

Degrees of Belief — II

HT2017 / Dr Teruji Thomas

Website: users.ox.ac.uk/~mert2060/2017/Degrees-of-belief

1 Conditionalisation

Where we have got to:

- * One reason to focus on credences instead of beliefs: response to evidence.
- * First core claim of orthodox Bayesian epistemology: *Probabilism* – rational credences can be quantified in a way that obeys the mathematics of probability.
- * But this doesn't tell us how credences should change over time.
Synchronic versus *diachronic* norms.

Basic Question: How should credences change over time? What diachronic norms are there?

1.1 Second core claim of orthodox Bayesian epistemology

Conditionalisation. Suppose you gain evidence E . Let Cr be your credences just before and Cr_{new} your credences just afterwards. Then, insofar as you are rational, for any proposition P

$$Cr_{\text{new}}(P) = \frac{Cr(P \text{ and } E)}{Cr(E)} \stackrel{\text{def}}{=} Cr(P|E)$$

(‘your old credence in P conditional on E ’).

Heuristically it's the proportion of cases in which E is true in which P is also true.

EXAMPLE 1. Clara initially has credence $1/2$ that it's sunny in Oxford, $1/2$ that it's sunny in London, and $1/3$ that it's both. She gains the evidence that it's sunny in London. Her new credence that it's sunny in Oxford is $\frac{1/3}{1/2} = 2/3$. (Roughly: she ‘thinks’ that $2/3$ of the time that it's sunny in London it's also sunny in Oxford.)

EXAMPLE 2. (THE BASE RATE FALLACY) There's a new blood test for a rare disease which is symptomless until it kills you. On a whim, you decide to get checked. The test is very accurate, in the following sense: everyone sick gets a positive result, and only 0.1% of healthy people get a false positive.

Your test is positive. How confident should you now be that you have the disease?

1.2 The Dutch Book Argument

Slight variation on last time:

Betting Principle: You're indifferent between having (a) a promise of $\pounds x$ if P is true, or (b) an extra $\pounds(x \text{ times } Cr(P))$.

The argument. Let PROM be a promise of $\pounds 1$ -if- P .

1. By the Betting Principle, if E comes about, you will be indifferent between PROM and $\pounds Cr_{\text{new}}(P)$.
2. To be consistent, you should, right now, be indifferent between a promise of PROM-if- E and a promise of $\pounds Cr_{\text{new}}(P)$ -if- E .

3. A promise of PROM-if- E is exactly the same as a promise of £1-if- P -and- E .
4. So you should, right now, be indifferent between a promise of £1-if- P -and- E and a promise of £ $Cr_{new}(P)$ -if- E .
5. You are, right now, indifferent between a promise of £1-if- P -and- E and £ $Cr(P \text{ and } E)$.
6. You should, right now, be indifferent between £ $Cr(P \text{ and } E)$ and a promise of £ $Cr_{new}(P)$ -if- E .

By the Betting Principle, (6) says that

$$Cr(P \text{ and } E) = Cr_{new}(P)Cr(E).$$

Rearranging, we find $Cr_{new}(P) = Cr(P \text{ and } E)/Cr(E)$.

1.3 Problem 1: Too much certainty?

According to conditionalisation, we become certain of the new evidence E : $Cr(E|E) = 1$. This is pretty strong; it's hard to imagine being completely certain of any ordinary empirical proposition. Indeed, if you are certain of E , you cannot ever, by conditionalisation, become less than certain of E .

A generalisation: 'Jeffrey Conditionalisation'

The agent inspects a piece of cloth by candlelight, and gets the impression that it is green, although he concedes that it might be blue or even (but very improbably) violet. If G , B , V are the propositions that the cloth is green, blue, and violet, respectively, then the outcome of the observation might be that...his degrees of belief in those same propositions...are .70, .25, and .05. If there were a proposition E in his preference ranking which described the precise quality of his visual experience in looking at the cloth, one would say that what the agent learned from the observation was that E is true...But there need be no such proposition E in his preference ranking; nor need any such proposition be expressible in the English language. (Jeffrey, p. 165)

Given the involuntary change in our credences about the colour of the cloth, 'Jeffrey conditionalisation' determines how our *other* credences ought to change. (Special case: if you became certain that the cloth was green, you would just conditionalise on G .)

BASIC ISSUE: Why .7, .25, and .05? In general: how should we understand 'evidence'?

One option for ordinary conditionalisation: your evidence at a given time is what you believe or (Williamson) what you *know*.

- See Jeffrey and Teller for arguments for Conditionalization and Jeffrey's generalization. See Williamson for evidence-as-knowledge.

1.4 Problem 2: Rational forgetfulness?

EXAMPLE. Clara just forgets what the weather forecast said. What should happen to her credences?

POSSIBLE RESPONSE: Forgetting is irrational?

Shangri La, Part I. (Arntzenius)

Some monks are taking Clara to Shangri La. A fair coin will be tossed by the monks to determine which path she will take: if heads, by the mountains,

if tails, by the sea. The coin lands heads. Clara journeys through the mountains on Sunday and enters Shangri La at midnight. On Monday she enjoys recalling the beautiful scenery.
How certain should Clara be on Monday that she was in the Mountains?

Shangri La, Part II

Something else Clara knows: if the coin had landed tails, and she had gone by sea, then, as soon as she entered Shangri La, the monks would have wiped her memories of the journey (and the coin toss) and replaced it with memories of a journey through the mountains. (Remember: the coin in fact landed heads.)

THE POINT: Clara's credences should change between Sunday and Monday, but not by conditionalisation. Moreover, there's nothing obviously non-ideal about her.

Reflection. Shangri-La is also a counterexample to another diachronic norm, proposed by van Frassen. Simplest case:

If you are certain that your future self will have credence $1/2$ that P is true, then you now have credence $1/2$ that P is true.

The idea is that your future self knows at least as much about P as you do – your future self is an 'expert', to whom you should defer. But this isn't always true.

Sleeping Beauty (Elga), Part I. It's Sunday morning. Clara knows that Dr Smith is going to give her a sedative and then wake her up on Monday morning. She also knows that, later on Monday, Dr Smith is going to toss a fair coin.
How confident should Clara be, when she wakes up on Monday, that the coin will land heads? LEWIS: One half, surely...

Sleeping Beauty, Part II. Something else Clara knows: A few minutes after she awakes on Monday, Dr Smith will give her a sedative, and wake her again on Tuesday. But *if the coin lands tails*, Dr Smith will first wipe Clara's memory of being awake on Monday.

ELGA: Clara's credence on Monday that the coin lands heads should be $1/3$, not $1/2$.

ARGUMENT 1. When Clara wakes on Monday, there are three scenarios that are equally supported by her evidence. Only one of them involves the coin landing heads.

ARGUMENT 2. If Clara is *told* that it is Monday, then she is definitely in the normal case, $\text{Cr}(\text{Heads}|\text{Monday}) = 1/2$. But being told it's Monday must make her more confident in Heads, since it rules out (Tuesday, Tails). So she must have $\text{Cr}(\text{Heads}) < 1/2$.

If Elga is right, then Clara's credences change from Sunday to Monday, not by conditionalisation, nor through any fault of her own.

Question. Grant that conditionalization is not the whole story; can we give a systematic account of cases like Sleeping Beauty?

ONE STRATEGY: GIVE UP ON DIACHRONIC NORMS. A synchronic alternative: 'At each time, apportion your credences to your evidence'. (If you gain new evidence, or you lose evidence, your credences should change. But there's no particular difference between learning and forgetting.)

A little more precisely: you've got an 'original' or '**ur-prior**' credence function Cr_0 . The norm is that, if at time t you have total evidence E , then your credence that P is true is $\text{Cr}_0(P|E)$.

- See Elga, Lewis, and Titelbaum for Sleeping Beauty; see Arntzenius for Shangri La and many other puzzles. See Meacham for the ur-prior move.

2 Bayesian Confirmation Theory.

Example. Even if a positive test shouldn't make you *very* confident that you are sick, it should make you *more* confident. In this sense, the positive test *is evidence* or *supports* or '*confirms*' that you are sick.

General Question of 'Confirmation Theory': When does a piece of evidence E support a hypothesis H ?

Basic Bayesian Answer: Learning E would support H just in case $\text{Cr}(H|E) > \text{Cr}(H)$. This also suggests various numerical measures of *how strongly* E supports H , for example

$$\text{Cr}(H|E) - \text{Cr}(H) \quad \text{or} \quad \frac{\text{Cr}(H|E)}{\text{Cr}(H)} \quad \text{or} \quad \log \text{Cr}(H|E) / \text{Cr}(H) \quad \text{or} \dots$$

Question. Does this notion of confirmation/support match with intuitions and/or scientific practice?

Some intially intuitive axioms:

Nicod's Condition. 'All F s are G s' would be supported by the observation of an F that is G .

EXAMPLE. Hypothesis: all ravens are black. Natural thought: let's prove it by going around and observing ravens.

Transitivity. If E supports H_1 , and H_1 supports H_2 , then E supports H_2 .

EXAMPLE. That you are vegan is evidence that you care about the environment. That you care about the environment is evidence that you would consider voting Green. So that you are vegan is evidence that you would consider voting Green.

The Bayesian notion of confirmation does *not* validate either of these axioms. But that's arguably a good thing —

EXAMPLES. The traditional 'paradox' for Nicod's condition is that 'All ravens are black' is logically equivalent to 'All non-black things are non-ravens'. By Nicod's condition we can confirm 'All non-black things are non-ravens', and (therefore?) 'All ravens are black', simply by observing a non-raven that is not black. And that seems ridiculous.

1. Observing black ravens:
 - (a) You are certain that there exists exactly one raven, but you are not sure whether it is black.
 - (b) (Titelbaum) You're at the zoo, visiting the Enclosure of Atypically Coloured Birds...
2. Observing non-black non-ravens?...
3. Transitivity?

Question. What can be said in favour of Nicod's criterion (or transitivity, etc.)? E.g. Are there recognizably 'normal' circumstances in which they are validated by the Bayesian notion of confirmation? Or is there a *different* notion of confirmation that validates them?

- See Hájek-Joyce and Howson-Urbach for overviews of the problems of confirmation theory and the Bayesian approach to them.

2.1 Problems: Subjectivity and Old Evidence

(1) Isn't there a more 'objective' notion of confirmation – one that isn't relative to a single person's credences?

SOME RESPONSES:

- * It's not that obvious that we need an objective standard for confirmation, rather than a widely shared one.
- * Instead of talking about confirmation relative to someone's actual credences, we can talk about confirmation relative to what credences they *ought* to have – this could be less subjective.

(2) ('The Problem of Old Evidence') The Bayesian tells us how much learning E would support H , but sometimes we say that E supports H even when we already know E . (If you know E , then $\text{Cr}(H|E) = \text{Cr}(H)$: no confirmation.)

EXAMPLE. 'The precession of the perihelion of Mercury supports the theory of General Relativity' – astrophysicists all think this, but they know all about the precession.

SOME RESPONSES:

- * Sometimes, in real life, we know E but we don't understand the relation between E and H . E.g. we know about the precession, but not that GR predicts it. Then learning the relationship makes us more confident of H . (Garber and others)
- * ' E supports H ' means something like 'If we didn't know about E , then learning E would make us more confident about H '. (Howson and others)

EXAMPLE (???) Clara is an astrophysicist. But if Clara didn't know about the precession, she wouldn't be an astrophysicist. She wouldn't have whatever other background knowledge you need to derive the precession from GR. Under such conditions, learning about the precession might not make her more confident in GR.

- See Glymour and Howson for the Problem of Old Evidence.

3 Summing up

1. Orthodox Bayesians accept Probabilism – a set of synchronic norms – but also some diachronic norms – especially Conditionalisation.
2. Various examples suggest that Conditionalisation isn't the whole story, and raise questions about how we should understand 'evidence.'
3. Still, conditionalisation gives a fairly attractive account of inductive reasoning, clarifying many traditional problems.
4. But it isn't completely straightforward to account for all the ways in which we talk about evidential support or confirmation.

Further Reading

Conditionalisation

Teller, P. (1973) 'Conditionalization and Observation', *Synthese* 26: 218–258.

[Considers several arguments for conditionalisation (and for Jeffrey conditionalisation), including the Dutch Book argument (pp. 222–225).]

Jeffrey, R. C. (1965) 'Probability Kinematics', chapter 11 in his *The Logic of Decision*; reprinted in Eagle.

[Jeffrey's own explanation of his generalisation of conditionalisation.]

Williamson, T. (2002) 'Evidence', chapter 9 in *Knowledge and Its Limits*.

[Defends the view that your evidence is what you know. In chapter 10 he uses it to develop a version of objective Bayesianism.]

Meacham, J. G. M. (2016) 'Ur-Priors, Conditionalization, and Ur-Prior Conditionalization'. *Ergo* 3 (17).

[Surveys different versions of the synchronic norm that we should apportion our credences to our current evidence, and the many advantages of this type of view.]

Self-Locating and Forgetting

Titelbaum, M. G. (2013) 'Ten Reasons to Care About the Sleeping Beauty Problem', *Philosophy Compass* 8 (11):1003-1017.

[A short survey with useful references connecting Sleeping Beauty to many wider issues, including anthropic reasoning and the interpretation of quantum mechanics.]

Elga, A. (2000). "Self-locating Belief and the Sleeping Beauty Problem". *Analysis*. 60 (2): 143–147. Reprinted in Eagle.

[Introduced Sleeping Beauty to philosophers, defending the 1/3 answer.]

Lewis, D. (2001). "Sleeping Beauty: reply to Elga". *Analysis*. 61 (3): 171–76.

[Defends the 1/2 answer.]

Arntzenius, F. (2003) 'Some Problems for Conditionalization and Reflection'. *Journal of Philosophy* 100 (7):356–370. Reprinted in Eagle.

[Analyses a variety of interesting cases involving the loss of evidence and the passage of time.]

Confirmation

Hájek A. and Joyce, J. M. (2008). 'Confirmation'. In: *The Routledge Companion to Philosophy of Science*. Ed. by Stathis Psillos and Martin Curd. Routledge, pp. 115–128

[An brief overview of different approaches to confirmation theory.]

Howson C. and Urbach, P. (1993) 'Bayesian versus non-Bayesian approaches', chapter 7 in their *Scientific Reasoning: The Bayesian Approach*. Reprinted in Eagle, pp. 222–249.

[A more comprehensive explanation of how Bayesian confirmation theory treats a variety of issues.]

Glymour, C. (1980) 'Why I am Not a Bayesian', chapter 3 in his *Theory and Evidence*; reprinted in Eagle.

[The origin of the 'Problem of Old Evidence'.]

Howson, C. (1991). 'The "old evidence" problem'. *British Journal for the Philosophy of Science* 42 (4):547–555.

[Defends a version of the 'counterfactual' solution to the Problem of Old Evidence; discusses some other solutions as well.]