

# Good just isn't good enough – Humean chances and Boltzmannian statistical physics

Claus Beisbart  
Institute for Philosophy  
University of Bern  
Länggassstr. 49a  
CH-3012 Bern, Switzerland  
Email: Claus.Beisbart@philo.unibe.ch

*May 2013, currently under review for the Proceedings of the ESF Lausanne workshop “New directions in the philosophy of science” in the PSE Series “The Philosophy of Science in a European Perspective”*

## Abstract

Statistical physicists assume a probability distribution over micro-states to explain thermodynamic behavior. The question of this paper is whether these probabilities are part of a best system and can thus be interpreted as Humean chances. I consider two Boltzmannian accounts of the Second Law, viz. a globalist and a localist one. In both cases, the probabilities fail to be chances because they have rivals that are roughly equally good. I conclude with the diagnosis that well-defined micro-probabilities under-estimate the robust character of explanations in statistical physics.

## 1 Introduction

Sometimes, just being good isn't good enough. Sometimes, you have to be better than your competitors. You have to be best, and, maybe, even by far more better than all other rivals. This, it seems, is not only true in sports, but also in metaphysics. For a probability to be a real-world chance, it has to be very good. It can't just be any old Bayesian degree of belief or so. It has to be best; in fact, it has to be much better than any other probabilities. This, at least, follows from David Lewis's account of chances. According to Lewis, chances are best probabilities, i.e., probabilities that occur in a best system.

The question of this paper is whether probabilities assumed in statistical physics are good enough to be chances. Are they real-world chances in the sense defined by Lewis?

Why ask this question? Well, the interpretation of probabilities<sup>1</sup> from statistical physics is a difficult problem. On the one hand, there are strong motives to take them to be ontic probabilities, i.e., probabilities that are independent of human needs and interests (Loewer 2001, pp. 611–612). On the other hand, probabilities used in classical statistical mechanics have to be compatible with determinism. But how can there be chances in a deterministic world (Loewer 2001)?

According to B. Loewer, there can be (Loewer 2001 and Loewer 2004). The idea is that a certain probability distribution over the initial condition of a classical Universe is good enough to yield chances. In the terms used by Callender (2011), Loewer takes a globalist strategy. The details of Loewer’s account have been criticized by Frigg (2008). Frigg & Hoefer (2010) and Frigg & Hoefer (2013) have thus pursued a different strategy, which is localist in the terms of Callender.

I will argue that a globalist strategy runs into problems, but not quite for the reason that Frigg (2008) provides. I will then argue that a localist strategy, though more promising in one respect, doesn’t do much better. The gist of my central objection is that good just isn’t good enough. Winsberg (2008) is also critical of Frigg and Hoefer, but not quite for the reasons given in my paper.

The plan is as follows. In section 2, I will shortly sketch the Humean approach to chances. Sec. 3 argues that macro probabilities concerning thermodynamic behavior may be part of the best system. Sec. 4.2 turns to statistical mechanics and examines the localist and the globalist strategies. Objections and replies are presented in Sec. 5. I draw my conclusions in Sec. 6.

This paper is subject to a few limitations. I will only consider classical statistical mechanics and assume that the world is deterministic (which I don’t take to be true). My question really is whether probabilities that statistical physicists assume for a classical world are chances. Further, I focus on Boltzmannian statistical mechanics in the way presented in Albert (2000). My focus is on the explanation of how equilibrium is approached. I will not consider approaches that rely on ergodicity or a variety thereof (see Lavis 2011, see Frigg & Werndl 2011b for new developments). Finally, as Meacham (2010), p. 1133 observes, there aren’t yet many works that cash out the details of a Humean account of chances in statistical physics (with some exceptions, e.g. Frigg & Hoefer 2013). This paper will try to advance things a bit in this respect. Nevertheless, it is still sketchy and I offer my apologies for this. The only excuse is that this is more or less the state of the art.

---

<sup>1</sup> See Hájek (1997), pp. 210–211 and Hájek (2010) for the interpretation of probabilities quite generally.

## 2 How to win the race: A short guide to Humean best system probabilities

According to Lewis, probabilities are chances if and only if they are in some sense best. But what does it mean to be best and what are the standards of the competition? In this section, I point out how I understand Lewis's proposal, and I introduce a few assumptions that will be important for my argument.

For Lewis, probabilities are best if they are part of the best system. Roughly, the best system provides something like an optimal pocket guide to the Universe. It is very informative, but simple.

In more detail, the systems in the competition consist of sentences. Some sentences are non-probabilistic, others assign probabilities to events or propositions.<sup>2</sup> The corresponding probabilities are purely formal, they are not yet supposed to refer to chances. As a consequence, the probabilistic sentences cannot be true or false. But otherwise, the competition is restricted to systems that are true in their non-probabilistic sentences.

The race is defined in terms of three pro-tanto desiderata. First, other things being equal, a system is better than another if it is *stronger*. In the simplest case, strength is just logical strength. But things are not always simple, in particular, logical strength does not apply to probabilistic sentences that lack definite truth values. Nevertheless, a system  $\Sigma$  is certainly stronger than another,  $\Sigma'$ , if it entails probabilities for all propositions to which  $\Sigma'$  ascribes probabilities; if it entails probabilities for propositions for which  $\Sigma'$  does not entail probabilities. The idea is that  $\Sigma$  makes progress compared to  $\Sigma'$  because it provides at least probabilities for some propositions with respect to which  $\Sigma'$  does not say anything at all.

Second, other things being equal, a system is better than another if it is *simpler* (it's a pocket guide for tourists, so it can't be too complicated!). The appeal of simplicity is clear enough, but it's also clear that more has to be said about simplicity. In this paper, we need only compare the simplicity of what I call *equiprobability models*. These are probability models that are zero everywhere apart from a region  $D$ , in which the probability density is a multiple of the Lebesgue measure.  $D$  is called the support of the model. I assume that an equiprobability model is simpler than another if the support can be identified using less information.

Third, other things being equal, a system is better than another if it fits better to the patterns of the actual world. For a very simple example, if a coin is flipped 10 times and lands heads four times, then a probability of .5 fits the pattern of outcomes much better than a probability of .6, because the actual outcomes of the trials are more probable under the first probability model than under the second. More generally, a probability model is the better the more likely the patterns of the real world are under

---

<sup>2</sup> Probabilities are assumed to be attached to propositions, but, for convenience, I will sometimes also speak of the probabilities of events.

the model. This idea is very difficult to generalize to probability models that live on infinite spaces (Elga 2004), but this will not be a problem in what follows. I will only need an assumption about equiprobability models with support regions that each include all actual data points relevant for the model. One of these models yields better fit if its support is a proper subset of the support of the other. A possible motivation is that the first model provides more information than the other because the actual patterns are more narrowly circumscribed.

The three criteria apply other things being equal, but often other things are not equal. Often, you can improve the fit of a probability model by moving to a less simple model. To obtain an overall comparison between both models and the corresponding systems, we have to strike a balance between the desiderata. Lewis assumes that, at least in many cases, there is only one rational way to do so. Overall, one system is better.

In this way, pairs of systems are compared. It is plausible to assume that there are limits to the quality of systems and that there is a best system. The probabilities in the best system win the race, and as a prize, they become chances.

But is it always possible to single out one winner? There may be two systems with two different probability models, but we may not be able to strike a balance between them in a unique rational way. Maybe, two different ways to strike the balance are equally rational or permissible. It's not clear which system is better overall. The most promising strategy to deal with this problem seems to deny that there are any chances in this case. The idea is that we only speak of chances if the competition singles out a unique winner that is by far more better than the other systems such that there can be no quarrels as to who the winner is.<sup>3</sup>

There are complications and problems that we can bracket in what follows.

First, for the purposes of this paper, the exact nature of the pattern to which probabilities are fitted can be left open if only the pattern is non-probabilistic.

Second, in this paper, I will not draw on the so-called Principal Principle (PP, for short; Lewis 1980). The principle is not proper part of the best-system account of chances (although it may follow from it, see e.g. Hoefer 2007), and its precise formulation is controversial (see e.g. Lewis 1994).

Third, as Lewis himself is aware (Lewis 1994, p. 479), simplicity may ultimately turn out to be a mind-dependent. If this is so, chances are not really objective. I will leave this problem at one side.

Fourth, Lewis submits that chances are incompatible with determinism (Lewis 1986b, pp. 117–121). But Loewer (2001), Loewer (2004), Hoefer (2007) and Frigg & Hoefer (2010) argue that indeterminism doesn't follow from the core of Lewis's account, but only if further commitments on the part of Lewis are taken into account. It is then

---

<sup>3</sup> This strategy is employed by Lewis concerning laws (Lewis 1994, p. 479). I'm not aware that he uses the same strategy concerning probabilities, but, in any case, it seems a very natural move. Cf. Winsberg (2008), p. 881.

suggested that these commitments should be given up. I will follow Loewer, Frigg and Hoefer in assuming that chances do not imply indeterminism.

Finally, Hoefer (2007) and Frigg & Hoefer (2010) ban non-probabilistic statements from their systems altogether and only consider probability models that are fitted to patterns in the real world. But non-probabilistic regularities do play an important role even if our main interest is in chances. For we do not assign chances to patterns that are fully described in terms of deterministic laws. Frigg and Hoefer have to rule this out by hand. Indeed, Hoefer (2007), pp. 563–564 requires that the patterns to be captured by chances look random. This requirement is not necessary if we allow for non-probabilistic statements in the systems.

### 3 Thermodynamics and chances

Our main question is whether the probabilities assumed in statistical mechanics are chances in the sense defined by Lewis. Some of these probabilities are supposed to explain the validity of the so-called Second Law of Thermodynamics. For the remainder of this paper, it is useful to have a brief look at thermodynamics. In this section, I will suggest that the Second Law may be a statement about chances.

This suggestion is absurd if the Second Law claims that entropy is always increasing until the equilibrium is attained. But such a claim would be too strong. In fact, micro-physics and statistical physics would not underwrite such a strong claim.

In a more cautious formulation, the second law reads (cf. Frigg 2008):

SL<sub>0</sub> For all times  $t$ , if an isolated macroscopic system has an entropy  $S(t)$  at time  $t$ , then, for every  $t' > t$ , the entropy  $S(t') \geq S(t)$  with a very high probability.

We may add that, very likely, the entropy increases unless the system is already in equilibrium. To keep things simple, I bracket this part of the Second Law, but nevertheless talk of entropy increase.

In this version, the Second Law is about probabilities. The question thus arises whether the probabilities are chances?

SL<sub>0</sub> is only qualitative because no values of the probabilities are specified. Lewis's account of chances, by contrast, only applies to quantitative probabilities. But SL<sub>0</sub> may only be a first shot for a more precise, quantitative version of the Second Law. Here is how the more quantitative version may look like:

SL For all  $N \in \mathbb{N}$ , for all energies  $E$ , for all times  $t$ , if an isolated macroscopic system of energy  $E$  with  $N$  degrees of freedom has an entropy  $S(t)$  at time  $t$ , then, for every  $t' > t$ , the entropy  $S(t') \geq S(t)$  with a probability of  $P(N, E, S(t))$ .

Here  $P(N, E, S)$  is a real function with values in  $[0,1]$ . For most values of  $N$ ,  $E$  and  $S$ ,  $P(N, E, S)$  is very close to one, but apart from this, the exact shape of the function is not yet known.

SL provides unique probabilities for the increase in entropy. The probabilities depend on the size of the system, its energy and its current entropy. Maybe, the function should also depend on other characteristics, e.g. on the strength of the most important interaction. It is not my task here to specify  $P(N, E, S)$ ; future physics should try to do so.

The question then is whether  $P(N, E, S)$  is a chance function. In a Lewisian framework, this is to ask whether  $P(N, E, S)$  is good enough to be a chance function? To be more precise, is it possible to pick a function  $P(N, E, S)$  such that it is part of the best system?

The best system in a classical world will certainly include the classical deterministic micro-dynamics. Call the system that only contains the deterministic micro-physical laws  $\Sigma(CM)$ . Add SL to  $\Sigma(CM)$  and call the resulting system  $\Sigma(CM, TD)$ . If  $\Sigma(CM, TD)$  is an improvement over  $\Sigma(CM)$ , and if it is so in an optimal way, as far as the description of entropy is concerned, then there is a strong reason to think that the best system includes  $\Sigma(CM, TD)$ , which is sufficient for  $P(N, E, S)$  specifying chances.<sup>4</sup>

To check whether the addition of SL improves over  $\Sigma(CM)$  in an optimal way, we have to apply our criteria.

- Strength:  $\Sigma(CM, TD)$  is stronger than  $\Sigma(CM)$  because a system is stronger than another if the Boolean algebra of propositions to which probabilities are assigned becomes larger. Admittedly, we would obtain at least as much strength if we added a non-probabilistic statement about the values of entropies in isolated systems. However, assuming that there are no deterministic laws about the increase of entropy, this statement would either be false or extremely complicated. Both ways it could not be part of the best system.
- Simplicity and fit:  $\Sigma(CM, TD)$  is less simple than  $\Sigma(CM)$  because we have added a sentence to  $\Sigma(CM)$  that is unconnected to others. In particular, it uses two predicates that do not occur otherwise in the system, viz. “isolated system” and “entropy”.<sup>5</sup> However, at least the predicate “isolated system” can easily be defined in terms of the micro-physics (the condition being that there are no interactions with other systems). Further, maybe, even entropy can be defined in terms of the micro-physics. In any case, the loss in simplicity is not huge if the form of  $P(N, E, S)$  is relatively simple.

Now we don’t know this form. As a consequence, neither simplicity nor fit can

---

<sup>4</sup> Things are in fact more complicated. Even if  $\Sigma(CM, TD)$  is an optimal improvement over  $\Sigma(CM)$ , as far as entropy is concerned, we may be able to move to an even better system that covers other aspects of the pattern and entails a different probability function for the increase of entropy. But this possibility will not matter for the purposes of my argument.

<sup>5</sup> Note that the word “system” is ambiguous in this paper, it either refers to systems of representations (e.g. Lewis’s best system) or to physical systems. I assume that readers will always be able to find out which sort of systems I refer to.

really be specified. But it seems at least possible that there is one unique model that provides reasonable fit, but is sufficiently simple and thus strikes a the unique rational balance between both desiderata.

If there is such a function, then adding it to  $\Sigma(CM)$  would provide an optimal improvement over  $\Sigma(CM)$  and  $P(N, E, S)$  specifies chances.

## 4 Probabilities in Boltzmannian statistical physics

### 4.1 Boltzmannian statistical physics

Boltzmannian statistical physics tries to explain the Second Law in the following way. Each isolated macroscopic system is assumed to consist of a great number ( $N$ ) of micro-constituents, typically atoms or molecules. The dynamics of the system is described using a  $6N$ -dimensional phase space. Since the system is isolated, energy is conserved, and the motion is confined to a  $(6N - 1)$ -dimensional hyperplane  $\Gamma$  in phase space. Each possible micro-state of the system corresponds to a point on the hyperplane; as time passes, the systems takes a trajectory in  $\Gamma$ .

From the macroscopic point of view, the system is described in a coarse-grained way only using macro-information, e.g. state variables such as temperature. Using the macro-information, we can define macro-states  $M_i$ . Each macro-state  $M_i$  can be realized by a great number of micro-states and thus corresponds to a subset of  $\Gamma$ , call it  $\Gamma_i$ . Altogether, the macro-states induce a partition of the energy hyperplane  $\Gamma$ .

The phase-space volume of a macro-state can be measured in a very natural way using the  $(6n - 1)$ -dimensional Lebesgue measure  $\mu$  over  $\Gamma$ . The entropy of macro-state  $i$  is defined as

$$S_B(M_i) = k \ln(\mu(\Gamma_i)) , \quad (1)$$

where  $k$  is the Boltzmann constant. The entropy of the system at time  $t$  is then defined via the unique macro-state  $M_{i(t)}$  in which the system is at  $t$ :

$$S_B(t) = k \ln(\mu(\Gamma_{i(t)})) . \quad (2)$$

At least for simple systems, this entropy can be shown to coincide with thermodynamic entropy (see Frigg & Werndl 2011a for details). In the following, I will always assume that this equality holds. I will thus drop the subscript ‘‘B’’ in the following.

The main idea then is to introduce a probability distribution over the micro-states and to derive the proposition that entropy increases with a very high probability. Roughly, the crucial probability model locates the system in its current macro-state and is flat or homogeneous within this macro-state. That is, within this macro-state, the probability measure follows the Lebesgue measure. Otherwise, it is zero. The region over which a non-zero probability density is assumed is called the support of the probability model.

The equilibrium state has a maximal phase space volume, and it is intuitive that most trajectories end up being there for a long time.

According to current wisdom, the story can only work if we additionally assume the Past Hypothesis, viz. that the initial state had a very low entropy (see Albert 2000 for an influential formulation). The trick is most often applied to the whole Universe, but we can also apply it just to the system under consideration (more on this below). In the remainder of this section, we will apply this strategy to an isolated system, be it the Universe or not.

Here are the details (I here orient myself after Frigg 2008): Let  $t_0$  be the initial time. For each macro-state  $i$  and each time interval  $\Delta t$ , define a region in phase space  $B_{i,\Delta t}$  as follows.  $B_{i,\Delta t}$  comprises all those points within  $\Gamma_i$  that evolve, within time interval  $\Delta t$ , into states corresponding to an entropy at least as high as  $S(M_i)$ .<sup>6</sup> Denote the initial low-entropy macro-state by  $M_{i_0}$ . Let  $\Gamma_t^0$  be that phase space region in which trajectories beginning in the Past State end up at  $t$  (the following would also work if the Past State were defined as a union of macro-states) and consider the following claim.

PD Phase space dynamics: For each  $\Delta t > 0$  for each  $t > t_0$ , for each  $i$ :  $\mu(B_{i,\Delta t} \cap \Gamma_t^0) \approx \mu(\Gamma_i \cap \Gamma_t^0)$ .

PD holds that most trajectories that start from the Past State and that are in a specific macro-state  $i$  at  $t$  run into regions of phase space with an entropy at least as high as that of  $i$ . Here, trajectories are “counted” using the Lebesgue measure. PD is an assumption about the dynamics. If PD or some suitable variant thereof holds true, it should follow from the micro-dynamics of the system. In this paper, I will take it that PD or a suitable variant thereof can be shown.

In addition, we need an assumption about probabilities over micro-states.

E For each time  $t > t_0$ , the probability that the system is in phase space region  $B$  is

$$\frac{\mu(B \cap \Gamma_t^0 \cap \Gamma_{i(t)})}{\mu(\Gamma_t^0 \cap \Gamma_{i(t)})}. \quad (3)$$

$M_{i(t)}$  is the current macro-state,  $\Gamma_{i(t)}$  the phase space region corresponding to it. E is a sort of equiprobability assumption because it assumes that the probabilities are homogeneous (or flat) concerning the Lebesgue measure. The homogeneous distribution extends over that part of the region of the current micro-state that can be reached from the Past State.

PD and E imply that, for every  $t$  and for every  $\Delta t$ , there is a high probability, that  $S(t + \Delta t) \geq S(t)$ . Thus, a qualitative version of the Second Law follows. Ideally, we can strengthen PD and E to obtain SL, but this will not be our concern in what follows.

---

<sup>6</sup> It is here assumed that the dynamics is homogeneous or stationary in time, i.e., if  $\mathbf{x}(t)$  is a possible trajectory in phase space, so is  $\mathbf{x}(t + T)$  with a constant  $T$ .



There may also be other improvements over our formulations of SL, PD and E. I do not think that this affects my arguments concerning probabilities.

E introduces probabilities over micro-states. Our question is whether they are chances. To answer this question, we have to distinguish between two strategies. Both underwrite E, but in different ways. Globalists assume a low-entropy state for the whole universe, assume a probability distribution over the micro-state and try to derive everything else. Localists postulate a low-entropy state and a certain probability distribution for each thermodynamic system (see e.g. Callender 2011 for the distinction). In the following, I will consider each strategy in turn.

The setting is always as follows. We start with the system  $\Sigma(CM, TD)$  and add certain probabilistic assumptions that lead to E. The question is whether the assumptions improve upon  $\Sigma(CM, TD)$  in an optimal way. If so, there is a good case that they form part of the best system of the world and thus are chances.

## 4.2 Globalism about micro-probabilities

Turn first to globalism as adopted by Loewer (2001) and Loewer (2004). Granted the deterministic laws that fully describe the dynamics of the world, we would gain a lot of information if we added sentences fully describing the initial condition. However, we will also pay a lot in simplicity, because information about initial condition has to take the form of a myriad of unconnected sentences (there is one particle at position  $x$  and with momentum  $p$ , and so on). Adding precise information about the initial condition will thus not improve the system overall. But there seems to be an alternative: We fit a simple probability distribution to the initial condition.

But how can we fit initial condition using a simple model? According to Loewer (see in particular the reconstruction by Frigg 2008), an equiprobability distribution is assumed within the region corresponding to the initial macro-state, which has a very low entropy according to the Past Hypothesis. Roughly, micro-states not compatible with the macro-information have zero probability of occurring, while micro-states within the right part of the phase space are equally likely.

Adding this probability model over the initial condition clearly adds strength. The model is also simple and fit seems reasonable. So we have made some progress by adding the model to  $\Sigma(CM, TD)$ . But maybe there is an alternative way of improving upon  $\Sigma(CM, TD)$ ?

We can improve fit when we further shrink down the support of the probability model. That is, we assume a flat probability distribution over a certain sub-region of the past low-entropy macro-state. That sub-region may be defined by the demand that a certain elementary particle has a kinetic energy larger than a particular value  $e_0$ , for instance.

If we do so, we have to pay in simplicity though because the new probability model is less simple than the old one. The new probability model is still an equiprobability

model, but it is less simple to characterize its support. To characterize the support, we begin with the support of the old probability model and further demand that the kinetic energy of a particular particle be  $e_0$ .

So overall, would we improve the system by shrinking down the support? This is a difficult question, and the answer is far from clear. The only thing we can say is that fit is considerably improved, but that there is a considerable cost in simplicity too. So it's not a case in which the right sort of balance favors one system rather than the other in a clear way.

But this is enough to conclude that the equiprobability model over the Past State can't provide chances. For as you will recall, good just isn't good enough for a best system. We need a system that is clearly optimal, but Loewer's isn't really. The consequence is that E as applied to the whole Universe isn't part of the best system. The probabilities aren't chances.

It may be objected that we do not have any information about the value of the kinetic energy of a particular particle. But this is immaterial for our purposes. Our ignorance only implies that we are not in a position to know the initial probability distribution. But our interest is in metaphysics, not in epistemology. The question is what the unique probability distribution over initial condition is. And for fixing this distribution, human needs and limitations do not matter (as Winsberg 2008 argues forcefully).

Further, under certain assumptions, there are two probability models that do not require the evaluation of micro-information, but that are still roughly equally good. For suppose that there are two macro-states with the same minimal phase-space volume and thus with the same minimal entropy. There are the two following options for a probability model: Either we assume a flat probability model over the actual past macro-state or we assume a flat probability model over the set union of both macro-states. The first model provides better fit, but the model is more complicated because the support region is more difficult to pick: It is not sufficient to demand that entropy be minimal, rather we also have to use additional macro-information to pick the actual Past State. The second model does not fit as well as the first, but it is simpler. Again, there is no clear winner between the models. Both improve the system  $\Sigma(CM, TD)$ , but we cannot tell which model provides a better improvement.<sup>7</sup>

Frigg (2008), p. 679 (cf. also Frigg & Hoefer 2013, Sec. 5.5) seems to think that there is in fact a probability model that is clearly optimal, viz. a Dirac delta distribution peaked at the actual initial condition. So, for each state  $\mathbf{x}$ , the chance distribution would be

$$p(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_0) , \tag{4}$$

where  $\mathbf{x}_0$  is the  $(6N - 1)$ -dimensional vector specifying the actual initial condition, i.e.

---

<sup>7</sup> Alternatively, we can shrink the support of the probability density over the initial condition by taking into account macro-information about later times, e.g. by conditioning on the macro-state one second after the initial time (cf. Frigg (2008), p. 679). Once again, we gain fit, but pay in simplicity.

the positions and momentums of all particles at time  $t_0$ . Clearly, the delta distribution provides optimal fit. Frigg further thinks that the delta distribution is still simple because it has a simple functional form. Overall, he concludes, the delta distribution is optimal.

However, in my view, Frigg overlooks a lot of costs. The delta distribution has a simple functional form, but to pick one particular delta function, you have to specify the location of the peak,  $\mathbf{x}_0$ , i.e., the whole initial condition, and this is not simple at all!<sup>8</sup>

The globalist approach runs into other troubles too. So far, we have only probabilities over the initial condition. But E refers to probabilities at later times. So we need introduce a dynamics of the probabilities.

There is a very natural way to fix a dynamics. We let the chances follow the micro-dynamics. Consider the probability that the system is within region  $B \subset \Gamma$  at time  $t$ . The probability of finding the system in  $B$  then is assumed to be

$$\mu(B \cap \Gamma_t^0) / \mu(\Gamma_t^0) . \quad (5)$$

where  $\Gamma_t$  is the dynamical development of the phase-space part corresponding to the Past State.

But this dynamics does not underwrite E. In E, the probability is additionally conditioned on the macro-state in which the system is at time  $t$ . There are thus at least two alternative dynamics of the probabilities.

From a Humean perspective, chances follow the dynamics that leads to the best system. We need the dynamics that strikes an optimal balance between simplicity and fit (strength is not an issue because both alternative dynamics are defined on the same Boolean algebra). Conditionalization on the current macro-state improves fit because we narrow down the range of possible micro-states. However, we have to pay in simplicity. The formula for the probability distribution itself becomes more complicated due to the conditionalization. Is there a clear balance between simplicity and fit that decides between both probability models? I do not think so. Again, we are faced with a situation in which there is no clear-cut balance between simplicity and fit. There is no unique winner. This means that both probability distributions do not specify chances.<sup>9</sup>

To be sure, if you have to guess what the real micro-state of the system is, then you do much better if you conditionalize on the present macro-state because you then take into account information that is pertinent to the case at hand. But we are here not guessing what the present micro-state is. We are in the business of finding a dynamics of chances that produces a best system. Better guesses go along with better fit, but fit is only one item on the list of desiderata.

Loewer goes even further and suggests that we condition on the whole macro *history* of the system under consideration. The effect is that fit is further improved while we have to pay in simplicity. Further, why not condition on further characteristics?

---

<sup>8</sup> I'm here in agreement with Meacham (2010), p. 1129 and Frisch (2011), pp. 1004–1005.

<sup>9</sup> Note that our natural dynamics of chances does not really mesh with PD. To explain the increase in entropy, a modification of PD would have to be combined with the natural dynamics.

The conclusion looming large here is that Humean considerations do not provide us with any clearly optimal solution to the problem of how to define a dynamics of chances. And good just isn't good enough. We need a best system that is clearly best, and not just a good one. If there isn't a best system, then there are no chances, or so I have suggested above. This means that the probabilities introduced by the globalist strategies cannot be interpreted as chances according to Lewis.

There is a further problem with the globalist strategy to underwrite E (e.g. Callender 2011, pp. 100–102). The account justifies E as applied to the whole Universe (the isolated system being the whole Universe). But what we are to explain is not only a global increase in entropy, but rather local increase in entropy in a lot of systems, e.g. in coffee cups together with their environments. Thus, the globalist approach only makes sense if E as applied to the Universe implies the right kind of probabilities about small systems. This is a very long way to go. Entropy is additive, and that global entropy increases is no more than to say that entropy is increasing on average. This leaves space for large pockets of decreasing entropy.

### 4.3 Localism about micro-probabilities

Turn now to the localist strategy as taken by Frigg & Hoefer (2013).<sup>10</sup> This is a different attempt to improve upon the system  $\Sigma(CM, TD)$ . The crucial move here is to apply E to all isolated thermodynamic systems. Here is the precise localist formulation of E:

E<sub>1</sub> For each system within the class of isolated thermodynamic systems, for each time  $t > t_0$ , the probability that the system is in region  $B$  is

$$\frac{\mu(B \cap \Gamma_t^0 \cap \Gamma_i)}{\mu(\Gamma_t^0 \cap \Gamma_i)}. \quad (6)$$

The initial time  $t_0$ , phase space and the respective partition into macro-states are now defined for each system  $s$  separately ( $t_0$  may that time at which the system is isolated from its environment). Note that we are implicitly quantifying over phase spaces of various dimensions because the systems under consideration differ in their numbers of degrees of freedom. Further, the systems will differ in their initial states, so we have to quantify over  $\Gamma^0$  as well. Note finally that E<sub>1</sub> quantifies over instances of time and thus fixes a probability distribution for every time. There is no need to postulate a dynamics of chances on top of this; the dynamics of chances is basically given by the micro-canonical distribution under the constraint of a low initial entropy.

E<sub>1</sub> is a very clever way to systematize patterns found in the mosaic of singular events. The idea is that there are patterns over and above the deterministic laws of succession,

---

<sup>10</sup> In this section, I focus on the localist strategy based upon the work of Albert. It closely parallels the globalist strategy as outlined above. Frigg & Hoefer (2013) take the localist strategy in a different way. But they run into problems parallel to those outlined in this section. See fn. 14 below for details.

but these patterns don't emerge from the initial conditions themselves; rather they arise because certain types systems arise in history and display a certain recognizable behavior. Ultimately, this is due to the initial condition of the Universe and the dynamics, but, even granted the deterministic laws, the pattern unfolding in history can't be translated to a simple pattern in the initial condition.<sup>11</sup>

The proposal then is that  $E_1$  is a part of the best system. As Frigg & Hoefer (2013), Sec. 5.5 note, from a Humean perspective, there is something healthy about  $E_1$ .  $E_1$  claims there to be a regularity and is thus often instantiated. The probability distribution that  $E_1$  claims for every isolated system can thus be obtained using a myriad of systems of the same type. The hope would be that this suffices to obtain a unique probability distribution by accounting for the relevant patterns in the Humean way.

But do we really improve upon  $\Sigma(CM, TD)$  in an optimal way if we add  $E_1$  to it? Let us take the criteria in turn.

- Strength: We certainly gain strength because the Boolean algebra over which probabilities are defined is extended.
- Simplicity: The probability distribution over micro-states is fairly simple.
- Fit:  $E$  quantifies over phase-space dimensions and thus over probability models. To determine fit, we have to take each phase space in turn. We begin, maybe, with a phase space of  $6N$  dimensions for some  $N$  and consider all possible energies  $E$ , continue with the next of  $6(N + 1)$  dimensions and so on. For each phase space and energy  $E$ , we consider the entirety of systems that can be characterized using the hyperplane with energy  $E$  in this phase space. Depending on how large the Universe is and on the distribution of the micro-systems, we may only obtain very few systems per phase space and energy  $E$  (if any at all) or a large number.

How good will fit be? This is a good question. Nobody has yet done the exercise to check whether systems with a  $6N$ -dimensional phase and with energy  $E$  scatter homogeneously within the range defined by the macro-variables or not, and the check is a difficult exercise too. So the correct answer seems to be that we don't know (cf. Frigg & Hoefer 2013, Sec. 4 for this point).

Altogether, since we don't know what fit is like, we can't really tell whether it pays off to add  $E_1$  to the system  $\Sigma(CM, TD)$ . So we can't really tell whether the probabilities are chances.

But suppose now for the sake of argument that the equiprobability models in the phase spaces provide good fit such that  $E_1$  leads to an improvement over  $\Sigma(CM, TD)$ . Even then, we face a problem. Consider the following alternative to  $E_1$ ,

---

<sup>11</sup> The same point applies to SL probabilities.

$E'_1$  For each system from the class of isolated systems, for each time  $t > t_0$ , the probability that the system is in region  $B$  is

$$\frac{\mu(B \cap \Gamma_t^0)}{\mu(\Gamma_t^0)}. \quad (7)$$

$E'_1$  does not take into account the current macro-state and doesn't delimit the equiprobability distribution to this range. In this way, fit becomes worse, while we gain in simplicity because the distribution is now less complicated (the model has less parameters). Is it better to add  $E'_1$  or  $E_1$  to  $\Sigma(CM, TD)$ ? This is a difficult question. It doesn't seem to be a clear case in which the balance between simplicity and fit favors one side without any doubts.

Maybe, we can even go the other way round and move to  $E''_1$  that conditionalizes on additional information (be it macro or not). Fit can be improved considerably in this way, but we have to pay concerning simplicity. If we strike the balance, it is not clear which hypothesis,  $E''_1$  or  $E_1$ , leads to a better system. And, recall, good just isn't good enough. So we haven't found an optimal system, and if there isn't, then there aren't Humean chances. In conclusion, both the globalist and the localist strategies don't yield chances over micro-states.<sup>12</sup>

## 5 Discussion points

Let me now discuss the argument and the results of the paper.

A natural reaction to them is as follows. Lewis's account of chances, it may be said, does not work quite generally. The competition for the best probabilities can't ever produce a winner. It's simply not possible to weigh simplicity against fit in a unique rational manner, in statistical physics as everywhere else. Or so it may be suggested.

I don't think that this is true. Lewis's account describes quite well what scientists do when they use probabilities to describe phenomena in nature. Sometimes they do come up with unique probabilities the values of which are not much debated. In these cases, Lewis's account should work. As an example, consider electrons with a spin parallel to a certain direction, call it  $z$ . Spin is later measured in a direction perpendicular to the  $z$ -direction, call it  $y$ . We restrict our attention to systems in which no forces affect spin between the preparation of an electron and the measurement. Consider now the outcomes of all these experiments. The outcomes define what we may call an empirical distribution. As is well-known, there is no way to predict the outcome of an individual measurement on the basis of information about the experimental set-up. Consequently, the outcomes cannot be inferred from initial conditions and laws of nature.<sup>13</sup> To include

---

<sup>12</sup> To be fair to Frigg & Hoefer (2013), we should note that they explicitly bracket the question as to whether their strategy yields unique best probabilities (Sec. 4).

<sup>13</sup> Bohmians, of course, deny this, but they nevertheless grant quantum-mechanical probabilities.

information about the outcomes in a system, we may compose a large list of all the individual outcomes, but this would render the system too complicated. The only way left to cover the outcomes is thus to assume a probability distribution. We fit a simple probability model to the empirical distribution. In our examples, the model has it that an electron has a probability of 50% to have a spin of  $+1/2$  in the  $y$ -direction.

The probabilities under consideration are chances as defined by Lewis if there is no alternative probability model that does at least roughly equally well on the desiderata. Let us thus try to find an alternative model that is roughly as good as our first one. We could increase fit, if we partitioned the electrons into two broad classes that displayed markedly different empirical distributions over spin measurements in the  $y$ -direction. We would then fit different probability models to the empirical distributions from both classes. But we would have to pay in simplicity because having probability models for the several sub-classes makes things more complicated. All in all, it might prove difficult to strike a balance between simplicity and fit, and we would end up in the same sort of mess that we encountered in statistical physics.

But can we indeed pick two classes with markedly different empirical distributions of spins? In principle, we can, we may for instance pick by hand those electrons with a positive  $y$ -spin. However, in this way we would lose a lot of simplicity. The reason is that a system with two probability models over spins has to specify to which class of electrons a probability model pertains. If we pick the classes by picking individual electrons by hand, as it were, the resulting system becomes too complicated. It will clearly be worse than the system with one probability model over all spins.

But maybe we can find an alternative system that partitions the entirety of outcomes in a much simpler way and nevertheless leads to two markedly different empirical distributions. The only way to do this is to pick the classes in terms of general physical characteristics. For instance, we may say that the first class includes those electrons with kinetic energies higher than  $e_0$ , whereas the second class comprises the other electrons. The crucial problem now is that there is no such way to partition the electrons in two classes with markedly different empirical distributions. At least as far as we know, there is no physical characteristic that correlates with spin in these kinds of experiments. In whatever way we partition the class using natural predicates built up from well-known physical characteristics, we always find roughly the same empirical distribution over spins. The distribution is invariant under any natural partition (under any partition defined using well-known physical characteristics such as mass ...). There is thus no way to improve fit by paying a bit of simplicity. Subsuming all electron spins under one general statement is the best way to account for them from the Humean perspective.

Things are different in the case of isolated systems, over which we quantify in statistical physics. There are always well-known natural variables that allow us to improve fit. The reason is that our systems are built up of micro-constituents and that we can use information about them to improve fit. This leaves us with a number of systems that

are about as good as the others. And good just isn't good enough, as you will remember by now.

The example in which we do have chances according to Lewis is from quantum mechanics, according to which fundamental processes are deterministic. But what about probabilities over macro-systems? Can Lewis's account ever produce chances at this level? A worry might be that it cannot, and if it cannot, then Lewis's account may seem implausible.

But I think, it is indeed possible that Lewis's account yields chances even for macro-events. The account yields chances if a certain type of macro-event (e.g. that a coin lands heads) arises because a large number of degrees of freedom interact in a very complicated way, such that the empirical distribution over the macro-events is invariant or almost invariant under partitions formed using simple combinations of physical characteristics.

This condition may be fulfilled for thermodynamic behavior. Consider the class of thermodynamic systems with a fixed number of degrees of freedom. There are a few parameters (energy, entropy, strength of interaction, density) that systematically influence the empirical distribution over increase in entropy. But once these parameters have been identified and taken into account in a model  $P(N, E, S, \dots)$ , it seems, there are no simple ways to subdivide the class further to improve fit. Part of the reason would be that increase in entropy is multiply realizable in a great many ways. If all this is correct, then there is a unique best probability model, and we do obtain chances. The question of whether the right sort of conditions obtain for there being chances over entropy increase cannot be answered in this paper, it would require more physics.

In any case, in statistical physics, things are different. In thermodynamics, we are interested in the probability of macro-events. The macro-events are multiply realizable, so there is an open question as to whether we could improve fit by defining suitable sub-classes. By contrast, it is essential for statistical physics to specify probabilities over micro-events, and statistical physicists always assume a homogeneous probability distribution at some level. They can always improve fit by playing around with the support of the probability model. As fit becomes better, costs in simplicity arise, and it is not clear what the overall best model is.

So far, I have considered the Boltzmannian account of the Second Law. I have argued that the equiprobability models that are assumed cannot be taken to be chance distributions. But maybe the Boltzmannian account can be improved upon. Maybe, there are actually much better probabilities that do in fact form part of a clear winner among the possible systems.

Here is one idea (see e.g. Winsberg 2008, Sec. 2; cf. also Frigg & Hoefer 2013, Sec. 3). Instead of introducing a flat probability distribution in some part of phase space, we rather partition each phase space into two regions, viz. one with initial conditions that lead to the right kind of thermodynamic behavior, and the other with initial conditions that do not. We take the relative frequencies or some elaboration of them from our



pattern and identify them with the respective probabilities. These probabilities are then used to explain why thermodynamic behavior is very common.

However, this strategy gives up the project of statistical physics as it is commonly conceived of because we are only left with time-less probabilities over certain histories. Statistical physics, by contrast, seems to be committed to probabilities that a system is in a certain micro-state at a specific time. Further, the explanatory value of the probabilities would be small. It is clear that, in a deterministic universe, probabilities over certain events can be translated into probabilities over prior initial conditions. We gain not much by explaining the former in terms of the latter.<sup>14</sup>

Some people may want to suggest that the problem is with Lewis's account of chances and not with the probabilities in Boltzmannian statistical physics. Maybe, there are indeed chances, but Lewis's account fails to account for them. I don't think that this is a helpful suggestion. Lewis's account is fairly weak, and something like his requirements seem necessary for chances. The argument in this paper is that even these necessary conditions are not fulfilled.

## 6 Conclusions

In a nutshell, the argument of this paper runs as follows:

1. Statistical physics is built upon some variant of the equiprobability assumption E.
2. If the probabilities from statistical physics are best system probabilities, then E has to implement probabilistic assumptions that optimally improve upon systems with deterministic laws and thermodynamics only ( $\Sigma(CM)$  or  $\Sigma(CM, TD)$ ).
3. Under both the globalist and the localist strategy, the pertinent probabilistic assumptions do not improve upon  $\Sigma(CM, TD)$  in an optimal way because there are rivals that are roughly equally good.
4. Thus, the probabilities in statistical physics are not best system probabilities or chances.

---

<sup>14</sup> For an alternative strategy, Frigg & Hoefer (2013) do not just partition a suitable region of each phase space in two regions; rather, they assume a flat probability distribution over the initial macro-state in each phase-space. This then is supposed to imply a very high probability for trajectories that manifest thermo-dynamic behavior. But this suggestion runs into the same types of problems we have seen before. First, as Frigg & Hoefer (2013), Sec. 4 themselves acknowledge, it is not known whether a flat distribution over micro-states yields good fit. Second, there is an alternative probability model that is roughly equally good. The model distinguishes two sub-regions within the initial macro-state in each phase-space with the help of a physical characteristics. We assume a flat probability model for each of these sub-regions. In this way we improve fit, but have to pay in simplicity. The advantages of this model do not clearly balance the disadvantages and vice versa.

To conclude, let me put my argument in a broader perspective.

First, it has sometimes been argued that Lewis's account doesn't really provide ontic chances, either as a general view of probabilities or as applied to probabilities in statistical physics. Maybe, as Lewis himself fears, the notion of simplicity will ultimately have to refer to humans in one or the other way. Or, maybe, Lewis's account works only regarding statistical physics if we restrict any admissible information to macro-information, which would in the end introduce the perspective of human beings (see Winsberg 2008). My argument is not of this type. My argument may even have force if the task is only to provide an objective interpretation of probabilities in the sense of epistemic objectivity. For the argument denies that the notion of a best system suffices to fix the values of the probabilities uniquely, which would be essential for objective probabilities. In some cases, lack of knowledge or other epistemic constraints may lead to one unique probability model, for instance if we don't have any knowledge about micro-conditions. But this need not always be so.

Second, in this paper, I have only considered the approach to equilibrium. Probabilities from equiprobability models are used for other purposes too, for instance to derive relationships between state variables (pressure, temperature, entropy) in equilibrium. It may be argued that, for these purposes, E is essential and that we cannot change the probabilities in the way suggested in this paper. A possible reply would be that there are alternative probability distributions that are roughly on par with the usual ones, that produce the same macro-results (including the Second Law), but that do roughly as well as the usual ones in the best system approach. The idea would be that the support region in phase space can be narrowed down even further using other physical variables and that this doesn't affect the macro results (cf. Uffink 2011, p. 49 for a related remark). But to show this is left to future work.

Third, let me address as final worry. As suggested in the introduction, there is some pressure to defend ontic probabilities in statistical physics, and most ontic interpretations of probabilities do not mesh well with statistical physics. How then can we understand probabilities in statistical physics?

The point may be pressed even more. According to my argument, probabilities used in statistical physics don't belong to an uncontroversially best system because other probability functions would do as well. But statistical physicists do use certain probabilities and there doesn't seem to be much doubt about this. They do assume the micro-canonical distribution if they consider an isolated system. How can I account for this?

As an answer I'd like to suggest that, at least for many purposes, different probability distributions would do as well. I think, it doesn't really matter for the explanation of the approach to equilibrium how exactly we condition on macro-information that we happen to have. The reason is that arguments in statistical physics are robust, the details don't really matter (see Uffink 2011, p. 49 for a somehow similar suggestion). There may be

several probability models that strike a good balance between simplicity and fit, but they would all underwrite the approach to equilibrium. I haven't shown this rigorously, but this seems to be a fruitful perspective on this problem. The idea would be that statistical mechanics explanations are robust and that a lot of details don't matter.

Callender (2001) has rightly reminded us that we shouldn't take thermodynamics too seriously. Maybe we shouldn't take certain distributions from statistical physics too seriously, either. Put succinctly, my point then is this: Good just isn't good enough, if we consider the best systems approach, but it's good enough for statistical physics.

## Acknowledgment

I'm grateful to my commentators Luke Glynn, Radin Dardashti, Karim P. Y. Thebault and Mathias Frisch (see <http://philsci-archive.pitt.edu/9794/>), to Georg Brun and to the participants at the Lausanne workshop for discussion. Thanks also to Michael Esfeld for the invitation and to Roman Frigg for sharing a yet unpublished manuscript with me.

## References

- Albert, D., *Time and Chance*, Harvard University Press, Cambridge (MA), 2000.
- Callender, C., *Taking Thermodynamics too Seriously*, *Studies in History and Philosophy of Science Part B* **32** (2001), no. 4, 539–553.
- Callender, C., *The Past History of Molecules*, in: *Probabilities in Physics* (Beisbart, C. & Hartmann, S., eds.), Oxford University Press, Oxford, 2011, pp. 83–113.
- Elga, A., *Infinitesimal Chances and the Laws of Nature*, *Australasian Journal of Philosophy* **82** (2004), 67–76.
- Frigg, R. & Hoefer, C., *Determinism and Chance from a Humean Perspective*, in: *The Present Situation in the Philosophy of Science*. (Dieks, D., Gonzalez, W., Hartmann, S., Weber, M., Stadler, F., & Uebel, T., eds.), Springer, Berlin and New York, 2010, pp. 351–37.
- Frigg, R. & Hoefer, C., *The Best Humean System for Statistical Mechanics*, manuscript, 2013.
- Frigg, R. & Werndl, C., *Entropy – a Guide for the Perplexed*, in: *Probabilities in Physics* (Beisbart, C. & Hartmann, S., eds.), Oxford University Press, Oxford, 2011, pp. 115–42.
- Frigg, R. & Werndl, C., *Explaining Thermodynamic-like Behavior in terms of epsilon-Ergodicity*, *Philosophy of Science* **78** (2011), no. 4, 628–652.
- Frigg, R., *Chance in Boltzmannian Statistical Mechanics*, *Philosophy of Science* **75** (2008), no. 5, 670–681.
- Frisch, M., *From Arbuthnot to Boltzmann: The Past Hypothesis, the Best System, and the Special Sciences*, *Philosophy of Science* **78** (2011), no. 5, 1001–1011.
- Hájek, A., 'Mises Redux'– *Redux. Fifteen Arguments Against Finite Frequentism*, *Erkenntnis* **45** (1997), 209–227.
- Hájek, A., *Interpretations of Probability*, in: *The Stanford Encyclopedia of Philosophy* (Zalta,

- E. N., ed.), spring 2010 ed., 2010,  
<http://plato.stanford.edu/archives/spr2010/entries/probability-interpret/>.
- Hofer, C., *The Third Way on Objective Probability: A Sceptic's Guide to Objective Chance*, *Mind* **116** (2007), 549–596.
- Lavis, D. A., *An Objectivist Account of Probabilities in Statistical Mechanics*, in: *Probabilities in Physics* (Beisbart, C. & Hartmann, S., eds.), Oxford University Press, Oxford, 2011, pp. 51–81.
- Lewis, D., *A Subjectivist's Guide to Objective Chance*, in: *Studies in Inductive Logic and Probability. Vol II* (Jeffrey, R. C., ed.), University of California Press, Berkeley, 1980, here quoted from the reprint in Lewis (1986a), pp. 84–113.
- Lewis, D., *Philosophical Papers. Vol. II*, Oxford University Press, New York, 1986.
- Lewis, D., *Postscripts to "A Subjectivist's Guide to Objective Chance"*, in: Lewis (1986a), pp. 114–132.
- Lewis, D., *Humean Supervenience Debugged*, *Mind* **103** (1994), 473–490, reprinted in Lewis, D. (1999), *Papers in Metaphysics and Epistemology*, Cambridge University Press, Cambridge.
- Loewer, B., *Determinism and Chance*, *Studies in History and Philosophy of Modern Physics* **32** (2001), 609–620.
- Loewer, B., *David Lewis's Humean Theory of Objective Chance*, *Philosophy of Science* **71** (2004), 1115–1125.
- Meacham, C. J. G., *Contemporary Approaches to Statistical Mechanical Probabilities: A Critical Commentary – Part ii: The Regularity Approach*, *Philosophy Compass* **5** (2010), no. 12, 1127–1136.
- Uffink, J., *Subjective Probability and Statistical Physics*, in: *Probabilities in Physics* (Beisbart, C. & Hartmann, S., eds.), Oxford University Press, Oxford, 2011, pp. 25–50.
- Winsberg, E., *Laws and Chances in Statistical Mechanics*, *Studies in History and Philosophy of Science Part B* **39** (2008), no. 4, 872–888.