



Kent Academic Repository

Williamson, Jon (2022) *Bayesianism from a philosophical perspective and its application to medicine*. International Journal of Biostatistics, 19 (2). pp. 295-307. ISSN 1557-4679.

Downloaded from

<https://kar.kent.ac.uk/99017/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1515/ijb-2022-0043>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

For the purpose of open access, the author(s) has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).



Jon Williamson*

Bayesianism from a philosophical perspective and its application to medicine

<https://doi.org/10.1515/ijb-2022-0043>

Received April 6, 2022; accepted October 3, 2022; published online December 9, 2022

Abstract: Bayesian philosophy and Bayesian statistics have diverged in recent years, because Bayesian philosophers have become more interested in philosophical problems other than the foundations of statistics and Bayesian statisticians have become less concerned with philosophical foundations. One way in which this divergence manifests itself is through the use of direct inference principles: Bayesian philosophers routinely advocate principles that require calibration of degrees of belief to available non-epistemic probabilities, while Bayesian statisticians rarely invoke such principles. As I explain, however, the standard Bayesian framework cannot coherently employ direct inference principles. Direct inference requires a shift towards a non-standard Bayesian framework, which further increases the gap between Bayesian philosophy and Bayesian statistics. This divergence does not preclude the application of Bayesian philosophical methods to real-world problems. Data consolidation is a key challenge for present-day systems medicine and other systems sciences. I show that data consolidation requires direct inference and that the non-standard Bayesian methods outlined here are well suited to this task.

Keywords: Bayesian networks; Bayesianism; direct inference; formal epistemology; objective Bayesianism; systems medicine.

Bayesianism is not a homogeneous theory. While Good [1, Chapter 3] counted 46,656 varieties of Bayesian, here I distinguish two: the Bayesian philosopher and the Bayesian statistician. As I explain in Section 1, their interests have diverged, and so have their resulting theories. Bayesian philosophers, in particular, see an important role for explicit, general direct inference principles in Bayesian theory, while Bayesian statisticians do not, by and large. I draw on recent work on direct inference to argue in Section 2 that this emphasis on direct inference, although well motivated, is in tension with the standard Bayesian tenet that conditional degrees of belief are conditional probabilities. Accommodating direct inference requires relaxing this tenet and moving to a slightly non-standard version of Bayesianism (Section 3). This move increases the divergence between the two Bayesian approaches. In Section 4, I explain how this non-standard version of Bayesianism can be applied to systems medicine, which needs to consolidate multiple datasets that only partially overlap in terms of the variables they measure.

1 Bayesianism from a philosophical perspective

Bayesian probabilities differ from frequencies in two main ways. Firstly, Bayesian probabilities are single-case. While a limiting relative frequency attaches to a potentially infinite reference class of outcomes, such as rolls of a specific die, a Bayesian probability attaches to a single outcome, such as the proposition that a particular roll of a specific die yields a 5. Second, Bayesian probabilities are epistemic. While a frequency is a proportion

*Corresponding author: Jon Williamson, Department of Philosophy and Centre for Reasoning, University of Kent, Canterbury, UK, E-mail: j.williamson@kent.ac.uk. <https://orcid.org/0000-0003-0514-4209>

of outcomes that have a certain attribute, a Bayesian probability measures an agent's rational degree of belief in a certain proposition.¹

We shall use $B_E(A)$ to denote the degree to which the agent in question believes A , supposing only E . Here A might be the proposition that the die's next roll yields a 5 and E might be some supposition about the die and the way it is thrown. Under the standard conception of Bayesianism, this conditional belief is a conditional probability:

CBCP. There is a prior probability function P_\emptyset such that $B_E(A) = P_\emptyset(A|E)$ for all A and E .

Here the prior function P_\emptyset encompasses the agent's degrees of belief in the absence of E .

When E is the agent's current evidence, CBCP motivates the principle of Bayesian conditionalisation, which says that current degrees of belief are obtained by conditionalising on E .

Conditionalisation. On evidence E , believe A to degree $B_E(A) = P_\emptyset(A|E)$.

Thus, all the agent's degrees of belief are determined by a single prior probability function P_\emptyset . Subjective Bayesianism holds that any prior is rationally permissible, while objective Bayesianism takes only certain priors to be appropriate.

For much of the 20th century, Bayesianism, though a fringe interest, had a common focus. Bayesians, whether in statistics, philosophy or the sciences, were largely interested in the foundations of statistical inference, e.g., the foundations of predictive inference and model selection. The rivalry between frequentism and Bayesianism was fierce, with each claiming to provide more cogent foundations for statistical inference.² The common focus and common foe gave the Bayesian community a sense of cohesiveness.

Since the 1990s, however, Bayesianism has moved into the mainstream, thanks in no small measure to advances in computational methods [3]. Being less preoccupied with the fight against frequentism, a divergence has taken place between Bayesian statistics and Bayesian philosophy. In Bayesian statistics, on the one hand, there has been a trend towards a pragmatic approach, with statistical methods being chosen on the basis of their inferential utility rather than their philosophical foundations [4]. This represents a move away from philosophy. In philosophy, on the other hand, Bayesians have taken up different questions, notably the application of Bayesianism to epistemology [5–7, Chapter 2] and to inductive logic [8–10]. Examples of questions of recent interest include: Does Bayesianism validate the claim that a greater variety of evidence is more confirmatory, *ceteris paribus* [11]? Can Bayesianism accommodate the use of old evidence to provide confirmation [12]? Can Bayesianism tell us how individual beliefs should be aggregated to form group beliefs [13]? Can Bayesianism measure how well beliefs cohere with one another [14]? Can Bayesianism accommodate growth in the space of possibilities under consideration [15]? While the use of Bayesianism to evaluate statistical hypotheses remains a topic of inquiry, it is now very much a minority interest in philosophy. Most philosophers are interested in Bayesianism insofar as it provides an account of rational belief.

One key way in which this divergence is manifested is in the use of explicit direct inference principles. The question arises as to the relationship between rational degrees of belief and non-epistemic probabilities such as frequencies. For example, how strongly should one believe that a die will yield a 5 or a 6 on the next roll given just that the limiting relative frequency of 5 or 6 in similar rolls of that particular die is $\frac{2}{3}$? Most Bayesian philosophers would say that one's degree of belief should be $\frac{2}{3}$ here: the die is heavily biased in favour of 5 or 6, and one's degree of belief should track this bias [7, §2.3]. But there is nothing in the machinery of Bayesianism as formulated above that picks out this degree of belief as uniquely rational. In order to deem the degree of belief $\frac{2}{3}$

¹ One point of divergence between Bayesian philosophy and Bayesian statistics surrounds the question whether probabilities attach to events or propositions. Bayesian philosophers tend to say that Bayesian probabilities attach to propositions, because Bayesian probabilities are degrees of belief and propositions are the objects of belief. Bayesian statisticians, on the other hand, tend to talk in terms of events, as that was the usual terminology during the development of the mathematical theory of probability. Not a lot hangs on this difference in terminology, as talk of events can be translated into talk of propositions (one can consider the proposition that an event obtains) and talk of propositions can be translated into talk of events (one can consider the event that the proposition is true, or the event that it is false). Hence I will not dwell further on this aspect of the divergence between Bayesian philosophy and Bayesian statistics.

² See Mayo [2] for a frequentist take on the "statistics wars".

to be rationally required, one needs to augment standard Bayesianism with an explicit direct inference principle that forces calibration of rational degrees of belief to available non-epistemic probabilities. One such direct inference principle is a version of the Principle of the Narrowest Reference Class, which requires calibration to frequencies in the narrowest reference class from which one has reliable statistics [16]:

PNRC. If X says that the frequency of attribute α in reference class ρ is x , and A says that attribute α holds of a particular member c of that reference class, and E is compatible with X and admissible, then $P_{\theta}(A|XE) = x$.

For E to be admissible here, it should contain no information about A as pertinent as X —in particular, no frequency of α in some other reference class that contains c and is as narrow as ρ .

Another direct inference principle, Lewis' Principal Principle, calibrates degrees of belief to single-case chances, rather than frequencies of repeatedly instantiable attributes [17]:³

PP. If X says that the chance of A is x , and E is any proposition that is compatible with X and admissible, then $P_{\theta}(A|XE) = x$.

For instance, given that the chance of the die yielding 5 or 6 on the next roll is $\frac{2}{3}$ and any other information that is compatible and admissible, one should believe that the die will yield 5 or 6 on the next roll to degree $\frac{2}{3}$.

Bayesian philosophers tend to endorse one or other of these two direct inference principles, or some variant of these principles.⁴ Direct inference is sometimes justified on the grounds that repeated infringements of a direct inference principle would expose an agent to eventual loss, if she were to bet according to her degrees of belief [22, §3.3].

Bayesian statisticians, on the other hand, do not tend to endorse an explicit, general direct inference principle. This is not to say that Bayesian statisticians deny the rationality of calibrating of Bayesian probabilities to non-epistemic probabilities. Indeed, in the standard Bayesian framework it will be permissible to do so, and Bayesian statisticians regularly draw on non-epistemic probabilities when devising a class of models to consider or when choosing a prior probability function. But they usually do so implicitly and on a case-by-case basis, without recourse to a general direct inference principle. For Bayesian philosophers, calibration is rationally required, while for Bayesian statisticians it is merely permissible.

There are several possible explanations for the dearth of explicit direct inference principles in Bayesian statistics. One is the influence of Bruno de Finetti, a pioneer of subjective Bayesianism, who denied the existence of the non-epistemic probabilities that are presupposed by direct inference principles [23]. Philosophers, on the other hand, tend to accept both epistemic and non-epistemic probabilities, perhaps under the influence of another pioneer of Bayesianism, Frank Ramsey, who accepted frequencies and direct inference [24, p. 50], [25]. A second possible explanation is the rivalry between Bayesianism and frequentism noted above: some Bayesian statisticians might view calibration to frequencies as too much of a concession to the frequentist. This rivalry has little current influence on Bayesian philosophers, however, who are usually more interested in rational belief than statistical methods. A third potential explanation is the fact that there are conditions under which degrees of belief are guaranteed to converge to frequencies in the long run, and this fact might seem to some to obviate the need for an explicit direct inference principle. These convergence results tend to be unsatisfying to philosophers due to their asymptotic nature and the fact that they only hold under certain conditions [26, Chapter 6]. Arguably, calibration to non-epistemic probabilities is required in the short run, and universally.

³ See Gillies [18, Chapters 5–7] for an introduction to the difference between the frequency interpretation of probability, which attaches probabilities to an infinite collective of outcomes, such as rolls of a die, and the propensity or chance interpretation, which attaches probabilities to a single outcome, such as a particular roll of a die.

⁴ In many other respects, Bayesian philosophers hold very diverse views. For example, they develop Bayesianism in many different ways and they differ with regard to fundamental questions, such as the relation between degrees of belief and outright beliefs [e.g., 19–21]. But a common interest in elucidating the relation between evidence and rational degree of belief has led to a surprising level of unity on the question of direct inference.

2 A problem for standard Bayesianism

In this Section 1 present a problem for the use of direct inference in a standard Bayesian framework. For further discussion of this and other problems for direct inference, see Wallmann and Williamson [27]; Wallmann and Hawthorne [28] and Williamson [29, §4].

Consider a die with colours as well as numbers on its faces—red or black, say. Suppose X is the proposition that the frequency of obtaining a red outcome in throws of this particular die is $\frac{2}{3}$ and that R says that the next throw will yield a red outcome. Suppose the remaining evidence E is compatible with X and admissible and, in particular, contains no information pertaining to whether the die is biased with respect to number or colour. Then PNRC would force:

$$P_{\emptyset}(R|XE) = \frac{2}{3} \quad (1)$$

Suppose H is the proposition that the throw yields a high score, i.e., that the numerical outcome of the throw is a 5 or a 6. In a Bayesian setting it will at least be rationally permissible to adopt the following degree of belief:

$$P_{\emptyset}(H|XE) = \frac{1}{3} \quad (2)$$

This assignment might be motivated by indifference: two out of six possible outcomes are favourable to H . Or it might be motivated by considering a wider reference class of throws of dice selected at random and reasoning that fair dice give frequency $\frac{1}{3}$ to 5 or 6 and that there are no grounds to suppose dice biased in favour of 5 or 6 are any more prevalent than dice biased against 5 or 6.

Now suppose one learns that red appears on faces 5 and 6 of the die and only on those two faces, so R and H have the same truth value, $R \leftrightarrow H$. In the light of this information, the Bayesian formalism requires that R and H be given the same degree of belief.⁵ Given this, it should at least be rationally *permissible* that the frequency information pertaining to R , which warrants degree of belief $\frac{2}{3}$, has more of a bearing on one's degree of belief in R than the subjective inclination to believe H to degree $\frac{1}{3}$:

$$P_{\emptyset}(R|XE(R \leftrightarrow H)) > \frac{1}{2} \quad (3)$$

Indeed, PNRC requires that $P_{\emptyset}(R|XE(R \leftrightarrow H)) = \frac{2}{3}$, because X specifies the value $\frac{2}{3}$ in the narrowest reference class for which frequency information is available.

The standard Bayesian framework faces a problem, however: Eqs. (1)–(3) are inconsistent, as is shown in the Appendix to this paper. Thus PNRC cannot coherently be implemented in standard Bayesianism.

This problem is not limited to PNRC—it applies equally to the Principal Principle. If we take X to say that the chance of R is $\frac{2}{3}$ then Eq. (1) is forced by PP. It remains rationally permissible to believe that the roll of the die will yield a 5 or 6 to degree $\frac{1}{3}$ (Eq. (2)), and to be influenced more by the chance of R than the subjective inclination to believe H to degree $\frac{1}{3}$, on learning that R and H have the same truth value (Eq. (3)). However, Eqs. (1)–(3) are inconsistent. Thus, neither direct inference principle can be properly implemented in the standard Bayesian framework.

The puzzle surrounding Eqs. (1)–(3) is symptomatic of the more general problem that the standard Bayesian framework fails to accommodate judgements about strength of evidence. In the case of the Principle of the Narrowest Reference Class, we need that frequencies in narrower reference classes are stronger determinants of degrees of belief than frequencies in wider reference classes, but such a requirement conflicts with the standard Bayesian framework. In the case of the Principal Principle, we need that chances are stronger determinants of degrees of belief than less informed subjective opinions, but such a requirement also conflicts with the standard Bayesian framework.

⁵ For any probability function P such that $P(R \leftrightarrow H) > 0$, $P(R|R \leftrightarrow H) = P(RH|R \leftrightarrow H) + P(R\bar{H}|R \leftrightarrow H) = P(RH|R \leftrightarrow H) + 0 = P(RH|R \leftrightarrow H) + P(\bar{R}H|R \leftrightarrow H) = P(H|R \leftrightarrow H)$. In the above example, it is clearly rationally permissible that $P(R \leftrightarrow H) > 0$.

3 An objective Bayesian resolution

The Bayesian philosopher is in a quandary. On the one hand, direct inference seems to be required for degrees of belief to count as rational. On the other, even simple direct inferences turn out to lead to inconsistency in the standard Bayesian framework. This problem motivates a move away from the standard Bayesian framework. In this Section 1 will describe a version of objective Bayesianism that can coherently accommodate direct inference while retaining as much as possible of the standard framework. While this move resolves the problem of Section 2, it does serve to increase the divergence of Bayesian philosophy from Bayesian statistics, which tends to retain the standard Bayesian framework.

The inconsistency of Section 2 can be diagnosed as follows [see [29]]. Recall from Section 1 that CBCP requires that there be a probability function P_\emptyset such that $B_E(A) = P_\emptyset(A|E)$ for all A and E . Although CBCP is a key component of the standard Bayesian framework, it imposes very strong constraints: every potential degree of belief needs to be encoded in a single prior probability function P_\emptyset . The use of direct inference, in the shape of PNRC or PP together with the requirement that stronger evidence should have more of a bearing on belief than weaker evidence, imposes many further constraints on this prior function. The problem is that, when taken together, all these constraints lead to inconsistency.

One of CBCP and direct inference must go. Direct inference is well motivated, philosophers argue. Hence, the inconsistency tells against CBCP.

Fortunately, there is a natural alternative to CBCP:

CBP. For any E , there is a probability function P_E such that $B_E(A) = P_E(A)$ for all A .

CBP takes conditional beliefs to be probabilities, but not conditional probabilities. CBP is much less restrictive than CBCP as it does not require that each of the functions P_E be reducible to a single function P_\emptyset .

Under this conception of conditional degrees of belief, our two direct inference principles can be formulated as follows:

PNRC'. If X says that the frequency of attribute α in reference class ρ is x , and A says that attribute α holds of a particular member c of that reference class, and E is compatible with X and admissible, then $P_{XE}(A) = x$.

PP'. If X says that the chance of A is x , and E is any proposition that is compatible with X and admissible, then $P_{XE}(A) = x$.

Now Eqs. (1)–(3) become:

$$P_{XE}(A) = \frac{2}{3} \quad (4)$$

$$P_{XE}(F) = \frac{1}{3} \quad (5)$$

$$P_{XE(A \leftrightarrow F)}(A) > \frac{1}{2} \quad (6)$$

While Eqs. (4) and (5) constrain the same probability function, P_{XE} , Eq. (6) constrains a different probability function $P_{XE(A \leftrightarrow F)}$. Thus no inconsistency arises: Eq. (6) cannot conflict with Eqs. (4) and (5), which are themselves mutually consistent.

So far, so good: direct inference can be coherently integrated into this modified Bayesian framework. But by rejecting CBCP we also lose Bayesian Conditionalisation, which is framed in terms of conditional probabilities (Section 1). An alternative means of determining $B_E(A)$ is provided by the Maximum Entropy Principle [30]:

MaxEnt. On evidence E , believe A to degree $B_E(A) = P_E^\dagger(A)$, where P_E^\dagger is the probability function, from all those that satisfy constraints imposed by E , that has maximum entropy, if there is such a maximum.

The entropy of a probability function P is defined as $H(P) \stackrel{\text{df}}{=} - \sum_{\omega \in \Omega} P(\omega) \log P(\omega)$. Here Ω is taken to be a finite partition of elementary outcomes, but the approach has also been extended to infinite domains.⁶

⁶ Williamson [10] extends this approach to propositions of a first-order predicate language. In such a framework, there are uncountably many elementary outcomes. Others have extended MaxEnt to uncountable domains while remaining within the standard Bayesian framework. See Jaynes [31], for example.

One can update degrees of belief simply by reapplying MaxEnt to new evidence. This method typically gives the same results as Conditionalisation in cases where the new evidence is expressible as a proposition in the domain [22, 32–34, Chapter 4]. The use of MaxEnt leads to a kind of objective Bayesianism—a version of Bayesianism in which there are strong constraints on degrees of belief even in the absence of evidence.⁷ MaxEnt can be justified on the grounds that maximising entropy minimises worst-case expected loss, when the losses incurred by one’s actions are logarithmic [35]. Logarithmic losses are the natural default choice, in the absence of any precise information about the losses that might actually be incurred [10, Chapter 9]. Where there is such information, another loss function, and hence another entropy function, may be more appropriate [36].

Let us recap. We have seen that Bayesian philosophy and Bayesian statistics have diverged. Bayesian statisticians became more pragmatic and less wedded to philosophical foundations. Bayesian philosophers, on the other hand, became interested in problems other than the foundations of statistical inference, and these problems led to a prominent role for direct inference. If we follow the latter route to its logical conclusion, we find a conflict between standard Bayesianism and direct inference, and hence a need to move to a non-standard Bayesian framework, such as the objective Bayesian approach outlined in this section. This move increases the gap between Bayesian philosophy and Bayesian statistics.⁸

We will see next that this kind of divergence is not necessarily a bad thing, because a diversity of approaches can lead to new solutions to important problems. For example, the objective Bayesian philosophical approach can lead to new ways of tackling important problems in medicine.

4 Application to medicine

While the evaluation of statistical hypotheses remains an important task, many of the scientific problems we face today have a different flavour: they are *data consolidation* tasks. A data consolidation task seeks to produce coherent models from evidence that is typically both very extensive and very heterogeneous. It is also common that these tasks are carried out by research projects with multiple goals, and that these different goals require different kinds of model. Thus these tasks seek to consolidate big data and diverse data by constructing multiple models. Data consolidation tasks are to be found in systems medicine, for example.

Systems medicine is related to systems biology. Systems biology studies systems of molecules and their causal interactions within the cell using data-intensive functional genomics techniques such as transcriptomics, metabolomics and proteomics [38]. Systems medicine applies systems biology to medicine. Systems medicine has two kinds of goal. In common with systems biology it has a theoretical goal, namely to discover pathophysiological mechanisms [39, 40]. But in common with medicine it also has a practical goal: diagnosis, prognosis and

⁷ Although MaxEnt was championed by Edwin Jaynes in the context of the standard Bayesian framework [30, 31], it is worth emphasising that it is used here in a non-standard Bayesian framework that appeals to CBP rather than CBCP. This is so as to properly incorporate direct inference into the framework. Jaynes rejected non-epistemic probabilities and hence rejected direct inference. While this version of objective Bayesianism imposes strong constraints on degrees of belief, it does not require that there is always a unique rational belief function [22, §9.3.1]. For example, if the evidence is just that a coin is biased in favour of tails, then this evidence yields the constraint $P_E(T) > 1/2$, where T is the proposition that the next toss of the coin will yield tails. In this case, no probability function that satisfies this constraint has maximum entropy: the nearer $P_E(T)$ is to $\frac{1}{2}$, the greater the entropy of P_E .

⁸ The gap between objective Bayesian philosophy and objective Bayesian statistics is wider still. This is because the objective Bayesian approach in statistics deviates from the standard framework in respects other than those considered above. Most notably, many objective Bayesian statisticians use improper priors—i.e., priors that are not probability functions. In addition, some objective Bayesian statisticians have abandoned the interpretation of Bayesian priors as degrees of belief, which represents another significant departure from standard Bayesianism [37, p. 160]. Thus there is a very large gap between the objective Bayesian philosophical position described in this section, which requires probability functions and which retains the epistemic interpretation of these functions, and the objective Bayesian approach in statistics, which often does not. See Mayo [2, Chapter 6] for criticisms of the objective Bayesian approach in statistics and its efforts to inform priors by appeal to frequencies. It is important to emphasise that the above objective Bayesian philosophical position is not intended to provide a general foundation for statistical inference—it is intended to provide an account of rational degree of belief.

treatment of individuals [41]. Because of this practical goal, systems medicine appeals to high-level clinical and environmental data in addition to the functional genomics datasets of systems biology.

This wealth of data poses a formidable data consolidation challenge [42, 43]. Large systems medicine projects often use many datasets to produce models for the purposes of diagnosis and prognosis (these prediction problems require a model of the associations in the data), predicting the effects of interventions (which requires a causal model) and explanation (which requires a mechanistic model). The relevant datasets typically overlap little: few variables are measured by more than one dataset. The challenge, then, is to construct a model that connects all the variables of interest, when no dataset measures them all together.

The data consolidation task can be thought of as a variety of “statistical matching” problem [44]. The problem is to produce a model that matches the datasets but extrapolates beyond the data to model the domain as an integrated whole. Consider a data consolidation task which seeks a model for the purposes of prediction (e.g., diagnosis and prognosis). Suppose the consolidation task has n datasets, DS_1, \dots, DS_n , measuring sets of variables V_1, \dots, V_n respectively, and that these datasets are consistent in the sense that the marginal distributions P_1^*, \dots, P_n^* determined by the datasets are satisfiable by some joint probability distribution defined over $V \stackrel{\text{df}}{=} V_1 \cup \dots \cup V_n$. Suppose further that each dataset is large enough that one is prepared to use the associated data distribution P_i^* as an estimate of the chance distribution on V_i in the underlying population. The data consolidation prediction task requires producing a model of some suitable joint distribution over V that matches the marginal distributions P_1^*, \dots, P_n^* . This joint distribution is then used for the purposes of prediction: for example, to diagnose the condition that is most probable given a patient’s clinical, genomic and environmental observations; or to prognose the outcome that is most probable given the patient’s condition and other observations.

From a Bayesian point of view, this is a direct inference problem. There is information about non-epistemic probabilities, encapsulated in P_1^*, \dots, P_n^* , and the task is to calibrate degrees of belief to these probabilities and determine a suitable belief function on V as a whole with which to draw reasonable predictions.

As we have seen, direct inference places this problem squarely within the remit of Bayesian philosophy, but requires a move away from the standard Bayesian framework. The version of objective Bayesianism introduced in the previous section provides a means to meet the data consolidation challenge: (i) the principles of objective Bayesianism can be used to determine a probability distribution over all the variables of interest; (ii) one can then construct a graphical model to represent and reason with this probability distribution, in order to perform predictive tasks such as diagnosis and prognosis. This approach proceeds as follows.

Firstly, objective Bayesianism can be used to determine a probability distribution over all the variables of interest. By direct inference, degrees of belief ought to be calibrated to the data distributions: $P_{E|V_i} = P_i^*$ for $i = 1, \dots, n$, where E is the available evidence, namely the n datasets, P_E is the belief function defined over the entire domain V , and $P_{E|V_i}$ is its restriction to the set V_i of variables measured by dataset DS_i . MaxEnt then requires that $P_E = P_E^\dagger$, the probability function, from all those that satisfy these constraints imposed by direct inference, that has maximum entropy, where the entropy is defined on the set Ω of possible assignments of values to all the variables in V . These constraints are consistent, closed and convex, and the entropy function is strictly concave, so there is guaranteed to be a unique entropy maximiser P_E^\dagger . According to objective Bayesianism, this is the function that one ought to use for prediction tasks such as diagnosis and prognosis.

The second step is to construct a graphical model that can be used to represent and reason with P_E^\dagger . A Bayesian network is often the model of choice for representing and reasoning with a probability distribution. This is specified by providing: (i) a directed acyclic graph whose nodes are the variables over which the probability distribution is defined and which represents the conditional probabilistic independence relationships satisfied by the distribution, and (ii) the probability distribution of each variable conditional on its parents in the graph. A Bayesian network that represents the maximum entropy function P_E^\dagger that is recommended by objective Bayesianism is called an *objective Bayesian net* or OBN [45].

One can construct an OBN from n consistent datasets using the following procedure [46, 47]. From each dataset learn the graphical structure of a Markov net that represents the dataset distribution P_i^* . (A Markov net is similar to a Bayesian net, but utilises an undirected graph instead of a directed acyclic graph to represent

Table 2: One genomic dataset of 502 cases from the Progenetix database.

1p31	1p32	1p34	2q32	3q26	4q35	5q14	7p11	8q23	20p13	Xp11	Xq13
0	0	0	1	-1	0	0	1	0	0	0	-1
0	0	1	1	0	0	0	-1	-1	0	0	0
...

Table 3: A further genomic dataset (119 cases with clinical annotation) from the Progenetix database.

Lymph nodes	1q22	1q25	1q32	1q42	7q36	8p21	8p23	8q13	8q21	8q24
0	1	1	1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
...

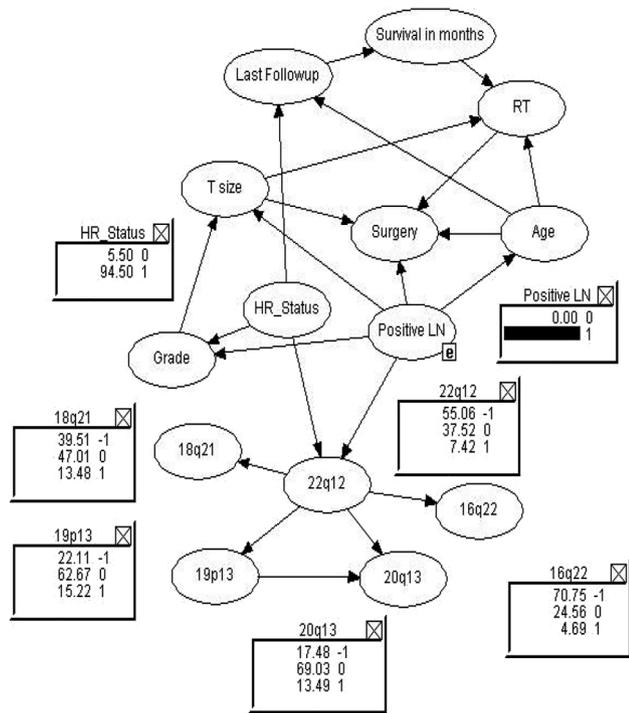


Figure 2: The graphical structure of the OBN produced by Nagl et al. [48].

We see, then, that although the non-standard version of objective Bayesianism of Section 3 is not intended to provide general foundations for statistical inference, it can nevertheless be applied to the data consolidation task, which is arguably one of the central challenges of our time and the key challenge for systems medicine. Arguably, this sort of approach is exactly the right way to solve a data consolidation prediction problem, for the following reasons. Prediction requires probabilities defined over the domain V as a whole. These probabilities cannot be thought of as estimates of non-epistemic probabilities, because the data is simply too sparse to have any confidence that such estimates would be accurate. Hence they must be interpreted as Bayesian probabilities: rational degrees of belief to be used as a basis for action. Moreover, the data consolidation prediction task is a statistical matching problem, and statistical matching requires direct inference.

Note that a standard Bayesian approach without direct inference would proceed by setting some prior P_{θ} on V and updating it on every record in every dataset and on frequencies reported in the relevant research literature. The behaviour of that sort of approach depends crucially on the choice of prior and the framework

cannot guarantee short-run calibration to the relevant non-epistemic probabilities, i.e., it cannot ensure that the updated probability function matches the dataset distributions P_i^* . In order to guarantee such a match, we would need an explicit, general direct inference principle, but then the problem of Section 2 arises: the standard Bayesian approach cannot coherently incorporate direct inference.⁹ The non-standard objective Bayesian approach can, on the other hand. This is why it is well suited to such problems.

5 Conclusions

Philosophers are rightly keen to incorporate direct inference into Bayesian theory: we regularly calibrate our degrees of belief to non-epistemic probabilities, where we have reliable estimates of these probabilities, and we judge failures to do so as irrational. Systems medicine provides a good example of the need for direct inference: the whole approach is predicated on the idea that we should defer to the data, and—at least where the data distributions provide consistent and reliable estimates of the underlying population frequencies—to the data distributions.

I have argued that a proper treatment of direct inference requires a move away from the standard Bayesian framework that is common in statistics. This move is not a seismic shift, however: degrees of belief are still probabilities in the modified framework, and its version of updating generalises Bayesian conditionalisation, even if conditional probabilities are no longer central to the revised approach.

The proposed approach is firmly in the objective Bayesian camp. Now, objective Bayesianism has been roundly criticised for failing to produce parameterisation-invariant priors on continuous spaces [see, e.g., 54, Chapter 9]. This criticism poses an important challenge to the application of objective Bayesianism to statistical parameter estimation, where the parameters in question are often continuous and where the problem formulation often underdetermines the parameterisation. It is much less of a concern for objective Bayesianism as used in philosophy, which tends to consider probability defined on a finite, indivisible partition Ω of possible worlds, or defined on a logical language—typically a propositional or first-order predicate language. On such spaces the uniform distribution emerges as the canonical probability function that is warranted in the total absence of evidence, and there is an important sense in which inferences are language invariant [10].

That a divide has opened up between Bayesian philosophy and Bayesian statistics is not in itself a problem. Different horses suit different courses. The divergence may even prove advantageous where the two approaches can provide complementary perspectives on new problems that arise.¹⁰ As we have seen, Bayesian philosophy and Bayesian statistics provide very different approaches to the data consolidation task, although this task is arguably a problem most naturally suited to Bayesian philosophy.

Acknowledgments: I am very grateful to Deborah Mayo, Michael Wilde and the anonymous referees for helpful comments and discussion.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

⁹ See Endres and Augustin [50] for an approach within the standard Bayesian framework that uses direct inference in the form of statistical matching. The use of “probability matching priors” has also been advocated in order to ensure a sort of direct inference [see, e.g., 51, 52]. These approaches are not immune to the problem of Section 2. The irony is that in the standard Bayesian framework one ought to *believe* that one is well calibrated to frequencies or chances [53], but, as we saw in Section 2, attempts to *be* well calibrated can lead to inconsistency.

¹⁰ For another example in the medical setting, consider the assessment of safety of medicines, where Price et al. [55] pursue an approach based on Bayesian statistics and De Pretis et al. [56, 57] develop an approach based on Bayesian philosophy. The use of complementary approaches can be thought of as a form of pluralism. Chang [58, Chapter 5], for one, argues for pluralism in science; see Ludwig and Rupy [59] for a survey. Gillies and Zheng [60] show how two fields can benefit when they dynamically interact with one another.

Research funding: This research was supported by funding from the Leverhulme Trust (grants RPG-2022-336 and RPG-2019-059) and the Deutsche Forschungsgemeinschaft (DFG, grant LA 4093/3-1).

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

Appendix

Here we see that Eqs. (1)–(3) are inconsistent [54, §3.1].

Suppose Eqs. (1) and (2) hold, so $P(R|XE) = \frac{2}{3}$ and $P(H|XE) = \frac{1}{3}$, where $P = P_\emptyset$. Suppose further that Eq. (3) holds. This presupposes that $P(R|XE(R \leftrightarrow H))$ is well defined, i.e., $P(XE(R \leftrightarrow H)) > 0$.

By Bayes' theorem,

$$\begin{aligned} P(R|XE(R \leftrightarrow H)) &= \frac{P(R \leftrightarrow H|RXE)P(R|XE)}{P(R \leftrightarrow H|XE)} \\ &= \frac{P(RH \vee \bar{R}\bar{H}|RXE)P(R|XE)}{P(RH \vee \bar{R}\bar{H}|XE)} \\ &= \frac{(P(RH|RXE) + P(\bar{R}\bar{H}|RXE))P(R|XE)}{P(RH|XE) + P(\bar{R}\bar{H}|XE)} \\ &= \frac{P(H|RXE)P(R|XE)}{P(RH|XE) + P(\bar{R}\bar{H}|XE)} \\ &= \frac{P(RH|XE)}{P(RH|XE) + P(\bar{R}\bar{H}|XE)} \\ &= \frac{1}{2} \end{aligned}$$

which contradicts Eq. (3). The last identity holds because $P(R\bar{H}|XE) + P(\bar{R}\bar{H}|XE) = P(\bar{H}|XE) = \frac{2}{3} = P(R|XE) = P(RH|XE) + P(\bar{R}\bar{H}|XE)$ and hence $P(\bar{R}\bar{H}|XE) = P(RH|XE)$.

References

1. Good IJ. Good thinking: the foundations of probability and its applications. Minneapolis: University of Minnesota Press; 1983.
2. Mayo DG. Statistical inference as severe testing: how to get beyond the statistics wars. Cambridge: Cambridge University Press; 2018.
3. Lenhard J. A transformation of Bayesian statistics: computation, prediction, and rationality. *Stud Hist Philos Sci* 2022;92:144–51.
4. Kass RE. Statistical inference: the big picture. *Stat Sci* 2011;26:1–9.
5. Bovens L, Hartmann S. Bayesian epistemology. Oxford: Oxford University Press; 2003.
6. Olsson EJ. Bayesian epistemology. In: Hansson SO, Hendricks VF, editors. *Introduction to formal philosophy*. Cham: Springer; 2018:431–42 pp. chapter 22.
7. Schupbach JN. Bayesianism and scientific reasoning In: *Elements in the Philosophy of Science*. Cambridge: Cambridge University Press; 2022.
8. Howson C. Hume's problem: induction and the justification of belief. Oxford: Clarendon Press; 2000.
9. Romeyn J-W. Bayesian inductive logic [Ph.D. thesis]. Groningen: University of Groningen Faculty of Philosophy; 2005.
10. Williamson J. *Lectures on inductive logic*. Oxford: Oxford University Press; 2017.
11. Landes J. The variety of evidence thesis and its independence of degrees of independence. *Synthese* 2021;198:10611–41.
12. Hawthorne J. Degree-of-belief and degree-of-support: why Bayesians need both notions. *Mind* 2005;114:277–320.
13. Thorn PD. The joint aggregation of beliefs and degrees of belief. *Synthese* 2020;197:5389–409.
14. Herzberg F. A graded Bayesian coherence notion. *Erkenntnis* 2014;79:843–69.
15. Mahtani A. Awareness growth and dispositional attitudes. *Synthese* 2021;198:8981–97.
16. Reichenbach H. *The theory of probability: an inquiry into the logical and mathematical foundations of the calculus of probability*, 2nd ed. Berkeley and Los Angeles: University of California Press; 1935. Trans. Ernest H. Hutten and Maria Reichenbach.
17. Lewis DK. A subjectivist's guide to objective chance. *Phil Pap* 1980;2:83–132.
18. Gillies D. *Philosophical theories of probability*. London, New York: Routledge; 2000.
19. Buchak L. Belief, credence, and norms. *Phil Stud* 2014;169:285–311.

20. Clarke R. Belief is credence one (in context). *Philosophers' Impr* 2013;13:1–18.
21. Jackson EG. The relationship between belief and credence. *Philos Compass* 2020;15:e12668.
22. Williamson J. In defence of objective Bayesianism. Oxford: Oxford University Press; 2010.
23. de Finetti B. Probabilismo. *Logos* 1931;14:163–219. English translation in *Erkenntnis* 31:169–223, 1989.
24. Ramsey FP. Truth and probability. In: Kyburg HE, Smokler HE, editors. *Studies in subjective probability*, 2nd ed. Huntington, New York: Robert E. Krieger Publishing Company; 1926:23–52 pp.
25. Ramsey FP. Miscellaneous notes on probability. In: Galavotti MC, editor. *Notes on philosophy, probability and mathematics*, 1st ed. Naples: Bibliopolis; 1928:275–6 pp.
26. Earman J. Bayes or bust? A critical examination of Bayesian confirmation theory. Cambridge MA: MIT Press; 1992.
27. Wallmann C, Williamson J. The principal principle and subjective Bayesianism. *Eur J Philos Sci* 2020;10:3.
28. Wallmann C, Hawthorne J. Admissibility troubles for Bayesian direct inference principles. *Erkenntnis* 2020;85:957–93.
29. Williamson J. Direct inference and probabilistic accounts of induction. *J Gen Philos Sci* 2023. <https://doi.org/10.1007/s10838-021-09584-0>.
30. Jaynes ET. Information theory and statistical mechanics. *Phys Rev* 1957;106:620–30.
31. Jaynes ET. Probability theory: the logic of science. Cambridge: Cambridge University Press; 2003.
32. Caticha A, Giffin A. Updating probabilities. *AIP Conf Proc* 2006;872:31–42.
33. Landes J, Rafiee Rad S, Williamson J. Determining maximal entropy functions for objective Bayesian inductive logic. *J Phil Logic* 2023. <https://doi.org/10.1007/s10992-022-09680-6>.
34. Seidenfeld T. Entropy and uncertainty. *Philos Sci* 1986;53:467–91.
35. Topsøe F. Information theoretical optimization techniques. *Kybernetika* 1979;15:1–27.
36. Grünwald P, Dawid AP. Game theory, maximum entropy, minimum discrepancy, and robust Bayesian decision theory. *Ann Stat* 2004;32:1367–433.
37. Irony TZ, Singpurwalla ND. Non-informative priors do not exist: a dialogue with José M. Bernardo. *J Stat Plann Inference* 1997;65:159–77.
38. Boogerd FC, Bruggeman FJ, Hofmeyr J-HS, Westerhoff HV, editors. *Systems biology: philosophical foundations*. Amsterdam: Elsevier; 2007.
39. Antony PM, Balling R, Vlassis N. From systems biology to systems biomedicine. *Curr Opin Biotechnol* 2012;23:604–8.
40. Wolkenhauer O, Auffray C, Jaster R, Steinhoff G, Dammann O. The road from systems biology to systems medicine. *Pediatr Res* 2013;73:502–7.
41. Galas DJ, Hood L. Systems biology and emerging technologies will catalyze the transition from reactive medicine to predictive, personalized, preventive and participatory (P4) medicine. *Interdiscipl Bio Cent* 2009;1:1–5.
42. Carusi A. Validation and variability: dual challenges on the path from systems biology to systems medicine. *Stud Hist Philos Sci C Stud Hist Philos Biol Biomed Sci* 2014;48:28–37.
43. Williamson J. Models in systems medicine. *Disputatio* 2017;9:429–69.
44. D'Ozario M, Di Zio M, Scanu M. *Statistical matching: theory and practice*. Chichester: Wiley; 2006.
45. Williamson J. Objective bayesian nets. In: Artemov S, Barringer H, d'Avila Garcez AS, Lamb LC, Woods J, editors. *We will show hem!* Essays in honour of Dov Gabbay, vol 2. London: College Publications; 2005:713–30 pp.
46. Landes J, Williamson J. Objective Bayesian nets from consistent datasets. In: Giffin A, Knuth KH, editors. *Proceedings of the 35th international workshop on Bayesian inference and maximum entropy methods in science and engineering*. Volume 1757 of American institute of physics conference proceedings. Potsdam, NY; 2016.
47. Landes J, Williamson J. Objective Bayesian Nets for integrating consistent datasets. *J Artif Intell Res* 2022;74:393–458.
48. Nagl S, Williams M, Williamson J. Objective Bayesian nets for systems modelling and prognosis in breast cancer. In: Holmes D, Jain L, editors. *Innovations in Bayesian networks: theory and applications*. Berlin: Springer; 2008:131–67 pp.
49. Fridlyand J, Snijders A, Ylstra B, Li H, Olshen A, Segraves R, et al. Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer* 2006;6:96.
50. Endres E, Augustin T. Statistical matching of discrete data by Bayesian networks. In: Antonucci A, Corani G, Campos CP, editors. *Proceedings of the eighth international conference on probabilistic graphical models*, vol 52; 2016:159–70 pp. *Proceedings of Machine Learning Research*.
51. Datta GS, Sweeting TJ. Probability matching priors. In: Dey DK, Rao CR, editors. *Bayesian thinking: modeling and computation*. *Handbook of statistics* 25. Amsterdam: Elsevier; 2005:91–114 pp.
52. Scricciolo C. Probability matching priors: a review. *J Ital Stat Soc* 1999;8:83–100.
53. Dawid AP. The well-calibrated Bayesian. *J Am Stat Assoc* 1982;77:604–13.
54. Howson C, Urbach P. *Scientific reasoning: the Bayesian approach*. In: Open Court, 2nd ed. Chicago IL; 1989.
55. Price KL, Xia HA, Lakshminarayanan M, Madigan D, Manner D, Scott J, et al. Bayesian methods for design and analysis of safety trials. *Pharmaceut Stat* 2014;13:13–24.
56. De Pretis F, Landes J, Osimani B. E-synthesis: a Bayesian framework for causal assessment in pharmacosurveillance. *Front Pharmacol* 2019;10:1317.

57. De Pretis F, Landes J, Peden W. Artificial intelligence methods for a Bayesian epistemology-powered evidence evaluation. *J Eval Clin Pract* 2021;27:504–12.
58. Chang H. Is water H₂O? Evidence, realism and pluralism. In: *Boston studies in the philosophy of science*. Dordrecht: Springer; 2012.
59. Ludwig D, Rupy S. Scientific pluralism. In: Zalta EN, editor. *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab. Stanford University; 2021.
60. Gillies D, Zheng Y. Dynamic interactions with the philosophy of mathematics. *Theoria* 2001;16:437–59.