Subjective probability and statistical physics

Jos Uffink
Trinity College, Cambridge CB2 1TQ UK*

July 6, 2009

1 Introduction

Statistical physics has a history stretching back 150 years. The subjective view on probability goes back three centuries. Any paper, such as this one, that aims to survey the interplay between these two themes in just a few pages must therefore necessarily be selective, and also subjective in its choice.

Before discussing the role of the subjective view on probability in statistical physics, it should be clear what is meant by ”subjective”. Particularly, because there are at least two main variants of the subjective/objective dichotomy. We shall meet both variants in our discusion of probability below.

A first sense in which a quantity, quality, attribute or predicate of a physical system (or event, or state) my be called ‘objective’ is that it corresponds to a property of the object (or event or state). For example, the mass, charge, position, etc. of a system are thought to be properties which are either inherent to the system or determined by the state of the system itself. Such attributes do not depend on us, the subjects who consider or observe such systems. Of course one may counter that the numerical values by which we measure mass, charge, position etc., do depend on a coordinate frame of reference, a choice of units and so on, so that, strictly speaking we are not dealing with properties of a system by itself, but rather with a relation between the system and some such a system of reference or units. But this does not alter the point that, even if so conceived, attributes of the kind just mentioned are are then determined by a factual relation between the system in and some given reference system. Whether this relation holds (or not) is considered a fact which does not depend on the beholder.

By contrast, when we say that a painting is beautiful, a piece of music is passionate, or a dish tasty, this is not thought to express an objective state of affairs, but rather a judgment or attitude from the person who considers it. Beauty, as the saying goes, is in the eye of the beholder. Subjective attributes of an object thus seem to reside in the mind perceiving or contemplating such objects rather than inherent in the object itself, or in relations of the object with fixed referential systems. This is the first main variant of the subjective/objective dichotomy.

*Permanent address: Institute for History and Foundations of Science, Utrecht University, the Netherlands, `j.b.m.uffink@uu.nl`

However, one often also reads that the theorems of mathematics or logic are 'objective', even though they do not refer to any concrete object at all. What seems to be intended in such cases is that they follow strictly from axioms and proof protocols, so that no one accepting these rules could dispute the validity of these results. They are not objective in the previous sense, i.e. referring to properties of physical objects — indeed, many mathematicians are Platonists and believe that mathematics purely resides in the mind or the realm of ideas — but only in the sense that there can be no rational disagreement about their validity.

Again, one could point out that objective validity is not so much an attribute of a theorem in isolation, but only in relation to a given system of axioms, rules of proof and so on, but this does not change anything essential. By contrast, then, the predicate 'subjective' is used in this second sense for statements that do not have a rationally compelling justification.

2 A Brief history of subjective probability

2.1 Jacob Bernoulli

The historical development of the theory of probability has been rather well-documented and studied (Hacking 1975, Hacking 1990, Krüger *et al.* 1990). The first main text in this field is Jacob Bernoulli's *Ars Conjectandi* (published posthumously in 1713). Indeed, it is the first source in which the term probability, in its modern quantitative sense, is coined.

Bernoulli distinguishes between objective and subjective certainty. According to him, every event in the past, present or future is objectively certain: the outcome of a throw of a die just as much as tomorrow's weather. Otherwise, he says, one could not make sense of Divine providence and omniscience. But certainty considered with respect to us, i.e. subjective certainty, may come in degrees. There are things we *know*, by revelation, reason, perception, experience or introspection or otherwise. These things have the highest degree of subjective certainty. But of other things we are less certain, our judgments might range from "practically certain", "probable", "doubtful", "improbable", etc.

Bernoulli defines probability as the degree of subjective certainty. He emphasised that probability does not reside in the facts or events that we assign probabilities to. Rather, probability assignments represent an epistemic judgement. The assignment or evaluation of probabilities is a human endeavour, and this is what constitutes the art of conjecturing.

Bernoulli also emphasized that a probability assignment to a given event may differ from person to person, depending on what information they possess. He relates, for example, a story of three ships setting sail from a harbour. At some point in time the news arrives that one of the three ships has been wrecked. Given this information, and nothing more, it might seem reasonable to assign an equal probability of 1/3 to each of the three ships for having suffered this fate. But then, Bernoulli says, suppose someone knows that exactly one ship had been poorly prepared: its captain was incompetent, its crew inexperienced, its maintainance had been neglected, an so on. For the person who has access to this information it would be more likely that precisely this ship is the one wrecked, rather than any of the other two. Of course, Bernoulli had no rules to offer for how much much the probability would have to change given such information or how to weigh the incompetence of the captain against the bad maintanaince of the ship and so on. But his qualitative message is clear enough.

For Bernoulli, probability is epistemic, it refers to human knowledge, human ignorance and expectation and probability assignments may differ from one person to the next, depending on what information they possess. It is not a matter of physics, or objective fact. In this respect, his view can be called a subjectivist. This is not to say that he would completely side with modern subjectivists in probability theory, like De Finetti (cf. section 3). For those authors, probability means a measure of a rational degree of belief, where rationality imposes some constraint of coherence, which entails that a person is bound to the usual axioms of probability theory, but otherwise entirely free to assign values in whichever way happens to represent their personal beliefs, which may well depend on say chauvinism, taste, etc. Indeed, there is no indication that Bernoulli conceived of the possibility that rational persons, possessing exactly the same information, might differ in their probability assignments. For him, probability is about 'knowledge' rather than 'belief'. So in the second sense of 'subjective', i.e. the sense which refers to the question whether all rational persons are assumed to agree with each other in their assignments, therefore, his view might be characterized as objective, or perhaps intersubjective.

2.2 The principle of insufficient reason

Bernoulli's view on the meaning of probability was more or less common to all later authors until and including Laplace, and often called "classical". It is subjective in the sense that it refers to a state of mind, but objective in the sense that it does assume a unique assignment by all rational minds in the possession of the same knowledge. But such a view can only be applied to practical cases if we have some additional rule that tells us how to assign numerical probability values in actual cases of knowledge. And that rule should obviously be equally objective, i.e., it should not allow for arbitrary personal differences of opinion.

Of course, this question, how to assign values to probabilities in a given state of knowledge, is the really hard one for the classical interpretation. In fact, there is only one obvious candidate for such a rule: the so-called Principle of Insufficient Reason (or Principle of Indifference). This principle states that when we contemplate a finite number of possible events, one and only one of them will occur, and our mind is equally undecided about which of will occur, i.e. if the information we possess does not provide any reason to believe that one or another event is favoured, we should assign them equal probability.

This principle is fraught with difficulty, and many authors have commented on its subtle and elusive meaning. For example, one might ask what is meant by "equally undecided", or no event being "favoured", or, as Laplace would put it, that all events in question are "equally possible". Reichenbach (1935) argued that this is just a thinly disguised way of stating that the events are equally probable, so that the principle would be circular. However, what Reichenbach overlooked was the intention behind the principle. The very point of the classical interpretation of probability is that a probability assignment should represent an epistemic judgment about a set of possible events. What the principle demands, therefore, is that if there happens to be a symmetry in our judgment, it should be reflected by a symmetry in the probability assignment. The principle is therefore not circular at all. Another question might be by what criteria we can reach the verdict that there is a symmetry in our judgment. This question, unfortunately, seems to lead to a general weakness in symmetry principles. Even the most sophisticated discussions of symmetry principles in physics or mathematic today state that a symmetry is an operation that leaves some "relevant" structure invariant, but without telling us when or why or by what criteria to determine what is and what is not "relevant". In the case at hand, i.e. the symmetry being an invariance of an epistemic judgment under the permutation of possible events, there might be several sources for such criteria: logical, conventional, empirical, etc.

Let us return to the history. Applications of the Principle of Insufficient Reason can be found throughout the early history of probability theory, long before the principle was named. For example, Huyghens argued that cases that happen "equally easy" have equal probability – but without specifying what he meant by the ease of happening.

Bernoulli was quite familiar with this type of argument, and argued, in what I take to be one of the most penetrating and insightful passages of his book, that it serves us well in all problems of games of chance. For the inventors of such games, Bernoulli argued, have designed such games such that only a given number of mutually exclusive possible events can occur, and that they all happen "equally easy". No doubt, Bernoulli is referring here to the fact that games of dice typically use cubical dice, so that the number of possible faces is fixed and rule that they must be shaken before being thrown. In card games, the number of cards in the deck is fixed and they are to be thoroughly shuffled before the game commences. Somehow these requirements warrant the conclusion that we shall have no reason to favour one face of the die above another, or any card being drawn from the deck above any other card.

However, Bernoulli's goal is to extend the calculus for games of chance to a more universal theory of probability that would be applicable in the areas of civil, moral, and economic affairs. And he was skeptical about the universal applicability of the above type of reasoning in such general problems. What if the number of possible cases is not fixed beforehand, or if no shuffling or shaking mechanism is in place? Consider, as he did, a case outside of the realm of games of chance, the case of Titius, say, a healthy and abled-bodied man of 40 years. What is the probability that Titius would die in the next ten years? Bernoulli considers what it would take to apply the principle here:

> " But what mortal, I ask, could ascertain the number of diseases, couting all possible cases, that afflict the human body in every one of its many parts and at every age, and say how much more likely one disease is to be fatal than another —plague than dropsy, for instance, or dropsy than fever— and on that basis make a prediction about life and death in future generations? [...] [I]t would be quite pointless to proceed along this road." (translation: Newman, 1956 Vol 3, pp. 1452-3).

He therefore distinguished between an *a priori* and an *a posteriori* method of assigning probabilities. The *a priori* method was based on the principle of insufficient reason; and according to Bernoulli its application was effectively restricted to games of chance. The *a posteriori* method was based on observed frequencies of the event in previous trials. Bernoulli argued that one could also form a judgment about the probability of an event a posteriori, on the basis of these observed frequencies in similar cases:

> There is, however another way that will lead us to what we are looking for and enable us to determine *a posteriori* what we cannot determine *a priori*, that is, to ascertain it from the results observed in numerous similar instances. It must be assumed in this connection that, under similar conditions, the occurrence (or nonoccurrence) of an event in the future will follow the same pattern as was observed for like events in the past. For example if we have observed that out of 300 persons of the same age and with the same constitution as Titius, 200 died within ten years while the rest survived, we can with reasonable certainty conclude that there is twice as many chances that he will have to pay his debt to nature within the ensuing decade as there are chances that he will live beyond that time. (ibid. p. 1453)

Bernoulli argued that this second method was neither new nor unusual. What he wanted to add, however, was the claim that by increasing the number of observations, one could increase the degree certainty to become as close to 1 as one wishes. This, he claimed was the consequence of a rightly celebrated theorem he proceeded to prove, nowadays often called the (weak) Law of Large Numbers.

Somewhat ahistorically, using the notation of Kolmogorov's formalism, the theorem is as follows:

> Law of Large Numbers Let $\langle X, \mathcal{A}, P\rangle$ be a probability space and take any event $A \in \mathcal{A}$. Let $\langle X^n, \mathcal{A}^n, P^n\rangle$ by the product probability space describing n independent and identically distributed repetitions of the original probability space. Let $X^n \ni \eta \mapsto \mathrm{relf}_n^A(\eta) \in \{0, 1/n, \ldots k/n, \ldots n\}$ be the function that counts the relative frequency of event A in n repetitions, i.e.
>
> $$\mathrm{relf}_n^A(\eta) = \frac{1}{n}\sum_{i=1}^{n} 1_A(x_i) \qquad \eta = (x_1, \ldots, x_n) \in X^n$$
>
> where 1_A is the characteristic function of A on X. Then:
>
> $$\forall \epsilon > 0 : \lim_{n \longrightarrow \infty} P^n(\{\eta \in X^n \ : \ |\mathrm{relf}_n^A(\eta) - P(A)| \geq \epsilon\}) = 1 \qquad (1)$$

However, Bernoulli's Law of Large Numbers fails to deliver what Bernoulli might have hoped it would, i.e. a method for assigning probabilities on the basis of observed data alone as an alternative to the Principle of Insufficient Reason.

The point is, perhaps, somewhat subtle. The Law of Large Numbers shows that when the number of repetitions increases, the probability of obtaining a sequence of events such that the relative frequency of A lies close to the given value of $P(A)$ becomes arbitrarily close to 1. It does *not* say that when the number of repetitions increases, the probability that an unknown value of $P(A)$ is close to the given value of the relative frequency of A, say $\text{relf}_n^A = \frac{k}{n}$, becomes arbitrarily close to one. Sometimes, the distinction between these two reading is obscured by writing the result of the theorem in the form

$$\lim_{n \longrightarrow \infty} P^n(|\frac{k}{n} - P(A)| \geq \epsilon) = 1$$

which suggests a symmetry between the value of the relative frequency, $\frac{k}{n}$ and $P(A)$. But in a more extensive notation as employed here, it should be clear that (1) is very different from

$$\forall \epsilon > 0 : \lim_{n \longrightarrow \infty} P^n(\{P(A) \in [0,1] : |P(A) - \text{relf}_n^A(\eta)| \geq \epsilon\}) = 1 \quad (2)$$

Indeed, one may debate whether the theory of probability should meaningfully allow the assignment of probabilities to probabilities. Bernoulli's work is still rather vague and uncommitted about the the domain for a probability assignment. He assigns probabilities to "events", or to "things" or "cases". He gives no discussion of the question whether these should be, say, outcomes of a repeatable chance set-up, states of a system, or even arbitrary propositions. In a liberal reading, one might argue that it is possible to assign a probability to anything a rational mind can be uncertain about, including perhaps propositions about an unknown probability.

On the other hand, the modern formalism developed by Kolmogorov is more specific about the domain of a probability measure. Its domain is a σ-algebra $\mathcal{A}$ of "events", all represented as subsets from an underlying space X. Now, this formalism is also largely indiffererent about whether these sets are thought of as representing outcomes of an experiment, states of affairs, or propositions. But it is quite clear that a proposition like "$0.49 \leq P(A) \leq 0.51$", although perfectly meaningful, itself does not designate an event in the algebra $\mathcal{A}^n$ over which the probability measure P^n is defined, and so the formula (2) is not meaningful, let alone that it would be implied by the Law of Large Numbers.

Hence, the claim that after a given number of observations of similar cases, and a given value of the relative frequency, namely $\frac{200}{300}$, we can be reasonably certain that the probability of event A (that Titius will die within a decade) is close to 2/3, as Bernoulli suggested, is a conclusion that simply cannot not be drawn from Bernoulli's theorem.

2.3 Bayes

Referend Thomas Bayes, in another posthumously published work, readressed the problem left open by Bernoulli. He opened his famous *Essay* of 1763 as follows:

> PROBLEM: *Given* the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of its happening in a single trial lies between any two degrees of probability that can be named.

And he elucidated:

"By *chance* I mean the same as probability"

In Bayes' construal of the problem, it is evident that he asks about the probability of a probability, given certain data. Essentially, and in modern notation, Bayes' solution is the following. Consider a probability space $\langle X, \mathcal{A}, P\rangle$ and an "unknown event" $A \in \mathcal{A}$ which has a likewise unknown probability $P(A) = p$.

Before we are given any data, we are uncertain about the value of p , except, of course, that it lies in the interval $[0, 1]$, but we might have some degree of beliefs, not only in whether event A occurs, or how often it occurs in n trials, but also in hypotheses H_{ab}, stating that the value of p lies between any two values a and b, where $0 \leq a \leq b \leq 1$. Let us represent these beliefs by some probability measure Pr[1], which for the hypotheses H_{ab} just mentioned, may be written in the form

$$Pr(H_{ab}) = \int_a^b \phi(p)dp \tag{3}$$

for some (possibly improper) probability density function ϕ.

Now suppose that new information comes along: in a series of n independent trials, the event A had occurred k times, and failed $n-k$ times. Let's call this evidence E_{kn}. According to Bayes, we should now replace our prior belief, as expressed by $\mathrm{Pr}(H_{ab})$ by a conditionalized probability density $\mathrm{Pr}(H_{ab}|E_{kn})$, which can be expressed, using Bayes' theorem, as

$$Pr(H_{ab}|E_{kn}) = \int_a^b \phi(p|E_{kn})dp = \int_a^b \frac{\mathrm{Pr}(E_{kn}|p)\phi(p)}{P(E_{kn})}dp \tag{4}$$

Now, on the right-hand side, $P(E_{kn})$ is just a normalization constant,

$$P(E_{kn}) = \int_0^1 P(E_{kn}|p)\phi(p)dp \tag{5}$$

so that in order to evaluate the requested expression, it remains to specify ϕ and $\mathrm{Pr}(E_{kn}|p)$. Now this latter expression represents our belief that the event E will occur, given the supposition that $P(A) = p$. One can plausibly argue that this probability should be equal to $\frac{n!}{(k!)(n-k)!}p^k(1-p)^{n-k}$. After all, this is the probability that A should happen k times and fail $n-k$ times in n independent trials, under the supposition that the probability of A is $P(A) = p$. This assumption is, essentially, what Lewis (1980) called the "Principal Principle", at least if one assumes that $P(A)$ represents an 'objective chance', (see below).

Finally, there is the question of determining ϕ. Here, Bayes falls back on an argument from insufficient reason. He argued that concerning such an event A, we have no more reason to expect that it should happen any particular number k of times in n trials rather than any other number, so that $\mathrm{Pr}(E_{kn})$ should be uniform for $k \in \{0, \ldots n\}$; or in other words: $\mathrm{Pr}(E_{kn}) = \frac{1}{n+1}$. His next step, then, is to note that this is exactly what results when we choose ϕ to be uniform, and to conclude that this is apparently the right choice for ϕ. Indeed, if one reads his assertion charitably, as intended to hold for all $n \in \mathbb{N}$, there is complete equivalence between

[1]The measure Pr will have as its domain the σ-algebra generated by the union of the σ-algebra $\mathcal{A}^n$ and the cylinder algebra generated by the propositions of the form H_{ab}, for $0 \leq a \leq b \leq 1$.

the claims (i) that ϕ is uniform and (ii) that the probabilities $\Pr(E_{kn}) = 1/(n+1)$ for all $n \in \mathbb{N}$.

Given the two ingredients just discussed we can now formulate Bayes' answer to the problem he stated. The probability that the probability $P(A) = p$ of an unknown event A should lie between a and b, given the fact that this event has happened exactly k times out of n trials is just

$$\Pr(H_{ab}|E_{kn}) = \frac{(n+1)!}{k!(n-k)!} \int_a^b p^k (1-p)^{n-k} dp \tag{6}$$

And from this result, one can also derive the probability that event A will happen on the next, $n+1$-th, independent trial, given that it happened k times in the previous n trials. That calculation, due to Laplace, gives the (in)famous *Rule of Succession* :

$$\Pr(A_{n+1}|E_{kn}) = \frac{(n+1)!}{k!(n-k)!} \int_0^1 p^{k+1} (1-p)^{n-k} dp = \frac{k+1}{n+2} \tag{7}$$

There are several remarks to make about Bayes' work. First, as already mentioned, he transformed Bernoulli's problem into one which is more readily focused on the question of assessing the probability of an unknown event in the light of certain observed data. And he provided a solution. However, his solution comes at a price: Bayes needed to attribute probabilities to probabilities.

Secondly, associated with this radical shift in framing the problem there is also a shift in the role of *a priori* and *a posteriori* assignments. Recall that for Bernoulli, these terms represented alternative methods, with the posterior method supposed to take over where the prior method could not be applied. For Bayes, however, the posterior and the prior are merely consecutive stages in a single method. The posterior probability $\Pr(H_{ab}|E)$ is just an update from the initial or prior probability assignment we already had antecendently to the arrival of the data, namely $\Pr(H_{ab})$. And these probabilities are determined by an appeal to the Principle of Insufficient Reason. thus, while Bernoulli argued that this principle only has limited validity, and that the consideration of empirical data would provide an alternative, Bayes' solution to the problem implies that there is, ultimately, no escape from this principle.

As we have seen, in Bayes' approach it is fundamental that one assigns probabilities to the values of probabilities. And this was something that Bernoulli, or other authors before Bayes, had never explicitly contemplated. This brings forth two further questions. First: Are both probabilities to be interpreted in the same way? Indeed, are the probability P and the "meta"-probability Pr both to be interpreted subjectively? There are two readings that suggest themselves. If the probability Pr represents our beliefs, it seems natural to think that these are beliefs about actual facts or 'states of the world' that may or may not obtain. Hence, since these beliefs are attributed to hypotheses such as H_{ab}, stating that $a \leq P(A) \leq b$, one would be inclined to see such hypotheses as having a similar status: they represent some objective fact unknown to us; something which can only be attributed to the event A and the way it is generated. Thus there would be an objective fact of the matter what the value of $P(A)$ actually is, even though we do not know it. That suggests $P(A)$ plays the role of an objective quantity, i.e. an "objective chance".

On the other hand, one could take the view that both probability and meta-probability are to be taken subjectively. If so, this leads us to another conundrum: if we do not know how to assign a probability value p to an event A, such that $P(A) = p$, can we then still make up our mind about how probable it should be that p takes on a particular value? Note that we are now asked, not to assess an objective issue, but rather to evaluate an epistemic judgement about our epistemic judgments.

In the 20th century, de Finetti in particular rejected the very basis of Bayes' approach. According to him, the notion of an unknown probability is utterly meaningless, or a "mythological pseudo-entity". The idea of assigning a probability to an unknown probability is even worse. For a radical subjectivist like de Finetti, this would involve the prospect of a rational agent, who is somehow unable to evaluate his or her degree of belief to assign to event A, but yet able to judge quantitatively how uncertain (s)he is about what these degrees of belief may be. He regarded this prospect as psychologically absurd.

Today, the name of Bayes is often associated with the subjective view on probability. We have already seen that this subjective view was in fact much older, and explicitly stated by Bernoulli. In contrast, Bayes was much more vague and uncommitted about what he understood by probability. According to the covering letter by Price, Bayes felt apologetic about his definition of probability, and aimed to avoid any dispute about the meaning of the word. Thus while subjectivists in probability theory, including those who radically oppose the notion of objective chance, like de Finetti, often call themselves 'Bayesians', it is ironic to note about Bayes' contribution is that, by introducing probabilities of probabilities, his paper is really the first to invite the notion of objective chance!

3 Modern subjectivism: Ramsey and de Finetti

Since the 1920s the subjective view on probability witnessed a revival which is continuing to the present. The main protagonists of this revival, F.P. Ramsey and B. de Finetti, and (somewhat later) L.J. Savage and R. Jeffrey, espouse a view of probability makes no claim that all rational minds, in the possession of the same information should agree upon their evaluation of probabilities. Rather, this view, sometimes called personalist, holds that any system of beliefs that a person wishes to entertain, whether based on empirical information, expert opinion or sheer prejudice is equally allowed. Thus, there is no need for a principle like that of Insufficient Reason.

The only requirement placed on rational beliefs is that the condition of *coherence*, i.e. that a rational mind should not judge as acceptable bets that one can never win but might loose. One of the main tenets of this approach is the famous Dutch Book Theorem, which states, essentially that this condition of coherence implies that such rational degrees of belief must obey the ordinary rules of the probability calculus (i.e. the Kolmogorov axioms, with the exception of σ-additivity). But the converse is also true: any probability measure obeying the ordinary rules is coherent, and hence represents a legitimate system of beliefs.

Here, I will focus mainly on De Finetti. De Finetti was a subjectivist of a more radical approach than Bayes or Bernoulli had ever been. He utterly rejected the notion of unknown probabilities, and hence the very basis of Bayes' approach to the above problem. De Finetti's main result is his well-known exchangeability theorem (which was actually prefigured by Chuprov and W.E. Johnson). Let us begin with a definition: consider some measurable space $\langle X, \mathcal{A} \rangle$ Suppose $a_1, a_2, a_3, \ldots$ is an infinite sequence of 'event variables', i.e. a_i take values in $\{A, \neg A\}$ for some given $A \in \mathcal{A}$. The sequence is called *exchangeable* just in case, for every $n \in \mathbb{N}$ the probability $P(a_1, \ldots a_n)$ is invariant under all permutations π of $\{1, \ldots, n)$. in other words:

$$P(a_1, \ldots, a_n) = P(a_{\pi(1)}, \ldots, a_{\pi(n)}).$$

Here, P is just any probability measure over the measurable space $\langle X^\infty, \mathcal{A}^\infty \rangle$. We do *not* presuppose that the sequence of events are independent and identically distributed!

If we take, as before, E_{kn} to stand for the event that in a sequence of n trials A has happened k times and failed $n - k$ times, de Finetti's theorem states:

EXCHANGEABILITY The sequence $a_1, a_2, a_3, \ldots$ is exchangeable if and only if there is a probability density function ϕ such that

$$P(E_{kn}) = \binom{n}{k} \int_0^1 p^k (1-p)^{n-k} \phi(p) dp \tag{8}$$

A few remarks about this theorem. First of all, one may note that this is almost as good as Bayes' approach, for the result (8) implies that we can calculate the probability that a next trial will show event A as

$$P(A_{n+1}|E_{kn}) = \frac{n+1}{k+1} \frac{\int_0^1 p^{k+1}(1-p)^{n-k}\phi(p)dp}{\int_0^1 p^k (1-p)^{n-k}\phi(p)dp} \tag{9}$$

and other such conditional probabilities in the same fashion. That is to say, the Exchangeability Theorem gives us the same result as Bayes, in the special case of choosing ϕ uniform. But in (8), of course ϕ might take any form. This absence of a rule deciding the choice of ϕ is of course concomitant with the radical subjectivist's rejection of the Principle of Insufficient Reason.

Secondly, note that the predicate "exchangeable", although attributed to a sequence of events, is not a property of the sequence but of a probability measure. Hence whether or not a sequence is exchangeable is as subjective as the probability itself, and reasonable persons are permitted to differ in their opinion about it.

Third, and perhaps most crucially, no unknown probability is posited in this theorem. The point is that even though we allow probabilities to be defined only on sequences of events, given exchangeability, our belief in future events is just the same *as if* we believed in the existence of some objective quantity p, attributing a probability density ϕ to its various values, and adopted the Principal Principle. But none of these ingredients are posited as postulates for how to deal with an event about which we know nothing; they are derived from the theorem above.

Of course, there is a price to be paid: instead of considering an event "concerning the probability of which we know nothing antecedently to any trials being made", De Finetti has to substitute the assumption that a sequence of such events are exchangeable. But this is hardly a drawback. Indeed, precisely because De Finetti's theorem is explicit in stating this assumption, his approach avoids many objections to Bayes' procedure.

For example, a well-known application to Bayesian updating, is the probability of the sun rising tomorrow, given that it his been observed to rise every morning for some period of time. This example is already discussed in Price's covering letter to Bayes' manuscript, and made particularly famous by Laplace, who calculated this probability in 6 decimals.

One of the well-known objections is that the calculation may seem to work plausibly if the data are formulated as

E_{nn}: The sun has risen every day for n consecutive days

and

A_{n+1}: The sun will rise tomorrow

The result, by the Rule of Succession (7), is $P(A_{n+1}|E_{nn}) = \frac{n+1}{n+2}$. Indeed, by the same analysis, the probability that the sun will continue to rise for another n days will be 2/3.

But if we replace the above by

E': I have seen the sun has risen every day for n consecutive days

A': I will see the sun rise tomorrow

the same formula seems to imply that the longer I have lived to see the sun rise, the more and more probable it becomes that I will continue to do so, leading to the conclusion that I the longer I have lived, the longer I should expect to be alive in the future.

While many, —but certainly not all!— authors accept the argument for $P(A_{n+1}|E_{nn})$ as sound, nobody accepts the same argument for $P(A'|E')$. Yet, on the Bayesian-Laplacian analysis it is hard to point out where the crucial distinction between the two cases lies. In De Finetti's analysis, however, it is quite clear that a sequence of days on which I am either alive or dead, would not be regarded as exchangeable. Thus, precisely because his theorem has a presupposition that may be false, the application of its result more clearly restricted and delineated than in the presentation of Bayes.

4 Subjective probability in statistical physics

Let us now turn to statistical physics. Roughly speaking, this involves the theory of heat as applied, in particular, to gases. As said earlier, there is a 150 year gap between the development of probability and its first appearance in this branch of physics. This is all the more remarkable because one of the first notable contributions to the theory of gases is by Daniel Bernoulli. Bernoulli considered gases as consisting of tiny molecules flying hither an thither, and explained pressure by their collisions on the walls. He retrieved the ideal gas law by assuming that temperature was proportional to their kinetic temperature. But, in contrast to later authors, he assumed that all particles would have a common speed.

Perhaps the most notable point is that Daniel Bernoulli did not even contemplate the idea that the theory of probability could have any bearing on this issue. He was of course quite familiar with the work of his uncle Jacob, and indeed himself one of the foremost probabilists of the 18th century (perhaps most famous for his discussion of the Petersburg Paradox). The fact that Daniel did not think that these two fields of his own expertise, gas theory and probability theory, were connected, underlines, in my opinion, that this was not an obvious way to breach this gap. Indeed, the most natural point of view for Daniel Bernoulli to take would have been that probability dealt with subjective certainty, and not with objective facts about physical systems like gases. Why should anyone mix such issues?

However, a century later things were different. Although Laplace and his followers still held the subjective nature of probability, they were eager to emphasize the role of probability in predicting the stable frequency of repeatable events. Indeed, around 1830, Adolphe Quetelet collected data about all kinds of events in society (marriages, suicides, crimes) or nature often split into the various months of the year, and demonstrated the remarkable stability in their frequencies. But Quetelet did not present these as subjective conclusions from a theory about epistemic judgment; he saw them as empirically grounded objective laws. Indeed he coined the phrase "social physics" to describe his enterprise, in order to convey the sense the society at large was somehow subject to fixed laws, just like the laws of physics, in spite of the apparent free will of all its members. Around 1830–1840 the idea emerged that such mass events were governed by statistical laws that had objective validity and did not care about personal subjective views. The main protagonists of this view were Mills, Boole, Leslie Ellis and Venn.

Maxwell was well aware of the theory of probability, and its classical interpretation. In a letter from 1850, he wrote

> "[T]he true logic for this world is the Calculus of Probability, which takes account of the magnitude of the probability (which is, or which ought to be in a reasonable man's mind). This branch of Math., which is generally thought to favour gambling, dicing and wagering, and therefore immoral, is the only "Mathematics for practical Men",[2] as we ought to be. [...]What is believing? When the probability (there is no better word found) in a mans mind of a certain proposition being true is greater than that of its being false, he believes it with a proportion of faith corresponding to the probability, and this probability may be increased or diminished by new facts." (Harman 1990, pp. 197–8.)

This remarkable statement was written when Maxwell was still a 19-year old student in Edinburgh, 10 years before he made his seminal contribution to statistical physics, in his *Illustrations of the dynamical theory of gases*.

However when Maxwell did start his work on gas theory, he introduced probability in a very different fashion. He defined the probability of a molecule to have their velocity between certain limits as the relative number of molecules in the gas having this velocity, i.e. as a an actual, and objective fact about the dynamical state of the gas, without any reference to a reasonable man's mind. It is hard to say, of course, how

[2] Maxwell refers here to a book by Olinthus Gregory (1825) with the same title. However, that book does not contain any discussion of probability theory; nor does the enlarged (1848) third edition of the book that Maxwell might well have had in mind.

or why Maxwell would reach such a drastic change of interpretation, but it does not seem farfetched to conjecture that his move from Edinburgh to Cambridge, where he would be exposed to the views of Leslie Ellis (like Maxwell, a fellow of Trinity College), or Boole and Venn, who were all critical of the classical interpretation, and espoused a frequency interpretation might have something to do with this (Garber 1973).

Nevertheless, in Maxwell's case, it seems that this professed adherence to an objective interpretation of probability went only so far and was supplemented by arguments that only seem to make sense from a classical viewpoint.

In particular, at the heart of Maxwell (1860) is an argument to show that this probability must follow a form, now known as Maxwell's distribution law.

If we write the velocity distribution function as $f(\vec{v})$, Maxwell assumed that this function (i) factorizes into some functions depending only on he three orthogonal components of the velocity vector:

$$f(\vec{v}) = \phi(v_x)\phi(v_y)\phi(v_z) \tag{10}$$

and (ii) is spherically symmetric, i.e., $f(\vec{v})$ depends on $\|\vec{v}\|$ only.

He showed that these two assumptions imply that

$$f(\vec{v}) = Ae^{-\|\vec{v}^2\|^2/B} \tag{11}$$

However appealing this elegant argument might have been, numerous later authors (Bertrand, Poincaré, Keynes a.o.) have regarded it as not entirely convincing. In fact, Maxwell himself was the first to criticize his own argument, calling it "precarious" in Maxwell (1867).

The main point is that the two conditions employed do not sit well with the intended frequency interpretation of the probability density function f. Maxwell's own cursory motivation in 1860 only refers to the claim that the three components of $\vec{v}$ are "independent & at right angles". If that is all, the motivation seems to rests on a conflation between the term 'independent' as it is used in linear algebra (i.e. linear independence) and probabilistic independence (i.e. factorization of probabilities.) A more promising analysis of the motivation would take his two requirements to reflect a priori desiderata on the probability distribution, motivated by an appeal to a close cousin of the Principle of Insufficient Reason, i.e. the statement that if we have no reason in our information that favours a correlation between events, we are entitled to regard them as probabilistically independent.

That Maxwell might have something like that in mind is suggested by the fact that when he described in 1867 his precarious 1860 argument, he stated assumption (i) as the assumption that the probability of a molecule having a velocity resolved parallel to x lying between given limits is not in any way affected by the *knowledge* that the molecule has a given velocity resolved parallel to y (Maxwell 1867, emphasis added).

It has been pointed out (see e.g. Gillispie 1963, Brush 1976, Vol. II, pp. 183–188) that Maxwells 1860 argument seems to have been heavily inspired by Herschels (1850) review of Quetelets work on probability. This review essay contained a strikingly similar argument, applied to a marksman shooting at a target, in order to determine the probability that a bullet will land at some distance from the target. What is more, Herschels essay is firmly committed to the classical interpretation of probability and gives the principle of insufficient reason a central role. Indeed, he explains the (analogue of) condition (10) as "nothing more than the expression of our state of *complete* ignorance of the causes of the errors [i.e. the deviation from the target] and their mode of action" (Herschel 1850, p. 398, emphasis in the original). If Maxwell indeed borrowed so much from Herschel, it could be that he would also have approved of this motivation of condition (10).

Now, Maxwell's (1867) argument to derive the Maxwell distribution was much more detailed and a clear improvement upon his previous one. This is not to say that the tension between the intended frequency interpretation of f and the assumptions motivated by insufficient reason, i.e. by appealing to the absence of reasons to assume otherwise, disappeared. However it would take me to far afield to analyse this argument. (See Uffink (2007, section 3.3) for more details.)

4.1 Boltzmann

Although the work of Ludwig Boltzmann made a profound contribution to the development of statistical physics, in this paper his work will only receive brief attention, because Boltzmann never championed a subjective view on probability. One ought to realize that even in the late 19th century, probability theory was not yet a subject commonly taught at universities, and Boltzmann was not exposed to any course on the theory as a student, nor is there any evidence that he read classical works like Bernoulli or Laplace. Instead, he seems to have derived his understanding of probability from Maxwell, mixing this with ideas of his own or perhaps derived from other sources. In any case, one does not see him mentioning or referring to any work on probability theory until the late 1890's (when he cited the book by Von Kries).

Nevertheless, it is clear that Boltzmann had less patience with the classical, subjective, notion of probability than Maxwell. Indeed, one of the few instances in which he criticised Maxwell was precisely Maxwell's (1867) appeal to Insufficient Reason (Boltzmann 1872). For Boltzmann, probability was an objective quantity. However, this does not mean that he had a clear and consistent concept of probability. Instead, Boltzmann used many different notions of probability in various stages of his work.

First, we can consider a given gas system consisting of many molecules and ask how many of the molecules have some property, e.g., their velocities within certain limits. One then calls the relative number of such molecules the probability of the event that a molecule meets this specification. This is the concept of probability that Boltzmann took over from Maxwell (1860).

A second notion of probability arises by concentrating on a given molecule in the gas and asking for the fraction of time during a long period which it has a certain property, (e.g. its velocity lying between given limits). Dividing this fraction by the duration of the period gives us another sense of probability. Both of these interpretations appear, for example, in Boltzmann 1868, 1872 and Boltzmann & Nabl (1904).

A third notion of probability is obtained by considering the gas as a whole and ask for it to have some property. In other words, we ask whether the state of the entire gas system has some microstate in a particular region of its phase space. Again, we can then consider the relative fraction of time during a very long period which the system has that property. This is a definition that Boltzmann used in the final section of (Boltzmann 1868).

The fourth notion of probability arises (cf. Boltzmann 1871, 1884, Maxwell 1879) by considering an *ensemble* of gas systems, and count the relative number of systems in the ensemble that have their state in a given region of phase space.

Yet another way of assigning probability to regions in phase space, used by Boltzmann (1877), is to take the relative size as measured by the standard (Lebesgue) measure, of such a region as the probability of that region. This is the view that most modern commentators associate with Boltzmann.

Boltzmann shifted rather easily between these notions, sometimes even within the same paper. Also, he sometimes would introduce intermediate positions, like the "time ensemble", i.e. an infinite sequences of copies of a gas system all prepared in the very same microstate, but at different initial times, say, one second apart (cf. Boltzmann 1894).

The main point to observe is that all these notions, except the last two, are entirely derived from dynamics and hence objective. What the probability of a property is (be it molecular or one of the entire gas) is determined by the dynamical details of the gas, i.e. its initial condition and its evolution equation. And in all cases our knowledge or belief about the system simply is not relevant for Boltzmann. Indeed a thoroughgoing subjectivist would prefer not to call Boltzmannian probabilities probabilities at all.

4.2 Ensembles: Gibbs and Tolman

As already indicated, the interpretation of probability is a bit more subtle for the ensemble considerations, which were pioneered by Boltzmann in 1871 and Maxwell 1879. In this case, probability refers to the number of elements in a collection of copies of the system that have been prepared 'similarly'. These numbers, however, are not determined by the dynamical details of the system and have to be specified independently. Nowadays this is usually done by postulating a probability measure on phase space that represents the ensemble density.

Now one could maintain, as Boltzmann clearly intended, that these probabilities still represent objective facts about the frequencies in such an ensemble, supposing that the ensemble is actually "there". However, later writers who explored ensembles in statistical physics more systematically, in particular Gibbs, saw the ensemble as only a fictional construction which is intended to represent our *knowledge* of the system. In that point of view, the ensemble, or its representative probability measure on phase space, is ultimately interpreted epistemically.

Perhaps the first to stress this epistemic point of view was (Tolman 1938). He argued, for example, that the microcanonical ensemble, i.e. the uniform probability measure over the energy hypersurface, is motivated by an appeal to insufficient reason, i.e. that if we know nothing about the state, apart from the value of its energy, then we should consider regions of equal size on this hypersurface as equally probable. He called this the "fundamental hypothesis of equal *a priori* probabilities". He argued that this hypothesis is supported by the well-known property of the microcanonical ensemble that it is stationary for isolated systems. Thus, the idea that we know nothing in particular about the system except its energy gives no information about its time of preparation.

4.3 Jaynes' Maximum Entropy Principle

In 1957 Edwin Jaynes proposed a much more extensively developed subjective approach to statistical physics, which, however, is also more controversial, also among fellow subjectivists. His adherence to a subjective view on probability is quite explicit. But he resisted a personalist interpretation of this concept, as expounded by Ramsey and De Finetti (cf. (Jaynes 1968)). Instead, his approach is closer to the classical view, in the sense that he considered it a definite desideratum that rational minds in the possession of the same information should assign the same probabilities. His major claim to fame is that he designed a principle to deal with import of observational data that is quite different from the Bayes-Laplace view, and generalizes the Principle of Insufficient Reason.

> MAXIMUM ENTROPY PRINCIPLE: Suppose in a given situation there are exactly n possible (mutually exclusive) events or states of a physical system. Our goal is to pick a probability distribution $p = (p_1, \ldots, p_n)$ from the set S of all probability distributions with n entries. suppose further that all the information we have concerning the system or events is encapsulated into a *constraint* $\mathcal{C}$ that can be modelled as allowing only a subset $C \subset S$ from the space of all probability distributions. Then the appropriate choice for p, given the information we possess, is the distribution $p \in C$ that maximize the Shannon entropy
>
> $$H(p) = \sum_{i=1}^{n} p_i \log p_i. \tag{12}$$

Clearly, in the simple case where we have no information beyond the fact that there are n possibilities, i.e. when $C = S$, this principle immediately leads to the choice $p = (1/n, \ldots, 1/n)$. This means that it contains the old principle of insufficient reason as a special case. However, it goes much further than that since it also prescribes a choice of p when we do have information that may favour one outcome above another. It is easy to show that there is, in fact, a unique distribution that maximizes the expression (12) as long as the constraint set C is convex, as is usually the case.

Extensions of this principle to the case where the number of possibilities are countably or uncountably infinite involve much more work, which I will skip in this context (see (Uffink 1995) for details).

However, I would like to dwell on the worry whether the MEP is compatible with Bayesian procedures. The issue is subtle, and I won't go into all aspects of the debate (see (Uffink 1996, Caticha & Giffin 2006) for details and references) except to point out the basic grounds of this worry.

The problem is, in essence, where the constraint comes from or what it represents. Note that, according to Jaynes' point of view, probability distributions represent epistemic judgments, and, therefore, to adopt a constraint $C \subset S$, means that we decide that our actual judgment should belong to some subset of all possible epistemic judgments. At first sight, this suggests that the constraint represents information about our own epistemic judgment. However, Jaynes often thinks of such a constraint as empirical information (his term is "testable information"). Typically, this information takes the form of data about the occurrence of some event, or a sequence of events. Thus, the question arises how information about the occurrence of an event can be successfully and faithfully represented as a constraint on epistemic judgments. What is more, the question arises how the MEP compares to a Bayesian procedure of conditioning on the same empirical data.

An example, due to Jaynes himself (Jaynes 1983, p 244), may clarify his stance on the problem.

> "Suppose a die has been tossed N times and we are told only that the average number of spots up was not 3.5 as one might expect for an "honest" die but 4.5. Given this information, *and nothing else* what probability should we assign to i spots in the next toss? Let us see what solution the Principle of Maximum Entropy gives for this problem, if we interpret the data as imposing the mean value constraints
>
> $$\sum_{i=1}^{6} i p_i = 4.5." \tag{13}$$

Let us denote the event that the average number of spots in N die throws is α as E_α, and the event that the die shows i spots at the next toss as A_i. In a Bayesian or De Finettian approach, one would take account of the event E_α by forming the conditional probability $P(A_i|E_\alpha)$. That is, in these approaches one would update a previous prior probability assignment $P(A_i)$ by conditioning on the observed data. But not so for Jaynes. As the above quotation shows, Jaynes interprets the data E_α, at least in the case $\alpha = 4.5$, as a constraint of the form

$$C_\alpha = \{p = (p_1, \ldots, p_6) \, : \, \sum_{i=1}^{6} i p_i = \alpha\} \tag{14}$$

and takes account of these data by choosing the distribution $p_i^\alpha \in S$ that has maximal entropy under the constraint C_α.

Now suppose that we had decided beforehand, i.e., antecendently to any trials being performed, to adopt this Maximum Entropy procedure, whatever the value of $\alpha \in [0, 6]$ the N tosses the experiment might yield. In that case, it is possible to determine, beforehand, what our possible epistemic judgments after the arrival of the data E_α would be. Indeed, it it will be some element from a collection

$$\{p^\alpha : p^\alpha \in C_\alpha, \max_{p \in C_\alpha} H(p) = H(p^\alpha), \ \alpha \in [0, 6]\}, \tag{15}$$

i.e. a one-parameter curve in S, each point of which will have a unique intersection with the constraint set C_α. But this curve is a very particular one-dimensional subset in the set S of all probability distributions. And a conflict with Bayesian or de Finettian approaches is now almost immediate.

For one could also consider the question of what probability to assign to the event A_i *before* any trials were made. Applying the MEP in that case , i.e. when no empirical information is available, would yield the judgment $p^{prior} = (1/6, \ldots, 1/6)$. Apparently, Jaynes' MEP principle requires one, in this scenario, to respond to the occurrence of the event E_α by shifting from an initial assignment p^{prior} to a posterior distribution p^α, i.e. to some point on the curve mentioned above. The question is thus how this shift would compare to Bayesian conditionalisation on the event E_α.

Now it is a general result, from the theorem of total probability, that the prior probability is always expressible as some convex mixture of my possible posterior probabilities. Thus:

$$p_i^{prior} = P(A_i) = \int P(A_i|E_\alpha)\rho(\alpha)d\alpha. \tag{16}$$

where ρ is some probability density for the events E_α.

But the shift from p^{prior} to p^α, according to the above MEP procedure, the initial belief p^{prior} is generally *not* representable as a proper mixture of the future beliefs p^α. Thus, these shifts are generally different from the shift in probability from prior to posterior from conditionalisation on the same empirical data.

There are two ways to respond to this problem. First one might wish to make a sharp distinction between assigning probability and changing or updating a probability assignment we already have (or between belief formation and belief kinematics). One may argue that the Maximum entropy Principle is designed for the first kind of problems while Bayesian conditionalisation is designed for the second. Thus, there would be a sharp division of labour between these two procedures without possibility of conflict.

Unfortunately, this response seems to restrict the applicability of the MEP as a method of statistical inference drastically. Typically, it may be the case that information arrives step by step, at different times and one would like to adapt a probability assignment several times. But according to the above division of labour, the MEP would not allow successive application. (As in the example above, where the MEP was applied first to determine the prior and then again to respond to the data E_α.)

The other response is to distinguish more sharply between events and constraints. As Jaynes argued in repsonse to the problem:

> "If a statement d refering to a probability distribution in a space S is testable (for example if it specifies the mean value $\langle f \rangle$ for some function f defined on S, then it can be used a a constraint in the PME; but it cannot be used as a conditioning statement in Bayes' theorem because it is not a statement about any event in S or any other space.
>
> Conversely, a statement D about an event in the space S^N (for example an observed frequency) can be used as a conditioning statement in applying Bayes' theorem, [...] but it cannot be used as a constraint in applying PME in space S because it is not a statement about [...] any probability distribution over S, i.e., it is not testable information in S. (Jaynes 1983, p. 250)"

Now, while I agree with Jaynes that a sharp distinction between events and constraints is advisable, it is hard to reconcile this distinction with Jaynes' own proposal in the example above to "interpret" a statement about the occurrence of some event as a statement about a constraint.

5 Discussion

The subjective view on probability in statistical physics has often been regarded as dangerous or unwanted precisely because it would introduce subjective elements in an otherwise objective scientific theory. Moreover, or so it is argued, such probabilities would fail to explain why things happen. For example (Popper 1982) argued that it is absurd to think that our beliefs could explain why molecules interact or why gases disperse when released into an evacuated chamber. Recently, Albert has echoed Popper's concern:

> "Can anybody seriously think that our merely being *ignorant* of the exact microconditions of thermodynamic systems plays some part in *bringing it about*, in *making it the case*, that (say) *milk dissolves in coffee*? How could that be?' (Albert 2000, 64, original emphasis)

But as argued by (Frigg 2007), the concern is misconceived. Of course, our beliefs or lack of knowledge do not explain or cause what happens in the real world. Instead, if we use subjective probability in statistical physics, it will represent our beliefs about what is the case or expectations about what will be the case. And the results of such considerations may very well be that we ought to expect gases to disperse, ice cubes to melt or coffee and milk to mix. They do not cause these events, but they do explain why or when it is reasonable to expect them.

What critics like Popper and Albert overlooked is what subjective interpreters have often been careful to stress: using a subjective interpretation of probability in statistical physics means that this theory is no longer purely a traditional theory of physics (understood as a description or explanation of physical systems, their properties, relations and dynamics). Rather, it means that one should conceive of the theory as belonging to the general field of Bernoulli's "art of conjecturing", or what is nowadays called statistical inference. That is: it will be a theory about epistemic judgments in the light of certain evidence. The theory is applied to physical systems, to be sure, but the probabilities specified do not represent or influence the physical situation: they only represent our state of mind.

So, in contrast to Popper and Albert, I would argue that the subjective view in statistical physics is clearly a coherent and viable option. This is not to say that it is without internal problems. As we have seen, Bayes and De Finetti differ in whether it is meaningful to speak of unknown probabilities, while Bayes' and Jaynes' approach are hard to reconcile since they model the role of evidence quite differently.

But apart from viability, one might also ask whether a subjective view on probability advances the theory of statistical physics in any way. Does it solve, or help us solve any problem beter than we would be able to do in an objective interpretation of probability?

There is at least one issue where one might expect the subjective view to be helpful. In ordinary statistical mechanics, i.e. adopting an objective view on probability, one is usually faced by the ergodic problem. That is to say, the question whether or not the relative time spent by the state of the system in a particular region in phase space is equal to the microcanonical ensemble average of that same region. The reason why this question is important in the objective view is that ensemble averages are, on the one hand, relatively easy to compute, but lack a solid objective interpretation, since they refer to a fictional construct, i.e. the ensemble. Time averages, on the other hand, are often notoriously hard to calculate but do refer to an objective property of a single system. If one can show that the time averages and ensemble averages coincide, one can, as it were, combine the best of both worlds, i.e. compute ensemble averages and equate them with objective properties of a single system.

Therefore, the Birkhoff-von Neumann ergodic theorem, which specifies under which conditions this equality holds, and generalizations of this theorem in ergodic theory are often regarded as highly relevant in an objective approach. In this approach, then, the assumption of dynamical properties (such as metrical transitivity, mixing, etc.) that form the basis of the ergodic theorem becomes a major issue, and their failure for realistic systems a serious problem in the theory. (cf. (Earman & Redei 1996, Uffink 2007).

In contrast, adherents to the subjective view often take pride that their view has no such qualms. The motivation for adopting an ensemble, in this view, is not that it has any empirical relevance but rather that it adequately reflects our epistemic judgment about the system. Hence, the choice of a microcanonical probability measure need not be justified by reference to dynamical properties like metrical transitivity at all, and the ergodic problem can simply be ignored, or so it is claimed.

However the issue is not that simple. For example, we have seen that Tolman motivated the choice of the microcanonical measure among other things by the idea that when we have no information about about the time at which the system has been prepared, we should adopt a probability measure that is stationary under time-translations. This argument would take us to the microcanonical measure, if the microcanonical measure were the unique stationary measure on phase space. But this is only the case (amongst those measures that are absolutely continuous with respect to the Lebesgue measure) iff the system is metrically transitive. In other words, the results and problems of ergodic theory are relevant to subjectivist and objectivist alike.

Indeed, the subjective view on probability can positively contribute to ergodic theory as is shown by von Plato's (1982) analysis of the ergodic decomposition theorem. This theorem says roughly that any stationary probability measure must be a convex mixture of ergodic probability measures. Von Plato showed that this can be seen as an analogy to De Finettti's exchangeability theorem for arbitrary dynamical systems (See von Plato (1982) and van Lith (2001b) for details.)

A more controversial issue in the question whether the subjective view may be superior to the objective view is Jaynes' claim that his Maximum Entropy approach would yield an "almost unbelievably short" proof of the second law of thermodynamics. Roughly speaking, his argument is as follows. Suppose that at some time t_0 a number of observables $(O_1, \ldots O_n)$ are measured and their values are found to be $(o_1, \ldots o_n)$. We describe the system by a probability density ρ obtained by maximzing the Gibbs entropy

$$S_G = -\int_\Gamma \rho(x) \log \rho(x) dx, \tag{17}$$

subject to the constraints[3]

$$o_i = \int_\Gamma O_i(x)\rho(x)dx \quad i = 1, \ldots, n. \tag{18}$$

Between time t_0 and t_1, the system undergoes some adiabatic process, e.g. with a time-dependent Hamiltonian. Solving Liouville's equation with ρ_0 as initial condition, we obtain the later distribution ρ_1. Use this distribution to calculate

$$o'_i := \int_\Gamma O_i(x)\rho_1(x)dx \quad i = 1, \ldots, n. \tag{19}$$

Now we allow the system to equilibrate, and calculate a new distribution $\hat{\rho}_1$ by maximizing the Gibbs entropy subject to the constraints

$$o'_i = \int_\Gamma O_i(x)\rho(x)dx \quad i = 1, \ldots, n. \tag{20}$$

Now, since the Gibbs entropy is invariant under the evolution by Liouville's equation, it is clear that $S_G(\rho_0) = S_G(\rho_1)$. But it is also clear that since ρ_1 is just one of the distributions that satisfy the constraints (20), we also have $S_G(\rho_1) \leq S_G(\hat{\rho}_1)$. Hence: $S_G(\rho_0) \leq S_G(\hat{\rho}_1)$, which, according to Jaynes proves the Second Law, i.e. entropy cannot decrease during adiabatic processes.

The argument has been criticised by several commentators before (Lavis & Milligan 1985, Earman 1986, Parker 2006) To me the main weakness seems to be to reference to a supposed process of equilibration. There is, of course, no dynamically allowed process that will transform the distribution from ρ_1 into $\hat{\rho}_1$ with higher Gibbs entropy. Therefore, the assignment of $\hat{\rho}_1$ can only be justified by a conscious decision to ignore or throw away information: namely the known values of the observables at an earlier time. But to derive an increase of entropy from a decision to ignore information one actually has, seems a strategy that both runs counter to the spirit of the neo-classical program and at the same time be unconvincing as a derivation of the second law, at least for those authors who feel that this represents an objective law of nature.

To sum up, a subjective view on probability in statistical physics is clearly a viable view to take, when dealing with ensembles. But substantial issues within this approach remain: whether one takes a Bayesian, a de Finettian or a Jaynesian approach on how to model the impact of empirical evidence. To me, De Finetti's approach is by far the most appealing. Moreover, claims that such a subjective view solves or dissolves the problems that appear in an objective view on probabilities in statistical physics are in my opinion premature.

References

Bayes, T. (1763). Essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* 53, 370–418; reprinted in *Biometrika* 45, pp. 293-315 (1958), and *Two Papers by Bayes*, with commentary by W. Edwards, (Deming, New York: Hafner, 1963).

[3] Note how, here too, Jaynes interprets observed data as constraints on probabilities, in defiance of the sharp boundary between the two proclaimed in the passage cited above.

Bernoulli, J. (1713). *Ars Conjectandi.* Basel: Thurnisius; also in: van der Waerden, B.L., 1975. *Die Werke von Jakob Bernoulli*, Band 3, edited by Naturforschende Gesellschaft in Basel. , Birkhäuser Verlag, Basel.

Boltzmann, L. (1868) Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten. *Wiener Berichte*, 58, 517–560. In (Boltzmann 1909) Vol. I, pp 49–96.

Boltzmann, L. (1871) Einige allgemeine Sätze über Wärmegleichgewicht. *Wiener Berichte*, 63, 679–711. In (Boltzmann 1909) Vol. I, pp 259–287.

Boltzmann, L. (1872) Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen *Wiener Berichte* 66 275–370. In (Boltzmann 1909) Vol. I, pp. 316–402.

Boltzmann, L. (1894) On the Application of the Determinantal relation to the kinetic theory of gases. In (Boltzmann 1909) Vol. III, pp. 520–525.

Boltzmann, L. and Nabl, J. (1907) Kinetische theorie der Materie. In F.Klein (ed). *Encyclopädie der Mathematischen Wissenschaften, Vol. V-1*, pp. 493–555 Leipzig: Teubner. .

Caticha, A. & Giffin, A. (2006). Updating Probabilities. In A. Mohammad-Djafari (ed.), *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*,AIP Conf. Proc. 872, 31 (2006) (http://arxiv.org/abs/physics/0608185).

Denbigh, K.G. & Denbigh J. (1985). Entropy in relation to incomplete knowledge. Cambridge: Cambridge University Press.

Earman, J. (1986), the problem of irreversibility. *PSA: Proceedings of the Biennal Meeting of the Philosophy of Science Association*, Vol. II, pp 226–233.

Earman, J. & Redei, M. (1996). Why ergodic theory does not explain the success of equilibrium statistical mechanics" *British Journal for the Philosophy of Science* 47 63–78.

Friedman, K.and Shimony, A. (1971) 'Jaynes's Maximum Entropy Prescription and Probability Theory, *Journal of Statistical Physics* 3, 381–384.

Frigg, R. (2007) Probability in Bolzmannian Statistical Mechanics', forthcoming in G. Ernst & A. Hüttemann (eds.): *Time, Chance and Reduction. Philosophical Aspects of Statistical Mechanics.* Cambridge: Cambridge University Press.

Garber, E. (1973). Aspects of the introduction of probability into physics *Centaurus* 17 11–40.

Garber, E. Brush, S.G., & Everitt C.W.F. (Eds.) (1986).*Maxwell on molecules and gases* Cambridge Mass.: MIT Press.

Gregory, O. (1825) *Mathematics for Practical Men* London : Baldwin, Cradock, and Joy. Third edition revised and enlarged by H. Law (London : J. Weale, 1848).

Gillespie, Ch. C. (1963) Intellectual factors in the background of analysis by probabilities, *Scientific Change*, ed. A.C. Crombie, London: Heinemann, 1963, pp. 431-453, 499-502.

Hacking, I., 1975. *The emergence of probability.* Cambridge: Cambridge University Press.

Hacking I. (1990). The Taming of Chance. Cambridge: Cambridge University Press.

Harman, P. M. (ed.) (1990) *The Scientific Letters and Papers of James Clerk Maxwell* Vol. I, 1846–1862. Cambridge: Cambridge University Press.

Herschel, J.F.W. (1850) Quetelet on probabilities. *Edinburgh Review.* Also in *Essays from the Edinburgh and Quarterly Reviews.* London: Longman, Brown, Green, Longmans, and Roberts, 1857, pp. 365–465.

Jaynes, E.T., (1965). Gibbs vs. Boltzmann entropies. *American Journal of Physics*, 33, 391–398. Also in (Jaynes 1983, pp. 77–86).

Jaynes, E.T. (1968) Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics* 4 227–241.

Jaynes, E.T. (1983). *Probability, statistics and statistical physics.* R. Rosenkrantz (Ed.) Dordrecht: Reidel.

Krüger, L., Daston, L.J. , Heidelberger, M., Gigerenzer, G. & Morgan M.S. (eds) (1990) *The Probabilistic Revolution* Vols. 1 and 2. Cambridge MA: MIT Press.

Lavis, D. & Milligan, P.J. (1985). The work of E.T. Jaynes on probability, statistics and statistical mechanics. *British Journal for the Philosophy of Science* 36, 193-210.

Lewis, D. (1980). A subjectivist's guide to objective chance. In R. C. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, Vol II. Berkeley: University of California Press, pp. 263–93.

Lith, J.H. van (2001a). Stir in stillness : a study in the foundations of equilibrium statistical mechanics. Ph.D Thesis, Utrecht University, igitur-archive.library.uu.nl/dissertations/1957294/title.pdf.

Lith, J.H. van (2001b). Ergodic Theory, Interpretations of Probability and the Foundations of Statistical Mechanics *Studies in History and Philosophy of Modern Physics* 32 581–594.

Newman, J.R. (1956) *The World of Mathematics*, (New York: Simon & Schuster). Reissued by Dover (2000).

Maxwell, J.C. (1860) Illustrations of the dynamical theory of gases. *Philosophical Magazine*, 19, 19–32; 20, 21–37. Also in (Garber, Brush & Everitt 1986, pp. 285–318).

Maxwell, J.C. (1867) On the dynamical theory of gases. Philosophical Transactions of the Royal Society of London, 157 49-88 (1867). Also in (Brush 1966) and (Garber, Brush & Everitt 1986, pp. 419–472).

Parker, D.N. (2006) *Thermodynamics, Reversibility and Jaynes' Approach to Statistical Mechanics* . Ph. D. Thesis University of Maryland.

Popper K.R. (1982) *Quantum theory and the schism in physics*, Totowa, New Jersey: Rowan and Littlefield.

Plato, J. von. (1982) The significance of the ergodic decomposition of stationary measures for the interpretation of probability. *Synthese* 53 419–432.

Quetelet, A.(1846). Lettres a S.A.R. le duc régnant du Saxe-Coburg et Gotha sur la théorie des probabilitités, Bruxelles, Hayez, 1846.

Tolman, R.C. (1938) *The Principles of Statistical Mechanics.* Oxford: Oxford University Press. Resissued by Dover, New York (1979).

Reichenbach, H. (1935) *Wahrscheinlichkeitslehre.* Leiden: A.W. Sijthoff.

Seidenfeld, T. (1986) 'Entropy and Uncertainty' *Philosophy of Science* 53, 467–491.

Shimony, A. (1985) The status of the principle of maximum entropy, *Synthese* 63, 55–74.

Stigler, S.M. *Statistics on the Table: The History of Statistical Concepts and Methods.* Cambridge, Mass.: Harvard University Press.

Stigler, S .M. (1982) Thomas Bayes's Bayesian inference *Journal of the Royal statistical Society Series A*, 145, 250–258.

Uffink, J. (1995) Can the Maximum entropy principle be explained as a consistency requirement? *Studies in History and Philosophy of Modern Physics* 26, 223–261.

Uffink, J. (1996) The constraint rule of the Maximum entropy principle. *Studies in History and Philosophy of Modern Physics* 27, 47–79.

Uffink, J. (2007) Compendium to the foundations of classical statistical physics. In J. Butterfield & J. Earman (Eds), *Handbook of the Philosophy of Science. Philosophy of Physics*, pp. 923–1047. Amsterdam: Elsevier, 2007.

Zabell, S. (2005) *Symmetry and its Discontents* Cambridge: Cambridge University Press.