★

# The heuristic use of conditionalisation

★

Jon Williamson*

**Abstract**

I provide an example in which Bayesian conditionalisation fails to validate plausible claims about rational permissibility. The example suggests that conditionalisation should be treated as a heuristic principle: useful in many situations, but not all. This raises the question of how to characterise the cases in which it is appropriate to use conditionalisation. I show that one can answer this question by appealing to the framework of objective Bayesian inductive logic to provide sufficient conditions for the applicability of conditionalisation.

**Keywords.** Bayesianism; Bayesian conditionalisation; objective Bayesian inductive logic; maximum entropy.

§1 provides an example of a situation in which conditionalisation arguably fails to validate normal standards of what is reasonable to believe—this suggests that conditionalisation should not be viewed as a universal updating principle, but rather as a heuristic principle that is appropriate in some circumstances and fails in others. §2 goes on to show that objective Bayesian inductive logic (OBIL) can provide an account of the scope of conditionalisation. Consequences of this argument are explored in §3. In particular, Bayesian conditionalisation should not be thought of as constitutive of Bayesianism; instead, Bayesianism needs to be grounded in a more general framework, such as OBIL, that can explain the successes and failures of conditionalisation.

## §1
## Bayesian conditionalisation and its limits

### §1.1. Bayesian conditionalisation

Suppose belief function $B$ represents your conditional degrees of belief: $B_E(A)$ is the degree to which you would believe proposition $A$ given just $E$, defined for all $A$ and $E$ in a given domain of propositions. Bayesianism standardly invokes the claim that conditional beliefs are conditional probabilities:

*CBCP.* For any rationally permissible conditional belief function $B$, there is some probability function $P$ such that $B_E(A) = P(A|E)$ for all $A$ and $E$.

*Department of Philosophy, Humanities Bridgeford Street, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK.

Bayesian conditionalisation can be thought of as CBCP together with the claim that your actual degrees of belief are conditional degrees of belief:

*ABCB.* For any time $t$, your degrees of belief at $t$ are the values encapsulated in your belief function $B_E(\cdot)$, where $E$ is your total evidence at $t$.

This latter claim is relatively uncontroversial, given the definition of $B_E(\cdot)$. Taken together with CBCP, it implies that your unconditional degrees of belief ought to be probabilities conditional on your total evidence:

*Conditionalisation.* For any time $t$, your degrees of belief at $t$ are the conditional probabilities $P(\cdot|E)$, where $E$ is your total evidence at $t$.

Conditionalisation (or its generalisation, Jeffrey Conditionalisation) is generally held to be constitutive of, or at least core to, Bayesianism (see, e.g., Earman, 1992, §2.1; Gillies, 2000, Chapter 4; Galavotti, 2005, §7.3; Easwaran, 2011; Joyce, 2011, §1.3; Lin, 2022, §1.2).

But is this central role warranted? Several objections to conditionalisation have been put forward in the literature, e.g., by Kyburg (1980, 2006); Bacchus et al. (1990); Talbott (1991); Earman (1992); Howson (1997) and Elga (2000) among others. These objections suggest that conditionalisation may not be universally warranted.

### §1.2. Red faces

In support of the claim that conditionalisation has its limits, consider the following apparent counterexample to CBCP.[1]

*Example* 1 (*Red faces*). Suppose that a fair six-sided die is to be rolled (proposition $X$) and that each face of the die is permanently coloured precisely one of red, yellow or green ($E$). Consider the outcome that the number rolled will be three or greater ($A$). It is reasonable to believe $A$ to degree $\frac{2}{3}$: given that the die is fair, each number has chance $\frac{1}{6}$ of being rolled, and four of the 6 numbers on the die are greater than or equal to three. So, if conditional degrees of belief are conditional probabilities, it is rationally permissible to set:

$$P(A|XE) = 2/3.$$

(One might be inclined to go further and maintain that it is rationally *required* that $P(A|XE) = 2/3$, by invoking the Principal Principle of Lewis (1980), which requires calibrating rational degrees of belief to available current chances. For the purposes of this example, though, we only need that this assignment be rationally permissible, and for this we do not need the Principal Principle. Those who reject the Principal Principle as a general requirement will accept that it is permissible that degrees of belief coincide with chances in scenarios such as this.)

[1]This counterexample develops examples of Hawthorne et al. (2017); Wallmann and Williamson (2020, §3.1) and Williamson (2023). Those earlier examples served a very different purpose, namely to shed light on David Lewis' Principal Principle: to argue that the Principal Principle implies the Principle of Indifference and to point to a tension between the Principal Principle and subjective Bayesianism. Here, it is CBCP and conditionalisation, not the Principal Principle, that is the focus of our attention. The purpose of the present example is purely to show that CBCP and conditionalisation cannot be construed as universal principles. The argument set out in this section requires neither the Principal Principle nor the Principle of Indifference.

Now consider an alternative outcome: that the colour rolled (i.e., the colour of the uppermost face) is red ($R$). It is arguably reasonable to believe $R$ to degree $\frac{1}{3}$, on the grounds that red is one out of the three possible colours and there is no evidence that favours one of these colours over any of the others. Thus it is permissible to set:

$$P(R|XE) = 1/3.$$

(Again, one might be inclined to go further and maintain that it is rationally *required* that $P(R|XE) = 1/3$, by invoking some Principle of Indifference. Here, we need only suppose that this assignment is rationally permissible, and for this we do not need any Principle of Indifference.)

Now suppose in addition that the red faces are precisely those that are numbered three or greater, so $A \leftrightarrow R$. Given that the die is fair, it is again clearly rationally permissible (required, perhaps) to believe $A$ to degree $\frac{2}{3}$:

$$P(A|XE(A \leftrightarrow R)) = 2/3.$$

Note that for these conditional probabilities to be well defined, it must also be rationally permissible to set $P(XE(A \leftrightarrow R)) > 0$, i.e., to assign some positive credence to the claim that the die is fair, faces 3-6 are red and faces 1-2 are yellow or green.

It turns out, however, that these conditional probability assignments are inconsistent—there is no probability function that satisfies them all:

*Theorem 2. There is no probability function $P$ such that $P(A|XE) = 2/3$, $P(R|XE) = 1/3$, $P(A|XE(A \leftrightarrow R)) = 2/3$ and $P(XE(A \leftrightarrow R)) > 0$.*

*Proof:* By Bayes' theorem, $P(A|XE) = \frac{2}{3}$, $P(R|XE) = \frac{1}{3}$ and $P(XE(A \leftrightarrow R)) > 0$ imply that:

$$\begin{aligned}
P(A|XE(A \leftrightarrow R)) &= \frac{P(A \leftrightarrow R|AXE)P(A|XE)}{P(A \leftrightarrow R|XE)} \\
&= \frac{P(AR \vee \bar{A}\bar{R}|AXE)P(A|XE)}{P(AR \vee \bar{A}\bar{R}|XE)} \\
&= \frac{\big(P(AR|AXE) + P(\bar{A}\bar{R}|AXE)\big)P(A|XE)}{P(AR|XE) + P(\bar{A}\bar{R}|XE)} \\
&= \frac{P(R|AXE)P(A|XE)}{P(AR|XE) + P(\bar{A}\bar{R}|XE)} \\
&= \frac{P(AR|XE)}{P(AR|XE) + P(\bar{A}\bar{R}|XE)} \\
&= \frac{1}{2},
\end{aligned}$$

but this contradicts the assignment $P(A|XE(A \leftrightarrow R)) = 2/3$. Note that the last identity above holds because $P(A\bar{R}|XE) + P(\bar{A}\bar{R}|XE) = P(\bar{R}|XE) = \frac{2}{3} = P(A|XE) = P(AR|XE) + P(A\bar{R}|XE)$, so $P(\bar{A}\bar{R}|XE) = P(AR|XE)$. □

It is thus not possible to use conditional probabilities to validate the judgements about rational permissibility in the red-faces example. This undermines CBCP, which identifies conditional beliefs and conditional probabilities. But Bayesian conditionalisation depends upon this identification, so the red-faces example also undermines Bayesian conditionalisation. To see this, suppose that the initial evidence is $XE$. If conditionalisation were a universal norm then it should be rationally permissible to initially believe $A$ to degree 2/3 and $R$ to degree 1/3, and to continue to believe $A$ to degree 2/3 after subsequently learning $A \leftrightarrow R$, at least when all the conditional probabilities are well defined. But we have seen that the corresponding conditional probabilities are inconsistent. So conditionalisation cannot be applied in this case.

### §1.3. Potential objections

The proponent of conditionalisation might object that it is too hasty to conclude that CBCP fails here: the fact that the above assignments of conditional probability are inconsistent shows only that we should reject at least one of them.

One might suggest, for example, that we should reject the assignment $P(R|XE) = 1/3$, on the grounds that this is the least well-evidenced of the three above assignments. One might thus grant that it is permissible that $P(A|XE) = 2/3$ and $P(A|XE(A \leftrightarrow R)) = 2/3$, because proposition $X$ tells us that the chance of $A$ is $\frac{2}{3}$, but argue that these values force $P(R|XE) \neq 1/3$.

This response leads to inconsistency, however. It turns out that if $P(A|XE) = 2/3$ and $P(A|XE(A \leftrightarrow R)) = 2/3$ (and these probabilities are well defined) then $P(R|XE) > 1/3$. Analogous reasoning suggests that you should believe that the outcome will be yellow to degree greater than 1/3, and similarly for the colour green. But then degree of belief greater than 1 is given to the proposition that the outcome is red or yellow or green, which is inconsistent with the axioms of probability, since $E$ implies that these three outcomes are mutually exclusive and exhaustive, so must have probability 1. Letting $Y$ and $G$ denote the outcomes yellow and green respectively,

*Theorem* 3. *There is no probability function $P$ such that*

$$P(A|XE) = P(A|XE(A \leftrightarrow R)) = P(A|XE(A \leftrightarrow Y)) = P(A|XE(A \leftrightarrow G)) = 2/3,$$

*supposing these probabilities to all be well defined.*

See Appendix 1 for a proof.

Since the above response fails, the proponent of conditionalisation might instead suggest that we should reject the assignment $P(A|XE(A \leftrightarrow R)) = 2/3$, on the grounds that if we grant that $P(A|XE) = 2/3$ and $P(R|XE) = 1/3$ then the proof of Theorem 2 can be used to argue that $P(A|XE(A \leftrightarrow R)) = 1/2$.

This much is indeed forced by the axioms of probability, but only this much. To use this fact to conclude that one is rationally required to believe $A$ to degree $\frac{1}{2}$, given $XE(A \leftrightarrow R)$, would beg the question: that would presuppose CBCP, the very principle that is at stake. In order to make good on this alternative line of response, one must somehow independently undermine the claim that it is rationally permissible to believe $A$ to degree $\frac{2}{3}$, given $XE(A \leftrightarrow R)$. There are two tactics one might try here: one might attempt to cast general doubt on any appeal to intuitions about rationality in cases such as these (e.g., Hart and Titelbaum, 2015; Titelbaum and Hart, 2020), or one might attempt to undermine this particular judgement of rational permissibility. The first option is unlikely to be of much help, however. As Landes et al. (2021) and Williamson (2022) argue, such intuitions are all we have to go on: we can only settle fundamental questions like those considered here by appealing to our normal informal standards of what is reasonable to believe. And in the red-faces example, CBCP cannot consistently validate generally held intuitions about rational permissibility. This clearly tells against CBCP and conditionalisation.

So let us turn to the second option: an attempt to undermine the specific judgement that it is permissible to believe $A$ to degree $\frac{2}{3}$, given $XE(A \leftrightarrow R)$. Recall that this judgement does not presuppose the Principal Principle: this is a claim about rational permissibility, while the Principal Principle asserts a rational requirement. Nevertheless, one might reasonably maintain that we take the permissibility claim to be intuitively true because the Principal Principle is broadly correct. Undermining the use of the Principal Principle in this instance might thereby undermine the judgement of rational permissibility. Let us consider the viability of this strategy.

Lewis (1980), when developing the Principal Principle, insisted that conditioned propositions should be 'admissible', i.e., that they should not provide relevant information that undermines setting your degrees of belief to available chances. Perhaps $E(A \leftrightarrow R)$ is inadmissible here, in which case the Principal Principle would not force $P(A|XE(A \leftrightarrow R)) = 2/3$, undermining the motivation for the assignment $P(A|XE(A \leftrightarrow R)) = 2/3$.

This would seem an implausible line of argument, however, in the light of Lewis' remarks on admissibility:

> If a proposition is entirely about matters of particular fact at times no later than $t$, then as a rule that proposition is admissible at $t$. Admissible information just before the toss of a coin, for example, includes the outcomes of all previous tosses of that coin and others like it. It also includes every detail—no matter how hard it might be to discover—of the structure of the coin, the tosser, other parts of the set-up, and even anything nearby that might somehow intervene. It also includes a great deal of other information that is completely irrelevant to the outcome of the toss. (Lewis, 1980, pp. 92–93.)

The proposition that *the faces of the die are (permanently) coloured red, yellow or green and the red faces of the die are precisely those numbered 3 or greater* is prototypically admissible: it is information about matters of fact up to the roll of the die in question, it is information about the structure of the die, and (we may suppose) it is completely irrelevant to the outcome $A$ of the roll of the die.

The proponent of conditionalisation might object that Lewis' remarks are too informal to be conclusive. Indeed, Pettigrew (2020, 612) suggests that a more precise account of admissibility should be applied to cases such as this:

*Levi-Admissibility.* $E$ is Levi-admissible for $A$ at $t$ if and only if, for all possible chance functions $ch$ at $t$, if $ch(E) > 0$, then $ch(A|E) = ch(A)$.

Is the proposition $E(A \leftrightarrow R)$ Levi-admissible for $A$? Yes—because whatever the chance of $A$ given $E(A \leftrightarrow R)$, the colouring of the faces is determined by the initial time $t$ and so already factored into the unconditional chance function $ch$. Consider a hypothetical example in which the colouring is relevant to the chance of $A$: suppose the yellow and green colourings on the faces numbered 1 and 2 are heavier than the red colouring, making the chance of $A$ greater than 2/3. Thus $ch(A|E(A \leftrightarrow R)) = x > 2/3$. But then it is also the case that $ch(A) = x > 2/3$, because the colouring is permanent and determined prior to $t$. Hence, $ch(A|E(A \leftrightarrow R)) = ch(A)$. This also holds for cases in which $x < 2/3$ or $x = 2/3$. Therefore, Levi-admissibility is ensured.

In sum, whether we appeal to Lewis' characterisation of admissibility or Levi-admissibility, $E(A \leftrightarrow R)$ turns out to be admissible. And with admissibility, the Principal Principle does indeed force the assignment $P(A|XE(A \leftrightarrow R)) = 2/3$. There is thus little prospect in appealing to questions of admissibility to undermine the judgement that is it permissible to believe $A$ to degree $\frac{2}{3}$, given $XE(A \leftrightarrow R)$.

Finally, it is worth noting that some authors have expressed reservations about the very idea of conditionalising on conditional propositions (see, e.g., Douven, 2012; Eva et al., 2020). Such doubts might plausibly extend to biconditionals, such as the proposition $A \leftrightarrow R$. But the restriction of conditionalisation to propositions other than conditionals merely supplies more grist to the mill. The purpose of the red-faces example is to show that conditionalisation does not have universal scope. If one cannot always conditionalise on conditionals or biconditionals then the scope of conditionalisation is limited. Therefore, our key conclusion goes through whether or not one can conditionalise on conditionals.

We see, then, that the above argument against the universality of conditionalisation is not so easy to undermine. If there are certain circumstances in which conditionalisation does not apply, it is perhaps best thought of a heuristic principle: convenient to use in a wide range of situations, but error-prone or sub-optimal in others. How can we characterise the situations in which it is safe to conditionalise? Clearly, no version of Bayesianism that presupposes CBCP as a universal principle can answer this question. One can, however, provide an answer this question by moving to a broader framework for inductive inference, as we shall now see.

## §2
## OBIL and the scope of conditionalisation

In this section, I shall introduce objective Bayesian inductive logic (in §2.1) and provide a result that sets out conditions under which conditionalisation is an appropriate rule of updating (in §2.2).

### §2.1. Objective Bayesian inductive logic

OBIL is an approach to inductive logic that appeals to the maximum entropy principle. The maximum entropy principle is usually applied to finite or continuous domains. In OBIL, however, the domain is countably infinite: the set of sentences of a first-order predicate language. Here, I shall sketch the rudiments of the approach—see Williamson (2010, 2017) for more detail.[2]

OBIL provides semantics for inductive entailment relationships of the form:

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx \psi^Y,$$

where $\varphi_1, \ldots, \varphi_k, \psi$ are sentences of the first-order language and $X_1, \ldots, X_k, Y$ are intervals of probabilities. OBIL interprets such an entailment relationship as holding just when every rationally permissible belief function that satisfies the premisses

[2] Note that there is long-standing controversy surrounding the application of the maximum entropy principle to uncountable domains (see, e.g., Shackel and Rowbottom, 2020). These worries do not pertain to the present context of the countable domain of sentences of a first-order predicate language. See Williamson (2010, 2017) for the motivation behind OBIL and a discussion of the ways in which it handles potential objections.

$\varphi_1^{X_1},\ldots,\varphi_k^{X_k}$ also satisfies the conclusion $\psi^Y$. OBIL takes the premisses to capture the available evidence and deems a belief function to be rationally permissible just when it is a probability function $P$, from all those that satisfy the evidential constraints $P(\varphi_1)\in X_1,\ldots,P(\varphi_k)\in X_k$, that is as equivocal as possible with respect to other propositions. In line with Jaynes (1957), the degree to which a belief function is equivocal is measured by its entropy.

Let us specify the framework more precisely. Here $\mathscr{L}$ is a pure first-order predicate language: it has finitely many relation symbols $U_1,\ldots,U_l$, denumerably many constant symbols $t_1,t_2,\ldots$ and variable symbols $x_1,x_2,\ldots$, but no function symbols or equality. Sentences $\theta,\varphi_i,\psi$ etc. are formed in the usual way using quantifiers $\forall,\exists$, and connectives $\neg,\wedge,\vee,\rightarrow,\leftrightarrow$. Atomic sentences $a_1,a_2,\ldots$ are ordered so that those involving constants $t_1,\ldots,t_n$ occur in the ordering before those involving $t_{n+1}$. The sublanguages $\mathscr{L}_n$ have all the syntactic apparatus of $\mathscr{L}$ but involve only the constants $t_1,\ldots,t_n$. The *$n$-states* $\omega\in\Omega_n$ of $\mathscr{L}$ are the states of $\mathscr{L}_n$, i.e., the sentences of the form $\pm a_1\wedge\cdots\wedge\pm a_{r_n}$, where $a_1,\ldots,a_{r_n}$ are the atomic sentences of $\mathscr{L}_n$.

A probability function $P$ is a function defined on the sentences of $\mathscr{L}$ such that:

*P1*: If $\tau$ is a deductive tautology, i.e., $\models\tau$, then $P(\tau)=1$.

*P2*: If $\theta$ and $\varphi$ are mutually exclusive, i.e., $\models\neg(\theta\wedge\varphi)$, then $P(\theta\vee\varphi)=P(\theta)+P(\varphi)$.

*P3*: $P(\exists x\theta(x))=\sup_m P\left(\bigvee_{i=1}^m\theta(t_i)\right)$.

Axiom P3, which is also known as *Gaifman's condition*, presupposes that each member of the domain of discourse is named by some constant symbol $t_i$.

A probability function is uniquely determined by its values on the $n$-states (Williamson, 2017, Chapter 2). The set of all probability functions on $\mathscr{L}$ is denoted by $\mathbb{P}$. We are particularly interested in the set of probability functions that satisfy the evidential constraints:

$$\mathbb{E}\stackrel{\text{df}}{=}\mathbb{P}[\varphi_1^{X_1},\ldots,\varphi_k^{X_k}]\stackrel{\text{df}}{=}\{P\in\mathbb{P}:P(\varphi_1)\in X_1,\ldots,P(\varphi_k)\in X_k\}.$$

In OBIL, constraints may be imposed by evidence directly ($\varphi$ imposes the constraint $P(\varphi)=1$, written $\varphi^1$ or simply $\varphi$) or by calibration to chances via an analogue of the Principal Principle (if $\varphi$ says that the chance of $\theta$ is in $X$, then $\varphi$ imposes the constraint $\theta^X$ as well as the constraint $\varphi^1$).[3]

Now we are in a position to say what it is for a probability function $P$ to be 'maximally equivocal'. We define the $n$-entropy:[4]

$$H_n(P)\stackrel{\text{df}}{=}-\sum_{\omega\in\Omega_n}P(\omega)\log P(\omega).$$

We then say that $P$ has greater entropy than $Q$ iff the $n$-entropy of $P$ eventually dominates that of $Q$, i.e., iff there is an $N\in\mathbb{N}$ such that for all $n\geq N$,

[3]Lewis' Principal Principle is formulated in terms of conditional probabilities. OBIL instead appeals to an analogue of the Principal Principle, the Calibration norm of Williamson (2010, §2.3.2), that is formulated in terms of unconditional probabilities. The difference between the two versions is not crucial for the purposes of the paper and I shall use 'Principal Principle' to refer to both the original principle and its objective Bayesian analogue.

[4]We adopt the usual convention that $0\log 0=0$. This convention is motivated by the fact that for $x>0$, $x\log x\rightarrow 0$ as $x\rightarrow 0$.

$H_n(P) > H_n(Q)$. The greater-entropy relation yields a partial ordering of probability functions, which may contain maximal elements (undominated functions) but need not necessarily contain a maximum element (a function that dominates all others). We thus define the maximally equivocal functions in $\mathbb{E}$ to be those with maximal entropy:

$$\text{maxent}\,\mathbb{E} \stackrel{\text{df}}{=} \{P \in \mathbb{E} : \text{ there is no } Q \in \mathbb{E} \text{ that has greater entropy than } P\}.$$

We can now provide semantics for OBIL:

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\approx} \psi^Y \text{ iff } P(\psi) \in Y \text{ for all } P \in \text{maxent}\,\mathbb{E},$$

as long as $\text{maxent}\,\mathbb{E} \neq \emptyset$. There are two cases in which $\text{maxent}\,\mathbb{E} = \emptyset$, and we need some conventions to cover these cases. The first case occurs when the premisses are unsatisfiable, $\mathbb{E} = \emptyset$. In that case, it is desirable to avoid 'explosion', i.e., the phenomenon that any conclusion follows, on the grounds that it is never rational to believe everything. We thus consider $\text{maxent}\,\mathbb{P}$ instead of $\text{maxent}\,\mathbb{E}$ when $\mathbb{E} = \emptyset$. $\mathbb{P}$ has a unique maximiser, namely the *equivocator function* $P_=$ defined by $P_=(\omega) = 1/|\Omega_n| = 1/2^{r_n}$ for all $\omega \in \Omega_n$ and $n \geq 1$. Hence an entailment relationship with inconsistent premisses holds iff $P_=(\psi) \in Y$. $P_=(\psi)$ will be referred to as the *measure* of $\psi$. The second case occurs where $\mathbb{E} \neq \emptyset$ but for any probability function in $\mathbb{E}$ there is another function with greater entropy. In this case it is not possible to maximise entropy, so we shall consider $\mathbb{E}$ instead of $\text{maxent}\,\mathbb{E}$. In this second case, then, the entailment relationship holds whenever $P(\psi) \in Y$ for all $P \in \mathbb{E}$.

We use '$\mathrel{|\!\!\stackrel{\circ}{\approx}}$' to refer to the ensuing objective Bayesian inductive entailment relation:

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\stackrel{\circ}{\approx}} \psi^Y$$

holds just when $[\![\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}]\!] \subseteq \mathbb{P}[\psi^Y]$, where $[\![\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}]\!]$ is the set of models of the premisses:

$$[\![\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}]\!] = \begin{cases} \text{maxent}\,\mathbb{E} & : \text{maxent}\,\mathbb{E} \neq \emptyset \\ \mathbb{E} & : \text{maxent}\,\mathbb{E} = \emptyset \neq \mathbb{E} \\ \text{maxent}\,\mathbb{P} & : \mathbb{E} = \emptyset. \end{cases}$$

If there are no premisses then $\mathbb{E} = \mathbb{P}$, so the equivocator function is used for inference: $\text{maxent}\,\mathbb{E} = \text{maxent}\,\mathbb{P} = \{P_=\}$. If $\mathrel{|\!\!\stackrel{\circ}{\approx}} \psi$ then $\psi$ is said to be an *inductive tautology*. Equivalently, $P_=(\psi) = 1$, i.e., $\psi$ has measure 1. If $\mathrel{|\!\!\stackrel{\circ}{\approx}} \neg\psi$ (i.e., $\psi$ has measure 0) then $\psi$ is an *inductive contradiction*. If $\not\mathrel{|\!\!\stackrel{\circ}{\approx}} \neg\psi$ (i.e., $\psi$ has positive measure) then $\psi$ is *inductively consistent*. Similarly, if $\not\mathrel{|\!\!\stackrel{\circ}{\approx}} \neg(\psi \wedge \theta)$ (i.e., $P_=(\psi \wedge \theta) > 0$) then $\psi$ is *inductively consistent* with $\theta$. If $\mathrel{|\!\!\stackrel{\circ}{\approx}} \psi \leftrightarrow \theta$ then $\psi$ and $\theta$ are *inductively equivalent*.

We have seen that there is a unique entropy maximiser in the case in which there are no premisses and in the case in which the premisses are unsatisfiable. There are various other situations in which there is known to be a unique entropy maximiser. One such situation is that in which the premisses have *satisfiable support* (see Landes et al., 2026):

*Definition* 4 (*Support*). Suppose $a_{i_1}, \ldots, a_{i_m}$ include all the atomic propositions that appear in sentence $\varphi$ of $\mathscr{L}$, and let $\Xi_\varphi \stackrel{\text{df}}{=} \{\pm a_{i_1} \wedge \ldots \wedge \pm a_{i_m}\}$ be the set of states of these atomic propositions. If $\varphi$ contains no atomic propositions, we take $\Xi_\varphi \stackrel{\text{df}}{=}$

$\{a_1, \neg a_1\}$. The *support* $\breve{\varphi}$ of $\varphi$ is the disjunction of states in $\Xi_\varphi$ that are inductively consistent with $\varphi$, i.e., the disjunction of $\xi \in \Xi_\varphi$ such that $\not\models^{\circ}\neg(\xi \wedge \varphi)$. Equivalently,

$$\breve{\varphi} \stackrel{\text{df}}{=} \bigvee\{\xi \in \Xi_\varphi : P_=(\xi \wedge \varphi) > 0\}.$$

A set of premisses $\{\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}\}$ has *satisfiable support* if the corresponding support constraints are satisfiable, $\mathbb{P}[\breve{\varphi}_1^{X_1}, \ldots, \breve{\varphi}_k^{X_k}] \neq \emptyset$.

## §2.2. The scope of conditionalisation

OBIL can be thought of as requiring a two-step process. One must first formulate the constraints imposed by evidence, and only then can one move on to inference in inductive logic:

$$\text{Evidence} \xmapsto[\text{formulation}]{\text{constraint}} \varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \xmapsto[\text{logic}]{\text{inductive}} [\![\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}]\!] \stackrel{?}{\subseteq} \mathbb{P}[\psi^Y]$$

OBIL conceives of learning as follows. New evidence imposes new constraints on rational degrees of belief, which may override or modify previous constraints. Thus on learning new information, the set of constraints (i.e., the set of premisses) changes. New inferences are made possible by maximising entropy with respect to the revised set of premiss constraints. There is no special updating rule that plays a role analogous to conditionalisation.

Nevertheless, learning in OBIL can accord with conditionalisation. It turns out that four conditions are sufficient to secure agreement between OBIL and conditionalisation:[5]

*Theorem 5 (OBIL and Conditionalisation). Suppose initial evidence imposes constraints $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ and $P^\dagger \in \text{maxent}\,\mathbb{E}$, where $\mathbb{E}$ is the set of probability functions satisfying these initial constraints. Let $\mathbb{E}'$ be the set of probability functions satisfying the constraints that result from subsequently learning some new evidence that can be expressed as a quantifier-free sentence $\theta$ of $\mathscr{L}$. If:*

1. $\mathbb{E}' = \mathbb{E} \cap \mathbb{P}[\theta]$,
2. $P^\dagger(\theta) > 0$,
3. $P^\dagger(\cdot|\theta) \in \mathbb{E}$,
4. *if* $P^\dagger(\neg\theta) > 0$, *then* $P^\dagger(\cdot|\neg\theta) \in \mathbb{E}$,

*then*

$$P^\dagger(\cdot|\theta) \in \text{maxent}\,\mathbb{E}'.$$

*If in addition the initial constraints have satisfiable support then $P^\dagger(\cdot|\theta)$ is the unique entropy maximiser, i.e.,*

$$\text{maxent}\,\mathbb{E}' = \{P^\dagger(\cdot|\theta)\}.$$

[5]This result generalises Seidenfeld (1986, Result 1), which applies only to finite domains, and Williamson (2017, Theorem 5.16), which applies to point-valued rather than interval-valued constraints.

A proof is provided in Appendix 2. A corollary, provided in Appendix 3, shows that condition (4) is guaranteed to be satisfied in the special case in which the constraints are point-valued, $X_1,\dots,X_k \in [0,1]$.

If one or more of the conditions of Theorem 5 fail and $P^{\dagger}(\cdot|\theta) \not\in \mathrm{maxent}\,\mathbb{E}'$ then conditionalisation is not an appropriate way to update. Consider first a failure of condition (1). In this case $\theta$ has complex consequences that cannot be captured solely by adding the constraint $P(\theta)=1$ to those imposed by the initial evidence. To give an example, suppose there is no initial evidence, and new evidence $\theta$ says that a thousand ravens have all been found to be black, $\theta = Br_1 \wedge \dots \wedge Br_{1000}$. $\theta$ arguably tells us something about the chance that the next raven is black: that this chance is close to 1. If so, then by the Principal Principle, $\theta$ imposes a constraint of the form $Br_{1001}^{[1-\epsilon,1]}$ for some small $\epsilon$. But $P^{\dagger} = P_{=}$ since there is no initial evidence, and $P_{=}(Br_{1001}|\theta) = 1/2$, which violates this extra constraint. Conditionalisation is thus inappropriate here because conditionalising on the new evidence $\theta$ would violate a constraint imposed by $\theta$.

If condition (2) fails then the conditional probability is undefined and conditionalising is not possible. If condition (3) fails, conditionalisation fails to satisfy the old constraints, which need to be satisfied by condition (1). Indeed, if either one of conditions (3) and (4) fails, then when ascertaining the truth or falsity of $\theta$ there will be no guarantee that conditionalisation will respect the old constraints.

The red-faces example violates condition 3. Here, the initial evidence is $XE$ and the new evidence $\theta$ is $A \leftrightarrow R$. Note that $XE$ imposes the constraint $P(A) = 2/3$ via the Principal Principle, so $\mathbb{E} = \mathbb{P}[XE, A^{2/3}]$ and $\mathbb{E}' = \mathbb{P}[XE, A^{2/3}, A \leftrightarrow R]$. Condition 3 fails because $P(\cdot|XE(A \leftrightarrow R))$ does not satisfy all the constraints imposed by $XE$: $XE$ imposes the constraint $P(A) = 2/3$, but we saw in the proof of Theorem 2 that $P(A|XE(A \leftrightarrow R)) = 1/2$, which violates this constraint. It is thus not safe to conditionalise on $A \leftrightarrow R$.

OBIL handles the red-faces scenario as follows. Initially we have evidence $XE$, where $E$ (which says that the colour rolled is precisely one of red, yellow or green) can be written $(R \vee Y \vee G) \wedge \neg(R \wedge Y) \wedge \neg(R \wedge G) \wedge \neg(Y \wedge G)$. As noted above, the initial evidence $XE$ imposes the further constraint $P(A) = 2/3$ via the Principal Principle. The function in $\mathbb{E} = \mathbb{P}[XE, A^{2/3}]$ that has maximal entropy is the probability function $P^{\dagger}$ such that:

$$P^{\dagger}(XAR\bar{Y}\bar{G}) = P^{\dagger}(XA\bar{R}Y\bar{G}) = P^{\dagger}(XA\bar{R}\bar{Y}G) = 2/9,$$

$$P^{\dagger}(X\bar{A}R\bar{Y}\bar{G}) = P^{\dagger}(X\bar{A}\bar{R}Y\bar{G}) = P^{\dagger}(X\bar{A}\bar{R}\bar{Y}G) = 1/9,$$

with all other states having zero probability. This function $P^{\dagger}$ does indeed satisfy $P^{\dagger}(A) = 2/3$ and $P^{\dagger}(R) = 1/3$, validating the first two judgements of rational permissibility in the red-faces example. Subsequently, on learning $A \leftrightarrow R$, OBIL selects the function in $\mathbb{E}' = \mathbb{P}[XE, A^{2/3}, A \leftrightarrow R]$ that has maximal entropy. This picks out the probability function $P^{\ddagger}$ defined by:

$$P^{\ddagger}(XAR\bar{Y}\bar{G}) = 2/3,$$

$$P^{\ddagger}(X\bar{A}\bar{R}Y\bar{G}) = P^{\ddagger}(X\bar{A}\bar{R}\bar{Y}G) = 1/6,$$

with all other states having zero probability. Again we have that $P^{\ddagger}(A) = 2/3$, validating the third of the three judgements of rational permissibility in the red-faces example. Thus OBIL validates all the judgements of rational permissibility in the red-faces scenario. Therefore, while the red-faces example poses a problem for conditionalisation, it does not pose a problem for OBIL.

## §3
## Consequences

I have argued that conditionalisation is a useful heuristic but not universally valid. Like other heuristics, it can lead to error in pathological situations. From the perspective of OBIL, the pathological situations for conditionalisation are situations in which one or more of the four conditions of Theorem 5 are violated.

Consequently, conditionalisation should not be taken to be constitutive of Bayesianism. Instead, we can understand Bayesianism in a broader way, as encompassing any theory that identifies rational degrees of belief and probabilities. A Bayesian interpretation of probability construes probabilities as rational degrees of belief, while in the other direction, Bayesian epistemology explicates rational degrees of belief as probabilities. It is in this broader sense that objective Bayesian inductive logic is Bayesian: it interprets inductive entailment relationships as entailment relationships between rational degrees of belief, and explicates these rational degrees of belief as probabilities.

A second consequence is that any version of objective Bayesianism (such as OBIL) that adopts a version of the maximum entropy principle need not presuppose a separate diachronic updating principle such as conditionalisation. The synchronic maximum entropy principle is sufficient to handle updating. Indeed, objective Bayesianism *should* not presuppose conditionalisation, because there are cases in which conditionalisation and the maximum entropy principle come apart. This consideration favours the objective Bayesianism of Williamson (2010) over the version developed by Jaynes (1957, 2003) and Rosenkrantz (1977).

Thirdly, this leaves subjective Bayesianism, which eschews principles such as the maximum entropy principle, in a quandary. Subjective Bayesianism invokes conditionalisation because it has no independent means of accommodating the learning of new information. But conditionalisation is not universally valid, as we have observed. Consequently, subjectivism provides, at best, an incomplete account of inductive inference.

Fourthly, the arguments of this paper can therefore be viewed as supporting 'time-slice rationality': the claim that the norms of rational belief formation are purely synchronic (Hedden, 2015; Cassell, 2025; Fischer, 2025). The red-faces example undermines the standard diachronic view of rational belief formation, which invokes conditionalisation. On the other hand, we have seen that OBIL, which invokes purely synchronic norms, is immune to this problem.

Finally, *red faces* poses a problem for the 'new paradigm' in the psychology of reasoning, which seeks to understand our reasoning by appeal to conditional probabilities and conditionalisation (Oaksford and Chater, 2020). The new paradigm analyses our use of conditional propositions in terms of conditional probabilities, for example. If it is to account for the heuristic role of conditionalisation and its failure in pathological situations, the new paradigm needs to move to a more general framework, such as OBIL.[6]

University of Manchester
UK
jon.williamson@manchester.ac.uk

[6]I am very grateful to Jürgen Landes and the anonymous referees for many constructive comments on earlier versions of this paper. §1 builds upon joint research with Christian Wallmann and §2 builds upon joint research with Jürgen Landes and Soroush Rafiee Rad.

**Funding**

This work was supported by UKRI [grant number MR/Z506291/1] and the Leverhulme Trust [grant number RPG-2022-336].

**Appendix 1: Proof of Theorem 3**

Suppose for contradiction that there is some probability function $P$ that satisfies these constraints.

First we show that $P(R|XE) > 1/3$.

Let:

$$z_0 \stackrel{\text{df}}{=} P(AR|XE)$$
$$z_1 \stackrel{\text{df}}{=} P(A\bar{R}|XE)$$
$$z_2 \stackrel{\text{df}}{=} P(\bar{A}R|XE)$$
$$z_3 \stackrel{\text{df}}{=} P(\bar{A}\bar{R}|XE).$$

Recall from the proof of Theorem 2 that

$$P(A|XE(A \leftrightarrow R)) = \frac{P(AR|XE)}{P(AR|XE) + P(\bar{A}\bar{R}|XE)}.$$

Since $P(A|XE(A \leftrightarrow R)) = 2/3$, we have that

$$\frac{2}{3} = \frac{z_0}{z_0 + z_3}$$

so,

$$z_0 = 2z_3.$$

Since $P(A|XE) = 2/3$,

$$z_0 + z_1 = 2/3.$$

The axioms of probability imply that

$$z_0 + z_1 + z_2 + z_3 = 1.$$

Let $x \stackrel{\text{df}}{=} P(R|XE) = z_0 + z_2$.

These constraints imply that

$$z_0 = 2x - 2/3$$
$$z_1 = 4/3 - 2x$$
$$z_2 = 2/3 - x$$
$$z_3 = x - 1/3.$$

Since the conditional probabilities are well defined, $P(A \leftrightarrow R|XE) > 0$, so

$$z_0 + z_3 = 3x - 1 > 0.$$

This implies that $x > 1/3$, as required.

Exactly analogous arguments can be used to show that $P(Y|XE) > 1/3$ and $P(G|XE) > 1/3$. But then $P(R \vee Y \vee G|XE) > 1$, infringing the axioms of probability.

## Appendix 2: Proof of Theorem 5

Condition (1) implies that $\mathbb{E}' = \mathbb{E}+\theta \stackrel{\text{df}}{=} \mathbb{E}\cap\mathbb{P}[\theta]$. Condition (3) implies that $P^\dagger(\cdot|\theta) \in \mathbb{E}'$.

There are two cases: either $P^\dagger(\theta) = 1$ or $P^\dagger(\theta) < 1$.

If $P^\dagger(\theta) = 1$, then $P^\dagger \in \mathbb{E}+\theta = \{P \in \mathbb{E} : P(\theta) = 1\}$ and $P^\dagger(\cdot|\theta) = P^\dagger(\cdot)$. Since $P^\dagger(\cdot|\theta) = P^\dagger \in \operatorname{maxent}\mathbb{E}$ and $P^\dagger(\cdot|\theta) \in \mathbb{P}[\theta]$, $P^\dagger(\cdot|\theta) \in \mathbb{E}' = \mathbb{E}\cap\mathbb{P}[\theta]$, as required.

If $P^\dagger(\theta) < 1$, then $P^\dagger(\neg\theta) > 0$ and the probability function $P^\dagger(\cdot|\neg\theta)$ is well defined.

Suppose for contradiction that $P^\dagger(\cdot|\theta) \notin \operatorname{maxent}(\mathbb{E}+\theta)$. Then $P^\dagger(\cdot|\theta)$ must be eventually dominated by some other probability function: there is some $S \in \mathbb{E}+\theta$ such that $S \neq P^\dagger(\cdot|\theta)$ and $H_n(S) > H_n(P^\dagger(\cdot|\theta))$ for sufficiently large $n$.

Define probability function $Q$ by:

$$Q(\cdot) \stackrel{\text{df}}{=} S(\cdot|\theta)P^\dagger(\theta) + P^\dagger(\cdot|\neg\theta)P^\dagger(\neg\theta).$$

Note that $Q \neq P^\dagger(\cdot|\theta)$ as can be seen by observing that $Q(\theta) = P^\dagger(\theta) < 1 = P^\dagger(\theta|\theta)$. Also, $Q \neq P^\dagger$ because $S(\cdot|\theta) = S \neq P^\dagger(\cdot|\theta)$. Since $Q$ is a convex combination of $S(\cdot|\theta) = S \in \mathbb{E}+\theta \subseteq \mathbb{E}$ and $P^\dagger(\cdot|\neg\theta) \in \mathbb{E}$ (by condition (4)), and $\mathbb{E}$ is convex, $Q \in \mathbb{E}$.

We now apply the chain rule for entropy (Cover and Thomas, 1991, §2.5). The chain rule can be expressed in terms of finite, quantifier-free partitions $\pi_1$ and $\pi_2$ of sentences of $\mathscr{L}$:

*Chain Rule.* $H_n^P(\pi_1,\pi_2) = H_n^P(\pi_1) + H_n^P(\pi_2|\pi_1)$,

where,

$$H_n^P(\pi_1,\pi_2) \stackrel{\text{df}}{=} -\sum_{\varphi\in\pi_1}\sum_{\theta\in\pi_2} P(\theta\wedge\varphi)\log P(\theta\wedge\varphi),$$

$$H_n^P(\pi_2|\pi_1) \stackrel{\text{df}}{=} -\sum_{\varphi\in\pi_1} P(\varphi) \sum_{\substack{\theta\in\pi_2 \\ P(\varphi)>0}} P(\theta|\varphi)\log P(\theta|\varphi).$$

Applying the Chain Rule twice, we have that for sufficiently large $n$,

$$\begin{aligned}
H_n(Q) &= -\sum_{\substack{\omega\in\Omega_n\\ \omega\models\theta}} Q(\omega)\log Q(\omega) - \sum_{\substack{\omega\in\Omega_n\\ \omega\models\neg\theta}} Q(\omega)\log Q(\omega)\\
&= -\sum_{\omega\in\Omega_n} Q(\omega\wedge\theta)\log Q(\omega\wedge\theta) - \sum_{\omega\in\Omega_n} Q(\omega\wedge\neg\theta)\log Q(\omega\wedge\neg\theta)\\
&= H_n^Q(\{\theta,\neg\theta\},\Omega_n)\\
&= H_n^Q(\{\theta,\neg\theta\}) + H_n^Q(\Omega_n|\{\theta,\neg\theta\})\\
&= -P^\dagger(\theta)\log P^\dagger(\theta) - P^\dagger(\neg\theta)\log P^\dagger(\neg\theta)\\
&\quad -P^\dagger(\theta)\sum_{\omega\in\Omega_n} S(\omega|\theta)\log S(\omega|\theta) - P^\dagger(\neg\theta)\sum_{\omega\in\Omega_n} P^\dagger(\omega|\neg\theta)\log P^\dagger(\omega|\neg\theta)\\
&= -P^\dagger(\theta)\log P^\dagger(\theta) - P^\dagger(\neg\theta)\log P^\dagger(\neg\theta)\\
&\quad -P^\dagger(\theta)\sum_{\omega\in\Omega_n} S(\omega)\log S(\omega) - P^\dagger(\neg\theta)\sum_{\omega\in\Omega_n} P^\dagger(\omega|\neg\theta)\log P^\dagger(\omega|\neg\theta)\\
&> -P^\dagger(\theta)\log P^\dagger(\theta) - P^\dagger(\neg\theta)\log P^\dagger(\neg\theta)\\
&\quad -P^\dagger(\theta)\sum_{\omega\in\Omega_n} P^\dagger(\omega|\theta)\log P^\dagger(\omega|\theta) - P^\dagger(\neg\theta)\sum_{\omega\in\Omega_n} P^\dagger(\omega|\neg\theta)\log P^\dagger(\omega|\neg\theta)\\
&= H_n^{P^\dagger}(\{\theta,\neg\theta\}) + H_n^{P^\dagger}(\Omega_n|\{\theta,\neg\theta\})\\
&= H_n^{P^\dagger}(\{\theta,\neg\theta\},\Omega_n)\\
&= H_n(P^\dagger),
\end{aligned}$$

where the first and last identities rely on the assumption that $\theta$ is quantifier-free, and the inequality follows by an application of our supposition that $H_n(S) > H_n(P^\dagger(\cdot|\theta))$ for sufficiently large $n$. However, that $H_n(Q) > H_n(P^\dagger)$ for sufficiently large $n$, where $Q \in \mathbb{E}$ and $Q \neq P^\dagger$, contradicts the assumption that $P^\dagger \in \mathrm{maxent}\,\mathbb{E}$. Hence $P^\dagger(\cdot|\theta) \in \mathrm{maxent}(\mathbb{E}+\theta)$ after all.

In the case in which, in addition to conditions 1-4 holding, the initial set $\{\varphi_1^{X_1},\ldots,\varphi_k^{X_k}\}$ of premisses has satisfiable support, the revised set of premisses $\{\varphi_1^{X_1},\ldots,\varphi_k^{X_k},\theta\}$ also has satisfiable support. We noted in §2.1 that support satisfiability implies uniqueness of the entropy maximiser. Since $P^\dagger(\cdot|\theta) \in \mathrm{maxent}\,\mathbb{E}'$, $P^\dagger(\cdot|\theta)$ must be that unique maximiser.

## Appendix 3. A corollary of Theorem 5

It is worth noting that we can drop condition (4) of Theorem 5 if we assume point-valued instead of interval-valued constraints:

*Corollary* 6. *Suppose* $\mathbb{E} = \mathbb{P}[\varphi_1^{c_1},\ldots,\varphi_k^{c_k}]$ *where* $c_1,\ldots,c_k \in [0,1]$, $P^\dagger \in \mathrm{maxent}\,\mathbb{E}$ *and* $\theta$ *is quantifier-free. If:*

1. $\mathbb{E}' = \mathbb{E}\cap\mathbb{P}[\theta]$,
2. $P^\dagger(\theta) > 0$,
3. $P^\dagger(\cdot|\theta) \in \mathbb{E}$,

*then*

$$P^\dagger(\cdot|\theta) \in \mathrm{maxent}\,\mathbb{E}'.$$

*Consequently, if in addition the initial constraints have satisfiable support then* $P^\dagger(\cdot|\theta)$ *is the unique entropy maximiser, i.e.,*

$$\mathrm{maxent}\,\mathbb{E}' = \{P^\dagger(\cdot|\theta)\}.$$

*Proof:* We need to show that condition (4) holds. Suppose $P^\dagger(\neg\theta) > 0$. We can see that $P^\dagger(\cdot|\neg\theta) \in \mathbb{E}$, because for $i = 1, \dots, k$:

$$\begin{aligned} P^\dagger(\varphi_i|\neg\theta) &= \frac{P^\dagger(\varphi_i \wedge \neg\theta)}{P^\dagger(\neg\theta)} \\ &= \frac{P^\dagger(\varphi_i) - P^\dagger(\varphi_i \wedge \theta)}{1 - P^\dagger(\theta)} \\ &= \frac{P^\dagger(\varphi_i) - P^\dagger(\varphi_i|\theta)P^\dagger(\theta)}{1 - P^\dagger(\theta)} \\ &\overset{(3)}{=} \frac{c_i - c_i P^\dagger(\theta)}{1 - P^\dagger(\theta)} \\ &= c_i \quad . \end{aligned}$$

□

## Bibliography


Bacchus, F., Kyburg, Jr., H. E., and Thalos, M. (1990). Against conditionalization. *Synthese*, 84(3):475–506.

Cassell, L. (2025). Time-slice epistemology for Bayesians. *Inquiry*, 68(2):428–451.

Cover, T. M. and Thomas, J. A. (1991). *Elements of information theory*. John Wiley and Sons, New York, second (2006) edition.

Douven, I. (2012). Learning conditional information. *Mind & Language*, 27(3):239–263.

Earman, J. (1992). *Bayes or bust? A critical examination of Bayesian confirmation theory*. MIT Press, Cambridge MA.

Easwaran, K. (2011). Bayesianism i: Introduction and arguments in favor. *Philosophy Compass*, 6(5):312–320.

Elga, A. (2000). Self-locating belief and the sleeping beauty problem. *Analysis*, 60(2):143–147.

Eva, B., Hartmann, S., and Rad, S. R. (2020). Learning from conditionals. *Mind*, 129(514):461–508.

Fischer, M. (2025). Time-slice Bayesianism as a potential solution to the problem of dilation and reflection for imprecise probabilities. In Destercke, S., Erreygers, A., Nendel, M., Riedel, F., and Troffaes, M. C. M., editors, *Proceedings of the Fourteenth International Symposium on Imprecise Probabilities: Theories and Applications*, volume 290 of *Proceedings of Machine Learning Research*, pages 148–157. PMLR.

Galavotti, M. C. (2005). *Philosophical introduction to probability*. CSLI Publications, Stanford.

Gillies, D. (2000). *Philosophical theories of probability*. Routledge, London and New York.

Hart, C. and Titelbaum, M. G. (2015). Intuitive dilation? *Thought: A Journal of Philosophy*, 4(4):252–262.

Hawthorne, J., Landes, J., Wallmann, C., and Williamson, J. (2017). The Principal Principle implies the Principle of Indifference. *British Journal for the Philosophy of Science*, 68:123–131.

Hedden, B. (2015). Time-slice rationality. *Mind*, 124(494):449–491.

Howson, C. (1997). Bayesian rules of updating. *Erkenntnis*, 45:195–208.

Jaynes, E. T. (1957). Information theory and statistical mechanics II. *Physical Review*, 108(2):171–190.

Jaynes, E. T. (2003). *Probability theory: the logic of science.* Cambridge University Press, Cambridge.

Joyce, J. M. (2011). The development of subjective Bayesianism. In Gabbay, D. M., Hartmann, S., and Woods, J., editors, *Inductive Logic*, volume 10 of *Handbook of the History of Logic*, pages 415–475. North-Holland, Amsterdam.

Kyburg, H. E. (1980). Conditionalization. *Journal of Philosophy*, 77(2):98–114.

Kyburg, H. E. (2006). Belief, evidence, and conditioning. *Philosophy of Science*, 73(1):42–65.

Landes, J., Rad, S. R., and Williamson, J. (2026). *Inductive logic: a Bayesian approach.* Oxford University Press, Oxford. Forthcoming.

Landes, J., Wallmann, C., and Williamson, J. (2021). The Principal Principle, admissibility, and normal informal standards of what is reasonable. *European Journal of Philosophy of Science*, 11(2):36.

Lewis, D. K. (1980). A subjectivist's guide to objective chance. In *Philosophical papers*, volume 2, pages 83–132. Oxford University Press (1986), Oxford. With postscripts.

Lin, H. (2022). Bayesian epistemology. In Zalta, E. N. and Nodelman, U., editors, *Stanford Encyclopedia of Philosophy.* Metaphysics Research Lab, Stanford University, Summer 2024 edition.

Oaksford, M. and Chater, N. (2020). New paradigms in the psychology of reasoning. *Annual Review of Psychology*, 71(1):305–330. PMID: 31514580.

Pettigrew, R. (2020). The Principal Principle does not imply the Principle of Indifference. *British Journal for the Philosophy of Science*, 71(2):605–619.

Rosenkrantz, R. D. (1977). *Inference, method and decision: towards a Bayesian philosophy of science.* Reidel, Dordrecht.

Seidenfeld, T. (1986). Entropy and uncertainty. *Philosophy of Science*, 53(4):467–491.

Shackel, N. and Rowbottom, D. P. (2020). Bertrand's paradox and the maximum entropy principle. *Philosophy and Phenomenological Research*, 101(3):505–523.

Talbott, W. J. (1991). Two principles of Bayesian epistemology. *Philosophical Studies*, 62:135–150.

Titelbaum, M. G. and Hart, C. (2020). The Principal Principle does not imply the principle of indifference, because conditioning on biconditionals is counterintuitive. *British Journal for the Philosophy of Science*, 71(2):621–632.

Wallmann, C. and Williamson, J. (2020). The Principal Principle and subjective Bayesianism. *European Journal for the Philosophy of Science*, 10(1):3.

Williamson, J. (2010). *In defence of objective Bayesianism.* Oxford University Press, Oxford.

Williamson, J. (2017). *Lectures on inductive logic.* Oxford University Press, Oxford.

Williamson, J. (2022). One philosopher's modus ponens is another's modus tollens: pantomemes and nisowir. *Metaphilosophy*, 53(2–3):284–304.

Williamson, J. (2023). Bayesianism from a philosophical perspective and its application to medicine. *International Journal of Biostatistics*, 19(2):295–307.