Bayesianism & Explanationism

Branden Fitelson

fitelson.org/berlin.pdf

There are two main kinds of epistemic requirements. First, let's examine these in the case of full (all-or-nothing) belief.

- **Correctness Requirements** (for full belief)
  - If p is false, then believing that p is incorrect.
  - Correctness requirements are *factive* — they involve relations between one's attitudes and *the facts* [27, 22, 28].
- **Rational Requirements** (for full belief)
  - If one believes both p and $\neg p$, then one's beliefs are *structurally* irrational (*viz.*, *incoherent*).
  - If one's total evidence K counter-supports p, then believing that p is *substantively* irrational.
  - Rational requirements are *non-factive* — they involve **either** (a) relations (of coherence) among one's attitudes (*structural* rationality) **or** (b) relations between one's total evidence K and one's attitudes (*substantive* rationality) [29].

This distinction also applies to Bayesian epistemology…

Let us suppose that our agent is equipped with a *confidence ordering* $\succeq$ over propositions. So that $p \succeq q$ iff S is *at least as confident* in the truth of p as in the truth of q.

And, $p \succ q$ iff S is *strictly more confident* in p than in q.

- **Correctness Requirements** (for comparative confidence)
  - If p is false and q is true, then $p \succ q$ is incorrect.
    - Heuristically, correctness requirements involve *agreement with the attitudes of an omniscient agent* [14, 12, 18].
- **Rational Requirements** (for comparative confidence)
  - Intransitive $\succeq$ relations are *structurally* irrational. *E.g.*, the *combination* of attitudes $\{p \succ q, q \succ r, r \succ p\}$ is *incoherent*.
  - If one's total evidence K supports p strictly more strongly than K supports q, then $q \succeq p$ is *substantively* irrational.

Bayesians also like to talk about *numerical degrees of confidence* (credences). Indeed, this will be our main focus today…

Let us suppose that our agent has (at each time t) *degrees of confidence* (*credences*) represented by a function $\mathrm{cr}_t(\cdot)$.

The function $\mathrm{cr}_t(\cdot \mid E)$ will reflect the agent's degrees of confidence (at t) — *on the indicative supposition that* E is true.

Finally, suppose our agent learns (exactly) E (with certainty) between t_0 and t_1. So, her transition from $\mathrm{cr}_{t_0}(\cdot)$ to $\mathrm{cr}_{t_1}(\cdot)$ reflects the upshot of learning (precisely) the content E.

Structural Bayesianism involves a Trinity of Requirements [23].

1. **Synchronic (Non-Suppositional) Probabilism.** At each time t, the agent's (non-suppositional) credence function $\mathrm{cr}_t(\cdot)$ should obey the (Komlogorov) probability axioms.
2. **Synchronic (Suppositional) Ratio Formula.** At each time t, the agent's (suppositional) credence function $\mathrm{cr}_t(\cdot \mid E)$ should obey the ratio formula $\mathrm{cr}_t(\cdot \mid E) = \frac{\mathrm{cr}_t(\cdot \,\&E)}{\mathrm{cr}_t(E)}$.
3. **Diachronic Conditionalization.** $\mathrm{cr}_{t_1}(\cdot)$ should equal $\frac{\mathrm{cr}_{t_0}(\cdot \,\&E)}{\mathrm{cr}_{t_0}(E)}$.

The Fourth Pillar of Structural Bayesianism *follows from* (1)–(3).

4. **Learning & Supposing.** $\mathrm{cr}_{t_1}(\cdot)$ should equal $\mathrm{cr}_{t_0}(\cdot \mid E)$.

de Finetti [3] gave Dutch Book Arguments for both (1) and (2).

Accuracy-dominance arguments for (1) are now popular [18]. One can also give an accuracy-dominance argument for (2) [10].

Lewis and others have aimed to adapt de Finetti's argument for (2) into a "diachronic Dutch Book" argument for (3) [19].

Like de Finetti, I view (1) and (2) as the *fundamental* Bayesian principles governing the (structural) rationality of credences.

I have always been more skeptical about the existence and nature of "diachronic coherence requirements." Indeed, there seem to be *many* reasons to worry about (3) & (4) [11, 30, 24].

I will focus, primarily, on the synchronic requirements (1) & (2), and how they (allegedly) interact with Explanationism.

There are also *substantive* Bayesian requirements. In general, these are of the following generic form: If one's total evidence at time t (K_t) is such and so, then $\mathrm{cr}_t(\cdot)$ should be thus and such.

Let $\{H_1, \dots, H_n\}$ be some partition of alternative hypotheses (putative explanations of E) entertained by an agent (at time t).

Substantive Bayesianism (some example requirements)

- **The Principle of Indifference** [7]. If K_t does not favor any H_i over any H_j, then $\mathrm{cr}_t(H_i)$ should equal $\mathrm{cr}_t(H_j)$, $\forall\ i, j$.
- **The Principal Principle** [17]. If K_t entails that the objective chance of H_i is c (and K_t doesn't contain/imply any inadmissible evidence), then $\mathrm{cr}_t(H_i)$ should equal c.
- **The Requirement of Total Evidence** [15, 1, 28]. $\mathrm{cr}_t(H_i)$ should be equal to the evidential probability $\Pr(H_i \mid K_t)$.

Next up: Explanationism. I will follow Douven's [5] recent discussion of (the various explications of) Explanationism.

As Douven [5] explains, there have been various historical views regarding the proper formulation of Explanationism.

Rather than rehearsing Douven's list of historical explications in detail, I will remain at a higher level of abstraction.

☞ The basic idea behind Explanationism is that some "epistemic credit" should accrue to a hypothesis H in virtue of its being the best explanation of E (among the available alternatives $\{H_k\}$).

The question in which I am interested is: What is the best way to accommodate this basic Explanationist idea — of "credit" accruing to E's best explanation — *within a Bayesian framework*?

van Fraassen [25] and Douven [5] maintain that a Bayesian should incorporate Explanationism *by revising some of the basic requirements of Structural Bayesianism: (1)–(3).*

I will focus on Douven's proposal, since it is more precise, and it can be couched in purely synchronic terms [as a revision of (2)].

Douven recommends that Bayesians *revise the Ratio Formula* (2) in such a way that the following alternative to Bayes's Theorem is adopted (for hypotheses $H_i \in \{H_1, \dots, H_n\}$ and evidence E).

EXPL. $$\mathrm{cr}_t(H_i \mid E) = \frac{\mathrm{cr}_t(H_i) \cdot \mathrm{cr}_t(E \mid H_i) + \mathfrak{c}(H_i, E)}{\sum_{k=1}^{n} [\mathrm{cr}_t(H_k) \cdot \mathrm{cr}_t(E \mid H_k) + \mathfrak{c}(H_k, E)]},$$

where $\mathfrak{c}(H, E) \in [0, 1)$ is H's "E-abductive credit score." And, $\mathfrak{c}(H, E) > 0$ *iff* H *best explains* E (o.w. $\mathfrak{c}(H, E) = 0$).[1]

Note that *if there is no best explanation* of E among the $\{H_k\}$, then **EXPL** *reduces to Bayes's Theorem* (since all of the credit scores $\mathfrak{c}(H_k, E)$ will be equal to zero in such a case).

Because **EXPL** leads to violations of Structural Bayesianism [*viz.*, (2)], Douven discusses various ways a defender of **EXPL** might respond to de Finetti's [3] Dutch Book argument for (2).

[1] This definition can't be quite right, since sometimes $\mathfrak{c}(H, E)$ will need to be ≥ 1 in order to emulate some Bayesian abductive updates. See Extras.

☞ There is a more elegant way to capture the idea of "abductive credit," which avoids the probabilistic incoherencies of [5, 25].

Note that there is something odd about thinking of **EXPL** as a *structural* requirement in the first place. **EXPL** has this form:

> **EXPL.** $\mathrm{cr}_t(H_i \mid E)$ should receive a boost — over and above the value prescribed by Bayes's Theorem — just in case $(\mathbb{A}_i)$ H_i is the best explanation of E (among the $\{H_k\}$).

Read literally, then, **EXPL** relates $\mathrm{cr}_t(H_i \mid E)$ to *the **fact*** that H_i best explains E, which makes **EXPL** a *correctness* requirement.

In order for **EXPL** to be a *structural* requirement, it would have to be stated *as a relation among the agent's credences.* To wit:

> **EXPL$_1$.** $\mathrm{cr}_t(H_i \mid E)$ should receive a boost — over & above its Bayes's Thm value — iff ***the agent is certain at*** t ***that*** $\mathbb{A}_i$.[2]

[2] We also have to be careful to allow this "boost" to occur *only once* — presumably, when $\mathbb{A}_i$ is *first learned*. Otherwise, H_i will be "over-boosted."

Even **EXPL$_1$** is arguably not a (pure) structural requirement, since it only constrains agents who *entertain* $\mathbb{A}_i$, for some H_i and E.

Pure structural requirements (*e.g.*, probabilism) do not presuppose anything about the *contents* of an agent's attitudes.[3]

☞ If **EXPL** is going to be a (non-vacuous) *rational* requirement, then it must presuppose that the agent *entertains* $\mathbb{A}_i$, for some H_i and E. And, in that case, I think a preferable explication exists.

> **Evidential Relevance of Abduction (ERA).** Let $\mathbb{A}_j$ assert that H_j is the best (or perhaps the *only* [4, 8]) explanation of E (among the available $\{H_k\}$). Then, it is — in *some* cases — *rationally required* that an agent's credence function at t_0 be such that
>
> $$\mathrm{cr}_{t_0}(H_j \mid E \,\&\, \mathbb{A}_j) > \mathrm{cr}_{t_0}(H_j \mid E \,\&\, \neg\mathbb{A}_j).$$

[3] One could also interpret **EXPL** as a relation between *S's total evidence at t* $[K_t]$ and their credences at t $[\mathrm{cr}_t(\cdot)]$. But, as with **EXPL$_1$**, we would have to be careful to state such a requirement so as to allow *only one* "boost" to occur, per hypothesis — presumably, when K_t comes to entail $\mathbb{A}_j$ *for the first time.*

I'm not the first one to propose something like **ERA.** Climenhaga, Hartmann *et. al.*, Lange, and Weisberg have all argued convincingly for similar principles [2, 16, 26, 4, 13, 8].

Roche & Sober [20] agree that **ERA** is the right *formulation* of Explanationism; but, they argue that **ERA** is *false* — *viz.*, that $\mathbb{A}_j$ is *never* relevant to H_j (on the supposition of the evidence E).

I think Climenhaga [2] and Lange [16] do a pretty good job of responding to Roche & Sober's skeptical argument [20].

The main thing I would add to Climenhaga's and Lange's trenchant responses Roche & Sober is the following point.

☞ In order to refute R&S's skepticism, all that is required is *a single example* in which $\mathbb{A}_j$ *is* relevant to H_j (given E).

I will close by discussing just such an example (involving Newton, Einstein, and the motion of Mercury), which is a well-known instance of The Problem of Old Evidence [6].

Let $E \overset{\text{def}}{=}$ what was (collectively) known about the precession of the perihelion of Mercury in $t_1 = 1914$.[4]

Let $H_1 \overset{\text{def}}{=}$ Newton's theory of planetary motion, and $H_2 \overset{\text{def}}{=}$ Einstein's theory of general relativity. This yields the following 3-element partition of hypotheses: $\{H_1, H_2, H_3 = \neg H_1 \,\&\, \neg H_2\}$.

If **ERA** applies in this case, then our agent should be such that

$$\mathrm{cr}_{t_0}(H_2 \mid E \,\&\, \mathbb{A}_2) > \mathrm{cr}_{t_0}(H_2 \mid E \,\&\, \neg\mathbb{A}_2).$$

Thus, after E was learned (between t_0 and t_1), we should have

$$\mathrm{cr}_{t_1}(H_2 \mid \mathbb{A}_2) > \mathrm{cr}_{t_1}(H_2 \mid \neg\mathbb{A}_2).$$

So, when $\mathbb{A}_2$ was subsequently learned in $t_2 = 1915$, *it* — and *not* E, which was *old evidence* at t_2 — provided a boost to H_2 [13, 8].

[4] Of course, E had been known for many years prior to 1914. It is widely agreed that t_0 would have been no later than 1882, when Newcomb calculated the Newtonian discrepancy in Mercury's precession precisely.

How would **EXPL**$_1$ handle this Old Evidence Problem?

Let us suppose that our agent became certain of $\mathbb{A}_2$ in $t_2 = 1915$.

Then, according to **EXPL**$_1$, H_2 receives a "positive credit score" [$\mathfrak{c}(H_2, E) > 0$] and a boost in credence (at t_2) — over and above what Bayes's Thm would recommend. Hence, **EXPL**$_1$ yields

$$\mathrm{cr}_{t_2}(H_2 \mid E) > \mathrm{cr}_{t_2}(H_2).$$

In this sense, according to **EXPL**$_1$, E *confirms* H_2 *in 1915*.

☞ I see (at least) two problems with this application of **EXPL**$_1$.

- It is *misleading* to say that E is what confirms H_2 in 1915. It is $\mathbb{A}_2$ — by ensuring $\mathfrak{c}(H_2, E) > 0$ — that is doing the work.
- In addition to the machinery of probability theory [modulo (2)], Douven also needs a "theory of credit scores." In this sense, the (Bayesian) **ERA** approach is *more parsimonious*. [See Extras for a detailed formal example, which brings this out.]

The simplest formal example of Douven *v.* Bayes involves a 2-element partition $\{H_1 = H, H_2 = \neg H\}$, evidence E, and an abductive claim $\mathbb{A}$ asserting that H explains E better than $\neg H$.

Our Bayesian agent will begin with a prior $\mathrm{cr}_{t_0}(\cdot)$ over the algebra generated by the three atoms $H, E, \mathbb{A}$. Between t_0 and t_1, they will learn E, and between t_1 and t_2 they will learn $\mathbb{A}$.

I will assume $\mathrm{cr}_{t_0}(\cdot)$ satisfies the following six (6) constraints.

(i) $\mathrm{cr}_{t_0}(\cdot)$ is regular.

(ii) E (initially) *confirms* H [$\mathrm{cr}_{t_0}(H \mid E) > \mathrm{cr}_{t_0}(H \mid \neg E)$].

(iii) **ERA** applies [$\mathrm{cr}_{t_0}(H \mid E \,\&\, \mathbb{A}) > \mathrm{cr}_{t_0}(H \mid E \,\&\, \neg\mathbb{A})$].

(iv) E is (*a priori*) *irrelevant* to $\mathbb{A}$. [$\mathrm{cr}_{t_0}(\mathbb{A} \mid E) = \mathrm{cr}_{t_0}(\mathbb{A} \mid \neg E)$].

(v) $\mathbb{A}$ is (*a priori*) *irrelevant* to H. [$\mathrm{cr}_{t_0}(H \mid \mathbb{A}) = \mathrm{cr}_{t_0}(H \mid \neg\mathbb{A})$].

(vi) $\mathrm{cr}_{t_0}(H) = \mathrm{cr}_{t_0}(E) = \mathrm{cr}_{t_0}(\mathbb{A}) = \frac{1}{2}$.

There are many priors that satisfy (i)–(vi). I will choose a relatively simple one for illustrative purposes.

$\mathbb{A}$	E	H	$\mathrm{cr}_{t_0}(\cdot)$	$\mathrm{cr}_{t_1}(\cdot) = \mathrm{cr}_{t_0}(\cdot \mid E)$	$\mathrm{cr}_{t_2}(\cdot) = \mathrm{cr}_{t_0}(\cdot \mid E \,\&\, \mathbb{A})$
T	T	T	7/32	7/16	7/8
T	T	F	1/32	1/16	1/8
T	F	T	1/32	0	0
T	F	F	7/32	0	0
F	T	T	3/32	3/16	0
F	T	F	5/32	5/16	0
F	F	T	5/32	0	0
F	F	F	3/32	0	0

Our Bayesian agent is such that $\mathrm{cr}_{t_2}(H) = \mathrm{cr}_{t_0}(H \mid E \,\&\, \mathbb{A}) = 7/8$.

Douven's rule for computing $\mathrm{cr}_{t_2}(H \mid E)$ will yield the Bayesian answer of $7/8$ here *only if* H's credit score is $\mathfrak{c}(H, E) = 1 \notin [0, 1)$.

In fact, we'll need to allow $\mathfrak{c}(H, E) \in [0, \infty)$. See my `PrSAT` [9] notebook `fitelson.org/douven.pdf (.nb)` for technical details.

☞ I would argue that this is how the values of $\mathfrak{c}(H, E)$ should be "reverse engineered." But, then, why *not* just go *fully Bayesian*?

[1] R. Carnap, *Logical Foundations of Probability*, 1950.
[2] N. Climenhaga, *How explanation guides confirmation*, 2017.
[3] B. de Finetti, *Foresight: Its Logical Laws, Its Subjective Sources*, 1937.
[4] R. Dawid, S. Hartmann, and J. Sprenger, *The No Alternatives Argument*, 2015.
[5] I. Douven, *The Art of Abduction*, 2022.
[6] E. Eells, *Bayesian Problems of Old Evidence*, 1990.
[7] B. Eva, *Principles of Indifference*, 2019.
[8] B. Eva and S. Hartmann, *On the Origins of Old Evidence*, 2020.
[9] B. Fitelson, *A Decision Procedure for Probability Calculus with Applications*, 2008.
[10] J.D. Gallow, *Learning and value change*, 2019.
[11] J.D. Gallow, *Updating for externalists*, 2021.
[12] A. Gibbard, *Rational Credence and the Value of Truth*, 2006.
[13] S. Hartmann and B. Fitelson, *A new Garber-style solution to the problem of old evidence*, 2015.
[14] J. Joyce, *Prospects for an Alethic Epistemology of Partial Belief*, 2009.
[15] J.M. Keynes, *A Treatise on Probability*, 1921.
[16] M. Lange, *Putting explanation back into "inference to the best explanation,"* 2020.
[17] D. Lewis, *A Subjectivist's Guide to Objective Chance*, 1980.
[18] R. Pettigrew, *Accuracy and the Laws of Credence*, 2016.
[19] R. Pettigrew, *The Dutch Book Arguments*, 2020.
[20] W. Roche and E. Sober, *Explanatoriness is evidentially irrelevant...*, 2013.
[21] J. Thomson, *Normativity*, 2008.
[22] M. Titelbaum, *Fundamentals of Bayesian Epistemology*, 2022.
[23] M. Titelbaum, *Quitting Certainties*, 2013.
[24] B. van Fraassen, *Laws and symmetry*, 1989.
[25] J. Weisberg, *Locating IBE in the Bayesian Framework*, 2009.
[26] B. Williams, *Internal and External Reasons*, 1981.
[27] T. Williamson, *Knowledge and its Limits*, 2000.
[28] A. Worsnip, *Fitting Things Together: Coherence and the Demands of Structural Rationality*, 2021.
[29] J. Zhao, V. Crupi, K. Tentori, B. Fitelson, and D. Osherson, *Updating: Learning vs. supposing*, 2012.