The Laplace-Jaynes approach to induction

Being part II of

"From 'plausibilities of plausibilities' to state-assignment methods"

P. G. L. Porta Mana,* A. Månsson, G. Björk

Kungliga Tekniska Högskolan, Isafjordsgatan 22, SE-164 40 Stockholm, Sweden

(Dated: 12 March 2007)

Abstract

An approach to induction is presented, based on the idea of analysing the context of a given problem into 'circumstances'. This approach, fully Bayesian in form and meaning, provides a complement or in some cases an alternative to that based on de Finetti's representation theorem and on the notion of infinite exchangeability. In particular, it gives an alternative interpretation of those formulae that apparently involve 'unknown probabilities' or 'propensities'. Various advantages and applications of the presented approach are discussed, especially in comparison to that based on exchangeability. Generalisations are also discussed.

PACS numbers: 02.50.Cw,02.50.Tt,01.70.+w

MSC numbers: 03B48,60G09,60A05

> Note, to head off a common misconception, that this is in no way to introduce a "probability of a probability". It is simply convenient to index our hypotheses by parameters [. . .] chosen to be numerically equal to the probabilities assigned by those hypotheses; this avoids a doubling of our notation. We could easily restate everything so that the misconception could not arise; it would only be rather clumsy notationally and tedious verbally.
>
> E. T. Jaynes, *Monkeys, kangaroos, and N* [1], p. 12

1 Dramatis personae, notatio, atque philosophia

We continue here our exploration of the notion of 'circumstance' and of its applications in plausibility theory. Through this notion we shall here approach no less than

*Email: mana@kth.se

the question of induction, i.e., of the prediction of unobserved evens from knowledge of observed (similar) ones. The study can be read independently of the previous one [2], although the two elucidate each other.

The following characters appear in this study:

The 'de Finettian': a scholar who conceives probability as 'degree of belief', pedantically insisting on the subjective nature of all plausibility assignments, and through the notion of infinite exchangeability makes sense of the frequentist's and propensitor's voodooistic practises. The views and tools of this scholar [e.g., 3–11] are supposed known to the reader.

The 'propensitor': a scholar who conceives of a 'physical disposition' of some phenomena to occur with definite relative frequencies, and call this physical characteristic 'propensity'. This scholar sometimes makes inferences about propensities by means of probability as degree of belief.

The 'frequentist': a scholar who conceives probability as 'limit relative frequency' of a series of 'trials' (or something like that). Frequentists are often propensitors at heart.

The 'plausibilist' or 'probability logician', who in this study will be your humble narrator: a scholar who conceives of probability as a formalisation and schematisation of the everyday notions of 'plausibility' and 'probability' — just as truth in formal logic is such a formalisation and schematisation of the everyday notion of 'truth' — and does not dwell more than so in its meaning. The views of this scholar are a distillate of the philosophies and views and/or the works of Laplace [12], Johnson [13], Jeffreys [14–16], Cox [17, 18], Jaynes [19–21], Tribus [22], de Finetti [6, 7], Adams [23, 24], Hailperin [25], and others [e.g., 11, 26–39].

Plausibilists agree with de Finettians on basically all points. The only difference is in emphasis: although they recognise the 'subjective' character of initial plausibility assignments — i.e., that these are matter of convention —, they also think that it is not such a big deal. It should in any case be no concern of plausibility theory, which is 'impersonal in the same sense as in ordinary logic: that different people starting from the same [assignments] would get the same answers if they followed the rules' (Jeffreys [16], reinterpreted). This will surely be self-evident when plausibility theory will reach its maturity, just as it is self-evident in formal logic today. In fact, the same subjective, conventional character is present — no more, no less — in formal logic as regards the initial truth assignments, those which are usually called 'axioms' or 'postulates'. Albeit in formal logic there is perhaps less place for disagreement amongst people about initial truth assignments, the possible choice being there between 'true' and 'false', or '0' and '1' only — and not amongst a continuum of possibilities [0, 1] as in plausibility theory.

The above remark is not meant to diminish the historical importance of de Finetti's Nietzschean work in plausibility theory. Emphasis on subjectivity may still be necessary sometimes. In this study we shall also use terms like 'judgement' to this purpose.

The following notation will be used:

$\mathrm{P}(A|\,B)$ denotes the plausibility of the proposition A conditional on B or, as we shall also say, in the *context B*. The proposition B, is supposed to express all knowledge (including beliefs), data, and 'working hypotheses' on the grounds of which we make the plausibility assignment to A. We are intransigent as regards the necessity and appropriateness of always specifying the context of a plausibility (as well as that of a truth!),[1] a point also stressed by Jeffreys [14, 15], Cox [17, 18], Jaynes [19–21], Tribus [22], de Finetti [47, § 4], Hájek [48], and apparently Keynes [26].

The remaining notation follows ISO [49] and ANSI/IEEE [50] standards.

2 General setting of the question

The question that we are going to touch is, in very general terms, that of induction: *In a given situation, given a collection of observations of particular events, assign the plausibility for another collection of unobserved events.* Note that temporal distinctions are not relevant (we did not say 'past' or 'future' events) and we shall not make any. Of course, in stating this general question we have in mind 'similar' events.[2] So let us call these events 'instances of the same phenomenon', following de Finetti's appropriate terminology [3, 4, 47]. We also suppose to know that each such instance can 'manifest itself' in a constant (and known) number of mutually exclusive and exhaustive 'forms', and there is a clear similarity between the forms of each instance (that is one of the reasons we call the events 'similar').

All this can be restated and made more concrete using a terminology that is nearer to physics; but we must keep in mind that the setting has not therefore become less general. We call the phenomenon a 'measurement',[3] and its instances 'measurement instances'. The forms will be called '(measurement) outcomes'. We can finally state our question thus: *In a given situation, given the observed outcomes of some instances of a particular measurement, assign the plausibility for the unobserved outcomes of some other instances of that measurement.*

We represent the situation, measurements, etc. by propositions. The situation by I, the measurement instances by $M^{(j)}$, the outcomes of the jth instance by $R_i^{(j)}$ (different instances of the same outcome are those with different j but identical i). Instances are thus generally denoted by an index (j) with $j = 1, 2, \ldots$; its range may be infinite or finite, a detail that will be always specified as it will be very important in later discussions. Other propositions will be introduced and defined later. With this representation, our question above simply becomes the assignment of the plausibility

$$\mathrm{P}(R_{i_{N+L}}^{(j_{N+L})} \land \cdots \land R_{i_{N+1}}^{(j_{N+1})} | M^{(j_{N+L})} \land \cdots \land M^{(j_{N+1})} \land R_{i_N}^{(j_N)} \land \cdots \land R_{i_1}^{(j_1)} \land I) \qquad (1)$$

for all possible distinct j_a, distinct i_a, $N \geqslant 0$, and $L > 0$.

[1]The context may also have a clarifying rôle in formal logic. Cf. e.g. the studies by Adams [23, 24, 40], Lewis [41, 42], Hailperin [25, 43], Barwise [44, 45], and Gaifman [46].

[2]But 'similar' in which sense? The answer to this question cannot be given by plausibility theory, but can be formalised within it, as shown later.

[3]Even more appropriate, but too long, would be 'measurement scheme'.

A more convenient notation. If you are wondering what those $M^{(j)}$ are doing in the context of the plausibility, consider that the plausibility of observing a particular outcome at the j instance of a measurement is, in general, the plausibility of the outcome *given that* the measurement is made, times the plausibility that the measurement is made (if the latter is not made the outcome has nought plausibility by definition):

$$\begin{aligned}\mathrm{P}(R|\,I) &= \mathrm{P}(R|\,M \wedge I)\ \mathrm{P}(M|\,I) + \mathrm{P}(R|\,\neg M \wedge I)\ \mathrm{P}(\neg M|\,I),\\ &= \mathrm{P}(R|\,M \wedge I)\ \mathrm{P}(M|\,I).\end{aligned}$$

But of course when we ask for the plausibility of an outcome we implicitly mean: *given that* the corresponding measurement is or will be made. We assume that knowledge of an outcome implies knowledge that the corresponding measurement has been made — symbolically, $\mathrm{P}(M^{(j')}|\,R^{(j'')} \wedge I) = 1$ if $j' = j''$ — hence there is no need of specify in the context the measurements of those outcomes that are already in the context. But the necessity remains of explicitly writing in the context the measurements of the outcomes outside the context. This can sometimes be notationally very cumbersome, and therefore we introduce the symbol $\mathfrak{M}$ with the following convention: $\mathfrak{M}$, in the context of a plausibility, always stands for the conjunction of all measurement instances $M^{(j)}$ corresponding to the outcomes on the left of the conditional symbol '|'. Thus, e.g.,

$$\mathrm{P}(R_5^{(7)} \wedge R_1^{(3)}|\,\mathfrak{M} \wedge R_8^{(2)} \wedge I) \equiv \mathrm{P}(R_5^{(7)} \wedge R_1^{(3)}|\,M^{(7)} \wedge M^{(3)} \wedge R_8^{(2)} \wedge I).$$

Note that $\mathfrak{M}$ has not a constant value (it is in a sense a metavariable); we indicate this by the use of a different typeface (Euler Fraktur).

With the new notational convention the plausibility (1) can be rewritten as

$$\mathrm{P}(R_{i_{N+L}}^{(j_{N+L})} \wedge \cdots \wedge R_{i_{N+1}}^{(j_{N+1})}|\,\mathfrak{M} \wedge R_{i_N}^{(j_N)} \wedge \cdots \wedge R_{i_1}^{(j_1)} \wedge I). \tag{1$_\mathrm{r}$}$$

3 The approach through infinite exchangeability

The propensitor (and, roughly in the same way, the frequentist as well) approaches the question of assigning a value to (1) by supposing that the outcome instances ('trials') are 'i.i.d.': that they are independently 'produced' with constant but 'unknown' propensities. As N in (1) becomes large, the relative frequencies of the outcomes *must* tend to the numerical values of the propensities. These frequencies can then be used as 'estimates' of the propensities, and so the estimated propensity for outcome R_i at an additional measurement instance can be given. This summary surely appears laconic to readers that have superficial knowledge of this practise. But this does not matter: it is the approach of the de Finettian that interests us.[4] Recall that for the

[4]There are authors who apparently keep a foot in both camps; see e.g. Lindley and Phillips' article [51]

de Finettian, and for the Bayesian in general, ‘probability’ is not a physical concept like ‘pressure’, but a logical (and subjective) one like ‘truth’. To mark this difference in concept we are using the term *plausibility*, which has a more logical and subjective sound.

How would a de Finettian approach the problem above? More or less as follows:

[A fictive de Finettian speaking:] ‘Before I know of any measurement outcomes, I imagine all possible *infinite* collections of instances of the measurement M. Let us suppose that I judge two collections that have the same frequencies of outcomes to have also the same probability. The probability distribution that I assign to the infinite collections of outcomes is therefore symmetric with respect to exchanges of collections having the same frequencies. Such a distribution is called *infinitely exchangeable*. De Finetti’s representation theorem [4–11, 52] (see also [13, 53, 54]) says that any L-outcome marginal of such distribution, where the outcomes $\{R_i\}$ appear with relative frequencies $\bar{L} \equiv (L_i)$, can be *uniquely* written in the following form:

$$\mathrm{P}(\underbrace{R_{i_L}^{(j_L)} \wedge \cdots \wedge R_{i_1}^{(j_1)}}_{R_1 \text{ appears } L_1 \text{ times, etc.}} | \mathfrak{M} \wedge I) = \int \Bigl(\prod_i q_i^{L_i}\Bigr)\, \mathit{\Gamma}(\bar{q}|\, I)\, \mathrm{d}\bar{q}, \tag{2}$$

where $\bar{q} \equiv (q_i)$ are just *parameters* — not probabilities! — satisfying the same positivity and normalisation conditions ($q_i \geqslant 0$, $\sum_i q_i = 1$) as a probability distribution, and $\bar{q} \mapsto \mathit{\Gamma}(\bar{q}|\, I)$ is a positive and normalised generalised function [55–57] (see also [58]), which can be called the *generating function* of the representation [cf. 52]. Example: I can write the probability of a collection of three measurement instances with two outcomes R_5 and one R_8 as

$$\mathrm{P}(R_5^{(j_1)} \wedge R_5^{(j_2)} \wedge R_8^{(j_3)} | \mathfrak{M} \wedge I) = \int q_5{}^2\, q_8\, \mathit{\Gamma}(\bar{q}|\, I)\, \mathrm{d}\bar{q}.$$

In particular, the probability for the outcome i in the whatever measurement instance j can be written as

$$\mathrm{P}(R_i^{(j)} | M^{(j)} \wedge I) = \int q_i\, \mathit{\Gamma}(\bar{q}|\, I)\, \mathrm{d}\bar{q}. \tag{3}$$

A propensitor or a frequentist would say that the right-hand side of eq. (2) represents the fact that the outcome instances are “independent” (therefore “their probabilities q_i are simply multiplied to give the probability of their conjunction”), “identically distributed” (therefore “the probability distributions q is the same for all instances”), and moreover “their probability is unknown” (therefore “the expectation integral over all possible probability distributions $\bar{q}$”). But de Finetti’s theorem shows that all these mathematical features are simply consequences of infinite exchangeability, and we do not need any “i.i.d.” terminology.

‘The generating function $\bar{q} \mapsto \mathit{\Gamma}(\bar{q}|\, I)$ is, by de Finetti’s theorem, equal to the limit

$$\mathit{\Gamma}(\bar{q}|\, I)\, \mathrm{d}\bar{q} = \lim_{L\to\infty} \mathrm{P}\Bigl(\begin{matrix}\text{‘All possible collections of } L \text{ outcomes with} \\ \text{frequencies in the range }]Lq_i, L(q_i + \mathrm{d}q_i)[\text{’}\end{matrix} \Bigm| \mathfrak{M} \wedge I\Bigr), \tag{4}$$

i.e., $\Gamma(\bar{q}|I)$ is equal to the probability — assigned by *me* — that my imagined infinite collection of outcomes has limiting relative frequencies equal to (q_i). My exchangeable probability assignment determines thus Γ uniquely. But de Finetti's theorem also says the converse, viz., any positive and normalisable Γ uniquely determines an infinitely exchangeable probability distribution. This allows me to specify my probability assignment by giving the function Γ instead of the more cumbersome probability distribution for infinite collections of outcomes.

'Let us suppose that I am now given a collection of N measurement outcomes, and in particular their absolute frequencies $\bar{N} \equiv (N_i)$. Given this evidence, the probability for a collection of further L unobserved measurement outcomes with frequencies (L_i) is provided by the basic rules of probability theory:

$$\mathrm{P}(\underbrace{R^{(j_{N+L})}_{i_{N+L}} \wedge \cdots \wedge R^{(j_{N+1})}_{i_{N+1}}}_{R_i \text{ appears } L_i \text{ times}} | \mathfrak{M} \wedge \underbrace{R^{(j_N)}_{i_N} \wedge \cdots \wedge R^{(j_1)}_{i_1}}_{R_i \text{ appears } N_i \text{ times}} \wedge I) = \frac{\mathrm{P}(R^{(j_{N+L})}_{i_{N+L}} \wedge \cdots \wedge R^{(j_{N+1})}_{i_{N+1}} \wedge R^{(j_N)}_{i_N} \wedge \cdots \wedge R^{(j_1)}_{i_1} | \mathfrak{M} \wedge I)}{\mathrm{P}(R^{(j_N)}_{i_N} \wedge \cdots \wedge R^{(j_1)}_{i_1} | \mathfrak{M} \wedge I)}. \tag{5}$$

Using de Finetti's representation theorem again, this probability can also be written as

$$\mathrm{P}(R^{(j_{N+L})}_{i_{N+L}} \wedge \cdots \wedge R^{(j_{N+1})}_{i_{N+1}} | \mathfrak{M} \wedge R^{(j_N)}_{i_N} \wedge \cdots \wedge R^{(j_1)}_{i_1} \wedge I) = \int \Big(\prod_i q_i^{L_i}\Big) \Gamma(\bar{q}| R^{(j_N)}_{i_N} \wedge \cdots \wedge R^{(j_1)}_{i_1} \wedge I) \, \mathrm{d}\bar{q}. \tag{6}$$

The function $\bar{q} \mapsto \Gamma(\bar{q}| R^{(j_N)}_{i_N} \wedge \cdots \wedge R^{(j_1)}_{i_1} \wedge I)$ is different from the previous one $\bar{q} \mapsto \Gamma(\bar{q}|I)$, but the two can be shown [e.g., 11] to be related by the remarkable formula

$$\Gamma(\bar{q}| R^{(j_N)}_{i_N} \wedge \cdots \wedge R^{(j_1)}_{i_1} \wedge I) = \frac{\Big(\prod_j q_j^{N_j}\Big) \Gamma(\bar{q}|I)}{\int \Big(\prod_j q_j^{N_j}\Big) \Gamma(\bar{q}|I) \, \mathrm{d}\bar{q}}, \tag{7}$$

which is *formally* identical with Bayes' theorem if we define

$$\Gamma(R^{(j_1)}_{i_1} \wedge R^{(j_2)}_{i_2} \wedge \cdots \wedge R^{(j_L)}_{i_L} | \bar{q}, I) \coloneqq q_{i_1} q_{i_2} \cdots q_{i_L} \quad \text{for all } L,\ i_a, \text{ and distinct } j_a, \tag{8}$$

which includes in particular

$$\Gamma(R^{(j)}_i | \bar{q}, I) \coloneqq q_i \quad \text{for all } j. \tag{9}$$

(Note that these are only *formal* definitions and not probability judgements, since the various Γs are not probability distributions.) Thus I can not only specify my infinitely exchangeable distribution by Γ, but also update it by 'updating' Γ. Moreover, for N enough large this function has the limit

$$\Gamma(\bar{q}|\underbrace{R^{(j_N)}_{i_N} \wedge \cdots \wedge R^{(j_1)}_{i_1}}_{R_i \text{ appears } N_i \text{ times}} \wedge I) \simeq \delta\Big(\tfrac{\bar{N}}{N} - \bar{q}\Big) \quad \text{as } N \to \infty. \tag{10}$$

Comparing with eq. (3), this means that

$$\mathrm{P}(R^{(j_{N+1})}_{i}|\, M^{(j_{N+1})} \wedge \underbrace{R^{(j_N)}_{i_N} \wedge \cdots \wedge R^{(j_1)}_{i_1}}_{R_i \text{ appears } N_i \text{ times}} \wedge I) \simeq \frac{N_i}{N} \quad \text{as } N \to \infty \quad (\text{with } j \neq j_a), \tag{11}$$

i.e., the probability I assign to an unobserved outcome gets very near to its observed relative frequency, as observations accumulate. This also means that two persons having different but compatible initial beliefs (i.e., different initial exchangeable distributions having the same support) and sharing the same data, tend to converge to similar probability assignments.'

The point of view summarised above by the de Finettian is quite powerful. It allows the de Finettian to make sense, without the need of bringing along ugly or meaningless metaphysical concepts, of those mathematical expressions very often used by propensitors (and frequentists as well) that are formally identical to formulae (3) ('expected propensity' or 'estimation of unknown probability'), (4) ('probability as limit frequency'), (8) (definition of 'propensity'), (11) ('probability equal to past frequency'). This point of view and the notion of exchangeability can moreover be generalised to more complex situations [5, 8–10, 52, 59–61][cf. also 62], leading to other powerful mathematical expressions and techniques.

4 Why a complementary approach?

We shall presently present and discuss another approach, not based on exchangeability or representation theorems, that can be used to give an answer to the question of induction, and to make sense of the propensitors' and frequentists' formulae and practise as well. Why did we seek an approach different from that based on infinite exchangeability? Here are some reasons:

(a) One sometimes feels 'uncertain', so to speak, about one's plausibility assignment. One can also say that some plausibility assignments feel sometimes more 'stable' than others. This is lively exemplified by Jaynes [21, ch. 18]:

> Suppose you have a penny and you are allowed to examine it carefully, convince yourself that it's an honest coin; i.e. accurately round, with head and tail, and a center of gravity where it ought to be. Then, you're asked to assign a probability that this coin will come up heads on the first toss. I'm sure you'll say 1/2. Now, suppose you are asked to assign a probability to the proposition that there was once life on Mars. Well, I don't know what

> your opinion is there, but on the basis of all the things that I have read on the subject, I would again say about 1/2 for the probability. But, even though I have assigned the same "external" probabilities to them, I have a very different "internal" state of knowledge about those propositions.
>
> To see this, imagine the effect of getting new information. Suppose we tossed the coin five times and it comes up tails every time. You ask me what's my probability for heads on the next throw; I'll still say 1/2. But if you tell me one more fact about Mars, I'm ready to change my probability assignment completely. There is something which makes my state of belief very stable in the case of the penny, but very unstable in the case of Mars.

An example similar to the Martian one is that of a trickster's coin, which always comes up heads or always tails (either because the coin is two-headed or two-tailed, or because of the tosser's skills). Not knowing which way the tosses are biased, you assign 1/2 that the coin will come up heads on the first observed toss. But as soon as you see the outcome, your plausibility assignment for the next toss will collapse[5] to 1 or 0. The 'uncertainty' in the plausibility assignments given to the penny and to the trickster coin's toss could be *qualitatively* pictured as in Fig. 1. We think that this 'uncertainty' in the plausibility assignment — or better, as Jaynes calls it, this 'difference in the internal state of knowledge' with regard to the propositions involved, is a familiar and undeniable feeling. There is in fact a rich literature which tries to take it into account by exploring or even proposing alternative plausibility theories based on probability intervals or probabilities of probabilities (see e.g. [63, 64][62, esp. § 3.1] and cf. [65, § 2.2][66, 67]).

(b) A plausibility assignment, according to de Finetti's approach, begins with a judgement of exchangeability. One is not concerned *within the formalism itself* with the motivation of such a judgement (de Finetti recognises that there usually is a motivation, but it is relegated to the informal meta-theoretical considerations). Yet such judgements are often grounded, especially in natural philosophy, on very important[6] reasonings and motivations which one would like to analyse by means of plausibility theory.[7]

(c) When the maximum number of possible observations (measurements) is finite *and small* it does not make sense to make a judgement of infinite exchangeability, not even as an approximation (cf. [9][11, § 4.7.1]). De Finettians can in this case use *finite* exchangeability [9, 10], but then they cannot make sense of the propensitor's formulae like (3) or (8), which the propensitor, however, is still entitled to use even in this case.

[5] Just like a quantum-mechanical wave-function.

[6] And, trivially, subjective.

[7] In a way there is a point in Good's statement [65, ch. 3] 'it seems to me that one would not accept the [exchangeability assumption] unless one already had the notion of physical probability and (approximate) statistical independence at the back of one's mind', although there is no need to bring 'physical probabilities' and 'statistical independence' along.

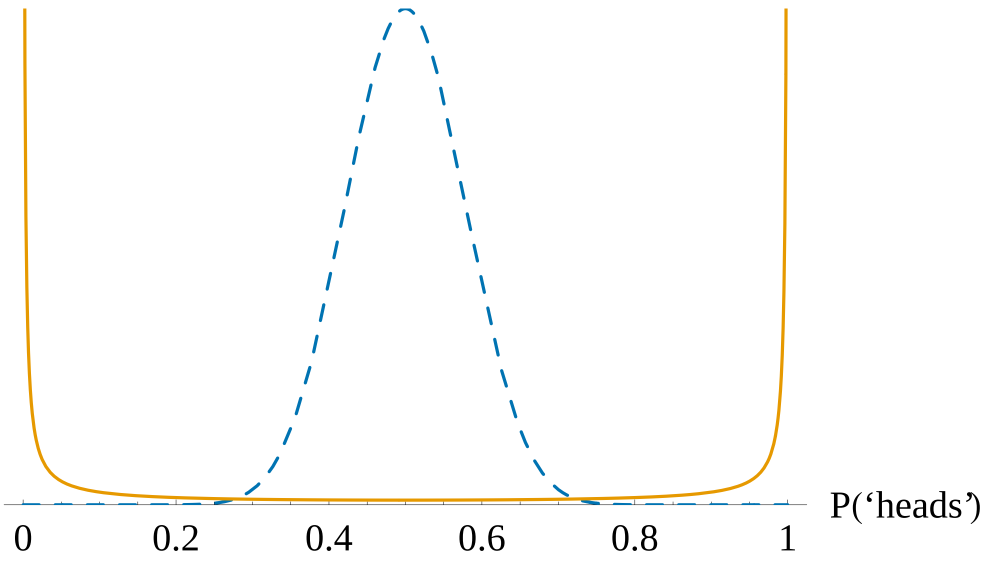

Figure 1: Qualitative illustration of the 'uncertainty' in the plausibility assignments for 'heads' in the case of a fair coin-toss (blue dashed curve, with maximum at $1/2$) and the toss of a trickster's coin (orange continuous curve, with maxima at 0 and 1). The illustration is made quantitative in § 6.2.

(d) Finally, another important reason is that the generalisation of de Finetti's theorem studied by Caves, Fuchs, and Schack [68–71] fails for some physical statistical models, viz. quantum mechanics on real and quaternionic Hilbert space. We consider this failure not as a sign that complex quantum theory is somehow 'blessed', but rather as a sign that either the theorem can be generalised in some other way that holds in any physical statistical model, or de Finetti's approach can be substituted or complemented by another one whose generalisation applies to any physical statistical model whatever. The point here is that de Finetti's theorem — just like the whole of plausibility theory — belongs to the realm of *logic*, not physics, and thus should apply to any conceivable physical theory that be logically consistent, even one that does not describe actual phenomena.

The alternative point of view to be presented meets all the points raised above: (a) It allows us to formalise within plausibility theory — i.e., without resorting to interval-valued plausibilities and the like — the intuitive notion of 'uncertainty' of a plausibility assignment, and even to quantify it. (b) It stems directly from an analysis of the conditions in which judgements of exchangeability usually originate and apply. (c) It can be used to interpret the propensitor's practise when the maximum number of possible observations is finite and small. (d) It readily generalises to any logically consistent physical statistical model, whether it apply to actual phenomena or not; in particular, within quantum theory it easily applies to such problems as quantum-state assignment, quantum-state 'teleportation', and others that involve the notion of 'unknown quantum state'.

This point of view, which we shall call the 'Laplace-Jaynes approach', stems from a re-reading and possibly a re-interpretation of Jaynes [1, esp. pp. 11, 12, 15] (cf. also [19, lect. 18][20, lect. 5][21, ch. 18]), Laplace [72], and de Finetti [3, § 20] (cf. also Caves [73]) and is strictly related to the 'approach through circumstances' presented in the previous paper [2].

It cannot be too strongly emphasised that this point of view is fully in line with plausibility theory, Bayesian theory, and de Finetti's point of view: Some plausibilities with particular properties will be introduced, based on particular judgements; but the acceptance or not of the latter is, as with judgements of exchangeability, always up to the individuals and their knowledge.

We now present this 'Laplace-Jaynes approach' and show how it can be used for the question of induction in a manner parallel to the approach through infinite exchangeability. We shall then discuss how it meets points a, b, c above (not necessarily in that order). The discussion of the fourth point is left to the next study of this series [74].

5 The Laplace-Jaynes approach

5.1 Introducing a set of circumstances

With the notation and the general settings of the introduction, let us recapitulate how we reason and proceed in the case of exchangeability. The situation I leads us to assign an infinitely exchangeable plausibility distribution to the collection of measurement outcomes. In other words, I is such that we consider two plausibilities like e.g. $\mathrm{P}(R_5^{(2)} \land R_7^{(4)} | \mathfrak{M} \land I)$ and $\mathrm{P}(R_5^{(4)} \land R_7^{(2)} | \mathfrak{M} \land I)$ as equal.

In § 4, point (b), we remarked that the approach through exchangeability is not formally concerned about those details of the situation which lead us to see a sort of analogy ar similarity amongst different measurement and outcome instances and thence make a judgement of exchangeability. Rather, the whole point is that this similarity is expressed by, or reflected in, the exchangeable plausibility assignment. As de Finetti says, 'Our reasoning will only bring in the events, that is to say, the trials, each taken individually; the analogy of the events does not enter into the chain of reasoning in its own right but only to the degree and in the sense that it can influence in some way the judgment of an individual on the probabilities in question' [4, p. 120]. And also:

> What are sometimes called *repetitions* or *trials*[8] of the same event are for us distinct events. They have, in general, some common characteristics or symmetries which make it natural to attribute to them equal probabilities, but we do not admit any *a priori* reason which prevents us in

[8]And what we call here *instances* (Authors' Note).

> principle from attributing to each of these trials [. . .] some different and absolutely arbitrary probabilities [. . .]. In principle there is no difference for us between this case and the case of *n* events which are not analogous to each other; the analogy which suggests the name "trials of the same event" (we would say "of the same phenomenon") is not at all essential, but, at the most, valuable because of the influence it can exert on our psychological judgment in the sense of making us attribute equal or very nearly equal probabilities to the different events.' [4, p. 113, footnote 1]

But although we agree on this terminology and its motivation, we do not agree on the unqualified and indiscriminate diminution of the importance of the similarity amongst measurement instances. Such similarity, especially in the natural sciences, often stems from or is traced back to similarities at deeper[9] levels of analysis and observation. It is by this process that natural philosophy proceeds.

So let us suppose that the considerations that lead us to a probability assignment for various measurement instances can be analysed — and formalised — at a deeper level.[10] More precisely, we suppose to have identified for each measurement instance j a set of diverse possible 'circumstances' $\{C_k^{(j)} : k = 1, \dots\}$ — all these sets having the same cardinality — with the following properties:

I. The first involves the plausibilities we assign to the circumstances:

$$\begin{aligned} &\mathrm{P}(C_k^{(j)} | D \wedge I) = \mathrm{P}(C_k^{(j)} | I) \quad \text{for all } k \text{ and all } D \text{ representing conjunctions of measurements (but \emph{not} of outcomes!) from the same or other instances,} \\ &\mathrm{P}(C_{k'}^{(j)} \wedge C_{k''}^{(j)} | I) = 0 \quad \text{if } k' \neq k'', \\ &\mathrm{P}(\bigvee_k C_k^{(j)} | I) = \sum_k \mathrm{P}(C_k^{(j)} | I) = 1, \end{aligned} \tag{I}$$

i.e., we judge the $\{C_k^{(j)}\}$ to be mutually exclusive and exhaustive, or in other words we are certain that one of them holds, but we do not know which. Moreover, we judge any knowledge of measurement instances (but not of outcome instances) as irrelevant for assessing the plausibilities of these circumstances, or in other words, simply knowing that a measurement is made does not give us clues as to the exact circumstance in which it is made.[11]

---

[9] We do not mean 'microscopic'.

[10] Using, once more, words by de Finetti: 'we enlarge the analysis to include our state of mind in relation to other events, from which it might or might not be *independent* and by which, consequently, it will or will not be modified if they occur, or if we learn of their occurrence' [75, § 28].

[11] This assumption can be relaxed.

II. The second involves the plausibility distribution we assign to the outcomes, conditional on knowledge of a circumstance:

$$\mathrm{P}(R_i^{(j)} | M^{(j)} \wedge D \wedge C_k^{(j)} \wedge I) = \mathrm{P}(R_i^{(j)} | M^{(j)} \wedge C_k^{(j)} \wedge I)$$
for all j, i, k, and all D representing a conjunction of measurements, measurement outcomes, and circumstances of instances different from j, (II)

which means that if we were certain about any circumstance $C_k^{(j)}$ for the instance j we should judge knowledge of any data concerning other instances as irrelevant for the assessment of the plausibility of the outcome of instance j. (The expression above is undefined if $C_k^{(j)}$ happens to be inconsistent with D, i.e. if $\mathrm{P}(D | \mathfrak{M} \wedge C_k^{(j)} \wedge I) = 0$; but this will not cause problems in the following analysis.)

III. The third property concerns the relationships amongst our plausibility assignments for the outcomes in different instances:

$$\text{for all } j', j'', \quad \mathrm{P}(R_i^{(j')} | M^{(j')} \wedge C_k^{(j')} \wedge I) = \mathrm{P}(R_i^{(j'')} | M^{(j'')} \wedge C_k^{(j'')} \wedge I) =: q_{ik}, \tag{III}$$

which expresses the fact that we see a similarity between outcomes and circumstances of different instances, like when we say 'the same outcome', 'the same measurement', or 'the same circumstance'. It is thanks to this property that we can often drop the instance index '(j)' and make sense of expressions like '$\mathrm{P}(R_i | M \wedge C_k \wedge I)$', which can stand generically for

$$\begin{aligned} \mathrm{P}(R_i | M \wedge C_k \wedge I) &:= \mathrm{P}(R_i^{(j)} | M^{(j)} \wedge C_k^{(j)} \wedge I) \quad \text{for any } j, \\ &\equiv q_{ik}. \end{aligned} \tag{12}$$

In the following, expressions like '$\mathrm{P}(R_i | M \wedge C_k \wedge I)$' will be understood in the above sense.

IV. The fourth property strengthen the similarity amongst instances, and is the one that makes induction possible:

$$\mathrm{P}(C_{k'}^{(j')} | C_{k''}^{(j'')} \wedge I) = \delta_{k' k''} \quad \text{for all } j', j''. \tag{IV}$$

This means that we believe that if a particular circumstance holds in a particular instance, then the 'same' circumstance (in the sense of (III) and (IV)) holds in all other instances. This property implies (and, with the help of (I), is implied by) the following:

$$\text{for all } j \text{ and } k, \quad \mathrm{P}(C_k^{(j)} | I) = \mathrm{P}(\bigwedge_l C_k^{(l)} | I) =: \gamma_k \tag{13}$$

for some γ_k; and

$$\mathrm{P}(\bigwedge_l C_{k_l}^{(l)} | I) = \begin{cases} \gamma_k & \text{if } k_l = k \text{ for all } l, \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

i.e., the plausibility that we assign to the jth instance of the kth circumstance is equal to that assigned to the collection consisting exclusively of 'repetitions' of the kth circumstance. Collections consisting of non-corresponding circumstances have nought plausibility. This allows us to define

$$\begin{aligned} \mathrm{P}(C_k | I) &:= \mathrm{P}(\textstyle\bigwedge_l C_k^{(l)} | I), \\ &\equiv \mathrm{P}(C_k^{(j)} | I) \quad \text{for any } j, \\ &\equiv \gamma_k, \end{aligned} \tag{15}$$

and

$$C_k := \bigwedge_l C_k^{(l)}. \tag{16}$$

Thanks to the definitions above the expression '$\mathrm{P}(C_k | I)$' can be used to unambiguously denote the plausibility that 'the "same" circumstance 'C_k' holds in all measurement instances'.

Note that in all the properties above the number of instances (i.e., the range of j) can be either infinite or finite.

Before further commenting the above properties, let us try to partially answer the question: what are these 'circumstances'? The most general and precise answer is: they are whatever (propositions concerning) facts you like that make you assign plausibilities satisfying properties (I)–(IV). No more than this would really need be said. But we can emphatically add that the circumstances need not concern 'mechanisms', 'causes', 'microscopic conditions', or the like; and that they do not[12] concern 'unknown probabilities'.[13] They concern details of the context that are unknown, but that (we judge) would have a great weight in our plausibility assignment for the $\{R_i^{(j)}\}$ if we only knew them; so great a weight as to render (practically) unimportant any knowledge of the details of other measurement instances. As for the choice of such details, we have complete freedom. A trivial example: the tosses of a coin which we have not examined, but which we know for sure to be two-headed or two-tailed. In this case the propositions 'Coin is two-headed (at toss j)' and 'Coin is two-tailed (at toss j)' form, for each j, a set of circumstances that satisfy property (I). But we also know that the coin is *the same* at all tosses, which implies property (IV); hence we can omit the specification '(at toss j)' without confusion. If we knew that the coin was two-headed (at toss j) we should assign unit plausibility to heads for all future or past tosses, i.e.,

$$\mathrm{P}(\text{'Heads at toss } j\text{'} | \mathfrak{M} \land \text{'Coin is two-headed (at toss } j\text{)'} \land I) = 1 \quad \text{for all } j,$$

[12]Cannot, in a logical sense; cf. Remark 2 in [2].

[13]But the circumstances *may* concern 'propensities', if the latter are intended as sorts of (ugly) physical concepts. This possibility is due to the generality of plausibility theory which, like classical logic, does not forbid you to bring along and reason about unreal or even preposterous concepts and entities, like 'nagas' [76], 'valier' [77], 'propensities', and 'wave-particles', provided you do it in a self-consistent way.

and we should consider knowledge of the outcomes of other tosses as irrelevant.[14] An analogous discussion holds for the two-tails possibility. Thus properties (II) and (III) are also satisfied.

This was an extreme example, for the plausibilities conditional on the circumstances were nought or one. But it needs not be so: other circumstances could lead to less extreme conditional plausibility judgements. In general, remaining within the coin toss example, we could ask: By which method is the coin tossed? What are its physical characteristics (two-headedness, centre-of-mass position, elastic and rigid properties, etc.)? Upon what is it tossed? Who tosses it?[15] Can there be any symmetries in my state of knowledge in respect of the situation? What are the consequences of the toss or of the outcomes? — and a set of circumstances could be distilled from the possible answers to these and other questions. Note in particular, with regard to the last two questions, that a circumstance may be a judgement of symmetry or may also be a *consequence*, in some sense, of the outcomes. Such kinds of circumstances are perfectly fine as long as they lead you to make a plausibility assignment on the outcomes and satisfy the properties (I)–(IV). This emphasises again the fact that a circumstance needs not be a sort of 'mechanism' or 'cause' of the measurements or of the outcomes. Note also that properties (I) and (IV) determine neither the plausibilities $\mathrm{P}(C_k^{(j)} | I)$ nor the $\mathrm{P}(R_i^{(j)} | M^{(j)} \wedge C_k^{(j)} \wedge I)$. These plausibilities are, a de Finettian would say, 'fully subjective'.

By properties (II) and (III) we can interpret and give a meaning to the locution 'independent and identically distributed events'. Such locution means only that we are entertaining some circumstances that, according to our judgement, render the conditional plausibilities of corresponding outcomes of different measurement instances equal, so that we need not specify the particular instance ('identically distributed'); and render knowledge of outcomes of other instances irrelevant, so that we can specify the conditional plausibility for an outcome independently of the knowledge of other outcomes ('independent'). The locutions 'unknown probability' and 'probability of a probability' will also be interpreted in a moment (§ 5.3).

5.2 Approaching the problem of induction through a set of circumstances

How do we face the question of induction with these 'circumstances' and the assumptions that accompany them? We answer this question in two steps.

Having introduced sets of circumstances $\{C_k^{(j)}\}$, we must assign (subjectively, a

[14]Because we can but expect all tosses to give heads. Note that if a toss has given or will give tails, this signals a *contradiction* in our knowledge. I.e., some of the data we have (about the coin or about toss outcomes) have to be mendacious. But this is a problem that does not concern plausibility theory. Like logic, it can give sensible answers only if the premises we put in are not inconsistent.

[15]See Jaynes' insightful and entertaining discussion [21, ch. 10] on these kinds of factors.

de Finettian would say) for all i and k the plausibilities

$$\mathrm{P}(R_i | M \wedge C_k \wedge I) \equiv q_{ik} \qquad (:= \mathrm{P}(R_i^{(j)} | M^{(j)} \wedge C_k^{(j)} \wedge I) \text{ for any } j), \tag{17}$$

$$\mathrm{P}(C_k | I) \equiv \gamma_k \qquad (:= \mathrm{P}(\textstyle\bigwedge_l C_k^{(l)} | I) \equiv \mathrm{P}(C_k^{(j)} | I) \text{ for any } j), \tag{18}$$

in the sense of eqs. (12) and (15). It is then a simple consequence of the rules of plausibility theory, together with the properties and definitions of the previous sections, that the plausibility we assign to any collection of measurement outcomes is given by

$$\mathrm{P}(\underbrace{R_{i_L}^{(j_L)} \wedge \cdots \wedge R_{i_1}^{(j_1)}}_{R_1 \text{ appears } L_1 \text{ times, etc.}} | \mathfrak{M} \wedge I) = \sum_k \Big[\prod_i \mathrm{P}(R_i | M \wedge C_k \wedge I)^{L_i}\Big] \mathrm{P}(C_k | I) \equiv \sum_k \Big(\prod_i q_{ik}^{L_i}\Big) \mathrm{P}(C_k | I). \tag{19}$$

Note how this plausibility assignment depends only on the frequencies (L_i) of the outcomes: it is an (infinitely) exchangeable assignment — although exchangeability was not our starting assumption.

Let us suppose that we are now given a collection of N measurement outcomes, and in particular their absolute frequencies $\bar{N} \equiv (N_i)$. Given this evidence, the plausibility for a collection of outcomes, with frequencies (L_i), of further L measurements is also derived by the basic rules of plausibility theory, from our initial plausibility assignments (17) and (18), using expression (19) and the properties of the circumstances:

$$\mathrm{P}(\underbrace{R_{i_{N+L}}^{(j_{N+L})} \wedge \cdots \wedge R_{i_{N+1}}^{(j_{N+1})}}_{R_i \text{ appears } L_i \text{ times}} | \mathfrak{M} \wedge \underbrace{R_{i_N}^{(j_N)} \wedge \cdots \wedge R_{i_1}^{(j_1)}}_{R_i \text{ appears } N_i \text{ times}} \wedge I) =$$
$$\sum_k \Big[\prod_i \mathrm{P}(R_i | M \wedge C_k \wedge I)^{L_i}\Big] \mathrm{P}(C_k | \wedge R_{i_N}^{(j_N)} \wedge \cdots \wedge R_{i_1}^{(j_1)} \wedge I) \equiv$$
$$\sum_k \Big(\prod_i q_{ik}^{L_i}\Big) \mathrm{P}(C_k | R_{i_N}^{(j_N)} \wedge \cdots \wedge R_{i_1}^{(j_1)} \wedge I), \tag{20}$$

with

$$\mathrm{P}(C_k | R_{i_N}^{(j_N)} \wedge \cdots \wedge R_{i_1}^{(j_1)} \wedge I) = \frac{\Big(\prod_i q_{ik}^{N_i}\Big) \mathrm{P}(C_k | I)}{\sum_k \Big(\prod_i q_{ik}^{N_i}\Big) \mathrm{P}(C_k | I)}. \tag{21}$$

These formulae present many similarities to the de Finettian's (6) and (7), and to the formally similar expressions used by 'frequentists' and 'propensitors'. But it should be noted that, in the present formulae, $q_{ki} := \mathrm{P}(R_i | \mathfrak{M} \wedge C_k \wedge I)$, $\mathrm{P}(C_k | I)$, and $\mathrm{P}(C_k | R_{i_N}^{(j_N)} \wedge \cdots \wedge R_{i_1}^{(j_1)} \wedge I)$ are *actual plausibilities*, not just parameters or positive and normalised generating functions. ('And they are fully subjective!', the de Finettian reiterates).

5.3 Coarse-graining the set of circumstances

The marked similarities can in fact be made into identities of form. To achieve this, we go back to the point where, after having introduced the circumstances $\{C_k^{(j)}\}$, we made the plausibility assignments

$$\mathrm{P}(R_i | M \land C_k \land I) \equiv q_{ik} \qquad (:= \mathrm{P}(R_i^{(j)} | M^{(j)} \land C_k^{(j)} \land I) \text{ for any } j), \tag{17$_\mathrm{r}$}$$

$$\mathrm{P}(C_k | I) \equiv \gamma_k \qquad (:= \mathrm{P}(\textstyle\bigwedge_l C_k^{(l)} | I) \equiv \mathrm{P}(C_k^{(j)} | I) \text{ for any } j), \tag{18$_\mathrm{r}$}$$

for all i, k (and j). Instead of proceeding as we did, we now follow the remark by Jaynes [1, p. 12], cited in the introductory epigraph:

> It is simply convenient to index our hypotheses by parameters [$\bar{q}$] chosen to be numerically equal to the probabilities assigned by those hypotheses; this avoids a doubling of our notation. We could easily restate everything so that the misconception could not arise; it would only be rather clumsy notationally and tedious verbally.

What Jaynes calls 'hypotheses' we have here called, more generically, 'circumstances'. Let us now group together those that lead to the same plausibility distribution for the outcomes. That is, fix a j, and form the equivalence classes of the equivalence relation

$$C_{k'}^{(j)} \sim C_{k''}^{(j)} \iff \text{for all } i,\ \mathrm{P}(R_i^{(j)} | M^{(j)} \land C_{k'}^{(j)} \land I) = \mathrm{P}(R_i^{(j)} | M^{(j)} \land C_{k''}^{(j)} \land I). \tag{22}$$

By the very method these classes are defined, each one can be *uniquely* identified by a particular set of values $(q_i) \equiv \bar{q}$ of the plausibility distribution for the outcomes. Note that owing to property (III) the value of $\bar{q}$ does not depend on j. Call therefore $\bar{q}^\sim$ the class identified by a particular $\bar{q}$, and denote membership of $C_k^{(j)}$ by '$k \in \bar{q}^\sim$' for short. Now let us take the disjunction of the circumstances in each class $\bar{q}^\sim$ and uniquely denote this disjunction by $S_{\bar{q}}^{(j)}$:

$$S_{\bar{q}}^{(j)} := \bigvee_{k \in \bar{q}^\sim} C_k^{(j)}. \tag{23}$$

It is easy to see that, owing again to property (I), each $S_{\bar{q}}^{(j)}$ yields a conditional distribution for the outcomes that is numerically equal to $\bar{q}$, i.e. numerically equal to all those yielded by the $C_k^{(j)}$ in $\bar{q}^\sim$:

$$\mathrm{P}(R_i^{(j)} | M^{(j)} \land S_{\bar{q}}^{(j)} \land I) = q_i \quad \text{for any } j. \tag{24}$$

Moreover, the $\{S_{\bar{q}}^{(j)}\}$ also satisfy properties (I)–(IV) as the circumstances $\{C_k^{(j)}\}$, and their plausibilities are easily obtained from those of the latter:

$$\mathrm{P}(S_{\bar{q}}^{(j)} | I) = \sum_{k \in \bar{q}^\sim} \mathrm{P}(C_k^{(j)} | I), \qquad \mathrm{P}(S_{\bar{q}} | I) = \sum_{k \in \bar{q}^\sim} \mathrm{P}(C_k | I), \tag{25}$$

where $S_{\bar{q}} \coloneqq \bigwedge_l S^{(l)}_{\bar{q}}$ in analogy with the definition (16).

These $S^{(j)}_{\bar{q}}$ are thus a sort of 'coarse-grained' circumstances. We shall call them thus, or also 'plausibility-indexed circumstances' sometimes. The $C^{(j)}_k$ can then be called 'fine-grained' circumstances when a distinction is necessary.

In general, we have a class $\bar{q}^{\sim}$ and a disjunction $S^{(j)}_{\bar{q}}$ for particular (vector) values of $\bar{q}$ only (depending on the initial choice of circumstances and on the initial assignments (17)). But we can formally introduce propositions $S^{(j)}_{\bar{q}'}$, all equal to the false proposition $(A \wedge \neg A)$, for the remaining values $\bar{q}'$. These propositions hence have nought plausibilities, $\mathrm{P}(S^{(j)}_{\bar{q}'} | I) \equiv 0$. The plausibilities $\mathrm{P}(R^{(j)}_i | M^{(j)} \wedge S^{(j)}_{\bar{q}'} \wedge I)$ are undefined, but they will appear multiplied by the former in all relevant formulae, and hence their product will vanish by convention. With this expedient we can consider the whole, continuous set $\{S^{(j)}_{\bar{q}} \mid \bar{q} \in \Delta\}$ for all possible distributions $\bar{q}$ which naturally belong to the simplex $\Delta \coloneqq \{(q_i) \mid q_i \geqslant 0, \sum_i q_i = 1\}$. We can thus substitute an integration $\int_\Delta \dots p(S_{\bar{q}}| \dots)\,\mathrm{d}\bar{q}$ for the summation $\sum_{\bar{q}} \dots \mathrm{P}(S_{\bar{q}}|\dots)$. The density $p(S_{\bar{q}}|\dots)$ will be a generalised function.[16]

We can therefore analyse the context I into the 'coarse-grained' circumstances $\{S^{(j)}_{\bar{q}}\}$, instead of the $\{C^{(j)}_k\}$. The rationale behind this is that those circumstances $C^{(j)}_k$ that belong to a given class $\bar{q}^{\sim}$ have all exactly the same effect on our judgement as regards the assignment and the update of the plausibilities of the outcomes. It is therefore not unreasonable to handle them class-wise. The special notation chosen for the different classes, '$S^{(j)}_{\bar{q}}$', the index '$\bar{q}$' in particular, is to remind us on the grounds of what plausibility judgements (the $\bar{q}$) we grouped the original circumstances thus in the first place. But it is only a notation, nothing more. Each $S^{(j)}_{\bar{q}}$ is a disjunction of propositions like, e.g., 'The coin is two-headed, or the tosses are made by a trickster with a predilection for heads, or …' — it is *not* a statement about the plausibilities (or the 'propensities') q_i.

The effect of this 'bookkeeping' notation is, however, surprising for the *form* our induction formulae (19)–(21) take when expressed in terms of the coarse-grained circumstances $\{S^{(j)}_{\bar{q}}\}$. The plausibility we assign to any collection of measurement outcomes takes now, by eq. (24), the form

$$\mathrm{P}(\underbrace{R^{(L)}_{i_L} \wedge \dots \wedge R^{(1)}_{i_1}}_{R_1 \text{ appears } L_1 \text{ times, etc.}} | \mathfrak{M} \wedge I) = \int \Bigl[\prod_i \mathrm{P}(R_i | M \wedge S_{\bar{q}} \wedge I)^{L_i}\Bigr] p(S_{\bar{q}}|I)\,\mathrm{d}\bar{q} \equiv \int \Bigl(\prod_i q_i^{L_i}\Bigr) p(S_{\bar{q}}|I)\,\mathrm{d}\bar{q}, \tag{26}$$

with value numerically equal to that of (19), and with the same remarks and conventions about the index j as made in § 5.1. The expression above is formally identical to the de Finettian's (2) with the correspondence $p(S_{\bar{q}}|I) \leftrightsquigarrow \Gamma(\bar{q}|I)$. In particular,

[16] We can avoid the expedient above if we like; then the integrals that follow must be understood in a measure-theoretic sense [78–85] (cf. also [86–88]), with '$p(S_{\bar{q}}|\dots)\,\mathrm{d}\bar{q}$' standing for appropriate singular measures.

the plausibility for the outcome of any instance takes the form

$$\mathrm{P}(R_i^{(j)} | M^{(j)} \wedge I) = \int q_i \, p(S_{\bar{q}} | I) \, \mathrm{d}\bar{q}, \tag{27}$$

formally identical to (3).

The plausibility conditional on the observation of N outcomes takes the form

$$\mathrm{P}(\underbrace{R_{i_{N+L}}^{(j_{N+L})} \wedge \cdots \wedge R_{i_{N+1}}^{(j_{N+1})}}_{R_i \text{ appears } L_i \text{ times}} | \mathfrak{M} \wedge \underbrace{R_{i_N}^{(j_N)} \wedge \cdots \wedge R_{i_1}^{(j_1)}}_{R_i \text{ appears } N_i \text{ times}} \wedge I) =$$

$$\int \Bigl[\prod_i \mathrm{P}(R_i | M \wedge S_{\bar{q}} \wedge I)^{L_i}\Bigr] p(S_{\bar{q}} | R_{i_N}^{(j_N)} \wedge \cdots \wedge R_{i_1}^{(j_1)} \wedge I) \, \mathrm{d}\bar{q} \equiv$$

$$\int \Bigl(\prod_i q_i^{L_i}\Bigr) p(S_{\bar{q}} | R_{i_N}^{(j_N)} \wedge \cdots \wedge R_{i_1}^{(j_1)} \wedge I) \, \mathrm{d}\bar{q} \tag{28}$$

with

$$p(S_{\bar{q}} | R_{i_N}^{(j_N)} \wedge \cdots \wedge R_{i_1}^{(j_1)} \wedge I) \, \mathrm{d}\bar{q} = \frac{\Bigl(\prod_i q_i^{N_i}\Bigr) p(S_{\bar{q}} | I) \, \mathrm{d}\bar{q}}{\int \Bigl(\prod_i q_i^{N_i}\Bigr) p(S_{\bar{q}} | I) \, \mathrm{d}\bar{q}}, \tag{29}$$

again formally identical to the de Finettian's (6) and (7).

Apart from the congruence between their mathematical *forms*, the above formulae and those derived from exchangeability have very different *meanings*. Whereas in the exchangeability approach the $\bar{q} \equiv (q_i)$ were just parameters, in the Laplace-Jaynes approach they are (numerical values of) *actual plausibilities*, viz. the $\mathrm{P}(R_i^{(j)} | M^{(j)} \wedge C_k^{(j)} \wedge I)$ or the $\mathrm{P}(R_i^{(j)} | M^{(j)} \wedge S_{\bar{q}}^{(j)} \wedge I)$. Whereas in the exchangeability approach $\bar{q} \mapsto \mathit{\Gamma}(\bar{q} | \dots)$ was only a generalised function, in the Laplace-Jaynes approach $\bar{q} \mapsto p(S_{\bar{q}} | \dots)$ is (the density of) an *actual plausibility distribution* — a distribution, however, not over 'probabilities' or 'propensities', but rather over propositions like 'The coin is two-headed, or the tosses are made by a trickster with a predilection for heads, or ...'.

Equations (28) and (29) (as well as (20) and (21)) show that the observation of measurement results has a double 'updating' effect in the circumstances approach. Not only are the plausibilities of unobserved results updated, but those of the circumstances as well; in fact, the former updating happens through the latter. This is related to what Caves calls 'learning through a parameter' [73], although there are no parameters here, only plausibilities. Also in the exchangeability approach, one could argue, is the generating function $\mathit{\Gamma}(\bar{q} | I)$ updated; but its updating only represents and reflects a change in our uncertainty or degree of belief *about the outcomes*, no more than that. In the Laplace-Jaynes approach, the updating of $p(S_{\bar{q}} | I)$ represents instead a further change in our uncertainty about events or phenomena other than the outcomes. We shall return to this in a moment.

We finally also see how the Laplace-Jaynes approach interprets and makes sense of the notions of 'unknown probability (or propensity)' and 'probability of a probability': what is unknown is not a probability, but which circumstance from a set of empirical ones holds; the second probability is therefore not about a probability, but about an empirical circumstance.

5.4 Limit for large number of observations

Let us suppose that the plausibility distribution $p(S_{\bar{q}}|\, I)\,\mathrm{d}\bar{q}$ does not vanish for any $\bar{q}$. This implies that the set of fine-grained circumstances $\{C_k\}$ is a continuum, but with analytical and topological care this case should not present particular difficulties.

That assumption being made, from eq. (29) we have that as the number N of observations increases, the updated distribution for the plausibility-indexed circumstances asymptotically becomes

$$p(S_{\bar{q}}|\, \mathfrak{M} \land \underbrace{R_{i_N}^{(j_N)} \land \cdots \land R_{i_1}^{(j_1)}}_{R_i \text{ appears } N_i \text{ times}} \land I) \simeq \delta\Big(\tfrac{\bar{N}}{N} - \bar{q}\Big) \quad \text{as } N \to \infty, \tag{30}$$

which is formally identical to the de Finettian's (10). It also follows that

$$\mathrm{P}(R_i^{(j_{N+1})}|\, M^{(j_{N+1})} \land \underbrace{R_{i_N}^{(j_N)} \land \cdots \land R_{i_1}^{(j_1)}}_{R_i \text{ appears } N_i \text{ times}} \land I) \simeq \frac{N_i}{N} \quad \text{as } N \to \infty, \tag{31}$$

exactly as in (11). Also in the Laplace-Jaynes approach, then, the plausibilities assigned to unobserved outcomes get nearer their observed relative frequencies as the number of observations gets larger (under the assumptions specified above). And persons sharing the same data and having compatible initial plausibility assignments tend to converge to similar plausibility assignments, as regards their respective sets of coarse-grained circumstances (as well as unobserved measurement outcomes. Cf. the discussion is §§ 6.2 and 6.4.

What happens if the distribution $p(S_{\bar{q}}|\, I)\,\mathrm{d}\bar{q}$ vanishes at, or in a neighbourhood of, the point $(q_i) = (N_i/N)$, as may be the case when the number of fine-grained circumstances $\{C_k\}$ is finite? It happens that the updated distribution gets concentrated at those $S_{\bar{q}}$ (and those C_k) with $\bar{q}$ (respectively (q_{ki})) nearest (N_i/N) and for which the initial plausibility does not vanish. One should make the adjective 'nearest' topologically more precise. There are also interesting results at variance with the asymptotic expression (11) when (N_i/N) cannot be obtained as a convex combination of those $\bar{q}$ (respectively (q_{ki})) whose related plausibilities do not vanish. All this is left to another study.

6 Discussion

6.1 Relations between the circumstance and exchangeability approaches

The Laplace-Jaynes approach and the infinite-exchangeability one do not exclude each other and may be used simultaneously in many problems. In fact, if in a given problem we can introduce 'circumstances', and the number of measurement instances is potentially infinite, the resulting distributions for collections of outcomes are then infinitely exchangeable and all the results and representations based on exchangeability also apply, beside those based on the circumstance representation. This leads to the following powerful proposition:

Proposition. *In the Laplace-Jaynes approach, if the number of measurement instances is potentially infinite then*

$$\mathrm{P}(\underbrace{R^{(j_L)}_{i_L} \wedge \dots \wedge R^{(j_1)}_{i_1}}_{R_1 \text{ appears } L_1 \text{ times, etc.}} | \mathfrak{M} \wedge D \wedge I) = \int \Bigl(\prod_i q_i^{L_i}\Bigr) \mathit{\Gamma}(\bar{q}| D \wedge I)\, \mathrm{d}\bar{q} = \int \Bigl(\prod_i q_i^{L_i}\Bigr) p(S_{\bar{q}}| D \wedge I)\, \mathrm{d}\bar{q}, \quad (32)$$

and in particular

$$\mathit{\Gamma}(\bar{q}| D \wedge I)\, \mathrm{d}\bar{q} = p(S_{\bar{q}}| D \wedge I)\, \mathrm{d}\bar{q}, \quad (33)$$

for any L, $\{i_l\}$*, and any D representing a (possibly empty) conjunction of measurement outcomes not containing the instances* $j_1, \dots, j_L$.

The equalities of the above Proposition may be read in two senses, whose meaning is the following: If in a given situation we have introduced circumstances $\{S_{\bar{q}}\}$ (or, equivalently, $\{C_k\}$) and assigned or calculated their plausibility density $p(S_{\bar{q}}| D \wedge I)$, then we also know, automatically, the generating function $\mathit{\Gamma}(\bar{q}| D \wedge I)$ of de Finetti's representation theorem, which we should have introduced had we taken an exchangeability approach. Vice versa, if we approach the problem through exchangeability and assign an infinitely exchangeable distribution to the possible infinite collections of outcomes, we have automatically assigned also the plausibility density $p(S_{\bar{q}}| D \wedge I)$ for *whatever* set of coarse-grained circumstances $\{S_{\bar{q}}\}$ we might be willing to introduce. The density of the Laplace-Jaynes approach and the generating function of the exchangeability one are, in both cases, numerically equal.

What is remarkable in the second direction of the Proposition is that the plausibilities of the coarse-grained circumstances are completely determined from our exchangeable plausibility assignment, even when we have not yet specified what those propositions are about! Although remarkable, this fact is a consequence of the particular way the coarse-grained circumstances are formed from the fine-grained ones. Note, moreover, that the plausibilities of the fine-grained circumstances $\{C_k\}$ would not be completely determined in general, even though their values would be constrained by eq. (25).

Some philosophical importance has the fact that $p(S_{\bar{q}}| \dots)$, numerically equal to $\mathit{\Gamma}(\bar{q}| \dots)$, is an 'actual' plausibility distribution, whereas the latter is only a generating function. It is practically impossible to specify most infinitely exchangeable plausibility distributions *directly*, and this is one of the main points and advantages of de Finetti's theorem: a de Finettian can specify an infinitely exchangeable distribution by making a (often necessary) detour and specifying $\mathit{\Gamma}$ instead. It is, however, a bit disconcerting the fact that one is almost always forced to specify the 'as if' instead of what according to de Finetti is 'the actual plausibility'. The Laplace-Jaynes approach gives instead a direct meaning to the generating function Γ *as a plausibility distribution*, so that one needs not feel embarrassed to specify it.

Can all situations which can be approached through infinite exchangeability also be approached from the Laplace-Jaynes point of view? Asking this means asking whether all situations for which we deem exchangeability to apply can be analysed into circumstances having the properties (I)–(IV). There is surely much place for discussion on the answer to this question, if by 'circumstances' we really mean, in some sense, 'interesting circumstances'. One could also wonder whether on a formal level the answer could be 'yes'. The reason is that the Laplace-Jaynes approach would seem to formally include the infinite-exchangeability one. To see how, make first a judgement of infinite exchangeability, and then introduce and define a set $\{C_{\bar{f}}\}$ of propositions stating that the limiting relative frequencies of an infinite collection of outcomes have certain values $\bar{f}$. This should be possible since the existence of this limit under exchangeability is guaranteed by de Finetti's theorem. Intuitively, such a set $\{C_{\bar{f}}\} =: \{C^{(j)}_{\bar{f}}\}$ for any j would seem to constitute (under a judgement of infinite exchangeability) a set of circumstances that satisfy properties (I)–(IV). If this were true then, whenever infinite exchangeability applied, the Laplace-Jaynes approach would be at least formally viable. But intuition is not a reliable guide here (the definition of a proposition like $C_{\bar{f}}$, e.g., would apparently involve a disjunction over the set of permutations of natural numbers), and the above reasoning is not mathematically rigorous. We have not the mathematical knowledge necessary to rigorously analyse and answer this sort of conjecture, and since we find it in any case uninteresting, let us not discuss it any further. It can be added, however, that any criticism from de Finettians with regards to the infinities involved in the argument would be like sawing the branch upon which they are sitting themselves, since we do not see how one can effectively make a judgement of infinite exchangeability without first considering possibly problematic propositions like $\{C_{\bar{f}}\}$.

6.2 'Unsure' and 'unstable' plausibiliy assignments

The presence of circumstances in the analysis of the problem, and the consequent fact that the plausibility of the outcomes can be decomposed as in eq. (27), suggest an interpretation of those feelings of 'uncertainty' and 'stability' about a plausibility assignment mentioned in § 4, point (a). A person might feel 'sure' about an assignment $1/2$ to heads in the situation I_s because of the following, perhaps unconscious, reasoning: 'If I knew that the coin was two-headed or the toss method favoured heads, I'd assign 1 to heads. If on the contrary I knew that the coin was two-tailed or the toss method favoured tails, I'd assign 0 to heads. If I knew that the coin was a usual one and the tosser knew nothing about tossing methods, I'd assign $1/2$ to heads. But the coin, although I gave to it a rapid glance only, seems to me quite common and symmetric; and I know that the tosser, a friend of mine, knows no tossing tricks. So I can safely exclude the first two possibilities, and I'm practically sure of the third. Yes, I'll give 1/2 to heads'. In this reasoning, the person entertains (for each j) a set of possibilities, like 'The coin is two-headed', 'The toss method favours tails', etc. From these a set of three plausibility-indexed circumstances $S_{\bar{q}}$, with $\bar{q} \equiv (q_{\text{heads}}, q_{\text{tails}})$,[17] is distilled, corresponding to the $\bar{q}$ values $(1, 0)$, $(1/2, 1/2)$, and $(0, 1)$. To these circumstances the person assigns, given the background knowledge I_s, the plausibilities

$$\begin{aligned} \mathrm{P}(S_{(1,0)} | I_s) &\approx 0, \\ \mathrm{P}(S_{(1/2,1/2)} | I_s) &\approx 1, \\ \mathrm{P}(S_{(0,1)} | I_s) &\approx 0. \end{aligned} \tag{34}$$

Then, according to the rules of plausibility theory, the total plausibility assigned to heads is

$$\begin{aligned} \mathrm{P}(\text{'heads'} | I_s) &= \sum_{\bar{q}} \mathrm{P}(\text{'heads'} | S_{\bar{q}} \wedge I_s)\ \mathrm{P}(S_{\bar{q}} | I_s), \\ &= 1 \times \mathrm{P}(S_{(1,0)} | I_s) + 1/2 \times \mathrm{P}(S_{(1/2,1/2)} | I_s) + 0 \times \mathrm{P}(S_{(0,1)} | I_s), \\ &\approx 1 \times 0 + 1/2 \times 1 + 0 \times 0 = 1/2. \end{aligned} \tag{35}$$

The fact to notice here is that amongst all the circumstances, those that lead to a conditional plausibility for heads near to the total plausibility, viz. 1/2, have together far higher plausibility than the others. This fact can be interpreted as the source of the feelings of sureness and stability associated to that final plausibility assignment: the person has conceived a number of hypotheses, and the total plausibility assigned on the grounds of these is practically equal to the plausibilities assigned conditionally on the most plausible hypotheses only. Put it otherwise, the most plausible circumstances are not discordant, all point more or less to the same plausibility assignment. Mathematically this is reflected in the fact that the distribution $\mathrm{P}(S_{\bar{q}} | I_s)$ is concentrated around the median.

The reasoning of a person who feels 'unsure' about an assignment 1/2 to heads in the situation I_u might go instead as follows: 'Had I known that the coin was a common one, I'd have assigned 1/2 to heads. But I have heard from reliable sources that this coin is not a common one, so I can safely exclude that possibility. If I knew that the coin was two-headed, I'd assign 1 to heads. If I knew that the it was two-tailed, I'd assign 0 to heads. But I'm completely unsure about these two possibilities! I have to give 1/2 to heads then'. Also in this reasoning the person introduces a collection of hypotheses and from these distills a set of three plausibility-indexed circumstances similar to those introduced by the 'sure' person. It is in the assessment of the plausibilities $\mathrm{P}(S_{\bar{q}} | I)$ that the two persons — owing to their different background

[17] Note, once more, that the $S_{\bar{q}}$ concern not plausibilities but statements like, in this case, 'The coin is a usual one and the tosser knows nothing about tossing methods', etc.

knowledges I_s, I_u — differ. In this case the assessment yields

$$\begin{aligned} \mathrm{P}(S_{(1,0)} | I_\mathrm{u}) &\approx 1/2, \\ \mathrm{P}(S_{(1/2,1/2)} | I_\mathrm{u}) &\approx 0, \\ \mathrm{P}(S_{(0,1)} | I_\mathrm{u}) &\approx 1/2, \end{aligned} \tag{36}$$

and therefore the unsure person assign to the outcome 'heads' the total plausibility

$$\begin{aligned} \mathrm{P}(\text{'heads'} | I_\mathrm{u}) &= \sum_{\bar{q}} \mathrm{P}(\text{'heads'} | S_{\bar{q}} \land I_\mathrm{u})\ \mathrm{P}(S_{\bar{q}} | I_\mathrm{u}), \\ &= 1 \times \mathrm{P}(S_{(1,0)} | I_\mathrm{u}) + 1/2 \times \mathrm{P}(S_{(1/2,1/2)} | I_\mathrm{u}) + 0 \times \mathrm{P}(S_{(0,1)} | I_\mathrm{u}), \\ &\approx 1 \times 1/2 + 1/2 \times 0 + 0 \times 1/2 = 1/2. \end{aligned} \tag{37}$$

The final assignment is the same as that of the sure person, but the most plausible circumstances for the unsure person would yield conditional plausibilities not near to the final value 1/2. The unsure person would have given 0 or 1 if a little bit more knowledge of the situation had been available.[18] In other words, the most plausible circumstances are discordant and point to different plausibility assignments, and this can be interpreted as the source of the feelings of unsureness and instability. In mathematical terms, the distribution $\mathrm{P}(S_{\bar{q}} | I_\mathrm{u})$ is not concentrated around the median.

The two examples above are very simple, the number of possible coarse-grained circumstances being only three. With a deeper analysis this number could be very large, and the assignments (34) and (36) would be replaced by effective continuous distributions like e.g.

$$p(S_{\bar{q}} | I_\mathrm{s})\, \mathrm{d}\bar{q} \propto (q_\mathrm{heads})^{20}\, (q_\mathrm{tails})^{20}\, \delta(1 - q_\mathrm{heads} - q_\mathrm{tails})\, \mathrm{d}\bar{q}, \tag{38}$$

$$p(S_{\bar{q}} | I_\mathrm{u})\, \mathrm{d}\bar{q} \propto (q_\mathrm{heads})^{-39/40}\, (q_\mathrm{tails})^{-39/40}\, \delta(1 - q_\mathrm{heads} - q_\mathrm{tails})\, \mathrm{d}\bar{q}, \tag{39}$$

whose graphs are those of Fig. 1, blue dashed curve for the first and orange continuous curve for the second. These are Dirichlet (or beta) distributions, known for their particular properties, for which see e.g. [1, 11, 53, 65, 89] (cf. also [51]).

The degree of 'sureness' or 'stability' about a plausibility assignment can thus be apparently captured, formalised, and even quantified by the plausibility distribution of a set of circumstances, $\mathrm{P}(S_{\bar{q}} | I)$. Graphs such as those of Fig. 1 have then a legitimate and quantitative meaning within (single-valued) plausibility theory: They do not represent 'probabilities of probabilities'; they represent plausibility distributions amongst different unknown empirical circumstances, grouped for simplicity according to the conditional plausibilities they lead to.

This interpretation is also consonant with the fact that, according to the results of § 5.4, eq. (30), as the observed data accumulate the distribution $\mathrm{P}(S_{\bar{q}} | I)$ gets more and more peaked around a particular value which gets also nearer the median, reflecting the fact that accumulation of data tends to make us surer of our plausibility assignments.

For further discussion and references, see [2].

[18] The assignment will in fact collapse onto one of those two values as soon as one toss is observed, leading to nought-or-one updated plausibilities for the circumstances $S_{(0,1)}$ and $S_{(1,0)}$.

6.3 Finite and small number of possible observations

As mentioned in § 4 with regard to point (c), there are situations in which the infinite-exchangeability approach cannot be adopted: situations where we know that the number of possible measurement instances is finite — and, in particular, small. In this case a de Finettian has three possible options as regards the application of exchangeability.

The first is to make a judgement of *finite* exchangeability and use the related representation theorem. It is known that the representation theorem for finite exchangeability has a different form from that for infinite exchangeability, eq. (2). The plausibility for a collection of outcomes takes in the finite case the form of a mixture of hypergeometric distributions ('urn samplings without replacement'). In the case of a finite but large number of possible observations this representation converges to the infinitely exchangeable one [9, 10][see also 90], so the de Finettian can make sense of the propensitor's 'unknown propensity' reasoning and of the related formulae, which are formally identical with (2)–(7), as approximations. But in the case of a small maximum number of possible observations the discrepancy between the two representations is too large, and if the 'propensitor' obstinately uses the 'unknown propensity' formulae, the de Finettian will then not be able to 'make sense' of them, as instead was the case with infinite exchangeability.

The second option is to use de Finetti's theorem as generalised by Jaynes [52], i.e. extended to generating functions of any sign, but restricting the choice of the latter to positive ones only. But this restriction would lack meaningful motivation: What kind of plausibility judgement would the restriction reflect? — For we must remember that the specification of the generating function is for the de Finettian only a detour to specify the plausibility distribution of the outcomes, which is the only 'actual' one. The restriction to positive generating functions would only seem to reflect the third option, to which we now turn.

The third option is to enlarge the finite collection of measurement instances with an infinite number of fictive ones, and then make a judgement of infinite exchangeability for the enlarged collection. With this artifice the formulae of the propensitor can be recovered, and seemingly with a meaningful motivation. It seems to us, however, that the introduction of very many (∞) fictive measurement instances, apart from being unpleasant, is formally inconsistent. For the fact that one *knows* that the number of measurement instances can but be finite, say at most n, must be included in the context I; viz., $M^{(j)} \land I$ is false for $j > n$. The plausibility of the outcomes of the fictive 'additional' ($j > n$) measurements, which is conditional on I, is then by definition nought, or undefined. One cannot even say 'Let us suppose that additional measurements were possible', because such a proposition conjoined with I would yield a false context and the plausibilities conditional on it would then be undefined. An alternative would be to eliminate from the context I those 'elements' that bound the number of measurement instances, so that the plausibilities of the fictive additional measurements would neither be nought nor undefined in the new, mutilated context. But mutilation of prior knowledge — granted its feasibility — can be dangerous.

The Laplace-Jaynes approach, on the other hand, applies unaltered to the case of a finite, even small, maximum number of measurement instances. This is clear if we look again at properties (II)–(IV): they nowhere require the instance index j to run to infinity, as cursively remarked directly after their enunciation. Hence, through the Laplace-Jaynes approach we can make straightforward sense of the propensitor's procedure also in these finite situations.

6.4 When is the Laplace-Jaynes approach useful? Making allowance for the grounds behind exchangeability assignments: some examples

We have seen in § 6.1 that the two approaches are not mutually exclusive but can coexist and strongly 'interact'. But in which sense are the two approaches complementary? How ought we to decide which approach to choose? Why? There is no definite answer: as de Finetti warns [75, § 20]:

> in certain enterprises it can be better to evaluate the probability of favourable outcomes all in a block, to see at a glance whether the investment is secure or insecure, and in others it is better to reach this conclusion starting from an analysis of the individual factors that are in play. There is no conceptual difference between the two cases. Someone who wants to estimate the area of a rectangular field can with the same right estimate the area directly in hectares, or estimate the lengths of the sides and multiply them: reasons of convenience, practicality and custom will make one method preferable to the other, the one that seems to us more trustworthy in relation to our capacity to judge, or we can follow both methods, or try other ones, and consider all of them in fixing our opinion.

In assigning a plausibility to some measurement outcomes we may simply want, depending on 'convenience, practicality and custom', to assign plausibilities to all possible collections of similar outcomes (indirectly, by specifying the generating function of de Finetti's theorem), update the plausibilities according to observations already made, and then marginalise for the instances of interest, as described in § 3. Or we may want to analyse the contexts of the measurements into more details and assign plausibilities to these, obtaining the final plausibilities for the outcomes of interest by the theorem on total probability, as described in § 5. Neither approach is more 'objective' than the other, for both must start from some set of (subjective, the de Finettian says) plausibilities that must be given without further analysis.

In situations such as that of a 'common' coin toss or die throw, we (the authors) usually prefer in general the approach through exchangeability. Although such situations could be analysed into more details and hypotheses about the way of tossing etc., we are not really interested in these and prefer to make an initial exchangeable plausibility assignment for future tosses, that we shall update according to observations made — the number of which can be very large (i.e., infinite exchangeability can be applied, at least as an approximation). One could say that it is the plausibilities '$\mathrm{P}(R_i^{(j)}|\dots)$' that are of interest, and we collect measurement data to be put in place of the dots in order to update those plausibilities. But there are examples, especially in the study of natural phenomena, where it is the details — the circumstances — that interest us most and that we want to investigate. In these cases measurement outcomes are collected not only for the sake of prediction of other, unknown ones, but also in order to modify our uncertainty about the circumstances, i.e., to change our initial plausibility assignments about them. In other words, the plausibilities '$\mathrm{P}(C_k|\dots)$' interest us as much as the '$\mathrm{P}(R_i^{(j)}|\dots)$', and measurement data are collected to update both.

A powerful feature of the approach through plausibility-indexed (i.e. coarse-grained) circumstances is that you can 'wait' to think about and to explicitly define the circumstances of interest until you have collected a large amount of observations. The 'procedure' is as follows:

1. *Imagine* to have introduced numerous circumstances $\{C_k\}$, to have assigned the plausibilities $\mathrm{P}(R_i| M \wedge C_k \wedge I)$, and to have performed the 'coarse-graining' of the circumstances into the set $\{S_{\bar{q}}\}$.

2. Assume to have assigned plausibilities $\mathrm{P}(C_k| I)$ such that the plausibility density $p(S_{\bar{q}}| I)$ over the coarse-grained circumstances is enough smooth and non-vanishing at every point. How much is 'enough' is related to the size of the collected data as for the next step.

3. Collect a large amount of outcome observations D. What counts as 'large' is related to the smoothness assumed in the previous step for the plausibility density over the coarse-grained circumstances.

4. Update your distributions according to the observations D. If $\bar{f} \equiv (f_i)$ are the relative frequencies of the observed outcomes, the updated plausibility distribution over the coarse-grained circumstances would now be concentrated around the circumstance labelled by $\bar{f}$, i.e., $p(S_{\bar{q}}| D \wedge I)\,\mathrm{d}\bar{q} \approx \delta(\bar{q} - \bar{f})\,\mathrm{d}\bar{q}$; cf. § 5.4. This means that you would judge $S_{\bar{f}}$ to be the most plausible circumstance given the evidence D. (But remember: the introduction of circumstances is up to now only imagined.)

5. Now re-examine the context as it was before the observations, and *really* introduce a set of circumstances of interest, respecting the assumptions in steps 1 and 2. In this process, however, you only need to concentrate on those circumstances conditional on which the outcomes of any measurement instance are near to $\bar{f}$, i.e. those for which $\mathrm{P}(R_i| M \wedge C_k \wedge I) \approx f_i$. This restriction is sensible since, as for step 4 above, these are the circumstances that have now become most plausible in the light of the observed data.

This process reflects e.g. what we do when, having tossed a coin many times and seen that we nearly always obtain heads, we say 'there's something peculiar going on here, let's see what it can be', and begin to look for circumstances (like who tossed the coin, how the coin was tossed, on which surface it was tossed, how it was manufactured, etc.) which presumably lead to the observed peculiar behaviour.

Examples in the same spirit, although more complex, abound in the natural sciences. Consider the following example from the study of granular materials [e.g., 91, 92]. These are made of a very large number of macroscopic (thermal agitation is unimportant) but small particles, that are hence considered point-like. These materials can then be studied by the methods of statistical mechanics. Often of interest is the particle motion resulting from an externally applied force and the particles' collisions. In the study by Rouyer and Menon [93] (to which we refer for details) we are interested in the measurement — our M — of the horizontal velocity component v of any particle. The possible measurement outcomes $\{R_v\}$ are the possible different values of v.[19] What is the plausibility distribution that we assign to the different outcomes? Depending on the circumstances, we should make different assignments. For example, if we knew that the particle collisions were elastic, the particle density enough low, long-range interactions negligible, and a couple more of details — all of which we represent by C_{MB} —, we should assign a Maxwell-Boltzmann, Gaussian distribution $p(R_v | M \wedge C_{\text{MB}} \wedge I)\,\mathrm{d}v \propto \exp(v^2/v_0^2)\,\mathrm{d}v$, with an appropriate constant v_0. (In the present case we know the collisions to be inelastic, so that C_{MB} is from the beginning known to be false; but let us disregard this to make the example more interesting.) If we knew [93] that the collisions were inelastic, knew that the density presented inhomogeneities, judged the particle density very relevant for the velocity distribution, and knew some other details — denote all them by C_{P} —, we should then assign a particular non-Gaussian plausibility distribution, according to the kinetic-theoretical considerations of Puglisi et al. [94]. Finally, if we knew [93] that the collision were inelastic, that energy was injected in the system homogeneously in space, and some other details — call them C_{NE} —, we should then assign another particular non-Gaussian plausibility distribution, according to the study by van Noije and Ernst [95]. (In the present case we know that energy is injected from the boundary, so that C_{NE} is known to be false; again, we disregard this.)

We do not consider other possible circumstances for the moment, but go on to collect a large amount of observations instead — call these D. The distribution of frequencies thus observed goes like $N_v/N\,\mathrm{d}v = \exp(v^\alpha/v_1^\alpha)\,\mathrm{d}v$ with $\alpha \approx 3/2$, a distribution very similar to that assigned on the grounds of C_{P}, and partially similar to that assigned conditional on C_{NE}. By eq. (11) we are thus led, in both the exchangeability and the Laplace-Jaynes approach, to assign

$$p(R_v^{(N+1)} | M^{(N+1)} \wedge D \wedge I)\,\mathrm{d}v \approx \exp(v^\alpha/v_1^\alpha)\,\mathrm{d}v \tag{40}$$

[19] Hence, the set of outcomes is continuous here but, with appropriate mathematical care, this is not so important.

and to say that

$$\Gamma(\text{‘}\exp(\nu^\alpha/\nu_1^\alpha)\text{’}\,|\,D \wedge I) = p(S_{\text{‘}\exp(\nu^\alpha/\nu_1^\alpha)\text{’}}\,|\,D \wedge I) \quad \text{has a very large value.} \tag{41}$$

You note that Γ has ‘$\exp(\nu^\alpha/\nu_1^\alpha)$’ as argument, which is also the index of S. This is because ν is a continuous variable, hence the vectors $\bar{q} \equiv (q_i)$ of our previous discussions become *functions* $Q(\nu)$ here. The function ‘$\exp(\nu^\alpha/\nu_1^\alpha)$’ thus represents a particular *value* of what is denoted by ‘$\bar{q}$’ in §§ 3 and 5.

The analysis has up to now involved both the exchangeability and the Laplace-Jaynes approaches, and at this point the exchangeability approach basically terminates. But the interesting part of the Laplace-Jaynes approach, instead, begins. In fact, from eq. (41) we are led to assign a vanishing plausibility to the circumstance C_{MB} and a small one to C_{NE} (which we already knew to be false, however) and a non-negligible plausibility to the circumstance C_{P}. This prompts other studies and experiments in order to update that plausibility so as to possibly decrease our uncertainty. Rouyer and Menon [93] actually do this, and eventually come to a vanishing plausibility for the circumstance C_{P}. Other circumstances are still to be formulated and introduced.

In the above typically physical example the ‘circumstances’ are indeed intended as ‘causes’ or ‘mechanisms’. But as we remarked in § 5.1, this needs not always be the case. A very interesting and curious example, for which we believe the circumstances cannot be interpreted as ‘causes’ or ‘mechanisms’, is provided by the study of the Newcomb-Benford law [96–99], here tersely described by Raimi [98, p. 521]:

> It has been known for a long time that if an extensive collection of numerical data expressed in decimal form is classified according to first significant digit, without regard to position of decimal point, the nine resulting classes are not usually of equal size. Indeed, [...] for the occurrence of a given first digit i ($i = 1, 2, \dots, 9$),[20] many observed tables give a frequency approximately equal to $\log_{10}[(i+1)/i]$. Thus the initial digit 1 appears about .301 of the time, 2 somewhat less and so on, with 9 occurring as a first digit less than 5 percent of the time. (We do not admit 0 as a possible first digit.)
>
> This particular logarithmic distribution of first digits, while not universal, is so common and yet so surprising at first glance that it has given rise to a varied literature, among the authors of which are mathematicians, statisticians, economists, engineers, physicists and amateurs.

This example can be analysed along the lines of the previous one. The ‘measurement’ M is the observation of the first digits in a given collection of numerical data, the possible ‘outcomes’ being $R_1, \dots, R_9$. Having made a judgement of exchangeability as regards the outcomes of any number of observations,[21] after many observations our

[20] ‘p’ in Raimi’s text.

[21] Since the number of data is finite, infinite exchangeability can here be used only as an approximation or an idealisation.

plausibility assignment will be very near to the observed frequencies, in this case approximately $(\log_{10}[(i+1)/i])$. Here the approach through exchangeability stops. But the Laplace-Jaynes approach does not. According to eq. (30) the circumstances that have acquired highest plausibility are those which lead us to assign the 'surprising' distribution $(\log_{10}[(i+1)/i])$. The 'varied literature' which Raimi mentions consists almost exclusively in searches and studies of such circumstances. The significant point is that many proposed ones concern not 'causes' or 'mechanisms', but symmetries.

A general and more thorough discussion of how the Laplace-Jaynes approach fits into the formalism of statistical physics will be given in our next study [74], together with an analysis of point (d) of § 4.

7 Generalisations

In the Laplace-Jaynes approach introduced in § 5.1 a set of possible circumstances is defined for each measurement instance, and there is a mutual correspondence of these sets, mathematically expressed in particular by eqs. (III) and (IV). These express the idea that there is a similarity amongst the measurement instances, for they can be analysed into the 'same' set of circumstances; and also the idea that the 'same', but unknown, circumstance holds in all instances. This approach can be generalised in different directions.

A first generalisation is to introduce a set of circumstances $\{C_\lambda\}$, not for each measurement instance, but for all of them *en bloc*. I.e., each circumstance C_λ concerns *all* measurement instances. This idea can be easily illustrated if different measurement instances are characterised by different times: then each C_λ can represent a possible 'history' of the details of the instances.[22] Modifying properties (I)–(III) and the subsequent sections by the formal substitution $C_k^{(\cdot)} \rightsquigarrow C_\lambda$, and dropping property (IV), most part of the analysis and discussion presented in the previous sections holds unchanged, or with minor adaptations, for this generalisation.

A second generalisation, with similarities with the preceding one, is made by dropping property (IV) only. This means that we do not assume the 'same' circumstance to hold at every measurement instance. All conjunctions $\bigwedge_l C_{k_l}^{(l)}$, where the k_l can differ for different l, can then have non-vanishing plausibilities (these conjunctions are obviously similar to the circumstances C_λ above; cf. footnote 22). With this

[22] Note, however, that different measurement instances do not need to be associated to different times in general, so the 'history interpretation' is just an example. A more general way to think of the circumstances $\{C_\lambda\}$ is to introduce for each measurement instance j a set of circumstances $\{C_k^{(j)}\}$, as in § 5.1, but without assuming any correspondence amongst the sets, not even the same cardinality. We then consider all possible conjunctions $\bigwedge_j C_{k_j}^{(j)}$, and each such conjunction can be taken to be by definition one of the C_λ.

generalisation formulae (19)–(21) do not hold; in their place we have

$$\mathrm{P}(R^{(j_L)}_{i_L} \land \cdots \land R^{(j_1)}_{i_1} | \mathfrak{M} \land I) = \\ \sum_{k_1,k_2,\dots} \mathrm{P}(R^{(j_1)}_{i_1} | M^{(j_1)} \land C^{(j_1)}_{k_{j_1}} \land I) \cdots \mathrm{P}(R^{(j_L)}_{i_L} | M^{(j_L)} \land C^{(j_L)}_{k_{j_L}} \land I)\ \mathrm{P}(\textstyle\bigwedge_l C^{(l)}_{k_l} | I) \equiv \\ \sum_{k_1,k_2,\dots} q_{i_1 k_{j_1}} \cdots q_{i_L k_{j_L}}\ \mathrm{P}(\textstyle\bigwedge_l C^{(l)}_{k_l} | I), \tag{19'}$$

$$\mathrm{P}(R^{(j_{N+L})}_{i_{N+L}} \land \cdots \land R^{(j_{N+1})}_{i_{N+1}} | \mathfrak{M} \land R^{(j_N)}_{i_N} \land \cdots \land R^{(j_1)}_{i_1} \land I) = \\ \sum_{k_1,k_2,\dots} \left[\prod_{l=1}^{L} \mathrm{P}(R^{(j_l)}_{i_l} | M^{(j_l)} \land C^{(j_l)}_{k_l} \land I) \right] \mathrm{P}(\textstyle\bigwedge_l C^{(l)}_{k_{j_l}} | R^{(j_N)}_{i_N} \land \cdots \land R^{(j_1)}_{i_1} \land I) \equiv \\ \sum_{k_1,k_2,\dots} q_{i_1 k_{j_1}} \cdots q_{i_L k_{j_L}}\ \mathrm{P}(\textstyle\bigwedge_l C^{(l)}_{k_l} | R^{(j_N)}_{i_N} \land \cdots \land R^{(j_1)}_{i_1} \land I), \tag{20'}$$

and

$$\mathrm{P}(\textstyle\bigwedge_l C^{(l)}_{k_l} | R^{(j_N)}_{i_N} \land \cdots \land R^{(j_1)}_{i_1} \land I) = \dfrac{q_{i_1 k_{j_1}} \cdots q_{i_N k_{j_N}}\ \mathrm{P}(\bigwedge_l C^{(l)}_{k_l} | I)}{\sum_{k_1,k_2,\dots} q_{i_1 k_{j_1}} \cdots q_{i_N k_{j_N}}\ \mathrm{P}(\bigwedge_l C^{(l)}_{k_l} | I)}. \tag{21'}$$

The plausibility distribution for the collection of outcomes is therefore generally *not* exchangeable (the frequentist and the propensitor would say that the events are not 'identically distributed', though independent).

The plausibilistic framework of the last generalisation is used in non-equilibrium statistical mechanics [100–104]. A generic circumstance $C^{(j)}_k$ represents a system's being in a ('microscopic') state k at the time t_j; $R^{(j)}_i$ represents the obtainment of the ith outcome of a ('macroscopic') measurement $M^{(j)}$ performed at time t_j; the plausibilities $\mathrm{P}(R^{(j)}_i | M^{(j)} \land C^{(j)}_k \land I)$ are given by the relevant physical theory for all i, j, k. We initially assign, usually by means of maximum-entropy or maximum-calibre principles, an initial plausibility distribution $\mathrm{P}(\bigwedge_l C^{(l)}_{k_l} | I)$ for all possible state dynamics $\bigwedge_l C^{(l)}_{k_l}$ that the system may follow. Then we update, from the history of the observed measurement outcomes, the plausibilities of the different dynamics and thence those of unobserved outcomes.

Finally, a third generalisation could be the introduction of multiple kinds of measurements instead of a single one M. The resulting representation, which should be easy for the reader to derive, would correspond to de Finetti's theorem for *partial* exchangeability [11, 105].

8 Conclusions

Beside exchangeability, there is another point of view from which the de Finettians and the plausibilists can approach the question of induction and thus also interpret and give meaning to the formulae of the propensitors and some of their locutions like 'unknown probability' or 'i.i.d.'. This point of view, which we have named the 'Laplace-Jaynes approach', is based on an analysis of the particular situation of interest into possible 'circumstances'. These 'circumstances' are propositions that concern details of an empirical nature; in particular, they are not and cannot be statements about probabilities. They do not necessarily involve notions of 'cause' or 'mechanism', but can concern symmetries of the situation under study, or consequences of the observed events. More generally, their choice is fully 'subjective', as are the plausibilities assigned to them. Their main required property is that when they are known they render irrelevant any observational data for the purpose of assigning a plausibility to unobserved events.

The Laplace-Jaynes approach can coexist with that based on exchangeability, and can be a complement or an alternative to the latter. It is particularly suited to problems in natural philosophy, whose heart is the analysis of natural phenomena into relevant circumstances (concerning especially, but not exclusively, the notion of 'cause'). It also allows — without resorting to alternative probability theories — an interpretation, formalisation, and quantification of feelings of 'uncertainty' or 'instability' about some plausibility assignments. Finally, this point of view is applicable in those situations in which the maximum number of possible observations is bounded (and, especially, small) and infinite exchangeability cannot therefore be applied.

The Laplace-Jaynes approach is also straightforwardly generalised to the case of more generic sets of circumstances, the case of non-exchangeable plausibility assignments (with applications in non-equilibrium statistical mechanics), and the case in which different kinds of measurements are present, usually approached by partial exchangeability.

Acknowledgements

PM non può ringraziare mai abbastanza Louise, Marianna, e Miriam per il loro continuo sostegno e amore. Affectionate thanks also to the staff of the KTH Biblioteket, especially the staff of the Forum Library — Elisabeth Hammam, Ingrid Talman, Tommy Westergren, Elin Ekstedt, Thomas Hedbjörn, Daniel Larsson, Allan Lindqvist, Lina Lindstein, Yvonne Molin, Min Purroy Pei, Anders Robertsson — for their patient and indefatigable work. Without them, research would hardly be possible. AM thanks Anders Karlsson for encouragement and advice.

Bibliography

Note: `arxiv` eprints are located at `http://arxiv.org/`.

[1] E. T. Jaynes, *Monkeys, kangaroos and N* (1996), `http://bayes.wustl.edu/etj/node1.html`; revised and corrected version of Jaynes [106]. (Er-

rata: in equations (29)–(31), (33), (40), (44), (49) the commas should be replaced by gamma functions, and on p. 19 the value 0.915 should be replaced by 0.0915).

[2] P. G. L. Porta Mana, A. Månsson, and G. Björk, *From "plausibilities of plausibilities" to state-assignment methods: I. "Plausibilities of plausibilities": an approach through circumstances* (2006), eprint `arXiv:quant-ph/0607111`.

[3] B. de Finetti, *Probabilismo*, Logos **14**, 163–219 (1931), transl. as [75]; see also [107].

[4] B. de Finetti, *Foresight: Its logical laws, its subjective sources*, in Kyburg and Smokler [108] (1964), pp. 93–158, transl. by Henry E. Kyburg, Jr. of [109].

[5] E. Hewitt and L. J. Savage, *Symmetric measures on Cartesian products*, Trans. Am. Math. Soc. **80**(2), 470–501 (1955).

[6] B. de Finetti, *Theory of Probability: A critical introductory treatment. Vol. 1* (John Wiley & Sons, New York, 1970/1990), transl. by Antonio Machi and Adrian Smith; first publ. in Italian 1970.

[7] B. de Finetti, *Theory of Probability: A critical introductory treatment. Vol. 2* (John Wiley & Sons, New York, 1970/1990), transl. by Antonio Machi and Adrian Smith; first publ. in Italian 1970.

[8] D. Heath and W. Sudderth, *De Finetti's theorem on exchangeable variables*, American Statistician **30**(4), 188–189 (1976).

[9] P. Diaconis, *Finite forms of de Finetti's theorem on exchangeability*, Synthese **36**(2), 271–281 (1977).

[10] P. Diaconis and D. Freedman, *Finite exchangeable sequences*, Ann. Prob. **8**(4), 745–764 (1980).

[11] J.-M. Bernardo and A. F. Smith, *Bayesian Theory* (John Wiley & Sons, Chichester, 1994).

[12] P. S. Laplace, (Marquis de), *Théorie analytique des probabilités* (M.me V.e Courcier, Paris, 1812/1820), 3rd ed., first publ. 1812; repr. in [110]; `http://gallica.bnf.fr/document?O=N077595`.

[13] W. E. Johnson, *Logic. Part III. The Logical Foundations of Science* (Cambridge University Press, Cambridge, 1924).

[14] H. Jeffreys, *Theory of Probability* (Oxford University Press, London, 1939/1998), 3rd ed., first publ. 1939.

[15] H. Jeffreys, *Scientific Inference* (Cambridge University Press, Cambridge, 1931/1957), 2nd ed., first publ. 1931.

[16] H. Jeffreys, *The present position in probability theory*, Brit. J. Phil. Sci. **5**(20), 275–289 (1955).

[17] R. T. Cox, *Probability, frequency, and reasonable expectation*, Am. J. Phys. **14**(1), 1–13 (1946).

[18] R. T. Cox, *The Algebra of Probable Inference* (The Johns Hopkins Press, Baltimore, 1961).

[19] E. T. Jaynes, *Probability Theory: With Applications in Science and Engineering: A Series of Informal Lectures* (1954–1974), `http://bayes.wustl.edu/etj/science.pdf.html`; lecture notes written 1954–1974; earlier version of [21].

[20] E. T. Jaynes, *Probability Theory in Science and Engineering* (Socony-Mobil Oil Company, Dallas, 1959), `http://bayes.wustl.edu/etj/node1.html`; see also [19].

[21] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, 1994/2003), ed. by G. Larry Bretthorst; first publ. 1994 at `http://bayes.wustl.edu/etj`. (Chapter 30, *Maximum entropy: matrix formulation*, of the original version of the book, not included in the printed one, is available at `http://web.it.kth.se/~mana/work/cc30e.pdf`). Earlier versions in [19, 20].

[22] M. Tribus, *Rational Descriptions, Decisions and Designs* (Pergamon Press, New York, 1969).

[23] E. W. Adams, *The Logic of Conditionals: An Application of Probability to Deductive Logic* (D. Reidel Publishing Company, Dordrecht, 1975).

[24] E. W. Adams, *A Primer of Probability Logic* (CSLI Publications, Stanford, 1998).

[25] T. Hailperin, *Sentential Probability Logic: Origins, Development, Current Status, and Technical Applications* (Associated University Presses, London, 1996).

[26] J. M. Keynes, *A Treatise on Probability* (MacMillan, London, 1921), 5th ed.

[27] F. P. Ramsey, *Truth and probability*, in Ramsey [28] (London, 1926/1950), pp. 156–198, repr. in [108, pp. 61–92], written 1926.

[28] F. P. Ramsey, *The Foundations of Mathematics and other Logical Essays* (Routledge & Kegan Paul, London, 1931/1950), ed. by R. B. Braithwaite, first publ. 1931.

[29] A. N. Kolmogorov, *Foundations of the Theory of Probability* (Chelsea Publishing Company, New York, 1933/1956), second English ed., transl. by Nathan Morrison, with an added bibliography by A. T. Bharucha-Reid first publ. in Russian 1933.

[30] B. O. Koopman, *The bases of probability*, Bull. Am. Math. Soc. **46**, 763–774 (1940), repr. in [108, pp. 159–172].

[31] B. O. Koopman, *The axioms and algebra of intuitive probability*, Ann. Math. **41**(2), 269–292 (1940).

[32] B. O. Koopman, *Intuitive probabilities and sequences*, Ann. Math. **42**(1), 169–187 (1941).

[33] G. Pólya, *Mathematics and Plausible Reasoning: Vol. I: Induction and Analogy in Mathematics* (Princeton University Press, Princeton, 1954).

[34] G. Pólya, *Mathematics and Plausible Reasoning: Vol. II: Patterns of Plausible Inference* (Princeton University Press, Princeton, 1954/1968), 2nd ed., first publ. 1954.

[35] H. Gaifman, *Concerning measures in first order calculi*, Israel J. Math. **2**, 1–18 (1964).

[36] D. Scott and P. Krauss, *Assigning probabilities to logical formulas*, in Hintikka and Suppes [111] (1966), pp. 219–264.

[37] H. Gaifman and M. Snir, *Probabilities over rich languages, testing and randomness*, J. Symbolic Logic **47**(3), 495–548 (1982).

[38] H. Gaifman, *Reasoning with bounded resources and assigning probabilities to arithmetical statements*, Synthese **140**(1–2), 97–119 (2003/2004), `http://www.columbia.edu/~hg17/`; first publ. 2003.

[39] P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with* Mathematica *Support* (Cambridge University Press, Cambridge, 2005).

[40] E. W. Adams, modus tollens *revisited*, Analysis **48**(3), 122–128 (1988).

[41] D. Lewis, *Probabilities of conditionals and conditional probabilities*, Phil. Rev. **85**(3), 297–315 (1976), see also [42].

[42] D. Lewis, *Probabilities of conditionals and conditional probabilities II*, Phil. Rev. **95**(4), 581–589 (1986), see also [41].

[43] T. Hailperin, *Probability logic*, Notre Dame J. Formal Logic **25**(3), 198–212 (1984).

[44] J. Barwise, *The Situation in Logic* (CSLI, Stanford, 1989).

[45] J. Barwise, *The situation in logic — II: Conditionals and conditional information*, Tech. Rep. CSLI-85-21, Center for the Study of Language and Information, Stanford (1985).

[46] H. Gaifman, *Vagueness, tolerance and contextual logic* (2002), `http://www.columbia.edu/~hg17/`.

[47] B. de Finetti, *Probability and exchangeability from a subjective point of view*, Int. Stat. Rev. **47**(2), 129–135 (1979).

[48] A. Hájek, *What conditional probability could not be*, Synthese **137**(3), 273–323 (2003).

[49] ISO, *Quantities and units*, International Organization for Standardization, Geneva, 3rd ed. (1993).

[50] IEEE, *ANSI/IEEE Std 260.3-1993: American National Standard: Mathematical signs and symbols for use in physical sciences and technology*, Institute of Electrical and Electronics Engineers, New York (1993).

[51] D. V. Lindley and L. D. Phillips, *Inference for a Bernoulli process (a Bayesian view)*, American Statistician **30**(3), 112–119 (1976).

[52] E. T. Jaynes, *Some applications and extensions of the de Finetti representation theorem*, in *Bayesian Inference and Decision Techniques with Applications: Essays in Honor of Bruno de Finetti*, edited by P. K. Goel and A. Zellner (North-Holland, Amsterdam, 1986), p. 31, `http://bayes.wustl.edu/etj/node1.html`.

[53] W. E. Johnson, *Probability: The deductive and inductive problems*, Mind **41**(164), 409–423 (1932), with some notes and an appendix by R. B. Braithwaite.

[54] S. L. Zabell, *W. E. Johnson's "sufficientness" postulate*, Ann. Stat. **10**(4), 1090–1099 (1982).

[55] Y. V. Egorov, *A contribution to the theory of generalized functions*, Russ. Math. Surveys (Uspekhi Mat. Nauk) **45**(5), 1–49 (1990).

[56] Y. V. Egorov, *Generalized functions and their applications*, in Exner and Neidhardt [112] (1990), pp. 347–354.

[57] A. S. Demidov, *Generalized Functions in Mathematical Physics: Main Ideas and Concepts* (Nova Science Publishers, Huntington, USA, 2001), with an addition by Yu. V. Egorov.

[58] M. J. Lighthill, *Introduction to Fourier Analysis and Generalised Functions* (Cambridge University Press, London, 1958/1964), first publ. 1958.

[59] B. de Finetti, *On the condition of partial exchangeability*, in Jeffrey [113] (1980), pp. 193–205, transl. by P. Benacerraf and R. Jeffrey of [105].

[60] P. Diaconis and D. Freedman, *A dozen de Finetti-style results in search of a theory*, Ann. Inst. Henri Poincaré (B) **23**(S2), 397–423 (1987).

[61] H. O. Georgii, ed., *Canonical Gibbs Measures: Some Extensions of de Finetti's Representation Theorem for Interacting Particle Systems* (Springer-Verlag, Berlin, 1979).

[62] F. Lad, J. M. Dickey, and M. A. Rahman, *The fundamental theorem of prevision*, Statistica **50**(1), 19–38 (1990).

[63] F. V. Atkinson, J. D. Church, and B. Harris, *Decision procedures for finite decision problems under complete ignorance*, Ann. Math. Stat. **35**(4), 1644–1655 (1964).

[64] E. H. Kyburg, Jr., *Bayesian and non-Bayesian evidential updating*, Artif. Intell. **31**(3), 271–293 (1987).

[65] I. J. Good, *The Estimation of Probabilities: An Essay on Modern Bayesian Methods* (The MIT Press, Cambridge, USA, 1965).

[66] D. Jamison, *Bayesian information usage*, in Hintikka and Suppes [114] (1970), pp. 28–57.

[67] I. Levi, *Information and ignorance*, Inform. Process. Manag. **20**(3), 355–362 (1984).

[68] C. M. Caves, C. A. Fuchs, and R. Schack, *Unknown quantum states: the quantum de Finetti representation*, J. Math. Phys. **43**(9), 4537–4559 (2002), eprint `arXiv:quant-ph/0104088`.

[69] C. M. Caves, C. A. Fuchs, and R. Schack, *Quantum probabilities as Bayesian probabilities*, Phys. Rev. A **65**, 022305 (2002), eprint `arXiv:quant-ph/0106133`.

[70] C. A. Fuchs and R. Schack, *Unknown quantum states and operations, a Bayesian view* (2004), eprint `arXiv:quant-ph/0404156`.

[71] C. A. Fuchs, R. Schack, and P. F. Scudo, *De Finetti representation theorem for quantum-process tomography*, Phys. Rev. A **69**, 062305 (2004), eprint `arXiv:quant-ph/0307198`.

[72] P. S. Laplace, (Marquis de), *Mémoire sur la probabilité des causes par les évènemens*, Mémoires de mathématique et de physique, presentés à l'Académie Royale des Sciences, par divers savans et lûs dans ses assemblées **6**, 621–656 (1774), repr. in [115], pp. 25–65; `http://gallica.bnf.fr/document?O=N077596`; transl. as [116].

[73] C. M. Caves, *Learning and the de Finetti representation* (2000), `http://info.phys.unm.edu/~caves/reports/reports.html`.

[74] P. G. L. Porta Mana, A. Månsson, and G. Björk, *A general bayesian framework for statistical physics, combining the circumstance approach to induction and the vector representation for physical theories: (Being part III of "From 'plausibilities of plausibilities' to state-assignment methods")* (2007), in preparation.

[75] B. de Finetti, *Probabilism: A critical essay on the theory of probability and on the value of science*, Erkenntnis **31**(2–3), 169–223 (1931/1989), transl. of [3] by Maria Concetta Di Maio, Maria Carla Galavotti, and Richard C. Jeffrey.

[76] J. L. Borges and M. Guerrero, *The Book of Imaginary Beings* (Vintage, London, 1957/2002), revised and enlarged ed., transl. by Norman Thomas di Giovanni in collaboration with Jorge Luis Borges; first publ. in Spanish 1957, second Spanish ed. 1967; `http://www.uiowa.edu/borges/vakalo/zf/home.html`.

[77] J. R. R. Tolkien, *The Silmarillion* (Ballantine Books, New York, 1977/1979), first publ. 1977.

[78] A. N. Kolmogorov and S. V. Fomin, *Measure, Lebesgue Integrals, and Hilbert Space* (Academic Press, New York, 1960/1962), transl. by Natascha Artin Brunswick and Alan Jeffrey; first publ. in Russian 1960.

[79] W. Rudin, *Real and Complex Analysis* (McGraw-Hill, London, 1970).

[80] C. Swartz, *Measure, Integration, and Function Spaces* (World Scientific Publishing, Singapore, 1994).

[81] D. H. Fremlin, *Measure Theory. Vol. 1: The Irreducible Minimum* (Torres Fremlin, Colchester, England, 2000/2004), `http://www.sx.ac.uk/maths/staff/fremlin/mt.htm`; first publ. 2000.

[82] D. H. Fremlin, *Measure Theory. Vol. 2: Broad Foundations* (Torres Fremlin, Colchester, England, 2001/2003), `http://www.sx.ac.uk/maths/staff/fremlin/mt.htm`; first publ. 2001.

[83] D. H. Fremlin, *Measure Theory. Vol. 3: Measure Algebras* (Torres Fremlin, Colchester, England, 2002/2004), `http://www.sx.ac.uk/maths/staff/fremlin/mt.htm`; first publ. 2002.

[84] D. H. Fremlin, *Measure Theory. Vol. 4: Topological Measure Spaces. Part I* (Torres Fremlin, Colchester, England, 2003/2006), `http://www.sx.ac.uk/maths/staff/fremlin/mt.htm`; first publ. 2003.

[85] D. H. Fremlin, *Measure Theory. Vol. 4: Topological Measure Spaces. Part II* (Torres Fremlin, Colchester, England, 2003/2006), `http://www.sx.ac.uk/maths/staff/fremlin/mt.htm`; first publ. 2003.

[86] E. J. McShane, *A unified theory of integration*, Am. Math. Monthly **80**(4), 349–359 (1973).

[87] A. M. Bruckner, *Creating differentiability and destroying derivatives*, Am. Math. Monthly **85**(7), 554–562 (1978).

[88] P. Zakrzewski, *Some set-theoretic aspects of measure theory*, Cubo Matemática Educacional **3**(2), 75–88 (2001), `http://www.mimuw.edu.pl/~piotrzak/publications.html`.

[89] M. R. Novick, *Multiparameter Bayesian indifference procedures*, J. Roy. Stat. Soc. B **31**(1), 29–64 (1969).

[90] S. L. Zabell, *The rule of succession*, Erkenntnis **31**(2–3), 283–321 (1989).

[91] H. M. Jaeger, S. R. Nagel, and R. P. Behringer, *Granular solids, liquids, and gases*, Rev. Mod. Phys. **68**(4), 1259–1273 (1996).

[92] P. G. de Gennes, *Granular matter: a tentative view*, Rev. Mod. Phys. **71**(2), S374–S382 (1999).

[93] F. Rouyer and N. Menon, *Velocity fluctuations in a homogeneous 2D granular gas in steady state*, Phys. Rev. Lett. **85**(17), 3676–3679 (2000).

[94] A. Puglisi, V. Loreto, U. Marini Bettolo Marconi, and A. Vulpiani, *Kinetic approach to granular gases*, Phys. Rev. E **59**(5), 5582–5595 (1999), eprint `arXiv:cond-mat/9810059`.

[95] T. P. C. van Noije and M. H. Ernst, *Velocity distributions in homogeneous granular fluids: the free and the heated case*, Granular Matter **1**(2), 57–64 (1998), eprint `arXiv:cond-mat/9803042`.

[96] S. Newcomb, *Note on the frequency of use of the different digits in natural numbers*, Am. J. Math. **4**(1/4), 39–40 (1881), benford.

[97] F. Benford, *The law of anomalous numbers*, Proc. Am. Philos. Soc. **78**(4), 551–572 (1938).

[98] R. A. Raimi, *The first digit problem*, Am. Math. Monthly **83**(7), 521–538 (1976).

[99] R. A. Raimi, *The first digit phenomenon again*, Proc. Am. Philos. Soc. **129**(2), 211–219 (1985).

[100] E. T. Jaynes, *The minimum entropy production principle*, Annu. Rev. Phys. Chem. **31**, 579 (1980), `http://bayes.wustl.edu/etj/node1.html`.

[101] E. T. Jaynes, *Macroscopic prediction*, in *Complex Systems — Operational Approaches*, edited by H. Haken (Springer-Verlag, Berlin, 1985), p. 254, `http://bayes.wustl.edu/etj/node1.html`.

[102] R. C. Dewar, *Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states*, J. Phys. A **36**(3), 631–641 (2003), eprint `arXiv:cond-mat/0005382`.

[103] R. C. Dewar, *Maximum entropy production and the fluctuation theorem*, J. Phys. A **38**(21), L371–L381 (2005).

[104] R. C. Dewar, *Maximum entropy production and non-equilibrium statistical mechanics*, in Kleidon and Lorenz [117] (2005), pp. 41–55.

[105] B. de Finetti, *Sur la condition d'équivalence partielle* (1938), transl. in [59].

[106] E. T. Jaynes, *Monkeys, kangaroos and N*, in *Maximum-Entropy and Bayesian Methods in Applied Statistics*, edited by J. H. Justice (Cambridge University Press, Cambridge, 1986), p. 26, see also the revised and corrected version [1].

[107] R. Jeffrey, *Reading* Probabilismo, Erkenntnis **31**(2–3), 225–237 (1989), see [3].

[108] H. E. Kyburg, Jr. and H. E. Smokler, eds., *Studies in Subjective Probability* (John Wiley & Sons, New York, 1964).

[109] B. de Finetti, *La prévision : ses lois logiques, ses sources subjectives*, Ann. Inst. Henri Poincaré **7**(1), 1–68 (1937), transl. as [4].

[110] P. S. Laplace, (Marquis de), *Œuvres complètes de Laplace. Tome septième : Théorie analytique des probabilités* (Gauthier-Villars, Paris, 1886), "Publiées sous les auspices de l'Académie des sciences, par MM. les secrétaires perpétuels"; `http://gallica.bnf.fr/notice?N=FRBNF30739022`.

[111] J. Hintikka and P. Suppes, eds., *Aspects of Inductive Logic* (North-Holland, Amsterdam, 1966).

[112] P. Exner and H. Neidhardt, eds., *Order, Disorder and Chaos in Quantum Systems: Proceedings of a conference held at Dubna, USSR on October 17–21, 1989* (Birkhäuser Verlag, Basel, 1990).

[113] R. C. Jeffrey, ed., *Studies in inductive logic and probability* (University of California Press, Berkeley, 1980).

[114] J. Hintikka and P. Suppes, eds., *Information and Inference* (D. Reidel Publishing Company, Dordrecht, 1970).

[115] P. S. Laplace, (Marquis de), *Œuvres complètes de Laplace. Tome huitième : Mémoires extraits des recueils de l'Académie des sciences de Paris et de la classe des sciences mathématiques et physiques de l'Institut de France* (Gauthier-Villars, Paris, 1891), "Publiées sous les auspices de l'Académie des sciences, par MM. les secrétaires perpétuels"; http://gallica.bnf.fr/notice?N=FRBNF30739022.

[116] P. S. Laplace, (Marquis de), *Memoir on the probability of the causes of events*, Stat. Sci. **1**(3), 364–378 (1986), transl. by S. M. Stigler of [72]; see also the translator's introduction [118].

[117] A. Kleidon and R. D. Lorenz, eds., *Non-equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond* (Springer, Berlin, 2005), with a Foreword by Hartmut Grassl.

[118] S. M. Stigler, *Laplace's 1774 memoir on inverse probability*, Stat. Sci. **1**(3), 359–363 (1986), introduction to the transl. [116].