David Mwakima

On the Quality of Perrin's Evidence

Draft of April 17, 2024

1 Introduction

Evidence occupies a central role in philosophy. Within philosophy of science specifically, evidence, and its quality, constitutes the basis on which the debate between realists and anti-realists as understood *today* is, or should be, adjudicated.[1] Within this debate, an informal notion of the "quality of evidence" from Jean Perrin's experimental work on Brownian motion has for decades been offered as an explanation of the shift in epistemic attitudes toward the atomic hypothesis among prominent scientists at the turn of the 20th century.[2] The shift in attitude was from viewing the atomic hypothesis as a merely instrumentally useful hypothesis (anti-realism) to viewing it as a well-established theory (realism). Assuming that the shift in epistemic attitudes was caused by the distinctive quality of Perrin's evidence, one goal for philosophers of science has been to characterize what makes evidence, such as Perrin's, good.

It is helpful to think of the various accounts that have been offered for evaluating evidence as falling along a spectrum. On one end of this spectrum, there are informal accounts provided by Quine (1976) and Maddy (2007, 398f, 406). On the other end, there are highly formal presentations in inductive logic, confirmation theory and formal epistemology.[3] In between these two extremes, there are semi-formal proposals due to Achinstein (2001) and Roush (2005), which approach the evaluation of evidence probabilistically, but independently of statistics. Finally, there are statistical approaches to the evaluation of evidence, which have slowly began penetrating philosophical discussions with some authors calling on philosophers to pay more attention to these statistical approaches to evaluating evidence.[4]

Here my goal is to characterize, from a *statistical perspective*, what made the evidence from Perrin's experiments on Brownian motion good. In order to accomplish this goal, I will focus specifically on Perrin's granule-displacement experiments that confirmed Einstein's diffusion model of the motion of suspended particles in a dilute solution. One reason for just focusing on the granule-displacement exper-

[1] See especially Psillos (2018), Psillos (2021) and Stanford (2021).

[2] See Glymour (1980), Salmon (1984), Mayo (1996), Maddy (1997), Achinstein (2002), Maddy (2007), Stanford (2009), Psillos (2011), Psillos (2014) and Smith and Seth (2020) for a sample of some of the recent views that have been offered on this topic.

[3] See Carnap (1962), Fitelson (2007) and Pettigrew (2016) for examples of some of the work that is done here. See Earman (1992) and Sprenger and Hartmann (2019) for a comprehensive overview and detailed bibliographies.

[4] See Edwards, Lindman, and Savage (1963), Edwards (1992), Royall (1997), Mayo (2000), Royall (2004), Forster and Sober (2004), Fitelson (2007), Sober (2008), Mayo and Spanos (2011), Gelman and Shalizi (2013), Mayo (2013), Morey, Romeijn, and Rouder (2013), Reid and Cox (2015), Morey, Romeijn, and Rouder (2016), Gelman and Hennig (2017), Mayo (2018), Rouder and Morey (2019), and Fletcher and Mayo-Wilson (forthcoming) for a representative sample.

iments is that the confirmation of Einstein's model by these experiments has been lauded by many as some of the most convincing evidence that came out of Perrin's laboratory.[5] Another reason for focusing on these experiments is that the observations can be modeled using well-known *statistical models* (see the next section). I will argue that the quality of Perrin's *statistical* evidence that confirmed Einstein's model was good because it was highly specific and discriminating. The specificity and discriminating character of his evidence can be understood using *Bayes Factors* (see section 4 below for a discussion of Bayes Factors).

Nevertheless, one may have the following worry about my strategy. The worry is that Perrin himself makes no statistical arguments using Bayes Factors or the other common measures of statistical evidence, which I discuss in the next section. In fact, Perrin did not even calculate probable errors. So isn't my account ahistorical in the sense that it leaves out the actual context of Perrin's experimental work? This context involved "eye-balling" the data for conclusions as was done in almost all 19th century chemical and epidemiological research, despite the widespread knowledge of Laplace's work on probability and of least squares and its connection with Gaussian distributions.

In response to this worry, let me say that my goal is to show what a Bayesian statistical account, using Bayes Factors, of Perrin's experimental work *would look like* and not to argue that, in fact, this is what Perrin *did*. The advantage of the account which I intend to provide is that it can illuminate some of the existing accounts of the quality of Perrin's evidence that have been offered in the literature while avoiding some of their shortcomings. Consider, for example, Salmon's argument for scientific realism about atoms and molecules. According to Salmon (1984, 213 – 227), Perrin provided strong evidence for the atomic hypothesis by compiling, in *Les Atomes* (1913), converging values for Avogadro's Number from a variety of independent experiments. He argues that it would be "an utterly astonishing coincidence" to have values for Avogadro's Number from different independent experiments all converging to approximately the same value unless there was a common causal explanation — atoms and molecules. As I will show in what follows (see section 6 below), one can use Bayesian statistical analysis to illuminate the force of, and intuition underlying, Salmon's argument. The specific improvement to what Salmon did is that not only do I single out all the experiments that Salmon cites, but I also provide explicit statistical models for them and show how the experiments are linked using a Bayesian *meta-analysis*.

Now consider more recent authors. On the one hand, Mayo (1996, Ch. 7) has a compelling discussion, from a Frequentist perspective, of the severe testing and statistical reasoning involved in Perrin's confir-

[5] See Smith and Seth (2020, 154). It is worth mentioning that Perrin performed at least three types of experiments on Brownian motion: vertical-gradient experiments, granule-displacement experiments and granule-rotation experiments. See Nye (1972) and Smith and Seth (2020) Chapter 4 and for a detailed discussion of these experiments. In Psillos (2011) and Psillos (2014), Psillos focuses on the vertical-gradient experiments. I do not wish to claim that my analysis of the experiment I focus on extends to these experiments. I believe that these other experiments would require a different account to make sense of how or whether they provided strong evidence and what this evidence was for.

mation of Einstein's model. She eschews any Bayesian characterization of this episode because Perrin's reasoning did not involve any explicit specification of prior credences.[6] On the other hand, Achinstein (2002), Psillos (2011) and Psillos (2014)'s arguments for scientific realism about atoms and molecules involve assumptions about what the value of the prior credences must be in order for their arguments to work. Smith and Seth (2020, 81, n. 13) find these assumptions ad hoc.[7] In what follows, unlike Mayo, I provide a Bayesian statistical perspective of Perrin's confirmation of Einstein's model. It is not my intention to criticize Mayo's illuminating error-statistical/severe testing perspective of Perrin's evidence; nor am I interested in the question of realism about atoms. My intention is to provide a correct retrospective Bayesian *statistical* perspective of Perrin's evidence for the first time (to the best of my knowledge) as an *alternative* perspective to Mayo's influential account of the same episode. At the same time, I will show how my approach avoids the objection of ad hoc specification of priors that has been raised against Achinstein and Psillos.

Here's how I have organized the rest of my paper. In the following section, I set the stage for what I mean by statistical evidence and why this matters for my account. This section is followed by another section with a detailed scientific and philosophical analysis of the reasoning or arguments involved in *specifying* the relevant theoretical models and statistical models that are involved in evaluating Perrin's evidence. In section 4, I give a quick overview of the Bayesian approach to statistical inference and Bayes Factors, which I use to characterize of the quality of Perrin's evidence. In section 5 and 6 I briefly discuss the pros and cons of using Bayes Factors to quantify statistical evidence and draw some lessons for philosophy of science. Remaining questions for my account are addresed in section 7 before I conclude.

2 What is Statistical Evidence?

Before proceeding, let me say more about what I mean by "statistical evidence" and why I want to characterize Perrin's evidence from a Bayesian statistical perspective. Statistical evidence is a form of evidence, somewhere between the semi-formal and formal accounts on the spectrum just mentioned. First, it is *evidence* because it shares what is common to the genus of evidence, namely, a capacity to impact our epistemic attitudes towards a claim or our dispositions to act, which are influenced by claims that we accept. The disjunction in the preceding sentence is important because on a widely-held, standard and non-technical understanding of the meaning of "evidence", evidence has to do with belief and must always impact our epistemic attitudes or beliefs about a claim. I am not disputing this understanding.

[6] See Mayo (1996, 232, 242).

[7] The problem of specifying prior credences is related to the problem of unconceived alternatives. It has been discussed by Roush (2005) and Stanford (2009) in connection to the problem of the "Catch-all Hypothesis". See section 6 below for how my account can avoid this problem.

Rather, I am claiming that it is too narrow because: (i) it leads us to preclude certain items in the world as evidence; and (ii) it prevents us from making certain comparisons we would like to make, for example, between statistical evidence in classical statistics and statistical evidence in Bayesian statistics. To broaden our discussion of evidence, I propose to refer to the widely-held, standard account of evidence as the *epistemic* role of evidence and distinguish this from the *indicative* role of evidence such as making decisions using statistical evidence. Second, it is a *form* or *kind* because it can be distinguished as a species under the broader genus of evidence. The species of evidence have differentiating features, which are suggested by the modifier or adjective. Some of the species (non-exhaustive) of this genus are: direct evidence, historical evidence, indirect evidence, legal evidence, observational evidence, and propositional evidence. One view in epistemology, in fact, is that evidence is propositional.[8] Propositional evidence is, roughly, a fact p — where p is a proposition — that confirms or justifies a given belief token.[9]

Of course there is some overlap between these species. Legal evidence is often propositional evidence since legal evidence is a fact that is admissible in a legal context such as a court of law. At the same time, legal evidence can include some pieces of historical evidence such as eye-witness reports or testimonies. Direct evidence often includes observational evidence such as seeing a smoking gun. Statistical evidence itself can offer indirect evidence for facts, i.e., propositional evidence.[10] One can try to make distinctions between all these kinds of evidence precise. But making these distinctions lies beyond the scope of this paper. I mention these various ways of talking about evidence to justify speaking of evidence more broadly and also insofar as it allows me to focus entirely on statistical evidence within *statistics*, as opposed to: (i) informal ordinary-life discussions of these other kinds of evidence in epistemology, and (ii) the formal discussions of evidence E, hypothesis H and theory T (where E, H and T are unqualified) that is the bread and butter of inductive logic or formal epistemology.

What distinguishes statistical evidence from evidence E in formal epistemology are two things: (i) the random character of statistical evidence, and (ii) the requirement of a statistical model of the observed data. In fact, these two distinguishing features of statistical evidence are linked. The random character of statistical evidence depends on the statistical model for data. Here's what I mean. A measure of statistical evidence is a real-valued function whose inputs are statistics. A statistic, by definition, is any function $t(\boldsymbol{x})$ of actual data $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. In the context of *parametric* statistical inference, the data is modeled as a realization of a finite vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ of observations, measurements etc. For i in $1, 2, \ldots, n$, each X_i is a random variable that has a probability distribution function $f(X_i; \boldsymbol{\theta})$

[8]See Williamson (2000, Ch. 9, 10) and Brown (2015) for the notion of propositional evidence. See Joyce (2004) for a helpful summary and appraisal of Williamson's account of evidence.

[9]Williamson (2000, 194) writes, "Propositions are the objects of propositional attitudes, such as knowledge and belief; they can be true or false; they can be expressed relative to contexts by 'that' clauses."

[10]Compare with Mayo (2018, 435)'s discussion of direct and indirect uses of probabilities.

that belongs to a *parametric family*, members of which are *identified* by the specific value of the parameter vector $\boldsymbol{\theta}$ they take. The model of the data is known as the *sampling model.* In the Frequentist approach to statistics, a statistical model just consists of the sampling model for the data. As we shall see in section 5 below, a Bayesian statistical model requires not just a *sampling model* for the data, but also a *prior model* on anything the data conditionally depend on that we would like to incorporate into our analysis. For example, suppose the experiment is to determine the coefficient of thermal expansion of a steel rod. Here the results of independent repeated measurements of the length of the rod (at a given temperature) can be modeled as the components of a vector $\boldsymbol{X}$ from the Normal or Gaussian parametric family of distributions. In the case of the normal or Gaussian parametric family, $\boldsymbol{\theta} = (\mu, \sigma^2)$ is a vector consisting of the mean (μ) and variance (σ^2) of the distribution. If $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ are the actual outcomes of those measurements, a statistic such as the mean $t(\boldsymbol{x}) = \bar{x} = \sum_i^n x_i/n$ is clearly a function of the data, which can be used in statistical inference: (i) to estimate μ; or (ii) to test hypotheses regarding μ.

Since a function of a random variable is a random variable, a statistic is a random variable. This means that a measure of statistical evidence is a random variable insofar as it takes random variables as inputs. This also means that regardless of what one's attitude towards the *evidential value* of p-values, confidence intervals, likelihood ratios, odds-ratios, Bayes Factors, and Mayo's Severity Function is; it remains the case that these are all measures of *statistical* evidence in the various schools of statistics on this picture.[11] To be sure, some of these measures do not have an *epistemic* evidential role.[12] The p-value, for example, is not a conditional probability assuming the null hypothesis is true. This is because on the Frequentist interpretation of probability, which is the interpretation that p-values are based on, statistical hypotheses are not repeatable events (so, statistical hypotheses cannot be modeled as random variables). Moreover, confidence intervals do not indicate our degree of belief or epistemic confidence about the boundaries within which a parameter might lie. The reason is that confidence intervals (which are based on the Frequentist interpretation of probability) are random variables while parameters (on the Frequentist approach to statistics) are unknown but not random.

The foregoing discussion is intended to restrict the scope of my paper and to provide some background to what I mean by "statistical evidence". By focusing on the evaluation of the quality of Perrin's evidence from a Bayesian statistical perspective, I am signaling two things. Firstly, I am signaling that I will be interested in Bayesian statistics not Bayesian formal epistemology. On the one hand, Bayesian statistics is one approach to statistical inference. Among the things that distinguishes Bayesian statistics, from say, classical Frequentist statistics, is that on the Bayesian approach one can consider prior distribution

[11]Mayo's Severity Function is discussed in Mayo (2018, 143ff.).

[12]Compare with Fletcher and Mayo-Wilson (forthcoming).

functions on parameters.[13] On the other hand, Bayesian formal epistemology or Bayesianism is a philosophy or school of thought that addresses questions in the theory of knowledge and confirmation theory. This philosophy has two distinguishing features: (i) an epistemic interpretation of probability as coherent graded beliefs or credences and (ii) the use of Bayes' theorem as an inductive rule through one form or another of conditionalization.[14] One reason for restricting my interest here is pragmatic, i.e., I don't have much to say about Bayesian formal epistemology. Another reason for focusing on Bayesian statistics is that Bayesian methods in epistemology sometimes mask the subtleties that underly actual Bayesian and non-Bayesian statistical modeling and inference. Given these reasons, I am signaling, secondly, my agreement with Mayo (2018, 73) who writes:

> [T]he Bayesian epistemologist invites trouble by not clearly spelling out corresponding statistical models. They seek a formal logic, holding for statements about radiation, deflection, fish, or whatnot. I think this is a mistake. That doesn't preclude a general account for statistical inference; it just won't be purely formal.[15]

In actual Bayesian statistical modeling, judicious choices of suitable prior probability distributions (to represent prior ignorance, for example) and Markov chain Monte Carlo methods to compute posterior distributions, make Bayesian methods in statistics very subtle business.[16] Elaborate tools for model checking and diagnostics are also being advocated for.[17] This means that care must be taken in going from the famous example, which serves to motivate Bayes' theorem by exposing base-rate fallacies in diagnostic medical testing, and the use of "Bayesian methods" in philosophy of science, where the distinction between Bayesian statistics and Bayesian epistemology isn't always made.[18]

Finally, by focusing on evaluating evidence from a statistical perspective I can distinguish between the following levels:

(1) Substantive or fundamental theories

(2) Theoretical models of these substantive theories

[13] Barnett (1999) gives a good overview of the various paradigms of statistical inference. Compare with Bernardo and Smith (2000) and Part III of Bandyopadhya and Forster (2011). For the information-theoretic approach to model selection and statistical inference see Burnham and Anderson (2002).

[14] See Sprenger and Hartmann (2019)

[15] A point related to the one Mayo makes here can be made using Woodward's distinction between *data* and *phenomena*. Woodward argues that one needs to use statistical methods, at least in science, to analyze data before one can infer that they have evidence for the existence or nonexistence of a phenomenon. See Woodward (2011), Bogen and Woodward (1988), and especially Woodward (1989, 409). More recently, Norton in Norton (2003) and Norton (2021) has also expressed his opposition to an entirely formal account of inductive inference.

[16] See Efron and Hastie (2016, Ch. 13) for a good discussion.

[17] See Morey, Romeijn, and Rouder (2013).

[18] See Magnus and Callender (2004) for discussion of the base-rate fallacy in the context of the realism-antirealism debate in philosophy of science.

(3) Statistical models of the theoretical models

I will return to this three-level distinction in section four and seven below. For now, I summarize the description and relationships between these levels. Sentences in a theory (substantive or not) specify constraints. For example, in the theoretical model of a substantive theory such as kinematics, a constraint can be that there is linear relationship between velocity and time for a body in uniform motion. A statistical modeler will use data to evaluate the validity of the constraints using a statistical model that captures those theoretical constraints. Because data collected from measurement processes typically involve errors or uncertainties, we need to use statistical methods to handle these uncertainties using the statistical models before drawing inferences about the theoretical models or substantive theories. Typically, at least in science, it is statistical evidence at the level of statistical models that (indirectly) impacts our beliefs about theoretical claims or hypotheses at level (1) and level (2).

In what follows, I make use of this three-level distinction in characterizing what made the quality of Perrin's evidence for the discontinuity of matter good (see section 4 and section 7 below). There is a hydrodynamical theoretical model given by Langevin's equation at the level of theoretical models, and a statistical model for the granule displacements based on Einstein's diffusion model for Brownian motion at the level of statistical models. It is the statistical evidence supporting the Gaussian distribution of granule displacements that I use to argue that Perrin had obtained strong statistical evidence for the discontinuous structure of matter, which is assumed in the derivation of Langevin's hydrodynamical model of Brownian motion. The discontinuous structure of matter in this hierarchy will be at the level of substantive or fundamental theories. I discuss all these interrelated parts of my account more fully in what follows, especially in section 7.

3 Perrin's Evidence

3.1 Einstein's Diffusion Model

In the previous section, I have said what I mean by statistical evidence and why that matters for my account. Here I want to say what Perrin's statistical evidence was. For this I need to say what the statistical model is and how it was supported by the data or observations that Perrin made. In this subsection I discuss what the statistical model is, in the next subsection I discuss what theoretical model it is a model of. In the next section I say why the evidence for this statistical model provided strong evidence for the theoretical model it is a model of.

The statistical model in Perrin's granule experiments on Brownian motion was suggested by Einstein's

Projections (in μm) comprised between	First Series		Second Series	
	n_i (found)	n_i(calculated)	n_i(found)	n_i(calculated)
0 and 1.7	38	48	48	44
1.7 and 3.4	44	43	38	40
3.4 and 5.1	33	40	36	35
5.1 and 6.8	33	30	29	28
6.8 and 8.5	35	23	16	21
8.5 and 10.2	11	16	15	15
10.2 and 11.9	14	11	8	10
11.9 and 13.6	6	6	7	5
13.6 and 15.3	5	4	4	4
15.3 and 17.0	2	2	4	2

Table 1: Calculated and observed n_i in two series of experiments

1905 diffusion model for Brownian motion.[19] Assuming the equipartition of energy among the three degrees of freedom and Van't Hoff's Law that extends the ideal gas law to dilute solutions; Einstein's diffusion model made a prediction for how many granules would be displaced from mean position (the origin) after a given time due to osmotic pressure. The prediction was that the number n_i of suspended granules between two fixed points a and b on the the x-axis that would be displaced from the mean (the origin) after a given time t would be given by the following formula.[20]

$$n_i = n \times \int_a^b \frac{1}{\sqrt{2Dt}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x^2}{2Dt}\right)} dx$$

The integrand is the familiar Gaussian or normal probability distribution function. This formula says that the predicted n_i is given by multiplying the area under this function (the definite integral between a and b), which is a probability, with the total number of suspended granules n. In other words, Einstein's diffusion model for Brownian motion predicted that the statistical model of the displacements of the n suspended granules is Gaussian or normal with a mean of 0 and a variance or mean squared displacement $\xi^2 = 2Dt$.[21] D is the diffusion coefficient and t is the time interval. Using Einstein's model, Perrin and his graduate student M. Chaudesaigues made the calculation for $t = 30$ seconds of the predicted number n_i of gamboge particles that would be displaced within intervals that were multiples of $1.7\mu m$. They then recorded the number n_i of particles observed within these intervals alongside their predicted values in Table 1.[22]

[19]See Einstein (1905)'s "On the Movement of Small Particles Suspended in a Stationary Liquid Demanded by the Molecular-Kinetic Theory of Heat" translated by A. D. Cowper and reprinted in Einstein (1956, 1 – 17).

[20]Although the displacements happen on the surface of the liquid, the equipartition of energy allows us to take projections of this displacement on the x-axis and analyze the displacement there.

[21]The notation of ξ^2 instead of σ^2 for mean squared displacement follows Smith and Seth (2020).

[22]This table is taken from Perrin (1910, 64 - 65).

The data above show a close match between predicted values by Einstein's model and observed values from Perrin's experiments. They confirm a statistical model. Specifically, the data confirmed that the statistical model for the displacements of n suspended granules on the x-axis after time t is Gaussian with a mean of 0 and variance (or *mean squared displacement*) of $\xi^2 = 2Dt$. It is standard practice among statisticians (both Bayesian and Frequentist) to write this statistical model using the following shorthand:

$$x(t) \sim N(0, 2Dt)$$

Here $x(t)$ denotes the displacement on the x-axis after time t. This shorthand is read as "The displacements on the x-axis are normally distributed with a mean of 0 and variance of $2Dt$." I shall make use of this shorthand freely in what follows.

But what is the significance of this confirmation? Since Perrin's experimental work was geared towards confirming the molecular-kinetic theory, the significance of this confirmation lies in what can be inferred to exist on the basis of this evidence. The questions now are: (1) what can be inferred to exist? And (2) how can we understand the quality of the evidence that justifies us in making those inferences? The answer to the first question involves showing that there must be some force sustaining the motion of the particles in Brownian motion. The answer to the second question involves making assumptions about the nature of this force and checking whether the nature of this force, as assumed, is supported by Perrin's evidence for Einstein's diffusion model. It will emerge in what follows that assumptions regarding the nature of this force can be used to specify *two* competing hydrodynamical models (models based on forces due to the surrounding liquid on which the Brownian particles are suspended). One of these hydrodynamical models is compatible with the liquid behaving like a continuum fluid at a scale immediately below that of Brownian particles, while the other is incompatible with the liquid behaving like a continuum fluid at that scale. It is this specification of competing models that I will use to support my argument that Perrin's evidence was good because it was highly specific and discriminating.

For these reasons, confirmation of the predicted statistical model for granule displacements from Einstein's diffusion model is important for my argument because the statistical evidence Perrin obtained supporting the statistical model, can be construed as specific and discriminating evidence for a *hydrodynamical* model, which is formulated assuming the existence of the force sustaining Brownian motion. To this end, in subsections 3.2 – 3.3, I will be concerned with answering the first question by analysing the reasoning involved in making the inference about what exists. In section 4 I will give my answer to the second question, which is: "why was the evidence that warranted this inference good?"

3.2 There are highly localized and irregular pressure fluctuations

In order to show that there must be some force sustaining the motion of the particles in Brownian motion, I will refer to Einstein's 1907 paper written in response to Svedberg's 1906 publication of the results he had obtained concerning Brownian motion.[23] This paper is important for my argument because one of the arguments in it is that there must exist random impulsive forces acting on suspended granules if hydrodynamics is to be reconciled with the kinetic theory of heat.[24] From a hydrodynamical point of view — the appropriate level of description of Brownian motion — the best candidate for these impulsive forces are *highly localized and irregular pressure fluctuations*.[25] Therefore, this paper contains the answer to our first question, which is: what can be inferred to exist? In this subsection I summarize the reasoning involved in Einstein's argument.

From the kinetic theory of heat, the mean velocity $\bar{v}$ of a suspended particle of mass m can be determined using

$$m\frac{\overline{v}^2}{2} = \frac{3}{2}k_BT \tag{1}$$

$k_B = \frac{R}{N_A}$ is Boltzmann's constant, R is the ideal gas constant, N_A is Avogadro's Number, and T is the absolute thermodynamic temperature. Equation (1) expresses the familiar idea that temperature is proportional to average kinetic energy. For particles in colloidal platinum solutions such as the ones Svedberg had prepared $\bar{v} = 8.6\text{cms}^{-1}$.[26]

Now suppose that the *only* force acting on the suspended particle undergoing Brownian motion is a viscous drag force, i.e., a force due to liquid friction that decelerates the particle. Newton's second law of motion for this particle of mass m is:

$$m\frac{dv}{dt} = -\zeta v \tag{2}$$

$\zeta = 6\pi r\eta$ is a damping term (i.e., a term that determines the rate of deceleration) that depends on η the viscosity of the liquid and r the radius of the spherical particle. This equation ignores the inertia of the particle and says that the dynamics of the suspended particle is only governed by Stokes' Law.[27]

[23] The Einstein paper is reprinted in Einstein (1956, 63 – 67) as "Theoretical Observations on the Brownian Motion."

[24] Compare with Munson, Young, Okiishi, and Huebsch (2009, 97) for the modern hydrodynamical modeling of pressure.

[25] Einstein was indeed using background information in order to arrive at an informative prior specification of the model space of Brownian motion. Some of this background information can be traced back to the work of the French physicist Léon Gouy. In 1889 and 1895, Gouy had performed careful experiments and published his results of these experiments thereby considerably narrowing down the space of theoretical explanations for Brownian motion. Nye writes, "Gouy excluded all exterior causes except the internal agitation of a liquid, and stated that [Brownian movement] is a direct and visible proof of the modern hypothesis of the nature of heat." See Nye (1972, 27 – 29). As will emerge below, the model space for Brownian motion can be narrowed down even further into a mutually exclusive and exhaustive set of two models: a continuity of matter model vs. a discontinuity of matter model.

[26] The assumptions in Einstein's calculations here are: $m = 2.5 \times 10^{-15}g$, $k_B = 1.38 \times 10^{-23}m^2kgs^{-2}K^{-1}$ and $T = 292K$.

[27] See Munson, Young, Okiishi, and Huebsch (2009, 493 – 500) for a derivation of Stokes' Law and the validity of the

If (2) is the hydrodynamical law governing Brownian motion, one can show that for a colloidal platinum particle suspended in water, it would take 3.3×10^{-7} seconds for it to lose one-tenth of its velocity. This means that at the macroscopic time-scale of a laboratory measurement (about 30 seconds), the particle would have lost almost all of its velocity through hydrodynamic friction or viscous drag.

But since Brownian motion was known to be incessant, equation (2) is not the true dynamical law governing the motion of a particle undergoing Brownian motion. This means that in order to maintain the average $\overline{v}$ demanded by the kinetic theory of heat (8.6cms^{-1}), the suspended particle must experience rapid impulses from *somewhere*. Here's how Einstein (1956, 66) puts it:

> We have to modify this conception [equation (2) above], we must assume that the particle gets new impulses to movement during this time by some process that is the inverse of viscosity, so that it retains a velocity which on an average is equal to $\sqrt{\overline{v^2}}$.

This modification was implemented by Langevin in 1908 leading to the celebrated *Langevin equation*.[28]

$$m\frac{dv}{dt} = -\zeta v + F(t) \tag{3}$$

In arguing that $F(t)$ exists, one cannot beg the question by assuming a priori that it is due to molecular impacts.[29] Notice as well that Einstein only says that we must assume "new impulses by some process." Compare this with how Smith and Seth (2020, 236) describe the situation:

> Pressure-gradients must be present in the liquid even though it is in thermodynamic equilibrium. The local pressure-gradients must be associated with and hence arising in conjunction with highly localized, extraordinarily rapid pressure fluctuations occurring continually throughout the liquid.

Therefore, the most one can say, from a hydrodynamical point of view, is that the source of $F(t)$ are *pressure impulses* or fluctuations on the suspended granules from the ambient fluid. These fluctuations happen locally, on extremely short time-scales, and haphazardly.

3.3 Evidence for discontinuity

In the previous subsection, I have answered the first question, namely: what can be inferred to exist? I showed that $F(t)$ must be included in an accurate hydrodynamical model for Brownian motion. I

assumption that inertia can be ignored.

[28]See the translation of the Langevin paper in Langevin (1997).

[29]See Smith and Seth (2020, 237ff) for some of the arguments why one must not assume this. Compare with Stanford (2009, 257) who discusses Roush (2005)'s modest atomic hypothesis.

now want to give an answer to the second question, namely: why was the quality of Perrin's evidence good? This involves showing that the confirmation of the Gaussian distribution variance $\xi^2 = 2Dt$ of the statistical model provides strong statistical evidence for a hydrodynamical model of Brownian motion that includes $F(t)$. But such a hydrodynamical model implies the discontinuity of matter at a level *immediately* below that of the suspended granules. Therefore, the main goal of this subsection is to show how Perrin's statistical evidence confirming the statistical model is also evidence for the discontinuity of matter.[30] I will use the implied discontinuity of matter by a hydrodynamical model that includes $F(t)$ in my argument that Perrin's evidence was specific and discriminating in section 4.

First, let me rewrite equation (3) above explicitly in terms of $x(t)$, i.e., displacement in the x-direction.

$$m\frac{d^2x}{dt^2} = -\zeta\frac{dx}{dt} + F(t) \tag{4}$$

Langevin solved this equation for the mean squared displacement $\overline{x(t)^2} = \xi^2 = 2Dt$ in an "infinitely more simple" way than Einstein. The solution of this equation obtained by Langevin himself (and more rigorously by others after him) was obtained by making two kinds of assumptions. One kind was based on the kinetic theory of heat, namely, that at thermal equilibrium the distribution of the velocities of suspended granules will be the Maxwell-Boltzmann distribution. This was a simplifying assumption since it meant that one could use $m\frac{\overline{v}^2}{2} = \frac{1}{2}k_BT$ instead of individual velocities for each particle. The other assumptions were statistical assumptions about $F(t)$. In fact, in solving the eponymous equation by making the following statistical assumptions about $F(t)$, Langevin was the first to employ methods which were later used for solving what are now called *stochastic differential equations*.[31]

(a) $\overline{F(t)} = 0$

Since $F(t)$ is a fluctuating force on the surface of a fluid at rest in thermal equilibrium, assumption (a) says that $F(t)$ must have zero mean even though it can vary widely and wildly across the surface of the fluid on very short timescales.

(b) $\overline{\langle F(t), F(t')\rangle} = 2\zeta k_BT\delta(t'-t)$

$\overline{\langle\cdot,\cdot\rangle}$ is the autocorrelation function in Mazo (2002)'s notation. By making it proportional to the Dirac delta function ($\delta(t'-t)$), assumption (b) says that the fluctuating forces are sharp, rapid and uncorrelated at different but very short timescales.[32] Some authors take condition (a) and (b) together as implying that $F(t)$ is a Gaussian white noise process.[33]

[30] See Stein (2021) where my framing of the Perrin case is very close to Poincaré's understanding of the atomic debates. The issue, as Poincaré saw it, was to decide between continuous and discontinuous approaches to chemical physics.

[31] For a rigorous discussion of the mathematics and methods involved in solving the Langevin equation, see Mazo (2002) and Coffey, Kalmykov, and Waldron (2004).

[32] See Arfken and Weber (2001, 84 – 88) for discussion of the Dirac delta function.

[33] See Gardiner (1983, 69) and compare with Mazo (2002, 59 – 63) and Coffey, Kalmykov, and Waldron (2004, 12).

(c) $\overline{\langle F(t), x(t) \rangle} = 0$

Finally, assumption (c) says that the fluctuating impulsive pressure forces are independent of position.

The key point here is that by assuming the kinetic theory of heat and using these three statistical assumptions about $F(t)$, Langevin solved equation (4) for the mean squared displacement $\overline{x(t)^2}$ obtaining $\overline{x(t)^2} = \xi^2 = 2Dt$. Now recall that Perrin and M. Chaudesaigues used Einstein's diffusion model to predict n_i. The very close match between observed n_i and predicted n_i that Perrin tabulated was evidence that the distribution of n_i on the x-axis is Gaussian with mean 0 and variance $\xi^2 = 2Dt$. Since the variance of this distribution can be obtained from the Langevin dynamical equation, evidence for Einstein's model is evidence for a hydrodynamical model of Brownian motion in which $F(t)$ has the properties (a) – (c).[34] But such a hydrodynamical model implies that the substructure of the ambient fluid at a scale immediately below that of the granules is discontinuous.

In order to see this discontinuity, recall that according to continuum fluid mechanics, the pressure $F(t)$ depends on, or is related to, position $x(t)$. This follows because the Navier-Stokes Equations together with the Continuity Equation are a set of four equations in four unknowns: three equations involving flow velocity components in the three spatial directions (Navier-Stokes) and one equation involving pressure gradients (the Continuity Equation).[35] This means the system of equations is determined. In particular, given $x(t)$ in a fluid and the flow velocities at $x(t)$ from Navier-Stokes, one can solve for $F(t)$ although a solution cannot always be obtained by analytical means. The existence of this solution assumes that $F(t)$ and $x(t)$ are *related* in the way specified by the system of equations. Since this system of equations characterizes the behavior of *continuous* fluids, the assumption that $F(t)$ is uncorrelated with $x(t)$ (assumption (c) above) is *incompatible* with the substructure of the ambient fluid surrounding a particle undergoing Brownian motion *having* a continuous description. Therefore, this substructure, at a scale immediately below the granules, must be discontinuous.[36]

Further, from the Continuity Equation, we know that pressure at one point in a fluid is transmitted to adjacent points continuously according to the pressure gradients within that fluid. This means that given some pressure gradient, the pressure $F(t)$ at a point $x(t)$ and time t *determines* the pressure $F(t')$ at a different but adjacent point $x'(t')$ at time t'. But the assumption that $F(t)$ and $F(t')$ are uncorrelated (assumption (b) above) at short-time scales is incompatible with the substructure of the ambient fluid being continuous at a scale below that of the granules.[37] Thus, obtaining data verifying $\xi^2 = 2Dt$ would be very unlikely assuming continuity of the ambient fluid at that scale. The fluid must be discontinuous

[34] Actually only properties (b) and (c) are required to infer the discontinuous structure of matter at a scale immediately below those of the granules.

[35] See Munson, Young, Okiishi, and Huebsch (2009, 42, 271).

[36] Compare with Smith and Seth (2020, 237)

[37] Compare with Smith and Seth (2020, 238).

at that scale.

Before moving on to the next section, let me be clear in order to avoid potential misunderstanding. It is known that hydrodynamics assuming continuum fluid mechanics depends on the type of fluid and the scale of the physical processes being modeled. Hydrodynamics may be formulated assuming a continuous fluid description (where Navier-Stokes and the Continuity Equation apply), if the associated *molecular mean free path* λ is small compared to a *typical length scale of the problem* L. The mean free path is the mean distance traveled by a molecule of the fluid between collisions with other molecules. A quantitative measure for identifying the scales at which continuity applies is provided by the dimensionless ratio $Kn = \frac{\lambda}{L}$ known as *Knudsen number*. So in saying that the assumption of continuity of the ambient fluid at that scale is incompatible with the data verifying $\xi^2 = 2Dt$, what I mean is that scales or levels of description matter. At the scale at which Brownian Motion is happening the assumption of continuity is no longer valid because Kn approaches 1, which is considered the upper limit of the continuum hypothesis.[38] The non-zero $F(t)$ is introduced at the microscale where Brownian motion is happening by the Langevin equation and the statistical assumptions for $F(t)$ only apply at this scale. But at this scale or level of description, because the pressure fluctuations $F(t)$ are uncorrelated with position $x(t)$ and at different times, a shift of perspective to incorporate discontinuity is required to explain the data from the experiments.

4 Why was Perrin's evidence good?

The previous section was helpful in specifying the hydrodynamical model space, which I will use to characterize the quality of Perrin's evidence using Bayes Factors. Before offering this characterization, let me first give an overview of what Bayes Factors are.

On a Bayesian statistical approach, the first thing to note is that we are justified (by the theorems of de Finnetti and later Diaconis and Freedman (1980) on *exchangeability*) to speak of probability distributions on *parameters* — effectively treating parameters as random variables. This turns out to have a huge payoff because it naturally leads to *hierarchical* models, where we use higher level probability models of the parameters, which appear on lower level sampling models of the data, to capture dependencies in our data, especially where such dependencies seem reasonable enough to capture. Using a hierarchical model, a Bayesian statistician will not only use the sampling model for the data $f(X_i|\boldsymbol{\theta})$ for $i = 1, 2, \ldots, n$; but also a prior model $\pi(\boldsymbol{\theta})$ for all the parameters $\boldsymbol{\theta}$ in the sampling model. Such a statistician may even have a prior on the models $p(M_j)$ for $j = 1, 2, \ldots, k$ themselves, possibly continuing this hierarchy as high up as they please using *hyperparameters*. We shall see in section 5 below how a Bayesian hierarchical model

[38]See section 1.2.2 in Katopodes (2018) and compare with section 9.9.2 in Rapp (2022).

can be used to analyse Salmon (1984)'s argument.

For now, in order to illustrate the fundamental ideas of a Bayesian statistical model and for the sake of my subsequent argument, suppose that we have observed data $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$. Suppose also that we are just comparing two theoretical models M_1 and M_2. For each model M_1 and M_2 there is a corresponding Bayesian statistical model, which has two parts. The first part, which is common to both Frequentist and Bayesian approachers, is a sampling model for the data. We write the sampling models as:

$$X_i \,|\, \boldsymbol{\theta}_1, M_1 \sim f(X_i \,|\, \boldsymbol{\theta}_1, M_1) \quad \text{for } i = 1, 2, \ldots, n$$
$$X_i \,|\, \boldsymbol{\theta}_2, M_2 \sim f(X_i \,|\, \boldsymbol{\theta}_2, M_2) \quad \text{for } i = 1, 2, \ldots, n$$

Here $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are a set of parameters in the statistical models associated with model M_1 and model M_2. The second part of a Bayesian statistical model, which is its distinguishing feature from a Frequentist statistical model, is a prior model on the parameters in the respective sampling models. We write the prior models as:

$$\boldsymbol{\theta}_1 \,|\, M_1 \sim \pi(\boldsymbol{\theta}_1)$$
$$\boldsymbol{\theta}_2 \,|\, M_2 \sim \pi(\boldsymbol{\theta}_2)$$

Now suppose that $p(M_1)$ and $p(M_2)$ are the prior probabilities on the theoretical models, then Bayesian statistical inference uses Bayes' theorem to find the posterior probabilities on the theoretical models as:

$$p(M_1 \,|\, \boldsymbol{\theta}_1, \boldsymbol{X}) = \frac{L(\boldsymbol{X} \,|\, \boldsymbol{\theta}_1, M_1)\pi(\boldsymbol{\theta}_1 \,|\, M_1)p(M_1)}{\sum_{i=1}^{2} L(\boldsymbol{X} \,|\, \boldsymbol{\theta}_i, M_i)\pi(\boldsymbol{\theta}_i \,|\, M_i)p(M_i)}$$
$$p(M_2 \,|\, \boldsymbol{\theta}_2, \boldsymbol{X}) = \frac{L(\boldsymbol{X} \,|\, \boldsymbol{\theta}_2, M_2)\pi(\boldsymbol{\theta}_2 \,|\, M_2)p(M_2)}{\sum_{i=1}^{2} L(\boldsymbol{X} \,|\, \boldsymbol{\theta}_i, M_i)\pi(\boldsymbol{\theta}_i \,|\, M_i)p(M_i)}$$

The terms $L(\boldsymbol{X} \,|\, \boldsymbol{\theta}_1, M_1)$ and $L(\boldsymbol{X} \,|\, \boldsymbol{\theta}_2, M_2)$ that appear in the expression for finding the posterior probabilities are the joint distributions of the data $\boldsymbol{X}$ under the respective models. These terms are called *likelihood functions* by statisticians.[39] Marginal likelihood functions involve integrating out the parame-

[39]Note that for a Frequentist, the likelihood functions are not conditional distributions; rather the likelihood functions are joint distributions of the data for a fixed value of the parameters $\boldsymbol{\theta}$. For this reason, a Frequentist will write the likelihood function as $L(\boldsymbol{X}; \boldsymbol{\theta}, M)$ to distinguish her approach from the Bayesian approach.

ters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in their respective parameter spaces Θ_1 and Θ_2. That is:

$$L(\boldsymbol{X} \mid M_1) = \int_{\Theta_1} L(\boldsymbol{X} \mid \boldsymbol{\theta}_1, M_1)\pi(\boldsymbol{\theta}_1 \mid M_1) d\boldsymbol{\theta}_1$$
$$L(\boldsymbol{X} \mid M_2) = \int_{\Theta_2} L(\boldsymbol{X} \mid \boldsymbol{\theta}_2, M_2)\pi(\boldsymbol{\theta}_2 \mid M_2) d\boldsymbol{\theta}_2$$

So we may write the posterior probabilities of M_1 and M_2 as:

$$p(M_1 \mid \boldsymbol{X}) = \frac{L(\boldsymbol{X} \mid M_1)p(M_1)}{\sum_{i=1}^{2} L(\boldsymbol{X} \mid M_i)p(M_i)} \tag{5}$$
$$p(M_2 \mid \boldsymbol{X}) = \frac{L(\boldsymbol{X} \mid M_2)p(M_2)}{\sum_{i=1}^{2} L(\boldsymbol{X} \mid M_i)p(M_i)} \tag{6}$$

when we are interested in the posterior probabilities just given the data $\boldsymbol{X}$.

Dividing equation (5) by equation (6) we get:

$$\frac{p(M_1 \mid \boldsymbol{X})}{p(M_2 \mid \boldsymbol{X})} = \frac{L(\boldsymbol{X} \mid M_1)p(M_1)}{L(\boldsymbol{X} \mid M_2)p(M_2)} \tag{7}$$

Rearranging the terms in equation (7) suitably we get:

$$\frac{\frac{p(M_1 \mid \boldsymbol{X})}{p(M_1)}}{\frac{p(M_2 \mid \boldsymbol{X})}{p(M_2)}} = \frac{L(\boldsymbol{X}|M_1)}{L(\boldsymbol{X}|M_2)} \tag{8}$$

The right hand side of equation (8) can be used to quantify the *relative predictive accuracy* of our models.[40] This quotient is the *Bayes Factor*. The left hand side of equation (8) quantifies the ratio with which our credences for each model *updated* given some data $\boldsymbol{X}$. The equality between the left and right hand side of equation (8) connects the *relative strength of evidence* — how much we are led to update our credences for the competing models given some data — with Bayes Factor, which is the relative predictive accuracy of our models.

A more intuitive way of thinking about the Bayes Factor is this: Bayes Factor quantifies the relative strength of our evidence for a given model in terms of the *specificity* and *discriminating character* of that evidence. This way of thinking is valid because there are two ways that the Bayes Factor can be high: either the numerator is high relative to the denominator or the denominator is low relative to the numerator. On the one hand, we can say that a relatively high value for the numerator quantifies the

[40]See Rouder and Morey (2019).

specificity of our evidence given M_1. A relatively high value is in effect saying that this is the sort of evidence that is *very likely* given M_1 when comparing it to M_2. The evidence is more specific to M_1 than M_2. On the other hand, we can say that a relatively low value for the denominator quantifies how discriminating our evidence is. A relatively low value indicates that this sort of evidence or data is *very unlikely* given M_2 when comparing it to M_1. The evidence discriminates against, or rules out, M_2. In sum, we can say that since the strength of our evidence is reflected by how much we are led to update our relative credences in light of it (the left hand side), the evidence in favor of some model is strong or good if it is specific and discriminating (the right hand side).

Here's how I apply Bayes Factor to evaluate the quality of Perrin's statistical evidence for the Gaussian distribution of displacements. Let M_1 be a hydrodynamical model for Brownian motion which includes $F(t)$ and in which the ambient fluid is discontinuous. The alternative M_2 is a hydrodynamical model for Brownian motion which does not include $F(t)$ with its assumed properties, i.e., the substructure of the ambient fluid even at a level immediately below that of the granules is a continuum. M_1 leads to the statistical model discussed in section 3.1. With this model specification, let the marginal likelihoods under M_1 and M_2 be $L(\boldsymbol{X} \mid M_1)$ and $L(\boldsymbol{X} \mid M_2)$, respectively.

On the one hand, $L(\boldsymbol{X} \mid M_1)$ is very high relative to $L(\boldsymbol{X} \mid M_2)$ because Perrin's evidence confirming that the statistical model of the Gaussian distribution of displacements has variance $\xi^2 = 2Dt$ was very specific under the assumption that matter is discontinuous at the scale of the particles undergoing Brownian motion, i.e., under M_1. By "specific" here I am referring to the close match between predicted values by the model and actually observed values in the two series of experiments (see Table 1). The specificity also arises because the Langevin dynamical equation leads to the variance $2Dt$ of the statistical model of displacements. Since D is related to Avogadro's number N_A by

$$D = \frac{RT}{N_A 6\pi r\eta}$$

the specificity of Perrin's evidence for the discontinuity of matter comes from here as well since we can use the estimated D from the distribution of displacements to estimate N_A and compare it to its theoretical value (see section 6 below).

On the other hand, because the Langevin dynamical equation in which $F(t)$ has the properties (a) – (c) is *incompatible* with the ambient fluid being a continuum, $L(\boldsymbol{X} \mid M_2)$ is very low relative to $L(\boldsymbol{X} \mid M_1)$. It is also relatively low because one of the parameters in the associated statistical model, D, is a function of N_A — a key quantity in the molecular-kinetic theory of gases and heat. But since we are assuming that M_2 is formulated within an alternative theoretical framework to the molecular-kinetic theory of gases and heat, M_2 cannot include a function of N_A as a parameter. And any theoretical model that did not include a function of N_A as a parameter would be incompatible with the statistical data on displacements, meaning the likelihood of that data given that model would be very low relative to an alternative that includes N_A.

Let me say a bit more here. My claim is that there is no other way to get the predicted mean square displacements unless one uses a hydrodynamical model with an $F(t)$ term satisfying conditions (a) – (c). Whether this term is needed to make sense of the data is the entire question at issue, and it is by showing that this is needed that Perrin establishes the relevant facts, summarized by "atoms exist" or the determination of N_A. In other words, I am claiming that the close match between theoretical predictions of N_A and actual estimates of N_A from statistical experiments would be very unlikely on theoretical models for Brownian motion that either don't include an $F(t)$ term or explicitly make use of a function of N_A in their derivations. An example of a theoretical model in this category would be Ostwald's *energetics*.[41]

At the same time, I believe that a continuum theorist (working in, say, continuum fluid dynamics) can still accept the statistical assumptions on $F(t)$ because she recognizes that Knudsen number is close to one at the scale of Brownian motion. So nothing I have said impugns continuum fluid dynamics. The pressure gradients at appropriate levels of descriptions required by continuum fluid dynamics are in fact smooth (or at least continuous) and the hydrodynamical phenomena that can be modeled faithfully by the Navier-Stokes and the Continuity Equation at macroscales can still be modeled as such. So it is true that one both has a continuum description of fluid dynamics at the macroscale, but also hydrodynamical models at the scale of Perrin's experiments on Brownian motion which, in fact, show that matter is discontinuous.

Taking together these two points about the relatively high and relatively low values of $L(\boldsymbol{X} \mid M_1)$ and $L(\boldsymbol{X} \mid M_2)$ respectively, it follows that $\frac{L(\boldsymbol{X} \mid M_1)}{L(\boldsymbol{X} \mid M_2)} \gg 1$, i.e., the resulting Bayes Factor comparing M_1 to M_2 is much greater than 1. But this large Bayes Factor means that the quality of Perrin's evidence for the discontinuity of matter at a scale immediately below that of the suspended granules was good because it was highly specific and discriminating.

5 Pros and Cons of using Bayes Factors

Although the Bayes Factor looks like a *likelihood ratio statistic* commonly found in Frequentist hypothesis testing, it is important to emphasize that Bayes Factor is *not* simply a likelihood ratio statistic. I mention this partly to justify my choice of analyzing Perrin's evidence from a Bayesian perspective as opposed to a non-Bayesian or Frequentist perspective. The first distinguishing feature is that we got Bayes Factor from

[41] Smith and Seth (2020, 333 - 341)

the ratio of *marginal* likelihoods — not likelihoods — by integrating out the parameters. This technique of marginalization is not possible within a Frequentist framework where prior models on parameters, which are required in order for this to work, are not considered. This technique turns out to offer the Bayes Factor more flexibility to compare all sorts of models with each other than is possible within the Frequentist framework. There are thus several advantages to using Bayesian methods to quantify the relative evidence we have for *any* theoretical models. The first advantage is that the probability distributions in the sampling models given the theoretical models do not necessarily have to belong to the same parametric family as is typically the case in the case of likelihood ratio based statistics in the Frequentist framework. Secondly, the vector of parameters are not necessarily nested, again, as is typically the case in other likelihood ratio based test statistics in the Frequentist framework.

These clear advantages have to be tempered with some of the well known disadvantages of using Bayes Factors for model comparison or quantifying the strength of evidence. First, Bayes Factors clearly depend on the prior model on parameters. We can see this by looking at how we calculated the marginal likelihoods for each model M_i for $i = 1, 2$

$$L(\boldsymbol{X} \mid M_i) = \int_{\Theta_i} L(\boldsymbol{X} \mid \boldsymbol{\theta}_i, M_i)\pi(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta}_i$$

Now the problem for Bayes Factors has to do with using *uninformative* priors on the parameters. An uninformative prior is a prior that is chosen in such a way that its influence on the posterior distribution is as small as possible. With an uninformative prior, statisticians want to eliminate as much bias as they can from their analysis. A typical uninformative prior on a parameter such as the mean μ of a continuous random variable X is a uniform distribution $\pi(\mu)$ over the entire real line $\mathbb{R}$, also known as a flat prior. The problem is that the uninformative prior in this case turns out to be an *improper prior*. The sense in which it is improper is that it is not a probability distribution function since it is not normalized, i.e., $\int_{\mathbb{R}} \pi(\mu)d\mu = \infty$. So the marginal likelihood functions are undefined in this case and so is the Bayes Factor. For this reason, Bayes Factors are highly sensitive to the prior model on *parameters*.

Another problem with Bayes Factors is that Bayes Factors are not *calibrated*. What this means, among other things, is that there is no way to tell what it means to say that a Bayes Factor is "large" in the same principled way that one can say a given likelihood ratio based statistic in Frequentist settings is large, where one considers the sampling distribution of the statistic. To be sure, there are *guidelines* based a scale given by Jeffreys (1961) (see Table 2) for how to *intepret* the magnitude of a Bayes Factor but this not the same thing as calibrating Bayes Factor.[42] In order to calibrate Bayes Factors one would need, not only to address the question of "largeness" but also to specify *how often* a given value of Bayes

[42] See Kass and Raftery (1995).

$B_{12} = \frac{L(\boldsymbol{X}\|M_1)}{L(\boldsymbol{X}\|M_2)}$	$2\log_{\mathrm{e}}(B_{12})$	Evidence against M_2
1 to 3	0 to 2	Not worth more than a bare mention
3 to 20	2 to 6	Positive
20 to 150	6 to 10	Strong
> 150	> 10	Very strong

Table 2: Guidelines for interpreting the magnitude of Bayes Factor

Factor is expected to occur with a certain choice of a statistical model.[43]

6 Some Lessons

In discussing the pros and cons of using Bayes Factors, I believe now is a good time to return to something I mentioned in the introduction, namely, that Bayes Factors can be used to draw out the force of Salmon (1984, 224)'s "coincidence argument" to a common cause — atoms and molecules. The best way (I can think of) to understand Salmon's argument from a Bayesian statistical perspective is to use a hierarchical Bayesian model for meta-analysis.

Here is why think a Bayesian hierarchical model for meta-analysis is appropriate here. The first reason involves what the goal of a meta-analysis is. As a method for summarizing and integrating the findings of research studies in a particular area, meta-analysis aims to provide a combined analysis of related studies in order to indicate the overall strength of the evidence for, say, a beneficial effect of a treatment under study, or the value of important parameters found in theoretical models. Essentially, meta-analysis involves pooling information across multiple studies each designed to address the same scientific question with the goal often being to estimate a single effect measure common to all studies.

This brings me to the second reason why the Bayesian hierarchical model is appropriate here. What is crucial for Salmon's argument is that the different experiments (5 in total) all lead to converging values of Avogadro's Number N_A. Referring to the four papers that Perrin published in 1908 in the *Comptes rendus* of the Académie des Sciences, Salmon (1984, 216) writes:

> It is of the greatest importance to *our* story to note that these papers included not only the precise value of Avogadro's number ascertained on the basis of his study of Brownian movement, but also a comparison of that value with the results of several other determina-

[43]The calibration of Bayes Factors is related to the problem raised by Mayo (2018) and others (see the discussion following O'Hagan (1995)'s paper) that Bayes Factor can be used to find evidence for a wrong model. Addressing these concerns lies beyond the scope of this paper, but is a fruitful avenue for further work on understanding the philosophy of Bayes Factors as a measure of evidence and is part of work-in-progress by the author of this paper.

Source of Theoretical Model	$N_A = g(\boldsymbol{\theta})$	Estimate for $N_A (\times 10^{22})$
Brownian Motion	$N_A = \frac{RT}{D6\pi r\eta}$	65 - 70.5
Alpha Decay	$N_A = \frac{F}{2e_0}$	62 - 71
X-ray diffraction	$N_A = \frac{8M}{\rho a^3}$	61.5
Blackbody Radiation	$N_A = \frac{R}{k_B}$	61 - 62
Electrochemistry	$N_A = \frac{QM}{me_0\nu}$	60 - 90

Table 3: The Five Ways of Determining N_A

> tions based upon entirely different methods, including Rutherford's study of radioactivity and Planck's work on blackbody radiation.

He writes, later, that:

> If there were no such micro-entities as atoms, molecules, and ions, then these different experiments designed to ascertain Avogadro's number would be genuinely independent experiments, and the striking numerical agreement in their results would constitute an utterly astonishing coincidence...We can say, very schematically, that the coincidence to be explained is the "remarkable agreement" among the values of N_A that result from independent determinations.

Given Salmon's own emphasis on the relatedness (they all lead to approximately the same value of N_A) and independence of the experiments, it makes sense to use a Bayesian *meta-analysis* using hierarchical models. The point I wish to make for understanding Salmon's argument (bracketing the issue of a common-cause) is that a Bayesian hierarchical model for meta-analysis has a distinct advantage when it comes to explaining this agreement between the different experiments whose goal is to estimate N_A; and when it comes to saying why the body of evidence constitutes strong evidence for theoretical models which include specific predictions for N_A.

Let us consider the five theoretical models that Salmon considers (See Table 3).[44] Let X_i for $i = 1, \ldots, 5$ denote the observed values of Avogadro's number from each study. We can then write the sampling model as:

$$X_i|\boldsymbol{\theta}_i \overset{ind}{\sim} f(X_i \,|\, \boldsymbol{\theta}_i) \qquad\qquad \text{for } \; i = 1, 2, \ldots, 5$$

This sampling model says that within each study $i = 1, 2, \ldots, 5$ the observed values of Avogadro's number X_i are jointly but independently distributed according to a distribution $f(X_i \,|\, \boldsymbol{\theta}_i)$ that depends on $\boldsymbol{\theta}_i$.

[44] The figures in the table are taken from Perrin (1910, 90). Compare with Smith and Seth (2020, 260), and Smith and Seth (2020, 369, n.8) for the values of N_A from the experiments on X-ray diffraction by the Braggs.

For example, in the Brownian motion displacement experiments, $\boldsymbol{\theta} = D$, the diffusion coefficient, which for given t can be estimated from the variance $\xi^2 = 2Dt$ of the normal distribution. In the blackbody radiation experiments the key parameter to be estimated is $\boldsymbol{\theta} = k_B$.

Another way of thinking about what the sampling model is saying is this. For each theoretical model, N_{A_i} is the key *theoretical* parameter whose value is to be determined using the X_i. N_{A_i} in turn is some function $g(\boldsymbol{\theta}_i)$ of the parameters $\boldsymbol{\theta}_i$ for $i = 1, \ldots, 5$, which appear in the statistical models associated with each theoretical model. Therefore, for each experiment the joint distribution of the observed Avogadro number X_i is $f(X_i|N_{A_i})$ where N_{A_i} is the "true value" for Avogadro's number in experiment i. This way of thinking is valid since $g^{-1}(N_{A_i})$ gives $\boldsymbol{\theta}_i$. See Table 3.

I do a meta-analysis using a Bayesian hierarchical model by placing a prior on N_{A_i} conditional on a theory that unifies all the studies. One candidate theory is the theory that matter is discontinuous. Call this theory M_1. According to this theory, the N_{A_i} are from a common distribution with unknown parameter N_A, which is the true value for Avogadro's number that we want to estimate by pooling from all the five experiments.[45] To complete the hierarchical Bayesian model I write:

$$
\begin{aligned}
N_{A_i} &\overset{iid}{\sim} h(N_A) \qquad\qquad \text{for } \ i = 1, 2, \ldots, 5 \\
N_A \,|\, M_1 &\sim \eta(N_A \,|\, M_1)
\end{aligned}
$$

Now suppose that $p(M_1)$ represents our epistemic probability that matter is discontinuous. In meta-analysis of the experimental estimates for N_A, we are interested in the posterior distribution:

$$
p(N_A|\boldsymbol{\theta}_i, X_i, M_1) \qquad\qquad \text{for } \ i = 1, 2, \ldots, 5
$$

By Bayes' theorem the posterior distribution is proportional to the joint distribution of N_A, X_i, $\boldsymbol{\theta}_i$ and M_1 for $i = 1, \ldots, 5$. By marginalizing out $\boldsymbol{\theta}_i$ from this joint distrubution we can write

$$
p(N_A|\boldsymbol{X}, M_1) \propto L(\boldsymbol{X}|N_A, M_1)\eta(N_A|M_1)p(M_1)
$$

where $L(\boldsymbol{X}|N_A, M_1)$ is now the marginal likelihood of the data.

I am now in a position to explain how my Bayesian meta-analysis explains why Salmon would find this convergence of values of N_A to be a very compelling argument for a common cause and how that impacts our epistemic beliefs. For fixed $\eta(N_A \,|\, M_1)$ and $p(M_1)$, the posterior distribution $p(N_A \,|\, \boldsymbol{X}, M_1)$

[45] As of 2018 the Committee on Data for Science and Technology (CODATA) places the value of Avogadro's Number at $60.2214076 \times 10^{22}$ mol^{-1}.

depends on the data through $L(\boldsymbol{X} \mid N_A, M_1)$. If the true N_A is 60×10^{22}, then the posterior distribution will have more mass around values of N_A that are around that number since all the X_i's are in fact close to 60×10^{22}.

Now suppose that we are interested in comparing the estimate of N_A conditional on M_1 to the estimate of N_A conditional on an alternative theory M_2. M_2 here can be a different unifying theory, which we want to use to carry out the same meta-analysis we have done using M_1. To achieve a jointly exhaustive set of theories, let M_2 be the theory that says matter is continuous at the microscale. Suppose we also assign the same prior probabilities $\eta(N_A \mid M_2)$ and $p(M_2)$ to M_2 as we did for M_1. In this case, it is precisely the Bayes Factor $\frac{L(\boldsymbol{X}|N_A,M_1)}{L(\boldsymbol{X}|N_A,M_2)}$ that will be used to compare the two models. On the one hand, the theoretical models that M_1 unifies in the meta-analysis all have very close and specific values for X_i conditional on N_A being around 60×10^{22}. In my account, I have shown that this specificity comes in because the parameters in the statistical models $\boldsymbol{\theta}_i$ are different functions of N_A. So using these parameters in the statistical models unified by M_1, we are *more likely* to observe values of X_i in Table 3 than we would if we used a different theory like M_2. A quantitative and precise way of saying this is that Bayes Factor $\frac{L(\boldsymbol{X}|N_A,M_1)}{L(\boldsymbol{X}|N_A,M_2)}$ in favor of M_1 would high. This is the explanation for why Salmon would find this convergence of values of N_A to be a very compelling argument for a common cause. It is important to emphasize, however, that I have offered this explanation without appealing to *any causal explanation*. Because my explanation has focused entirely on considering the relevant *statistical models*, I believe that I can avoid the problem of unconceived alternatives which face accounts of causal explanations such as Salmon's. I believe that this constitutes a significant virtue of my account. I return to this point below.

I can now also say how I avoid the objection of ad hoc specification of priors. I do this by distinguishing the question I am asking from that which Achinstein and Psillos are asking. My question is: what made the quality of Perrin's statistical evidence good? Achinstein and Psillos's question is: how were prominent scientists convinced of the truth of the atomic hypothesis in light of Perrin's evidence? Both Achinstein and Psillos answer their question in terms of how prior credences were updated in light of the data. Let M_1 be the atomic hypothesis and M_2 be an alternative to it. They need $p(M_1)$ to be 0.5 or "not too low" in order that $p(M_1|\boldsymbol{X}) > p(M_2|\boldsymbol{X})$ after updating. These are the prior specifications that Smith and Seth find ad hoc. Because I am asking a different question, the answer to my question depends entirely on Bayes Factor $\frac{L(\boldsymbol{X}|M_1)}{L(\boldsymbol{X}|M_2)}$. Thus, by asking a different question and characterizing the strength of evidence in terms of relative predictive accuracy using Bayes Factor, I can avoid making ad hoc specification of priors.

7 Some remaining questions

There are several questions one may ask at this point. First, one may ask: how faithful is my specification of the model space to this historical episode? In performing his experiments, Perrin was certainly focused on confirming the molecular-kinetic theory or the existence of atoms and molecules. In fact, philosophers who discuss Perrin are also often concerned with the implications of Perrin's work for the existence question. So why am I not presenting the models involved in this historical episode in terms of those that assert the existence of atoms and molecules and those that don't?

Here are my reasons. Perrin cannot possibly have provided strong evidence for atoms and molecules because the actual details of atomic structure require quantum mechanics, which was unavailable to Perrin in 1908 – 1913. This means that if Perrin was providing definitive evidence for anything, it is for something that is compatible with both classical physics (including the kinetic theory) and quantum physics. My specification of the model space has this compatibility.

Essentially, by emphasizing the need to focus on statistical evidence, I am capable of making fine distinctions between the following levels, which I also described briefly in section 2:

(1) Substantive or fundamental theories

(2) Theoretical models of these substantive theories

(3) Statistical models of the theoretical models

The action in the atomic debates I have zoomed in on happens at level (3) — statistical models of the theoretical models. My argument is that Perrin provided strong statistical evidence for the statistical model of the Gaussian distribution of displacements. Using a meta-analysis of the five ways for determining N_A, I have also argued that Perrin could rightly claim to have provided strong evidence that N_A is around 60×10^{22} mol^{-1}. The strength of evidence for the theoretical models is inherited *upwards* if one believes that the statistical models are adequate for capturing the theoretical predictions and constraints demanded by the theoretical models at level (2). For example, if one accepts that the variance of the statistical model for the displacements is $2Dt$, then one accepts the statistical model as an adequate model for checking the theoretical constraints imposed by Einstein's diffusion model and the hydrodynamical model for Brownian motion that leads to the Langevin equation. In this case, the strong statistical evidence accrues to the theoretical model higher up in level (2). Moving up the hierarchy to substantive theories is more complicated for reasons that Stanford (2009) has discussed having to to do with the problem of unconceived alternatives and the "Catch-all Hypothesis".

In considering how to move the strength of Perrin's statistical evidence up to the level (1), I chose to specify the model space at level (1) as M_1 — discontinuity of the ambient fluid immediately below the granules, and M_2 — continuity of the ambient fluid immediately below the granules; because this is the only retrospective, and consistent, way of saying that Perrin provided strong evidence for *something* without begging the question or failing to consider all serious alternatives. This model specification at level (1) has the advantage of including dichotomous and exhaustive models, which means it avoids the "Catch-all Hypothesis" problem (i.e., the problem of exhaustively specifying, in model space, the *logical complement* of a given hypothesis in order to compute $L(\boldsymbol{X})$) discussed in Stanford (2009) in relation to the problem of unconceived alternatives.[46]

Second, and finally, one may ask: if in order to compute Bayes Factors one needs to specify at least *two* competing models, what is (or are) the alternative statistical model(s) that I am comparing *pairwise* with Einstein's statistical model? Einstein's statistical model says $\xi^2 = 2Dt$ and I say that the evidence for this model gives a high Bayes Factor in favor of M_1. But what would the Bayes Factor be if we considered explicit alternatives to Einstein's statistical model? For example, why not have $\xi^2 = 2Dt^2$ or $\xi^2 = 2D\sqrt{t}$ or $\xi^2 =$ constant for all t or $\xi^2 = 2Dt^{-1}$ and so on? The challenge raised by this last question is pressing when one realizes that there are infinitely many *explicit* alternative statistical models that one can specify and that this list of explicit alternative models cannot be exhaustively specified or enumerated. If so, then I have still not shown that the "Catch-all Hypothesis" problem can be avoided.

Here's why I believe the problems raised by this question for my account are only apparent. Recall that at level (2) both Einstein and Langevin were led to their derivations by assumptions about the nature of $F(t)$ that sustains Brownian motion. Both found that assuming discontinuity, which is exhibited by the random pressure fluctuations $F(t)$ at the micro-scale; the mean squared displacement of colloidal particles ξ^2 would have to *increase* with time. This means that on alternative models, which did not include $F(t)$, the mean squared displacement of the colloidal particles would have to either (i) *decrease* over time; or (ii) *remain constant*. Thus, the exhaustive hydrodynamical model space (discontinuous vs. continuous) can be used to partition the statistical model space as follows:

1. Models with increasing ξ^2 over time t
2. Models with decreasing ξ^2 over time t
3. Models with $\xi^2 = 0$ or some constant for all t

As a partition, it is exhaustive of the space and we can thus avoid the "Catch-all Hypothesis" problem. Further, statistical models with decreasing ξ^2 over time t would require ξ^2 to be a monotonically decreasing function of time. This would mean that rather than spread out over time according to Einstein's diffusion model (from regions of high concentration to regions of low concentration), the particles in Brownian

[46] Compare Smith and Seth (2020, 238 – 239).

motion would be clustering or moving closer to the mean position (origin). This is very improbable given the facts we know about diffusion. So models with decreasing ξ^2 over time t can be ruled out. This leaves models with increasing ξ^2 over time t and models with $\xi^2 = 0$ or some constant for all t. Models with $\xi^2 = 0$ or some constant are impossible because it would mean no concentration gradient or osmotic pressure for diffusion to take place. This means that the only possibilities for *any* model in the statistical model space for Brownian motion are ones in which ξ^2 increases over time t. One does not need to specify *explicitly* what the form of ξ^2 in these models has to be as the alternative statistical model to the actual one in which $\xi^2 = 2Dt$. The reason is that the only *free parameter* to be estimated in these statistical models is D, the diffusion coefficient. Let M_1^* be another statistical model for the displacements in which ξ^2 is a *different* increasing function of time and D^* be the estimated diffusion coefficient for this model that is within ε distance of the estimated diffusion coefficient D in the statistical model M_1 with $\xi^2 = 2Dt$. Then Perrin's statistical evidence shows that M_1^* will be practically indistinguishable from M_1. This means that they will both have high a Bayes Factor in their favor compared to *any* model in which ξ^2 is constant or a decreasing function of time and we may choose either M_1 or M_1^*.

8 Conclusion

In conclusion, let me recapitulate the main points of my paper. I have argued that the quality of Perrin's statistical evidence can be characterized as specific and discriminating using Bayes Factors. My argument involved focusing on the data involved in Perrin's confirmation of the statistical model of displacements of particles in Brownian motion according to Einstein's diffusion model. While focusing on this statistical model, I also analyzed the space of hydrodynamical models for Brownian motion carved out by the Langevin dynamical equation, which leads to the variance, $\xi^2 = 2Dt$, of the statistical model. This specification of the theoretical model space was the crucial step in arguing that Perrin provided strong evidence for the discontinuity of matter, while also avoiding the "Catch-all Hypothesis" problem and the ad hoc specification of priors objection that has been directed at Bayesian perspectives of this historical episode.

References

Achinstein, Peter (2001). *The Book of Evidence*. Oxford University Press.

Achinstein, Peter (2002). "Is There a Valid Experimental Argument for Scientific Realism?" *The Journal of Philosophy*, 99, no. 9:470–495.

Arfken, George B. and Weber, Hans J. (2001). *Mathematical Methods for Physicists (5th Edition).* Harcourt Academic Press.

Bandyopadhya, Prasanta S. and Forster, Malcom R. (Eds.) (2011). *Handbook of Philosophy of Science Vol. 7 Philosophy of Statistics.* Elsevier.

Barnett, Vic (1999). *Comparative Statistical Inference.* John Wiley & Sons, Ltd. Third Edition.

Bernardo, José M. and Smith, Adrian F. M. (2000). *Bayesian Theory.* John Wiley & Sons.

Bogen, James and Woodward, James (1988). "Saving the Phenomena". *The Philosophical Review*, 97, no. 3:303–352.

Brown, Jessica (2015). "Evidence and Epistemic Evaluation". In *Oxford Studies in Epistemology Vol. 5* (edited by Tamar Szabó Gendler and John Hawthorne). Oxford University Press.

Burnham, Kenneth P. and Anderson, David R. (2002). *Model selection and Multimodel Inference: A practical Information-theoretic Approach.* Springer New York.

Carnap, Rudolf (1962). *Logical Foundations of Probability.* The University of Chicago Press.

Coffey, William, Kalmykov, Yu P., and Waldron, J. T. (2004). *The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry, and Electrical Engineering*, volume 14. World Scientific.

Diaconis, Persi and Freedman, David (1980). "Finite Exchangeable Sequences". *The Annals of Probability*, pages 745–764.

Earman, John (1992). *Bayes or Bust?: A Critical Examination of Bayesian Confirmation Theory.* MIT Press.

Edwards, A. W. F. (1992). *Likelihood: Expanded Edition.* The John Hopkins University Press.

Edwards, Ward, Lindman, Harold, and Savage, Leonard J (1963). "Bayesian Statistical Inference for Psychological Research." *Psychological review*, 70, no. 3:193.

Efron, Bradley and Hastie, Trevor (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science.* Cambridge University Press.

Einstein, Albert (1956). *Investigations on the Theory of the Brownian Movement (Translated by A.D. Cowper).* Dover.

Fitelson, Branden (2007). "Likelihoodism, Bayesianism, and Relational Confirmation". *Synthese*, 156, no. 3:473–489.

Fletcher, Samuel C. and Mayo-Wilson, Conor (forthcoming). "Evidence in Classical Statistics". In *Routledge Handbook of the Philosophy of Evidence* (edited by Maria Lasonen-Aarnio and Clayton Littlejohn). Routledge. Preprint available online at `http://philsci-archive.pitt.edu/16191/1/Evidence_in_Classical_Statistics%20%286%29.pdf` Retrieved on 10/25/2021.

Forster, Malcom and Sober, Elliott (2004). "Why Likelihood?" In *The Nature of Scientific Evidence: Statistical, Philosophical and Empirical Considerations* (edited by Mark L. Taper and Subhash R. Lele). The University of Chicago Press.

Gardiner, C. W. (1983). *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer.

Gelman, Andrew and Hennig, Christian (2017). "Beyond Subjective and Objective in Statistics". *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180, no. 4:967–1033.

Gelman, Andrew and Shalizi, Cosma Rohilla (2013). "Philosophy and the Practice of Bayesian Statistics". *British Journal of Mathematical and Statistical Psychology*, 66, no. 1:8–38.

Glymour, Clark (1980). *Theory and Evidence*. Princeton University Press.

Jeffreys, Harold (1961). *The Theory of Probability*. Oxford University Press.

Joyce, Jim (2004). "Williamson on Evidence and Knowledge". *Philosophical Books*, 45, no. 4.

Kass, Robert E and Raftery, Adrian E (1995). "Bayes Factors". *Journal of the American Statistical Association*, 90, no. 430:773–795.

Katopodes, Nikolaos D (2018). *Free-Surface Flow:: Shallow Water Dynamics*. Butterworth-Heinemann.

Langevin, Paul (1997). "On the Theory of Brownian Motion (Translated by Don S. Lemons and Anthony Gythiel from "Sur la Théorie du Mouvement Brownien," Fr. Acad. Sci. (Paris) 146, 530–533 (1908))". *American Journal of Physics*, 65, no. 11:1079–1081.

Maddy, Penelope (1997). *Naturalism in Mathematics*. Oxford University Press.

Maddy, Penelope (2007). *Second Philosophy: A Naturalistic Method*. Oxford University Press.

Magnus, Paul D. and Callender, Craig (2004). "Realist Ennui and the Base Rate Fallacy". *Philosophy of Science*, 71:320 – 338.

Mayo, Deborah G. (1996). *Error and the Growth of Experimental Knowledge*. University of Chicago Press.

Mayo, Deborah G. (2000). “Experimental Practice and an Error Statistical Account of Evidence”. *Philosophy of Science*, 67:S193–S207.

Mayo, Deborah G. (2013). “Discussion: Bayesian methods: Applied? Yes. Philosophical defense? In flux”. *The American Statistician*, 67, no. 1:11–15.

Mayo, Deborah G. (2018). *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. Cambridge University Press.

Mayo, Deborah G. and Spanos, Aris (2011). “Error Statistics”. In *Handbook of Philosophy of Science Vol. 7 Philosophy of Statistics* (edited by Prasanta S. Bandyopadhya and Malcom R. Forster), pages 153–198. Elsevier.

Mazo, Robert M. (2002). *Brownian Motion: Fluctuations, Dynamics, and Applications*. Oxford University Press.

Morey, Richard D., Romeijn, Jan-Willem, and Rouder, Jeffrey N. (2013). “The Humble Bayesian: Model Checking from a Fully Bayesian Perspective”. *British Journal of Mathematical and Statistical Psychology*, 66, no. 1:68–75.

Morey, Richard D., Romeijn, Jan-Willem, and Rouder, Jeffrey N. (2016). “The Philosophy of Bayes Factors and the Quantification of Statistical Evidence”. *Journal of Mathematical Psychology*, 72:6–18.

Munson, Bruce R., Young, Donald F., Okiishi, Theodore H., and Huebsch, Wade W. (2009). *Fundamentals of Fluid Mechanics (6th Edition)*. John Wiley & Sons, Inc.

Norton, John D (2003). “A Material Theory of Induction”. *Philosophy of Science*, 70, no. 4:647–670.

Norton, John D (2021). *The Material Theory of Induction*. University of Calgary Press.

Nye, Mary Jo (1972). *Molecular reality: A perspective on the scientific work of Jean Perrin*. American Elsevier Inc: New York.

O’Hagan, Anthony (1995). “Fractional Bayes factors for Model Comparison”. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, no. 1:99–118.

Perrin, Jean (1910). *Brownian Movement and Molecular Reality (Translated from the 1909 8th Series of Annales De Chimie et de Physique by F. Soddy)*. Taylor and Francis.

Pettigrew, Richard (2016). *Accuracy and the Laws of Credence*. Oxford University Press.

Psillos, Stathis (2011). “Making Contact with Molecules: On Perrin and Achinstein”. In *Philosophy of Science Matters: The Philosophy of Peter Achinstein* (edited by Gregory J. Morgan). Oxford University Press.

Psillos, Stathis (2014). “The View from Within and the View from Above: Looking at van Fraassen’s Perrin”. In *Bas van Fraassen’s Approach to Representation and Models in Science* (edited by Wenceslao J. Gonzalez), pages 143–166. Springer.

Psillos, Stathis (2018). “Realism and Theory Change in Science”. In *The Stanford Encyclopedia of Philosophy* (edited by Edward N. Zalta). The Metaphysics Research Lab, Stanford University.

Psillos, Stathis (2021). “From the Evidence of History to the History of Evidence: Descartes, Newton, and Beyond”. In *Contemporary Scientific Realism: Challenges from the History of Science* (edited by Timothy Lyons and Peter Vickers). Oxford University Press.

Quine, W. V. O. (1976). “Posits and Reality”. In *The Ways of Paradox and Other Essays*, pages 233–241. Random House Inc. Originally published in 1955.

Rapp, Bastian E (2022). *Microfluidics: Modeling, Mechanics and Mathematics*. Elsevier.

Reid, Nancy and Cox, David R. (2015). “On Some Principles of Statistical Inference”. *International Statistical Review*, 83, no. 2:293–308.

Rouder, Jeffrey N. and Morey, Richard D. (2019). “Teaching Bayes’ theorem: Strength of Evidence as Predictive Accuracy”. *The American Statistician*, 73:186–190.

Roush, Sherrilyn (2005). *Tracking Truth: Knowledge, Evidence, and Science*. Oxford University Press.

Royall, Richard (1997). *Statistical Evidence: A Likelihood Paradigm*, volume 71. Chapman & Hall.

Royall, Richard (2004). “The Likelihood Paradigm for Statistical Evidence”. In *The Nature of Scientific Evidence: Statistical, Philosophical and Empirical Considerations* (edited by Mark L. Taper and Subhash R. Lele). The University of Chicago Press.

Salmon, Wesley C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.

Smith, George E. and Seth, Raghav (2020). *Brownian Motion and Molecular Reality: A Study in Theory-Mediated Measurement*. Oxford University Press.

Sober, Elliott (2008). *Evidence and Evolution: The Logic Behind the Science*. Cambridge University Press.

Sprenger, Jan and Hartmann, Stephan (2019). *Bayesian Philosophy of Science*. Oxford University Press.

Stanford, P. Kyle (2009). "Scientific Realism, The Atomic Theory, and The Catch-all Hypothesis: Can We Test Fundamental Theories Against All Serious Alternatives?" *The British Journal for the Philosophy of Science*, 60, no. 2:253–269.

Stanford, P. Kyle (2021). "Realism, Instrumentalism, Particularism: A Middle Path Forward in the Scientific Realism Debate". In *Contemporary Scientific Realism: Challenges from the History of Science* (edited by Timothy Lyons and Peter Vickers). Oxford University Press.

Stein, Howard (2021). "Physics and Philosophy Meet: The Strange Case of Poincaré". *Foundations of Physics*, 51, no. 3:1–24.

Williamson, Timothy (2000). *Knowledge and Its Limits*. Oxford University Press.

Woodward, James (1989). "Data and Phenomena". *Synthese*, 79, no. 3:393–472.

Woodward, James (2011). "Data and phenomena: a restatement and defense". *Synthese*, 182, no. 1:165–179.