

# Typicality and Notions of Probability in Physics\*

Sheldon Goldstein

Departments of Mathematics, Physics, and Philosophy – Hill Center  
Rutgers, The State University of New Jersey  
110 Frelinghuysen Road  
Piscataway, NJ 08854-8019, USA  
E-mail: oldstein@math.rutgers.edu

January 20, 2011

## 1 Introduction

A variety of notions of probability, playing different roles, are relevant in physics. One crucial notion, typicality, while not genuinely probabilistic at all, is arguably the mother of them all.

There are lots of different words for probability. Here are some: chance, likelihood, distribution, measure. There are also a variety of different notions of probability:

- Subjective chance (Bayesian?)
- Objective chance (propensity?)
- Relative frequency, empirical (pattern)
- A mathematical structure providing a measure of the size of sets (Kolmogorov)

Sometimes these are presented as competing notions. That's not my intention here. I wish only to emphasize at this point that when one speaks of probability it is a good idea to be clear about which notion one has in mind.

---

\*I am very pleased to dedicate this article to the memory of Itamar Pitowsky. Itamar was an exceptionally creative philosopher and scientist, and one with a sophisticated understanding of mathematics. In recent years he had become very interested in typicality, a subject on which he was working at the time of his death.

My main concern in this paper, however, is with typicality, a notion that, while extremely important for understanding probability, is not really a notion of probability at all. The logic of typicality is this. Many important phenomena, in physics and beyond, while they cannot be shown to hold without exception, can be shown to hold with very rare exception, suitably understood. Such phenomena are said to hold typically; a proof that they do so is a typicality proof.

Regarded as mathematics, such results can be very interesting, with prizes awarded for their achievement. Of course the practical relevance of such results is that if some observed behavior has been shown to hold with rare exception, one should not be surprised if no exceptions are seen and one will tend to feel justified in regarding the behavior as explained.

It must be admitted, however, that as a matter of logic such practical conclusions don't follow. If exceptions exist there is nothing that would preclude the exceptional cases from being the only cases we ever encounter. Nonetheless, science could make little if any progress without invoking appeals to typicality, at least implicitly.

Here is an important example of a typicality statement:

One should not forget that the Maxwell distribution is not a state in which each molecule has a definite position and velocity, and which is thereby attained when the position and velocity of each molecule approach these definite values asymptotically. . . . It is in no way a special singular distribution which is to be contrasted to infinitely many more non-Maxwellian distributions; rather it is characterized by the fact that by far the largest number of possible velocity distributions have the characteristic properties of the Maxwell distribution, and compared to these there are only a relatively small number of possible distributions that deviate significantly from Maxwell's. Whereas Zermelo says that the number of states that finally lead to the Maxwellian state is small compared to all possible states, I assert on the contrary that by far the largest number of possible states are "Maxwellian" and that the number that deviate from the Maxwellian state is vanishingly small. (Ludwig Boltzmann, 1896 [1])

Notice that this statement of Boltzmann involves probability ("distribution") and typicality. Boltzmann is saying here that states with Maxwellian probabilities are typical ("by far the largest number of possible states are 'Maxwellian' . . . the number that deviate from the Maxwellian state is vanishingly small"). This illustrates an important source of confusion in this business: that typicality statements often concern probabilities, making it all too easy to conflate typicality and probability.

## 2 History

There has been a revival of interest in typicality among physicists and philosophers in recent years. However the recognition of the importance of the notion is not new. That goes back to the very beginnings of probability theory in the 18th century. What I shall describe here of the relevant ancient history I've learned from Glenn Shafer [2, 3]. Notice in what follows how some of the founding fathers of probability theory struggled to finesse the gap between an event having extremely small size as measured in some natural way and the event being impossible, or certain to fail.

### 2.1 Ancient History (< 1950)

- Jakob Bernoulli, in his great work *Ars Conjectandi* (1713), writes that “Because it is only rarely possible to obtain full certainty, necessity and custom demand that what is merely morally certain be taken as certain.”
- Antoine Cournot (1843) writes that “A physically impossible event is one whose probability is infinitely small. This remark alone gives substance—an objective and phenomenological value—to the mathematical theory of probability.” This later became known as *Cournot's principle*.
- According to Paul Levy ( $\approx 1919$ ), Cournot's principle is the only connection between probability and the empirical world. He calls it “the principle of the very unlikely event.”
- Hadamard refers instead to “the principle of the negligible event.”
- Kolmogorov, in his *Foundations of Probability* (1933), Chapter 1, §2, *The Relation to Experimental Data*, writes that “Only Cournot's principle connects the mathematical formalism with the real world.”
- Similarly Borel ( $\approx 1948$ ) writes that “The principle that an event with very small probability will not happen is the only law of chance.”

### 2.2 Modern History (> 1950)

Notice that while the probabilists did not refer to “typical” or “typicality,” that notion, or something very much in its vicinity, is what they had in mind. In more recent years the “t”-word has been used quite frequently, most often, curiously, in connection with probability in quantum mechanics. I hope the following quotations help convey the idea of the method of appeal to typicality.

In order to establish quantitative results, we must put some sort of measure (weighting) on the elements of a final superposition. This is necessary to be able to make assertions which hold for almost all of the observer states described by elements of the superposition. We wish to make quantitative statements about the relative frequencies of the different possible results of observation—which are recorded in the memory—for a typical observer state; but to accomplish this we must have a method for selecting a typical element from a superposition of orthogonal states. . . .

The situation here is fully analogous to that of classical statistical mechanics, where one puts a measure on trajectories of systems in the phase space by placing a measure on the phase space itself, and then making assertions . . . which hold for “almost all” trajectories. This notion of “almost all” depends here also upon the choice of measure, which is in this case taken to be the Lebesgue measure on the phase space. . . . Nevertheless the choice of Lebesgue measure on the phase space can be justified by the fact that it is the only choice for which the “conservation of probability” holds, (Liouville’s theorem) and hence the only choice which makes possible any reasonable statistical deductions at all. (Hugh Everett, III, 1957 [4, page 460])

Then for instantaneous macroscopic configurations the pilot-wave theory gives the same distribution as the orthodox theory, insofar as the latter is unambiguous. However, this question arises: what is the good of *either* theory, giving distributions over a hypothetical ensemble (of worlds!) when we have only one world.

. . . a single configuration of the world will show statistical distributions over its different parts. Suppose, for example, this world contains an actual ensemble of similar experimental set-ups. . . . it follows from the theory that the ‘typical’ world will approximately realize quantum mechanical distributions over such approximately independent components. The role of the hypothetical ensemble is precisely to permit definition of the word ‘typical.’ (John S. Bell, 1981 [5, page 129])

### 3 Typicality in Statistical Mechanics

If there is a branch of physics in which typicality is most prominently used it is probably statistical mechanics. And the most famous use of typicality in statistical mechanics concerns Boltzmann’s equation. Moreover one could scarcely have a better illustration of the point of and the need for a typicality argument than in the transition from Boltzmann’s presentation of 1872 to

his presentation in 1877. Boltzmann (1872) claimed that (at low density) the state of a gas *must* evolve in accord with his equation. Boltzmann (1877) claimed, in effect, only that it would typically do so. Here are some details.

Boltzmann’s equation is an evolution equation for a function  $f(\mathbf{q}, \mathbf{v}, t)$ , where  $\mathbf{q}$  is a point in physical space,  $\mathbf{v}$  is a velocity, and  $t$  of course is time. Boltzmann analyzed the behavior of a certain function of  $\mathbf{q}$  and  $\mathbf{v}$  that provides an efficient summary of the most important details of the state of a gas, namely the empirical one-particle distribution  $\rho_{emp}(\mathbf{q}, \mathbf{v}) \equiv f_X(\mathbf{q}, \mathbf{v})$ , giving basically the density of particles of the gas that are at or near  $\mathbf{q}$  with velocity more or less  $\mathbf{v}$ .

Here, for an  $N$ -particle system,  $X = (\mathbf{q}_1, \mathbf{v}_1, \dots, \mathbf{q}_N, \mathbf{v}_N)$  is the point in the  $N$ -particle phase space describing the detailed state of the gas. The subscripts “emp” and  $X$  on  $\rho$  and  $f$  are to emphasize that  $f_X$  is indeed an empirical distribution, determined by the phase point  $X$ , and not a probability distribution that describes a random system or a hypothetical ensemble of systems. As the phase point  $X(t)$  evolves according to the Hamiltonian dynamics for the system,  $\rho_{emp}$  evolves accordingly:  $\rho_{emp}(\mathbf{q}, \mathbf{v}, t) \equiv f_{X(t)}(\mathbf{q}, \mathbf{v})$ .

What Boltzmann claimed to have shown in 1872 is that for a low density gas it must be the case that  $f_{X(t)}(\mathbf{q}, \mathbf{v})$  is well approximated by a solution  $f(\mathbf{q}, \mathbf{v}, t)$  to Boltzmann’s equation. On the basis of an analysis of that equation using his  $H$ -function

$$H(f(\mathbf{q}, \mathbf{v}, t)) = \int f(\mathbf{q}, \mathbf{v}, t) \log f(\mathbf{q}, \mathbf{v}, t) d\mathbf{q}d\mathbf{v}.$$

Boltzmann then argued that for large times  $t$ ,  $f(\mathbf{q}, \mathbf{v}, t)$ —and hence also  $f_{X(t)}(\mathbf{q}, \mathbf{v})$ —will approach the distribution that minimizes  $H$ , namely the equilibrium distribution—the Maxwellian distribution— $f_{eq}(\mathbf{q}, \mathbf{v}) \propto e^{-\frac{1}{2}m\mathbf{v}^2/kT}$ , where  $k$  is Boltzmann’s constant and  $T$  is the temperature of the gas.

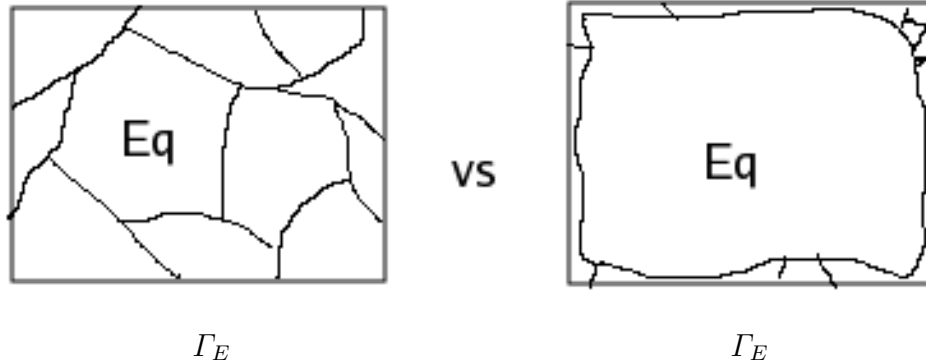
Because of Loschmidt’s reversibility objection, by 1877 Boltzmann had realized that his earlier claim could not be right. He concluded that he had shown, not that  $f_{X(t)}(\mathbf{q}, \mathbf{v})$  is, approximately, a solution to Boltzmann’s equation for all initial phase points  $X(0)$ , but only for most of them. More precisely, he concluded that he had shown that given any distribution function  $f(\mathbf{q}, \mathbf{v})$ , even one that is non-Maxwellian and that does not correspond to equilibrium,  $f_{X(t)}$  will approximate a solution to Boltzmann’s equation for the overwhelming majority, suitably understood, of initial phase points for which  $f_{X(0)}$  is (approximately)  $f$ —the overwhelming majority of phase points in the macrostate defined by  $f$ . In other words, in 1877 Boltzmann argued that the evolution of a gas in accord with Boltzmann’s equation, while not inevitable, is typical. (Boltzmann’s proof was not rigorous. Almost a century later, a rigorous typicality proof, valid only for short times, was found by Oscar Lanford [6].)

More important for our understanding of the origin of thermodynamics, in 1877 Boltzmann arrived at a far deeper appreciation of why a gas will

tend to approach a state of equilibrium, in which nothing seems to change. Crucial to this understanding is the notion of macrostate, alluded to above. The macrostate  $\Gamma_f = \{X \in \Gamma_E \mid f_X(\mathbf{q}, \mathbf{v}) \approx f(\mathbf{q}, \mathbf{v})\}$  corresponding to  $f$  is the set of phase points, in the energy surface  $\Gamma_E$  of phase points having energy  $E$ , that are all macroscopically like  $f$ —in the sense that the macro-variable  $f_X$  is approximately  $f$ . The phase points in the same macrostate are thus very similar from a macroscopic perspective.

The most important fact about these macrostates, recognized by Boltzmann, concerns their sizes as measured using the natural volume measure on the phase space, Lebesgue or Liouville measure. It is, in fact, this natural volume measure that provides a sufficiently precise notion of “overwhelming majority” for his typicality claim.

Here are two depictions of the partition of  $\Gamma_E$  into macrostates (corresponding to different choices of  $f$ ):



One special macrostate is singled out here by “Eq,” indicating the equilibrium macrostate  $\Gamma_{f_{eq}}$ , which is larger than all the others. The crucial fact is that the depiction on the left is utterly misleading, giving a very wrong sense of the relative sizes of the macrostates.

The depiction on the right is much better. But in fact the equilibrium macrostate is so very much larger than the other macrostates that no picture could adequately depict the difference in sizes. In fact, as Boltzmann showed, at low density  $|\Gamma_f| \sim e^{-NH(f)}$ . For a macroscopic system, with particle number  $N \sim 10^{20}$  or greater, this means that the overwhelming majority of the points of  $\Gamma_E$  are in  $\Gamma_{f_{eq}}$ , the ratio of the size of a non-equilibrium macrostate to that of the equilibrium macrostate being ridiculously small, of order  $10^{-10^{20}}$ .

The depiction on the right illustrates another one of Boltzmann’s typicality results—that equilibrium is typical: the overwhelming majority of the phase points  $X$  in the energy surface  $\Gamma_E$  correspond to a gas having equilibrium properties, in the sense that  $f_X$  is approximately  $f_{eq}$ . This typicality result, which is easy, should not be confused with the typicality result for Boltzmann’s equation, which is very hard. For the former, “overwhelming

majority” is relative to the entire energy surface, while for the latter it is mainly relative to the incredibly small non-equilibrium macrostates.

Be that as it may, the estimates associated with the depiction do in fact provide us with a good rule of thumb for the relative size of atypical events of all sorts in statistical mechanics: corresponding to the ratio  $10^{-10^{20}}$ . You should at least not be surprised when events corresponding to sets of possibilities that are so small aren’t observed.

Besides the two typicality results that I’ve mentioned there are many others in statistical mechanics—either proven already or awaiting a rigorous proof. Some examples are the second law of thermodynamics, the derivation of hydrodynamic equations, approach to equilibrium in quantum mechanics, and the universality of the canonical ensemble in quantum mechanics (canonical typicality). And outside of statistical mechanics there is, for example, the origin of quantum randomness in Bohmian mechanics, to which I now turn.

## 4 Bohmian Mechanics

Bohmian mechanics [7–10] is arguably the simplest formulation of non-relativistic quantum mechanics. It concerns the dynamics of a system of particles, with positions  $\mathbf{Q}_1, \dots, \mathbf{Q}_N$ , defining a configuration  $Q$ . This dynamics is determined by the usual quantum mechanical wave function  $\psi$ , itself evolving, as in standard quantum mechanics, according to Schrödinger’s equation. In the simplest case, of particles without spin,  $\psi$  is a function  $\psi(\mathbf{q}_1, \dots, \mathbf{q}_N)$  of the possible positions of the particles. The joint evolution of  $\psi$  and  $Q$  is deterministic. Nonetheless, as a consequence of a typicality analysis, the usual quantum probabilities, given by  $|\psi(q)|^2$ , govern the results of observations in a Bohmian universe.

*Quantum equilibrium*, corresponding to the quantum equilibrium distribution  $\rho_{qe}(q) = |\psi(q)|^2$ , should be thought of, in this regard, as roughly analogous to thermodynamic equilibrium, corresponding to the Maxwellian  $f_{eq} \propto e^{-\frac{1}{2}mv^2/kT}$ . A proper understanding of quantum equilibrium probabilities and of thermodynamic equilibrium probabilities both require that we appreciate that there are a variety of conceptually different probabilistic objects relevant to the analysis, as I shall explain later. They also require that we appreciate that there are, in both cases, two different sorts of systems to be dealt with: a large system, for thermodynamics a gas in a box, and for Bohmian mechanics the entire universe; and a small subsystem of the large system, which in both cases we will take here, for simplicity, to be a single particle.

## 4.1 The Wave Function of a Subsystem

Consider a 1-particle subsystem of an  $N$ -particle Bohmian universe. Let's denote by  $\mathbf{Q}$  the position of the particle and by  $Q_{env}$  the configuration of the rest of the particles of our universe—the configuration of the *environment*. Let  $\Psi$  be the wave function of the universe. It is a function  $\Psi(\mathbf{q}, q_{env})$ . (Here we've used, as is common in Bohmian mechanics, lower case letters to indicate possible values, generic values, as opposed to the actual values, denoted with capital letters.)

The appropriate notion for the wave function of our subsystem is given by the *conditional wave function*

$$\psi(\mathbf{q}) = \Psi(\mathbf{q}, Q_{env}),$$

a function of the generic position of our particle obtained by plugging the actual configuration of its environment into the wave function of the universe. Note that  $\psi$  need not be normalized—its absolute square integral over all of space need not be 1. Whenever  $\psi$  appears as part of a probability formula it should be regarded as having been normalized via multiplication by the appropriate positive real number. Note also that because of the dependence on the actual configuration of the environment, which inherits its own typically complicated evolution from the Bohmian evolution of the configuration of the universe, the wave function of our particle depends on time in a somewhat complicated way:

$$\psi_t(\mathbf{q}) = \Psi_t(\mathbf{q}, Q_{env}(t)),$$

with  $\Psi$  itself, as a solution of Schrödinger's equation, depending on  $t$ . As a consequence of this evolution, the wave function of a subsystem in Bohmian mechanics can evolve in a variety of ways. In particular it will evolve according to Schrödinger's equation when the system is suitably decoupled from its environment, and will undergo collapse of the wave packet in the appropriate measurement situations.

For our purposes here, the most important fact about the conditional wave function is that it provides us with the probability distribution of our subsystem—in fact in a variety of senses. The most basic sense in which it does so is expressed in the following simple mathematical fact:

$$P(\mathbf{Q}(t) \in d\mathbf{q} | Q_{env}(t)) = |\psi_t(\mathbf{q})|^2 d\mathbf{q} \tag{1}$$

Here  $P$  is the probability distribution on initial configurations of the universe (at a time, say, shortly after the big bang) given by  $|\Psi(q)|^2$ . This *fundamental conditional probability formula* of Bohmian mechanics says that for such a random universe the conditional distribution of the position of a particle at any time, given its environment at that time, depends only on its conditional wave function at that time, and does so via the usual Born, quantum equilibrium, probability formula.



As a consequence of this formula it follows via a typicality analysis [10] that for the overwhelming majority—in the sense of the measure  $P$ —of initial configurations of a Bohmian universe, the empirical distribution for the positions of particles (and for larger subsystems) in suitable real world ensembles of systems having given conditional wave function  $\psi$  is (approximately) the quantum equilibrium distribution  $|\psi|^2$ . In short, quantum equilibrium is typical.

## 5 Probability and Typicality

In a typical typicality analysis in physics—and arguably in any serious application of probability to the real world—probability structures play several quite different roles, the most important of which are the following:

- $\rho_{emp}$ : empirical distribution (relative frequency)
- $\rho_{th}$ : theoretical distribution (idealization,  $N \rightarrow \infty$ )
- $P$ : measure for typicality

It is the empirical distribution that describes a real world pattern of events that is responsible for what we observe. The theoretical distribution is an idealization providing a good approximation to the empirical distribution,  $\rho_{emp} \approx \rho_{th}$ , in the limit of large ensembles of subsystems.  $P$  is a probability distribution on the big system containing the subsystems. It is via a law of large numbers kind of analysis using  $P$  that one can show that it typically happens that  $\rho_{emp} \approx \rho_{th}$ , with typicality defined in terms of  $P$ .

Many different probability distributions  $P$  define the same sense of typicality. This is because, insofar as typicality is concerned, the detailed probability of a set is not relevant; all that matters is which sets have very large measure and which very small. Nonetheless, it is often the case that a particular choice of  $P$  is special. It is for such a choice that one in fact can most efficiently carry out the relevant analysis. For this special  $P$  the theoretical distribution will be in some sense a conspicuous part of  $P$ , meaning that:

$$\rho_{th}(\mathbf{x})d\mathbf{x} = \rho^P(\mathbf{x})d\mathbf{x} = "P(\mathbf{X} \in d\mathbf{x})" .$$

Here I use  $\mathbf{X}$  and  $\mathbf{x}$  for the subsystem variables. (I shall use  $X$  and  $x$ —without bold—to refer to the variables for the big system.)  $\rho^P$  is a sort of marginal distribution of the subsystem, arising from the distribution  $P$  of the big system, and the quotation marks are to indicate that  $\rho^P$  is often only “sort of” a marginal, and not always an actual marginal. For example the relevant  $\rho^P$  in Bohmian mechanics is a conditional marginal, just as is suggested by the fundamental conditional probability formula (1).

For the second typicality result in statistical mechanics mentioned above, that “equilibrium is typical,”  $P(dx)$  is the microcanonical ensemble on  $\Gamma_E$ , the uniform distribution over the energy surface. However, to illustrate the points I wish to make here it would be better to make a different but physically equivalent choice for  $P$ , namely the canonical ensemble, given by

$$P(dx) \propto e^{-H(x)/kT} dx .$$

And we shall assume we are dealing with the simplest case, that of non-interacting particles, with  $H = \sum_i \frac{1}{2} m \mathbf{v}_i^2$ . In this case  $P = \prod_i f_{eq}(\mathbf{v}_i)$  is simply the product over all the particles of the equilibrium distribution for each particle.

Then [with  $\mathbf{x} = (\mathbf{q}, \mathbf{v})$ ] we have that

$$\rho_{emp}(\mathbf{x}) \equiv \rho_{emp}^{(X)}(\mathbf{x}) = f_X(\mathbf{q}, \mathbf{v})$$

(the precise definition of  $\rho_{emp}$  is  $\rho_{emp}^{(X)}(\mathbf{q}, \mathbf{v}) = \frac{1}{N} \sum_{(\mathbf{q}_i, \mathbf{v}_i) \in X} \delta(\mathbf{q} - \mathbf{q}_i) \delta(\mathbf{v} - \mathbf{v}_i)$ ) and that

$$\rho_{th}(\mathbf{x}) \equiv f_{eq}(\mathbf{x}) = \rho^P(\mathbf{x}) \propto e^{-\frac{1}{2} m \mathbf{v}^2 / kT} .$$

In particular the theoretical distribution here is a factor of  $P$ —a piece, as it were, of the measure for typicality.

In Bohmian mechanics (writing  $x$  for  $q$ , etc.) we have that

$$P(dx) = |\Psi(x)|^2 dq$$

$$\rho_{emp}(\mathbf{x}) \equiv \rho_{emp}^X(\mathbf{x}) = \frac{1}{N} \sum_{\mathbf{x}_i \in X} \delta(\mathbf{x} - \mathbf{x}_i)$$

and

$$\rho_{th}(\mathbf{x}) \equiv \rho_{qe}(\mathbf{x}) = \rho^P(\mathbf{x}) = |\psi(\mathbf{x})|^2 .$$

Here too the theoretical distribution is sort of a piece of the measure for typicality.

A typicality analysis binds tightly together these three very different probabilistic objects. This is particularly so for the special choice of  $P$ , a choice for which  $P$  has some nice properties—more on these shortly—the simplest such being that  $P$  be a product measure as above. When  $P$  is thus “nice” one can show via a law of large numbers type analysis that (when  $N$  is large)  $\rho_{emp} \approx \rho^P$ ,  $P$ -typically—that for the  $P$ -overwhelming majority of points  $X$ ,  $\rho_{emp}^X$  is approximately the theoretical distribution, with the latter itself being a quasi-marginal of the measure for typicality  $P$ . We shall say in this situation that  $P$  is *statistically transparent*.

## 5.1 Ergodicity and Statistical Transparency

Statistical transparency is closely connected to the notion of ergodicity [11]. That is because it is often the case that  $\rho_{emp}$  is more or less a time-average or space-average—the sorts of things with which ergodicity is concerned. The ergodicity of  $P$  (under either space or time translation) implies that these averages agree with the phase averages, i.e., with the theoretical distribution  $\rho^P$  arising from  $P$ . Thus we can more or less identify statistical transparency with the ergodicity of  $P$ .

This does, however, have to be taken with a grain of salt, since the space averages relevant to ergodicity would be infinite system averages (so  $N = \infty$ ) or for time averages, infinite time averages ( $T = \infty$ ), idealizations that might not exactly match the typicality analysis under consideration. We shall however ignore this point, abusing mathematics a bit, and simply pretend without qualification that statistical transparency can be identified with ergodicity—that it is the ergodicity of  $P$  that makes it special, so that we have statistical transparency.

## 5.2 Symmetry and Statistical Transparency

There is another way in which  $P$  might be special: among all measures defining the same sense of typicality, it might be one that is symmetric, and the only one that is.

The relevant symmetry here depends upon whether  $\rho_{emp}$  involves space or time averages. In the former case the symmetry is that of spatial-translation invariance (remember we are pretending that our system is suitably idealized, and thus spatially infinite if necessary), in the latter case that of time-translation invariance. Suppose  $P$  is in this sense symmetric. The set of measures  $\tilde{P}$  defining the same sense of typicality as  $P$  are those of the form  $\tilde{P}(dx) = g(x)P(dx)$  obtained from  $P$  by multiplying it by a positive function  $g$  that is bounded above and away from zero below. And if, as we are pretending,  $\rho_{emp}$  involves infinite space or time averages, the set of probability measures equivalent to  $P$  in the sense of typicality could now be taken to correspond to the requirement that  $g$  be positive, with integral with respect to  $P$  equal to 1—i.e., to the set of probability measures equivalent to  $P$  in the sense of measure theory.

The connection between symmetry and statistical transparency is then this:  $P$  is the only symmetric probability measure in the class of equivalent ones precisely in case there is statistical transparency. That's because of the connection between ergodicity and statistical transparency just discussed together with the fact that a system is ergodic precisely in the case of there being a unique symmetrical  $P$  in the equivalence class.

### 5.3 Predictive Typicality and Ergodicity

The discussion has so far taken for granted that we have unambiguous (space or time) averages, i.e., that  $\rho_{emp}$  is unambiguous. This *predictive typicality*, we should remark, is a property of the typicality class itself, and not a characterization of a special member of that class. Predictive typicality is more or less equivalent to the requirement that the typicality equivalence class has an ergodic member  $P$ . This will of course be the member of the class that most directly expresses the observed probabilities  $\rho_{emp}$ .

### 5.4 The Good, the Bad, and the Ugly

To summarize, the three probability measures,  $\rho_{emp}$ ,  $\rho_{th} = \rho^P$ , and  $P$ , involved in the usual typicality analysis are intimately related: We have that

$$\rho_{emp} \leftrightarrow \rho^P \leftrightarrow P,$$

conveying that the observed probability distribution  $\rho_{emp}$ , which of course varies from trial to trial of the same experiment, is well approximated by the theoretical distribution  $\rho^P$  (which of course is the same for all trials of an experiment), the latter being a conspicuous part of the measure for typicality  $P$ .

This is both good and bad. It is good, because it suggests a nice simplicity, inasmuch as it means that for many practical purposes one need worry about just one probability measure and not three. At the same time it is bad, because the simplicity is a misleading simplicity, since the three probability measures are conceptually of entirely different natures, despite their closeness for practical purposes. And the consequences of the conflation of three such very different notions—the discussions and analyses in which crucial distinctions between very different objects are not properly recognized—can be quite ugly.

The confusion is probably greater still with regard to the typicality analysis for Bohmian mechanics, which can be summarized like so:

$$\rho_{emp} \leftrightarrow \rho_{th}^{\psi} \leftrightarrow P^{\Psi}$$

Here  $\rho_{th}^{\psi} = |\psi|^2$  and  $P^{\Psi} = |\Psi|^2$  are given by the very same formula, with the only difference being  $\psi$  versus  $\Psi$ . Unless one appreciates the great difference between the wave function  $\psi$  of a subsystem and that of the universe  $\Psi$ , this can make it difficult to accept that the probabilistic objects involved are so very different.

## 6 Two Directions for Typicality Research

I have discussed here the method of appeal to typicality and given some examples. I have indicated that, while typicality is fundamentally not a

version of probability, it can nonetheless easily seem to be one. But we have not attempted here to justify the conclusions that scientists arrive at by appealing to typicality. In particular I have not explained why what is typical should be expected to happen.

Nor shall I do so here: a systematic analysis would require that we deal with some of the most fundamental issues in the philosophy of science, such as the meaning and nature of scientific explanation. I do feel, however, that a comprehensive philosophical analysis of scientific explanation and the logic of appeal to typicality would be most welcome. (Some gestures in this direction can be found in section 6 of [12].)

## 6.1 Types of Typicality

I would, however, like to mention here three distinctions between types of typicality that are relevant to how strongly typicality seems to compel our expectations: (i) natural versus axiomatic, (ii) continuum versus finite, and (iii) hypothetical versus actual. The measure of typicality might be natural, like the uniform distribution over the space of possibilities, naturally expressed; or it might be merely stipulated axiomatically. The set of possibilities might be finite, or it might be a continuum (it of course might also be infinite but not a continuum). The possibilities might be merely possibilities—they might be hypothetical—or they might, as with many-worlds, be all actual.

Other things being equal, typicality corresponding to the first type of each pair seems to more strongly compel our expectations. For example, the notion of most elements of a finite set seems entirely unambiguous, corresponding to counting measure, whereas with a continuum one might be able to argue that there are a variety of reasonable senses of most. The worst case in this regard is that of axiomatic typicality with a finite set of actualities. (It might well be that the only way typicality can be persuasively applied to the case in which the possibilities are in fact actual is within the Humean approach to law and probability advocated by Barry Loewer [13]; however, the goal in this approach is more modest: description rather than explanation.)

## 6.2 Typicality not Given by Probability

I will conclude by putting on the table a possibility afforded by the recognition that typicality is not probability. While it is usually the case that typicality is defined using a probability measure, a different way of deciding which sets are large and which small, for example one that is given by a set function that violates the axioms of probability, is feasible. Such a wider notion of typicality could be used for the formulation of new types of physical theories.

Along such lines not much has yet been done. But Murray Gell-Mann and James Hartle [14] have noted with regard to their decoherent histories

version of quantum mechanics that insofar as their decoherence functional fundamentally is used to define, in effect, typicality (though they don't use that word) the fact that it may end up violating the axioms of probability in a limited sort of way need not concern us. And Bruno Galvan [15] has proposed a trajectory based version of quantum mechanics that, unlike Bohmian mechanics, is defined solely in terms of a typicality that is not based on probability. While Galvan's theory is a bit odd, it does have the virtue of seeming to exploit only traditional quantum mechanical structure.

The possibility of a typicality liberated from probability might be a great source of inspiration for theory formation. I think that this possibility would have pleased Itamar very much.

## 7 Acknowledgements

This work was supported in part by NSF Grant DMS-0504504.

## References

- [1] L. Boltzmann: *Annalen der Physik* **57**, 773 (1896); reprinted and translated as Chapter 8 in S.G. Brush: *Kinetic Theory* (Pergamon, Oxford, 1966)
- [2] G. Shafer: Why did Cournot's principle disappear?. *Talk at Ecole des Hautes Etudes en Sciences Sociales, Paris. May 19, 2006*, <http://www.glennshafer.com/assets/downloads/disappear.pdf>
- [3] G. Shafer and V. Vovk: The Sources of Kolmogorov's Grundbegriffe. *Statistical Science* **21**, 70–98 (2006)
- [4] H. Everett, III: "Relative State" Formulation of Quantum Mechanics. *Reviews of Modern Physics* **29**, 454–462 (1957)
- [5] J.S. Bell: *Speakable and Unsayable in Quantum Mechanics*, (Cambridge University Press, Cambridge, 1987).
- [6] O. E. Lanford III: Time Evolution of Large Classical Systems. In J. Moser (ed.): *Dynamical Systems, Theory and Applications*, Lecture Notes in Physics 38, 1–111, (Springer, Berlin, 1975)
- [7] S. Goldstein: Bohmian Mechanics. In E. N. Zalta (ed.): *Stanford Encyclopedia of Philosophy*, published online by Stanford University (2001), <http://plato.stanford.edu/entries/qm-bohm/>
- [8] D. Bohm: *Physical Review* **85**, 166 (1952).

- [9] D. Bohm: *Physical Review* **85**, 180 (1952).
- [10] D. Dürr, S. Goldstein and N. Zanghì: *Journal of Statistical Physics* **67**, 843 (1992), <http://arxiv.org/abs/quant-ph/0308039>
- [11] V. I. Arnold and A. Avez: *Ergodic Problems of Classical Mechanics*, (Addison-Wesley, Reading, 1989)
- [12] S. Goldstein, J. L. Lebowitz, R. Tumulka, N. Zanghì: Long-Time Behavior of Macroscopic Quantum Systems: Commentary Accompanying the English Translation of John von Neumann’s 1929 Article on the Quantum Ergodic Theorem. *The European Physical Journal H: Historical Perspectives on Contemporary Physics* **35**, 173–200 (2010), <http://arxiv.org/abs/1003.2129v1>
- [13] B. Loewer: David Lewis’ Theory of Objective Chance. *Philosophy of Science* **71**, 1115–1125 (2004)
- [14] M. Gell-Mann and J. B. Hartle: Quantum mechanics in the light of quantum cosmology. In W. Zurek (ed.): *Complexity, Entropy, and the Physics of Information*, (Addison-Wesley, Reading, 1990)
- [15] B. Galvan: Typicality vs. Probability in Trajectory-Based Formulations of Quantum Mechanics. *Foundations of Physics* **37**, 1540–1562 (2007)