The Conceptual Role of 'Temperature' in Statistical Mechanics: Or How Probabilistic Averages Maximize Predictive Accuracy*

Malcolm R. Forster†‡
Department of Philosophy, University of Wisconsin-Madison

and I. A. Kieseppä
Department of Philosophy, University of Helsinki

---

*

† Send requests for reprints to: Malcolm Forster, Department of Philosophy, 5185 Helen C. White Hall, 600 North Park Street, Madison, WI 53706; email: mforster@facstaff.wisc.edu; homepage: http://philosophy.wisc.edu/forster.

‡ We wish to thank Marty Barrett, Bob Batterman, Ellery Eells, Berent Enç, Branden Fitelson, Dan Hausman, Alexei Krioukov, Stephen Leeds, Alan Macdonald, Larry Shapiro, and Elliott Sober for helpful discussions on previous drafts and related topics.

ABSTRACT: If scientific reduction requires that micro-theories explain the *truth* of macro-theories, then reduction appears to be an unfulfilled goal of science even in the best examples. If reduction is viewed more liberally as requiring only that the micro-theory explains the *predictive accuracy* of macro-theories, then there may be real examples of reduction.

1 Introduction

Nagel (1961) thought that thermodynamics *reduces* to the physics of molecular motions in the sense that thermodynamics is deducible from Newtonian mechanics with the help of 'bridge' laws that link terms such as 'temperature' to microphysical properties such as kinetic energy. Any deduction is, by definition, *truth-preserving*, so if a macro-theory is deducible from a micro-theory and the micro-theory is true, then the macro-theory is truth. A reduction in Nagel's sense would therefore explain the empirical success of the macro-theory. But do such reductions actually occur in science?

Kitcher (1984) looks at the relationship between molecular genetics and classical Mendelian genetics and concluded that there is no reduction between them. Waters (1990) looks at exactly the same example and says that there is an emerging reductive relationship between them. Today's reductionists, who hold onto the idea that micro-theories *explain the success* of macro-theories, recognize that the explanation is not instantiated by any straightforward deductive relationship.

We explore a different possibility. Perhaps micro-theories succeed only in explaining the *empirical* success of macro-theories, rather than their *truth*. If this view is correct, then there is no need for bridge laws, for the demand for bridge laws is driven by the demand for logical connectivity. Then there would be no need for 'temperature' to be defined in terms of kinetic energy. There would be no need for macro-quantities to reduce to, or supervene, on the micro-properties of the system in order for the explanation to succeed. Instead, quantities like 'pressure' and 'temperature' may be defined as probabilistic averages.

The empirical success of probabilistic theories cannot be characterized in terms of their empirical adequacy in the sense of van Fraassen (1980) because empirical adequacy requires that the observational consequences of a theory are true (probabilistic consequences are not *observed* to be true

or false). So, it is a non-trivial task to define empirical success in such a way that Newtonian mechanics or quantum mechanics can explain the empirical success of statistical mechanical laws. We describe a theorem that shows that if a macro-system is modeled in terms of an *ensemble* of micro-states, then the maximization of the *predictive accuracy* (in the sense defined by Forster and Sober, 1994) is achieved by a probabilistic theory.

Probabilistic modeling is increasingly common outside of statistical mechanics. For example, neuroscientists determine the probability of a neuron spiking given a particular stimulus. In a more recent development, Rieke *et al*. (1997) introduce ensembles and invert the probability—that is, they seek to determine the probability environmental stimuli given a particular neuronal response, for it is the information available *to the organism about the environment* that is relevant to the organism's survival.

Naturally, any example of probabilistic modeling that inverts probabilities will make use of Bayes theorem, and for this reason Rieke *et al*. (1997) describe their approach as Bayesian. All probabilistic modeling is likely to use of Bayes theorem (because it is a theorem of probability), but not all probabilistic modeling is Bayesian. In fact, the use of ensembles obviates the need for priors probabilities, which is the hallmark of Bayesianism. The ensemble approach, coupled with the optimization of predictive accuracy, may provide a more objective understanding of 'Bayesian' models in science.

Probabilities are now widely used to make sense of causal claims like 'smoking causes lung cancer' (*e.g*., Eells 1991, Hausman 1998, Pearl 2000). But if the world is deterministic, where do the probabilities come from? If the underlying theory is quantum mechanics, then it is still unclear how one derives macroscopic probabilities about smoking and lung cancer from quantum mechanical probabilities. Ensemble probabilities arise independently of whether the micro-world is deterministic

or indeterministic. That is why most thermodynamic laws are equally well grounded in Newtonian mechanics and quantum mechanics.

Exactly the same kind of philosophical dilemma lies at the heart of the on-going debate about the statistical character of evolutionary theory (Mills and Beatty 1979; Sober 1984, 118-134; Beatty 1984; Rosenberg 1985; Horan 1994; Brandon and Carson 1996; Graves *et al*. 1999; Glymour 2001; Rosenberg 2001). Rosenberg (2001) follows Sir Ronald Fisher, one of the founders of population genetics, in claiming that the statistical character of evolutionary theory is analogous to statistical mechanics. If this is correct, and the ensemble approach to statistical mechanics (Gibbs 1902) is the correct one, then evolutionary fitness should be defined, not as a propensity, nor as a quantum mechanical probability, but as the average survival rate in an ensemble of populations. The advantage of the ensemble approach is the same as in the example of causal modeling: It works equally well whether the micro-theory is deterministic or not.

We shall not study these examples here. Rather, our aim is to thoroughly understand some very simple examples in statistical mechanics—beginning with one particle in a one-dimensional box. In section 2, we define the notions of 'ensemble' and 'theory', and characterize the payoff of macro-theories relative to ensembles. Section 3 focuses on predictive accuracy as the criterion of success, and proves that theories in the Gibbsian sense are theories that maximize predictive accuracy (in the sense defined in terms of the Kullback-Leibler information). In section 3, we explain how the macro-theory is related to the micro-physics by the minimization of the Kullback-Leibler discrepancy. Because this is not a deductive relationship, quantities such as 'pressure' that appear in macro-laws, such as the ideal gas law, need not supervene on the underlying microstate (section 4). Section 5 extends this discussion to the statistical mechanical concept of 'temperature' in an ideal gas, and critically examines the philosophical lore that temperature is identical to the mean kinetic energy.

2 Ensembles, Theories, and Payoffs

In this section, we explain our definitions using a simple example: Imagine a single Newtonian particle (a 'billiard ball') of mass m bouncing around inside a box. To simplify the mathematics, suppose that the box is a long and narrow cylinder of length L. Effectively, the box is one-dimensional. There are no electric or gravitational fields, so the kinetic energy of the particle is a constant through time. A minor exception to this occurs when the particle bounces off the ends of the box. If these collisions are elastic (that is, the particle has the same kinetic energy after the interaction ceases) and quick, then we can adopt the idealization that the particle *instantly* reverses its momentum when it hits a wall.

Denote the momentum of the particle by p, which is a function of time. When the particle is moving to the right, p is positive. When it hits the rightmost wall, its momentum changes to $-p$. The momentum *lost* by the particle is equal to its initial momentum minus its final momentum; namely, $2p$. This lost momentum is experienced externally as an *impulse* on the external wall of the container. The impulse per unit time per unit area of the container produced over a finite time window is a measurable macroscopic property of the system, called *pressure*. Similarly, the particle loses momentum by an amount $-2p$ when it strikes the leftmost wall of the container. While the momentum of the particle is not conserved, its the kinetic energy is conserved because $\frac{1}{2} mv^2$ $(= p^2/2m)$ does not change when the velocity, or the momentum, changes sign. We assume that the *total* energy of the system (in this case, a single particle) is fixed, which we denote by E. Thus, the *magnitude* of the momentum is a constant over time, although its sign depends on whether the particle is moving to the left or to the right.

Let q denote the position of the particle in the box along the length of the cylinder. Like p, q is also a function of time. Because the particle is confined to the box, $0 \le q \le L$. The ordered pair (q, p) denotes a possible *microstate* of the system at a particular time. The set of all possible microstates is

known as the *phase space* of the system, commonly denoted by Γ. The pair of values (q, p) is the microstate of a one-particle system, and is represented by a point in Γ, called a *phase point*. Now let us focus our attention on a particular time, which we choose to be $t = 0$. How should we represent the system in *macroscopic* terms (given we do not know the microstate of the system at that time)? The principal macroscopic quantity is the total energy of the system, E. By fixing this value, we restrict the set of possible microstates of the system to the *energy surface* associated with the energy value E, where the energy surface is the set of all points in Γ such that $p = \pm\sqrt{2mE}$ and $0 \leq q \leq L$.

Given an initial phase point at $t = 0$, Newton's laws of motion determines the microstate at all times t. The state of the system as a function of time is referred to as the *phase trajectory* of the system under natural motion, where 'natural motion' refers to that determined by Newton's laws. For each initial phase point, there is a unique trajectory, and for each trajectory there is a unique initial phase point.

For our purposes, it is useful to think of Newtonian mechanics as specifying the *probability* of later states conditional on an initial state at $t = 0$. These probabilities are trivially 0 or 1, but their interpretation as probability measures allows us to introduce statistical ideas in a general way. The following definition is a modified version of a definition in Blackwell (1953, p. 265).

> **Definition 1:** A *system type* is an ordered pair $S = (\Lambda, \Omega)$ where $\Lambda = (\lambda_1, \lambda_2, ..., \lambda_n)$ is for some $n \in \mathbb{N}$ an n-tuple of probability measures on the same collection of subsets of the set Ω.

In order to explain the definition clearly, suppose that time is discrete, so that our system evolves discontinuously from t_0 to t_1, and from t_i to t_{i+1}, for any integer i, where each time increment is equal to Δt. Clearly, we may allow Δt to be as small as we like. Now let γ_i denote a phase point that the

system could occupy at time t_i. Then Ω is the set of all logically possible sequences $(\gamma_0,\gamma_1,\ldots,\gamma_i,\ldots)$, where each γ is a phase point in Γ. Each element of Ω is therefore a *trajectory* in phase space. Λ is a set of probability measures on Ω. In our example, a member of Λ, call it λ, will assign probability 1 to one particular trajectory and probability 0 to all other trajectories of Ω. Because there is a one-to-one correspondence between possible initial states and possible trajectories, each λ also assigns probability 1 to one initial state (a phase point γ on the energy surface) and 0 to all others. It is useful to be able to switch back and forth between the two interpretations of λ. For now, it is sufficient to think of λ as specifying a possible microstates of the system at time 0.

Since the microstate is unknown, we may also think of the λ as a hidden variables of the kind postulated in hidden variable interpretations of quantum mechanics. If the λ were to specify non-trivial probability measures, then the hidden variable would be a *stochastic* hidden variable. Our formalism is the same as that required to derive the famous Bell inequalities (Bell 1964), and subsequent generalization of Bell's inequalities in the case of stochastic hidden variables.

Given a particular system type, we may now say something about what it means for a hypothesis to be *successful*. In our discussion we assume that the micro-theory under consideration, in our case Newtonian mechanics, provides a true characterization of the system *type* of the system under consideration. In order to simplify the mathematics, we further assume that there are a finite number of possible states, and that time is discrete.

Now consider a set *H* of hypotheses *h* about the token system. In the statistical mechanics example, these hypotheses will determine a probability distribution on Ω. Since the λ in Λ are probability measures on Ω, they are in *H*, but these are not the only hypotheses in *H*. Since a hypothesis is about a *token* system (rather than the system type), which is characterized by a particular λ (a particular

trajectory in the case of Newtonian mechanics), it is appropriate that the payoff of the hypothesis depends on which λ actually applies to the token system.

We introduce $pay(h|\lambda)$ to denote the payoff of *h* given that the system is in state λ. The payoff may be truth, empirical adequacy, predictive accuracy, or some other measure of success. However, we are only interested in truth-related payoffs, which we call *epistemic* payoffs. We define these by the requirement that a true and complete hypothesis has the highest payoff possible. Truth, empirical adequacy and predictive accuracy are epistemic payoffs in this sense because there is no hypothesis that has greater truth, empirical adequacy, or predictive accuracy, than the true hypothesis. Here is a more precise definition of the idea just described:

Definition 2 A: Suppose that $S = (\Lambda, \Omega)$ is an system type, and *h* is any hypothesis. Let *pay* be a function that associates any pair (h, λ) with a real number $pay(h|\lambda)$ where λ is one of the distributions in Λ. Then the vector

$$pay(h) = \left(pay(h|\lambda_1), pay(h|\lambda_2), \ldots, pay(h|\lambda_n) \right)$$

is called the *payoff vector* of the hypothesis *h* relative to the system type.

Every hypothesis has its payoff, in our sense, independently of the beliefs, desires, or actions of scientists. We are interested in the *achievements* of scientific hypotheses in an objective sense—whether the achievement is intended or unintended, known or unknown, is immaterial.

Definition 2 B: If *S* is an system type and the payoff function is such that $pay(h|\lambda) = 1$ whenever the probability distribution determined by *h* over Ω is equal to λ, and $pay(h|\lambda) = 0$ otherwise, then the payoff function is said to *simple*.

For simple payoffs, $pay(h|\lambda)$ is 1 if h determines λ, and 0 if h does not determine λ. Simple payoffs are *too* simple in the sense they fail to define what is achieved by *false* hypotheses. It is not our view that simple payoffs make no sense, or that they do not play a role in science. Our claim is that other payoff functions are needed to provide a complete account of how science works, especially those branches of science that include probabilistic modeling.

Consider the outcome of an observation of the system, such as a measurement of pressure, volume, or temperature. In our framework, an observational outcome is represented as an element x of X, where X is a partition of the space Ω. Recall that Ω is the set of all possible trajectories. Suppose, for the sake of illustration, we measure the pressure during the interval from t_0 to t_9. The observed pressure at the end of the box depends on the number of impacts during the interval and the magnitude of each impulse on the wall. It does not depend on the exact time at which the impulses occur during the interval. Every trajectory in Ω will determine a unique outcome of a pressure measurement, so there is a one-to-one correspondence between values of pressure and a set of mutually disjoint and exhaustive set of subsets of Ω. Such a set of subsets is called a *partition* of Ω. By representing observational outcomes in this way, we are ensuring that all measurement *outcomes* supervene on the trajectory of the system (that is, given any trajectory, there is a unique measurement outcome). In our simple example, measurement outcomes also supervene on the initial state of the system (because a unique trajectory is determined by the initial state). However, we stress that the quantities introduced by a *hypothesis* need not supervene on the microstate, especially if they are defined in terms of ensemble averages. This is a key point for us.

It is also possible for x to encode information about how the system was set up, such as the volume of the container, the energy of the system, and the number of particles. In a one-particle system, x may

determine the ordered pair of numbers (E,V). For each particular value of (E,V) there corresponds a set of phase space trajectories and the set of all these sets is a partition of Ω. Clearly, the system type is different depending on which parameters we consider to be 'fixed'. Our intention is that there are narrower and broader system types, just as the apple on your table is simultaneously a token of the type 'apple' and 'fruit'. The important point is that different hypotheses apply to systems that have different energies and volumes. We therefore introduce the notion of a *theory*, which maps narrow system types (picked out by *x*) to different hypotheses.

Definition 4: Suppose that (Λ,Ω) is an system type, and let X be a partition of the space Ω. Now consider a function T that maps each value of *x* into a hypothesis in *H*. We call *T* a *theory* (or *macro-theory* if there is some chance of confusing it with the underlying micro-theory). Moreover, the payoff of the theory *T* given λ is equal to the payoffs for the hypotheses $T(x)$ weighted by the probability of *x* when the microstate is λ. That is,

$$pay(T|\lambda) = \sum_{x\in X} \lambda(x)\, pay(T(x)|\lambda).$$

In words, the payoff of a theory is the expected payoff of the hypotheses that it postulates. We are assuming that underlying microstate λ fixes the value of parameters like the number of molecules, the total energy, and the volume of the container. So, if the theory is well defined relative to a broad system type, then $\lambda(x) = 0$ if *x* is not consistent with λ. In this way, the sum over all values of *x* is restricted to a subset of values of *x*.

We have defined a theory as a mapping from a partition *X* into a set of hypotheses. A more general definition of a theory might define a theory as set of mappings from different partitions. Quantum mechanics might be viewed as theory of this kind where the probabilistic hypotheses mapped from

different partitions do not fit together as one might expect on the basis of classical intuitions (as Bell's theorems show). While we do not intend to develop this idea here, we want our definition of 'theory' to allow for such possibilities.

As an illustration of a theory in our sense, suppose that we restrict our attention to a single particle with energy E confined to a one-dimensional cylinder of length L, so that the system type is characterized by the set of points on an energy surface (where each point specifies the position and momentum of the particle at time $t = 0$). Further suppose that x records the pressure exerted by the particle on the rightmost end of the cylinder (which has area A) averaged over an interval of time from $t = 0$ to $t = \Delta t$, where Δt is small enough so that there is either 1 impact or 0 impacts during this time. Then the pressure is either $2|p|/(A\Delta t)$ or 0. What theory should we postulate in order to maximize the payoff in the sense of Definition 4? Is there a theory that is optimal for all values of λ? Note that T cannot pick the true microstate in response to the measurement outcome because the empirical facts underdetermine the micro-facts. A theory must map to a *unique* hypothesis from each x, so theories, by our definition, are *not* underdetermined by the empirical facts. Herein lies the difference between a macro-theory and a micro-theory.

We know that the two possible outcomes of the pressure observation divides the set of possible trajectories into two subsets, which we may label Λ_1 and Λ_0, respectively. Consider the theory that 'guesses' at some particular λ in Λ_1 when the pressure is non-zero and some member of Λ_0 when the pressure is 0. This macro-theory will be wrong most of the time. A second macro-theory may postulate the disjunction of all the λ in Λ_1 or the disjunction of all the members of Λ_0. While many logicians may favor this answer, statistical mechanical theories provide a different nature. Statistical mechanics postulates a probability distribution over Λ_1, or a probability distribution over Λ_0,

depending on whether the observed pressure is non-zero or zero, respectively. But where does this probability distribution come from, and in what sense is the choice of one probability distribution better than another? The answer relies on the concept of an ensemble.

A common misconception about ensembles is that they are just sets of phase points (*i. e.*, system types). However, this is not how ensembles are defined in statistical mechanics. The concept of an ensemble goes one step further in assigning a probability distribution over a set of micro-states.

Definition 5: (Ensemble) The pair $\mathcal{E} = (S, \mathbf{p})$ is *an ensemble of systems* if (i) $S = \left((\lambda_1, \lambda_2, \ldots, \lambda_n), \Omega\right)$ is an system type and (ii) $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ is an *n*-tuple of non-negative real numbers that satisfies the condition $p_1 + p_2 + \ldots + p_n = 1$.

In statistical mechanics, the *equilibrium* ensemble is a fundamental kind of ensemble. An equilibrium ensemble is, by definition, an ensemble that does not change under natural motion. An ensemble assigns a probability to the phase points on an energy surface at $t = t_0$. Each of those points will move along a phase space trajectory (which also lies on the energy surface, since energy is time invariant). Thus the phase point γ_1 may evolve to γ_2 at time t_1. The probability of the state γ_1 at t_0 becomes the probability of γ_2 at time t_1. For the ensemble to be time invariant (stationary), the probability of γ_2 at t_1 must be the same as its probability at t_0. It follows that γ_1 and γ_2 must have the same initial probabilities at t_0. In fact, for any two states γ_i and γ_j, if γ_i can evolve into γ_j after some interval of time, then an equilibrium ensemble must assign γ_i and γ_j the same initial probability.

In our example of a single particle in a one-dimensional box, it is easy to see that the system cycles through every point on its energy surface (which means that the system is *ergodic*).

Therefore, the equilibrium ensemble of an ergodic system is unique, and if the state space is finite, it assigns an equal probability to every point on the energy surface.

The theorem is only true when the system is ergodic and that the number of points on the energy surface is finite. Nevertheless, it illustrates the general point that the requirement that an ensemble is an equilibrium ensemble greatly constrains the number of ensembles that are used in statistical mechanics. If a system is created by allowing two initially separated gases to mix, then after a period of time (called the relaxation time), the system is represented by an equilibrium ensemble. So, even in non-equilibrium statistical mechanics, the assumption that the initial and final ensembles are equilibrium ensembles is a powerful constraint. We shall examine a simple example of this in section 5.

In classical statistical mechanics, there is a theorem, called Liouville's theorem, that implies the existence of an equilibrium ensemble in the sense defined (although in practice, it may be too complicated to calculate). Gibbs (1902) called it the *microcanonical* ensemble. The microcanonical distribution is not usually the only equilibrium ensemble (see Earman and Rédei, 1996). Nevertheless, even when uniqueness fails, the determination of the microcanonical ensemble, and the characterization of thermodynamic properties in terms of a microcanonical averages works very well in a surprising variety applications. Why it works when uniqueness fails is an important problem in the foundations of statistical mechanics.

Part of the justification for the use of microcanonical ensembles (*e. g.*, Khinchin 1949) is that the choice is empirically testable in the following sense. An ensemble assigns a probability distribution to measurable quantities, such as pressure. In our running example, the measurement of pressure over an interval of time Δt is either $2|p|/(A\Delta t)$ or 0. If we perform the measurement on a large number of systems that are macroscopically indistinguishable (in the relevant way, as dictated by the system

type), then we shall observe each outcome with a given relative frequency. Assuming that the test sample is randomly drawn from the ensemble (according to the ensemble probabilities), then the ensemble will assign a probability to the total data (*e. g.*, the microcanonical probability of the non-zero pressure outcome is $\frac{1}{2}(p/m)(\Delta t/L)$). The judgment whether the ensemble probabilities 'fit' sufficiently well with the observed distribution is a common kind of judgment made by statisticians. For large data sets, the judgment is often uncontroversial. Granted, a good fit between the ensemble probabilities and the empirical data doesn't *uniquely* justify the ensemble because there will be many ensembles that assign the same probabilities to the data (have the same likelihood). But uniqueness is not crucial. So long as the microcanonical ensemble, or the non-equilibrium ensemble derived from it, fits the data, then we have an empirical justification for its use.

We grant that justifying the *use* of microcanonical ensembles is not the same thing as *explaining* why they work (Batterman, 1998). But it does highlight the point that explaining why the microcanonical ensemble works need not require that we explain why the microcanonical ensemble is the *only* ensemble that works, for there is no point in trying to explain something that is not true.

The main problem addressed this section is very different: Given that tokens of a system type are represented by a particular ensemble, what is the *best theory* that can be applied to the tokens systems of that type? In particular, how should our representation of the system change in response to the empirical information (x) that we have about the system?

A Bayesian will reply that if the observation of a non-zero pressure restricts the class of initial microstates from Λ to Λ_1, then the initial ensemble distribution should be conditionalized on Λ_1, where Λ_1 represents the proposition that one of the states in Λ_1 is the true initial state of the system. The Bayesian is assuming *carte blanche* that the conditionalization of probabilities is the right way to respond to all empirical evidence.

In our view, the optimality of a theory depends on which epistemic payoff is being optimized, where the optimal theory is determined by the payoff function according to the following definition.

Definition 6: (Optimality) Let us suppose that $\mathcal{E} = (S, \mathbf{p})$ is an ensemble, where $S = (\{\lambda_1, \lambda_2, \ldots, \lambda_n\}, \Omega)$, and $\mathbf{p} = (p_1, p_2, \ldots, p_n)$. Further, let X be a partition of Ω and x an arbitrary member of X such that a theory T that maps every element of X into some hypothesis h. The payoff for T relative to $\mathcal{E}$ is

$$pay_{\mathcal{E}}(T) = p_1\, pay(T|\lambda_1) + p_2\, pay(T|\lambda_2) + \cdots + p_n\, pay(T|\lambda_n).$$

Further, we say that a *theory T** is *optimal* relative to $\mathcal{E}$ if the quantity $pay_{\mathcal{E}}(T)$ is maximized when $T = T^*$.

The most important feature of the definition is only apparent after it is combined with Definition 4:

$$pay_{\mathcal{E}}(T) = \sum_{i=1}^{n} p_i\, pay(T|\lambda_i) = \sum_{i=1}^{n} p_i \sum_{x \in X} \lambda_i(x)\, pay(T(x)|\lambda_i).$$

Now reverse the order of the summation, so that

$$pay_{\mathcal{E}}(T) = \sum_{x \in X} \sum_{i=1}^{n} p_i \lambda_i(x)\, pay(T(x)|\lambda_i).$$

Thus, for each x, the term

$$\sum_{i=1}^{n} p_i \lambda_i(x)\, pay(T(x)|\lambda_i)$$

depends only on the hypothesis that T selects *when the observational outcome is x*. An arbitrary theory is therefore free to choose a hypothesis for one x quite independently of another value of x. Therefore, one can *construct* the optimal theory by mapping each value of x to the *hypothesis* that is optimal *for each x*. The problem decomposes into manageable sub-problems.

A second important consequence of the definition is that the optimal theory may depend only on some parts of the ensemble. Suppose that in the case of our one-particle system, the ensemble is a

weighted average of sub-ensembles, where in each sub-ensemble the values of the total energy, the number of particles, and the volume of the container are fixed. In this case, Λ will include phase points in different phases spaces and on different energy surfaces. It is an interesting consequence of our definition that the optimal theory does not depend on the weights used to join the sub-ensembles. For $\lambda_i(x)$ will be 0 for all λ_i that are not members of the relevant sub-ensemble, so the optimal hypothesis mapped from *x* will depend only on the *relative* weights of the microstates within the relevant sub-ensemble. This is one instance in which different ensembles lead to a single optimal theory.

3 The Maximization of Predictive Accuracy

This section examines the consequences of our definitions for one particular choice of payoff function—the Kullback-Leibler (K-L) information discrepancy. Fundamentally, the K-L information is a measure of the discrepancy between two probability distributions defined on the same event space. So, if q is an arbitrary probability distribution on Ω, and λ is the micro-state of the system, then the K-L information is a measure of the discrepancy of q relative to λ:

$$\Delta_{KL}(q|\lambda) = \sum_{\omega\in\Omega} \lambda(\omega)\log\lambda(\omega) - \sum_{\omega\in\Omega} \lambda(\omega)\log q(\omega),$$

where ω is an element of Ω. A fundamental theorem about the K-L discrepancy is that discrepancy is greater than or equal to 0, and equal to 0 if and only if $q(\omega) = \lambda(\omega)$ for almost all ω (where 'almost all' means 'for all ω except a set of measure zero'). This implies that the K-L payoff is an *epistemic* payoff in the sense defined earlier—if λ represents the complete truth about the token system under consideration, then no other hypothesis has a higher payoff.

Since the first term is constant (independent of q), the discrepancy is minimized if and only if the quantity

$$\sum_{\omega \in \Omega} \lambda(\omega) \log q(\omega)$$

is maximized. So, we define the K-L payoff of q relative to λ as

$$pay_{KL}(q|\lambda) = \sum_{\omega \in \Omega} \lambda(\omega) \log q(\omega).$$

The idea behind this definition is that if ω is the true trajectory of the system, then q receives of score according to the probability that it gives ω, the higher the probability the higher the score. The score is $\log q(\omega)$, which is an increasing function of $q(\omega)$ (the higher the likelihood, the higher the log-likelihood). Furthermore, given if the true probability distribution is given by λ, then the probability of ω actually being the case is $\lambda(\omega)$. Therefore the *objectively* expected score for q given λ, calculated in accordance with the probabilities given by λ is equal to the K-L payoff. The K-L payoff is the same as what Forster and Sober (1994) define as predictive accuracy.

Another interpretation of the K-L payoff is that it is equal to the amount of information (in the information-theoretic sense) contained in q relative to λ. So the K-L payoff is intimately related to measures of information and entropy as they commonly arise in statistical mechanics, although we shall not specify the connection here.

When we turn our attention to determining which *theory* has the greatest K-L payoff, then the situation is very different. For there is there is no *theory* that can map every element of the partition X to the true hypothesis in every instance (for such a mapping would be one-to-many). So, for each value of x, a theory must select a hypothesis that is as *close as possible* to all the λ compatible with x. An ensemble specifies the relative weights of these λ, and this allows a unique determination of the distribution q_x that is the closest to the true λ *on average*. When 'closest' is defined in terms of the K-L discrepancy, then the 'best' theory maps x to q_x, such that q_x maximizes the quantity

$$\sum_{\omega\in\Omega} p_x(\omega)\log q_x(\omega),$$

where $$p_x(\omega)=\sum_{i=1}^{n} w_i(x)\lambda_i(\omega)$$

and $$w_i(x)=p_i\,\lambda_i(x)\Big/\sum_{j=1}^{n} p_j\lambda_j(x).$$

The optimal hypothesis, q_x, is therefore $q_x(\omega)=p_x(\omega)$, or

$$q_x(\omega)=w_1\lambda_1(\omega)+\cdots+w_n\lambda_n(\omega),$$

where $$w_1+\cdots+w_n=1.$$

That is, the optimal way of compromising amongst the different λ is achieved by averaging over the λ.

In the special case when the underlying theory is deterministic, each value of $\lambda_i(x)$ is 0 or 1. So, if we define Λ_x as the set of λ such that for all $\lambda(x)=1$, then

$$q_x(\omega)=\Pr_{\mathcal{E}}(\omega|\Lambda_x),$$

where $\Pr_{\mathcal{E}}(\omega|\Lambda_x)$ is the ensemble probability of ω determined by averaging over the probabilities $\lambda(\omega)$ for all those λ in Λ_x according to the *relative* weights of the λ in Λ_x. While this is the same answer that would be obtained by the Bayesian conditionalization of the ensemble probabilities on Λ_x, *we obtained the answer on the basis of a more fundamental principle*, and the answer would be different if the payoff function were different.

It should be noted that the probability measure q_x does not (normally) belong to the set $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. This is especially clear in the deterministic case, where all the λ always assign probabilities of 0 and 1, while q_x assigns probabilities between 0 and 1.

We realize that our result may appear to be circular. For we began with an ensemble in which q_x is the relative measure of states within the sub-ensemble picked out by *x*, and then we proved that the optimal theory should map *x* to q_x. We agree that if the ensemble were different, then the optimal theory would be different, which means is that the justification of the theory requires a justification of the ensemble. But this is not a fatal objection, for we have already described how the ensemble probabilities are tested empirically. And now we may tell the same story about the justification of the theory. Our result is only circular in the innocuous sense that we end up at the same place we began. It is not circular in the *logical* sense, because the theorem depends non-trivially on the choice of payoff function.

If the payoffs were *simple*, then q_x would always have a payoff of 0, and a theory that 'guessed' at the true λ would have a slightly greater payoff. Thus, ensemble averaging cannot be justified on the basis of simple payoffs. The K-L payoff function works, in part, because it takes account of the closeness of false hypotheses to the truth.

If 'closeness to the truth' were defined differently (for example, if the K-L payoffs were defined relative to a more narrow domain of prediction), then the optimal theory may be different. Far from being a weakness of our approach, it might explain why scientists make use of different theories for different purposes.

4 The Emergence Macroscopic Laws

In this section, we apply the definitions and theorems of the previous section to the example of a particle in a narrow cylinder, where we consider a very broad system type that allows the length of the cylinder (*L*), the area of the ends (*A*), the mass of the particle (*m*), and the total energy (*E*) to vary. Consider the equilibrium ensemble constructed as a weighted average of all the possible

microcanonical sub-ensembles of the system. Ω is the set of *all* possible trajectories of systems compatible with this general type. Now suppose that each x in the partition X determines a precise value for all of the parameters (L, A, m, E). The optimal theory for this ensemble is such that $q_x(\omega_i) = p_i$, for all i, where p_i denotes the microcanonical probability of the initial state λ_i *within its sub-ensemble*, and ω_i is the trajectory that evolves from λ_i. What are the properties of this optimal theory?

This is where the story becomes interesting. To *understand* a probabilistic theory of this kind we need to see how the different probabilistic hypotheses fit together. One thing we can do is to see if we can discover *laws* that connect to mean values of measurable quantities, such as pressure, with other macroscopically observable quantities such as the volume V and the total energy E.

One can discover one such law by calculating the mean pressure. We have already noted that the pressure within a small interval of time Δt will be either $2|p|/(A\Delta t)$ or 0, where the microcanonical probability of the positive pressure outcome is $\frac{1}{2}(p/m)(\Delta t/L)$, and the probability of the zero outcome is $1-\frac{1}{2}(p/m)(\Delta t/L)$. The ensemble average of the pressure is therefore

$$\overline{P} = \tfrac{1}{2}(p/m)(\Delta t/L)\,2|p|/(A\Delta t) = 2\left(p^2/2m\right)/(LA) = 2E/V\,,$$

where $V = LA$ is the volume of the cylinder. Thus, the *mean* pressure is not only independent of the time interval over of the measurement, and independent of the area over which the pressure is measured, and independent of the mass of the particle, but it is determined in a very simple way by two macroscopically accessible quantities—the total energy E and the volume of the container V.

In fact, if we were to broaden the system type to include any number of particles confined to the 1-dimensional container of arbitrary shape then we could still prove that

$$\overline{P}V = 2E\,.$$

If we further generalize the derivation to the case a 3-dimensional container, then the impacts of the molecules are distributed over its surfaces in 3 dimensions, which reduces the mean pressure by a factor of ⅓. We then arrive at the law

$$\bar{P}V = \tfrac{2}{3}E,$$

which is one way of writing the *ideal gas law*.

This law is a (partial) characterization of the macro-*theory*, as we have defined it, because it tells us how the (mean) pressure depends on x. In fact, as far as pressure is concerned, the ideal gas law is very nearly a complete characterization of what the theory says about pressure when the number of particles is large because the variance of the pressure is very nearly zero (provided that the pressure is averaged over an reasonably large area and time interval). That is why classical thermodynamics was successfully developed as non-statistical theory. Nevertheless, if we ask what can be derived *from the microphysics*, the answer is: A probabilistic *theory* in which the *mean* pressure is given by the ideal gas law.

It is philosophically important that this answer applies uniformly to cases in which the variance is negligible as well as to cases in which the variance is not negligible (such as in Brownian motion). For it is the uniformity of the account that allows one to apply the philosophical lessons drawn from one science to another.

In statistical mechanics, there is a clear distinction made between phase functions, which represent quantities that supervene on the microstate (such as the *observed* pressure) and ensemble averages such as the mean pressure. Phase functions are random variables[1], while ensemble averages are not. There

[1] A random variable is a term used in statistics to refer to any variable whose values are governed by a probability distribution.

are many ambiguities and confusions that result from the inattention to this distinction. For example, one often hears that the laws of thermodynamics, like the ideal gas law, are *statistical* laws. But does this mean that 'pressure' is a phase function, and that the ideal gas law is an approximation? Or does it mean that pressure is defined as an ensemble average, and is therefore statistical in that sense of being defined as a probabilistic average?

In our view, the ideal gas law relates the mean pressure (ensemble average) to the energy and volume of the gas—it is an *exact* law about ensemble *averages*. One might say that it is true *of the ensemble*, and this way of talking is harmless so long as it is clearly understood that it is not the same as saying that the law is true of the token system under consideration. The distinction is vitally important in cases where the variance of the phase function is not negligible (such as in Brownian motion). For when the variance is not negligible, one must refer to the full probabilistic theory to fully explain the statistics features of the phenomenon. And when the variance is nearly zero, that fact should be explained as well.

Statistical mechanics makes use of phase functions and ensemble averages alike—there is no exclusion of one in favor of the other. The conceptual role of 'pressure', *as it appears in the idea gas law*, is not to predict the observed pressure in a simple-minded way (even though it does a good job of that when the variance is small). Rather, its function is to describe a feature *of the probabilistic theory*. The theory is not true of the token system, even though it is optimal in a truth-related sense. It has a degree of predictive accuracy that is explained by the micro-theory, and it passes empirical tests in the same way as Kepler's theory of planetary motion coupled with an assumption about the distribution of observational errors. In both cases, the available empirical information (x) is unable to pick out the true trajectory of the system, and so the best theory is a probabilistic theory for the same reason.

One might complain that the discovers of the ideal gas law (*e.g.*, Boyle and Charles) did not make a distinction between phase functions and ensemble averages. Hence, the law derived from the macro-theory does not have the meaning intended by the discovers of the ideal gas law. Far from being an objection to our view, it supports what we are saying: The discovers of the ideal gas law may well have believed that it is a *true* description of nature, and therefore they may have *believed* that the quantities supervene on the underlying micro-state. But the fact remains that Newtonian mechanics does not explain why the ideal gas law is *true* of the token system, because in general it is not. Rather, Newtonian mechanics explains why the ideal gas law is empirically successful, and this explanation relies on a clear distinction between phase functions and ensemble averages.

'Temperature' is not only the most common example cited in discussions of scientific reduction, it is also different from the example of pressure in some interesting ways. It provides an important test case, which we consider next.

5 Temperature of an Ideal Gas

It would be a mistake to automatically assume that there is a single concept of 'temperature' in statistical mechanics. Like the case of gravitational mass and inertial mass, it may turn out that different concepts are equivalent in some sense, but this should not be assumed to be true *a priori*. In this section, we focus our attention on the concept of temperature as it applies to an ideal gas. This is the case that philosophers have in mind when they claim that temperature is identical to the mean kinetic energy, so the example is well worth considering.

We begin our story with a theorem about the equipartition of kinetic energy, which holds for microcanonical ensembles (so our discussion is also restricted to systems at *equilibrium*). The theorem states that mean kinetic energy is the same for each molecule in the gas, where 'mean' refers

to the microcanonical average.[2] The 'mean kinetic energy' does *not* refer to the total kinetic energy divided by the number of particles, for that is not what the equipartition theorem is about (and this point is crucial, as we shall see). The theorem implies that we may introduce a new quantity, $\overline{e}$, to denote the microcanonical average kinetic energy of an arbitrary particle. In an ideal gas, the particles have no potential energy except for the negligibly short periods of time during collisions (this is what distinguishes an ideal gas from a condensed gas, or a liquid, or a solid). So, for an ideal gas at equilibrium, we see that $\overline{e} = E/n$, where E is the total energy and n is the number of particles. Therefore, the ideal gas law can be re-written in the form

$$\overline{P}V = \tfrac{2}{3} n\overline{e} ,$$

where the bars remind us that the quantities are ensemble averages.

Now suppose two equilibrium gases are initially separated in volumes V_1 and V_2, respectively, by a thermally insulated partition (so that the gas particles cannot exchange energy through the partition). Then each gas is represented by a microcanonical ensemble, and by the equipartition of kinetic energies, the particles in each gas have the same mean kinetic energies as all other particles *on the same side of the partition*. Let these mean kinetic energies be $\overline{e}_1$ and $\overline{e}_2$, respectively, where $\overline{e}_1 \neq \overline{e}_2$. Consequently, $E_1/n_1 \neq E_2/n_2$, where E_1 and E_2 are the total energies of the two gases, and n_1 and n_2 are the numbers of particles in each gas.

---

[2] If some of the molecules are diatomic then they have rotational kinetic energy as well as translational kinetic energy ($\equiv \frac{1}{2} mv^2$). The equipartition theory says that the energy is equally partitioned according to the degrees of freedom, where a monatomic molecule has 3 degrees of freedom while a diatomic molecule has 5 degrees of freedom. The ideal gas temperature is then proportional to the mean kinetic energy per degree of freedom. We shall ignore this complication in what follows.

Now remove the sliding partition between the gases. After a sufficient period of time, the mixture of the gases is represented in terms of a new microcanonical ensemble. When we apply the equipartition theorem to the new ensemble, we infer that all gas particles have the same mean kinetic energy, which we denote by $\overline{e}$. By a simple calculation, $\overline{e}$ is related to $\overline{e}_1$ and $\overline{e}_2$ by the equation $\overline{e} = \left(n_1\overline{e}_1 + n_2\overline{e}_2\right)/\left(n_1 + n_2\right)$. Thus, the transition from the initial equilibrium state to the final equilibrium state is characterized by the evolution of the mean kinetic energy of molecules of the first gas from $\overline{e}_1$ to $\overline{e}$ and of the second gas molecules from $\overline{e}_2$ to $\overline{e}$.

This transition is more commonly described in terms of the concept of temperature, in the following way. First define the initial temperatures of the gases and their final temperature by the equations

$$kT_1 = \tfrac{2}{3}\overline{e}_1,\ kT_2 = \tfrac{2}{3}\overline{e}_2, \text{ and } kT = \tfrac{2}{3}\overline{e},$$

where *k* is Boltzmann's constant. (Note that according to these definitions, the ideal gas law may be re-written in its most familiar form, $\overline{P}V = nkT$.) If we now retell the previous story, the two gases start out at different temperatures, T_1 and T_2, and end up at a single final temperature

$$T = \left(n_1T_1 + n_2T_2\right)/\left(n_1 + n_2\right),$$

which is an arithmetic average of the initial temperatures. Thus, we have shown that two initially separated ideal gases with different temperatures will evolve towards an equilibrium state with an in-between temperature when they are brought into thermal contact. (The hotter gas will cool down, and the cooler gas will warm up.) This is a paradigmatic instance of the second law of thermodynamics.

In our view, temperature is *not defined as a phase function*, but in terms of an ensemble average. And so we claim that the temperature of an ideal gas is directly proportional to the mean kinetic energy only in the sense in which the mean refers to the microcanonical average of the kinetic energy of a *single* particle. The reference to a single particle does not present a problem because the mean kinetic

energies of all the particles are equal at equilibrium (by the equipartition theorem). The conceptual role of 'temperature' is to describe how the probabilistic hypotheses applied by the theory at different times fit together. 'Temperature' is similar to 'pressure' in this regard.

In order for 'temperature' to play its role, it is not *required* that it supervene on the microstate of the system. What makes this an interesting test case is that the ideal gas temperature does in fact supervene on the microstate of the system. For it is a numerical fact that $kT = \frac{2}{3} E/n$, where E is the total energy, and n is the number of particles in the gas. So, one *could* define the temperature of an ideal gas as a function of the microstate (in the trivial sense that E and n are constant functions of the microstate). Then temperature would be proportional to the mean kinetic energy in the other sense—namely, the sense in which the 'mean kinetic energy' is the numerical average of the *actual* kinetic energies of *all* the particles in the system.

It is *not* our view that this alternative concept of temperature is illegitimate. It may have a role to play. But it *is* our view that it cannot *replace* the concept of temperature qua ensemble average. A preliminary point is that if temperature were defined by the equation $kT = \frac{2}{3} E/n$, then the composite system in our previous example would attain its equilibrium temperature *immediately* after the partition is removed. That is, the phase function concept pays no heed to the idea that final temperature of the system evolves from an exchange of energy between the hotter gas and the cooler gas, and that this takes some time.

It would also leave physicists with no way of predicting the temperature in many important applications. For suppose that divider between V_1 and V_2 is re-inserted after the mixed gases have reached equilibrium. What is the temperature of the re-separated gases? The phase space temperature would depend on number and energies trapped on each side of the partition, and this will vary from one token system to the next. But the thermodynamical concept of temperature, if it is to serve its purpose,

should abstract away from details of the positions and energies of the molecules at any particular time. The ensemble average concept of temperature serves this purpose. It doesn't matter if a disproportionate number of molecules are trapped on one side of the partition after the divider is re-inserted. The mean kinetic energies of all particles are the same (by the equipartition theorem) no matter where they are located, so the temperatures of re-separated gases is the same as the final equilibrium temperature. Its value is determined independently of micro-state.

A related point is that temperature defined as an ensemble average can be *manipulated*, while temperature in the phase function sense cannot. For example, if a system with an unknown temperature is placed in contact with a large thermal reservoir (defined as a system with an extremely large number of particles) with temperature T, then we can say that the final equilibrium temperature of the smaller system is also T (because $\left(n_1T_1 + n_2T\right)/\left(n_1 + n_2\right) \simeq T$ if n_2 is much greater than n_1). The manipulation results from the tendency of systems to exchange energies in a way that ensures that an equilibrium ensemble is predictively accurate in a sense of our definition.

If there is to be a mutually exclusive choice between phase function and ensemble average definition of 'temperature', then physicist have voted with their feet. We have attempted to explain why the physicists' choice makes sense not merely in terms of economy of effort, but also in terms the best job that can be done in terms of the truth-related payoff of predictive accuracy (the Kullback-Leibler payoff) on the basis of the *available* evidence (x).

6 Concluding Remarks

Our project has been to understand the relationship between the micro-theories and macro-theories. The argument against there being a reductive relation between them is summarized as follows:

1. Macro-theories abstract away from the micro-details.

2. If a macro-theory is true, then the quantities mentioned in its laws are real.
3. If the quantities are real, then they are either definable in terms of micro-quantities or they refer to emergent properties.
4. If the quantities are definable in terms of micro-quantities, then the macro-theory does not abstract away from the micro-details.
5. If the quantities are emergent, then the micro-theory cannot explain the truth of the macro-theory.
6. *Therefore*, the micro-theory cannot explain the truth of the macro-theory.

The weakest link in the argument is premise 4. For it may be that quantities like 'pressure' and 'temperature' are defined in a way that successfully abstracts away from the micro-details, while they are definable in terms of micro-quantities *at the same time*. We grant that this is a logical possibility. In fact, Boltzmann proposed a hypothesis in 1871 (Brush 1983, p. 66), called the ergodic hypothesis (not to be confused with the ergodic theorem)[3], which would have solve the problem if it were only true (of more than a few systems). In Boltzmann's formulation, the ergodic hypothesis states that "a mechanical system will eventually pass through *all* microstates before returning any microstate a second time" (Brush 1983, p. 66). If the ergodic hypothesis were true, then the long-run time averages of any micro-quantity would be the same for every microstate, and therefore equal to their microcanonical average (Khinchin 1949, chapter III). Since a *time* average supervenes on the microstate (because the microstate determines the system's trajectory), it would have followed that the

[3] Reichenbach (1956, p. 78) is guilty of this confusion when he cites a proof of the ergodic theorem to justify the claim that we have "a derivation of the probability metric [the microcanonical distribution] from causal laws alone." For those who wish to learn more about the ergodic *theorem*, we recommend von Plato (1982).

microcanonical averages supervene on the microstate. So, there was a time when equilibrium thermodynamics looked like an example of reduction in the classical sense.

However, the most recent consensus is that the ergodic hypothesis is false (Earman and Rédei 1996, Batterman 1998), and so there is no such argument for reductionism in the traditional sense. Moreover, the truth of the ergodic hypothesis would only have ensured the supervenience of *microcanonical* averages, leaving the use of non-equilibrium ensembles unexplained.

Callender (1997) recognizes the failure of the ergodic hypothesis, and the success of the ensemble approach to statistical mechanics, but recommends that physicists devote more attention to Boltzmann's problem in the hope of repudiating the anti-reductionist position.

In the meantime, there is another philosophical problem that has not received enough attention: If statistical mechanical quantities are not actually defined in terms of micro-quantities *at the present time* (if there are no known 'bridge laws' that succeed in bridging the logical gap between micro-theories and macro-theories), then the success of the macro-theory is not *presently* explained by the micro-theory. So, what kind of explanatory success has statistical mechanics *achieved* during the last century? To this, a traditional kind of reductionist would have to say 'nothing very much at all'.

Our response has been to shift the focus away from explaining the *truth* of macro-theories towards explaining their *predictive accuracy*. For this goal does not *require* that the quantities like 'pressure' and 'temperature' supervene on the microstate. In terms of this weaker goal, we claim that statistical mechanics has achieved a great deal, and we have supported this claim by analyzing a simple, but realistic example.

References

Batterman, Robert W. (1998), "Why Equilibrium Statistical Mechanics Works: Universality and the

Renormalization Group", *Philosophy of Science*, 65: 183-208.

Beatty, John (1984), "Chance and Natural Selection", *Philosophy of Science*, 51: 183-211.

Bell, John S. (1964), "On the Einstein-Podolsky-Rosen Paradox", *Physics* 1: 195-200.

Blackwell, David (1953), "Equivalent Comparisons of Experiments", *Annuls of Mathematical Statistics*, 24: 265-272.

Brandon, Robert and Scott Carson (1996), "The Indeterministic Character of Evolutionary Theory: No 'No Hidden Variable Proof, but No Room for Determinism Either", *Philosophy of Science*, 63: 315-337.

Brush, Stephen G. (1983), *Statistical Physics and the Atomic Theory of Matter, from Boyle and Newton to Landau and Onsager*, Princeton University Press.

Callender, Craig (1997), "Reducing Thermodynamics to Statistical Mechanics: The Case of Entropy", *Journal of Philosophy*, XCVI, July 1999, 348-373.

Earman, John and M. Rédei (1996), "Why Ergodic Theory Does Not Explain the Success of Equilibrium Statistical Mechanics", *British Journal for the Philosophy of Science*, 47: 63-78.

Eells, Ellery (1991), *Probabilistic Causality*, Cambridge: Cambridge University Press.

Forster, M. R. and Elliott Sober (1994), "How to Tell when Simpler, More Unified, or Less *Ad Hoc* Theories will Provide More Accurate Predictions", *British Journal for the Philosophy of Science*, 45: 1-35.

Gibbs, J. Willard (1902; reprinted 1981), *Elementary Principles in Statistical Mechanics*, Woodbridge, Conn.: Ox Bow Press.

Graves, Leslie, Barbara L. Horan, and Alex Rosenberg (1999), "Is Indeterminism the Source of the Statistical Character of Evolutionary Theory," *Philosophy of Science*, 66: 140-157.

Glymour, Bruce (2001), "Selection, Indeterminism and Evolutionary Theory", *Philosophy of Science*, 68: 536-544.

Hausman, Dan M. (1998): *Causal Asymmetries*. Cambridge: Cambridge University Press.

Horan, Barbara L. (1994), "The Statistical Character of Evolutionary Theory", *Philosophy of Science*, 61: 76-95.

Khinchin, A. I. (1949), *Mathematical Foundations of Statistical Mechanics*, translated from the Russian by G. Gamov, Dover Publications, Inc., New York.

Kitcher, Philip (1984), "1953 and All That. A Tale of Two Sciences", *The Philosophical Review*, 93: 335-375.

Kullback, S. and R. A. Leibler (1951), "On Information and Sufficiency", *Annals of Mathematical Statistics*, 22: 79-86.

Mills, S. and J. Beatty (1979), "The Propensity Interpretation of Fitness," *Philosophy of Science*, 46: 263-286.

Nagel, E. (1961), *The Structure of Science*, New York: Harcourt, Brace and World.

Pearl, Judea (2000), *Causality*: *Models, Reasoning, and Inference*, Cambridge: Cambridge University Press.

Reichenbach, Hans (1956), *The Direction of Time*, Berkeley: University of California Press.

Rieke, Fred, D. Warland, R. van Steveninck and W. Bialek (1997), *Spikes*: *Exploring the Neural Code*, Cambridge, Mass.: MIT Press.

Rosenberg, Alex (1985), "Is the Theory of Natural Selection a Statistical Theory?", *Canadian Journal of Philosophy* (Suppl.), 14: 187-207.

Rosenberg, Alex (2001), "Discussion Note: Indeterminism, Probability and Randomness in Evolutionary Theory", *Philosophy of Science*, 68: 536-544.

Sober, Elliott (1984), *The Nature of Selection: Evolutionary Theory in Philosophical Focus*, Cambridge, Mass: MIT Press.

van Fraassen, Bas (1980), *The Scientific Image*, Oxford: Oxford University Press.

von Plato, Jan (1982), "The Generalization of de Finetti's Representation Theorem to Stationary Probabilities", *PSA 1982*, Volume 1, East Lansing, Michigan: Philosophy of Science Association.

Waters, C. Kenneth (1990), "Why the Anti-reductionist Consensus Won't Survive the Case of Classical Mendelian Genetics", in A. Fine, M. Forbes, and L. Wessels (eds.) *PSA 1990*, Vol. 1: 125-139.