates in traditional epistemology? Each of these topics suggests a bridge waiting to be built.

To some extent, such progress awaits a better understanding of the relationship between traditional claims about belief/knowledge and Bayesian claims about degrees of belief, which is still controversial. As we have said, we are not sanguine about the prospects of a reduction in either direction, although reduction is surely not the only way to offer illumination. And even if the two epistemologies continue on separate tracks, still developments in one can provide inspiration or heuristic guidance for the other.

2. More bridges between Bayesian epistemology and philosophy of science. Bayesianism started as a confirmation theory. And indeed, the formal machinery to address confirmation-theoretical questions is highly developed. Our ambitious goal, however, should be to develop a full-fledged Bayesian philosophy of science. Here is an incomplete list of questions that should be addressed to this end:

(i) Which stance does Bayesianism take in the realism debate? Is it neutral to the debate, or does it favor a version of scientific realism or antirealism? (Douven 2005, Earman 1996)

(ii) Can scientific theory change be understood in Bayesian terms? (Earman 1996)

(iii) Can Bayesianism help to characterize the overall structure of science? Is it epistemically advantageous to aim for unified theories? (Myrvold 2003)

(iv) Can a Bayesian reading of Inference to the Best Explanation be given? Lipton (2004) argues that explanatory considerations are encoded in the likelihoods, and not in the priors.

(v) How can scientific idealizations be understood in Bayesian terms? Addressing this question is important as idealizations are ubiquitous in science. The trouble begins when we attach a prior probability of zero to an idealized (hence false) statement. The posterior is then arguably also zero, which arguably renders Bayesianism useless.

3. Bayesian Social Epistemology. Bayesian epistemology, as we have presented it so far, shares one important feature with traditional epistemology: it is individualistic, i.e. it is concerned with one agent, who has beliefs and who updates her beliefs in the light of new evidence. However, the doxastic unit could well be a community comprising several individuals, or more. Kuhn (1962) argued that it is the entire scientific community that accepts or rejects a paradigm. Or think of a jury that has to come up with a consensual verdict in a murder case. In light of examples such as these, a new field—social epistemology—has been established (Goldman 1999, Kitcher 1993). While much work in this field is informal, formal tools have recently been developed that address issues in social epistemology. Especially noteworthy is the work on judgment aggregation (List and Puppe 2009) that comprises investigations inspired by the discursive dilemma. Bovens and Rabinowicz (2006) and Hartmann and Sprenger (2009) have given epistemic analyses of various aggregation rules studied in this context. It is hoped that this work will eventually develop into a full Bayesian account of group judgment and group decision-making. Other topics of current interest include the debate about rational disagreement (Feldman and Warfield 2009) and, related to this, theories about consensus and compromise formation.

To sum up: Bayesian epistemology is an exciting and thriving research program. There's plenty more work for Bayesians to do![1]

References

Armendt, B. (1980). Is There a Dutch Book Argument for Probability Kinematics? *Philosophy of Science* 47: 583-588.

BonJour, L. (1985). *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.

Bovens, L. and S. Hartmann (2003). *Bayesian Epistemology*. Oxford: Oxford University Press.

Bovens, L. and W. Rabinowicz (2006). Democratic Answers to Complex Questions: An Epistemic Perspective. *Synthese* 150: 131-153.

Chater, N. and M. Oaksford (eds.) (2008). *The Probabilistic Mind: Prospects for Bayesian Cognitive Science.* Oxford: Oxford University Press.

Crupi, V., K. Tentori and M. Gonzalez (2007). On Bayesian Measures of Evidential Support: Theoretical and Empirical Issues. *Philosophy of Science* 74(2):

[1] We thank John Cusbert, Kevin Korb, Ralph Miles, Jonah Schupbach, Declan Smithies and Jan Sprenger for very helpful comments.

229-252.

Crupi, V., B. Fitelson and K. Tentori (2008). Probability, Confirmation and the Conjunction Fallacy. *Thinking and Reasoning* 14(2): 182-199.

Diaconis, P. and S. L. Zabell (1982). Updating Subjective Probability. *Journal of the American Statistical Association* 77(380): 822-830.

Douven, I. (2005). Wouldn't it Be Lovely: Explanation and Scientific Realism, *Metascience* 14: 338-343.

Douven, I. and W. Meijs (2007). Measuring Coherence. *Synthese* 156: 405-425.

Earman, J. (1996). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, Mass.: MIT Press.

Edwards, W., Lindman, H., and Savage, J. (1963). Bayesian Statistical Inference for Psychological Research. *Psychological Review* 70: 193-242.

Eells, E. and B. Fitelson (2000). Measuring Confirmation and Evidence. *Journal of Philosophy* 97: 663-672.

Elga, A. (2000). Self-Locating Belief and the Sleeping Beauty Problem. *Analysis* 60(2): 143-47.

Eriksson, L. and A. Hájek (2007). What Are Degrees of Belief? *Studia Logica* 86, Special Issue Formal Epistemology I, B. Fitelson (ed.), 183-213.

Feldman, R. and T. Warfield (eds.) (2009). *Disagreement*. Oxford: Oxford University Press.

Fitelson, B. (1999). The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity. *Philosophy of Science* 66 (Proceedings): S362-78.

Franklin, A. (1990). *Experiment, Right or Wrong*. Cambridge: Cambridge University

Press.

Gaifman, H. and M. Snir (1982). Probabilities over Rich Languages. *Journal of Symbolic Logic* 47: 495-548.

Gärdenfors, P. and H. Rott (1995). *Belief Revision*, in *Handbook of Logic in Artificial Intelligence and Logic Programming. Volume IV: Epistemic and Temporal Reasoning*, eds. D.M. Gabbay, C.J. Hogger und J.A. Robinson. Oxford: Oxford University Press, 35-132.

Glymour, C. (1980). *Theory and Evidence*. Princeton: Princeton University Press.

Goldman, A. (1999). *Knowledge in a Social World*. Oxford: Clarendon Press.

Good, I. J. (1971). Letter to *The American Statistician* 25: 62-63.

Hájek, A. and N. Hall (2002). Induction and Probability. In *The Blackwell Guide to the Philosophy of Science,* P. Machamer and M. Silberstein (eds.). Oxford: Blackwell, 149-172.

Hartmann, S. and J. Sprenger (2009). Judgment Aggregation and the Problem of Tracking the Truth. *Synthese*.

Howson, C. and P. Urbach (2007). *Scientific Reasoning: The Bayesian Approach.*

Jaynes, E. (2003). *Probability Theory: The Logic of Science.* Cambridge: Cambridge University Press.

Jeffrey, R. (1983). *The Logic of Decision*. 2nd ed. Chicago: University of Chicago Press.

Jeffrey, R. (1992). Probable Knowledge. In *Probability and the Art of Judgment*, Cambridge: Cambridge University Press.

Joyce, J. (1998). A Nonpragmatic Vindication of Probabilism. *Philosophy of Science* 65: 575-603.

Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Joyce, J. (2008). Bayes' Theorem. *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, E. Zalta (ed.).

Kitcher, P. (1993). *The Advancement of Science. Science without Legend, Objectivity without Illusions*. New York: Oxford University Press.

Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer. English translation: *Foundations of the Theory of Probability*. New York: Chelsea 1950.

Korb, K. and A. Nicholson (2004). *Bayesian Artificial Intelligence*. Boca Raton: CRC Press.

Kuhn, T. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Lewis, D. (1980). A Subjectivist's Guide to Objective Chance. In R. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, Vol II., Berkeley: University of California Press, 263-293.

Lewis, D. (1999). Why Conditionalize? In *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lipton, P. (2004). *Inference to the Best Explanation*. London: Routledge.

List, C. and C. Puppe (2009). Judgment Aggregation: A Survey. In: P Anand, C. Puppe and P. Pattanaik (eds.), *Oxford Handbook of Rational and Social Choice*. Oxford: Oxford University Press.

Mehta, N., S. Rajiv and K. Srinivasan (2004). Role of Forgetting in Memory-Based Choice Decisions: A Structural Model. *Quantitative Marketing and*

Economics 2(2), 107-140.

Myrvold, W. (2003). A Bayesian Account of the Virtue of Unification. *Philosophy of Science* 70**:** 399-423.

Neapolitan, R. (2003). *Learning Bayesian Networks*. Prentice Hall.

Nozick, R. (1981). *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

Oaksford, M. and N. Chater (2007). *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.

Olsson, E. (2009). *Against Coherence: Truth, Probability, and Justification*. Oxford: Oxford University Press.

Pearl, J. (2000). *Causality*. New York: Cambridge University Press.

Ramsey, F. P. (1931). Truth and Probability. In R. B. Braithwaite (ed.), *Foundations of Mathematics and other Essays*. London: Routledge, 156-98; reprinted in D. H. Mellor (ed.), *F. P. Ramsey: Philosophical Papers*, Cambridge: Cambridge University Press 1990.

Savage, L. J. (1954). *The Foundations of Statistics.* John Wiley.

Shogenji, T. (2009). The Degree of Epistemic Justification and the Conjunction Fallacy. *Synthese*.

Sosa, E. (1999). How to Defeat Opposition to Moore. *Philosophical Perspectives 13*: 141-153.

Spirtes, P., C. Glymour, and R. Scheines (2001). *Causation, Prediction and Search*. Cambridge, Mass.: MIT Press.

Tversky, A. and D. Kahneman (1983). Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review* 90:

293-315.

van Fraassen, B. (1984). Belief and the Will. *Journal of Philosophy* 81: 235-256.

Williamson, J. (2005). *Bayesian Nets and Causality: Philosophical and Computational Foundations*. Oxford: Oxford University Press.

Zynda, L. (1995). Old Evidence and New Theories. *Philosophical Studies* 77: 67-95.