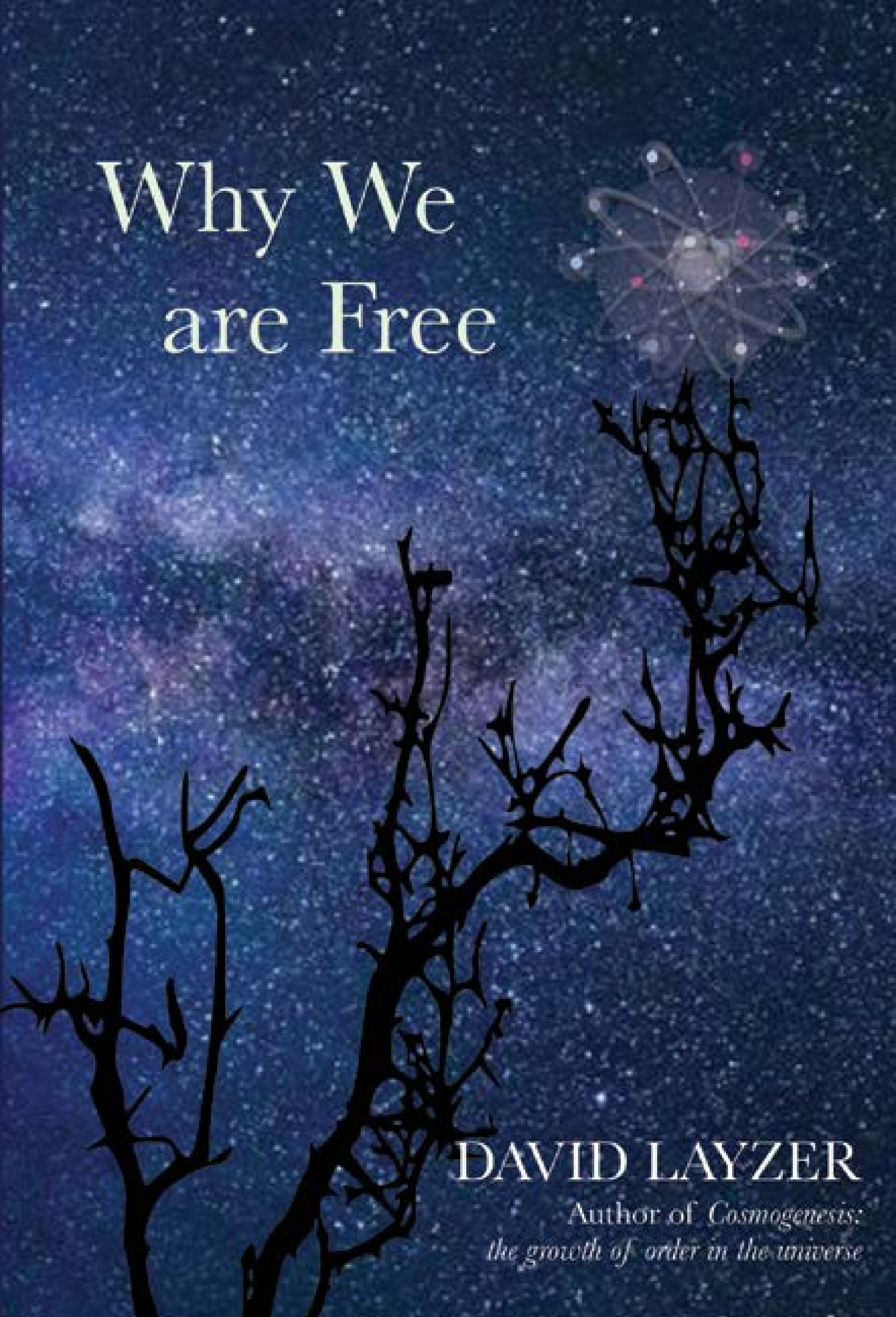

Why We
are Free
DAVID LAYZER
Author of *Cosmogenesis:*
the growth of order in the universe

Why We are Free

Why We are Free

Consciousness, Free Will and Creativity in
A Unified Scientific Worldview

DAVID LAYZER
Author of *Cosmogenesis:*
The Growth of Order in the Universe
First Donald H. Menzel
Professor of Astrophysics
Harvard University

Anthony Aguirre and Bob Doyle, Editors

First edition, 2021
© 2019, David Layzer

All rights reserved. No part of this book may be reproduced in any form by electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without the prior permission of The Information Philosopher.

Publisher's Cataloging-in-Publication Data
(Prepared by The Donohue Group, Inc.)

Names: Layzer, David, 1925-2019, author. |
Aguirre, Anthony, 1973- editor. | Doyle, Bob, 1936- editor.

Title: Why we are free : consciousness, free will and creativity in a unified scientific worldview / David Layzer, first Donald H. Menzel Professor of Astrophysics, Harvard University ; Anthony Aguirre, UC San Diego and Bob Doyle, Harvard University, editors.

Description: First edition. | Cambridge, MA, USA : I-Phi Press, 2021. | Includes bibliographical references and index.

Identifiers: ISBN 9780983580256 (paperback)

Subjects: LCSH: Free will and determinism. |
Cosmology. | Philosophy and science.

Classification: LCC BJ1461 .L39 2021 | DDC 123.5--dc23
1. Free Will. 2. Cosmology. 3.Philosophy. I. Title.

I-Phi Press
Cambridge, MA, USA

For my undergraduate and graduate students,
from whom, over many decades,
I learnt more than they learned from me

Preface

David Layzer was an astrophysicist and cosmologist at Harvard who from the 1960s onward generated key insights into the evolution and structure of the universe, many of which arguably remain under-appreciated. He was among the first to carefully and incisively analyze the arrow of time in cosmology, he formulated a subtle, distinct, and fascinating version of the "cosmological principle," and he provided key insights into the growth of both *entropy* and *order* through cosmic evolution. In his later career he focused on the connection between cosmic foundations of randomness, the growth of order, and the creativity and freedom inherent in biological evolution and mental processes. *Why We are Free* ties all of these threads together. The work speaks for itself, but readers may find some historical and scientific contextualization of each of them useful.

Layzer's cosmological work began at a time when our view of cosmic history was far less settled. In the early 1960's he was invited to a conference on the nature of time at Cornell University organized by the founders of the steady-state theory of the universe, Hermann Bondi, Tommy Gold, and Fred Hoyle.

The steady-state founders went beyond Newton's "cosmological principle" that space is homogeneous and isotropic, the same in all places and in all directions. They added their "perfect cosmological principle," claiming the universe is eternal and appears the same at all times. This appeared to eliminate what Arthur Stanley Eddington in 1928 called "Time's Arrow," which Eddington had associated with the apparently *irreversible* increase in entropy demanded by the second law of thermodynamics.

At the Cornell conference Layzer proposed his "strong cosmological principle," that the homogeneity and isotropy of matter and space applies both statistically and exactly. The subtle but foundational implication of this assumption is that the cosmological principle is not just a convenient approximation to a more precisely knowable cosmic state, but that the statistical description is complete and cannot be improved upon. This enormously reduces, for example, the information required to specify the initial cosmic state.

Subsequently, Layzer argued that the strong cosmological principle implies a fundamental uncertainty inherent in any finite physical system: because the statistical description of the universe is complete, the precision with which any subsystem can be described is limited. He called this implication "primordial randomness." He later came to believe that this very uncertainty is at the root of quantum uncertainty. Sketched in his book *Cosmogenesis*, this idea was revisited by Layzer[1], and by Max Tegmark and one of us[2], eventually splitting into two versions of a "Cosmological interpretation of quantum mechanics, " similar in spirit but different in details and context.

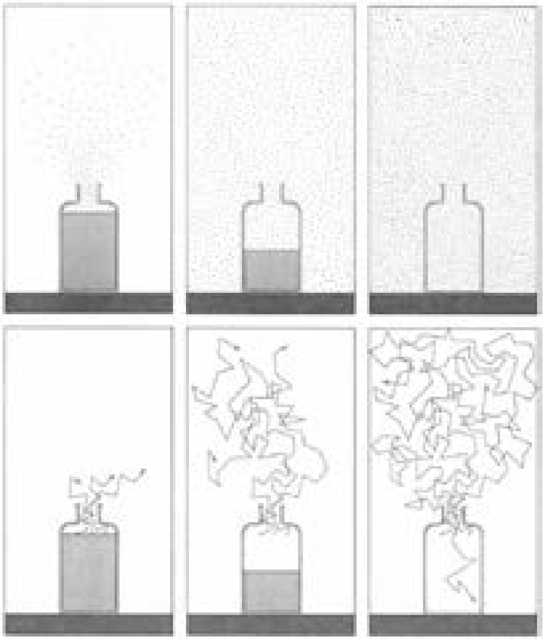
Layzer also showed in detail (particularly in the Journal article) the different types of arrows of time, and some of their relations: alongside Eddington's identification of increasing entropy as the fundamental time's arrow, Layzer now added another arrow he called the "historical arrow of time; this was in addition to the so-called "radiation arrow" or "electromagnetic arrow" (the asymmetry between emission and absorption processes) and the so-called "cosmological arrow" (the expansion of the universe).

The strand of Layzer's work addressing the "arrow of time" also began early in the origins of modern cosmology. The two most important publications were in *Scientific American* in 1975 and the *Astrophysical Journal* in 1976. In the former article, Layzer introduced his famous example of the perfume bottle being opened and the perfume dispersing into the room.

1 See https://arxiv.org/abs/1008.1229 for Layzer's elegant and detailed exposition.

2 Aguirre, Anthony, and Max Tegmark. "Born in an infinite universe: a cosmological interpretation of quantum mechanics." Physical Review D 84.10 (2011): 105002.

Layzer likened the time-reversed process to a movie played backwards. If the "initial" state contained "hidden" microscopic information corresponding to a reversal of the momentum of each perfume molecule at a certain time, classical dynamics says that the perfume will make its way back into the bottle. This is a powerful visual image illustrating Josef Loschmidt's reversibility objection to Ludwig Boltzmann's *H*-Theorem.



Source: "The Arrow of Time" *Scientific American*, December, p.57 (1975)

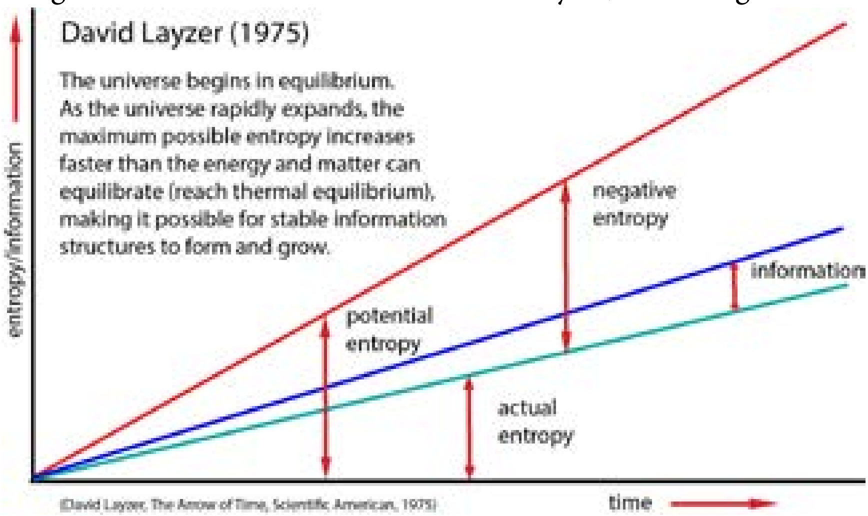
In Layzer's view, primordial randomness precludes just the sort of "hidden information" that would allow such a reversal.

Beyond its clarity, what most distinguished Layzer's thinking about time's arrow in cosmology was his resolution of a very basic paradox: the universe appears to start in a dense, high-energy, equilibrium state. Yet the second thermodynamic law implies that entropy increases. So how does the universe ever leave equilibrium and so generate all of the ordered and highly non-equilibrium structures it clearly contains? Layzer's resolution, in his words from the 1975 article:

> Suppose that at some early moment local thermodynamic equilibrium prevailed in the universe. The entropy of any region would then be as large as possible for the prevailing values of the mean temperature and density. As the universe expanded from that hypothetical state the local values of the mean density and temperature would change, and so would the entropy of the region. For the entropy to remain at its maximum value (and thus for equilibrium to be maintained) the distribution of energies allotted to matter and to radiation must change, and so must the concentrations of the various kinds of particles. The physical processes that mediate

> these changes proceed at finite rates; if these "equilibration" rates are all much greater than the rate of cosmic expansion, approximate local thermodynamic equilibrium will be maintained; if they are not, the expansion will give rise to significant local departures from equilibrium. These departures represent macroscopic information; the quantity of macroscopic information generated by the expansion is the difference between the actual value of the entropy and the theoretical maximum entropy at the mean temperature and density.[3]

This core and clear insight — that information is the gap between realized and possible entropy, and that the latter increases in an expanding universe — has even now not fully penetrated the thinking of many working in the field. Going further, Layzer persuasively argues that this dynamic, in which would-be equilibrating interactions fail to maintain equilibrium, pervades and underlies the growth of order in the universe. For Layzer, this along with the



strong cosmological principle provided a clear explanation of the arrows of time, with the cosmological principle ensuring statistical equilibrium at the earliest time, and the time-varying rates of various processes during cosmic evolution doing the rest. Again in his words:

> As a result the big bang is an exceedingly gentle process; local equilibration processes easily keep pace with the changing macroscopic conditions of temperature and density during the first fraction of a microsecond. It is only for this brief initial phase in

3 "The Arrow of Time" *Scientific American*, December, p.68 (1975)

> the evolution of the universe that local thermodynamic equilibrium can be assumed, but from that assumption it follows that the expansion of the universe has generated both macroscopic information and entropy. Thus the cosmological arrow, the historical arrow and the thermodynamic arrow all emerge as consequences of the strong cosmological principle and the assumption that local thermodynamic equilibrium prevailed at or near the initial singularity. Remarkably, neither of these assumptions refers directly to time or temporal processes.[4]

For Layzer, these cosmological considerations tied directly to basic issues about the nature of complex systems including biological ones. He explored these ties in detail in his 1990 book, *Cosmogenesis*, which lays out a full vision for how the enormous amount and quality of structure comes into being, first through cosmic processes and then via life and evolution that can "utilize" cosmic and astrophysical order. That book also began Layzer's study of free will, which through his arguments he connected to randomness in the earliest time.

In his later years, Layzer developed through a series of talks and unpublished papers a compelling view of how the freedom and creativity of living and mental systems can coexist with the (seemingly) rigid determinism of natural law. It has been argued in many places that randomness — quantum or otherwise — does not bear upon free will. Layzer forcefully disputes this, arguing that the same randomness and growth of order that pervades the universe underlies living phenomena, with the randomness providing the fuel (but not in itself an explanation) for genuine novelty and choice.

In *Why We Are Free*, his last work (especially page 153), we see Layzer 's concise telling of this entire story.

Anthony Aguirre, UC Santa Cruz

Bob Doyle, Harvard University

January, 2021

4 "The Arrow of Time" *Scientific American*, December 1975, p.57

Contents

What This Book Is About

No biological capacity is more distinctively human than our ability to shape future events in ways that accord with our needs, wants, and desires. Yet many mainstream scientists and philosophers deny the existence of such a capacity. They subscribe to a scientific worldview called physicalism, or materialism, according to which reality comprises just those objects and relations that figure in our strongly confirmed physical laws. Francis Crick, co-discoverer with James Watson of DNA's double helix, summarized this idea in the opening paragraph of his book *The Astonishing Hypothesis, the Scientific Search for the Soul*:. It is the alien idea that each person is no more than a vast assembly of nerve cells and their molecules, that we are "nothing but a pack of neurons.

This book describes an alternative to physicalism in which our joys and our sorrows, our memories and our ambitions, our sense of personal identity and free will, are just as real as the objects and relations of the world physics describes. Like physicalism, the proposed worldview rests on our strongly confirmed physical laws. I argue that these laws don't need to be supplemented by additional laws that expand the domain of natural science to include consciousness and subjectivity. Physicalism, I argue, rests on Newton's assumption in The Principia that place is absolute. This assumption implies, for instance, that the present value of the mass per unit volume at the Sun's position has a definite value (though one we can't know with infinite precision).

This book proposes to replace the assumption that place is absolute by a pair of less obvious cosmological assumptions: The first of these, the strong cosmological principle, says that a complete description of the Universe doesn't privilege any position or direction in space. It implies that a complete description of the Universe wouldn't tell us the values of macroscopic quantities. Instead it would specify probability distributions of the possible values of macroscopic quantities. It would interpret the probability of any given range of possible values of a macroscopic quantity as the relative frequency of that range in an infinitely extended sample of the Universe. The second assumption, the assumption of primordial randomness, says that the probability distributions that characterize the Universe's initial state were maximally random. Randomness in this context is synonymous with statistical entropy, a property of probability distributions introduced in 1872 by Ludwig Boltzmann as the counterpart in the atomic theory of gases to Rudolf Clausius's thermodynamic entropy. The assumption of primordial randomness implies that the Universe has evolved from a state of complete disorder, so it contradicts the widely believed claim that the entropy of the Universe never decreases. This book argue that this generalization of the strongly confirmed law of entropy non-decrease for undisturbed macroscopic systems is untenable.

Chance and the openness of the future are pervasive features of prescientific views of the natural world. Physicalism tells us that they are illusory. In a scientific worldview that incorporates the strong cosmological principle and the assumption of primordial randomness chance is indeed pervasive and the future is indeed largely open. But how do physics and biology themselves fare under the new scientific dispensation?

I will argue that the two cosmological hypotheses allow a more unified view of macrophysics, microphysics, and statistical physics (which links macrophysics to microphysics). The implications of the two cosmological hypotheses for biology are more striking. Physicalism purports to reduce biology to physics. But physicalism attributes macroscopic chance to human ignorance. By contrast, macroscopic chance plays a central role in evolution, both in genetic variation (for example in genetic recombination) and natural selection (think of extinction events). Consequently, although biological processes are governed by physical laws, biology doesn't reduce to physics. I will argue that the strong cosmological principle and the assumption of primordial randomness together with our strongly confirmed physical laws provide a framework for modern evolutionary theory. They also provide a framework for understanding consciousness, creativity, and free will as biological phenomena.

I
Physics, Biology, and Physicalism

A physical theory needn't be intuitively plausible, but it must make testable predictions that agree with experiment or observation. For example, quantum mechanics describes electrons as mathematical points endowed with mass and electric charge. It assigns these points spatial coordinates and asserts that when an electron is in a definite physical state its coordinates don't have definite values. None of this makes intuitive sense. How could a mathematical point have a finite mass? How could a point's coordinates not have definite values? Nevertheless, unambiguous rules link quantum mechanics' abstract mathematical laws to the outcomes of a wide range of experiments, and the close agreement between the observed and predicted outcomes of these experiments leaves no room for doubt that the laws are, at least, very nearly correct. Experience isn't just the arbiter of scientific theories, though. It is itself a subject of scientific investigation. Neuroscientists investigate how perception, memory, cognition, and other mental states and processes are linked to physical structures and processes in the brain. Such studies have produced convincing evidence that the mathematical laws that govern physical phenomena also govern the biological processes that accompany mental processes.

But mental processes seem to be more than the physical processes that accompany them. All experience is someone's experience. My experience of seeing a red ball is distinct from your experience of seeing what we both believe to be the same ball. It may seem obvious that this aspect of experience – consciousness – is as much a part of the natural world as matter and energy. Yet there is no scientific or philosophical consensus about how – or even whether – consciousness fits into a scientific picture of the world.

Some scientists and philosophers argue that the subjective aspect of consciousness – the aspect that makes your consciousness different from mine – isn't part of a scientific picture of the world. Francis Crick, co-discoverer with James Watson of DNA's double helix, summarized this position in the opening paragraph of his book *The Astonishing Hypothesis, the Scientific Search for the Soul*:

> The Astonishing Hypothesis is that "You," your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules. As Lewis Carroll's Alice might have phrased it: "You're nothing but a pack of neurons." This hypothesis is so alien to the ideas of most people alive today that it can truly be called astonishing.[1]

Crick's astonishing hypothesis implies that biology not only rests on physics – something few if any contemporary biologists would dispute – but also reduces to physics: the natural world contains just those entities mentioned in our strongly confirmed physical theories, nothing more. I'll refer to this scientific worldview as *physicalism*.

Among the features of conscious life that physicalism reduces to the behavior of assemblies of brain cells is free will: our felt ability to alter the course of events through our deliberate acts. Though habits and deep-seated preferences dictate many of the choices we make in our daily lives, we take it for granted that the actions that flow from decisions we have thought about long and hard help shape the future. We think the world would have been different in ways that matter to us if we had decided and acted differently. Law, ethics, and widely held opinions about the aims of education all presuppose that our deliberate actions spring from a kind of freedom other animals don't enjoy. Our laws punish theft, but we don't regard a dog that steals another dog's bone as a lawbreaker. We think torture is wrong, but we don't think a cat is acting unethically when it tortures a mouse. We train our pets, but we believe children should

1 Crick, Francis. *The Astonishing Hypothesis* (New York: Simon and Schuster 1994) p. 3

be brought up not just to behave in certain ways but also to make good decisions. None of these beliefs and attitudes would be tenable if we weren't capable of making choices and decisions that affect the future course of events. Yet decision-making is a biological process, and biological processes are also physical processes, governed by the physical laws that prevail in physics and chemistry laboratories and in stars and galaxies. These laws are deterministic in the following sense: They connect a complete description of the present physical state of an isolated, or undisturbed, physical system to a complete description of any of the system's past and future states. The laws of quantum physics are just as deterministic in this sense as those of classical, or pre-quantum, physical theories – theories that deal with macroscopic and astronomical phenomena. Quantum mechanics' law of change, like its classical counterparts, links the present state of an undisturbed physical system to each of the system's future states.

Unlike classical states, quantum states of undisturbed systems aren't directly observable. The standard formulation of quantum mechanics links them to measurement outcomes by a rule that Paul Dirac in *The Principles of Quantum Mechanics* called a "general assumption." Given a system's quantum state, this rule (discussed in more detail below) enables one to calculate not only the possible outcomes of a measurement of any of the system's physical properties, such as position or momentum or energy, but also the relative frequency of each possible outcome in a long series (or large collection) of identical measurements. This feature of quantum mechanics – the unpredictability of individual measurement outcomes, along with the predictability of possible measurement outcomes and their relative frequencies in long series of identical measurements – is what physicists mean by "quantum indeterminism."

The philosopher John Searle assumes, as do many or perhaps most physicists, that "all indeterminism in nature is quantum indeterminism."[2] He then argues that "consciousness is a feature of nature that manifests indeterminism" and concludes that quantum indeterminism must underpin free will. While I agree with Searle that free will requires "indeterminism in nature," I think the conclusion that quantum indeterminism underpins free will is untenable, for two reasons. First, as far as I know, nothing in neuroscience suggests that

2 Searle, John R. *Freedom and Neurobiology* (New York, Columbia University Press, 2004) p. 74

the neural processes accompanying decision-making are, or resemble, quantum measurements. Second, quantum measurements, unlike free acts, never have novel and unpredictable outcomes; their possible outcomes are entirely predictable. Yet if we have the ability to help shape future events through our deliberate actions, as I believe we do, the future must in some ways be open. Either free will is an illusion or determinism is false.

Like Crick, the evolutionary biologist Edward O. Wilson has embraced the first alternative:

> The self, an actor in a perpetually changing drama, lacks full command of its own actions. It does not make decisions solely by conscious, purely rational choice. Much of the computation in decision-making is unconscious – strings dancing the puppet ego. Circuits and determining molecular processes exist outside conscious thought. They consolidate certain memories and delete others, bias connections and analogies, and reinforce the neurohormonal loops that regulate subsequent emotional response. Before the curtain is drawn and the play unfolds, the stage has already been partly set and much of the script written.
>
> The hidden preparation of mental activity gives the illusion of will. We make decisions for reasons we often sense only vaguely, and seldom if ever understand fully. Ignorance of this kind is conceived by the conscious mind as uncertainty to be resolved; hence freedom of choice is ensured. An omniscient mind with total commitment to reason and fixed goals would lack free will.

But if free will is illusory, why is the illusion so strong? Wilson explains:

> …Confidence in free will is biologically adaptive. Without it the mind, imprisoned by fatalism, would slow and deteriorate. Thus in organismic time and space, in every operational sense that applies to the knowable self, the mind does have free will.[3]

Psychologist Daniel Wegner agrees with Wilson that free will is an illusion. He argues that the neural processes underlying voluntary actions have two distinct outcomes: the action itself and awareness of willing the action. Is this awareness the cause of the action? From a thorough examination of the evidence bearing on this question, Wegner concludes the answer is no:

3 Wilson, E.O. *Consilience*, (New York, Knopf: 1998) p. 130

> It usually seems that we *consciously* will our voluntary actions, but this is an illusion. ... *Conscious* will arises from processes that are psychologically and anatomically distinct from the processes whereby mind creates action. (pages1-2) … If a team of scientific *psychologists* … somehow had access to all the information they could ever want, … they *could* uncover the mechanisms that give rise to all your behaviors …. [4]

Behind this thought is the image of a complex physical system, such as the solar system, whose physical states are completely determined by physical laws and antecedent conditions. Indeed Wegner ends his book with a quotation from Einstein:

> If the moon, in the act of completing its eternal way around the earth were gifted with self-consciousness, it would feel thoroughly convinced that it was traveling its way of its own accord. ...So would a Being, endowed with higher insight and more perfect intelligence, watching man and his doings, smile about man's illusion that he was acting according to his own free will.[5]

The philosopher Thomas Nagel has argued that physicalism can't be reconciled with the existence of consciousness:

> The existence of consciousness seems to imply that the physical description of the universe, in spite of its richness and explanatory power, is only part of the truth, and that the natural order is far less austere than it would be if physics and chemistry accounted for everything. If we take this problem seriously, and follow out its implications, it threatens to unravel the entire naturalistic world picture. Yet it is very difficult to imagine viable alternatives."[6]

The subtitle of *Mind and Cosmos* states that "the materialist neo-Darwinian conception of nature is almost certainly false."

If evolutionary theory is a purely physical theory, then it might in principle provide the framework for a physical explanation of the appearance of behaviorally complex animal organisms with central nervous system. But subjective consciousness, if it is not reducible to something physical, would not be part of this story; it would be left completely unexplained by physical evolution – even if the physical

4 Wegner, D.M. *The Illusion of Conscious Will*. (Cambridge: MIT Press 2002) p. 29

5 Einstein, A. 1995. Quoted in Home, D. and A. Robinson, "Einstein and Tagore: Man, nature and mysticism", *Journal of Consciousness Studies* 2, 167-169

6 Nagel, Thomas, *Mind and Cosmos*: *Why the Materialist Neo-Darwinian Conception of Nature Is Almost Certainly False* (Cambridge, Belknap Press of Harvard University) p. 35

evolution of such organisms is in fact a causally necessary and sufficient reason for consciousness.

Yet Nagel doesn't argue that consciousness lies outside the scope of natural science. Rather, he suggests, "[i]t makes sense to seek an expanded form of understanding that includes the mental but that is still scientific – i.e. still a theory of the immanent order of nature."[7]

Ernst Mayr, an architect of contemporary evolutionary theory, would have agreed. He wrote that "living systems ... have numerous properties that are simply not found in the inanimate world."[8]

> ...I believe that a unification of science is indeed possible if we are willing to expand the concept of science to include the basic principle and concepts of not only the physical but also the biological sciences. Such a new philosophy of science will need to adopt a greatly enlarged vocabulary – one that includes such words as biopopulation, teleonomy, and progress. It will have to abandon its loyalty to a rigid essentialism and determinism in favor of a broader recognition of stochastic processes, a pluralism of causes and effect, the hierarchical organization of much of nature, the emergence of unanticipated properties at higher hierarchical levels, the internal cohesion of complex systems, and many other concepts absent from – or at least neglected by – the classical philosophy of science.[9]

Mayr emphasized the role of chance in evolution:

> Evolutionary change in every generation is a two-step process: the production of genetically new individuals and the selection of the progenitors of the next generation. The important role of chance at the first step, the production of variability is universally acknowledged, but the second step, natural selection, is on the whole viewed rather deterministically: Selection is a non-chance process. What is usually forgotten is the important role chance plays even during the process of selection. In a group of sibs it is by no means necessarily only those with the most superior genotypes that will reproduce. Predators mostly take weak or sick prey individuals, but not exclusively, nor do localized natural catastrophes (storms, avalanches, floods) kill only inferior individuals. Every founder population [the parent population of

7 Nagel, Thomas, "The Core of '*Mind and Cosmos*,'" New York Times, August 18, 2013

8 Mayr, E., *Toward a New Philosophy of Biology*. (Cambridge: Harvard University Press, 1988) p. 1

9 ibid, p. 21

> a new species] is largely a chance aggregate of individuals, and the outcome of genetic revolutions, initiating new evolutionary departures, may depend on chance constellations of genetic factors. There is a large element of chance in every successful colonization. When multiple pathways toward the acquisition of a new adaptive trait are possible, it is often a matter of a momentary constellation of chance factors as to which one will be taken.[10]

Physicalism's picture of the natural order has no room for the kinds of pervasive and objective randomness that evolution requires, according to Mayr and other evolutionary biologists.

Physicalist views of evolution characterize it as a mechanical process – a process whose outcomes are predictable, at least in principle. Some evolutionary biologists argue that evolution is a creative process – one that brings into being novel and unpredictable forms of organized complexity. This was the central theme of the philosopher Henri Bergson's masterwork, *L'Évolution créatrice* (*Creative Evolution*, 1907). Bergson's books combined philosophical and scientific erudition with an engaging and persuasive literary style. They won him a wide and enthusiastic readership. In 1927 he was awarded the Nobel Prize for Literature. But Bergson didn't believe that creative biological processes are consistent with well-established physical laws. He argued that physical laws couldn't govern either evolution or consciousness because they rest on an oversimplified characterization of time. Physics represents time by a line and moments in time by points on the line. In his doctoral dissertation *Time and Free Will* Bergson argued that this static representation deprives time of its essential dynamic character. As time manifests itself in creative biological contexts, including evolution and free human acts, it is a dynamic process, "pure becoming." Evolution, Bergson argued, transcends scientific description. It is driven by an *élan vital* or vital impulse. Because that impulse informs all living organisms, Bergson argued, in *Introduction to Metaphysics*, that we can grasp it through an effort of intuition even if we can't describe it in the language of science.

The *élan vital* didn't survive twentieth-century advances in our understanding of biology. Experimental and observational evidence leaves little room, if any, for doubting that physical laws apply across the board – to living organisms and physical systems alike. Biology's

10 Mayr, E. 1983. *The American Naturalist* 121: 324-33; reprinted in *Toward a New Philosophy of Biology*. pp. 148-159

current picture of the living world also dispenses with "biotonic" laws – laws that apply only to living organisms.

But while Bergson's interpretation of evolution is no longer tenable, his description of it as a creative process remains compelling. Embracing that description while rejecting Bergson's metaphysics, Theodosius Dobzhansky, like Mayr a principal architect of current evolutionary theory, argued that evolution is indeed "a creative process, in exactly the same sense in which composing a poem or a symphony, carving a statue, or painting a picture are creative acts":

> Can the word 'creative' be validly applied to a process that has no foresight and no ability to devise means to a chosen goal? …Evolution [like artistic creation] not only brings novelties into being, but these novelties present embodximents of new ways of life. …Every new form of life that appears in evolution can, with only moderate semantic license, be regarded as an artistic embodiment of a new conception of living.[11]

Like artistic creativity, evolutionary creativity requires randomness, or indeterminism, to be an objective feature of the natural world. Quantum indeterminism isn't enough. If Mayr's and Dobzhansky's views of evolution are correct, macroscopic indeterminism – the prevalence of chance at macroscopic levels of description – must be a pervasive and objective feature of the macroscopic world, as it seems to be of the world of experience. But it has no place in physicalism's deterministic picture of physical reality, as described by Pierre Simon Laplace in 1814:

> We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.[12]

11 Dobzhansky, Theodosius.. "Chance and Creativity in Evolution" in *Studies in the Philosophy of Biology*, edited by F.J. Ayala and T. Dobzhansky. (Berkeley: University of California Press 1974) p. 329.

12 Laplace, Pierre Simon, *A Philosophical Essay on Probabilities*, translated into English from the original French 6th ed. by Truscott, F.W. and Emory, F.L., (New York, Dover Publications 1951) p.4

In his essay "Chance" (1908) the mathematician, physicist, and philosopher Henri Poincaré reaffirmed this worldview:

> Every phenomenon, however trifling it be, has a cause, and a mind infinitely powerful and informed concerning the laws of nature could have foreseen it from the beginning of the ages. If a being with such a mind existed, we could play no game of chance with him; we should always lose.

Yet in Mayr's and Dobzhansky's views of evolution, all living organisms are winners in games of chance whose outcomes not even "a mind infinitely powerful and informed concerning the laws of nature" could have foreseen." Physicalism equates chance to imperfect human knowledge of laws or initial conditions. By contrast, Mayr's and Dobzhansky's views of evolution require chance events to be unpredictable in principle. Physical laws allow us to predict future states of complex systems if we know enough about their present state. By contrast, evolution gives rise to forms of complex order that aren't prefigured or implicit in earlier states of the universe.

Randomness is synonymous with disorder. Its antithesis, in biological contexts, is organized complexity. Because physicalism's account of physical reality doesn't accommodate macroscopic randomness, neither does it accommodate organized complexity and the processes that give rise to it (creative processes). A scientific worldview that accommodated consciousness and evolution, as conceived by Mayr, Dobzhansky, and other contemporary evolutionary biologists, would also allow us to characterize randomness and organized complexity in an expanded language of physics.

This book describes such a worldview and its underpinnings.

II
Cosmological Assumptions

Natural science's account of the world combines a small set of strongly confirmed physical laws with supplementary conditions that characterize more or less idealized models of physical systems to which the laws apply. Supplementary conditions are of two kinds: *initial conditions* characterize a system at a particular moment; *boundary conditions* describe the system's interaction or absence of interaction with its environment. Each of our strongly confirmed laws applies to many possible systems and processes. For example, astronomers apply Newton's law of gravitation and his laws of motion to idealized models of astronomical systems ranging from the Earth-Moon system to clusters of galaxies. A handful of deep, abstract, and strongly confirmed physical laws govern an unlimited variety of models of system and processes characterized by freely chosen initial and boundary conditions. A theory's testable predictions serve to test both its laws and its initial conditions.

The most fundamental supplementary conditions are cosmological assumptions – assumptions about both the physical universe and the laws. Physicalism assumes that our strongly confirmed physical laws are correct – or at least excellent approximations to still deeper and more highly unified laws. But physicalism isn't a consequence of this assumption alone. It depends also on the assumption that at each moment the physical universe is in a definite physical state:

Physicalism's underlying assumption. A complete description of the physical universe assigns definite values to the physical quantities that characterize macroscopic systems. Thus it fully describes and individuates every macroscopic physical system.

Together with the fact that our strongly confirmed physical laws link a description of any given state of an undisturbed system to which the laws apply to descriptions of the system's earlier and later states, this assumption, implies that the outcomes of nearly all physical processes – indeed all processes other than quantum measurements – are predictable in principle. In other words, the assumption allows us to pass from the (uncontested) proposition that our strongly confirmed physical laws are deterministic – that they link the present state of any undisturbed system to which the laws apply to any of the system's past or future states – to the proposition that these laws plus initial and boundary conditions that conform to physicalism's underlying assumption determine future states of undisturbed systems, including the universe itself. It allows us to pass from the determinism of physical laws to cosmic determinism.

Of course, experiments and observations don't provide the exact values of physical quantities that have continuous ranges of possible values, like position, mass, and temperature. Random measurement errors – a consequence of human ignorance – locate the values of such quantities within small subranges. Improved experiments and observations reduce these subranges. And since measurement errors can't be eliminated entirely, the measured values of physical quantities can't distinguish the Sun from other suns or the Galaxy from other galaxies. Which is why both Laplace's and Poincaré's definitions of determinism, quoted above, refer to an imaginary being that knows the precise values of all relevant physical quantities.

The universe envisaged by Newton, Laplace, and Poincaré is made up of particles of various sizes and shapes, moving and interacting in ways governed by Newton's laws of motion. In such a universe every particle (or its center of mass) has a definite position at each moment, defined by a set of three real numbers – the particle's rectangular coordinates in a fixed but arbitrary coordinate system. But cosmologists now agree that the early universe wasn't a sea of classical particles. It was a sea of particles governed by quantum laws. One strongly confirmed prediction of these laws is that two particles of the same kind, such as a pair of electrons or neutrinos or photons, are indistinguishable in a way that has no classical counterpart. Two classical particles may have identical intrinsic properties, such as mass, electric charge, and spin. Yet they are nevertheless distinguishable: one is here, the other is there. By contrast, the laws of quantum physics don't assign each member of an assembly of elementary particles of the same kind its own position or, more generally, its own single-particle state. The assembly has distinct quantum states, but the individual particles that compose it do not. As a result, states of the early universe *could* be fully described in ways that don't privilege any spatial position or direction. For example, a state of the early universe *could* be a state fully characterized by its temperature and the relative concentrations of elementary particles.

If the initial conditions that characterize the early universe don't privilege any point or direction in space, the same will be true of the conditions that characterize the universe at later times, because the laws that link earlier to later states don't introduce a privileged position or direction. So a complete description of the physical universe can't contain descriptions of this star or that galaxy. Instead it describes what I'll call *cosmological ensembles* – infinite collections of near-replicas uniformly and isotropically distributed in space. Thus geophysics isn't just about Earth; it is about an infinitely dispersed collection of planets whose observable properties are indistinguishable from those of Earth.

I'll refer to the assumption that a complete description of the universe doesn't privilege any spatial position or direction as *the strong cosmological principle.* It is a strengthened version of an assumption that underlies most contemporary theories (and observations) of the astronomical universe:

> *The cosmological principle.* There are coordinate systems relative to which a statistical description of the distribution of matter and motion in the universe at any given moment doesn't privilege any spatial position or direction.

The strong version of this assumption replaces the word *statistical* by the word *complete*:

> *The strong cosmological principle.* There are coordinate systems relative to which a *complete* description of the distribution of matter and motion in the universe at any given moment doesn't privilege any spatial position or direction.

The strong cosmological principle implies that, contrary to the assumption underlying physicalism, a complete description of the physical universe doesn't describe individual physical systems such as the Sun and Earth. Instead it describes infinite collections of replicas – *cosmological ensembles*. The systems that make up an ensemble are uniformly and isotropically distributed in space. The ensemble itself is fully characterized by a set of probabilities, or *probability distribution*, of its members' physical attributes. The probability that an attribute has a given value or a value that lies in a given range of values has an objective interpretation: it is the fraction of replicas in a cosmological ensemble in which the attribute has that value or a value in the given range.

The strong cosmological principle relies on quantum physics. It couldn't hold in a universe made up of particles governed by the laws of classical physics. Consider the simplest model of such a universe: a random, statistically uniform distribution of identical particles. In such a universe the distance between any given particle and (say) its nearest neighbor, measured in a fixed unit of length, is a real number – a number represented by a point on the number line. This number characterizes the particle uniquely, because the probability that a second particle has the same distance from its nearest neighbor is zero. (As the mathematician Georg Cantor showed in 1891, there are infinitely more real numbers in any real-number interval than there are particles in the universe.) The picture of the natural world described in this book rests on a pair of related cosmological assumptions. The first is the strong cosmological principle. The second is the *assumption of primordial randomness*:

> *Primordial randomness.* The probability distributions that characterize the earliest state of the astronomical universe

> to which our present physical laws apply are maximally random.

What does *randomness* mean in this context? As discussed in more detail below, every probability distribution has two complementary attributes: *randomness* and *information*.

Randomness is closely related to – but not identical with – *entropy*. Between 1850 and 1865 Rudolf Clausius discovered a previously unnoticed property of effectively isolated macroscopic systems that have relaxed into a macroscopically uniform, unchanging state called *thermal equilibrium*. He named this property *entropy* and proved that the second law of thermodynamics (the science of heat and its transformations) is equivalent to the statement that the entropy of a closed, or isolated, system never decreases.

Although randomness is a property of any probability distribution, Ludwig Boltzmann introduced it, in 1872, in a particular physical context. He showed that it is the counterpart of Clausius's entropy in the molecular theory of gases – a theory that seeks to found thermodynamics on what was then the speculative assumption that a gas is a collection of particles whose motions and interactions are governed by Newton's laws of motion. Boltzmann's *H* theorem states that the randomness of an isolated sample of an ideal gas never decreases.

In wider contexts randomness is a measure of disorder. Its complementary property, *information*, is the amount by which a probability distribution's information falls short of its largest possible value. It is a measure of order.

The assumption of primordial randomness contradicts a proposition advanced by Rudolf Clausius in 1865: "The entropy of the world tends toward a maximum." This statement has been widely accepted. As astrophysicist Arthur S. Eddington explained in an influential book, *The Nature of the Physical World*,

> The practical measure of the random element which can increase in the universe but never decrease is called entropy. …Entropy continually increases. …The law that entropy always increases – the second law of thermodynamics – holds, I think, the supreme position among the laws of Nature. If someone points out to you that your pet theory of the universe is in disagreement with Maxwell's equations – then so much for Maxwell's equations. If it is found to be contradicted by observation – well, these experimentalists do

> bungle things sometimes. But if your theory is found to be against the second law of thermodynamics I can give you no hope; there is nothing for it but to collapse in the deepest humiliation.[1]

More recently the mathematician Roger Penrose likewise endorsed Clausius' formulation of the law, arguing that the universe began to expand from a highly organized state of exceedingly low entropy and has been becoming progressively more disorganized.[2] But let's take a closer look. The first of thermodynamics' two main laws says that heat and mechanical energy, or work, are interconvertible at a fixed rate of exchange. The second law imposes restrictions on the conversion of heat into work. In one of its forms it states that no cyclic heat engine can convert heat drawn from a heat reservoir at a fixed temperature entirely into work. Clausius defined the entropy of a macroscopic system in terms of its changes when a macroscopic system accepts heat from or delivers heat to an external heat reservoir. He showed that the Second Law is equivalent to the statement that the entropy of a closed (or undisturbed or isolated) system never decreases. The two laws underpin predictions about all macroscopic processes involving heat flow, including the life-sustaining energy transactions between living organisms and their environments.

Both have been strongly confirmed.

At first sight it may seem easy to generalize the definition of entropy and the law of entropy change from macroscopic systems to the universe. Seemingly, we can represent the universe as a collection of weakly interacting macroscopic systems for each of which entropy is a well-defined quantity that never decreases. To calculate the entropy of the universe we simply add up these contributions. The resulting quantity, the "entropy of the universe", never decreases. But, as Enrico Fermi emphasized in his *Lectures on Thermodynamics* and as we'll discuss in more detail below, this recipe works "only if the energy of the system is the sum of the energies of all the parts and if the work performed by the system during a transformation is equal to the sum of the amounts of work performed by all the parts."[3] The systems experimental physical scientists and biologists study usually

1 Eddington, Arthur S. *The Nature of the Physical World*, (Ann Arbor: University of Michigan Press 1958) p. 74

2 Penrose, Roger. *The Emperor's New Mind*. (Oxford: Oxford University Press 1989) Chapters 7, 8

3 Fermi, Enrico. *Thermodynamics*, (New York: Prentice-Hall 1937) p. 53

satisfy these conditions. But astronomical systems don't satisfy the first condition because the gravitational potential energy of a system held together by the mutual gravitational attraction of its parts isn't equal to the sum of the parts' gravitational potential energies. For example, the gravitational potential energy of a uniform gas sphere of given mass density is proportional not to its mass but to the square of its mass.

That Clausius's generalization of his law doesn't apply to self-gravitating systems shows up in the fact that self-gravitating systems have negative heat capacity: the mean temperature of a self-gravitating gas cloud increases when heat, in the form of radiation, *leaves* the cloud. The law of entropy growth implies, however, that any system to which the law applies has positive heat capacity: its mean temperature increases when heat flows into it.

Since the universe is made up of self-gravitating systems, it doesn't satisfy the additivity condition. So we can't extend the law of entropy non-decrease to the universe as a whole.

By contrast, any probability distribution that characterizes the state of the universe has a well-defined randomness. So it does make sense to postulate that the randomness of a statistical description of the universe tends toward a maximum. I think this assumption is implicit in the physicalist picture of the natural world, which contains only the entities mentioned in physical theories, and I think that may be how Eddington and Penrose interpreted Clausius's dictum about the entropy of the universe.

This book argues that a picture of the natural world based (in part) on the alternative assumption that the universe has expanded from a primordial state of zero information or maximum randomness is more inclusive than physicalism's world picture. I'll argue that it accommodates not only consciousness but also creative processes, including free will and biological evolution itself.

The case for the two cosmological assumptions rests in part on the claim that they characterize the physical universe as simply as possible. I argue below that these assumptions throw light on three longstanding problems in physics: the problem of time's arrow – the apparent conflict between the irreversibility of macroscopic processes involving heat flow and the fact that the underlying microscopic laws do not distinguish between the direction of the past and the direction of the future; the measurement problem of quantum mechanics – the apparent conflict between the indeterminacy of quantum measurement outcomes and the determinacy of the physical laws that govern the physical processes involved in measurements; and the interpretation of probabilities in the statistical theories that link macrophysics and microphysics.

III

The Universe's Large-Scale Structure

Isaac Newton envisioned the universe as an infinitely extended assembly of particles, each of which is in a definite place at each moment. Replying to a question from the theologian Nicholas Bentley, who was preparing a set of lectures intended to defend religion from atheism, Newton, a devout theist, wrote, in 1692:

> As to your first query, it seems to me that if the matter of our sun and planets and all the matter in the universe were evenly scattered throughout all the heavens, and every particle had an innate gravity toward all the rest, and the whole space throughout which this matter was scattered was but finite, the matter on the outside of the space would, by its gravity, tend toward all the matter on the inside, and by consequence, fall down into the middle of the whole space and there compose one great spherical mass. But if the matter was evenly disposed throughout an infinite space, it could never convene into one mass; but some of it would convene into one mass and some into another, so as to make an infinite number of great masses, scattered at great distances from one to another throughout all that infinite space. And thus might the sun and fixed stars be formed, supposing the matter were of a lucid nature. But how the matter should divide itself into two sorts, and that part of it which is fit to compose a shining body should fall down into one mass and make a sun and the rest which is fit to compose an opaque body should coalesce, not into one great body, like the shining matter, but into many little ones; or if the sun at first were an opaque body like the planets, or the planets lucid bodies like the sun, how he alone would be changed into a shining body whilst all they continue opaque, or all they be changed into opaque ones whilst he remains unchanged, I do not think explicable by mere natural causes, but am forced to ascribe it to the counsel and contrivance of a voluntary Agent.

Newton doesn't mention this cosmogonic hypothesis in his masterwork, T*he Mathematical Principles of Natural Philosophy* (the *Principia*), which had appeared five years earlier. In *Relativity, the Special and the General Theory,* a popular exposition of relativity, Einstein points out that Newton's model of the astronomical universe, in which matter is distributed more or less uniformly throughout infinite Euclidean space, is inconsistent with his (Newton's) law of gravitation:

> [Newton's theory] requires that the universe should have a kind of center in which the density of stars is a maximum, and that as we proceed outwards from this center the group density of the stars should diminish, until finally, at great distances, it is succeeded by an infinite region of emptiness. The stellar universe ought to be a finite island in the infinite ocean of space.[1]

Einstein appends a proof of this claim:

The proof rests on the fact that Newton's law of gravitation can be expressed as a picture of "lines of gravitational force." Every particle is the terminus of such lines. Their number is a fixed multiple of the particle's mass. If mass were uniformly distributed in infinite Euclidean space, the number of lines of force passing through unit area of any given sphere – and hence the magnitude of the gravitational force at any point on the sphere – would be proportional to the sphere's radius and would therefore increase without limit as the radius increased, "which [Einstein writes] is impossible."

Einstein hoped that his own theory of spacetime structure and gravitation *would* apply to Newton's simplified model of the physical universe – an infinitely extended, uniform distribution of mass/energy. In 1917, two years after his account of the general theory of relativity appeared in print, he summarized his efforts to construct a description of the universe based on that model. He concluded that the equations that link the structure of spacetime to its contents – the theory's "field equations" – do *not* have a solution that describes a motionless, infinitely extended, uniform distribution of mass/energy.

Einstein searched for a way to reconcile general relativity with his preferred model of the astronomical universe. General relativity

1 Einstein, A., Relativity, the Special and the General Theory, fifteenth edition, (Crown, New York, 1952) p. 106

incorporates Bernhard Riemann's mathematical theory of curved n-dimensional continuums. That theory allows a three-dimensional space to be curved both globally and locally, like a model of Earth's surface. If physical space had constant positive curvature it would be the three-dimensional "surface" of a four-dimensional ball. Like a sphere (the two-dimensional surface of a three-dimensional ball), such a space is unbounded – it has no edge – yet finite. Disappointingly, the field equations seemed to have no solution that describes a finite quantity of matter spread evenly throughout a finite space of constant positive curvature. Convinced that a plausible model of the universe must be uniform and isotropic apart from local irregularities, Einstein now took a step he later regretted. He modified the field equations of his 1915 theory in a way that would allow them to have such solutions. The 1915 theory links the structure of space-time to the distribution of mass/energy in a way that involves only a single physical constant – the constant that appears in Newton's law of gravitation. The field equations of the modified theory contain an additional term, whose coefficient is a second constant, the *cosmological constant*.

Five years later, in 1922, the mathematician Alexander Friedmann wrote a paper claiming that Einstein's original field equations do in fact have solutions that describe uniform, unbounded distributions of mass – solutions Einstein had overlooked. Einstein's mistake, Friedmann claimed, had been to assume that such a distribution must be *static*. True, there was as yet no observational evidence that the distribution of matter was not static on large scales, though there soon would be. But Friedmann was a mathematician, not a physicist or astronomer. So he asked himself: Do Einstein's original field equations have solutions that describe a uniform, unbounded, but *non-static* distribution of mass? He found that they do. The mass distribution may be infinitely extended or, if space has positive curvature, it may occupy a finite volume. If the medium is infinitely extended, a solution of the field equations describes a universe that expands forever from a state of nominally infinite density in the finite past. If the distribution of mass/energy and the space it occupies are finite, a solution (of the same field equations) cycles endlessly through alternate periods of expansion and contraction. These solutions of Einstein's 1915 field equations describe models of the physical universe that, so far as I know, no astronomer, philosopher, or writer of science fiction had previously described. At first Einstein rejected Friedmann's reasoning, but on reflection he admitted he had been mistaken. He wholeheartedly embraced Friedmann's solution of the cosmological problem and urged the editors of the journal to which Friedmann had submitted his paper to accept it (which, of course, they did). Einstein also decided that it had been a mistake to add an extra term to his original field equations. Ironically, most contemporary cosmologists believe the extra term is needed to account for astronomical evidence.

"The Realm of the Nebulae"

Using a telescope he himself had designed and built, Galileo discovered in 1610 that the Milky Way, a diffuse band of light that encloses a great circle on the dome of the sky, is made up of myriad "suns" too faint for the naked eye to resolve.

In 1750 Thomas Wright, an instrument maker, suggested that just as the Earth's orbit lies near the central plane of a flattened system of planets circling the Sun, the stars that make up the Milky Way belong to a flattened system of Sun-like objects whose members, including the Sun, circle a distant center.

Immanuel Kant embraced Wright's hypothesis and, in his 1755 book *Universal Natural History and Theory of the Heaven*, greatly extended it. Kant suggested that the stars are suns, many of them surrounded by flattened systems of planets, which, like Jupiter and Saturn, are themselves surrounded by flattened systems of satellites; that not only the Milky Way but also the "nebulae" – elliptical cloud-like objects scattered among the stars – are very distant stellar systems with the same structure as planetary systems – a massive central object surrounded by a disc-like family of satellites, all revolving about the central object in the same sense; and that these stellar systems themselves belong to still larger unbounded system with the same structure.

Kant argued that all these systems could have formed by a single process: a cloud of dispersed matter collapses under the combined action of Newton's gravitational attraction and a hypothetical repulsive force. (Curiously, because he was an admirer and expounder of Newtonian physics, Kant didn't recognize that Newton's theory makes a repulsive force superfluous. As the mathematician Pierre

Simon Laplace (1749 – 1827) later showed, unless a cloud has zero angular momentum, or spin, it collapses into a flattened system in which the centrifugal, or radially outward, acceleration of a particle circling the central mass balances the radially inward gravitational acceleration of the particle produced by matter inside the particle's orbit.) Astronomers disagreed about whether the "nebulae" are relatively nearby gas clouds or distant stellar systems. In 1920 astronomers Harlow Shapley and Heber Curtis held a famous debate on this question. Curtis defended Kant's view that the nebulae are stellar systems comparable in all respects to our own system, the Milky Way. Shapley argued that they are gas clouds in the Milky Way.

In the 1920s the astronomer Edwin Hubble and his colleagues, using what was then the world's most powerful telescope, the recently completed 100-inch reflecting telescope at the Mount Wilson Observatory in southern California, made systematic observations of the nebulae that resolved the debate. Longexposure photographs resolved bright stars in the largest and brightest of the nebulae, the Andromeda Nebula, and analyses of photographic and spectroscopic data confirmed that it was a stellar system comparable in size and stellar composition with our own stellar system, thus confirming Kant's hypothesis. The nebulae were indeed galaxies (a term suggested by Shapley). In *The Realm of the Nebulae* Hubble wrote:

> Investigations of the observable region as a whole have led to two results of major importance. One is the homogeneity of the region – the uniformity of the large-scale distribution of the nebulae. The other is the velocity-distance relation. [2]

The first result depended on a discovery by the astronomer Henrietta Swan Leavitt (1868-1921). Some stars are close enough to have measurably different directions when the Earth is on opposite sides of the Sun. From measurements of this difference, the star's parallax, one can deduce its distance, much as Thales deduced the distance of a ship at sea from measurements of its parallax and the measured distance between the two points of observation. Astronomers then deduce the star's true brightness, or luminosity, from its apparent brightness and the fact that an object's apparent brightness diminishes like the reciprocal of the square of its distance. Conversely, if an astronomer knew a celestial object's luminosity she could infer

2 Hubble, Edwin, *The Realm of the Nebulae*, (Yale University Press, 1936)

its distance from a measurement of its (apparent) brightness. Leavitt studied 1777 variable stars in the Magellanic Clouds, a pair of dwarf galaxies that orbit our own galaxy. Among these variable stars were Cepheids, pulsating stars whose brightness varies periodically. She found that the mean brightnesses of the Cepheids she studied were tightly correlated with their periods. Since stars in the Magellanic Clouds are all at nearly the same distance from Earth, Leavitt's discovery meant that the luminosities of the Cepheids she studied were tightly correlated with their periods.

To calibrate Leavitt's period-luminosity relation astronomers use Cepheids that are close enough to have measurable parallaxes. They then use the calibrated relation between distance and luminosity to infer the distances of more distant Cepheids from their (apparent) brightnesses. In 1923, two years after Leavitt's death, Hubble discovered the first of several Cepheids in the Andromeda Nebula, the brightest of the external galaxies. Its measured brightness and Leavitt's period-luminosity relation played the decisive role in establishing that the Andromeda Nebula is a stellar system comparable in size and stellar composition with our own galaxy.

Hubble's assumption that Cepheid-like variables in external galaxies satisfy the same relation between true brightnesses and period of variation as Cepheids in our own galaxy exemplifies an assumption he called "the principle of uniformity." Using that assumption to interpret measurements of the brightnesses of galaxy images on photographs made with the 100-inch reflector, Hubble concluded:

> There is no evidence of a thinning-out, no trace of a physical boundary. There is not the slightest suggestion of a supersystem of nebulae isolated in a larger world. Thus, for the purposes of speculation, we may apply the principle of uniformity, and suppose that any other equal portion of the universe, selected at random, is much the same as the observable region. We may assume that the realm of the nebulae is the universe and that **the observable universe is a fair sample.** (emphasis added)

In Hubble's hands the assumption that "any other equal portion of the universe, selected at random, is much the same as the observable region" proved to be an invaluable research tool. As he understood, "much the same" doesn't mean "exactly the same." Hubble's successors have devoted much time, effort, and ingenuity to refining and extending the principle of uniformity. For example, they have discovered that there are different kinds of Cepheids, with slightly but significantly different period-luminosity relations, and that some of the differences depend on systematic differences in chemical composition between different stellar populations. Such differences, however, are consistent with a more general cosmological assumption, mentioned earlier: the *cosmological principle*:

> Relative to a suitably defined system of spacetime coordinates, the average properties of galaxies and their distribution in space at any given moment are the same everywhere and in all directions.

Or more briefly:

> *The astronomical universe is statistically homogeneous and isotropic.*

The second of Hubble's "two results of major importance" was the velocity-distance relation. Whereas the first result (that the "external nebulae", or galaxies, are stellar systems comparable to our own) depended on brightness measurements, the second depended on measurements of the displacements of absorption lines in the spectra of distant galaxies.

Astronomical Spectroscopy

By passing a beam of sunlight through a glass prism, Newton separated it into what he called "colored rays." He found that differently colored rays were deflected through different angles when they passed through the prism. Newton's successors established that the colored rays are electromagnetic waves with definite frequencies, ranging from 430×10^{12} cycles per second at the red end of the spectrum to 770×10^{12} cycles per second at the violet end.

The spectrum of sunlight is crossed by thousands of dark lines, each of which results from the absorption and subsequent isotropic reemission of light by a specific kind of atom or molecule. The lines produced by a single atom or molecule are very narrow, but those produced by a macroscopic gas sample – a portion of the solar atmosphere, for example – are broadened in wavelength by the Doppler effect associated with the relative line-of-sight motions of the atoms or molecules in the sample. (The Doppler effect increases the frequency of monochromatic light emitted by an approaching source and decreases the frequency of light emitted by a receding source. The fractional change in frequency is proportional to the velocity of approach or recession.)

The spectrum of a distant galaxy superimposes contributions from the galaxy's stars (and other emitters of light). The breadth of an absorption line measures the dispersion of the relative line-of-sight velocities of the stars. The displacement of the line's center from its rest-value measures the line-of-sight velocity of the galaxy's center of mass (or, more accurately, "center of light"), relative to the telescope.

Extending pioneering measurements of galaxy redshifts by Vesto Slipher, Hubble and his colleague Milton Humason found that the absorption lines in the spectra of very distant galaxies were always displaced toward the red end of the spectrum. Moreover, this displacement, or red shift, had a systematic component, proportional to the galaxy's distance: the points on a graph of measured redshift against estimated distance hugged a straight line. And because the amount of scatter didn't increase systematically with estimated distance, Hubble could attribute deviations from the straight-line relation to gravitational accelerations arising from local non-uniformities in the spatial distribution of galaxies – nonuniformities whose average properties didn't depend on distance from the observer.

The cosmological principle implies that there's nothing special about our galaxy's place in the universe of galaxies. It also implies that the population of galaxies extends indefinitely in all direction. The proportionality between the average line-of-sight velocity of a galaxy and its distance from the observer then has a straightforward interpretation:

> The unbounded distribution of galaxies is undergoing a uniform expansion – an expansion that looks the same from every vantage point.

Hubble announced the velocity-distance relation in 1929. At the time, he didn't know that his discovery confirmed Alexander Friedmann's seven-year-old prediction, based on the general theory of relativity and the assumption that matter is more or less evenly distributed in an unbounded universe.

The Early Universe

In an expanding universe the mass density, averaged over scales greater than the largest scales on which the mass distribution is non-uniform, decreases steadily with time. So as we look back in time, we see the average mass density steadily increasing. Eventually it becomes equal to, and then surpasses, the average mass density of galaxy clusters. At these and earlier times galaxy clusters couldn't yet have existed. As we look back still further in time, we reach epochs in which galaxies themselves couldn't have existed. Still earlier, the average cosmic mass density exceeded stellar densities. Even earlier, the average mass density would have been so high that atoms and molecules couldn't have existed. Finally, at the earliest times at which our current physical theories apply, the universe would have been a more or less uniform distribution of elementary particles and photons. This conclusion seems inescapable if we assume that the average properties of the astronomical universe are the same everywhere and that quantum theory and Einstein's theory of gravitation apply in their respective domains.

In 1964 physicists Arno Penzias and Robert Wilson at the Crawford Hill location of Bell Telephone Laboratories discovered what came to be called the cosmic microwave background. With a radio telescope they had built to observe electromagnetic radiation emitted by astronomical objects with wavelengths ranging from tenths of a centimeter to a few centimeters, they detected a diffuse radiation field whose intensity never varied and was the same in all directions. The spectrum of the radiation was indistinguishable from that of radiation in a cavity whose walls are maintained at a temperature around three degrees above absolute zero (on the Kelvin scale). Later more accurate measurements confirmed this conclusion and placed the temperature at 2.725 K.

In the most widely accepted model for the origin of the cosmic microwave background, the temperature of the background radiation has steadily decreased from very high values near the beginning of the cosmic expansion:

> At the time when [the temperature] $T \approx 10^{12}$, the universe contained photons, muons, antimuons, electrons, positrons, neutrinos and antineutrinos. In addition, there was a very small nucleonic

> contamination, with neutrons and protons in equal numbers. All of these particles were in thermal equilibrium.[3]

In other models the primordial universe was cold, and the cosmic microwave background came into being later.[4]

Quantum Indistinguishability

Elementary particles and photons come under the jurisdiction of quantum mechanics. A statistically uniform distribution of elementary particles and photons in thermal equilibrium differs profoundly from a statistically uniform and random distribution of particles governed by the laws of classical, or non-quantum, physics. A statistically uniform and random distribution of classical particles with given mass density and temperature has infinitely many microscopically distinct realizations, or microstates. But a statistically uniform distribution of identical particles governed by quantum mechanics is fully characterized by the distribution's mass density and its temperature. It has a single microstate. This conclusion follows from a deep, consequential, and non-intuitive feature of the quantum world: *quantum indistinguishability*. I'll argue shortly that quantum indistinguishability underpins this essay's main argument – *that randomness is an objective and pervasive feature of the physical universe*.

What is quantum indistinguishability? Two classical particles may have identical properties, such as mass, electric charge, and spin. Yet they are nevertheless distinguishable: one is here, the other is there. Now consider a statistically uniform assembly of identical classical particles. At any given moment the distance between a given particle (or its center of mass) and the particle's nearest neighbor, expressed as a multiple of a fixed unit of distance, is a real number – a number represented by a point on the number line or by a non-terminating decimal. The probability that any other particle in the assembly has the same distance from its nearest neighbor is zero. For the probability that the first n digits in the decimal expansions of two randomly selected real numbers coincide is $1/10^n$, which approaches zero as n increases without limit. So each member of an infinitely extended, statistically uniform and random distribution of identical particles is distinguishable in principle from all other members by the relative positions of its neighboring particles (or even of one of

3 Weinberg, S. Gravitation and Cosmology. (New York: Wiley 1972)p. 528.

4 Aguirre, A. 1999. Astrophysical Journal, 521, 17-29.

them). Such an assembly is in a definite microstate at each moment; and the number of possible microstates is infinite in the same way that the number of real numbers in an interval of the number line is infinite.

Like classical particles of the same kind, quantum particles of the same kind have identical intrinsic properties – mass, electric charge, spin, and magnetic moment. But they are also indistinguishable in a more radical way that depends on a feature of the mathematical description of quantum states that has no counterpart in experience or in pre-quantum physics. This new kind of indistinguishability – quantum indistinguishability – strongly influences observable, macroscopic properties of matter and light. For example, it is behind the fact that atoms in the same column of the periodic table of chemical elements (such as hydrogen, lithium, sodium, and potassium) have similar chemical properties (for example, the named elements are all monovalent). It is also behind the fact that the distribution of photon energies in a box whose walls are maintained at a fixed temperature differs from the distribution of the kinetic energies of gas particles in the same box.

To understand quantum indistinguishability you need to know two things about the grammar of state vectors, the mathematical objects that represent quantum states. (1) Two state vectors that differ only in sign represent the same quantum state. (2) Exchanging the labels of two particles in a state vector that represents the state of an assembly of identical particles must either leave the state vector unaltered or change its sign (that is, multiply it by −1). These are the only possibilities the rules of quantum mechanics allow. They define two classes of elementary particles, called *fermions* (after Enrico Fermi) and *bosons* (after Satyendra Nath Bose). Electrons, protons, and neutrinos are fermions – the joint state vector of an assembly of fermions changes sign when you exchange the labels of two particles in the assembly. Helium atoms in their states of lowest energy and photons are bosons. The joint state vector of an assembly of bosons doesn't change when you exchange the labels of two particles in the assembly. Which of the two classes an elementary particle belongs to depends on its spin. Particles of spin 1/2, 3/2, ... in units of $h/2\pi$ are fermions, particles of spin 0, 1, 2, ... in the same unit are bosons.

An assembly of fermions has different observable properties from an assembly of bosons, and both collections have different observable properties from an assembly of identical classical particles. Some examples:

— From plausible but then-controversial statistical assumptions James Clerk Maxwell, in 1860, deduced the distribution of molecular kinetic energies in a gas sample in thermal equilibrium. He found that each molecular-velocity component has a "normal distribution," represented by the bell-shaped curve that represents the distribution of random measurement errors, and that the width of the distribution is proportional to the temperature of the gas sample. Much later, experiments confirmed his prediction.

— In 1900 Max Planck devised a formula that closely fitted recently improved measurements of the spectrum (frequency distribution) of thermal radiation (light in a box whose walls are maintained at a fixed temperature). In view of Einstein's photon hypothesis (now a firmly entrenched feature of quantum physics) we can think of Planck's law as describing the equilibrium distribution of photon energies. It differs markedly from Maxwell's law for the distribution of the energies of classical particles in thermal equilibrium.

— Two fermions can't occupy the same single-particle state. Wolfgang Pauli discovered an instance of this rule in 1925. To explain the periodic structure of the periodic table of chemical elements he made two proposals. He suggested that four, rather than three, quantum numbers characterize the state of an electron in an atom. (The first three quantum numbers correspond to the fact that we need three numbers to specify a classical particle's orbit. The fourth quantum number, which had two possible values, turned out to characterize the component of the electron's spin along a fixed direction.) Pauli's second proposal was the rule, known as Pauli's exclusion principle: two electrons can't occupy the same quantum state.

— Because several bosons can occupy the same single-particle states, the members of a collection of bosons in thermal equilibrium (a dilute gas of helium atoms, for example) crowd into the single-particle state of lowest energy at temperatures close to absolute zero.

The Strong Cosmological Principle

The cosmological principle is an assumption about the large-scale structure of the physical universe. It says that there are spacetime coordinate systems in which a description of the universe phrased entirely in terms of probability distributions and average values doesn't discriminate between positions in space or between directions in space. For example, at each moment the mass density, averaged over a sufficiently large region, has the same value everywhere. So too does the average value of the squared difference between the mass density at a point and its (position-independent) average value.

All coordinate systems in which a statistical description of the universe doesn't discriminate between spatial positions or directions have a common time coordinate. In this respect the cosmological principle may seem to reinstate Newton's universal time. In Newton's physics time, and with it the notion of rest, are "absolute" because his laws of motion hold in, and only in, particular spacetime coordinate systems all of which are at rest relative to one another. Einstein's special theory of relativity (1905) abolished absolute time and absolute rest. It demanded that physical laws in no way discriminate between a spacetime coordinate system in which Newton's laws hold and any other coordinate system whose motion is unvarying in speed and direction relative to the first system. Coordinate systems in which Newton's laws hold are called inertial systems. The principle of special relativity requires physical laws to take the same form in all inertial coordinate systems if they are written in a mathematical language devised by Einstein's former teacher, Hermann Minkowski. Minkowski spacetime replaces the Euclidean space plus absolute time of Newton's theory. Newtonian physics becomes a limiting case of special-relativity physics, approximately valid for particle speeds much smaller than the speed of light in empty space. Beginning in 1928 with P.A.M. Dirac's relativistic generalization of Schrödinger's equation, discussed in more detail below, special relativity became a pillar of quantum theory.

The general principle of relativity revokes the privileged status of inertial spacetime coordinate systems. It requires physical laws to take the same form in all coordinate systems that assign the same squared spacetime interval (the squared time interval minus the squared distance interval) to every pair of neighboring point events. As Einstein explains in his popular exposition Relativity, the Special and the General Theory,[5]in his more technical lectures *The Meaning of Relativity*,[6] and in his comprehensive journal article *The Foundations of the General Theory of Relativity*,[7]a consistent working-out of the general principle of relativity leads to a unique theory of gravitation and spacetime structure. (The path wasn't easy, though, even for Einstein. One wrong turn, known as the "hole problem" put him off course for two years.[8])

The existence of a preferred family of spacetime coordinate systems for the universe as a whole doesn't clash with the general principle of relativity. The latter constrains the *laws* governing spacetime structure and gravitation; the cosmological principle characterizes a particular system to which the laws apply: the astronomical universe.

The assumption that there exists a system of spacetime coordinates relative to which no statistical property of the physical universe defines a preferred position or direction in space is often viewed as defining an idealized model of the universe, like the assumptions that define idealized models of stars and galaxies. Unlike those assumptions, however, it *could* hold exactly. If the early universe is statistically uniform and isotropic, quantum indistinguishability implies that its *complete* description – and hence a complete description of all subsequent states – doesn't privilege any spatial position or direction. I will refer to the hypothesis that it does hold exactly as the *strong cosmological principle*. This book argues that it is the missing element in a picture of the natural world that incorporates our

5 Einstein, Albert, *Relativity, the Special and the General Theory*, 15th edition, , (Crown Publishers, New York, 1952)

6 Einstein, Albert, *The Meaning of Relativity*, 5th edition, (Princeton, NJ, Princeton University Press, 1953)

7 Einstein, Albert, *The Foundations of the General Theory of Relativity*, in *The Principle of Relativity*, (Methuen and Company 1923, reprinted by Dover Publications)

8 Norton, John D., "The Hole Argument", *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2015/entries/spacetime-holearg/>

strongly confirmed physical laws. It draws support from (at least) three arguments.

First, precise and extensive observations of the cosmic microwave background and of the spatial distribution and line-of-sight velocities of galaxies have so far produced no compelling evidence of deviations from statistical homogeneity and isotropy, now or in the past. Astronomical observations provide little support if any for the view that the cosmological principle is merely an approximation or an idealization, like the initial conditions that define models of stars and galaxies.

Second, the initial conditions that define the universe don't have the same function as those that define models of astronomical systems. A theory of stellar structure must apply to a range of stellar models because stars have a wide range of masses, chemical compositions, spins, and ages. But there's only one universe. The strong cosmological principle defines its simplest models. In that respect it is more like a law than an initial condition.

Finally, the strong cosmological principle accounts for what is sometimes called "Mach's principle": local inertial reference frames – frames in which Newton's and Maxwell's laws as well as their special-relativistic generalizations hold – are unaccelerated relative to a frame defined by the cosmic microwave background and the distribution and motions of distant galaxies. Einstein's theory of gravitation predicts this coincidence *provided the cosmological principle holds exactly*. Astronomical evidence supports this prediction. It indicates that local inertial reference frames are indeed unaccelerated relative to a coordinate system in which the cosmic microwave background is equally bright, on average, in all directions and the spatial distribution of galaxies is statistically homogeneous and isotropic. If the distribution of energy and momentum on cosmological scales were not statistically homogeneous and isotropic, there would be no preferred cosmological frame and hence no obvious explanation for the observed relation between local inertial frames and the frame defined by the cosmic microwave background and the spatial distribution and line-of-sight velocities of galaxies. If the cosmic medium was in thermal equilibrium at very early times, it lacked structure on all macroscopic scales. Its state was maximally random. (As discussed below, thermal equilibrium is a state of maximum randomness.) But as the universe expanded, structure emerged. I have discussed one possible scenario for this process elsewhere.[9] The cosmic distribution of mass/energy became progressively less random, more orderly.

9 Layzer, David, *Cosmogenesis: The Growth of Order in the Universe*, (Oxford *University* Press, 1990)

IV
Entropy and Its Law

Heat as a Form of Energy

From the middle of the eighteenth century to the middle of the nineteenth century scientists who invented and used devices for measuring the heat released or absorbed in chemical reactions and in the compression and expansion of gas samples disagreed about the nature of what they were measuring. Some held that heat is a conserved substance, "caloric," analogous to mass. When a body gains caloric its temperature rises; when it gives up caloric its temperature falls. And in a heat engine, such as James Watt's steam engine, caloric does work on a movable piston – for example, by raising a weight – when the engine's working substance expands and heat "falls" from a higher to a lower temperature. Newton, late in the seventeenth century, had championed a different view. It rested on what many scientists, then and until the opening years of the twentieth century, regarded as a speculative hypothesis: that the universe is made up of invisible, indivisible particles moving and interacting in otherwise empty space. Put forward by Leucippus in the fifth century BCE and elaborated by his pupil Democritus, the atomic hypothesis was made the basis of a comprehensive naturalistic philosophy by Epicurus (341 – 270 BCE). In the first century BCE it became the subject of a long poem, *On the Nature of Things*, by the Roman poet and philosopher Lucretius. Stephen Greenblatt, in *The Swerve*,[1] has argued that this widely read poem played an important role in the emergence of the modern world.

The atomic hypothesis suggested that a flow of heat into an otherwise undisturbed system, measured in an appropriate unit, produces an equal change in the sum of the kinetic energy associated with the motions of the system's hypothetical atoms and the potential energy associated with the forces the atoms exert on one another. Newton's laws of motion imply that the sum of these two kinds of energy is constant in time; an increase of one of them is accompanied by an equal decrease in the other. Inflows of heat increase an otherwise undisturbed system's internal energy; outflows diminish it. In the early 1840s Robert von Mayer (in Germany), James Joule (in England), and Ludwig Colding (in Denmark) independently advocated this hypothesis. In 1843 James Joule described experiments that lent it strong support. Using a calorimeter, he measured the quantity of heat generated by the viscous (internal frictional) dissipation of internal motions created by a paddle wheel immersed in water and driven by a descending weight. He found that the heat generated in this process, measured by a rise in temperature of the water, was consistently and accurately proportional to the quantity of mechanical energy that disappeared, measured by the decrease in the height of a weight whose descent drove the paddle wheel. And the constant of proportionality between the heat generated and the mechanical energy (in this case, gravitational potential energy) that disappeared always had the same value up to experimental error. In short, Joule concluded, heat has a fixed "mechanical equivalent." It is a kind of energy.

This proposition became the first of the two most basic laws of the new science of thermodynamics. It allowed scientists to extend the definition of energy from its original domain, mechanics, to include the domain of thermodynamics. They could now attribute a new property – internal energy –to an undisturbed macroscopic system that has relaxed into a macroscopically unchanging equilibrium state, and extend the scope of Newtonian mechanics' principle of conservation of energy, which up until then had applied only to systems whose internal motions don't generate heat, to any undisturbed system.

1 Greenblatt, Stephen , *The Swerve*, (New York: Norton and Company 2009)

The Second Law

The second of thermodynamics' two mains laws imposes a limitation on devices that, like the steam engine and the internal combustion engine, convert heat into mechanical energy. Its first version was a theorem stated and proved by the engineer and physicist Sadi Carnot (1796 - 1832) in an essay entitled "Reflections on the Motive Power of Heat" published in 1824. The essay opens:

> Everyone knows that heat can produce motion. That it possesses vast motive-power no one can doubt, in these days when the steam engine is everywhere so well known.

But despite its ubiquity in nature and its importance in industry,

> [t]he phenomenon of the production of motion by heat has not been considered from a sufficiently general point of view. We have considered it only in machines the nature and mode of action of which have not allowed us to take in the whole extent of application of which it is susceptible. In such machines the phenomenon is, in a way, incomplete. It becomes difficult to recognize its principles and study its laws. In order to consider in the most general way the principle of the production of motion by heat, it must be considered independently of any mechanism or any particular agent. It is necessary to establish principles applicable not only to steam engines but to all imaginable heat engines, whatever the working substance and whatever the method by which it is operated.

To understand "the production of motion by heat," Carnot argues, it isn't enough to study heat engines experimentally. The principle underlying heat engines isn't an empirical generalization; it's an exact mathematical law. Carnot's essay lays bare this principle. It is known as Carnot's theorem.

A quarter of a century later, William Thomson (who became Lord Kelvin) and Rudolf Clausius independently deduced the second of thermodynamics' two main laws from a slightly but significantly emended version of Carnot's theorem. Thomson based his thermometer-independent definition of temperature (*thermodynamic* *temperature*) on Carnot's theorem, and Clausius based his definition of entropy and his derivation of the law of entropy change on Carnot's theorem and Thomson's definition.

In the following paragraphs I argue that the argument that led Clausius from Carnot's theorem to the conclusion that the entropy of the universe tends toward a maximum over-generalizes and over-extends a series of correct inferences. I conclude that his law of entropy change applies in an important but limited domain but doesn't apply to systems held together by the mutual gravitational attraction of their particles (self-gravitating systems). In particular, it doesn't apply to stars, galaxies, and the physical universe.

Neither Clausius's deduction of his law nor the argument that it doesn't apply to the physical universe involves mathematics beyond elementary algebra. Since that argument plays a pivotal role in the following discussion, I'll try to describe both Clausius's route to entropy and its law and my critique of the law's most general form in enough detail to allow readers with a modest mathematical background to make an informed judgment about the critique's validity.

Carnot's Theorem

Carnot's derivation of the theorem from which William Thomson and Rudolf Clausius deduced the second law of thermodynamics rests on two solid empirical notions – temperature and heat-gain (or heat-loss), both of which are unambiguously measurable (though in arbitrary units).[2] It also rests on two mistaken theoretical assumptions – that heat is a conserved substance and that it does work by "falling" from a higher to a lower temperature, just as water in a watermill does work when it falls from a higher to a lower elevation.

Carnot imagined an ideal cyclic heat engine, an engine that converts heat into work and whose working substance – the analogue of the steam in a steam engine – returns to its initial state at the end of each cycle. The engine's working substance is a gas enclosed in

2 Carnot's contemporaries measured a body's *temperature* by the volume of a sample of air or mercury in thermal contact with the body at a given pressure. Experimenters discovered that different "thermometric substances" yielded different but *interconvertible* temperature measures. *Heat gain*. When heat flows into (or out of) a body, its temperature rises (or falls) – or else it undergoes a change of state, as when ice melts or water turns into steam. Experimenters discovered that different "calorimetric substances" yielded different but *interconvertible* measures of heat gain.

a cylindrical cavity bounded at one end by a movable piston. The working substance is intermittently in contact with two heat reservoirs: a source of heat, analogous to the boiler in a steam engine, and a sink of heat, analogous to the condenser in a heat engine. As the piston moves back and forth, the gas in the cylinder alternately expands and contracts. When it expands it does work on the outside world – for example, by raising a weight. When it contracts, the descending weight does work on the gas.

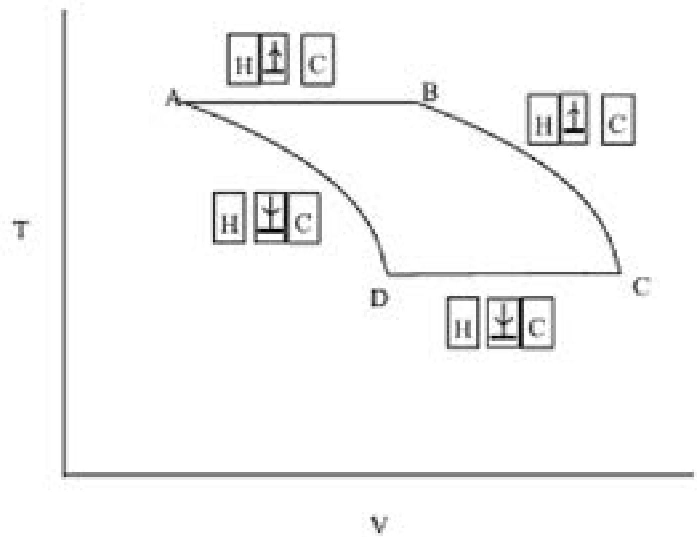


Figure 1 The Carnot Cycle

A point whose vertical coordinate is the working substance's temperature and whose horizontal coordinate is its volume represents the state of the working substance. At the beginning of a cycle the working substance is in the state represented by point A. Its temperature is equal to that of the hot reservoir (H), with which it remains in contact as its volume increases to the value represented by the horizontal coordinate of point B. As the working substance expands from state A to state B it does work on the piston and draws heat from the hot reservoir. Between states B and C the working substance neither gains nor loses heat but continues to expand and do work on the piston. Between states C and D it is in contact with and delivers heat to the cold reservoir (C) while the

> piston does work on it. In the final phase of the cycle, DA, the working substance neither gains nor loses heat as the piston does work on it. State D is chosen so that the fourth phase returns the working substance to its initial state. In the reverse cycle, ADCBA, the working substance expands, doing work on the piston while thermally isolated (AD), continues to expand while accepting heat from the cold reservoir (DC), contracts, having work done on it by the piston while thermally isolated (CB), and contracts further to its initial state while delivering heat to the hot reservoir (BA).

Carnot made his imaginary engine as efficient as possible by postulating not only that no mechanical energy is wasted by friction between the engine's moving parts but also that when heat flows between a heat reservoir and the working substance the two are at exactly the same temperature, thus ensuring that all of the heat withdrawn from the hot reservoir is transferred to the cold reservoir. He also assumed that when the working substance is thermally isolated it expands so slowly that at each moment it is in an equilibrium state characterized by definite values of its temperature and volume.

These idealizations not only make Carnot's ideal engine as efficient as possible. Crucially, they also make it reversible. Operating in its reverse mode, a Carnot engine acts as a refrigerator: it absorbs mechanical energy from the alternately falling and rising weight, extracts heat from the cold reservoir, and delivers heat to the hot reservoir.

Because he accepted the caloric hypothesis Carnot assumed that in the course of a cycle his ideal engine transfers all the heat it withdraws from the hot reservoir to the cold reservoir. Operating in its reverse mode, it transfers all the heat it withdraws from the cold reservoir to the hot reservoir.

Carnot defined the engine's efficiency as its mechanical-energy output during a cycle divided by the quantity of caloric transferred from the hot reservoir to the working substance (and eventually to the cold reservoir). He then asked and answered the question on which his fame rests: *Can two ideal heat engines operating between the same pair of heat reservoirs have different values of this ratio?*

Suppose this were possible. Let the more efficient engine transfer a quantity Q of caloric from the hot to the cold reservoir while doing a quantity of work W on its surroundings. We can then use the less efficient engine, operating in its reverse, or refrigerator, mode, to transfer the same amount of caloric Q from the cold reservoir back to the hot reservoir while doing a smaller quantity of work, W'. At the end of a cycle the composite engine is back in its initial state. Neither reservoir has gained or lost heat. But a quantity $W - W'$ of mechanical energy has appeared; the weight that rises during the expansion phase and falls during the compression phase is higher at the end of a cycle of the composite engine than it was at the beginning of the cycle. So if we could find two Carnot engines with different efficiencies we could build a perpetual-motion machine, a machine that creates mechanical energy out of nothing.

Carnot assumed that a perpetual-motion machine can't exist and concluded that all ideal heat engines operating between two heat reservoirs have the same efficiency. This conclusion is called *Carnot's theorem*.

Of course, the idealizations that define an ideal Carnot engine can't be realized in practice. Friction can't be entirely eliminated, heat doesn't flow between bodies at precisely the same temperature, and while a thermally insulated gas sample is expanding or contracting it is never in a state of equilibrium; it doesn't literally pass through a sequence of equilibrium states. Yet, as Carnot argued, actual heat engines can be made to resemble an ideal Carnot engine so closely that they would indeed create mechanical energy out of nothing if that were possible. So while Carnot's theorem rests on assumptions that can't be realized in practice, it is experimentally testable. And those assumptions allowed Carnot to discover the principle behind the motive power of heat – that all Carnot cycles operating between two heat reservoirs at given temperatures have the same efficiency.

From Carnot's Theorem to the Second Law

Carnot's theorem rests on the false assumption that heat is a conserved substance, like mass or energy in Newtonian physics. Having been convinced by Joule's experiments that heat and mechanical energy are interconvertible at a fixed rate of exchange, William Thomson and Rudolf Clausius independently set out to discover what became of Carnot's argument when they replaced the false caloric assumption with the newly established interpretation of heat as a form of energy and the assumption that energy, in its new inclusive form, is conserved.

Carnot's conclusion that ideal Carnot engines working between a given pair of heat reservoirs all have the same efficiency survived this replacement. So did the idea behind Carnot's proof – hooking up ideal engines with different efficiencies, one operating in its direct mode, the other in its reverse mode. But Thomson and Clausius could not now deduce from Carnot's argument that if two Carnot engines had different efficiencies they could be coupled to make an engine that created mechanical energy, for the First Law implies that if in the course of a cycle the gas in the cylinder withdraws a quantity of heat Q_2 from the hot reservoir and delivers a quantity of mechanical energy, or work, W to its environment, it must transfer a quantity of heat $Q_1 = Q_2 - W$ to the cold reservoir; in the course of a cycle the engine does an amount of work on its environment equal to the difference between the heat it extracts from the hot reservoir and the heat it delivers to the cold reservoir:

$$W = Q_2 - Q_1$$

We can continue to define an engine's efficiency as the ratio W/Q_2, but that ratio now becomes

$$W/Q_2 = 1 - Q_1/Q_2,$$

Suppose now that two ideal Carnot engines working between a given pair of heat reservoirs had different efficiencies. As in Carnot's argument, we could then build a composite engine in which the first, more efficient, engine operates in its direct mode and the second, less efficient, engine in its reverse mode. The second engine accepts a quantity of heat Q_1^* from the cold reservoir, delivers a quantity of heat Q_2^* to the hot reservoir, and absorbs a quantity of mechanical energy $Q_2^* - Q_1^*$.

Suppose we make Q_1^* equal to Q_1. Then, because the second engine is less efficient than the first engine, the preceding displayed formula shows that Q_2^* must be less than Q_2, so that $Q_2 - Q_2^* > 0$; and the difference $Q_2^* - Q_1^*$, the work absorbed by the second, less efficient engine during a cycle, is less than $W = Q_2 - Q_1$, the work done by the first, more efficient engine. At the end of a cycle the composite engine has withdrawn a positive quantity of heat $Q_2 - Q_2{}^*$ from the hot reservoir and done an equal quantity of work on its surroundings. I*t has converted heat drawn from a single reservoir into mechanical energy, leaving the world otherwise unchanged.* Alternatively, we can arrange matters so that the second engine absorbs all the mechanical energy the first engine delivers. *Then the composite engine extracts heat from the cold reservoir and delivers an equal quantity of heat to the hot reservoir.*

Thomson and Clausius independently surmised that both of these italicized statements describe impossible engines. This surmise quickly earned the status of a physical law, the second main law of the new science of thermodynamics. It has two equivalent forms:

> —*No cyclic engine can withdraw heat from a single source and convert it into mechanical energy, leaving the world otherwise unchanged.* In other words, a cyclic engine can't be perfectly efficient. It can't transform all the heat it withdraws from a heat source into mechanical energy; it needs a heat sink, to which it conveys some of the energy extracted from the hot reservoir.
>
> —*No cyclic engine can transfer heat from a cooler reservoir to a warmer reservoir, leaving the world otherwise unchanged.* In other words, a cyclic engine that transfers heat from a cooler reservoir to a warmer reservoir needs to be supplied with mechanical energy. Even if it is as efficient as possible, it must extract less heat from the cold reservoir than it deliver to the hot reservoir, the difference being equal to the mechanical energy that disappears in the course of a cycle.

Thomson and Clausius drew two remarkable inferences from the Second Law. Thomson used it to define temperature in a way that made it independent of thermometers. Clausius used the Second Law and Thomson's definition of temperature to define entropy and formulate the law of entropy change.

Thermodynamic Temperature

Before Thomson used Carnot's theorem to redefine temperature, scientists had defined a body's temperature as the reading of a thermometer that has been in thermal contact with the body long enough for both to reach equilibrium. They had formulated empirical laws that relate the volume and pressure of a gas sample to the temperature that would be measured by an ideal-gas thermometer – a thermometer that uses an ideal gas as the substance that expands or contracts when it absorbs or releases heat.

Since the work W done by a Carnot engine during a cycle is the difference $Q_2 - Q_1$ between the quantities of heat extracted from the hot reservoir and delivered to the cold reservoir, Carnot's theorem tells us that the engine's efficiency W/Q_2 depends only on the ratio Q_2/Q_1. This ratio takes an especially simple form if the working substance of a Carnot engine is an ideal gas. (The internal energy of an ideal gas sample depends only on the sample's temperature, and the sample's pressure, volume, and temperature are related by the "equation of state" $PV = nRT$, where R is a constant that has the same value for ideal gas samples of any chemical composition, and n is the number of gram-molecules (moles) in the sample.) From this definition of an ideal gas one can deduce that

$$Q_1/Q_2 = T_1/T_2 .$$

But Carnot's theorem shows that the efficiency of an ideal Carnot engine – and hence the ratio Q_1/Q_2 – doesn't depend on any property of the working substance. Thomson accordingly proposed, in 1848, that physicists *define* the ratio T_1/T_2 between the temperatures of two bodies as the ratio Q_1/Q_2 between the quantities of heat extracted from and delivered to the bodies when they act as (or are in thermal equilibrium with) the heat reservoirs in an ideal Carnot engine. The preceding displayed equation then shows that temperature ratios defined in this way coincide with temperature ratios measured by an ideal-gas thermometer. As Thomson explained,

> In the present state of physical science, therefore, a question of extreme interest arises: *Is there any principle on which an absolute thermometric scale can be founded?* It appears to me that Carnot's

> theory of the motive power of heat enables us to give an affirmative answer.
>
> The relation between motive power and heat, as established by Carnot, is such that *quantities of heat*, and *intervals of temperature*, are involved as the sole elements in the expression for the amount of mechanical effect to be obtained through the agency of heat; and since we have, independently, a definite system for the measurement of quantities of heat, we are thus furnished with a measure for intervals according to which absolute differences of temperature may be estimated.[3]

While the first law of thermodynamics makes heat independent of calorimeters, Thomson's definition, based on Carnot's theorem and the first law of thermodynamics, makes temperature ratios independent of thermometers.

Thomson's definition fixes the zero-point of the temperature scale but leaves the unit of temperature undefined. Thomson suggested that the unit of temperature be chosen to make the freezing and boiling temperatures of water at standard pressure be separated by one hundred units. The resulting unit of temperature, denoted by °K or just K, is called the Kelvin in his honor.

Entropy and the Law of Entropy Change

Thomson's definition of temperature paved the way for Rudolf Clausius's discovery, in 1854, of a previously unnoticed property of isolated macroscopic systems in thermal equilibrium. In 1865 he named this property *entropy*. Clausius deduced from the second law of thermodynamics that the entropy of an undisturbed system, unlike its energy and its mass, may change with time, and he proved that it never decreases. Later he extrapolated this law to the largest macroscopic system, the physical universe:

> The entropy of the world tends toward a maximum.

After discussing Clausius's definition of entropy and his derivation of its law from the second law of thermodynamics, I'll argue that this extrapolation is invalid.

We can rewrite the equation that Thomson used to define temperature,

3 Thomson, William (Lord Kelvin), *Philosophical Magazine* October 1848

$$Q_1/Q_2 = T_1/T_2,$$

as

$$Q_2/T_2 - Q_1/T_1 = 0\,.$$

This equation is suggestive. Suppose that the working substance in Carnot's imaginary engine absorbs a small quantity of heat, denoted by dQ, as it transitions between two neighboring equilibrium states. While the working substance is in contact with the hot reservoir at temperature T_2, these small heat transfers add up to Q_2; while it is in contact with the cold reservoir at temperature T_1, they add up to $-Q_1$. The preceding displayed equation tells us that in the course of a complete cycle the changes dQ/T add up to zero.

This conclusion suggests the question: Do the changes dQ/T add up to zero around *any* cyclic sequence of reversible transitions between equilibrium states of *any* macroscopic system? If the answer is yes, we can further infer that the sum of the small quantities dQ/T along any reversible path connecting any two equilibrium states, A and B, of the system, is the difference $S_B - S_A$ between the values of a property S of the system in these states.

To see why, suppose that S does indeed return to its initial value when the state of the system it refers to undergoes any reversible cyclic process. Suppose that one series of reversible changes, ABC, passes through an arbitrary state B and another series of reversible changes, CDA, brings the system back to its initial state A via another arbitrary state D. Since we are assuming that the quantities dQ/T sum to zero around any reversible path, they sum to zero along the path ABCDA. So the sum along the path CDA is the negative of the sum along the path ABC. But the sum along the path CDA is the negative of the sum along the reverse path ADC, because each summand in the sum of incremental changes dQ/T along ADC is the negative of the corresponding summand in the sum along CDA –when you reverse a heat inflow it becomes a heat outflow and vice versa. So the sum of the quantities dQ/T has the same value along the paths ABC and ADC – two arbitrarily selected paths connecting two given states, A and C. Which is what we set out to prove.

So if the quantity dQ/T sums to zero around any closed sequence of reversible changes, it represents a small change in a property of the system – a quantity that, like temperature, pressure, and internal energy, has the same value whenever the system is in an equilibrium state. Clausius named this property entropy. But does the quantity dQ/T sum to zero around any closed sequence of reversible changes?

To show that it does, Clausius invented a simple but ingenious argument.

Clausius's Argument

Recall that Carnot's proof of his theorem (that all ideal Carnot engines working between two heat reservoirs at given temperatures have the same efficiency) deploys two Carnot engines. One operates in its direct mode, withdrawing heat from the hot reservoir, depositing heat in the cold reservoir, and doing work on the piston. The other operates in its reverse mode, withdrawing heat from the cold reservoir, depositing heat in the hot reservoir, and having work done on it by the piston. Thomson's and Clausius's revised versions of Carnot's proof also use coupled pairs of Carnot engines, one member of each pair operating in its reverse mode. Clausius's proof that the quantity S returns to its initial value when the state of the macroscopic system it refers to undergoes any reversible cyclic process uses a generalized version of this idea.

Suppose a macroscopic system undergoes a sequence of n small, not necessarily reversible changes. At the kth change the system accepts a small quantity of heat dQ_k from a reservoir R_k at temperature T_k. dQ_k can be positive, negative, or zero. (So the word "accepts" doesn't have its everyday meaning in this context; but the new meaning will help us keep the signs straight.) The nth and final change brings the system back to its initial state, so the sequence of n changes constitutes a cycle. We can make n large enough so that the discrete sequence of changes approximates a continuous sequence as closely as we wish.

Clausius now imagines a collection of n ideal Carnot engines. The kth engine works between reservoir R_k at temperature T_k – the temperature of the system during the kth change – and a reservoir R_0 used by all n Carnot engines. The kth engine replaces the heat drawn by the system from the reservoir R_k by delivering an equal amount of heat dQ_k to R_k and it withdraws a quantity of heat dQ_{k0} from the common reservoir R_0. Thomson and Clausius showed in their revised version of Carnot's proof that

$$dQ_k/T_k = dQ_{k0}/T_0 \quad (k = 1, 2, \ldots, n)$$

Summing these n equations, we get:

$$\Sigma_k\, dQ_k/T_k = (\Sigma_k dQ_{k0})/\, T_0\, ,$$

where the symbol Σ_k indicates that the quantity on its right is to be summed over all values of the index k between 1 and n. The sum $\Sigma_k\, dQ_{k0}$ represents the total quantity of heat that has been withdrawn from the reservoir R_0 in the course of a cycle. Call this quantity Q_0.

At the end of a cycle the reservoirs with labels from 1 to n are back in their initial states, and so is the system. But a quantity of heat Q_0 has been withdrawn from the reservoir R_0. Suppose Q_0 is positive. Since energy is conserved, the cyclic process just described must produce an equal quantity of mechanical energy. But the Second Law says this can't happen; heat drawn from a single source can't be transformed entirely into mechanical energy. So Q_0 must be non-positive.

The last displayed equation then tells us that the sum $\Sigma k\, dQ_k/T_k$ is also non-positive:

$$\Sigma_k\, dQ_k/T_k \leq 0 \text{ for any cyclic process} \quad (1)$$

If the cycle is reversible, the preceding equation still holds if we reverse the sign of each of the quantities dQ_k. But this is impossible unless the sum $\Sigma_k\, dQ_k/T_k$ equals zero:

$$\Sigma_k\, dQ_k/T_k = 0 \text{ for a reversible cyclic process} \qquad (2)$$

Earlier we saw that If S is a quantity that changes by an amount dQ/T when the system it refers to reversibly absorbs a quantity of heat dQ, and if S returns to its initial value when the system undergoes any reversible cyclic process, then the difference $S_C - S_A$ doesn't depend on the path that connects A to C. Equation (2) says that S does indeed returns to its initial value when the system undergoes any reversible cyclic process. So S is indeed a physical property of the system, the property Clausius named entropy.

The preceding argument defines the entropy S of a system in thermal equilibrium through the amount by which S changes when the system passes from one equilibrium state to another via a series of reversible changes. Consequently, S is defined only up to an additive constant, which we may identify with the entropy of the system in an arbitrary reference state.

Next suppose that an isolated, or undisturbed, macroscopic system evolves from equilibrium state A to equilibrium state C via an irreversible process. Because the system is isolated, it doesn't interact with the outside world; no heat flows into or out of it and no work is done on or by it.

For example, imagine a box divided into two compartments by a removable partition. Suppose that initially, one compartment is filled with air at temperature T_2, the other with air at a lower temperature T_1, and assume (as can always be arranged) there's no pressure difference between the compartments. After the partition is removed, the sample settles into an equilibrium state at a single uniform temperature between T_1 and T_2. If during this process a quantity of heat Q flows from the set of molecules that make up the subsample that was initially at temperature T_2 to the set that make up the subsample that was initially at temperature T_1, the entropy of the whole sample changes by an amount $(Q/T_2 - Q/T_1)$. In this example the entropy of the system increases (because $T_2 > T_1$). We want to show that the entropy of any isolated, or undisturbed, system either increases or doesn't change when it passes from one state of thermal equilibrium (in which all its macroscopic properties have definite, unchanging values) to another.

Call the initial equilibrium state A and the final equilibrium state C. Now construct a cycle by adding a *reversible* sequence of changes CDA, in which heat can flow into or out of the system and work can be done on or by it. (Any two equilibrium states can be joined by infinitely many reversible sequences of reversible changes, represented by smooth curves joining the points that represent the two states.) Applying equation (1) to the cycle ACDA gives:

$$\Sigma_{\mathrm{AC}}\, \mathrm{d}Q_k/T_k + \Sigma_{\mathrm{CDA}}\, \mathrm{d}Q_k/T_k \leq 0$$

Because no heat flows into or out of the system during the irreversible leg AC, the first sum vanishes. The second sum is $S_A - S_C$, the entropy change between C and A. So

$$S_C \geq S_A \qquad (3)$$

When a closed, or isolated, system evolves between two equilibrium states, its entropy cannot decrease; it must either increase or stay the same.

This is the law of entropy change. Clausius deduced it, in the way I've described, from the Second Law of thermodynamics. His argument shows that if the entropy of an isolated system were to decrease, we could build a device – the device described in Clausius's proof – that converts all the heat drawn from a single source into mechanical energy.

The law doesn't say that the entropy of an isolated system must increase during an irreversible change in the system's state; it merely says that it cannot decrease. For example, if an isolated system consists of two subsystems with initially different temperatures, the law tells us that heat can flow only from the warmer to the cooler subsystem, because heat flow in the opposite direction would decrease the system's entropy. But the law doesn't tell us that heat *must* flow between the subsystems. As in our earlier example, the two subsystems might be separated by a membrane that doesn't conduct heat.

A separate physical law governs heat flow. It contains a parameter that characterizes the thermal conductivity of a medium at each point and in every direction. If you change the sign of the time coordinate in this law, and hence the direction in which time increases, you get a different law – one that predicts that heat flows *up* a temperature gradient and thus violates the law of entropy change.

More generally, separate macroscopic laws govern the processes through which an isolated system evolves from one equilibrium state to another through a sequence of irreversible processes. The law of entropy change constrains the laws that govern these processes, but doesn't, on its own, require that irreversible processes cause the entropy of an isolated system not initially in thermal equilibrium to increase.

Defining Entropy for Systems Not in Thermal Equilibrium

What is the scope of Clausius's definition of entropy? So far we've defined entropy for homogeneous systems, such as uniform gas samples, in thermal equilibrium. The preceding proof of the law of entropy non-decreases applies to two states, A and C, of such a system connected by an irreversible path. Now consider an undisturbed nonuniform gas sample. Can we assign entropies to its states between A and C and apply the law of entropy non-decrease to them? For example, can we assign entropy to an isolated gas sample whose temperature varies smoothly from point to point?

Clausius defined the entropy S of a system in thermal equilibrium through its changes between neighboring equilibrium states. The First Law of thermodynamics tells us that if the volume V of a gas sample enclosed in a cylinder fitted with a movable piston changes by an amount $\mathrm{d}V$ and a positive or negative quantity of heat $\mathrm{d}Q$ flows into the sample, its internal energy U changes by an amount dU given by:

$$\mathrm{d}Q = \mathrm{d}U + P\,\mathrm{d}V.$$

Here P is the pressure the sample exerts on the piston. If the piston has area A and travels a short distance dx, it does work $PA\mathrm{d}x$ or $P\mathrm{d}V$. So the preceding equation says that the work done by the sample when its volume changes must come from either the sample's internal energy or from heat that flows into the sample or from a combination of the two. The preceding equation shows that unlike $\mathrm{d}U$ and $\mathrm{d}V$, $\mathrm{d}Q$ isn't the change in a quantity that characterizes the state of the gas sample; although heat is a form of energy and energy is conserved, a gas sample doesn't hold heat in a separate account. It does, though, hold entropy in a separate account. Clausius's theorem enables us to replace $\mathrm{d}Q$ by $T\mathrm{d}S$, where $\mathrm{d}S$, like $\mathrm{d}U$, $\mathrm{d}V$, and $\mathrm{d}T$, denotes a small change in a quantity that characterizes the state of the gas sample. The preceding equation then takes the form

$$T\mathrm{d}S = \mathrm{d}U + P\mathrm{d}V \qquad \text{or}$$

$$\mathrm{d}S = (\mathrm{d}U/T) + (P/T)\mathrm{d}V$$

Suppose now that we can draw imaginary surfaces separating the region occupied by an inhomogeneous gas sample into cells in each of which the temperature and pressure are (nearly) uniform. We will be able to do this if the gas sample isn't too chaotic. But if the entropy of the sample is to equal the sum of the entropies of our imaginary cells, two more conditions must be satisfied: The sample's energy U must be *additive*: it must equal the sum of the energies of the parts. And work performed by the system must be equal to the sum of the amounts of work $P\mathrm{d}V$ performed by the parts. Then, and only then, can we equate the change $\mathrm{d}U$ in the sample's energy to the sum of the energy changes of the parts. Enrico Fermi emphasized this condition in his lectures on thermodynamics:

> The entropy of a system composed of several parts is very often equal to the sum of the entropies of all the parts. This is true if the energy of the system is the sum of the energies of all the parts

> and if the work performed by the system is equal to the sum of the amounts of work performed by all the parts. Notice that these conditions are not quite obvious and that in some cases they may not be fulfilled. Thus, for example, in the case of a system composed of two homogeneous substance, it will be possible to express the energy as the sum of the energies of the two substances only if we can neglect the surface energy of the two substances where they are in contact. The surface energy can generally be neglected only if the two substances are not very finely divided; otherwise it can play a considerable role. [4]

Energy is usually additive under laboratory conditions. Parcels of air in the atmosphere, whose energies include a height-dependent gravitational potential energy, also satisfy the additivity condition, as do parcels of gas in the interiors of stars. But the Earth as a whole and the Sun as a whole don't satisfy the condition, because the gravitational potential energy of a system held together by the mutual gravitational attraction of its parts isn't a sum of contributions each of which depends only on local conditions. This is a consequence of what physicists call the "long-range" character of Newtonian gravitational force – the fact that it decreases only with the inverse square of the separation between mutually attracting particle. This makes gravitational potential energy a nonlocal property, in contrast to the energies associated with the short-range forces that come into play when two gas atoms or molecules collide, or the binding energies of atoms in gas molecules or of atoms and molecules in crystals. So Clausius's definition of entropy doesn't apply to planets, stars, galaxies, or other self-gravitating systems. Consequently, it doesn't apply to the universe.

An isolated macroscopic gas sample settles into thermal equilibrium – a macroscopically quiescent state of uniform temperature. Could a self-gravitating system be in such a state? It could be in a state of *mechanical* equilibrium, in which the weight of every small parcel of gas balances the pressure force the surrounding gas exerts on the parcel's boundary. But a calculation based on Archimedes's law of hydrostatic equilibrium shows that if the system also had uniform temperature – if it were in thermal as well as mechanical equilibrium – it would extend to infinity and have infinite mass. *A finite self-gravitating system can't evolve toward a state of uniform temperature – a state without internal heat flow.*

4 Fermi, Enrico, *Thermodynamics*, Prentice-Hall, New York, 1937, p. 53.

In a star, energy is liberated by thermonuclear reactions in the deep interior and escapes from the surface layers in the form of light and neutrinos. As a result, the star's structure and its energy slowly change. In a protostar whose central temperature is not yet high enough for thermonuclear reactions to liberate energy, the loss of energy by radiation from the surface layers is uncompensated, so the protostar's energy decreases. This has a surprising consequence. The *virial theorem* (like entropy and its law, discovered and named by Rudolf Clausius) implies that when a self-gravitating system of particles is in mechanical equilibrium, its gravitational energy, a negative quantity, is equal in magnitude to twice the system's thermal energy (the combined kinetic energy of its particles). The sum of the system's gravitational energy and its thermal energy is consequently a negative quantity equal in magnitude to its thermal energy (or half its gravitational energy). So when the system *loses* energy, its thermal energy, and hence its mean temperature, *increase*. Thus a self-gravitating system in mechanical equilibrium has negative heat capacity (the ratio between a small inflow of heat – in the present example a negative inflow – and the consequent small increase in temperature). In contrast, the second law of thermodynamics predicts (and experiments confirm) that when a system (to which the law applies) loses energy, its temperature drops. Of course, as we've just seen, the domain of the Second Law doesn't include self-gravitating systems – and for basically the same reason that a self-gravitating gas mass grows hotter as it radiates away energy: gravitational potential energy is non-local and hence non-additive.

Rudolf Clausius's discovery of entropy and his reformulation of the Second Law as the law of entropy change enabled him to recast thermodynamics as a deductive mathematical science, an adjunct to Newtonian particle physics and Newtonian fluid mechanics. That science has proved to be enormously useful, and its predictions have been uniformly successful in a very wide domain. But, as we've seen, that domain is limited to macroscopic systems whose structure is sufficiently smooth and whose energy can be expressed as a sum of the energies of nearly uniform subregions. As long as "the entropy of the universe" remains undefined, we can assign no meaning to Clausius's statement that the entropy of the universe tends toward a maximum.

V
Atomism

Thermodynamics' two main laws are generalizations about the outcomes of possible experiments. The first law says that in any experiment in which heat appears in a closed (or isolated or undisturbed) system a strictly proportional quantity of mechanical energy, such as gravitational potential energy, disappears. The second law states that it is impossible to build a device whose only effect is to transfer heat from a cooler to a warmer body; or, equivalently, to build a device whose only effect is to transform heat drawn from a single heat reservoir into mechanical energy. Some of Clausius's contemporaries succeeded in constructing a theory based not on these empirical generalizations but on Newton's laws of motion together with an assumption that, in the mid-nineteenth century, seemed far less secure than thermodynamics' two empirical laws: that matter consists of invisible and indivisible particles in motion. Not until the early 1900s did that assumption, the atomic hypothesis, become the atomic fact (as the physicist Richard Feynman put it in his *Lectures on Physics*).

Atomic theories of macroscopic systems and processes have a statistical character; they are formulated in the mathematical language of probability theory. Ludwig Boltzmann showed that the statistical description of an isolated sample of an ideal gas has a counterpart to entropy. As we'll discuss, this counterpart, sometimes called *statistical entropy*, is well defined in a much broader domain than Clausius's entropy. It is a measure of the randomness inherent in a statistical description. So for the sake of clarity, I'll usually call it *randomness* rather than statistical entropy.

While probability theory itself is uncontroversial, its interpretation in statistical theories based on the atomic hypothesis raises a question that goes to the heart of the problem of free will versus determinism: *Do the probabilities that figure in these theories represent incomplete knowledge? Or do they represent a kind of objective indeterminacy distinct from quantum indeterminacy?* To answer this question we'll need to take a closer look at atomism and statistical physics.

The Roots of Atomism

In the fifth century BCE Leucippus and his pupil Democritus (460 – 370 BCE) conjectured that the world consists of indivisible particles, or atoms, moving about in otherwise empty space. Atoms came in a variety of sizes and shapes. Some had hooks or barbs on their surfaces, allowing them to form molecules; others, such as water atoms, were smooth. Atoms of fire and soul were small and round and moved at great speeds.

Epicurus (341–270 BCE) based his philosophy on this thoroughly materialistic picture of the world. But he amended it in one important way. In his ethical philosophy Epicurus assumes that we are free to shape our conduct. But how does freedom fit into a world in which our thoughts and actions are determined by invisible atoms and their motions? Epicurus suggested that there is an element of randomness in atomic motions. Atoms occasionally swerve from their paths, and these swerves have an irreducibly random character. This assumption allowed Epicurus to reconcile his materialistic picture of the world with an ethical philosophy based on human freedom.

How did the atomists come up with their picture of the world? In his *History of Western Philosophy* Bertrand Russell suggested that the atomic hypothesis was a lucky guess:

> By good luck, the atomists hit on a hypothesis for which, more than two thousand years later, some evidence was found, but their belief, in their day, was nonetheless destitute of any solid foundation.[1]

This judgment rests on a widely accepted view of the "scientific method" in the physical sciences:

> Scientists do experiments and make observations. These may suggest hypotheses that have testable implications. Further experiment or observations then confirm or disconfirm these hypotheses. Those that survive become physical laws.

The same view of the scientific method underlies the definition of physical laws in the *Oxford English Dictionary*, 3d edition: Physical laws are statements "inferred from particular facts, applicable to a defined group or class of phenomena, and expressible by the statement that a particular phenomenon always occurs if certain conditions be present." Definitions like these reflect a view of physical laws that was popular (though not uncontroversial) among physical scientists in the late nineteenth century. Its most famous and influential advocate was the physicist and philosopher of physics Ernst Mach. Mach held that physicists should abjure metaphysical notions, such as Newton's "absolute" space and "absolute" time; they should rely entirely on objective experimental and observational facts. Einstein tells us that Mach's elaboration of this view in *The Science of Mechanics* helped motivate his own search for a theory that would relate the structure of spacetime to its contents.

But Mach went further. He held that physical laws are nothing more than economical summaries of experimental and observational data. So he rejected theoretical constructs that went beyond the data. These constructs included atoms and Einstein's principle of special relativity. During the opening years of the twentieth century, Mach's view of physical laws fell out of favor. One group of experiments, which included experiments designed to test Einstein's predictions about the random motions of microscopic particles suspended in a liquid (Brownian motion), showed that atoms actually exist. Another group of experiments confirmed special relativity's predictions about phenomena involving particles traveling at nearly the speed of light. Though physicists continued to insist, and still insist, that scientific hypotheses stand or fall by their testable predictions, special relativity and the experimental vindication of the atomic hypothesis showed that not all scientific hypotheses are empirical generalizations. Could Leucippus and Democritus have viewed atomism as a scientific hypothesis? Would they even have understood the notion of a scientific hypothesis?

About Leucippus we know only that he was Democritus's mentor. Democritus, however, was famous in his day as a mathematician as well as a philosopher. He wrote treatises on number theory, geometry, and astronomy. And he was no run-of-the-mill mathematician (if such people exist). According to the historian of Greek mathematics Thomas Heath, Democritus "already had the idea of a solid being the sum of an infinite number of parallel planes, or indefinitely thin laminae [plates], indefinitely near together; a most important anticipation of the same thought which led to such fruitful results in Archimedes." The thought in question is the idea underlying the integral calculus, an early version of which Archimedes used to calculate the area of a segment of a parabola and the surface area and volume of a sphere. Two centuries before Archimedes, Democritus used the same idea to prove that the volume of a circular cone is one-third the product of its height and the area of its base, and that the volume of a pyramid is likewise one-third the product of its height and the area of its base. So Democritus was, to say the least, an exceptionally creative and insightful mathematician.

Nowadays many mathematicians specialize in subjects remote from the concerns of the natural sciences. But Greek mathematics, physics, and astronomy were closely interwoven strands of a single project. Heath tells us that these subjects "were born together at the beginning of the sixth century."[2] Collections of geometric and arithmetic facts, rules, and algorithms had existed for millenniums in Egypt and Mesopotamia before Thales of Miletus (about 624 – 547 BCE), after visiting Babylon and Egypt and acquainting himself with some of their mathematical treasures, invented a new kind of mathematics: He proved the first theorems.

A theorem is a statement about mathematical objects and relations accompanied by a proof. Written out in full, a proof is a chain of logical deductions that connects the statement to be proved to a small fixed collection of unproved statements, or axioms, and a small fixed collection of terms, such as *point* and *line* in plane geometry, that are implicitly defined by the axioms that mention them. Logical deductions preserve truth: if the premises of a deduction are true,

2 Heath, T. L.. *A History of Greek Mathematics.* (Cambridge University Press, 2013).

then so is its conclusion. So if geometry's axioms are true, then so are its theorems. Greek mathematicians and their successors until well into the nineteenth century regarded the axioms of number theory and geometry as self-evident truths. By contrast, many *theorems* about numbers and geometric objects were far from self-evident. Some were amazing. For example, Greek mathematicians discovered and proved that the lines connecting the vertices of any triangle to the midpoints of the opposite sides intersect in a single point, which divides each of these lines into two segments one of which is half the length of the other.

Again, Pythagoras (about 570 - 490 BCE) or one of his followers discovered that *geometric* atomism is false: a side and a diagonal of a square have no common measure; it isn't possible to choose a unit of length so that the lengths of a side and a diagonal are both whole-number multiples of that unit. And since this claim could be proved, it had to be true.

Mathematics is a creative enterprise. Theorems and their proofs need to be invented. Mathematicians imagine properties of numbers or geometric objects that might be true; then they imagine possible routes to a proof. For its first practitioners the method of conjecture and proof was a way of arriving at hidden truths about the world, more reliable than divine revelation or authority. And its successes seemed to reveal that numbers and geometric forms underpin reality. Mathematics was the language of the new sciences of physics and astronomy. Even the parts of Greek philosophy that were not explicitly mathematical aspired to the logical rigor epitomized by mathematical demonstrations. Plato held that the study of mathematics was a necessary propaedeutic to the study of philosophy. The motto "Let no one unversed in geometry enter here" was inscribed over the door of his Academy.

Greek mathematicians didn't draw a sharp distinction, as we do today, between pure mathematics and mathematical physics. They recognized that mathematics deals with idealized objects – that dimensionless points, perfectly straight lines, and perfectly smooth planes exist nowhere in the world of experience. But mathematical discoveries and the method of conjecture and proof seemed to reveal hidden and often surprising truths not only about idealized mathematical objects and relations but also about the physical world.

Theorems about plane figures and solids are also testable propositions about their inexact real-world counterparts. And experience invariably supported these propositions to within the accuracy of the measurements. Mathematics as a deductive science (rather than as a set of rules, formulas, and algorithms) was seen as a new and powerful tool in humankind's struggle to understand and control nature, allowing its practitioners to glimpse a hidden reality that undergirds experienced reality but can be described only in mathematical language.

Thales, reputed to have been the first mathematician, devised a method for finding the distance of a ship at sea and a method for calculating the heights of the pyramids, both based, presumably, on the theorem that pairs of corresponding sides of similar triangles have the same ratio. Both methods also rely on identifying light rays or lines of sight with the straight lines of theoretical geometry. By combining geometry with observation, Greek astronomers conjectured and then proved that the Earth and the Moon are spheres. They also arrived at the modern explanation of solar and lunar eclipses. In the fourth century BCE, Heraclides suggested that the daily motions of the stars are produced by the Earth's rotation (rather than by the rotation of a solid Earth-centered sphere in which the stars are embedded). A century later Aristarchus anticipated Copernicus's model of the planetary system, in which the Moon is a satellite of Earth, Earth is a planet, and the planets are satellites of the Sun. Aristarchus also invented and applied a method of deducing the ratio of the Sun's and the Moon's distances from observations of their directions when the Moon is exactly half full. His account marks the debut of trigonometric ratios such as sines and tangents.

Archimedes took over the method and style of Euclid's *Elements* (composed around 300 BCE) and extended its vocabulary to include the notions of mass, force, and mechanical equilibrium. He invented the notion of center of mass, discovered the principle of the lever (objects at opposite ends of a weightless beam balance if the ratio of their weights is equal to the reciprocal of the ratio of the distances of their centers of mass from the fulcrum), and founded the science of hydrostatics, which includes the proposition known as Archimedes' principle: An object wholly or partly submerged in water experiences an upward, or buoyant, force equal in magnitude to the weight of the water it displaces; so the weight of a submerged object is equal to its weight in air minus the weight of an equal volume of water; if the difference is negative, the object experiences a net upward force.

Archimedes's *The Method of Mechanical Theorems* used his method of infinitesimals, along with what we'd now call physical concepts – the notion of center of mass and the principle of the lever – to calculate areas and volumes of *geometric* objects. For Archimedes, number theory, geometry, and mechanics were interlocking parts of a single mathematical structure. Archimedes was famous in the ancient world for putting the physical principles he had discovered to practical uses. His inventions include the screw pump (for bailing ships), block-and-tackle pulley systems, and a mechanical model of the solar system. He also invented devices that were used to defend his native Syracuse, in Sicily, from the invading Romans: a powerful catapult, a device for lifting ships out of the water and then dropping them (Archimedes's claw), and (perhaps) arrays of either lenses or parabolic mirrors for focusing the Sun's rays on an invader's ships and setting them on fire.

To return to the atomic hypothesis, we can perhaps think of it, in the context of Greek mathematics and natural science, as the response to a question that is both philosophical and scientific. One way to state the question is to ask what would happen if you split a drop of water in two, then split each of the resulting drops in two, and kept on in this way. Leucippus and Democritus might perhaps have reasoned as follows.

Either the process of splitting will go on forever or it will come to an end. If every substance is indefinitely divisible into smaller and smaller parts with the same physical properties, the only possible answer to the question "Why do different substances have different properties?" is "They just do." If the splitting process eventually ends, we have a choice. Either the process ends with a large collection of indivisible water droplets or with a large collection of objects that aren't water droplets. The first possibility doesn't help us to understand why alcohol differs from water. If, as I'm assuming, Leucippus and Democritus were seeking to understand such differences, they would have opted for the second possibility – that if you continue to split a drop of water you will eventually arrive at particles whose properties don't coincide with *but serve to explain* the properties of water. In the historical context I have sketched, these would have been objects describable in the vocabulary of Euclidean solid geometry. If you ask the right question – a question that would have arisen quite naturally in the context of Greek mathematics, physics, and natural philosophy – the atomic hypothesis seems almost inevitable.

We can interpret the question "What happens when you keep splitting a bit of matter into smaller and smaller parts?" as either a philosophical or a scientific question, depending on the kind of answer we're looking for. Physicists ask questions whose answers can be expressed in the vocabulary of mathematics, even if they have to invent new mathematics to do so. Biologists ask questions that fit into a conceptual framework based in part on the premises that all living organisms are genetically related and have evolved from a single ancestral population. Empiricists ask questions that can be answered by an appeal to sensations and perceptions; metaphysicians, by an appeal to basic metaphysical principles, such as Gottfried Wilhelm Leibniz's principles of sufficient reason (everything has a reason) and identity of indiscernibles (distinct objects can't have identical properties).

Atomism Becomes a Testable Hypothesis

The atomic hypothesis had immediate explanatory value. It supplied a starting point for an explanation of why water becomes a solid at low temperatures and becomes a vapor at high temperatures. It explained why you can mix wine and water in various proportions to make homogeneous liquids with experimentally distinguishable properties. It explained why a sample of brine exposed to air eventually turns into salt crystals. Newton, with not much more *direct* evidence than Democritus had had, was a convinced atomist. To quote the historian of the physical sciences Alan E. Shapiro:

> The corpuscular theory of matter was thus for Newton not a hypothesis but a demonstrated principle established with as much certainty as the existence of God or the theory of gravitation. He cites two principal sorts of evidence in its support: various substances penetrate the pores [empty spaces between atoms] of bodies, like water into vegetable and animal matter, and quicksilver [mercury] into metals; and transparency, which shows that

> light passes through the pores of a great variety of bodies [such as thin layers of gold][3]

While Newton considered his view that matter consists of atoms to be "a demonstrated principle" on a level with his law of gravitation, he considered another of his strongly held views – that light rays are streams of particles – to be a hypothesis in need of experimental support – which his own experiments consistently *failed* to provide.

Newton's view of atoms and their role in observable phenomena differed from Democritus's in a crucial respect: Newton assumed that the motions and interactions of atoms obey his three laws of motion. Thus enriched, Democritus's picture of the physical world became a testable scientific hypothesis. It not only offered qualitative explanations of some phenomena but could also serve as the basis for testable predictions about the outcomes of precise measurements. Prominent among the empirical rules Newton hoped the atomic hypothesis could explain was Boyle's law.

In 1659 Robert Boyle, assisted by Robert Hooke, built an improved version of Otto von Guericke's air pump, and with its help began a series of experiments on what Boyle called "the spring of the air" – the fact that a finite sample of air resists compression and also expands to fill any airtight enclosure. Their experiments supported the quantitative relation between the pressure and the volume of an enclosed gas sample (at a given temperature) that is now called Boyle's law: the pressure, or force per unit area, the sample exerts on the walls of its container is inversely proportional to the sample's volume.

To account for this remarkably simple rule, Newton proposed that air atoms repel one another. He showed that this hypothesis would yield Boyle's law if (a) each gas atom repelled only its nearest neighbor and (b) the repulsive force was inversely proportional to the distance between the two atoms. "But whether elastic fluids do really consist of particles so repelling each other," he wrote in the *Principia*, "is a physical question" – that is, a question to be settled by experiment.

And experiment did settle it. Although Newton's model accounted for Boyle's law, another of its predictions contradicted an empirical

3 Shapiro, A.E. *Fits, Passions and Paroxysms: Physics, Method and Chemistry and Newton's Theories of Colored Bodies and Fits of Easy Reflection.* (Cambridge University Press, 1993).

law established by Guillaume Amontons between 1700 and 1702: *If an air sample's volume is held constant, its pressure increases with its temperature.* Newton's model predicts instead that the force a gas sample exerts on the walls of its container depends only on the average number of atoms per unit volume and not on the temperature.

In his *Hydrodynamica*, published in 1738, Daniel Bernoulli proposed an atomic model of gases that not only accounted much more simply for Boyle's law than Newton's model but also predicted Amontons' law and other, as-yet-undiscovered, empirical laws.

Like Newton, Bernoulli postulated that in a uniform gas sample the number of gas particles per unit volume has the same value everywhere. But instead of postulating that the particles of a quiescent gas sample are themselves motionless, as Newton had done, he postulated that they have large random velocities. He imagined that each particle travels at constant speed in a fixed direction until it collides with another particle or with a wall, after which it sets off in a new direction. He assumed that the duration of a collision is much shorter than the time interval between collisions, as it is in collisions between billiard balls. (Bernoulli's successors assumed further that particle collisions are elastic: the combined momentum of a pair of colliding particles is the same before and after a collision and so is their combined kinetic energy.)

Newton had attributed the expansive tendency of a gas sample to a repulsive force between the gas's constituent particles. In Bernoulli's model the fact that a gas sample expands to fill a box of any size follows directly from Newton's first law of motion: a particle on which no force acts continues to move in a straight line with constant speed; collisions with other particles deflect a gas particle but can't confine it to a finite region.

Appealing to Newton's second and third laws of motion, Bernoulli equated the force an enclosed gas sample exerts on a wall of its container to the rate at which gas particles transfer momentum (mass times velocity) to the wall when they strike it and rebound. He assumed that such encounters are elastic, conserving momentum. The force per unit area, or pressure, is then proportional to the number of particles per unit volume and to two-thirds the average value of a particle's kinetic energy (half the product of its mass and the square of its velocity). So Bernoulli's model predicts the following simple relation between the pressure, P, the number of gas particles per unit volume, n, and the average kinetic energy of a particle, $\langle \frac{1}{2} mv^2 \rangle$:

$$P = \tfrac{2}{3}\, n \,\langle \tfrac{1}{2}\, mv^2 \rangle \,\langle E \rangle \,\langle E \rangle$$

If we now *assume* that the temperature T of a gas is some constant multiple of the mean kinetic energy of the gas molecules, Bernoulli's formula becomes

$$P = constant \times nT = constant \times (N/V)T,$$

where N is the number of particles in a sample, V is the sample's volume, and n is the number of particles per unit volume. This relation includes not only Boyle's law and Amontons' law (which says that the pressure in a gas sample whose volume is held constant increases with its temperature) but also Charles's law (1780), which says that at constant pressure the temperature of an ideal-gas sample is proportional to its volume. Moreover, because the displayed equation doesn't depend on any property of the gas particles, it contains Amedeo Avogadro's hypothesis (1811): *Equal volumes of different ideal gases at the same temperature and the same pressure contain equal numbers of particles (N in the displayed equation).*

We can't call the preceding argument a *derivation* of the gas laws, because it merely *assumes* that the average kinetic energy of the gas particles is proportional to the gas sample's temperature, as defined by Thomson through Carnot's theorem. It doesn't explain why. As we'll see, a full derivation of the ideal-gas law didn't appear until over a century after Bernoulli proposed his model.

Atoms and Molecules

During the eighteenth century chemists devised experimental methods for distinguishing homogeneous "substances" from mixtures, and for distinguishing "elements" – substances that couldn't be broken down into simpler substances –from "compounds." They perfected the chemical balance and used it to discover that elements combine to form compounds in fixed proportions by weight. They also discovered that weight (or mass) is conserved when a compound is formed or when it is separated into elements.

During the first decade of the nineteenth century John Dalton (1766 – 1844) realized that simple assumptions about the atomic composition of elements and compounds could bring greater order to the wealth of experimental data chemists had built up during the preceding century. He postulated that every chemical element consists of identical, unchanging atoms that differ from the atoms of other elements, and that chemical compounds consist of identical molecules made up of atoms of the compound's constituent elements.

These postulates immediately accounted for a striking regularity in the data. Sometimes two elements combine to form different compounds. For example, carbon and oxygen combine to form the gases we now call carbon monoxide and carbon dioxide. Experiments showed that if samples of the two compounds contain equal masses of carbon, the mass of oxygen in the heavier sample is twice the mass of oxygen in the lighter sample. To take another stock example, arsenic and oxygen likewise form two compounds. If samples of these compounds contain equal masses of arsenic the ratio between the masses of oxygen in the two samples is exactly 5/3 – again a ratio of two small whole numbers. These examples suggest, and experiment confirms, the following rule: If element #1 forms two (or more) different compounds with element #2, then samples that contain the same mass of element #1 contain masses of element #2 that are in the ratio(s) of small whole numbers. Before Dalton, this finding seemed inexplicable. Dalton's postulates make the explanation obvious: *Every molecule made up of atoms of different kinds contains a small integral number of each kind.*

Dalton's axioms by themselves didn't allow chemists to predict the proportions of elements in every compound. In the preceding examples you might guess, correctly, that the two oxides of carbon are, in modern notation, CO and CO_2, but you might also guess, incorrectly, that the oxides of arsenic contain a single arsenic atom (in fact they contain two). Dalton suggested that a "rule of simplicity" should guide the choice of chemical formulas for compounds. But efforts to follow this rule led to inconsistencies. Then in 1808, the same year in which Dalton's *A New System of Chemical Philosophy* appeared, Joseph Gay-Lussac (1778 – 1850) published a new empirical law, the law of combining volumes. Suppose a sample of gas #1 and a sample of gas #2 combine to form a compound. Gay-Lussac measured the *volumes* of the reactants (gas #1 and gas #2) and the product *at the same pressure and temperature.* He found that the measured volumes always had the ratio of small whole numbers. For example, suppose that a sample of hydrogen and a sample of oxygen react to form water vapor, with no hydrogen or oxygen left over. The volumes of the hydrogen and oxygen samples and the volume of the water vapor produced by their reaction, all measured at the same temperature and pressure, turn out to have the ratios 2:1:2.

Three years later, in 1811, Avogadro published his conjecture that equal volumes of dilute gases at the same temperature and the same pressure contain equal numbers of particles (atoms or molecules). Dalton's axioms then imply that in the preceding example two atoms of hydrogen combine with one atom of oxygen to make two molecules of water. So each water molecule ends up with half an oxygen atom – or, as we now say, half an oxygen molecule.

But Dalton and many of his contemporaries refused to consider the possibility that the "atoms" of oxygen (or hydrogen or nitrogen or chlorine) might be what we now call diatomic molecules. Before 1925, when quantum mechanics began to explain chemical bonds and chemical reactions, chemists had no firm basis for understanding what holds molecules together. They thought in terms of inherent affinities between atoms. Such affinities, many believed, couldn't exist between atoms of the same kind. Finally, in 1861, Stanislao Cannizzaro (1826 – 1910), ignoring this prejudice, used Avogadro's hypothesis to infer from experimental data on combining volumes a great many mutually consistent molecular formulas of gaseous elements and compounds – enough to convince most of his colleagues that the hypothesis was sound.

This brief account of the atomic hypothesis before it became the atomic fact suggests what is missing, or at least underemphasized in the conventional account of "the scientific method." Radical and consequential scientific hypotheses are rarely if ever "suggested" by experimental and observational data. They are, as Einstein put it, free creations of the human mind. And, like other creative acts, they help bring into being novel and unforeseen kinds of order.

Maxwell's Statistical Interpretation of Thermal Equilibrium

A box one centimeter on a side filled with air at standard temperature and pressure contains roughly 3×10^{19} (30 billion billion) molecules. A completely detailed Newtonian description of this collection of particles would specify, among other things, the three position coordinates and three velocity components of each molecule. Yet experiments show that two undisturbed samples of the same gas, or of the same sample at different times, have the same measurable properties if the samples have the same temperature and the same pressure. What accounts for the vast disparity between how much data we need to characterize the state of an undisturbed gas sample at the molecular level of description and how much we need to characterize the sample's macroscopic, or thermodynamic, state?

Thermodynamics interprets equilibrium states of an isolated system as states of maximum entropy. So to understand thermal equilibrium at the molecular level we need to know the molecular-level counterpart of entropy.

Daniel Bernoulli postulated that the distribution of molecular positions and velocities in a gas sample in thermal equilibrium is uniform (the same at all positions) and isotropic (the same in all directions). Now, statistical uniformity and isotropy are aspects of randomness, or lack of order, at the molecular level. This in turn may suggest that *randomness, appropriately defined, is the molecular counterpart of thermodynamic entropy and that thermal equilibrium is the condition in which the positions and velocities of gas molecules are distributed as randomly as possible*. To pursue this suggestion we need to make the notion of a distribution (of molecular positions or velocities) more precise, and we need to define randomness.

Collisions between the molecules of an enclosed, undisturbed gas sample redistribute their positions and velocities. Experiments show that an undisturbed gas sample settles into a state in which its mass density, and hence the number of molecules per unit volume, has the same value everywhere, up to measurement error. We can plausibly assume that in thermal equilibrium the gas molecules in any small region are traveling in random directions. But what about their speeds?

We might guess that in thermal equilibrium the molecules all have the same speed. But this can't be right, because when we apply Newton's laws to collisions between, say, billiard balls, we find (and experience confirms) that their post-collision speeds differ in general from their pre-collision speeds. Newton's laws require only that the combined momentum and the combined kinetic energy of colliding particles have the same values before and after a collision – provided the collision converts a negligible part of the pair's initial kinetic energy into the particles' internal energies. So molecular collisions tend to randomize the distribution of molecular speeds. Or so it may seem intuitively. But what, precisely, does *randomize* mean in this context? And what is the "most random" distribution of the velocities of molecules in an undisturbed gas sample? In a short paper that appeared in 1860 James Clerk Maxwell offered answers to these questions.

To frame the question he began by representing a gas particle's velocity by a point in what he called *velocity space*. By introducing a Cartesian reference frame – three mutually perpendicular lines passing through the same point (the coordinate origin) – we can assign a particle three Cartesian coordinates, x, y, z, and we can assign its velocity three Cartesian components: u parallel to the x axis, v parallel to the y axis, and w parallel to the z axis. We can think of u, v, w as the *position* coordinates of a particle's representative point in *velocity* space.

The velocities of particles in a macroscopic gas sample are represented by a dense swarm of points in velocity space. To represent the distribution of these points we begin by dividing velocity space into identical cells whose edges are parallel to the coordinate axes. By the "occupation number" of a cell I mean the number of particles whose representative points lie in that cell at a given moment. By the "fractional occupation number" I mean the occupation number divided by the number of particles in the sample. Maxwell assumed that there is a range of cell sizes for which the fractional occupation numbers of contiguous cells are nearly equal and each fractional occupation number is proportional to the cell volume. If these conditions are met we can approximate the fractional occupation number of a cell in velocity space with dimensions du, dv, dw centered on the point (u, v, w) by the product $f(u,v,w)\mathrm{d}u\mathrm{d}v\mathrm{d}w$ of a smoothly varying function of position in velocity space, f(u,v,w), and the cell's volume, du dv dw. We can think of the fractional-occupation-number density f(u,v,w) as the mass density of a smoothly varying "probability fluid." (In the present context *probability* is synonymous with *fractional occupation* number.) The "mass" of probability fluid doesn't change with time – probability fluid is conserved – because the quantities

$f(u,v,w)\,du\,dv\,dw$ summed over all the cells in velocity space, add up to 1.

Maxwell expressed the assumption that the particle velocities in a gas sample in thermal equilibrium are randomly distributed by two mathematical conditions:

1. *The distribution is isotropic.* This condition implies that the value of f at a point (u, v, w) depends only the point's distance from the origin (0, 0, 0) and hence only on the quantity $u^2 + v^2 + w^2$, the square of that distance. That is,

$$f(u,v,w) = F(u^2 + v^2 + w^2) \text{ for some function } F$$

2. *A molecule's velocity components in mutually perpendicular directions are uncorrelated.* Let $g(u)\,du$ denote the proportion, or fraction, of the particles in the sample for which u lies in a given interval du. Then the proportions with v and w in given intervals must be $g(v)\,dv$ and $g(w)\,dw$, since there are no preferred directions in velocity space. And if velocity components in mutually perpendicular directions are uncorrelated, the proportion of the molecules in the gas sample for which u, v, w lie in given intervals must be the product of the three individual proportions:

$$F(u^2 + v^2 + w^2) = g(u)\,g(v)\,g(w).$$

This equation has the solution

$$F(u^2 + v^2 + w^2) = C^3 \exp[-\,a(u^2 + v^2 + w^2)],$$

Here a and C are constants and exp is the exponential function: $\exp(x) = e^x$, the number e raised to the power x. ($e = 2.718 \ldots$, the limit of $(1 + 1/n)^n$ as n increases without limit.)

The fractional occupation numbers $f(u,v,w)dudvdw$, summed over all the cells in velocity space, must add up to 1. This requirement determines the constant C:

$$f(u,v,w) = (a/\pi)^{3/2} \exp[-\,-\,a(u^2 + v^2 + w^2)] \qquad (1)$$

Maxwell's two randomness conditions therefore determine the distribution of particle velocities up to a single parameter, the constant a in (1).

Now recall Bernoulli's formula for the pressure of an ideal gas,

$$P = \tfrac{2}{3}\, n\, \langle E \rangle,$$

where n is the number of particles per unit volume, $E = \tfrac{1}{2} m(u^2 + v^2 + w^2)$, the kinetic energy of a particle, and $\langle E \rangle$ is the average value of E. This formula becomes the equation of state of an ideal gas, $P = nT$, if we define the temperature T of an ideal gas as $\tfrac{2}{3} \langle E \rangle$.

Using Maxwell's formula (1) for the distribution of particle velocities, we can calculate the average value of the squared particle velocity, and hence the average kinetic energy, which we then set equal to T. This calculation gives $a = m/2T$, so Maxwell's formula becomes

$$f(u,v,w) = (m/2\pi T)^{3/2} \exp[-E/T] \qquad (2)$$

If we want to measure temperature on the Kelvin scale, we must replace T in this formula by kT, where k is a constant, called Boltzmann's constant, whose value depends on the units in which we choose to measure mass, length, and time.

Because Maxwell's formula was not suggested by observational or experimental evidence but instead followed from mathematical conditions that Maxwell devised to make precise the notion of molecular disorder, it had a mixed reception. Most supporters of the atomic hypothesis considered the assumption that the equilibrium distribution of particle velocities doesn't favor any position or direction in space plausible, but many balked at the assumption that in thermal equilibrium the components of a gas particle's velocity are statistically uncorrelated.

Nevertheless, Maxwell and other advocates of atomism used the formula to construct quantitative theories of heat conduction, molecular diffusion, and viscous dissipation of internal motions in gases. These theories made testable prediction, which matched experiment up to experimental error.

Experimental physics didn't reach a stage in which Maxwell's formula (2) could be *directly* tested for another half-century. Since then, physicists have devised many direct tests. In one of the conceptually simplest of these a molecular beam on its way to a detector passes through narrow slits in two discs spinning at slightly different rates around a common axis parallel to the direction of the beam. The pair of spinning discs acts as a speed filter, like a pair of timed traffic lights. For a given difference between the rates at which the discs are spinning, a molecule that gets through a slit in the first disc also gets through a slit in the second disc if, and only if, its speed falls within narrow limits. This arrangement sorts molecules in the beam by their speeds.

The Doppler effect supplies another way of sorting molecules in a hot gas by their speeds. Each atom or molecule in a hot gas emits light in narrow bands of frequency, called *emission lines*. An observer equipped with a spectroscope sees light whose intensity in each narrow frequency band is the sum of the intensities of the light waves emitted in this frequency band by the gas's individual atoms. Because each atom is moving relative to the observer, its light is shifted at each frequency by an amount proportional to the line-of-sight component of its velocity. Each emission line emitted by a hot gas is accordingly shifted in frequency and broadened: its frequency is shifted by an amount proportional to the average line-of-sight velocity of the atoms; it is broadened by the atoms' thermal velocities – their velocities relative to the average velocity of the gas molecules. The profile (intensity versus frequency) of an emission line has the same shape as the distribution of line-of-sight velocities in the gas. If that distribution is Maxwellian, so is the emission-line profile.

Experiments like these show that Maxwell's formula accurately characterizes the velocity distribution of gas molecules in thermal equilibrium. Thus they support the statistical assumptions on which the formula rests – assumptions that lend specificity to the notion of a "random" distribution of molecular velocities.

In 1867 Maxwell revisited the problem of the equilibrium distribution of molecular velocities.[4] Referring to his earlier assumption that mutually perpendicular components of molecular velocity are independently distributed, he wrote: "As this assumption may appear precarious, I shall now determine the form of the [molecular-velocity distribution] function in another way." Maxwell's new argument was part of a detailed account of how encounters between gas particles redistribute energy and momentum. It rests on the assumption that *the incoming velocities of colliding molecules are statistically uncorrelated*, a state of affairs Ludwig Boltzmann later called "molecular

4

chaos." Like the "precarious" assumption in Maxwell's first derivation of formula (2), molecular chaos posits the absence of statistical correlations – but between the velocities of different molecules in a gas sample instead of between velocity components of the same molecule.

The 1867 paper contained another major result. Maxwell considered two intermingled samples of different gases. He assumed that initially both gases are in thermal equilibrium, each at its own temperature. He then showed that encounters between the molecules of the two gases tend to erase differences between their average kinetic energies. Earlier we saw that Bernoulli's relation $P = \frac{2}{3}\, n\langle E\rangle$ becomes the equation of state of an ideal gas if we define the temperature T of a gas in thermal equilibrium as a constant multiple of its average molecular kinetic energy $\langle E\rangle$. Maxwell's argument shows that the average molecular kinetic energy of a gas sample in thermal equilibrium does indeed have the defining property of temperature: energy exchanges between two intermingled gas samples with different average molecular kinetic energies tends to equalize these average molecular energies. Notice that this crucial link between Daniel Bernoulli's atomic theory of an ideal gas and the empirical gas laws was supplied not by experiments or observations but by the *predicted* outcome of a *thought* experiment.

Entropy of an Ideal-gas Sample

We've seen that certain thermodynamic properties of an ideal-gas sample in thermal equilibrium have counterparts in the atomic model:

Thermodynamic property	Atomic counterpart
Pressure	average momentum transfer per unit area
Temperature	average particle kinetic energy
Energy	total particle kinetic energy
Mass density	particle mass × average number density
Mass	total particle mass

Conspicuously absent from this list is entropy. In 1872 Ludwig Boltzmann filled this gap. He defined a statistical counterpart of thermodynamic entropy that is also a measure of randomness in systems and conditions for which entropy can't be defined. I'll refer to this measure as *randomness*. It's also called *statistical entropy*. Randomness is fundamentally a mathematical property of abstract probability distributions. It is a precise mathematical counterpart of the word's colloquial meaning.

In one way Boltzmann's definition of randomness is less general than Clausius's definition of thermodynamic entropy. Clausius's definition of entropy applies to macroscopic systems in thermal equilibrium or local thermal equilibrium. Boltzmann's definition of randomness applies to states of an ideal-gas sample characterized by arbitrary distributions of molecular position and velocity.

Boltzmann's 1872 paper also purported to demonstrate that the growth of an isolated gas sample's randomness is a consequence of Newton's laws of motion, applied to the atomic model of an ideal gas sample. Although, as we'll see, Boltzmann's proof of this claim was flawed, it broke new ground. It raised the possibility that Clausius's law of entropy change is a special case of a far broader and more deeply rooted generalization about processes that destroy (and, as I'll argue later, also create) order in the physical universe.

Boltzmann's Formula for the Randomness of a Gas Sample

We can derive a formula for the (thermodynamic) entropy of an ideal-gas sample in thermal equilibrium by combining Clausius's definition of entropy with the equations that relate the energy and pressure of such a sample to its temperature, the number of particles in the sample, and the sample's volume.

We saw earlier that the first law of thermodynamics and Clausius's definition of the entropy S imply the following formula for the change in S between neighboring equilibrium states of an enclosed gas sample:

$$\mathrm{d}S = \mathrm{d}U + P\,\mathrm{d}V$$

Let's apply this formula to an ideal-gas sample. If we choose a temperature scale in which

$$U = N\,T \quad \text{and} \quad P = (N/V)\,T$$

then $$\mathrm{d}S = N(\mathrm{d}T/T + \mathrm{d}V/V). \tag{3}$$

Since $d[\log(x/x_0)] = dx/x$,

$$S = N[\log(T/T_0) + \log(V/V_0)] + \text{constant} \qquad (4)$$

T_0 and V_0 in equation (4) are arbitrary constants. They need to be there because the argument of the logarithm function – the quantity x in the expression $\log x$ – must be a positive pure number; it mustn't depend on how we choose the units of distance, time, and mass. Any reference temperature T_0 and reference volume V_0 will do because, as we've seen, entropy is undefined up to an additive constant, and changing the values of T_0 and V_0 merely adds a constant to the right side of formula (4).

In Newtonian mechanics six quantities define the state of a particle – its three position coordinates x, y, z and its three velocity components u, v, w. If we know the laws that govern the forces acting on a particle and are given the values of these six quantities at a single moment, we can use Newton's law of motion to calculate their values at any later or earlier moment. Boltzmann united Maxwell's velocity space with position space to form a six-dimensional *state space*. The state of each particle in a gas sample containing N particles is represented by a point in this six-dimensional space, and a microstate of the whole sample is represented by a cloud of N such points.

To construct a statistical description of the state of the cloud Boltzmann divided the six-dimensional state space into cells of equal six-dimensional volume dx dy dz du dv dw. A statistical description of the state of the cloud of representative points assigns each cell a fractional occupation number

$f(\mathrm{x},y,z,u,v,w)\, dx\, dy\, dz\, du\, dv\, dw$ or $f(\tau)d\tau$, where the Greek letter τ (tau) stand for the six coordinates of a representative point in a molecule's state space and $d\tau$ stands for the cell volume $dx\, dy\, dz\, du\, dv\, dw$ in this space. We assume we can make the dimensions of a cell so small and the value of N is so large that $f(\tau)$ varies smoothly from cell to cell, like the density of a continuous fluid.

If the assembly of N identical particles is a gas sample of volume V in thermal equilibrium, the particles are uniformly distributed in space, so the fraction of the total number of particles N contained in a region of volume dV is dV/V. The particles also have a Maxwell distribution in velocity. Finally, Boltzmann assumed that a particle's position isn't correlated with its velocity.

Then $f(\tau)\mathrm{d}\tau = (m/2\pi T)^{3/2}\,(1/V)\,\exp[-E/T]\,\mathrm{d}\tau$.

Take the negative logarithm of both sides of this formula, noting that $\log(x^k) = k\log x$, where k is any positive or negative real number:

$$-\log\,[f(\tau)\mathrm{d}\tau] = E/T + \log(\mathrm{T}/\mathrm{T}_0)^{3/2} + \log(V/V_0), \qquad (5)$$

where $T_0^{3/2} = (m/2\pi)^{3/2}\,(\mathrm{d}u\;\mathrm{d}v\;\mathrm{d}w)$, $V_0 = \mathrm{d}x\;\mathrm{d}y\;\mathrm{d}z$.

Only the first term on the right side of equation (5), E/T, varies from cell to cell (E is the kinetic energy of a molecule; its average value $\langle E\rangle$ is $3T/2$). Multiply equation (5) through by the fractional occupation number of a cell, $f(\tau)\mathrm{d}\tau$, sum over all the cells, and denote the resulting average by angle brackets $\langle\;\rangle$:

$$\langle\text{-}\log[f(\tau)\mathrm{d}\tau\rangle = \langle E/T + \log(T/T_0)^{3/2} + \log(V/V_0)\rangle$$

$$\langle\text{-}\log[f(\tau)\mathrm{d}\tau]\;\rangle = \langle E/T\rangle + \log(T/T_0)^{3/2} + \log(V/V_0)$$

$$= 3/2 + \log(T/T_0\,)^{3/2} + \log(V/V_0)$$

Now recall that

$$S = N[\log(T/T_0)^{3/2} + \log(V/V_0)] + \text{constant} \qquad (4)$$

From equations (5) and (4) we conclude that

$$S = N\,\langle\text{-}\log f(\tau)\mathrm{d}\tau\rangle\;\text{-}\log f(\tau)\mathrm{d}\tau \qquad (6)$$

Like the statistical counterpart of the energy E of an ideal-gas sample consisting of N identical particles, the statistical counterpart of the sample's thermodynamic entropy S is N times an average value – the average value of the quantity – $\log f(\tau)\mathrm{d}\tau$, the negative logarithm of the fractional occupation number. This quantity isn't the average value of a physical property of the gas particles. Instead it characterizes a *probability distribution* – the distribution of the points that represent single-particle states in Boltzmann's six-dimensional position-velocity space.

Boltzmann conjectured that formula (6) defines the atomic counterpart of entropy not just for samples of an ideal gas in thermal equilibrium but of *any* smooth distribution of single-particle states in position- velocity space. When the particles are uniformly distributed in space and have a Maxwell velocity distribution, the formula coincides with the formula for the thermodynamic entropy of an ideal gas. But formula (6) defines the randomness of a much wider class of macroscopic states of an undisturbed gas sample than Clausius's definition of thermodynamic entropy. For example, it defines the randomness of a gas sample in which all the particles have the same speed. Boltzmann now asked: With the statistical counterpart of entropy defined in this way, does Clausius's law of entropy change have a counterpart in his and Maxwell's statistical description of an ideal-gas sample?

Boltzmann's Transport Equation and His *H* Theorem

In his 1867 paper Maxwell used Newton's laws of motion to study the trajectories of colliding particles. He then used the results of that study to describe how particle collisions cause the average values of single-particle properties like kinetic energy to change with time. Boltzmann deepened and extended Maxwell's theory. He asked: How do particle collisions change the distribution of fractional occupation numbers in six-dimensional position-velocity space?

Boltzmann's answer to this question, his *transport equation*, equates the rate at which the fractional occupation number of a given cell in the position-velocity space of a particle in an ideal-gas sample changes with time to the difference between the rate at which particles enter the cell and the rate at which they leave the cell.

Boltzmann's transport equation became the starting point of a flourishing discipline, kinetic theory, which he and his successors used to construct testable statistical theories of processes such as heat conduction, viscous dissipation of relative fluid motions, and the diffusion of one gas through another.

Boltzmann also used his description of how molecular collisions alter the populations of cells in position-velocity space to prove his *H* theorem (Boltzmann denoted randomness by $-H$):

> *The randomness of an undisturbed gas sample not initially in thermal equilibrium increases until it reaches its largest possible value, which corresponds to thermal equilibrium.*

As you'd expect, Maxwell's form of the molecular-velocity distribution maximizes the distribution's randomness.

Boltzmann's proof of the H theorem is logically rigorous if one accepts its premises. Unlike the law of entropy non-decrease, it applies to any isolated gas sample, not just to samples in local thermal equilibrium. And unlike Clausius's law of entropy non-decrease, the H theorem says that the randomness of an undisturbed gas sample not in thermal equilibrium actually increases with time if it isn't already as large as it can be.

Criticism of the *H* Theorem

Many of Boltzmann's contemporaries rejected his statistical counterpart of entropy and his *H* theorem. Some of the critics rejected the atomic hypothesis and everything based on it. Advocates of Energetics, a school whose members considered the idea that heat is molecular motion to be unscientific, were especially energetic in their criticisms. The introduction to Part I of Boltzmann's *Lectures on Gas Theory* (1896)[5] contains a long list of "works on Energetics"; the foreword to Part II (1898) contains a supplementary and equally long list of "attacks on the theory of gases." Among the distinguished critics it cites are Ernst Mach, Pierre Duhem, Henri Poincaré, Robert Mayer, Wilhelm Ostwald – and V. Lenin.

But the criticism Boltzmann took most seriously was raised in 1876 by Josef Loschmidt. Loschmidt argued that the *H* theorem cannot show what it purports to show – that the randomness of an undisturbed gas sample not initially in thermal equilibrium increases with time – because it relies on a description of particle encounters based on Newton's laws of motion. These laws are time-reversible: they don't change when one replaces the time coordinate *t* by its negative, thereby reversing the direction in which *t* increases. So if we replace *t* by –*t* in Boltzmann's account of how molecular encounters alter the populations of cells in position-velocity space *and also replace his description of the sample's initial state by a description of its final state*, his proof of the *H* theorem becomes a proof that the randomness of an undisturbed gas sample decreases with time.

5 Ludwig Boltzmann, 1895. *Lectures on Gas Theory*, translated by Stephen G. Brush, (New York, Dover Publication, 2011)s

Here's how Boltzmann describes Loschmidt's objection in his *Lectures on Gas Theory*.

> Let a gas be enclosed by absolutely smooth, elastic walls. Initially there is an unlikely but molecular-disordered state – for example, all the molecules have the same velocity. After a certain time the Maxwell velocity distribution will nearly be established. We now imagine that at time t, the direction of the velocity of each molecule is reversed, without changing its magnitude. The gas will now go through the same sequence of states backwards. We have therefore the case that a more probable [more highly random] distribution evolves into a less probable [less random] one, and the quantity *H* [the negative of randomness] increases as a result of collisions.[6]

We can rephrase Loschmidt's objection to Boltzmann's derivation of his *H* theorem as a question: How can molecular collisions, governed as they are by time-reversible laws, produce an irreversible macroscopic change in a gas sample? As Boltzmann points out immediately following the passage just quoted, the time-reversed picture differs from the picture envisaged in his proof of the H theorem in a crucial respect: The proof assumes (as did Maxwell's earlier proof that collisions between the molecules of intermingled gases tend to equalize their temperatures) that *the incoming (or initial) velocities of colliding molecules are uncorrelated. This assumption can't be true in the time-reversed picture imagined by Loschmidt.* Suppose, Boltzmann writes, that at some initial moment all the molecules have the same speed *a*. Because a collision leaves the combined kinetic energy of the collision partners unchanged, the collision partner of a molecule that emerges from its first collision with a velocity *b*, say, must have speed $\sqrt{(2a^2 - b^2)}$. So in the time-reversed picture, molecules with initial speeds *b* must collide only with molecules with speed $\sqrt{(2a^2 - b^2)}$. The assumption that the initial velocities of colliding particles are uncorrelated doesn't hold in the time-reversed picture.

Yet this assumption is crucial not only to Boltzmann's proof of the *H* theorem. It is a special case of an assumption that – as Boltzmann emphasized – underlies the entire molecular theory of ideal gases, from Bernoulli to Boltzmann. *The molecular theory of gases rests on the assumption that the states of individual molecules in a gas sample are uncorrelated.* If this assumption weren't true, at least to an excellent approximation, we couldn't characterize the state of an ideal-gas

6 ibid. p. 58

sample by the distribution in position-velocity space of the points that represent possible states of a single molecule. Since experience strongly supports the predictions of the molecular theory of gases, it also supports the assumption that single-particle states in an isolated gas sample are statistically uncorrelated. Why then is the assumption true, or at least very nearly true?

Before addressing that question, we need to consider a theory that extends Boltzmann's theory of an undisturbed ideal-gas sample to any undisturbed system of particles whose motions and interactions conform to Newton's laws – Josiah Willard Gibbs's *statistical mechanics*.

Gibbs's Statistical Mechanics

The theory expounded by Gibbs in his *Elementary Principles in Statistical Mathematics, Developed with Special Reference to the Rational Foundations of Thermodynamics* (1902) rests squarely on ideas introduced by Maxwell and Boltzmann but takes a giant step beyond them. Boltzmann, like Maxwell, assumed that the particles of a gas sample are statistically independent; the probability of finding a particle in a particular microstate doesn't depend on the microstates occupied by the remaining $N - 1$ particles. He could therefore represent a macrostate of an undisturbed gas sample by a probability distribution of the microstates of a single gas particle. Gibbs dropped this assumption. His theory represents the macrostates of any N-particle system – not just samples of an ideal gas – by a probability distribution of N-particle microstates. (That is, it assigns a non-negative number to every possible N-particle microstate, and these numbers add up to 1.) Because Gibbs didn't assume that the particles of a macroscopic system have statistically independent probability distributions, this probability distribution – a function of $6N$ position coordinates and velocity components – doesn't factor into N identical functions of a single particle's six position coordinates and velocity components.

Gibbs represented the possible microstate of an undisturbed system of N interacting particles by a point in a 6N-dimensional position-momentum, or phase, space. His account of how the system evolves relies on a reformulation of Newton's laws of motion published by William Rowan Hamilton in 1833.

Hamilton's version of Newtonian mechanics revealed a hidden symmetry between the roles of position and momentum in dynamical processes. It also underpins the most cogent formulation of the principles of quantum mechanics.

Following Boltzmann's example, Gibbs partitioned his phase space into identical cells. He represented a probability distribution of an undisturbed system's microstates by a large (in principle infinite) collection of points distributed among these cells, and he identified the probability that a microstate's representative point lies in a given cell with the cell's fractional occupation number (or its limiting value as the number of cells increases without limit).

Gibbs called the collection of N-particle microstates and the associated probability distribution an *ensemble*. The members of a Gibbs ensemble are *imaginary* N-particle systems. As he wrote, "Let us imagine a great number of independent systems, identical in nature but differing in phase, that is, with respect to their configuration and velocity."[7] Gibbs called the fractional occupation numbers of cells in the $6N$-dimensional phase space probabilities and denoted them by the letter P. But whereas the occupation numbers in Boltzmann's theory represent numbers of gas-particles in an actual gas sample, Gibbs's occupation numbers don't have a concrete interpretation: they refer to imaginary replicas of an N-particle system.

Having defined an ensemble, Gibbs asked: Which ensembles are the statistical counterparts of thermal equilibrium?

Because a macroscopic system has finite spatial extent and finite energy, the points that represent its possible microstates occupy a finite region of the $6N$-dimensional phase space. Since we can make the number of these points as large as we please – the points represent imaginary replicas after all – we can represent the swarm of points that represents them by a continuous fluid of variable density. If we stipulate that the total mass of the fluid is 1, the probability that a representative point lies in any given region of the phase space equals the mass of the fluid within the region. As the system evolves, the density of probability fluid changes smoothly at each point.

7 Gibbs, J. Willard, *Elementary Principles in Statistical Mathematics*, *Developed with Special Reference to the Rational Foundations of Thermodynamics,* (New Haven: Yale University Press, 1902)

Gibbs now deduces from Hamilton's equations of motion that these changes have a remarkable property. Consider the moving point in 6N-dimensional position-momentum space that represents a particular evolving microstate. Gibbs proved that *the density of probability fluid* ***at that moving point*** *is constant in time*. He called this conclusion "the fundamental equation of statistical mechanics."

> To visualize this conclusion, imagine a swirling fluid of variable mass density in physical space. Suppose the color of the fluid varies from white to black through shades of grey in such a way that the fluid is darker where it is denser. If the motion of the fluid preserved the density of every sufficiently small fluid element, its color would remain unchanged as it moved and changed shape.
>
> The actual motion of a fluid of variable density, such as air, doesn't have this property. For example, a sound wave in air produces alternate compressions and rarefactions; fluid elements become smaller and denser as they contract, larger and less dense as they expand. Why does the motion of probability fluid in 6*N*-dimensional position-momentum space differ in this way from the motion of a real fluid like water or air in physical space?
>
> In important ways the two kinds of motion are similar. Newton's laws of motion govern the flow of a real fluid; they also govern – though less directly, of course – the flow of probability fluid in 6*N*-dimensional phase space. Moreover, both flows are conservative: one conserves the mass of physical fluid; the other conserves the mass of probability fluid. Why, then, are the volume and density of an element of probability fluid in 6*N*-dimensional position-momentum space constant in time, while the volume and density of an element of a compressible fluid in physical space vary with time?
>
> Consider an imaginary closed surface, such as a sphere, in physical space. Because mass is conserved, during any given time interval the mass of fluid inside the surface changes by an amount equal to the inflow of mass across the bounding surface (counting outflows as negative

> inflows). The same is true of conserved fluids in Euclidean spaces with any number of dimensions. (A Euclidean space of n dimensions is a space in which the squared distance between two points is equal to the sum of the squares of the n Cartesian-coordinate differences.) But conservation of mass does *not* entail that the volume of a fluid element (and hence its density) never changes.
>
> Now consider a closed surface in a $6N$-dimensional position-momentum space. As before, during any given interval of time the mass of probability fluid inside the surface changes by an amount equal to the inflow of mass across the surface. But the mass of probability fluid that flows across a surface element of unit area in unit time has two parts. The first part results from the fact that the $3N$ position coordinates of a point in $6N$-dimensional position-momentum space that represents an evolving microstate change by small amounts during a short time interval. The second part is new: it results from the fact that the point's $3N$ momentum coordinates also change by small amounts during that time interval. When we calculate the probability current we need to add these two contributions. And when we do that, using Hamilton's version of Newton's laws, we find that the volume, as well as the mass, of an element of probability fluid centered on a point that represents an evolving microstate never changes.

Earlier we denoted by τ the set of six coordinates of a point in the six-dimensional position-velocity space of a gas particle's state space. We denoted by $\mathrm{d}\tau$ a volume element in that space; and by $\mathrm{f}(\tau)\mathrm{d}\tau$ the fractional occupation number, or probability, of a cell of volume $\mathrm{d}\tau$ centered on the point with coordinates τ. Let's now use the same notation for points, volume elements, and fractional occupation numbers in the $6N$-dimensional position-momentum space of an undisturbed macroscopic system containing N particles. Gibbs's "fundamental equation of statistical mechanics" then says that $\mathrm{f}(\tau)$, evaluated at a moving point $\tau(t)$ that represents an evolving microstate of the system never changes. As before, define the randomness

S of the probability distribution $\{f(\tau)d\tau\}$ as the probability-weighted average, or mean, of the negative logarithm of the probability $f(\tau)d\tau$: $\langle$-log[f(τ)d$\tau\rangle$

$$S = \langle -\log \langle f(\tau)d\tau \rangle$$

Gibbs's fundamental theorem implies that if τ (t) denotes the $6N$ coordinates of a moving point that represents an evolving microstate of an undisturbed system, then the mass of probability fluid in an evolving cell, $f[\tau(t)]d\tau(t)$, doesn't depend on the time t.

So neither do its logarithm and the negative mean of its logarithm, the randomness S:

> The randomness of the probability distribution that characterizes an evolving macrostate of an undisturbed macroscopic system is constant in time.

In marked contrast, the entropy of an undisturbed macroscopic system may increase with time.

Boltzmann showed that an undisturbed gas sample is in thermal equilibrium when the randomness of the probability distribution of single-particle microstates has the largest value compatible with a given value of the mean particle energy $\langle E\rangle$. He also showed that the single-particle randomness, multiplied by the number of particles N in a gas sample, plays the same role in the statistical theory of an ideal-gas sample as Clausius's entropy does in thermodynamics. Gibbs proved the analogous propositions for an undisturbed system composed of particles whose motions and interactions are governed by Newton's laws of motion:

> An undisturbed macroscopic system is in thermal equilibrium when the randomness of the probability distribution of its microstates has the largest value compatible with the mean energy E of the imaginary replicas that make up the ensemble that represents the system's macrostate. Call this largest value S_{max}. *S_{max} plays the same role in Gibbs's statistical description of equilibrium states as Clausius's entropy does in thermodynamics.*

Gibbs also showed, as Maxwell had done for an ideal gas, that the parameter T in Maxwell's formula for the randomness- maximizing probability distribution of microstates has the defining property of temperature: when two initially undisturbed macroscopic systems characterized by randomness-maximizing probability distributions with different values of T are brought together and allowed to interact, the system with the larger value of T loses energy to the system with the smaller value, and the systems' combined randomness increases during this process.

After 1925, quantum mechanics replaced Newtonian mechanics as the fundamental theory of matter and radiation on small scales. But thanks to deep structural similarities between quantum mechanics and Hamilton's formulation of Newton's laws, the mathematical framework of Gibbs's theory survived this replacement virtually unchanged. Gibbs's theory, modified to fit the new microphysics, has succeeded brilliantly in predicting the observable properties of matter and radiation in thermal equilibrium.

It also has an important new feature. The six-dimensional position-momentum space, or phase space, of a particle governed by quantum mechanics differs from the phase space of a classical particle. We can partition a classical particle's phase space into arbitrarily small cells of equal (six-dimensional) volume. In contrast, cells in the phase space of a quantum particle have a minimum volume, determined by Werner Heisenberg's indeterminacy relations: the product of the indeterminacies of a position coordinate and the corresponding momentum component can't be less than $h/4\pi$, where h is Planck's constant. So the volume of a cell in the phase space of a quantum particle can't be less than $(h/4\pi)^3$, and the volume $d\tau$ of a cell in the $6N$-dimensional phase space of a collection of N identical quantum particles can't be less than $(h/4\pi)^{3N}$. Now recall that Boltzmann's randomness, like Clausius's entropy, is defined only up to an additive constant; only changes in randomness and entropy are well defined. In Boltzmann's theory this is because we can partition the position-momentum space of a classical particle into cells of arbitrarily small volume. The smaller we choose a cell's volume, the smaller the cells' fractional occupation numbers; the smaller the cells' fractional occupation numbers, the larger the average value of the logarithm of their reciprocals – Boltzmann's randomness. For quantum particles Heisenberg's indeterminacy relations imposes a lower limit on the volume of a cell in position-momentum space, and hence an upper limit on the randomness of an N-particle gas. *Quantum mechanics makes randomness – not just its changes – well defined.* I'll argue later that randomness is an objective property of the physical universe.

Boltzmann's molecular theory of gases replaces thermodynamics' description of the equilibrium states of ideal-gas samples by a statistical theory based on the assumption that gases consist of particles moving and interacting in ways governed by Newton's laws of motion. Gibbs's theory extends Boltzmann's theory to arbitrary macroscopic systems. Both theories have enjoyed great predictive success. Yet, as the preceding discussion has shown, they don't fit smoothly together. Nor does either theory fit smoothly with thermodynamics:

> — In Boltzmann's theory the randomness of the probability distribution that characterizes a macrostate of an undisturbed gas sample containing N particles equals the randomness of the probability distribution that characterizes a macrostate of a single gas particle, multiplied by N. This is a consequence of the assumption that the motions of gas particles are statistically uncorrelated. But as discussed below, Gibbs's theory shows that this relation almost never holds. And even if it were to hold at one moment, it would immediately break down, because, as Boltzmann himself emphasized, encounters between gas particles create statistical correlations between the probability distributions of their velocities.
>
> — How can we reconcile Boltzmann's proof that the randomness of an undisturbed gas sample not initially in thermal equilibrium always increases with the fact, pointed out by Loschmidt, that the proof rests on a *time-reversible* description of the motions of gas particles in the sample?
>
> — How can we reconcile Gibbs's proof that the randomness of a closed system is constant in time with Boltzmann's proof that the randomness of an undisturbed sample of an ideal gas increases with time (unless it already has its largest possible value)?

— Gibbs's proof that the randomness of a closed system is constant in time also clashes with Clausius's law of entropy non-decrease. Yet, as we saw, the randomness of the probability distribution that characterizes an equilibrium state of an ideal-gas sample coincides with the system's entropy.

— Finally, Boltzmann's theory supplies a physical interpretation of the probabilities it assigns particle microstates. It locates the *N* points that represent the microstates of a gas sample's *N* particles in a *single particle's* six-dimensional position-velocity space, and it identifies the probabilities of single-particle states with the fractional occupation numbers of (sufficiently small) cells in that space. Gibbs's statistical mechanics, in contrast, leaves the probabilities it assigns the microstates of a macroscopic system uninterpreted. It identifies system microstates with fractional occupation numbers of cells in the 6*N*-dimensional position-momentum space of an *N*-particle system, but the systems whose representative points are distributed among these cells are imaginary replicas, or, as Erwin Schrödinger called them, "mental copies" of a single, real system. "Now what on earth could it mean, physically," Schrödinger asked, "to distribute a given amount of energy over [a collection of] mental copies?"[8]

We can gain insight into these issues by using some mathematical properties of randomness stated and proved by Claude Shannon in 1948 in "*A Mathematical Theory of Communication,*" the paper that launched information theory.[9]

Conditional Randomness and Correlation Information

Let $\{p_i\}$ denote a discrete set of non-negative numbers p_i that add up to 1. These are the defining properties of probabilities p_i and a probability distribution $\{p_i\}$. The index *i* labels an individual "event" in a discrete set of events $\{i\}$. (It can also label a possible value A_i of a *random variable A*, a mathematical object that has a discrete set

8 Schrödinger, Erwin, *Statistical Thermodynamics*, (Cambridge University Press, 1948) p. 3

9 Shannon, C.E., *The Bell System Technical Journal* (Vol. 27, July, October, 1948) pp. 379–423, 623–656,

of possible values A_1, A_2, ... with corresponding probabilities p_1, p_2, ... These might be the possible outcomes of a measurement of a physical quantity A.)

We interpret the quantity

$$S = -\Sigma_i\, p_i \log p_i$$

as a measure of the probability distribution's randomness.

Denote by S_{max} the largest value of S for which the set of probabilities $\{p_i$ satisfies a given constraint. For example, we might require the mean of a molecule's kinetic energy, $\Sigma_i\, p_i\, E_i$, to have a given value. Boltzmann's theory characterizes the state of thermal equilibrium of a gas sample by the probability distribution that maximizes S subject to this constraint.

We interpret the difference

$$I = S_{max} - S$$

as a measure of a probability distribution's *information*. A probability distribution is maximally informative if $S = 0$, that is, if one of the system's possible state has probability 1 and the rest have probability zero. It is least informative if S has the largest possible value that is consistent with given constraints on the probability distribution. If the number of possible states is finite and equal to n and there are no constraints on the probability distribution, the most random distribution assigns the same probability $1/n$ to each state, and $S = \log n$.

Suppose that the error δ of a particular measurement has a certain probability distribution. Assume that the mean error is 0 and that the average of the squared error, or mean square error, is some real number σ^2. The most random probability distribution of measurement errors δ – the distribution whose randomness is as large as possible subject to these conditions – is then $f(\delta) = (1/\sqrt{(2\pi\, \sigma^2)}) \exp[-\delta^2/2\sigma^2]$. The probability that the error lies between δ and $\delta + d\delta$ is $f(\delta)d\delta$. Maxwell's theory predicts that each component of a molecule's velocity in a gas sample in thermal equilibrium is distributed according to this formula. So the present definition of randomness coincides with Maxwell's.

Now consider a pair of random variables, A, whose possible values are $A_1, A_2, \ldots,$ and B, whose possible values are $B_1, B_2, \ldots$. We assign each pair of possible values A_i, B_j *a joint probability* $P(A_i, B_j)$. Summed over all pairs of indices i, j, these joint probabilities add up to 1. Summed over the possible values B_j of B, the joint probabilities $P(A_i, B_j)$ add up to $P(A_i)$, the probability that A takes the value A_i regardless of the value taken by B:

$$P(A_i) = \Sigma_j P(A_i, B_j). \qquad \text{(a)}$$

Similarly, $$P(B_j) = \Sigma_i P(A_i, B_j). \qquad \text{(b)}$$

We can think of the joint probabilities as entries in a table whose rows are labeled by the index i and whose columns are labeled by the index j. The entries in the ith row sum to $P(A_i)$. So if we divide each of these entries by $P(A_i)$ we get a new set of nonnegative numbers that sum to 1. These are the conditional probabilities $P(B_j|A_i)$; $P(B_j|A_i)$ is the probability that B takes the value B_j given that A has the value A_i :

$$P(B_j|A_i) = P(A_i, B_j)/P(A_i), \quad \Sigma_j P(B_j|A_i) = 1 \quad (i = 1, 2, \ldots)$$

Similarly, the entries in the *j*th column sum to $P(B_j)$, so if we divide the *i*th entry in the *j*th column by $P(B_j)$ we get the conditional probability $P(A_i|B_j)$, the probability that A takes the value A_i given that B has the value B_j :

$$P(A_i|B_j) = P(A_i, B_j)/P(B_j), \Sigma_i P(A_i|B_j) = 1 \quad (j = 1, 2, \ldots)$$

Since the logarithm of a product is the sum of the logarithms of the factors,

$$\begin{aligned}\log P(A_i, B_j) &= \log P(A_i) + \log P(B_j|A_i)\\ &= \log P(B_j) + \log P(A_i|B_j)\end{aligned}$$

If you insert these expressions for log $P(A_i, B_j)$ into the definition of S you will find, after a short calculation:

$$\begin{aligned}S(\{P(A_i, B_j)\}) &= S(\{P(A_i)\}) + \Sigma_i P(A_i)\, S(\{P(B_j|A_i)\})\\ &= S(\{P(B_j)\}) + \Sigma_i P(B_j)\, S(\{P(A_i \mid B_j)\})\end{aligned}$$

or, more succinctly,

$$S(A, B) = S(A) + \Sigma_i P(A_i)S(B|A_i)$$
$$= S(B) + \Sigma_j P(B_j)S(A| B_j)$$

The joint randomness of a pair of random variables *A, B* equals the randomness of one of them plus the weighted average value of the conditional randomnesses of the other.

The random variables *A, B* are said to be (*statistically*) *independent* or *uncorrelated* if, for every pair of indices *i, j* the joint probability $P(A_i, B_j)$ is the product of the individual probabilities:

$P(A_i, B_j) = P(Ai)\ P(Bj)$, just in case *A, B* are independent, or uncorrelated, random variables (RVs)

And this condition implies that the randomness of the joint probability distribution of uncorrelated random variables equals the sum of the randomnesses of the individual probability distributions:

$$S(\{P(Ai, Bj)\}) = S(\{P(Ai)\}) + S(\{P(Bj)\})$$

or more succinctly,

$$\mathrm{S(A, B) = S(A) + S(B)} \quad \text{if A, B are uncorrelated RVs} \ .$$

Now consider the set of all joint probability distributions $\{P(A_i, B_j)\}$ that have a given pair of individual probability distributions $\{P(A_i)\}$, $\{P(B_j)\}$ that satisfy formulas (a), (b) above. Using a standard algorithm of elementary calculus (the method of Lagrange multipliers), one can find the joint probability distributions $\{P(A_i, B_j)\}$ that maximizes the randomness $S(\{P(A_i, B_j)\})$, or $S(A, B)$, subject to these conditions. It turns out to be the joint probability distribution $\{P(A_i)\ P(B_j)\}$.

> The randomness of the joint probability distribution of two random variables has its largest possible value if the variables are uncorrelated.

According to the preceding definition of information, the difference

$$S(\{P(A_i)P(B_j)\}) - S(\{P(A_i, B_j)\})$$

is positive unless A and B are uncorrelated, in which case it vanishes. The difference represents *correlation information* – information associated with correlations between the random variables

A, B. To repeat: Unless the random variables *A* and *B* are statistically independent, the randomness of their joint probability distribution falls short of its largest possible value, the sum of the statistical entropies of the probability distributions of the individual variables, by a quantity that represents information associated with statistical correlations between the variables. (Two random variables are correlated if the probability that one of them takes one of its possible values depends on the value taken by the other random variable.)

If we have three random variables, *A, B, C*, a straightforward generalization of the preceding argument shows that the randomness of their joint probability distribution falls short of its largest possible value, the sum of the randomnesses of the probability distributions of the individual variables, by a quantity that represents information associated with statistical correlations between the random variables – correlations between the pairs *A* and *B, B* and *C, A* and *C*, as well as triple correlations. With the concept of correlation information in hand, we can now resolve the conflicts between Boltzmann's and Gibbs's theories. We can also get a deeper understanding of Loschmidt's objection to the *H* theorem and Boltzmann's reply to that objection.

Boltzmann and Gibbs

Gibbs represented the possible macrostates of an undisturbed system of *N* interacting particles by probability distributions of *N*-particle microstates. The preceding discussion shows that (a) the randomness S of such a probability distribution takes its largest possible value when the microstates of the *N* particles have mutually independent probability distributions; (b) the *N*-particle probabilities are then products of single-particle probabilities, and *S* is *N* times the randomness of the single-particle probability distribution; and (c) the difference between S and its largest possible value represents information associated with correlations between variables that refer to different particles.

Boltzmann assumed, in effect, that correlation information is *initially* absent in an *N*-particle gas sample; the *N*-particle randomness is then initially equal to *N* times the single-particle randomness. Gibbs proved, however, that the *N*-particle randomness of any isolated system is constant in time. Boltzmann's response to

Loschmidt's criticism of the *H* theorem shows that he understood that particle collisions create statistical correlations between particles. His proof of the *H* theorem shows that particle collisions *initially create correlation information* – that in an undisturbed gas sample, an increase in single-particle randomness is accompanied by an equal increase in correlation information. *Collisions transform single-particle information into correlation information*.

Boltzmann's proof of the *H* theorem would be valid if, and only if, the correlation information created by the decay of single-particle information immediately left the system – if, say, it leaked into the system's surroundings.

And if correlation information tended to leak into a nominally undisturbed system's surroundings, the following discussion shows that we would also be able to resolve the conflict between Clausius's law of entropy non-decrease and Gibbs's theorem that the randomness of the probability distribution that characterizes a macroscopic system's macrostate is constant in time. We'd then be led to ask: *Under what conditions does correlation information leak out of a nominally undisturbed macroscopic system? And is there any reason to suppose that such conditions regularly occur?*

Gibbs's Theorem and Clausius's Law

Gibbs's definition of the randomness of an undisturbed macroscopic system applies to a wider class of macroscopic systems than Clausius's definition of thermodynamic entropy. Yet Gibbs proved that the randomness of an isolated, or undisturbed, system is constant in time, whether or not the system is in a state of thermal equilibrium. In contrast, the system's thermodynamic entropy can increase with time if the system is not initially in thermal equilibrium. Gibbs's proof that the randomness of an undisturbed system is constant in time is simple and straightforward. His "fundamental equation of statistical mechanics" shows that both the volume and the "mass" of an element of probability fluid centered on a point that represents an undisturbed system's evolving microstate are constant in time.

In Chapter XII of his *Elementary Principles of Statistical Mechanics* Gibbs addresses this conflict between his theory and Clausius's law of entropy non-decrease. He argues that the probability fluid in

$6N$-dimensional phase space changes with time like a stirred liquid. Suppose the liquid contains coloring matter of variable density.

> If we give the liquid any motion whatever … the density of the coloring matter at any same point [i.e., any point moving with liquid] will be unchanged.... Yet no fact is more familiar to us than that stirring tends to bring a liquid to a state of … uniform densities of its components. … [10]

Imagine the region occupied by the liquid to be divided into cells of equal volume. No matter how small we make these cells, the distribution of coloring matter will eventually appear to be uniform on that scale. In language introduced by Gibbs's successors, the "coarse-grained" probability distribution becomes increasingly uniform while the "fine-grained" distribution never changes. Some of Gibbs's successors therefore proposed to identify thermodynamic entropy with the coarse-grained randomness of the probability distribution that characterizes a macrostate of an undisturbed system in Gibbs's theory..

Gibbs's argument doesn't, however, provide a criterion for the scale of coarse-graining. More important, while his analogy between the flow of probability fluid in phase space and the flow of a colored liquid is illuminating, it doesn't supply a *reason* for the temporal asymmetry of the process. Because the equations that govern the motion of an ideal fluid are time-reversible, any description of the mixing process is likewise reversible if we exclude the effects of molecular diffusion (as Gibbs does). In a single brief passage Gibbs addresses this problem. He suggests that the conflict we've been discussing reflects an inability of mathematical models to capture physical reality:

> But while the distinction of prior and subsequent events may be immaterial with respect to mathematical fictions, it is quite otherwise with respect to the events of the real world. It should not be forgotten, when our ensembles are chosen to illustrate the probabilities of events in the real world, that while the probabilities of subsequent events may often be determined from the probabilities of prior events, it is rarely the case that the probabilities of prior events can be determined from those of subsequent events, for we are rarely justified in excluding the consideration of the antecedent probability of the prior events. [11]

---

10 Gibbs *Elementary Principles* pp. 144-5

11 Gibbs *Elementary Principles* p. 150

Whereas a realistic picture of the physical world distinguishes between the directions of the past and the future, Gibbs's statistical mechanics, like Boltzmann's kinetic theory of gases, rests on mathematical laws – Newton's laws of motion – that don't make this distinction.

Toward a Resolution

Although Clausius's law of entropy non-decrease rests on an empirical generalization (that no cyclic engine can convert heat drawn from a single source into mechanical energy and leave the rest of the world unchanged), no physical law has been confirmed more strongly or in more different ways. Yet Gibbs's statistical mechanics, which supplies what Gibbs in the subtitle of his book called "a rational foundation" for thermodynamics, predicts that the statistical counterpart of the entropy of an undisturbed system is constant in time. If both Clausius and Gibbs are right, as I think they are, it follows that the systems for which Clausius's law holds aren't, strictly speaking, undisturbed. For if they were, and if the domain of statistical mechanics includes the domain of thermodynamics (as it surely does), Gibbs's theorem implies that the law of entropy non-decrease is false. So we're led to ask: When are we justified in treating macroscopic systems as if they didn't interact at all with the outside world?

We can think of any macroscopic system as part of a larger macroscopic system. If we take *that* system to be large enough, we are surely justified in neglecting *its* interaction with its environment when we describe what is happening in the original system. Now recall Gibbs's remark that "while the distinction of prior and subsequent events may be immaterial with respect to mathematical fictions, it is quite otherwise with respect to the events of the real world." Gibbs's statistical mechanics applies to *undisturbed* macroscopic systems – systems that do not interact at all with the outside world. The Second Law, in contrast, refers *directly* to the world of experience.[12] It refers neither explicitly nor implicitly to undis-

12 "We, therefore, put forward the following proposition as being given directly by experience: It is impossible to construct an engine which will work in a complete cycle, and produce no effect except the raising of a weight and the cooling of a heat-reservoir." Max Planck, Treatise on Thermodynamics, trans. Alexander Ogg. Third edition translated from the seventh German edition. Dover Publications, New York, 1945. 9 N. G. van Kampen in Fundamental Problems in Statistical Mechanics,

turbed systems, which are mathematical fictions. This line of reasoning suggests that by investigating the environments in which macroscopic systems are usually embedded and how the nominally isolated systems of experimental macrophysics interact with these environments, we may be able to do three things: reconcile Gibbs's theorem on the constancy of entropy's statistical counterpart with the law of entropy non-decrease; reconcile Gibbs's theorem with Boltzmann's *H* theorem and with the postulate of molecular chaos, which underpins both Maxwell's and Boltzmann's gas theories; and counter Loschmidt's objection to the *H* theorem.

Physicists recognized early on that they could gain insight into the mathematical regularities that lie behind appearance by minimizing interactions between the objects they studied in their laboratories and the outside world, and by ignoring these interactions in their theories. Galileo's experimental and theoretical studies of the (almost) frictionless motion of bodies sliding down inclined planes illustrate this strategy. But whereas macroscopic systems can be effectively insulated against exchanges of heat, mechanical energy, and momentum with their environments, it's much harder to insulate them against exchanges of *information* (in the technical sense defined earlier) with their environments. Suppose now that nominally undisturbed macroscopic systems are normally embedded in highly random environments, and consider an isolated system composed of a nominally undisturbed macroscopic system S and an extensive (but bounded) random environment E – a system characterized by a probability distribution of microstates whose randomness has its largest possible value. By Gibbs's theorem, the randomness of the combined system S + E is constant in time. But, as Boltzmann's proof of his *H* theorem implies, collisions between molecules of S and molecules of E at the interface between S and E create molecular correlations. Information associated with these correlations flows from S to E. Hence the randomness of S increases, while the randomness of E decreases by an equal amount. The transfer of correlation information from S to its environment preserves the mutual statistical independence of the single-particle probability distributions and causes the randomness of the *N*-particle distribution to increase steadily with time – a stronger – as well as more widely applicable –statement than that provided by the law of entropy non-decrease.

But what justifies the assumption that (actual) macroscopic systems are embedded in random environments? The assumption of primordial randomness implies that

> A complete description of a macroscopic system's initial state contains just the information created by the system's history, which includes an account of how the initial state was prepared.

This supplies a framework – but only a framework – for addressing what the physicist Nico Van Kampen[13] has called "the main problem in the statistical mechanics of irreversible processes:" What determines the choice of macrostates and macroscopic variables?

Van Kampen has also emphasized the role of "repeated randomness assumptions" in theories of stochastic processes:

> This *repeated randomness assumption* is drastic but indispensable whenever one tries to make a connection between the microscopic world and the macroscopic or mesoscopic levels. It appears under the aliases "*Stosszahlansatz*," "molecular chaos," or "random phase approximation," and it is responsible for the appearance of irreversibility. Many attempts have been made to eliminate this assumption, usually amounting to hiding it under the rug of mathematical formalism.[14]

To derive his transport equation and his *H* theorem, Boltzmann had to assume (following Maxwell) that the initial velocities of colliding molecules in a closed gas sample are uncorrelated. This assumption can't hold for non-equilibrium states, because when the information that characterizes a non-equilibrium state decays, residual information associated with molecular correlations comes into being at the same rate. Indeed Poincaré proved that if you wait long enough, the (classical) microstate of an undisturbed system whose particles move and interact in ways governed by Newton's laws of motion eventually returns to a state that approximates the system's initial state arbitrarily closely. However, because a gas sample's reservoir of microscopic information is huge, molecular chaos may

13 N. G. van Kampen in *Fundamental Problems in Statistical Mechanics, Proceedings of the NUFFIC International Summer Course in Science at Nijenrode Castle*, The Netherlands, August, 1961, compiled by E. G. D. Cohen (Amsterdam, North-Holland, 1962), especially pp. 182-184.

14 N.G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3d edition (Amsterdam, Elsevier, 2007), p. 58.

prevail to a good approximation for periods much shorter than the Poincaré recurrence time, which is typically much greater than the age of the universe.

Another approach to the problem of justifying repeated randomness assumptions starts from the remark that no gas sample is an island unto itself. Almost sixty years ago, J. M. Blatt[15] argued that the fact that actual gas samples interact with their surroundings justifies the assumption that correlation information is permanently absent in a nominally closed gas sample. The walls that enclose a gas sample are not perfectly reflecting. When a gas molecule collides with a wall, its direction and its velocity acquire tiny random contributions. These leave the single-particle probability distribution virtually unaltered, but they alter the histories of individual molecules, thereby disrupting multi-particle correlations. Blatt distinguished between states of true equilibrium, characterized (in the vocabulary of the present essay) by an information-free probability distribution, and quasi-equilibrium states, in which single-particle information is absent but correlation (or residual) information is present. With the help of a simple thought experiment, he argued that collisions between molecules of a rarefied gas sample and the walls of its container cause an initial quasi-equilibrium state to relax into true equilibrium long before the gas has come into thermal equilibrium with the walls. Earlier, P.J. Bergmann and J.L. Lebowitz[16] constructed and investigated detailed mathematical models of the relaxation from quasi-equilibrium to true equilibrium through external "intervention." More recently, T.M. Ridderbos and L.G. Redhead[17] have expanded Blatt's case for the interventionist approach. They constructed a simple model of a famous experiment, the spin-echo experiment[18], in which a macroscopically disordered state evolves into a macroscopically ordered state. They argued that in this experiment (and more generally) interaction between a nominally isolated macroscopic system and its environment mediates the loss of correlation information.

15 J. M. Blatt, *Prog. Theor. Phys.* 22, 745 (1959)

16 P. J. Bergmann and J. L. Lebowitz, *Phys. Rev.* 99, 578 (1955); J. L. Lebowitz and P. J. Bergmann, Annals of Physics 1, 1 (1959)

17 T. M. Ridderbos and M. L. G. Redhead, *Found. Phys.* 28, 1237 (1988)

18 E. L. Hahn, *Phys. Rev.* 80, 580 (1950)

Blatt noted that this "interventionist" approach "has not found general acceptance."

> There is a common feeling that it should not be necessary to introduce the wall of the system in so explicit a fashion. ... Furthermore, it is considered unacceptable philosophically, and somewhat "unsporting," to introduce an explicit source of randomness and stochastic behavior directly into the basic equations. Statistical mechanics is felt to be a part of mechanics, and as such one should be able to start from purely causal behavior.[19]

Derivations of Boltzmann's *H* theorem and its generalization that rely on decoherence[20] exemplify and significantly extend the interventionist approach. These derivations assume that macroscopic systems are initially in definite quantum states but interact with random environments. Under these conditions, environmental interactions transfer information very effectively from the system to its surroundings. Microscopic information that is wicked away from the system disperses outward, eventually getting lost in interstellar and intergalactic space.

Because the microstates of undisturbed systems evolve reversibly, a theory that assigns macroscopic systems (or the universe) definite quantum states cannot provide a framework for theories that distinguish in an absolute sense between the two directions of time. The historical approach sketched in this book offers such a framework because by linking the initial states of macroscopic systems to states of the early universe it supplies a cosmological context for theories of irreversible processes. Because the probability distributions that characterize the initial states of macroscopic systems depend on their histories, there can be no genuine laws about initial conditions, only historical generalizations. For example, macroscopic systems cannot usually be prepared in definite quantum states; but physicists have succeeded in preparing superconducting quantum interference devices (SQUIDs) in superpositions of macroscopically distinct quantum states. Again, macroscopic systems are usually microscopically disordered; but Hahn's spin-echo experiment showed that this is not necessarily the case. A historical narrative based on the strong cosmological principle and the assumption of primordial random-

19 J. M. Blatt, *Prog. Theor. Phys.* 22, 745 (1959), p. 747

20 M. Schlosshauer, *Decoherence and the Quantum-to-Classical Transition*, corrected 2d printing (Berlin, Springer, 2008)

ness predicts that macroscopic systems are *usually* embedded in random environments and that their initial states *usually* contain only information associated with the values of macroscopic variables. However, as the spin-echo experiment shows, macroscopic systems *can* be prepared in states that do contain microscopic information.

VI
A Unified Picture of the Physical World

The Measurement Problem

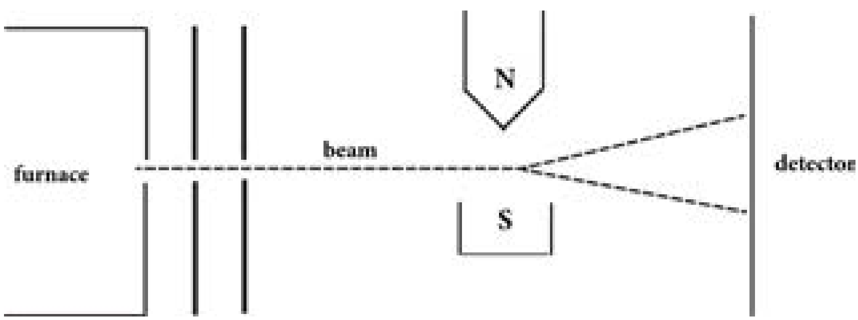


Figure 2 . The Stern-Gerlach experiment

Some atoms – silver, for example – have magnetic moments: in a magnetic field they behave as if they were tiny bar magnets. In 1922, three years before the birth of quantum mechanics and two years before Samuel Goudsmit and George Uhlenbeck proposed that electrons have an intrinsic spin and an intrinsic magnetic moment, physicists Otto Stern and Walter Gerlach devised a now classic experiment to measure the strength of a silver atom's magnetic moment. (Because silver atoms have a single valence electron, they have a magnetic moment equal to the electron's magnetic moment.) Their experimental setup included a vertical magnetic field whose field-lines diverged from top to bottom. When a silver atom traveling in a horizontal direction enters such a field, classical electromagnetic theory predicts that it is deflected upward or downward by an amount proportional to the cosine of the angle its magnetic moment makes with the field. In the Stern-Gerlach experiment (Figure 2) the atom, after emerging from the diverging magnetic field, strikes a detector, where it leaves a permanent trace. The height of this trace above or below the point at which the atom would have struck the detector if it hadn't been deflected by the magnetic field is proportional to the vertical component of the atom's magnetic moment. So if we knew the initial orientation of the atom's magnetic moment we could infer its magnitude from the height of the trace. This is the quantity Stern and Gerlach set out to measure.

The experiment was done not with a single silver atom but with a narrow beam of silver atoms. The atoms had nearly the same horizontal velocities but randomly oriented magnetic moments. So Stern and Gerlach expected that the heights of the traces produced by individual atoms would be distributed smoothly along a vertical strip, whose length would be a calculable multiple of a silver atom's magnetic moment. To their surprise the traces of individual atoms clustered around the points that corresponded to the two extreme orientations of the magnetic moment, as if half of the magnetic moments had been initially parallel to the field, half antiparallel. Yet the conditions of the experiment ensured that before the atoms entered the field their magnetic moments were randomly oriented. What was going on?

Quantum mechanics eventually answered the question. Its laws predict that under the conditions of the Stern-Gerlach experiment the magnetic moment of a silver atom must exhibit one of two equal values, parallel or antiparallel to the magnetic field. But while this prediction, along with countless other similar predictions, agrees with experiment, physicists still – after more than 90 years – don't agree about how to interpret the theory on which it rests. They disagree about how that theory relates to classical physics and the world of experience. That disagreement in turn reflects, in part, a diversity of opinion about the nature of chance and its role in physical theories – questions that also underlie the issue of free will versus determinism.

The Classical World and the Quantum World

Classical physics both refines and extends commonsense notions about the physical world. The world of classical physics is made up of objects with measurable properties, usually called "dynamical variables" or just "physical quantities." Physical quantities have definite values, which represent the outcomes of ideal error-free measurements and which define the physical state of the system they refer to. For example, at each moment a classical particle has a definite position, specified by its three position coordinates in some coordinate system, and a definite velocity, specified by its three velocity components in that coordinate system; and these six quantities define its physical state, which varies smoothly with time in a way governed by Newton's laws of motion and appropriate force laws, including Newton's law of gravitation.

Some of the physical quantities of classical physics have counterparts in quantum theory, where they go by the same names. Among these properties are position, velocity, energy, linear momentum, and angular momentum. Many relations between classical quantities also have quantum counterparts. For example, the formulas that express a particle's momentum, kinetic energy, potential energy and angular momentum in terms of its position and velocity have the same form in classical physics and quantum physics.

But the mathematical objects that represent the quantum counterparts of classical quantities don't always have definite numerical values when the system they refer to is in a definite state. For example, while the electron in a hydrogen atom can be in a state of definite energy, neither the quantum counterparts of its position coordinates nor the quantum counterparts of its momentum components have definite values in such a state.

Because quantum dynamical variables can't in general be assigned definite numerical values, the outcome of an ideal error-free measurement of a quantum dynamical variable that refers to a quantum system in a definite state isn't predictable in general. It does, though, have a predictable set of *possible values*. These are called the physical quantity's *eigenvalues*. For example, the Stern-Gerlach experiment shows that the magnetic moment of a silver atom has just two eigenvalues, $+h/4\pi$ and $-h/4\pi$, where h is the constant introduced in 1900 by Max Planck in his semi-empirical formula for the frequency distribution of thermal radiation.) Each eigenvalue of a physical quantity is associated with one or more *eigenstates*.

> An ideal measurement of a physical quantity has a predictable outcome just in case the system to which the quantity refers is in one of the quantity's eigenstates. The measurement then yields the corresponding eigenvalue.

This statement is the most basic of the rules that link the mathematical formalism of quantum mechanics to the mathematical formalisms of classical theories and to experience.

Different physical quantities that refer to the same system needn't, and in general don't, have the same eigenstates. For example, the energy eigenstates of the electron in a hydrogen atom are not eigenstates of either the electron's position coordinates or its velocity components. Even if an electron is in a state of definite energy it doesn't have a definite orbit.

An undisturbed, or isolated, system needn't be in an eigenstate of *any* of the system's physical quantities, including its energy. But every state of the system can be expressed as a *superposition* of the eigenstates of any physical quantity that pertains to the system. That is, the wavefunction (or state vector) that represents a given state can be expressed as a linear combination – a sum of multiples – of wavefunctions (or state vectors) that represent eigenstates of any given physical quantity that pertains to the system. The coefficients in such linear combinations are complex numbers, numbers of the form $a + bi$, where a and b are real numbers and i denotes the square root of -1. For example, we can express the wavefunction that represents an arbitrary state of the electron in a hydrogen atom as a sum of complex multiples of wavefunctions that represent states of definite energy.

The mathematical language of quantum theory expresses these highly non-intuitive (and philosophically opaque) properties of physical quantities and physical states of molecules, atoms, subatomic particles, and other quantum systems precisely and unambiguously. These are the language's main features:

> *Quantum states*. Every isolated, or undisturbed, physical system has its own state space. The mathematical objects

that inhabit state space are called wavefunctions or, more generally, state vectors. Each state vector represents a possible state of the system. For example, an electron has two possible spin states, represented by a pair of state vectors. If a vertical magnetic field is present, an experiment designed to exhibit an electron's magnetic moment, which is proportional to its spin, will show either that the magnetic moment and the spin point up or that they point down. Spin-up and spin-down are the only two possible outcomes of such an experiment, and in a series of identical experiments they occur equally frequently.

State vectors combine with one another and with numbers much like vectors in Euclidean space.

Vectors and their algebra. Let **u** denote a vector in Euclidean space. It is defined by its direction and length. We can picture it as a free-floating arrow, untethered to any particular point. If c is a positive real number, $c\mathbf{u}$ is also a vector, represented by an arrow whose length is c times the length of the arrow that represents **u**. If c is a negative real number, $c\mathbf{u}$ is the vector whose direction is opposite to the direction of u and whose length is $|c|$ times the length of **u**, where $|c|$ denotes the absolute value of c. If $\mathbf{u}_1$ and $\mathbf{u}_2$ are any two vectors, their sum $\mathbf{u}_1 + \mathbf{u}_2$ is represented by an arrow you can construct by joining the arrows that represent $\mathbf{u}_1$ and $\mathbf{u}_2$ end to tip in either order (while preserving their lengths and directions), then joining the free end to the free tip.

State vectors in an undisturbed physical system's state space have exactly the same algebraic properties, except that the constant c in the preceding account is a *complex* number. Thus if ψ_1 and ψ_2 are state vectors in the state space of a given system and a and b now denote any complex numbers, then $a\psi_1 + b\psi_2$ is also a state vector in the same space, representing one of the system's possible states. The state vector $a\psi_1 + b\psi_2$ represents a *superposition* of the states represented by ψ_1 and ψ_2.

Superpositions of quantum states have no counterpart in the language of classical physics (or in ordinary language). They are responsible for the most counterintuitive features of the quantum world. For example, the cat in Erwin Schrödinger's famous thought experiment is in a superposition of the states "dead" and "alive" until its interaction with the world of experience puts it into one state or the other, just as in the Stern-Gerlach experiment a silver atom's interaction with the measuring apparatus puts into a definite classical state, "up" or "down."

> *Observables*. The language of quantum physics represents physical quantities by mathematical objects called *observables*, a subclass of a broader class of objects called *operators*. Like the variables that represent physical quantities in Newtonian physics, observables (as well as other operators) can be added to one another, multiplied by complex numbers, and multiplied by one another. And relations between Newtonian variables have counterpart in relations between the corresponding observables. *But multiplication of operators (including observables) is non-commutative*: if O_1 and O_2 are operators, O_1O_2 need not equal O_2O_1. In particular, the observables that represent a particle's position coordinates don't commute with the operators that represent the corresponding momentum components. Heisenberg's indeterminacy relations are a consequence of the fact that "conjugate" physical quantities, like a position coordinate and the corresponding momentum component or like a particle's energy and its time coordinate, don't commute.

Physics texts and papers in physics journals often distinguish the quantum counterparts of classical variables by adding a circumflex ˆ or a suffix, like $_{op}$, to the symbol that represents the corresponding classical variable, if there is one, writing $\hat{E}$ or E_{op} for the quantum counterpart of energy E. But physicists customarily use the same words, such as position, momentum, and energy, for classical variables and their quantum counterparts, relying on context to distinguish between them. In nontechnical contexts the failure to distinguish between classical variables and their quantum counterparts can lead to confusion. Consider the Stern-Gerlach experiment. Stern and

Gerlach thought of silver atoms as classical objects. They thought they were measuring the vertical position coordinates of silver atoms in the beam by recording the positions at which the atoms struck a detector. In 1923, when Stern and Gerlach carried out their experiment, the phrases "the vertical position coordinate of a silver atom" and "the height at which an atom strikes the detector" referred to the same physical quantity, the *z* coordinate of a silver particle. But once quantum mechanics had been formulated as a mathematical theory, this was no longer true. "The vertical position coordinate of a silver atom" became a quantum observable, represented by an operator, while "the height at which an atom strikes the detector" remained a classical quantity. What Stern and Gerlach measured, and what the classical theory of the motion of a tiny bar magnet in an inhomogeneous magnetic field (incorrectly) predicted, were values of a classical position coordinate. What quantum theory predicts, through a rule that links the quantum world to the world of experience, is the average value of the *classical counterpart* of a quantum observable.

Before they strike the detector, each silver atom is in a superposition of two distinct position states, corresponding to the fact that a silver atom's magnetic moment has just two eigenstates. But the atom isn't in two places at the same time. Nor is it strictly accurate to say that the detector records the atom's position. Before the atom interacts with the detector its position is represented by an operator and doesn't have a definite value. Afterwards the same atom does have a definite position, which coincides with a spot on a photographic plate.

More generally, in a quantum measurement a system in a definite quantum state interacts with a macroscopic measuring device, such as a photographic plate or a particle detector, that is in a definite classical state. The interaction causes the combined system, consisting of the measured system plus the measuring device, to evolve rapidly from its initial state, in which the measured system is in a definite quantum state and the measuring device is in a definite classical state, to a final joint state, in which the measured system is in an eigenstate of the measured quantity and the measuring device is in a corresponding "pointer state." The rapid evolution of the combined system during a quantum measurement is sometimes called a "jump."

Here's how Dirac describes the measurement process in the fourth edition of *The Principles of Quantum Mechanics* (1967; first edition 1930):

> When we measure a real dynamical variable ξ, the disturbance involved in the act of measurement causes a jump in the state of the dynamical system. From physical continuity, if we make a second measurement of the same dynamical variable ξ immediately after the first, the result of the second measurement must be the same as that of the first. Thus after the first measurement has been made, there is no indeterminacy in the result of the second. Hence, after the first measurement has been made, the system is in an eigenstate of the dynamical variable ξ, the eigenvalue it belongs to being equal to the result of the first measurement. This conclusion must still hold if the second measurement is not actually made. In this way we see that *a measurement always causes the system to jump into an eigenstate of the dynamical variable that is being measured, the eigenvalue this eigenstate belongs to being equal to the result of the measurement.*1 (italics added)

Dirac doesn't describe the rapid evolution of the combined system in a quantum measurement. He fills this gap by what he calls a "general assumption." Consider an observable *O* and a physical state *s* of the system of which *O* is a property. Dirac's general assumption *equates the average outcome of a large number of identical ideal measurements of O when the system O pertains to is in state s to a real number constructed according to a simple rule from the operator O and the state vector s.*

Dirac's general assumption contains – and extends in the simplest possible way – our earlier basic rule: if a system is in an eigenstate of the observable *O*, an ideal measurement of *O* yields the corresponding eigenvalue of *O*. It implies further that if the system isn't in an eigenstate of *O*, the outcome of an ideal measurement is one of *O*'s eigenvalues. Finally, it yields a formula for the relative frequencies of the possible measurement outcomes in an infinite series of identical measurements of *O* when the system is in state s.

Dirac's general assumption links the mathematical formalism of quantum mechanics to measurement outcomes efficiently and unambiguously. It has enabled physicists to carry out experimental

1 Dirac, P.A.M., *The Principles of Quantum Mechanics*, fourth edition (revised), (Clarendon Press, Oxford, 1967,)p. 46

tests of quantum theories of molecules, atoms, and subatomic particles and of macroscopic objects such as crystals, conductors, and semi-conductors – objects whose observable properties depend critically on their quantum microstructure. Up until now all such predictions have been confirmed to the accuracy of the models and calculations on which they rest.

What is missing from Dirac's account of measurement is a description of the measurement process itself – the rapid evolution, during a measurement, of the joint state of the combined system. Can quantum theory supply such a description? This is the measurement problem. In a recent popular article, "The Trouble With Quantum Mechanics," the physicist Steven Weinberg succinctly defines the core of the problem:

> If we regard the whole process of measurement as being governed by the equations of quantum mechanics, and these equations are perfectly deterministic, *how do probabilities get into quantum mechanics?*[2]

"The Quantum-to-Classical Transition"

Introductory textbooks on quantum mechanics usually have a section entitled "The Hydrogen Atom." The hydrogen atom the textbooks describe is a fictitious object because, unlike actual hydrogen atoms, it doesn't interact with the outside world. Yet the quantum theory of an undisturbed hydrogen atom accurately predicts the measurable properties of real hydrogen atoms, such as the frequencies at which a hot sample of atomic hydrogen gas emits light and a cool sample absorbs light.

In contrast, an account of quantum measurement that neglects the interaction between the measuring device and its environment fails spectacularly. For, as Weinberg emphasized, an undisturbed quantum system, like an undisturbed classical system, evolves deterministically; but the measurement postulate, Dirac's "general assumption," predicts (and experiments confirm) that in general a quantum measurement has more than one possible outcome. Thus we can neglect the interaction between a physical system and its environment, at least in a first approximation, in some circumstances but not in others. What, precisely, are these circumstances?

An undisturbed quantum system can be in a state that doesn't change with time and in which its energy, like the energy of an undisturbed classical system, has a definite value. But unlike an undisturbed classical system, whose energy has a continuous range of possible values, an undisturbed quantum system that occupies a bounded region of physical space has a discrete set of possible energies and a corresponding discrete set of possible quantum states. In 1970 physicist Hans Dieter Zeh pointed out that *we can assign the system a definite quantum state if and only if the difference between the energy of that state and the energies of neighboring states greatly exceeds the energy associated with the system's interaction with its surroundings.* If this condition isn't fulfilled, transitions between neighboring energy levels occur so rapidly that the system can no longer be assigned a state of definite energy.

Atoms and molecules in dilute gases satisfy this condition because the intervals between their low-lying energy levels typically exceed the energies associated with the interaction between an atom or molecule and its neighbors. In contrast, as I mentioned earlier, macroscopic systems have very closely spaced energy levels. Under ordinary experimental conditions the intervals between adjacent energy levels of a macroscopic object are *much* smaller than the energy associated with the object's interaction with its environment. So even if the laws of quantum mechanics apply to isolated, or undisturbed, macroscopic objects, as experimental evidence strongly suggests they do, these objects cannot, under ordinary conditions, be assigned definite quantum states.

This remark is the starting point of *decoherence theories* of quantum measurement. These theories seek to refine and improve an account of quantum measurement proposed in 1932 by the mathematician John von Neumann. Von Neumann set out to replace Dirac's general assumption by an account of the measurement process itself. He idealized the combined system in a quantum measurement – the measured system plus the measuring apparatus – as an undisturbed system in a definite quantum state. He also assumed that if the measured system is initially in an eigenstate of the measured observable, the measurement leaves it in that state and puts the measuring apparatus into a corresponding *quantum* state that is *macroscopically* distinguishable from the quantum states associated with other possible measurement outcomes. If the measured system's initial, or pre-measurement, state is not an eigenstate of the measured observable, the rules of quantum mechanics allow us to represent it as a superposition of eigenstates of the measured observable – a sum of complex multiples of the state vectors that represent the measured observable's eigenstates. Schrödinger's equation then predicts that the combined system evolves into the same superposition of combined-system states. Each state of the combined system in this superposition is a possible outcome state, in which the measured system is in one of the eigenstates of the measured observable and the measuring apparatus is in a correlated quantum state. Moreover, the squared magnitude of each coefficient in the superposition is the probability of the corresponding measurement outcome, as predicted by Dirac's general assumption.

But Dirac's general assumption implies, and experiment confirms, that a quantum measurement leaves the measured system in *one or another* of the measured quantity's eigenstates, not in a superposition of all of them.

There's a more subtle difficulty with von Neumann's account of a quantum measurement as I've described it so far. Each coefficient in the superposition that represents the measured quantum system's initial state is a complex number; it has a magnitude and a phase.[3] A measurement governed by Dirac's general assumption preserves information about the magnitudes of the coefficients in the superposition that represents the measured quantum system's initial state: their squares are the probabilities of the outcomes represented by the state vectors they multiply.

But Dirac's account of a quantum measurement implies that it destroys information about the relative phases of the coefficients. In von Neumann's account, a quantum measurement preserves information about both the magnitudes *and* the relative phases of the coefficients in the superposition that represents the evolved joint state of the measured system and the measuring apparatus.

3 Consider a complex number $a + bi$, where a and b are real numbers and i is the square root of −1. We can think of a and b as the rectangular coordinates of a point in a plane, the complex-number plane. The angle between the x-axis and the line joining the origin to the point that represents a complex number is called the *phase* of that number. Two complex numbers have the same phase if and only if the line joining their representative points passes through the origin of the complex-number plane.

There's also a more subtle difficulty with von Neumann's account: It contradicts the fact, stressed by Niels Bohr, that the measuring apparatus in a quantum measurement is a macroscopic system in a definite *classical* state. Von Neumann's account of a quantum measurement assumes that the measuring apparatus is initially in a definite quantum state. Decoherence theories emend von Neumann's account not by addressing this difficulty but by postulating that the measuring apparatus interacts with a random environment (in calculations, either thermal radiation or a tenuous gas). Instead of assuming, as von Neumann did, that the combined system is undisturbed (and hence in a definite quantum state), decoherence theories assume that a supersystem consisting of the combined system plus a bounded portion of the postulated random environment can be treated as an undisturbed system in a definite quantum state. A quantum measurement then transfers information about the relative phases of the state vectors of the combined system (measured system plus measuring device) to the environment. Model calculations by Wojciech H. Zurek, Erich Joos and Hans Dieter Zeh, by Maximilian Schlosshauer, and by other authors, have shown that interaction between the measuring apparatus and the random environment rapidly decoheres macroscopically separated states of the combined system, randomizing the relative phases of the coefficients in the superposition of combined-system state vectors predicted by von Neumann's account:

> Formally, decoherence can be viewed as a dynamical filter on the space of quantum states, singling out those states that, for a given system, can be stably prepared and maintained, while effectively excluding most other states, in particular, nonclassical superposition states of the kind popularized by Schrödinger's cat.[4]

By predicting the loss of coherence between different possible outcomes of a quantum measurement, decoherence theory improves von Neumann's account. But as the authors mentioned above have stressed, it doesn't predict that quantum measurements have definite but unpredictable outcomes. And, as I've already emphasized, it ignores Bohr's requirement that the measuring apparatus is in a definite classical state before and after a quantum measurement.

4 Schlosshauer, Maximilian, "The quantum-to-classical transition and decoherence," arXiv:1404.2635v1 [quant.phys] 9 April 2014

Von Neumann's Collapse Hypothesis and Attempts to Avoid It

Von Neumann completed his account of quantum measurement by postulating that no sooner does the superposition of outcome states produced by interaction between the measured system and the measuring apparatus come into being than it "collapses" randomly onto one of these outcome states, the probability of each possible outcome being equal to the squared magnitude of the corresponding coefficient in the superposition. With this additional assumption, von Neumann's account reproduces – but doesn't explain – Dirac's general assumption. It replaces Dirac's "jump" of the joint state of the combined system by a "collapse" of the combined system's wavefunction. Von Neumann assumed that this collapse is a real physical process. But quantum mechanics doesn't describe it. Does quantum mechanics need to be modified or extended to include it?

Von Neumann suggested that the experimenter's awareness of the outcome of a quantum measurement causes the superposition of outcome states to collapse. Other prominent physicists, including Eugene Wigner, embraced this view. Wigner interpreted a system's wavefunction as representing the physicist's knowledge about the system: "Given any object, all the possible knowledge concerning that object can be given as its wave function."[5]

> [T]he impression which one gains at an interaction [with a quantum system] may, and in general does, modify the probabilities with which one gains the various possible impressions at later interactions. In other words, the impression which one gains at an interaction, called also the result of an observation, modifies the wave function of the system. The modified wave function is, furthermore, in general unpredictable before the impression gained at the interaction has entered our consciousness: it is the entering of an impression into our consciousness which alters the wave function because it modifies our appraisal of the probabilities for different impressions which we expect to receive in the future.[6]

In an essay entitled "The Copenhagen Interpretation of Quantum Theory" Werner Heisenberg (1958) argued that the indeterminacy of quantum measurement outcomes arises from "our incomplete knowledge of the world." Just before a measurement begins, the

5 Wigner later changed his mind.

6 Wigner, Eugene, "Remarks on the Mind-Body Question" in *Symmetries and Reflections*, (Cambridge, MA. MIT Press, 1970) p. 173

measured object is in a definite quantum state. Then, in Heisenberg's account, it interacts with a macroscopic measuring device.

> This influence introduces a new element of uncertainty, since the measuring device is necessarily described in the terms of classical physics; such a description contains all the uncertainties concerning the microscopic structure of the device which we know from [statistical] thermodynamics, and since the device is connected with the rest of the world, it contains in fact the uncertainties of the microscopic structure of the whole world. These uncertainties may be called objective in so far as they are simply a consequence of the description in the terms of classical physics and do not depend on any observer. They may be called subjective in so far as they refer to our incomplete knowledge of the world. … It is for this reason that the result of the [measurement] cannot generally be predicted with certainty; what can be predicted is the probability of a certain result …[7]

Physicists Christopher Fuchs and Rüdiger Schack have proposed a related interpretation of quantum theory, which they call Quantum Bayesianism or QBism. (Bayesian statistics interprets probabilities as degrees of belief. Its methods enable one to update his or her degree of belief in a hypothesis in the light of new evidence.) It rests in part on an earlier paper they co-authored with Carlton Caves. Physicist N. David Mermin, a prominent supporter of this interpretation, has written:

> QBism attributes the muddle at the foundations of quantum mechanics to our unacknowledged removal of the scientist from the science.[8]

Much of this muddle is associated with the 'wavefunction' that quantum mechanics assigns to a physical system. This irritatingly uninformative term reveals the lack of clarity present in the field from its very beginning in 1925. People argue to this day about whether wavefunctions are real entities, like stones or ripples on a pond, or mathematical abstractions that help us to organize our thinking, like the calculus of probabilities.

Fuchs and Schack adopt the latter view. They take a wavefunction to be associated with a physical system by an agent – me, for example,

7 Heisenberg, W. *Physics and Philosophy: The Revolution in Modern Science*, (London: George Allen & Unwin, 1958.)

8 Mermin, N. David, *Nature*, 26 March 2014, "QBism puts the scientist back into science"

based on my past experience. I use the wavefunction, following rules laid down by quantum mechanics, to calculate the likelihood of what I might experience next, should I choose to probe further. Depending on what I then perceive, I can update the wavefunction on the basis of that experience, allowing me to better assess my subsequent expectations.[9]

At the opposite pole from QBism are emended versions of Schrödinger's equation that describe the "collapse of the wavefunction" as a real physical process. G.C. Ghirardi, A. Rimini, and T. Weber (1986) have proposed one such theory, and Philip Pearle (1989) has described a modified version of it. Both versions add to Schrödinger's equation a stochastic term that causes a superposition of quantum states to collapse onto one of its components. The stochastic term represents an interaction between a nominally undisturbed system and a purpose-built and otherwise unobserved space-filling stochastic field, reminiscent of the ether of pre-relativity physics. These theories contain adjustable parameters whose values are chosen to ensure that superpositions representing possible states of atoms and molecules do not collapse within observable periods of time whereas superpositions representing impossible states – superpositions whose components are separated by macroscopic distances – do not persist long enough to be detected experimentally.

Still other attempts to explain why measurements have definite but unpredictable outcomes take as their starting point the remark that under typical experimental conditions *no* bounded system (including the supersystems considered in decoherence theories) can be idealized as an undisturbed system in a definite quantum state. In 1957 Hugh Everett III, then a doctoral student of John A. Wheeler, suggested a radical answer to the question: Why do quantum measurements have definite but unpredictable outcomes? He postulated that quantum mechanics is universally valid: it applies to arbitrarily large undisturbed physical systems. But only the universe is an undisturbed physical system. So, Everett reasoned, we should identify the supersystem consisting of the measured system, the measuring device, and the environment with the universe. A quantum measurement then causes the premeasurement state of the universe to evolve into a superposition of universe-states in each of which the

9 Caves, C. M., Fuchs, C. A., & Schack, R. (2002). "Quantum probabilities as Bayesian probabilities." *Physical Review* A, 65(2), 022305.

measured system is in an eigenstate of the measured observable and the rest of the universe is in a correlated quantum state. Everett now postulated that the universe-states that occur in this superposition represent states of distinct but equally real universes in each of which a particular measurement outcome has occurred. A measurement causes the universe in which it occurred to split into many – sometimes infinitely many – replicas, differing only in the outcome of the measurement that caused the splitting.

There is an extensive literature about what this account might (or could) mean, and how (or whether) it can be reconciled with the rest of physics, including general relativity and the physics of the everyday world of medium-sized objects. As far as I know, these issues haven't been resolved.

The theories of quantum measurement I've been describing assume either that some bounded part of the universe can be idealized as being in a definite quantum state or that the universe itself is in a definite quantum state. None of these theories successfully predicts that an ideal quantum measurement leaves the measuring apparatus in a definite pointer state and the measured system in a corresponding quantum state, an eigenstate of the measured physical quantity. They leave Weinberg's question "*How do probabilities get into quantum mechanics?*" unanswered. My suggested answer to this question also answers a question asked in 1908 by the mathematician and theoretical physicist Henri Poincaré in his essay "Chance": How do probabilities get into classical mechanics?

Poincaré's "Chance"

Poincaré begins his essay by contrasting the "ancients'" view of chance with the "modern" view:

> To begin with, what is chance? The ancients distinguished between the phenomena which seemed to obey harmonious laws, established once for all, and those that they attributed to chance, which were those that could not be predicted because they were not subject to any law. In each domain the precise laws did not decide everything, they only marked the limits within which chance was allowed to move. In this conception, the word chance had a precise, objective meaning; what was chance for one was also chance for the other and even for the gods.

> But this conception is not ours. We have become complete determinists, and even those who wish to reserve the right of human free will at least allow determinism to reign undisputed in the inorganic world. Every phenomenon, however trifling it be, has a cause, and a mind infinitely powerful and informed concerning the laws of nature could have foreseen it from the beginning of the ages. If a being with such a mind existed, we could play no game of chance with him; we should always lose.
>
> For him, in fact, the word chance would have no meaning, or rather there would be no such thing as chance. That there is for us is only on account of our frailty and our ignorance. And even without going beyond our frail humanity, what is chance for the ignorant is no longer chance for the learned. Chance is only the measure of our ignorance. Fortuitous phenomena are, by definition, those whose laws we are ignorant of.[10]

No sooner does Poincaré arrive at this answer to his opening question than he explains why it won't do. Chance, he argues, must be "something more than the name we give to our ignorance," because some chance phenomena, while individually unpredictable, obey mathematical laws. "Laws of chance" relate the probabilities of measurement outcomes and other macroscopic phenomena to the frequencies of these outcomes in large (or infinite) collections of identical trials. For example, probability theory allows us to calculate the probability of k heads in N tosses of a coin, for $k = 0, 1, 2, ..., N$, if we are given the probability p of heads for a single toss. It predicts that the probability of pN heads approaches 1 as N increases without limit.

Poincaré now asks, "What are the defining properties of chance processes whose outcomes are governed by probabilistic laws?" He considers three examples of such processes.

The first is an idealized cone balanced on its tip. If no force other than gravity acts on it, the cone remains in this state forever, but any disturbance, however slight, causes it to topple over. If we try to balance a real cone on its tip we will fail. The cone will topple in an unpredictable direction. But if we repeat the experiment many times, using exactly the same experimental arrangement to fix the cone's initial position and motion, the direction in which it tips will be more or less uniformly distributed between 0 and 2π radians (0 and 360 degrees). If the cone tended to topple in a particular direction,

10 Poincsré, H. *Science and Method*, (New York, Dover Publications, 2003) p.64-5

we would look for the cause of this asymmetry in a corresponding asymmetry of the cone itself or of the experimental arrangement that determines its initial state. In this example, as Poincaré put it. "slight differences in the initial conditions produce very great differences in the final phenomena." Roulette, another of Poincaré's examples, exemplifies the same principle. Small uncontrollable differences in the relative speed of the ball and the roulette wheel produce unpredictable differences in the number of red and black compartments it passes over before coming to rest. But probability theory predicts that the fraction of N spins with the outcome red approaches ½ as N increases.

Poincaré discusses other processes that display extreme sensitivity to small changes in their initial conditions. In the 1880s he had discovered that in some regions of the planetary system, a small change in an asteroid's initial position or velocity causes its orbit to diverge at an exponential rate from the original orbit. Such orbits are predictable in principle but not in practice. They are said to be *chaotic*; and the phenomenon they exemplify is called deterministic chaos. Students of chaos theory have found that mathematically analogous processes occur in a wide variety of complex systems, including the Earth's atmosphere and the biosphere.

The processes Poincaré discusses have unpredictable outcomes not because we don't know the laws that govern them – we do – but because, unlike an "infinitely powerful and informed" mind, we don't know enough about their initial conditions. Yet we correctly predict that the cone topples in a random direction, that in the long run the roulette ball comes to rest in a red compartment as often as it does in a black compartment, and that in the planetary system small bodies in chaotic orbits are randomly distributed along the ecliptic (the great circle in which the plane of Earth's orbit intersects the celestial sphere). Why do these statistical predictions succeed?

Consider roulette. If we knew the initial speeds of the ball and the wheel we could predict the color of the compartment in which it came to rest. But the initial speeds have ranges of possible values. Define subranges of these ranges so that initial relative speeds in the same subranges correspond to the same outcome. If the number of these subranges is large enough and if the initial relative speeds are smoothly – not necessarily uniformly – distributed over the whole range of possible initial speeds, then the probabilities of the outcomes red and black will be nearly equal, because any smooth curve has nearly the same height at nearly equal horizontal coordinates. Thus, Poincaré concludes, red and black occur equally frequently in a long series of identical spins of the wheel because two conditions are satisfied: (a) small changes in a relevant initial condition produce large differences in the outcome (i.e., the number of same-outcome subranges in the range of initial conditions is very large); and (b) the distribution of initial conditions is smooth, so that contiguous subranges have (nearly) equal probabilities of being realized. Poincaré's other examples admit a similar analysis.

Condition (a) defines a class of chance processes that have statistically predictable outcomes. Poincaré's remaining task was to explain why, in processes that satisfy this condition, the curve that represents the distribution of the relevant initial conditions is smooth.

Poincaré argued that the distribution of initial conditions of chance processes that have statistically predictable outcomes is the "outcome of a long previous history," during which irregularities on the smallest scales have been smoothed out by "complex causes" working over a long period of time.

Poincaré doesn't appeal to the second law of thermodynamics to justify this claim. (As I argued earlier, such an appeal would be illegitimate because the Second Law doesn't apply to the physical universe.) He does, though, connect the absence of irregularities on very small scales to the temporal asymmetry of macroscopic processes – the fact that macroscopic processes such as heat flow, unlike molecular motions and collisions, are governed by laws that distinguish between the direction of the past and the direction of the future. He discusses in some detail how different the course of events would look to a being traveling backward in time and concludes: "Lengthy explications [of such differences] perhaps would aid in the better comprehension of the irreversibility of the universe." In the end, he leaves the smoothness of the curve that represents the probability distribution of the cone's initial tilt unexplained.

Poincaré's Cone in Light of the Strong Cosmological Principle

Poincaré assumes that the initial positions and velocities of the cone's axis, the roulette ball, and the asteroid have definite values. In other words, these systems are in definite (classical) microstates. Since we don't know the precise values of the physical quantities that characterize classical microstates we assign probabilities to them.

A probability distribution of initial microstates defines a (classical) macrostate. A macrostate is the result of an experimental or observational protocol. An experiment or observation gives us information about the macrostate of the measured or observed system.

Newton's laws of motion connect initial microstates to final microstates. Hence they connect the probability distribution that characterizes a system's initial *macrostate* to the set of its possible final *macrostates*. What Poincaré couldn't explain was why we can successfully represent the probability distributions of microstates that characterize the initial macrostates of the cone, the roulette ball, and the asteroid by smooth curves.

In a universe in which the strong cosmological principle holds, macroscopic systems don't in general have definite initial microstates. Instead a macroscopic system's initial macrostate is characterized by a probability distribution of microstates fashioned by the system's history. The most recent episode of such a history describes the act of measurement or observation, and the probability distribution it creates typically contains particular kinds of information but not others. An experimental arrangement designed to put the cone into its initial state imposes limits on the cone's initial tilt and initial angular velocity but creates little or no additional information. The device that sets the roulette wheel and the ball in motion determines their ranges of initial speeds but nothing more. And plausible scenarios for the origin of the solar system don't specify the initial distribution of the positions and velocities of each of the myriad small bodies it contains in enough detail and with enough precision to produce a non-uniform distribution of their present celestial longitudes. If the strong cosmological principle holds, *the probability distributions that characterize the initial macrostates of the macroscopic systems in Poincaré's examples are represented by smooth curves because in each case the system's history didn't create the kinds of information that would produce deviations from smoothness.*

Instead of imagining an infinite series of identical trials whose initial conditions are defined by a probability distribution of (initial) microstates, we can imagine an *experimental ensemble* – an infinite collection of identical cones uniformly distributed in space, each in a definite microstate. We stipulate that the fraction of cones in every range of microstates equals the fraction of identical trials in which the cone's initial microstate lies in that range. That fraction is the probability assigned to that range by the probability distribution that characterizes the cone's initial macrostate. Such a collection evolves into one in which the fraction of cones in every range of microstates equals the fraction of trial outcomes in which the cone's final microstate lies in that range.

This mode of description extends to experimental systems the indeterminacy in position that the strong cosmological principle attributes to naturally occurring systems. It says that physics doesn't describe particular systems but only uniformly dispersed and isotropically oriented systems defined by their histories.

The strong cosmological principle and the assumption of primordial randomness also explain what Poincaré called "the irreversibility of the universe." In a history of the physical universe based on these initial conditions, randomness is the primordial condition. Subsequently, cosmogonic processes create information associated with the probability distributions that characterize galaxies, stars, planets, and other self-gravitating systems. Some of this information decays in processes governed by the second law of thermodynamics, and thereby fuels the creation of new kinds of information. For example, the burning of hydrogen to helium in the Sun's core creates the sunlight that supports life on Earth. Do these inferences about the strong cosmological principle still hold in a universe whose microstructure is governed by quantum mechanics rather classical physics? Can the strong cosmological principle resolve the problem of time's arrow and the measurement problem of quantum mechanics?

Quantum Indeterminacy and Quantum Measurement

To extend Poincaré's account to macroscopic systems whose microstates are governed by quantum physics we need a way of linking measurement outcomes, which give us information about a measuring device's classical macrostate, to the microstates of the measured system, described in the language of quantum physics. Dirac's general assumption and its generalization[11] supply such a link.

Let *s* denote a microstate of a system *S*, and let *O* denote a property S. Suppose *S* is in the microstate *s*.

> Dirac's general assumption equates a real number, denoted by $\langle s|O|s\rangle$, that depends on the state *s* and the observable *O* to the average result of a large number of measurements of *O* when the system is in state *s*.

Let α denote a macrostate of the macroscopic system *S*. Gibbs's statistical mechanics characterizes such macrostates by probability distributions of the system's microstates. Let s denote a microstate of *S*, and let $p_\alpha(s)$ denote the probability assigned to that microstate by the probability distribution that characterizes the macrostate α. Dirac's general assumption implies that the weighted average of the quantities $\langle s|O|s\rangle$ – the sum of products $\langle p_\alpha(s)\ \ s|O|s\rangle$ – equals the average result of a large number of measurements of *O* when the system is in the macrostate α:

> The average of the quantities $s|O|s$, weighted by the probability that the microstate s is in the macrostate α, equals the average result of a large number of measurements of *O* when the system is in the macrostate α. (1)

In a description of measurement that comports with the strong cosmological principle we interpret the initial conditions that characterize a macroscopic system *S* as characterizing an experimental ensemble of systems, and the preceding rule becomes:

> The weighted average in rule (1) equals the average value of the macroscopic counterpart of *O* in an experimental ensemble of systems in the macrostate α. (2)

We can now describe quantum measurements in a way that differs slightly but significantly from von Neumann's account as emended by decoherence theory. Decoherence also plays an essential role in the new account: it randomizes the relative phases of wavefunctions of microstates of the combined system that belong to dif-

11 Dirac, P.A.M., *The Principles* , pp. 132-133

ferent pointer states. But the new account, unlike von Neumann's account as emended by decoherence theory, doesn't assume that the combined system plus a random section of its environment is in a definite quantum microstate. Instead it assumes that the combined system plus a random section of its environment is in the superposition of microstates that, in the quantum version of Gibbs's statistical mechanics, represents a macrostate of the combined system. A straightforward calculation then shows that the experimental ensemble that represents the premeasurement state of the combined system rapidly evolves into an experimental ensemble of postmeasurement states in each of which the measured system is in an eigenstate of the measured observable and the measuring apparatus is in the corresponding pointer state.

Does this account of quantum measurement explain how probability gets into quantum mechanics? Does it solve the measurement problem?

One can argue that it does *not* explain how probability gets into quantum mechanics, because it rests on the strong cosmological principle, which *posits* that a complete description of the physical world specifies only probabilities of physical states and physical quantities. Posits are not explanations.

Yet the strong cosmological principle doesn't have an *ad hoc* character. As I mentioned earlier, it defines the simplest possible model of the universe that is consistent with observations of the cosmic microwave background and of the spatial distribution and line-of-sight velocities of galaxies. It also accounts for the fact that reference frames in which Newton's and Maxwell's laws as well as their special-relativistic generalizations hold are unaccelerated relative to a frame defined by the cosmic microwave background and the distribution and motions of distant galaxies. Equally importantly, the strong cosmological principle supplies a single, objective interpretation of probability as it occurs in quantum physics, in statistical mechanics, and in cosmology. Does the preceding account of measurement solve the measurement problem? One can argue that it doesn't.

Dirac's general assumption equates the number denoted by $\langle s|O|s\rangle$ to the average result of a large number of measurements of

O. Von Neumann's account of measurement was intended to replace this assumption by a theory of measurement. It failed to do so, even when emended by decoherence theory. The preceding account of measurement unlike the emended von Neumann account, successfully predicts that quantum measurements have definite outcomes. *But it relies on Dirac's general assumption* (which underlies the quantum version of Gibbs's statistical mechanics). Doesn't this mean that it leaves the measurement problem unresolved?

Yes and no. Yes if you assume that there are as-yet undiscovered mathematical laws of which the laws of quantum physics and classical physics (including general relativity) are limiting cases. If such laws existed, Dirac's general assumption wouldn't be an assumption; it would be a theorem. In their present forms, however, quantum physics and classical physics have unbridgeable differences. Quantum superpositions have no classical counterpart; and quantum physics can't describe curved spacetime. At the same time, special relativity is a pillar of quantum mechanics. The last point is worth a more detailed discussion.

Special Relativity and Quantum Mechanics

Einstein based his special theory of relativity (1905) on two assumptions: a) All inertial reference frames – frames in which Newton's laws of motion and Maxwell's laws of electricity and magnetism hold – enjoy equal status. b) The speed of light in empty space has the same value in every direction and in every inertial frame. Minkowski accordingly set the speed of light in empty space equal to 1, so that in his new geometry distance intervals and time intervals are always measured in the same unit. He then defined the squared space-time interval between two point-events in spacetime as the squared time interval between the point-events and the squared distance interval (given by Pythagoras's theorem as the sum of their squared position-coordinate differences); and he postulated that this quantity has the same value in every unaccelerated reference frame. Finally, he stipulated that the algebraic statements that express physical laws should take the same form in all unaccelerated, or inertial, reference frames.

These postulates greatly extended the domains of mechanics and electromagnetism. The laws of pre-relativity physics became limiting cases of relativistic laws, valid only for particles traveling with speeds much less than the speed of light. And experiments invariably confirmed the predictions of theories that comported with the principles of special relativity. Nevertheless, the earliest versions of quantum mechanics, Werner Heisenberg's "matrix mechanics" of 1925 and Erwin Schrödinger's "wave mechanics" of 1926, didn't comport with these principles: like Newton's laws of motion, they assumed that space is Euclidean and that the time interval between any two moments has the same value in all allowed coordinate systems.

Schrödinger's wave equation is the quantum counterpart of Newton's equation of motion for an electron in an external electric field. Schrödinger and other founders of quantum theory tried to formulate a relativistic generalization of the wave equation – an equation that would have the same form in all unaccelerated, or inertial, reference frames. But they ran into a formidable mathematical obstacle. Schrödinger's equation is "linear": any sum of numerical multiples of two (or more) solutions of the equation that satisfy the same boundary conditions) is likewise a solution. There were compelling reasons to require the equation's relativistic generalization to share this property. But that requirement seemed impossible to satisfy.

In 1928 Paul Dirac discovered a way around this difficulty. He devised a wave equation whose coefficients were not numbers but 4×4 matrices – rectangular arrays of (real or complex) numbers that can be added to and multiplied by one another according to certain rules. These matrix-coefficients weren't arbitrary: they were determined by the requirement that the relativistic wave equation be consistent with special relativity's version of the law of energy conservation. Three years earlier, Heisenberg had used matrices to represent an electron's position coordinates and momentum components, but no physical law with matrices as coefficients had previously been suggested.

Schrödinger's equation has a single solution that satisfies given boundary conditions. It represents the quantum state of a point-like charged particle. Dirac's relativistic wave equation has not one but four solutions.

One pair of these solutions describes electrons with an extra, internal degree of freedom. Dirac proved that the extra degree of freedom shows up in two ways: as an intrinsic angular momentum, or spin;

and as an intrinsic magnetic moment. A classical point charge has neither an intrinsic angular momentum nor an intrinsic magnetic moment. Dirac showed that his relativistic wave equation predicts that electrons have both. It also predicts that the spin has two possible values, $h/4\pi$ and $-h/4\pi$, where h is Planck's constant. And it predicts the ratio between the intrinsic magnetic moment and the spin. The predicted ratio has twice the value classical electromagnetic theory predicts for a spinning charged sphere. Experiments confirmed both predictions with high accuracy. For example, Dirac's theory predicts that each energy level of an electron in an external magnetic field splits into a pair of levels, one with the electron's magnetic moment parallel to the magnetic field line at its position, the other antiparallel; and it predicts the magnitude of the splitting.

Dirac's wave equation also made new predictions about the energy spectrum of hydrogen. Schrödinger's equation predicts that the electron in a hydrogen atom has a discrete set of negative energy states. The energy of the nth state, for $n = 1, 2, 3, \ldots$, is inversely proportional to n^2. Schrödinger's equation also predicts the (negative) constant of proportionality. Both predictions agree closely with experiment. Dirac's relativistic theory predicts that the energy levels have a fine structure that depends on a second quantum number, j, which also takes integer values. This prediction matched precise experimental measurements.

The second pair of solutions to Dirac's relativistic wave equation was harder to interpret than the first pair. It seems to describe electrons in impossible states of negative total energy (including the rest energy mc^2). Dirac eventually concluded that the negative-energy solutions represent a previously unknown particle with the same mass and spin as the electron and an equal but opposite charge. A short time later, experimental physicists discovered a particle with precisely these characteristics, the positron.

Despite its successes, Dirac's relativistic theory of the electron raised important technical problems. "[T]hese problems were all eventually to be solved (or at least clarified) through the development of quantum field theory."[12]At the same time, quantum field theory reinforced the link between quantum mechanics and special rela-

12 Weinberg, Steven, *The Quantum Theory of Fields*, (Cambridge University Press, 1995,) Volume 1, p. 14

tivity. In the preface to his magisterial two-volume *The Quantum Theory of Fields*, Steven Weinberg writes: "The point of view of this book is that quantum field theory is the way it is because ... it is the only way to reconcile the principles of quantum mechanics ... with those of special relativity."[13]

Special relativity, in turn, is a local approximation to general relativity, Einstein's unified theory of spacetime and gravitation. In Minkowskian spacetime the squared spacetime interval between two events equals the squared time interval between the events minus the squared distance interval, which in turn equals the sum of the three squared coordinate intervals. Gauss showed that the geometry of a smoothly curved surface could be derived from a formula that expresses the squared distance between two neighboring points on the surface as a sum of multiples of the squares of the coordinate intervals in a system of curvilinear coordinates, in which the multiples vary smoothly with position. Riemann, Gauss's pupil, extended his theory to curved spaces with any number of dimensions. Riemann's theory expresses the squared distance between two neighboring points as a sum of multiples of the squares of the coordinate intervals in an n-dimensional system of curvilinear coordinates; the multiples vary smoothly with position. Finally, Einstein's theory expresses the squared distance between two neighboring events in spacetime as the difference between a multiple of the squared time interval between the events and a sum of multiples of the squared space intervals in a four-dimensional system of curvilinear coordinates; each of the four multiples is a smooth function of the four coordinates. Einstein's field equations relate these multiples and their first and second derivatives to the spacetime distribution and flow of mass/energy. *Quantum mechanics presupposes that the (undisturbed) systems it describes are embedded in the spacetime Einstein's general theory of relativity describes.*

For decades physicists have tried to construct a mathematical structure that would contain general relativity and quantum mechanics as limiting cases. Yet, as we've seen, quantum mechanics needs general relativity; and general relativity doesn't explain the microstructure of the physical world. The preceding discussion shows that in a universe that comports with the strong cosmological principle, quantum

13 ibid. p. xxi

mechanics and general relativity fit smoothly together, linked by the quantum version of Gibbs's statistical mechanics and the ensemble version (2) of Dirac's general assumption (1). In this scheme chance arises from the postulated existence of space-time coordinate systems relative to which a complete description of the physical world privileges no position or direction, and manifests itself in quantum indeterminacy, in the probabilities that relate microstates (governed by quantum mechanics) and macrostates (governed by classical physics), and in the largely indeterminate character of the initial conditions that define macroscopic objects and processes.

VII
Living World

Physicalism rests on the assumption that the physical quantities that occur in classical, or non-quantum physical theories, have definite values even when we don't or can't know them precisely. It implies a modified form of determinism that allows for quantum indeterminacy. As an alternative to physicalism I've proposed a pair of cosmological hypotheses: the strong cosmological principle and the assumption of primordial randomness. I've argued that these hypotheses define a framework for physics that supplies physical interpretations of chance in quantum mechanics and statistical mechanics. Now I want to argue that this framework allows us to view biology as an autonomous science – one that, though based on strongly confirmed physical laws, doesn't reduce to physics. In particular the new framework accommodates consciousness and, more generally, the inner lives of living organisms.

Biological Complexity

Although the same strongly confirmed physical laws govern physical and biological processes, Ernst Mayr has argued that living systems exhibit kinds of complexity that have no counterpart in the physical world. The organization of living systems

> endows them with the capacity to respond to external stimuli, to bind or release energy (metabolism), to grow, to differentiate, and to replicate. Biological systems have the further remarkable prop-

> erty that they are open systems, which maintain a stead-state balance in spite of much input and output. This homeostasis is made possible by elaborate feedback mechanisms, unknown in their precision in any inanimate system.[1]

Mayr immediately points out that "complexity in and of itself is not a fundamental difference between organic and inorganic systems." But organic complexity has "extraordinary properties not found in inert matter"[2]:

> The complexity of living systems exists at every hierarchical level, from the nucleus to the cell, to any organ system (kidney, liver, brain), to the individual, to the species, the ecosystem, the society. The hierarchical structure within an organism arises from the fact that the entities at one level are compounded into entities at the next higher level – cells into tissues, tissues into organs, and organs into functional systems.[3]

Moreover, "systems at each hierarchical level have two properties. They act as wholes ... and their characteristics cannot be deduced from the most complete knowledge of the components taken separately ..."[4] In other words, systems at each level are characterized by emergent properties – properties that evince novel and unpredictable kinds of order.

By contrast, the kinds of order that physicists study are predictable, at least in principle. As an example consider the orderly structure of hemoglobin, the molecule in the blood of vertebrates that carries oxygen from the lungs to the rest of the body. Quantum mechanics predicts the structure of hemoglobin. It also predicts that the molecule has two distinct conformations with different oxygen-binding affinities. But chemists, as chemists, don't ask how these two conformations came about – why they occur in the blood of vertebrates – or why the active form is found in the lungs but flips into the inactive form in muscle tissues undergoing exertion. Quantum mechanics doesn't address these questions. More generally, it doesn't seek to explain the *functions* and *histories* of specifically biological kinds of complexity – kinds of complexity that emerge in the course of evolution.

---

1 Mayr, Ernst, *Toward a New Philosophy of Biology* (Cambridge, Harvard University Press, 1988) p. 14

2 ibid.

3 ibid.

4 ibid, p15

Chance and Order

Earlier I quoted Mayr on the importance of chance in evolution. Chance, he points out, plays a central role both in genetic variation and natural selection. It is likewise central to our pre-scientific views of individual human experience and human history. Yet in the physicalist world-view only quantum measurements have objectively unpredictable outcomes, and quantum measurements are largely irrelevant to both genetic variation and natural selection. In the physicalist worldview the unpredictability of events that are not quantum measurement reflects our ignorance of relevant initial conditions. In contrast, in a scientific worldview that comports with the strong cosmological principle chance and the assumption of primordial randomness chance is objectively real at all levels of description, from the molecular to the cosmological. Randomness is the raw material from which processes governed by deterministic mathematical laws fashion myriad novel varieties of physical and – especially – biological order

Randomness and Information Have a Hierarchical Structure

In a scientific theory that comports with the strong cosmological principle physical quantities are associated with probability distributions. Such a theory assigns a physical quantity's possible values probabilities and assigns each of these values (or each small range of values) a probability. It interprets these probabilities as relative frequencies in a cosmological ensemble. Following Boltzmann, I defined the *randomness* of a probability distribution as the mean, or probability-weighted average, of the negative reciprocal of the probability, and the probability distribution's *information* as the amount by which the distribution's randomness falls short of its largest possible value – the value that meets given constraints. Claude Shannon showed that the information of a probability distribution is the sum of contributions associated with statistical correlations between pairs, triples, quadruples, and so on of the events to which the distribution assigns probabilities.

If the assumption of primordial randomness holds, the probability distributions that characterized the universe at the earliest times when our present strongly confirmed physical laws held were maximally random; they contained no information. As the universe expanded, physical processes governed by quantum mechanics and by the laws of general relativity created information in the form of multiparticle correlations. In particular, gravitational instability – the tendency of relatively dense regions to expand more slowly than less dense regions – caused the initially uniform distribution to become progressively more clumpy.[5] This process creates information on very large scales, leaving randomness on molecular and small macroscopic scales untapped.

Once self-gravitating systems have begun to form, gravitational collapse – the tendency of relatively dense regions to expand more slowly than less dense regions –creates spatial inequalities of mass density and temperature. Thus gravitational collapse diminishes randomness and creates information. It is the ultimate source of the information that, in accordance with Clausius's law of entropy non-increase, fueled the origin of, and continues to sustain, life on Earth.

In experimental physics and chemistry, the experimenter uses a local source of information to create the information that characterizes an experimental setup. As Bohr emphasized, experimental setups must be capable of being described in the language of classical physics. So descriptions of experimental setups don't contain quantum-level information. Of course, experimental *predictions* can and do involve quantum mechanics.

(Think of the Stern-Gerlach experiment.)

Biophysicists and biochemists use the experimental and theoretical methods of physics and chemistry to study specific biological phenomena. Their efforts leave little room for doubt if any that biological structures and processes are also physical structures and processes, governed by the same strongly confirmed physical laws. Yet the theory of biological evolution, within which accounts of specific biological structures and processes are necessarily embedded (recall Dobzhansky's dictum that nothing in biology makes sense except in the light of evolution), can't be construed as a physical theory. Physical theories make testable predictions about the outcomes of experiments or observations based on assumptions about initial

5 For a detailed account of one version of this process, see David Layzer, *Cosmogenesis: The Growth of Order in the Universe*, (Oxford, Oxford University Press, 1990)

conditions. In contrast, evolution gives rise to novel and unpredictable kind of order.

How do evolutionary explanations differ from explanations offered by physical theories? Evolutionary theories, like theories in the physical sciences, seek to show that the present state of a system could have evolved from a hypothetical earlier state through processes governed by physical laws. Laplace's theory of the origin of the solar system, for example, sought to explain why the system consists of planets that circle the Sun in the same direction. The theory purported to show that an initially structureless gas cloud governed by Newton's laws would evolve into a system with these features. Analogously, population genetics, the predictive arm of evolutionary theory, successfully models the evolutionary acquisition of relatively simple traits like industrial melanism, "the darkness—of the skin, feathers, or fur—acquired by a population of animals living in an industrial region where the environment is soot-darkened."[6]But population genetics doesn't aspire to predict, for example, that some species of fish will evolve into land-dwelling animals. We don't yet have a sufficiently detailed and reliable description of the evolutionary precursors of land-dwelling animals. And, as Mayr has emphasized, the evolutionary history of land-dwelling animals is punctuated, and to a considerable extent determined by, chance events that have major but unpredictable evolutionary consequences. The story that connects a structureless gas cloud to a model of the solar system can't have an evolutionary counterpart because chance events break the causal chain.

What Is Life?

Although the origin of life is still highly speculative, biologists' understanding of the history of life leaves little room for doubt that all living organisms descended from small populations of membrane--enclosed, approximately self-replicating collections of molecules. What distinguished these primordial organisms from nonliving molecular assemblies of similar composition and complexity?

Unlike nonliving molecular assembles, every living organism, no matter how simple, has interests and *needs* that set it apart from and put it into a particular relationship with its environment. A living

6 "Industrial Melanism", *Encyclopedia Britannica*.

organism must regulate its interaction with its surroundings in ways that allow it, among other things, to:

- replicate itself more or less accurately;
- maintain processes that allow it to import building materials and high-grade (i.e. low-entropy) information and export waste products and low-grade information;
- maintain close-to-optimal relative chemical abundances
- (homeostasis); protect itself against predators.
- promote the wellbeing and reproductive success of other, not necessarily related, individuals and groups.

These and other interests and needs not only distinguish organisms from nonliving systems. They also distinguish living organisms from one another. Every living organism has a unique set of interests and needs as well as a unique history.

The observation that sensitivity to external stimuli, sentience, and consciousness are ubiquitous in the living world suggests that *the emergence of enclosed, approximately self-replicating molecular assemblies coincided with the emergence of something like a point of view – a rudimentary version of subjectivity*. Inner lives enable living organisms to respond in unified and creative ways to the challenges and opportunities presented by the external world and to exploit the opportunities offered by the ubiquity of chance in the macroscopic world.

This hypothesis may seem daring but it's hard to avoid. Because evolution is a fact and consciousness and sentience are biological attributes, they must be the product of evolutionary processes, like bilateralism or the structure of hemoglobin. Since the most essential feature of consciousness is its subjective character, we can hardly avoid assuming that subjectivity came into being with our earliest ancestors, the first enclosed, approximately self-regulating molecular aggregates with interests and needs that include those mentioned above. How primitive forms of subjectivity evolved into more complex forms then becomes a scientific problem, like the origin of feathers.

Why then do many contemporary neuroscientists and philosophers deny the reality of consciousness as a feature of physical reality distinct from the neural processes that underpin it? The thesis that elementary particles and photons are the ultimate and only constituents of physical reality is neither required nor implied by any strongly confirmed physical law. One can deny it without accepting the strong cosmological principle and the assumption of primordial randomness. But someone who does accept these assumptions has a strong motive to reject physicalism because it implies (and physicalism denies) that chance, randomness, and order, along with the novel kinds of order that biological evolution fashions from the raw material of randomness are as real and objective as mass and electric charge.

Creativity

In a paper entitled "Evolution as a Creative Process" Theodosius Dobzhansky wrote: "A living organism resembles a work of art, and the evolutionary process resembles the creation of a work of art."[7]The creation of a work of art involves analogues of random genetic variation and natural selection. But I think it involves something more. Creative processes not only rely on macroscopic randomness, a feature of the physical world that is a consequence of the strong cosmological principle and the assumption of primordial randomness. They also exploit this feature. Consider the following example of the creative process, supplied by a member of my family.

My brother Bob is a lyric poet as well as a neurologist. I asked him to write down an account of how he writes a poem. This is what he wrote:

HOW I WRITE A POEM

by Robert Layzer

A new poem often starts with a phrase that evokes some interest or feeling, or points to a subject that might be interesting or moving. Sometimes the phrase describes something I saw or heard in the environment, or a painting that I am attached to; or it could be a fragment of a thought or a feeling.

7 Dobzhansky, T., *Proc. Ninth Int. Congress Genetics*, 1954, pp. 435-449

Rarely, the poem may emerge and be completed within a day. More often nothing much happens for a while-hours or days-while I repeat the phrase silently or aloud now and then. Then, out of nowhere, new phrases and sentences start to cluster around the original idea, like crystals coming out of solution. I begin to see a theme, and that suggests more words to enlarge the theme. My conscious mind becomes more involved as I shape and plan the structure and content. Pruning and revising may take place as the poem develops, or after it is nearly finished. This is a much more conscious process; now I am reading the poem in the context of the history of literature, as if I were a critic.

During the composition, I try to listen for unexpected words or ideas, and to the internal assonances and other harmonies that tell a story of their own. In fact, the sound of the poem may turn out to be the main carrier of meaning, and if the poem is successful I may not understand the real "message" for some time afterward. But if I'm successful, the ending often gives a sense of a question answered.

The creative process described in this account generates candidates for selection by a process that is both random and constrained: only candidates in certain broad categories, which themselves depend on the work being created, present themselves for selection. The goal of a creative process isn't known beforehand; the process isn't teleological. But once the process has reached its endpoint, its product can be seen to satisfy "fitness criteria" that were already in place; it "gives a sense of a question answered"; "everything comes together, and you say 'that's it.'" The sustained exploratory effort in artistic creation has a counterpart in Bergson's philosophical account of evolution: the *élan vital*. It also has counterparts in modern accounts of biological evolution. Notably, biologists John Gerhart and Marc Kirschner have proposed a theory of phenotypic variation in evolution – a theory of how differences between anatomical, physiological, and behavioral traits of organisms have evolved – that differs radically from older accounts, in which phenotypic variation and the divergence of genetic lineages result mainly from the action of natural selection on uncorrelated variations of individual genes:

> Most anatomical and physiological traits that have evolved since the Cambrian are, we propose, the result of *regulatory* changes in the usage of various members of

> a large set of conserved core components that function in development and physiology. Genetic change of the DNA sequences for regulatory elements of DNA, RNAs, and proteins leads to heritable regulatory change, which specifies new combinations of core components, operating in new amounts and states at new times and places in the animal. These new configurations of components comprise new traits.[8]

As I mentioned, artistic creation is an exploratory process. An artist searches for and eventually finds targets in a large "space" of imagined structures. Kirschner and Gerhard have argued that evolution is an exploratory process in this sense:

> As the name [Exploratory Processes] implies, some conserved core processes appear to search and find targets in large spaces or molecular populations. Specific connections are eventually made between the source and target. These processes display great robustness and adaptability and, we think, have been very important in the evolution of complex animal anatomy and physiology. Examples include the formation of microtubule structures, the connecting of axons and target organs in development, synapse elimination, muscle patterning, vasculogenesis, vertebrate adaptive immunity, and even behavioral strategies like ant foraging. All are based on physiological variation and selection. [9]

Kirschner and Gerhart have given a more extended (and highly readable) account of their view of evolution in *The Plausibility of Life*.[10]

In 1980 I discussed the hypothesis that evolution is a process of hierarchic construction, involving not just constraints on deleterious mutations but also what Kirschner and Gerhart call "deconstraints." The only example I then knew about was the adaptive vertebrate immune response. I append the abstract of that paper to this chapter.

Free Will

In his introduction to *The Oxford Handbook of Free Will* philosopher Robert Kane writes:

> [D]ebates about free will in the modern era (since the seventeenth century) have been dominated by two questions, not one – the "Determinist Question": "Is determinism true?" and the "Compatibility Question": "Is free will compatible (or incompatible) with determinism?"[11]

Some scientists and philosophers have argued that free will must be an illusion because genuine freedom – the capacity to influence the course of future event – would be incompatible with determinism; and science, they argue, is deterministic: past conditions and physical laws determine future events. Although quantum mechanics predicts that certain processes – quantum determined measurements – have unpredictable outcomes, such processes don't begin to account for the kind of unpredictability required by a robust conception of human freedom.

If we replace physicalism's assumption that classical physical quantities have definite values by the strong cosmological principle and the assumption of primordial randomness, the answer to Kane's first question – Is determinism true? – is *no*. Initial conditions in biological processes ranging from evolution to individual development and cultural evolution are products of histories punctuated by unpredictable events. And these processes have a creative character. They give rise to novel and unpredictable forms of order.

Since human behavior is a province of biology, we can now replace Kane's second question ("Is free will compatible (or incompatible) with determinism?") by the question "Is free will compatible with biology?" If this book's central argument is correct, the answer is an unambiguous yes.

Appendix, Chapter VII

Abstract. The present theory offers a unified solution to three closely related evolutionary problems. (1) Why does an evolving population explore only a small fraction of the accessible pathways in genotype space? (2) Conventional ideas about genetic variation suggest that major adaptive shifts, which involve large numbers of separate but functionally related genetic changes, have vanishingly small probabilities of occurrence and require long periods of time. Yet the fossil record indicates that such shifts are neither slow nor uncommon. (3) Studies in comparative morphology and comparative embryology indicate that evolution is a process of hierarchic construction. How is the principle of hierarchic construction related to the basic postulates of evolutionary theory? According to the theory elaborated in this paper, the genomes of both unicellular and multicellular organisms contain two functionally distinct systems of genes: an α system, which encodes a program for the organism's development and which includes regulatory genes and other "modifiers" as well as structural genes; and a β system, which encodes a strategy for adaptive variability and whose elements regulate the rates of genetic recombination and of genic and chromosomal mutations. Corresponding to these two systems are two classes of adaptations: α adaptations, which enhance the fitness of their possessors, and β adaptations, which increase the expectation of fitness in the descendants of their possessors. Sexual reproduction is the most familiar example of a β adaptation. It is noteworthy that the most primitive known form of sexuality, that found among bacteria, is characterized by genetic mechanisms serving to regulate genetic recombination in precisely the manner postulated here for β genes. The β system serves to direct evolutionary flows in genotype space into potentially adaptive channels. It also serves to stabilize and buffer highly adapted genetic structures against the potentially disruptive effects of accidental variations. At the genetic level of description, a major adaptive shift corresponds to the ascent of a fitness peak in a multidimensional subspace of genotype space. The β system serves to focus evolutionary flow in this subspace in the direction of steepest ascent, thereby ensuring rapid and coordinated evolution of the genetic systems involved. (A mathematical framework that should make it possible to construct numerical models of the evolutionary

process just described is given by Layzer[12]) The principle of hierarchic construction emerges as a natural consequence of selection-regulated genetic variation. Hierarchic units are defined by their covariability (or costability). The "otherwise inexplicable tendency of organisms to adopt ever more complicated solutions to the problem of remaining alive" (Medawar 1967, pp. 99-100) results from the action (and progressive elaboration) of a genetic system that promotes just those kinds of genetic variation that result in the growth of functionally hierarchic genetic systems. The present theory also throws light on several specific evolutionary problems, including the interpretation of genic polymorphism, the variability of evolutionary rates inferred from the fossil record, the evolution of "pseudoexogenous" and "trivial" adaptation, and the problem of speciation. Applications to the evolution of social behavior are discussed elsewhere.[13]

12 Layzer, David. "A macroscopic approach to population genetics." *Journal of Theoretical Biology* 73.4 (1978): 769-788.

13 Layzer, D., *The American Naturalist*, 115, 6, June 1980

In the following several pages, we present two excerpts from Layzer's important 1990 book, *Cosmogenesis*, and two of his unpublished manuscripts on free will written in 2010 and 2011, some years before *Why We are Free*.

Cosmogenesis, Chapter 1 (excerpt)

The Problem of Order

Around the beginning of the sixth century b.c., in the prosperous trading center of Miletus on the Aegean coast of present-day Turkey, a handful of Greek thinkers made the first recorded efforts to construct a rational or nonmythical account of how order arose in the world. It seemed obvious to them that the world could not have sprung from emptiness. "Nothing," they were fond of saying, "comes from nothing." It seemed equally obvious that the world didn't spring into being ready-made: the *kosmos*, they all agreed, must have evolved from a primordial chaos. made: (In classical Greek, kosmos means both "order" and "world.") But how? West- em science and Western philosophy have their roots in the efforts of Thales, Anaximander, Anaximenes, and their successors to answer this question.

Philosophers have long since ceded the question to natural scientists, who, following science's oldest and most fruitful methodological precept, divide and conquer, have separated it into more specific questions:

• How can we account for the permanence, stability, and orderliness of crystals, molecules, atoms, and subatomic particles?

• How did the complex hierarchic structure of the astronomical Universe come into being?

• What is the origin of biological organization in all its manifestations, from DNA to the human mind?

• And why do most kinds of order tend to crumble and decay?

These four questions are central to four great divisions of natural science: quantum physics, cosmology, biology, and macroscopic physics or thermodynamics. These disciplines have yielded a deeper and more detailed understanding of order in its varied manifestations than anyone could have anticipated, even fifty years ago. There are gaps, of course. Theoretical physicists are still striving to unify the laws of physics; the origin of life is still a mystery; and the origin of astronomical systems remains an area of speculation and controversy. These problems lie at the frontiers of modern science, and they are getting a lot of attention.

But another kind of incompleteness in natural science's picture of the world has received less attention: *the four great pieces of the picture don't quite fit together.* Each piece, although still incomplete, is remarkably coherent. Each piece is connected to other pieces. But the connections aren't smooth. There are deep unresolved conflicts between quantum physics and macroscopic physics, between macroscopic physics and cosmology, and between the physical sciences and biology-

Conflicts and Paradoxes

The relation between quantum physics, which describes the invisible world of elementary particles and their interactions, and macroscopic physics, which describes the world of ordinary experience, has perplexed physicists since the birth of quantum physics in 1925. Viewed as a system of mathematical laws, quantum physics includes macroscopic physics as a limiting case. By that I mean that quantum physics and macroscopic physics make the same predictions in the domain where macroscopic physics has been strongly corroborated (the macroscopic domain), but quantum physics also successfully describes the behavior and structure of molecules; atoms, and subatomic particles (the microscopic domain). Yet from another point of view, macroscopic physics seems more fundamental than quantum physics. As we will see later, the laws of quantum physics refer explicitly to the results of measurement. But every measurement necessarily has at least one foot in the world of ordinary experience:

it has to be recorded in somebody's lab notebook or on magnetic tape. So quantum physics seems to presuppose its own limiting case — macroscopic physics. This is the mildest of several paradoxes that have sprung up in the region where quantum physics and macrophysics meet and overlap.

The relation between macrophysics and cosmology is also problematic. The central law of macroscopic physics — the second law of thermodynamics — was understood by its inventors, and is still understood by most scientists, to imply that the Universe is running down — that order is degenerating into chaos. How can we reconcile such a tendency with the fact that the world is full of order — that it is a kosmos in both senses of the word. Some scientists say, "The contradiction is only apparent, The Second Law assures us that the Universe is running down, so it must have begun with a vast supply of order that is gradually being dissipated. But this way of trying to resolve the difficulty takes us from the frying pan into the fire, because, as we will see, modern cosmology strongly suggests that the early Universe contained far less order than the present-day Universe.

Astronomical evolution and biological evolution are both stories of emerging order. Nevertheless, the views of time and change implicit in modern physics and modern biology are radically different. The physical sciences teach us that all natural phenomena are governed by mathematical laws that connect every physical event with earlier and later events. Imagine that every past and future event was recorded on an immense roll of film. If we knew all the physical laws, we could reconstruct the whole film from a single frame. And in principle there is nothing to prevent us from acquiring complete knowledge of a single frame.

This worldview is epitomized in a much-quoted passage by one of Newton's most illustrious successors, the mathematician and theoretical astronomer Pierre Simon de Laplace (1749-1827):

> We ought then to regard the present state of the Universe as the effect of its previous state and the cause of the one that follows. An intelligence that at a given instant was acquainted with all the forces by which nature is animated and with the state of the bodies of which it is composed would — if it were vast enough to submit these data to analysis — embrace in the same formula the move-

> ments of the largest bodies in the Universe and those of the lightest atoms: Nothing would be uncertain for such an intelligence, and the future like the past would be present to its eyes. The human mind offers, in the perfection it has been able to give to astronomy, a feeble idea of this intelligence.

Much the same view of the world was held by Albert Einstein:

> The scientist is possessed by the sense of universal causation. The future, to him, is every whit as necessary and determined as the past.

Most contemporary physical scientists would probably agree with Laplace and Einstein. The world they study is a block universe, a four-dimensional net of causally connected events with time as the fourth dimension. In this world, no moment in time is singled out as "now." For Laplace's Intelligence, the future and the past don't exist in an absolute sense, as they do for us.

How does life, regarded as a scientific phenomenon, fit into this worldview?

> A modern Laplacian might reply: Living organisms are collections of molecules that move and interact with one another and with their environment according to the same laws that govern molecules in nonliving matter. A supercomputer, supplied with a complete microscopic description of the biosphere and its environment, would be able to predict the future of life on Earth and to deduce its initial state. Implicit in the present state of the biosphere and its environment are the precise conditions that prevailed in the lifeless broth of organic molecules in which the first self-replicating molecules formed, And implicit in the conditions that prevailed in that broth and its environment is every detail of the living world of today.

If you believe that living matter is subject to the same laws as nonliving matter and few, if any, contemporary biologists would dispute this assertion - this argument may seem compelling. Yet it clashes with two key aspects of the evolutionary process as described by contemporary evolutionary biologists: randomness and creativity.

Randomness is an essential feature of the reproductive process. In nearly every biological population, new genes and new combinations of genes appear in every generation. Reproduction, whether sexual or asexual, involves the copying of genetic material (DNA). In all modern organisms the copying process is astonishingly accurate. But it isn't perfect. Occasionally there are copying errors, and these have a random character. In sexually reproducing populations there is another source of randomness: the genetic material of each individual is a random combination of contributions from each parent.

The *creative* factor in biological evolution is natural selection, the tendency of genetic changes that favor survival and reproduction to spread in a population, and of changes that hinder survival and reproduction to die out. From the raw material provided by genetic variation, natural selection fashions new biological structures, functions, and behaviors.

A mainstream physicist might reply that the apparent randomness of genetic variation is just a consequence of human ignorance — our inability to understand exceedingly complex but nevertheless completely determinate causal processes — and that evolution is "creative" only in a metaphorical sense. According to this view, evolution merely brings to light varieties of order prefigured in the prebiotic broth.

There is an even more fundamental difference between the physical and the biological views of reality: the physicist's picture of reality seems impossible to reconcile with subjective experience. For there is nothing in the neo-Laplacian picture that corresponds to the central feature of human experience, the passage of time. We humans must watch the film unwind, but Laplace's Intelligence sees it whole. Nor is there anything that corresponds to the aspect of reality (as we experience it) that Greek philosophers called becoming, as opposed to the timeless being of numbers, triangles, and circles. The universe of modern physics is an enormously expanded and elaborated version of the perfectly ordered but static and lifeless world we encounter in Euclid's *Elements*, of which it is indeed a direct descendant. The biologist's world seems entirely different. Life, as we experience it, is inseparable from unpredictability and novelty.

Freedom and Necessity

What is the relation between being and becoming? Is the future as fixed and immutable as the past? What is chance? These questions bear on one of the perennial problems of Western philosophy, the problem of freedom and necessity.

Each of us belongs to two distinct worlds. As objects in the world that natural science describes we are governed by universal laws. To Laplace's Intelligence we are systems of molecules whose movements are no less predicable and no more the results of free choice than the movements of the planets around the Sun. but as the subjects of our own experience we see the world differently; not as bundles of events frozen into the block universe of Laplace and Einstein like flies in amber, but as the authors of our own actions, the molders of our own lives. However strongly we may believe in the universality of physical laws, we cannot suppress the intuitive conviction that the future is to some degree open and that we help to shape it by our own free choices.

This conviction lies at the basis of every ethical system. Without freedom there can be no responsibility. If we are not really free agents — if our felt freedom is illusory — how can we be guided in our behavior by ethical precepts? And why should society punish some acts and reward others? The Laplacian worldview tends to undermine the basis for ethical behavior.

Judeo-Christian theology faces a similar problem. Although Laplace's Intelligence is not the Judeo-Christian God — Laplace's Intelligence observes and calculates; the Judeo-Christian God wills and acts ("Necessitie and chance approach not mee, and what I will is Fate," says the Almighty in Milton's Paradise Lost)— they contemplate similar universes. Nothing is uncertain for an all-knowing God, and the future, like the past, is present to His eyes. But if we cannot choose where we walk, why should those who take the narrow way of righteousness be rewarded in the next life while those who take the primrose path are consigned to the flames of hell?

Theologians have not, of course, neglected this question. Augustine, for example, argued that God's foreknowledge (or more accurately, God's knowledge of what we call the future) doesn't cause events to happen and is therefore consistent with human free will. Other theologians have embraced the doctrine of predestination and argued that free will is indeed an illusion. Still others have taken the position that divine omniscience and human free will are compatible in a way that surpasses human understanding.

Reconciling the scientific and ethical pictures of the world was a concern of the first scientists. Our scientific picture of the world was foreshadowed by Greek atomism, a theory invented by the natural philosophers Leucippus and Democritus in the fifth century B.C. According to this theory, the world is made up of unchanging, indestructible particles moving about in empty space and interacting with one another in a completely deterministic way. Like modern biologists, Democritus believed that we, too, are assemblies of atoms. Yet Democritus also elaborated a system of ethics based on moral responsibility. He taught that we should do what is right not from fear, whether of punishment or of public disapproval or of the wrath of gods, but in response to our own sense of right and wrong. Unfortunately, the surviving fragments of Democritus's writings don't tell us how or whether he was able to reconcile his deterministic picture of nature with his doctrine of moral responsibility.

A century later, another Greek philosopher with similar ideas about physical reality and moral responsibility faced the same dilemma. Epicurus (341-270 B.C.) sought to reconcile human freedom with the atomic theory by postulating a random element in atomic interactions. Atoms, he said, occasionally "swerve" unpredictably from their paths. In modern times, Arthur Stanley Eddington and other scientists have put forward more sophisticated versions of the same idea. According to quantum physics, it is impossible to predict the exact moment when certain atornic events, such as the decay of a radioactive nucleus, will take place. Eddington believed that this kind of microscopic indeterminism might provide a scientific basis for human freedom:

> It is a consequence of the advent of quantum theory that physics is no longer pledged to a scheme of deterministic laws. . . . The future is a combination of the causal influences of the past together with unpredictable elements. . [S]cience thereby withdraws its moral opposition to free will.

But neither Epicurus nor Eddington explained what the "freedom" enjoyed by a swerving atom or a radioactive atomic nucleus has to do with the freedom of a human being to choose between two courses of action. Nor has anyone else.

Cosmogenesis, Chapter 15 (excerpt)

Chance, Necessity, and Freedom

To be fully human is to be able to make deliberate choices. Other animals sometimes have, or seem to have, conflicting desires, but we alone are able to reflect on the possible consequences of different actions and to choose among them in the light of broader goals and values. Because we have this capacity we can be held responsible for our actions; we can deserve praise and blame, reward and punishment. Values, ethical systems, and legal codes all presuppose freedom of the will. So too, as P. F. Strawson has pointed out, do “reactive attitudes” like guilt, resentment, and gratitude. If I am soaked by a summer shower I may be annoyed by my lack of foresight in not bringing an umbrella, but I don’t resent the shower. I could have brought the umbrella; the shower just happened.

Freedom has both positive and negative aspects. The negative aspects — varieties of freedom from — are the most obvious. Under this heading come freedom from external and internal constraints. The internal constraints include ungovernable passions, addictions, and uncritical ideological commitments. The positive aspects of freedom are more subtle. Let’s consider some examples.

1. A decision is free to the extent that it results from deliberation. Absence of coercion isn’t enough. Someone who bases an important decision on the toss of a coin seems to be acting less freely than someone who tries to assess its consequences and to evaluate them in light of larger goals, values, and ethical precepts.

2. Goals, values, and ethical precepts may themselves be accepted uncritically or under duress, or we may feel free to modify them by reflection and deliberation. Many people don't desire this kind of freedom and many societies condemn and seek to suppress it. Freedom and stability are not easy to reconcile, and people who set a high value on stability tend to set a correspondingly low value on freedom. But whether or not we approve of it, the capacity to reassess and reconstruct our own value systems represents an important aspect of freedom.

3. Henri Bergson believed that freedom in its purest form manifests itself in creative acts, such as acts of artistic creation. Jonathan Glover has argued in a similar vein that human freedom is inextricably bound up with the "project of self-creation." The outcomes of creative acts are unpredictable, but not in the same way that random outcomes are unpredictable. A lover of Mozart will immediately recognize the authorship of a Mozart divertimento that he happens not to have heard before. The piece will "sound like Mozart." At the same time, it will seem new and fresh; it will be full of surprises. If it wasn't, it wouldn't be Mozart. In the same way, the outcomes of self-creation are new and unforeseeable, yet coherent with what has gone before.

Although philosophical accounts of human freedom differ, they differ surprisingly little. On the whole, they complement rather than conflict with one another. What makes freedom a philosophical problem is the difficulty of reconciling a widely shared intuitive conviction that human beings are or can be free (in the ways discussed above or in similar ways) with an objective view of the world as a causally connected system of events. We feel ourselves to be free and responsible agents, but science tells us (or seems to tell us) that we are collections of molecules moving and interacting according to strict causal laws.

For Plato and Aristotle, there was no real difficulty. They believed that the soul initiates motion — that acts of will are the first links of the causal chains in which they figure. With few exceptions, modern neurobiologists have rejected the view of the relation between mind and body that this doctrine implies. They regard mental processes as belonging to the natural world, subject to the same physical laws that govern inanimate matter. The differences between animate and inanimate systems and between conscious, and nonconscious nervous processes are not caused by the presence or absence of nonmaterial substances (the breath of, life, mind, spirit, soul) but by the presence or absence of certain kinds of order. This conclusion is more than a profession of scientific faith. It becomes unavoidable once we accept the hypothesis of biological evolution, without which, as Theodosius Dobzhansky remarked, nothing in biology makes sense. The evolutionary hypothesis implies that human consciousness evolved from simpler kinds of consciousness, which in turn evolved from nonconscious forms of nervous activity. There is no point in this evolutionary sequence where mind or spirit or soul can plausibly be assumed to have inserted itself "from without." It seems even more implausible to suppose that it was there all along, although, as we saw earlier, some modem philosophers and scientists have held this view.

Karl Popper and other philosophers have tried to resolve the apparent conflict between free will and determinism by attacking the most sacred of natural science's sacred cows, the assumption that all natural processes obey physical laws.

In asserting that there may be phenomena that don't obey physical laws, these philosophers are obviously on safe ground. But the assumption of indeterminism doesn't really help. A freely taken decision or a creative act doesn't just come into being. It is the necessary — and hence law-abiding — outcome of a complex process. Free actions also have predictable — and hence lawful - consequences; otherwise, planning and foresight would be futile. Thus every free act belongs to a causal chain: it is the necessary outcome of a deliberative or creative process, and it has predictable consequences.

Some physicists and philosophers have suggested that quantal indeterminacy may provide leeway for free acts in an otherwise deterministic Universe. Freedom, however, doesn't reside in randomness; it resides in choice. Plato and Aristotle were right in linking Chance and Necessity as "forces" opposed to design and purpose in the Universe.

Thus freedom seems equally inconsistent with determinism and indeterminism. Thomas Nagel has suggested that it isn't even possible to give a coherent account of our inner sense of freedom:

When we try to explain what we believe which seems to be undermined by a conception of actions as events in the world - determined or not — we end up with something that is either incomprehensible or clearly inadequate.

"The real problem," Nagel says, "stems from a clash between the view of action from inside and any view of it from outside." Yet the intuitive view of what it means to be free doesn't rest on introspection alone. We recognize other people's spontaneity and creativity even — or especially — when it is of such a high order that we can't imagine ourselves capable of it. We can apprehend the exquisitely ordered unpredictability of Mozart's music without beginning to be able to imagine what it would be like to compose such music. And even subjective impressions of freedom, unlike subjective impressions of pain or of self, aren't hard to describe. Consider the process of making a decision. Shall I do A or B? My head says A; my heart says B. I agonize. I try to imagine the consequences first of A, then of B. Suddenly, a new thought occurs to me: C. Yes, I'll do C. The essential aspect of such commonplace experiences is that their outcomes aren't determined in advance but are created by the process of deliberation itself, a process unfolding in time. All creative processes have this character.

Such processes, however, go on not only in people's subjective awareness but also in their brains. Conscious experience gives us a fragmentary and unrepresentative view of its underlying cerebral processes, but there is no reason to suppose that the view is deceptive. On the contrary, modern techniques of imaging brain activity suggest that there is a high degree of structural correspondence between consciousness and brain activity.

If, then, the outcome of a deliberative or creative process seems undetermined at the outset, if it seems to us that such processes create their outcomes, perhaps the reason is that the outcomes of the underlying cerebral processes are, in some objective sense, undetermined, are, in some objective sense, created by the processes themselves.

I will argue that the neural processes that give rise to subjective experiences of freedom are indeed creative processes, in the sense, that they bring into the world kinds of order that didn't exist earlier and weren't prefigured in earlier physical states. These novel and unforeseen products of neural activity include not only works of art, but also the evolving patterns of synaptic connections that underlie the intentions, plans, and projects that guide our commonplace activities. Although consciousness gives us only superficial and incomplete glimpses of this ceaseless constructive activity, we are aware of it almost continuously during our waking hours. This awareness may be the source of — or even constitute — the subjective impression that we participate in molding the future.

Much of the argument that supports this view has already been given in earlier chapters. Let me now try to pull it together around the following three questions:

1. Do all law-abiding processes have predetermined outcomes?

2. What does it mean to say that a physical process creates its outcomes?

3. How is this kind of creativity related to creativity in contexts relevant to the problem of human freedom?

[Answer to question 1]: Do all law-abiding processes have predetermined outcomes? Outcomes are determined by laws plus initial conditions. They are undetermined to the extent that the initial conditions are unspecified.

[Answer to question 2]: A theory of cosmic evolution requires initial conditions. The simplest initial conditions is that the Universe began to expand from a purely random state — a state wholly devoid of order. From this postulate, we can easily deduce the Strong Cosmological Principle. The inference hinges on the fact that none of our present physical laws discriminates between different points in space or between different directions at a point. (A physicist would say, "The laws are invariant under spatial translations and rotations.") This implies that no physical process can introduce discriminatory information. So if information that would discriminate between positions or directions is absent at a single moment, it must be absent forever. In short, if the Strong Cosmological Principle is valid at any single moment, it must be valid for all time.

If the Universe began to expand from a state of utter randomness, how did order come into being? Before reviewing our answer to this question, we have to recall how we dealt with the concept of order itself.

The two key ideas needed to formulate an adequate scientific definition of order were put forward by Ludwig Boltzmann.

The first idea is the distinction between microstates and macrostates. Macrostates are groups of microstates, defined by their statistical properties. For example, the microstates of a gas may be assigned to macrostates defined by density, temperature, and chemical composition. Proteins may be assigned to macrostates defined by biological fitness. Boltzmann's second key idea was to identify the randomness or entropy of a macrostate with the logarithm of the number of its microstates. Supplementing this definition of randomness, we defined the order or information of a macrostate as the difference between its potential randomness or entropy (the largest value of the randomness or entropy consistent with given constraints) and the actual value. Thus maximally random macrostates have zero order and maximally ordered macrostates have zero randomness. According to these definitions, a physical system far removed from thermodynamic equilibrium (the macrostate of maximum randomness) is highly ordered. So is a protein whose biological fitness can't be improved by changes in its sequence of amino acids: it belongs to a very small subset of the class of polypeptides of the same length.

These definitions of randomness and order are important not just, or even primarily, because they lend precision to the corresponding intuitive notions in a wide range of scientific contexts. They are important primarily because they are adapted to theoretical accounts of the growth and decay of order. Boltzmann himself proved (under restrictive assumptions) that molecular interactions in a gas not already in its most highly random macrostate increase its randomness. In Chapter 8 we saw how the cosmic expansion generates chemical order (chemical abundances far removed from those that would prevail in thermodynamic equilibrium); in Chapter 9 we discussed the origin and growth of structural order in the astronomical Universe; and in Chapters 10 and 11 we saw how random genetic variation and differential reproduction generate the biological order encoded in genetic material.

Astronomical and biological order-generating processes are hierarchically linked in the manner discussed in Chapter 2. Each process requires initial conditions generated by earlier processes. For example, the first self-replicating molecules needed an environment that provided high-grade energy, molecular building blocks, and catalysts. High-grade energy was supplied, directly or indirectly, by sunlight, produced by the burning of hydrogen deep inside the Sun. To understand why hydrogen is so abundant, we have to go back to the early Universe, when the primordial chemical composition of the cosmic medium was laid down by an interplay between nuclear reactions and the cosmic expansion. Apart from hydrogen, the atoms that make up biomolecules (carbon, oxygen, and nitrogen are the most common) were synthesized in exploding stars far more massive than the Sun. So, too, were inorganic catalysts like zinc and magnesium. Finally, the emergence of an environment favorable to life as we know it resulted from planet-building processes, for which we still lack an adequate theory.

Although some of the specific order-generating processes we have discussed are speculative or controversial, the general principles underlying the emergence of order from chaos seem more secure. In particular, we can now understand why, in spite of the second law of thermodynamics, the Universe is not running down. The Second Law states that all natural processes tend to increase randomness. In an ordinary isolated system, the growth of randomness leads inevitably to a decline of order, because the sum of randomness and order is a fixed quantity.

The Universe, however, is not an ordinary isolated system. Because space is expanding, the sum of randomness and order is not a fixed quantity; it tends to increase with time. Hence a gap may open up between the actual randomness of the cosmic medium and its maximum possible randomness. This gap represents a form of order. Chemical order (as evidenced by the prevalence of hydrogen) emerges when equilibrium-maintaining chemical reactions can no longer keep pace with the cosmic expansion. Structural order (in the form of astronomical systems) emerges when the uniform state of an expanding medium becomes unstable—that is, less than maximally random.

By making randomness an objective property of the Universe, the Strong Cosmological Principle also objectifies the timebound varieties of order, which consist in the absence of randomness. The infinitely detailed world picture of Laplace's Intelligence is devoid of macroscopic order. It contains no objective counterpart to astronomical or biological order. Laplace's Intelligence is an idiot savant. It knows the position and velocity of every particle in the Universe;

but because this vast fund of knowledge (or its quantal-counterpart) is complete in itself, there is no room in it for information about stars, galaxies, plants, animals, or states of mind. In this book I have argued that the external world — the world that natural science describes — is fundamentally different from the universe of Laplace and Einstein, which is given once and for all in space and time (or in spacetime). It is a world of becoming as well as being, a world in which order emerged from primordial chaos and begot new forms of order. The processes that have created and continue to create order obey universal and unchanging physical laws. Yet because they generate information, their outcomes are not implicit in their initial conditions.

Creative Processes

All order-generating processes may be said to be creative, but some seem to deserve the label more than others. For example, the evolution of chemical order in the early Universe seems less creative than the evolution of biological order. To gain insight into this difference, let's compare the evolution of a star cluster with the evolution of a biological population. Suppose we are given a statistical description of the cluster's initial state and asked to calculate its subsequent evolution. To do the calculation, we have to assign an initial position and velocity to each star. This can be done in many different ways that are consistent with the given statistical description of the initial state, and different assignments will yield different evolutionary trajectories. But if the number of stars is large, these evolutionary trajectories diverge very little, because each star responds to the combined attraction of all the others, and the combined attraction is insensitive to statistical fluctuations in the cluster's initial state.

Now consider a biological population. Suppose we knew everything that could in principle be known about the population's initial state, including the genotypes of all the organisms belonging to the population. Suppose we also had the ability to simulate on a supercomputer every relevant aspect of the evolutionary process.

Could we then predicts what genotypes would be present in the population at some later time?

No — at least not for a population undergoing significant evolutionary change. The reason is that evolutionary outcomes are very sensitive to some of the random genetic changes brought about by mutation and genetic recombination. Suppose we could enumerate all the possible outcomes of every mutational and recombinational event and assign a probability to each of them. We would then be able, in principle, to construct a complete statistical description of our evolving population. This description would encompass a vast number of qualitatively distinct, multiply branching pathways, each with only a tiny probability of being realized. It would therefore contain very little information about the history of any given population. A prediction about the outcome of a horse race that assigns small and nearly equal probabilities of winning to each of a large number of entrants isn't very informative.

Biological evolution, therefore, not only generates order and information, but does so in an essentially unpredictable way. This, I suggest, is an essential element of every truly creative process. A creative process not only generates order, but does so in an essentially unpredictable way.

We don't yet fully understand the biological basis of creative human activity, but I find the analogy with biological evolution compelling. In Chapter 14 I suggested that higher mental processes are mediated by a cyclic process in which the brain constructs, tests, and modifies internal representations. It is tempting to speculate that the process by which internal representations are constructed has a strong random component, in addition to systematic components that are built up in the course of individual development and that constrain and channel the random component. The systematic components would play a role analogous to that of beta genes in the evolutionary theory sketched in Chapter 11. They would be responsible for the elements of an artist's work that we recognize as his or her individual style.

[Answer to question 3]: Creative human activity is unpredictable in the same way and for the same reasons that biological evolution is unpredictable. Unpredictability, however, is only one aspect of human freedom. We are free because we are, to a considerable extent, the authors of our own lives, and because every human life is something new under the Sun. That is what Democritus and Socrates believed; and if the picture I have sketched in this book is correct in its main outlines, it is also one of the lessons of modern science. Our awareness of the openness of the future and of our own ability to help shape it reflects a deep property of objective reality.

The scientific worldview sketched in the preceding pages offers an alternative to reductionism in both its physical and its biological forms. It shows us that the Universe is more than a collection of elementary particles governed by immutable mathematical laws. Order and the processes that bring order into being lie at the heart of reality. Biological evolution, cultural evolution, and individual human lives not only are the most prolific sources of order in the known Universe, but also are creative. Because of them, the future is genuinely open.

Naturalizing Libertarian Free Will[1]

David Layzer

Department of Astronomy, Harvard University

Libertarian free will is incompatible with the thesis that physical laws and antecedent conditions determine events other than the outcomes of quantum measurements. This thesis is not a consequence of physical laws alone. It also depends on an assumption about the conditions that define macroscopic systems (or, in some theories, the universe): the assumption that these systems (or the universe) are in definite microstates. This paper describes a theory of macroscopic initial conditions that is incompatible with this assumption. Standard accounts of classical and quantum statistical mechanics characterize the initial conditions that define macroscopic systems by probability distributions of microstates, which they customarily interpret as representing incomplete descriptions. In the theory described below, the initial conditions that define a macroscopic system are shaped by the system's history and represent objective indeterminacy. This interpretation is a consequence of a simple but far-reaching cosmological assumption supported by astronomical evidence and by methodological and theoretical arguments. It entails a picture of the physical universe in which chance prevails in the macroscopic domain (and hence in the world of experience). Because chance plays a key role in the production of genetic variation and in natural selection itself, evolutionary biologists have long advocated such a picture. Chance also plays a key role in other biological processes, including the immune response and visual perception. I argue that reflective choice and deliberation, like these processes and evolution itself, is a creative process mediated by indeterminate macroscopic processes, and that through our choices we help to shape the future.

1 This unpublished manuscript was written in 2010.

I. The presumption of determinism

The proposition that physical laws and antecedent conditions determine the outcomes of all physical processes (other than quantum measurements) is widely regarded as the cornerstone of a naturalistic worldview. Defenders of libertarian free will[2] who accept this proposition must therefore choose between two options:

(1) They can argue (against prevailing opinion among neurobiologists) that some brain processes involved in choice and decision-making are, in effect, quantum measurements – that is, that they involve interactions between a microscopic system initially in a definite quantum state and a macroscopic system that registers some definite value of a particular physical quantity.

(2) They can argue that our current physical laws – specifically, the laws of quantum mechanics – need to be revised. For example, the physicist Eugene Wigner[3] has argued that Schrödinger's equation must be modified to account for certain conscious processes (perceiving the outcome of a quantum measurement).

This paper explores a third possibility: that the presumption of determinism is false.

The laws that govern macroscopic processes and the initial conditions that define the systems to which these laws apply have a statistical character. They involve probability distributions of microstates governed by the laws of quantum mechanics (or classical mechanics when it applies). Physicists have usually assumed that macroscopic systems can be idealized as closed systems in definite microstates. They have justified the use of probability distributions to describe macroscopic states by arguing that it is nearly always impossible to prepare a macroscopic system in a definite microscopic state or to discover through measurements which microscopic state a macroscopic system is actually in. I will argue that the probability distributions that figure in statistical mechanics represent not incomplete knowledge but an objective absence of certain kinds of microscopic information. The argument rests on a pair of

2 For a modern defense of libertarian free will and a critical discussion of the issues that separate libertarian free will from compatibilist accounts, see Robert Kane, *The Significance of Free Will,* (Oxford, Oxford University Press, New York, 1996)

3 Wigner, E. P. "Two Kinds of Reality" in *Symmetries and Reflections*, (Cambridge, MIT Press, 1967)

cosmological assumptions: An exhaustive description of the state of the cosmic medium shortly after the start of the cosmic expansion (a) does not privilege any position or direction in space and (b) contains little or no (statistical) information.

Within physics proper, these assumptions draw support from two main arguments. They contribute to an understanding of quantum measurement; and they serve to ground the statistical assumptions that underlie derivations of the statistical laws that govern irreversible macroscopic processes. The picture of the physical world that emerges from these cosmological assumptions is more hospitable than the usual picture to the worldview of biology. For example, it is consistent with the view that chance events, such as the meteor impact that caused the dinosaurs to die out, have played a prominent role in the history of life, and the view that key processes in evolution, the immune response, and perception have objectively indeterminate outcomes. I will argue that some of the brain processes that mediate deliberation and choice likewise have indeterminate outcomes.

II. Making room for indeterminism

How can a picture of the physical world that incorporates existing physical laws fail to predict that the outcome of every physical process (quantum measurements excepted) is predictable in principle? Before addressing this question, I need to review the standard case for determinism.

II.A. A closer look at determinism

By "determinism" I mean the thesis that physical laws connect a fully described initial state of a closed system to each of the system's subsequent states.[4] Our current physical laws are domain-specific: different physical laws govern the microscopic, macroscopic, and astronomical/cosmological domains. In addition, the conditions under which we can regard a system as effectively closed are not the

4 A closed system is one that in no way interacts with the outside world. All physical systems (except the universe) interact to some extent with the outside world, but many can be idealized as closed for specific purposes. For example, celestial mechanics treats the solar system as closed: it ignores gravitational forces exerted by stars other than the Sun.

same in the three domains. So we need to consider the domains separately.

II.A.1. The microscopic domain

This is the domain of quantum mechanics. Quantum laws are deterministic: the present state of a closed quantum system determines its future state. This assertion may seem to contradict the fact that quantum measurements have unpredictable outcomes, but it does not. In a quantum measurement, a quantum system (for example, an atom) interacts with a (necessarily) macroscopic measuring device. So quantum measurement has one foot in the quantum domain and one in the macroscopic domain. The quantum measurement problem consists in formulating an account of quantum measurement that bridges the two domains.

II.A.2. The macroscopic domain

This is the domain populated by middle-sized objects, the objects of ordinary experience.

Macroscopic theories include classical mechanics, the physics of fluids, electromagnetic theory, the theory of heat and its transformations, and theories governing irreversible transport processes such as heat flow and diffusion.

It is a familiar fact that *knowable* macroscopic initial conditions do not always determine a macroscopic system's future states. For example, the science of dynamical meteorology does not allow one to make reliable long-range weather forecasts. To take a simpler example, cited by Poincaré[5] in a famous essay on chance I will return to later, we cannot predict the direction in which a cone initially balanced on its tip will topple. The cone's initial state is one of unstable equilibrium. How such states evolve depends sensitively on the initial conditions. Poincaré discovered a class of *dynamical* processes in celestial mechanics that likewise depend sensitively on initial conditions. This class includes chaotic (as opposed to regular) orbits in systems of mutually gravitating particles. Two regular orbits in

5 Henri Poincaré, "Le Hasard" in *Science et Méthode* (Flammarion, Paris,1908). Reprint: Paris: Éditions Kimé. English translation: James R. Newman, ed., *The World of Mathematics*, (New York, Simon and Schuster, 1956) Page references are to the English translation.

closely similar initial states diverge slowly; two chaotic orbits in closely similar initial states diverge at an exponential rate.

Poincaré assumed that the cone and the particle in its chaotic orbit were initially in definite (classical) microstate, governed by deterministic laws, even if we do not or cannot know what these states are. Since the microscopic laws of motion (for Poincaré, Newton's laws) are deterministic, he concluded that determinism prevails in the macroscopic, as well as the microscopic, domain. Thus in Poincaré's picture of the world, chance was an intersubjective rather than a truly objective phenomenon. The replacement of classical descriptions and laws by quantum descriptions and laws left this conclusion unchanged.

II.A.3. The astronomical/cosmological domain.

This is the domain of Einstein's theory of gravitation, general relativity, of which Newton's theory is a limiting case. General relativity is a deterministic theory. Its field equation link classically described states of the universe and of effectively closed systems (stars and galaxies, for example) at different times. But it presupposes theories that describe the physical state of the cosmic medium and physical conditions in stars and galaxies; thus it presupposes the theories that prevail in the macroscopic domain. These theories in turn presuppose theories of the microscopic processes that underlie macroscopic processes. For example, the Maxwell-Boltzmann theory of gases assumes that a gas is composed of particles whose motions and interactions are governed by Newton's laws of motion.

Physicists customarily assume that macroscopic laws not only depend on the deterministic laws governing microscopic processes but also that they are reducible to these. That, in a nutshell, is the scientific case for determinism. In each domain physical laws and appropriate initial conditions uniquely determine the future states of closed systems. It depends crucially on the assumption that the statistical laws of macrophysics are reducible to the deterministic laws that govern closed systems in definite microstates.

II.B. Breaking the grip of determinism

II.B.1. The strong cosmological principle

I will argue that indeterminacy enters physics not through the measurement postulate of quantum mechanics but via a cosmological initial condition I call the strong cosmological principle. It has two parts:

a) There exists a system of spacetime coordinates relative to which an exact statistical description of the universe does not privilege any position or direction in space.

b) This description is exhaustive.

The assumption that the average properties of the universe do not determine a preferred position or direction in space is usually called the cosmological principle. It is conventionally viewed as an approximate simplifying assumption that defines one of many possible models of the universe. Part a) asserts that, contrary to this interpretation, the cosmological principle is an exact property of the universe.

Part b) needs a word of explanation. Theoretical physicists take it for granted that our current physical laws are *incomplete*. Quantum mechanics, broadly interpreted to include particle physics and quantum field theory, is manifestly in need of unification; and theorists have been engaged for decades in efforts to unify quantum mechanics and general relativity. In addition, there are phenomena whose explanations will not (theorists believe) require new or modified laws but for which adequate theories do not yet exist. A standard example is the phenomenon of turbulence in classical fluid dynamics. Another example is quantum measurement. In these respects our current description of the universe is incomplete.

But that description involves more than laws. It also involves auxiliary conditions – initial and boundary conditions – that physicists need to specify in order to apply physical laws to specific phenomena. Auxiliary conditions serve to define physical systems, from atoms and subatomic particles to the astronomical universe. The specification of initial and boundary conditions is the part of the description to which the adjective exhaustive in part b) applies. I intend it to imply that when you have specified all possible spatial averages of relevant physical quantities (or enough of them to allow you to calculate the rest), you are done; nothing remains to be specified.

A statistical description that satisfies part a) cannot be exhaustive if the universe is finite. For example, a finite universe would have a largest galaxy, whose mass would be part of an exhaustive description and whose center of mass would be a privileged position.

Part b) would also be false if a deterministic classical theory, rather than quantum physics, prevailed in the microscopic domain, even in an infinite universe. For suppose the particles were randomly distributed. An exhaustive description would specify the distance between the centers of mass of each particle and its nearest neighbor. With probability 1, no two of these distances would be equal, because the number of particle pairs is countably infinite while the set of positive real numbers is uncountably infinite. So a description that merely specified statistical properties of the distribution could not be exhaustive. By contrast, a statistical quantum description of an infinite medium that conforms to the strong cosmological principle *could* be exhaustive.

If the strong cosmological principle holds at a single instant, it holds at all times, because our present fundamental laws are invariant under spatial translation and rotation; that is, they treat all positions in space and all directions at a point in space impartially. The strong cosmological principle extends this kind of invariance to the auxiliary conditions that define the universe. This extension, though unorthodox, is consistent with the abundant observational evidence that bears on it – measurements of the brightness of the cosmic microwave background and of the spatial distribution and line-of-sight velocities of galaxies.

It is also supported by a methodological argument: The initial conditions that define the universe don't have the same function as those that define models of stars and galaxies. We need a theory of stellar structure to apply to a wide range of stellar models because stars have a wide range of masses, chemical compositions, angular momenta, initial magnetic fields, and ages. But there is only one universe.

Finally, the strong cosmological principle contributes to the solution of a purely scientific puzzle that goes back to Newton. Experiment shows that local inertial coordinate systems (coordinate systems relative to which free particles are unaccelerated) are themselves unaccelerated relative to the coordinate system defined by observations of the cosmic microwave background and the spatial distribution and line-of-sight velocities of galaxies. Newton conjectured that the "fixed stars" are uniformly distributed throughout infinite Euclidean space and that they, along with a particular inertial reference frame, are at rest relative to "absolute space." But Newton's own theory of gravitation does not throw light on this coincidence. Einstein's theory of gravitation does. It predicts that local inertial reference frames are unaccelerated relative to the (essentially unique) coordinate system defined by the cosmological principle.

II.B.2. Cosmology and probability

An exhaustive description of an infinite universe that satisfies the strong cosmological principle has a statistical character. It cannot specify, say, the value of the mass density at a particular point at some moment, because such a specification would privilege that point; even in an infinite universe the probability of finding another point at which the mass density had exactly the same value would be zero. Instead, an exhaustive description specifies (among other things) the probability that at any given point the value of the mass density lies in a given range.

This probability has an exact physical interpretation: It is the spatial average of a quantity that takes the value 1 at points where the mass density lies in the given range and takes the value 0 elsewhere. That is, it is the fractional volume (or, more generally, normed measure) of the set of points where the mass density lies in the given range.

With this interpretation of probability, Kolmogorov's axioms of probability theory become true statements about physical space, and probability itself acquires a physical interpretation: The indeterminacy inherent in probabilistic statements reflects the indeterminacy of position in a description of the universe that satisfies the strong cosmological principle.

II.B.3. Randomness and order

Two related properties of probability distribution – statistical entropy and information –play important roles in the following discussion. The statistical entropy of a discrete probability distribution $\{p_i\}$ is the negative mean of the logarithm of the probability $(-\Sigma\, p_i \log p_i)$. As Boltzmann and Gibbs showed, it is the statistical counterpart of thermodynamic entropy. Shannon emphasized that it is a measure of the randomness of discrete probability distributions in other scientific contexts and made it the centerpiece of his mathematical theory of information.

Information is a measure of order. Different authors define it differently. I define it here as the amount by which a probability distribution's randomness falls short of the largest value that is consistent with appropriate constraints. If this largest admissible value is fixed, an increase in the statistical entropy of the probability distribution that characterizes a physical state is accompanied by an equal decrease in its information. But we will encounter situations in which the largest admissible value of the statistical entropy is not fixed and in which statistical entropy and information both increase.

Statistical mechanics characterizes (macroscopic) states of thermodynamic equilibrium by maximally random, or information-free, probability distributions. The probability distributions that characterize non-equilibrium states contain information. This can take a number of qualitatively different forms. For example, in a closed gas sample, the relative concentrations of chemically reactive constituents may deviate from their equilibrium values; the temperature may not have the same value at every point in the sample; the fluid velocity may not vanish everywhere; vortices may be present; and so on.

As discussed below, I take randomness to be the primordial condition of the cosmic medium. The cosmic expansion and, later, other physical processes create information. But in general the probability distributions of microstates that characterize macrostates of macroscopic systems contain only a tiny fraction of the information needed to specify a definite microstate of the system; they assign comparable probabilities to an enormous number of microstates. Thus macroscopic systems cannot in general be assigned definite microstates. Because microstates evolve deterministically, this condition is necessary for macroscopic indeterminacy.

III. Growth and decay of order

Current theories of the early universe posit that shortly after the beginning of the cosmic expansion, the cosmic medium consisted of uniformly distributed elementary particles whose relative concentrations corresponded at each moment to thermodynamic equilibrium at the temperature and density prevailing at that moment.[6] The average properties of the cosmic medium during this early period depended on the values of physical constants that figure in theories of elementary particles and on the parameters needed to define a general-relativistic model of a uniform, unbounded medium. Density fluctuations, with a spectrum that depends on a small number of additional parameters, may also have been present. But there is no evidence (to my knowledge) to suggest that additional *non-statistical* information was needed to fully describe the state of the cosmic medium during this period.

But if the early universe was information-free, how did the information-rich structures that form the subject matter of astronomy, geology, biology, and human history come into being?

III.A. Cosmic evolution and the Second Law

Some physicists have held that a scenario in which information comes into being is impossible because it would contravene the second law of thermodynamics. Rudolf Clausius, who introduced the concept of entropy and formulated its law, expressed the law in these terms: "The entropy of the universe goes toward (*strebt nach*) a maximum." This has often been interpreted to mean that universe must have begun in a highly complex and orderly state and is becoming increasingly uniform and undifferentiated. I think this interpretation is mistaken, for two reasons.

First, in an expanding universe, entropy production is not necessarily accompanied by a loss of information. An important example, discussed below, is the genesis of chemical information – information associated with the far-from-equilibrium relative abundances of the chemical elements.

---

6 Weinberg, S. *Gravitation and Cosmology*, (New York, John Wiley and Sons, 1972) Chapter 15

Second, I think it is questionable whether the notion of entropy applies beyond the macroscopic domain in which it was originally defined and in which it has been exhaustively tested and confirmed. In particular, the following argument suggests to me that it doesn't apply to self-gravitating systems.

In classical thermodynamics, entropy is an additive[7] property of extended systems in thermodynamic equilibrium. The definition of entropy extends naturally to systems in local, but not global, thermodynamic equilibrium and to composite systems whose components are individually in global or local thermodynamic equilibrium. It also applies to systems in fixed external gravitational, electric, or magnetic fields. The law of entropy growth predicts, and experiments confirm, that as a closed system that meets these criteria evolves, its entropy increases. Experiments also confirm that eventually a closed system settles into an unchanging state, called thermodynamic equilibrium, in which its entropy is as large as possible. This state is uniquely determined by the values of two of the system's macroscopic properties, such as temperature and volume, and by the values of quantities that are conserved by particle reactions at the given temperature and volume. Can the notions of entropy and its law, so defined, be extended to a self-gravitating gas cloud?

We can regard a sufficiently small region of such a cloud as a gas sample in an external gravitational field. We can therefore calculate its entropy, and conclude from the Second Law that physical processes occurring within the region generate entropy. If the cloud's large-scale structure is not changing too rapidly, we can then conclude that local thermodynamic equilibrium prevails. We can even *define* the entropy of the cloud as the sum of the entropies of its parts.

But because gravity is a long-range force, the cloud as a whole does not behave like a closed gas sample in the laboratory. Although heat flow, mediated by collisions between gas molecules, tends to reduce temperature differences between adjacent regions, the cloud does not relax into a state of uniform temperature.[8] Instead, it relaxes

7 The value of an additive property of an extended system is the sum of its values for the weakly interacting parts into which we can imagine the system to be divided. Internal energy and entropy are additive in this sense; temperature is not.

8 For a cloud of finite mass, an equilibrium state of uniform temperature does not exist.

– through fluid motions driven by interplay between gravitational forces and pressure gradients – toward a state of *dynamical* equilibrium (or quasi-equilibrium). If the cloud has zero angular momentum, the quasi-equilibrium state has spherical symmetry. The mass density and temperature are constant on concentric spheres and increase toward the center of the cloud. But unlike any macroscopic system to which the Second Law applies, a self-gravitating gas cloud has negative heat capacity: adding heat to the cloud causes its mean temperature to *decrease*; withdrawing heat causes its mean temperature to increase. So if the cloud can radiate, it grows hotter as it cools.

In such a gas cloud, microscopic processes convert kinetic energy associated with particle motions into radiation, which makes its way toward the surface of the cloud and eventually escapes into space. The loss of energy causes the cloud to contract. As a result, its mean temperature increases and the temperature gradient between the center and the surface becomes steeper. Eventually the core of the cloud becomes hot enough to ignite nuclear reactions that burn hydrogen into helium. These energy-releasing (and entropy-producing) reactions drive the core toward chemical equilibrium, simultaneously creating new deviations from thermodynamic equilibrium in the form of a composition gradient between the core and its envelope. When the core temperature becomes so high that the rate at which hydrogen-burning nuclear reactions in the core release energy balances the rate at which energy is being lost at the surface, the contraction stops and a slowly changing, far-from-equilibrium state ensues. The Sun is now in such a state.

Local entropy-generating processes play an essential role in the process just sketched. Heat flow and radiation tends to level temperature gradients, molecular diffusion tends to level concentration gradients, internal friction (viscosity) tends to level fluid-velocity gradients. All these processes generate entropy. But the mutual gravitational attraction of its parts continually steepens temperature and concentration gradients within the cloud. While Newton's theory of gravitation allows one to extend the notion of gravitational potential energy, initially defined for an object in the Earth's gravitational field, to a self-gravitating cloud, no such extension has, to my knowledge, been shown to be possible for the notion of entropy.

III.B. Chemical order

The history of life begins with the first self-replicating molecules or molecular assemblies, but its prehistory begins with the formation of hydrogen in the early universe. The burning of hydrogen into helium in the core of the Sun releases energy at the same rate as energy is lost in the form of sunlight (and neutrinos). Solar photons drive metabolic, biosynthetic, and reproductive processes in plants and algae, which store energy in information-rich molecular forms that animals use to fuel the chemical processes that mediate their basic biological functions. Life as we know it exists because hydrogen is so plentiful. It is plentiful because only about a quarter of it was consumed early on in reactions like those that keep the Sun and other stars of similar mass shining.

In the very early universe, reactions between elementary particles were so fast that they were able to maintain near-equilibrium relative particle abundances corresponding to the instantaneous values of the rapidly decreasing temperature and mass density. As the universe continued to expand, its rate of expansion decreased, and so did the rates of the equilibrium-maintaining particle reactions; but the latter decreased faster than the former. Eventually, the reaction rates became too slow to maintain the relative concentrations of particle kinds near their equilibrium values. Relative concentrations of particle kinds close to those that prevailed during that era were frozen in. Meanwhile, the universe continued to expand, its temperature and mass density continued to fall, and the gap between the actual statistical entropy per unit mass of the cosmic medium and the value appropriate to thermodynamic equilibrium widened. This gap represents chemical information.

Like the gradients of temperature and mass density in stars, chemical information was created in the early universe by a competition between processes on vastly different scales, governed by different physical laws. Entropy-generating particle reactions, governed by the laws of quantum mechanics, drove the chemical composition toward equilibrium; the cosmic expansion, governed by the laws of Einstein's theory of gravitation, drove the chemical composition of the cosmic medium away from equilibrium.

Some of the chemical information created in the early universe is now being converted into biological information. Because the Sun and other stars of similar mass formed from hydrogen-rich material, they are long-lived sources of light in the visible, infrared, and ultraviolet regions of the electromagnetic spectrum. The energies of photons emitted by the Sun reflect the temperature of the light-emitting layers – roughly 6,000 degrees on the Kelvin scale. The energies of individual photons that reach the Earth are therefore much greater than those of the ambient photons, whose energies correspond to temperatures around 300 degrees Kelvin. This disparity makes it possible for solar photons to drive the chemical reactions that maintain the cells of photosynthetic organisms in far-from-equilibrium states.

Every link in the chain of processes that connects the early universe to the creation of information in human brains (including several I haven't mentioned[9]) is now understood in some depth and detail. The only novel feature of the preceding account is the claim that information is an objective component of the world that physical theories describe.

IV. Macroscopic systems and history

Molecules, atoms, and subatomic particles do not have individual histories. Two hydrogen atoms may be in different quantum states, but otherwise they are indistinguishable.[10] By contrast, objects of ordinary experience are individuals, each with its own history. I will argue that the statistical laws that govern macroscopic processes likewise have a historical character. Unlike the ahistorical laws of quantum mechanics and general relativity, they depend on the

9 For example, the provenance of elements such as zinc that play an essential metabolic role. Heavy elements form in the cores of stars much more massive than the Sun. These short-lived stars explode, thereby "enriching" the interstellar medium within which the Sun evolved from a proto-star.

10 Particles of the same kind are indistinguishable in a stronger sense that defies visualization: the quantum state of a collection of particles of the same kind – electrons, say, or photons – must be described in mathematical language that does not allow one to distinguish between members of the collection. For example, we may say that one of a pair of electrons is in state *a* and the other is in state *b* but we may not say that electron 1 is in state *a* and electron 2 in state *b*. A wave function that assigns electron 1 to state *a* and electron 2 to state *b* does not represent a possible two-electron state.

initial and boundary conditions that define the systems they apply to; and these conditions are expressed by probability distributions descended from the probability distributions that characterize the early cosmic medium.

Boltzmann's *transport equation* illustrates these claims. Boltzmann represented the state of a closed sample of an ideal gas by a probability distribution of the (classical) microstates of a single molecule in the sample, a microstate being defined by a molecule's position and velocity. His transport equation relates the rate at which this probability distribution changes with time to the rates at which molecular collisions throughout the sample populate and depopulate single-molecule states.

From the transport equation Boltzmann inferred that a quantity he called H – the negative of the single-molecule probability distribution's statistical entropy – decreases with time until it reaches its smallest admissible value. Thereafter it doesn't change. The probability distribution that characterizes this state of *statistical equilibrium* has a unique form (derived earlier by Maxwell) that depends on two parameters: the number of molecules per unit volume and the statistical counterpart of absolute temperature.

Maxwell and Boltzmann identified[11] statistical equilibrium with thermodynamic equilibrium. Boltzmann showed that the quantity $-H$ is the statistical counterpart of thermodynamic entropy. Thus the H theorem is the statistical counterpart, for an ideal gas, of the Second Law.

Boltzmann's transport equation is the starting point for derivations of laws governing specific irreversible processes in gases, such as heat flow and molecular diffusion. These derivations yield formulas that relate the coefficients that appear in these laws (thermal conductivity and diffusivity) to properties of the gas molecules and the laws that govern their interaction. This enables physicists to deduce the values of parameters that characterize the structure and

11 Classical thermodynamics rests on the empirical rule that a nominally isolated macroscopic system or a system in thermal contact with a heat reservoir at a fixed temperature eventually settles into a state in which all measurable properties of the system have unchanging values. This state is called thermodynamic equilibrium. Because the definition of entropy relies on the notion of thermodynamic equilibrium, the second law of thermodynamics cannot be used to *explain* the empirical rule.

interaction of gas molecules from macroscopic measurements. The internal consistency of such efforts leaves little room for doubt that Boltzmann's theory is substantially correct.

Boltzmann's theory has also served as a model for theories of irreversible processes in other macroscopic systems, such as dense gases, liquids, solids, and plasmas; and these theories, too, enjoy strong experimental support.

Yet for all their practical success, Boltzmann's theory and other statistical theories of irreversible processes raise foundational issues that are still debated today.

IV.A. Objectivity versus intersubjectivity

The second law of thermodynamics and the laws that govern irreversible processes are widely held to be observer-independent features of the external world. Statistical theories link these macroscopic laws to microphysics. Yet these theories characterize macrostates by probability distributions that are standardly interpreted as reflecting an observer's incomplete knowledge of the objective physical situation (in which every closed macroscopic system is in a definite microstate).

By contrast, an account of cosmic evolution that incorporates the strong cosmological principle makes information and statistical entropy objective properties of the physical world, created and destroyed by historical processes linked to conditions that prevailed in the early universe. History determines what kinds of information are (objectively) present in the probability distributions that characterize initial macrostates and what kinds are objectively absent. By situating statistical theories of irreversible processes within a historical account of initial and boundary conditions, we eliminate the taint of subjectivity. The problem of justifying assumptions about initial and boundary conditions for particular macroscopic systems and classes of macroscopic systems remains, of course, but it is now a soluble scientific problem.

IV.B. Entropy, Boltzmann entropy, and Gibbs entropy

Boltzmann's theory applies to ideal gases; Gibbs's statistical mechanic applies not only to samples of an ideal gas but to any closed system of N particles governed by the laws of classical mechanics. Its quantum counterpart, quantum statistical mechanics, preserves the overall structure of Gibbs's theory and its main theorems.

Like Maxwell and Boltzmann, Gibbs identified thermodynamic equilibrium with statistical equilibrium. Boltzmann's theory reproduces the classical thermodynamic theory of an ideal gas, with the statistical entropy of the probability distribution of a single molecule's microstates in the role of thermodynamic entropy and a parameter that characterizes the maximum-statistical-entropy probability distribution in the role of absolute temperature. Gibbs's theory reproduces the whole of classical thermodynamics, with the statistical entropy of the probability distribution of a closed macroscopic system's microstates in the role of thermodynamic entropy and a parameter that characterizes the maximum-statistical-entropy probability distribution in the role of absolute temperature. So Boltzmann's theory may appear at first sight to be a limiting case of Gibbs's. But Boltzmann proved that the statistical entropy of the single-molecule probability distribution increases with time (unless it has already reached its largest admissible value), while Gibbs proved that the statistical entropy of the probability distribution of the microstates of the sample as a whole is constant in time.

The resolution of this apparent contradiction is unproblematic. It hinges on a mathematical property of statistical entropy. The statistical entropy of an N-particle probability distribution can be expressed as the sum of two contributions. The first contribution is N times the statistical entropy of the single-particle distribution. The second contribution is associated with statistical correlations between molecules of the gas sample. The constancy of the N-particle statistical entropy is consistent with the growth of the single-particle contribution. Taken together, Boltzmann's H theorem and Gibbs's proof that the statistical entropy of the N-particle probability distribution is constant in time imply that the second contribution – the contribution associated with intermolecular correlations – decreases at a rate that exactly compensates the growth of the first contribution. In terms of information: the decline of single-particle information in a closed gas sample is matched by the growth of correlation information.

Although Boltzmann correctly identified the thermodynamic entropy of a closed gas sample with the single-particle statistical entropy,[12] his derivation of the *H* theorem – the statistical counterpart of the Second Law as applied to an ideal-gas sample – had a technical flaw. The derivation rests on an assumption (known as the *Stosszahlansatz*[13]) that cannot in fact hold for a closed gas sample. A stronger form of this assumption states that correlation information – the amount by which the information of the single-molecule probability distribution, multiplied by *N*, falls short of the information of the *N*-molecule probability distribution – is *permanently absent*. As we've just seen, this assumption cannot be true, because the decay of single-molecule information creates correlation information at the same rate. So even if correlation information is initially absent, it cannot be permanently absent.

The persistence of correlation information in a closed system poses a threat to the very notion of thermodynamic equilibrium. Poincaré proved that a closed, initially non-uniform dynamical system cannot relax into a permanent state of macroscopic uniformity; after an initial period of relaxation, it must eventually return to a non-uniform state resembling its initial state. For example, if the left half of a gas sample is initially at a higher temperature than the right half, heat flow equalizes the temperature throughout the sample; but eventually an uneven temperature distribution resembling the initial distribution reasserts itself. The "recurrence time" – the precise meaning of "eventually" – depends on how closely the revenant distribution resembles the original. The closer the resemblance, the longer the recurrence time. In any case, though, it greatly exceeds the age of the universe.

---

12 The single-particle probability distribution is, in Gibbs's phrase, a "coarse-grained" version of the N-particle probability distribution. But the spin-echo experiment, discussed below, shows that in some experimental situations the coarse-grained entropy of a nominally closed system *decreases* with time.

13 The *Stosszahlansatz*, introduced by Maxwell, says the incoming velocities of molecules that are about to collide are statistically uncorrelated; their joint probability distribution is the product of the individual (marginal) probability distributions.

Although the assumption that correlation information is permanently absent cannot be true, some authors have argued that the *Stosszahlansatz* may nevertheless hold to a good approximation for periods that are long by human standards but short compared with the Poincaré recurrence time.

Another approach to the problem of initial conditions starts from the remark that no gas sample is an island unto itself. Can the fact that actual gas samples interact with their surroundings justify the assumption that correlation information is permanently absent in a nominally closed gas sample? And if so, is it legitimate to appeal to environmental "intervention"?

J. M. Blatt[14] has argued that the answer to both questions is yes. The walls that contain a gas sample are neither perfectly reflecting nor perfectly stationary. When a gas molecule collides with a wall, its direction and its velocity acquire tiny random contributions. These leave the single-particle probability distribution virtually unaltered, but they alter the histories of individual molecules, thereby disrupting multi-particle correlations.[15] Blatt distinguished between true equilibrium (in the present terminology, sample states characterized by an information-free probability distribution) and quasi-equilibrium, in which single-particle information is absent but correlation information is present. He argued, with the help of a simple thought-experiment, that collisions between molecules of a rarefied gas sample and the walls of its container cause an initial quasi-equilibrium state to relax into true equilibrium long before the gas has come into thermal equilibrium with the walls. Walls impede the flow of energy much more effectively than they impede the outward flow of correlation information.

Bergmann and Lebowitz[16] have constructed and investigated detailed mathematical models of the relaxation from quasi-equilibrium to true equilibrium through external "intervention." More recently, Ridderbos and Redhead[17] have expanded Blatt's case for the interventionist approach. They construct a simple model

14 J. M. Blatt, *Prog. Theor. Phys*. 22, 1959, p. 745

15 Thus in applying Poincaré's theorem, we must include the walls and their surroundings in the closed dynamical system to which the theorem refers; and this step nullifies the theorem's conclusion: "eventually" becomes "never."

16 P. J. Bergmann and J. L. Lebowitz, *Phys. Rev*. 99, 1955, p. 578; J. L. Lebowitz and P. J. Bergmann, *Annals of Physics* 1, 1959, p. 1

17 T. M. Ridderbos and M. L. G. Redhead, *Found. Phys*. 28, 1988, p. 1237

of E. L. Hahn's spin-echo experiment,[18] in which a macroscopically disordered state evolves into a macroscopically ordered state. They argue that in this experiment, and more generally, interaction between a nominally isolated macroscopic system and its environment mediates the loss of correlation information by the system.

Blatt noted that the interventionist" approach "has not found general acceptance."

> There is a common feeling that it should not be necessary to introduce the wall of the system in so explicit a fashion. ... Furthermore, it is considered unacceptable philosophically, and somewhat "unsporting", to introduce an explicit source of randomness and stochastic behavior directly into the basic equations. Statistical mechanics is felt to be a part of mechanics, and as such one should be able to start from purely causal behavior.[19]

The historical account I have been sketching supplies a hospitable context for the interventionist approach. Macroscopic systems are defined by their histories, which determine the kinds and quantities of information in the probability distributions that (objectively) characterize their initial states. A historical account also justifies the necessary assumption that the environments of macroscopic systems are usually characterized by maximally random, or information-free, probability distributions of microstates.

IV.C. The origin of irreversibility

The laws of statistical macrophysics discriminate between the direction of the past and the direction of the future; the laws that govern the microscopic processes that underlie macroscopic processes do not. What is the origin of irreversibility, and how can it be reconciled with the reversibility of microscopic processes?

The conflict between microscopic reversibility and macroscopic reversibility is rooted in the idealization of perfect isolation. This idealization works for atoms, molecules, and subatomic particles, but it fails for macroscopic systems, for two related reasons. It implies that macroscopic systems are in definite microstates, and it detaches the histories of individual physical systems from the history of the universe.

18 Hahn, E. L. *Phys. Rev.* 80, 1950, p. 580

19 Blatt, 1959,. p. 747

In the account I have been sketching, historical processes supply the initial conditions that define physical systems and their environments. These initial conditions are expressed by probability distributions, which a historical narrative links to the probability distributions that objectively characterize the early universe. This narrative supplies the directionality (the arrow of time) that characterizes the laws of statistical macrophysics but not their underlying microscopic laws. The probability distributions that characterize the initial states of physical systems and their environments also help to shape the macroscopic laws that govern these systems (the derivation of Boltzmann's transport equation depends on statistical assumptions about the gas samples to which the equation applies). For this reason, statistical macroscopic laws lack the universality of microscopic laws. They apply to particular classes of macroscopic systems under broad but limited ranges of physical conditions.

V. Indeterminism

The case for macroscopic determinism rests on the widely held view that macroscopic descriptions are just microscopic descriptions seen through a veil of ignorance. On this view, macroscopic systems are actually in definite microstates, but we lack the resources to describe them at that level of detail. I have argued that the strong cosmological principle entails a different view of the relation between macrophysics and microphysics. Historical processes link the probability distributions that characterize macroscopic systems in statistical theories like those of Boltzmann and Gibbs to a statistical description of the early universe. The probability distributions that figure in that description represent an objective absence of information about the microstructure of the early cosmic medium. We cannot assign macroscopic systems definite microstates (except at temperatures close to absolute zero) not because we lack the resources to determine them but because an exhaustive description of the universe that comports with the strong cosmological principle does not contain that kind of microscopic information.

Once we give up the idea that a macroscopic description of the world is just a blurred version of an underlying microscopic description, we can no longer infer macroscopic determinism from microscopic determinism. Instead we must consider specific macroscopic processes whose outcomes might be unpredictable in principle. The most promising candidates are processes whose outcomes scientists conventionally attribute to chance (while maintaining that objectively, chance doesn't exist).

V.A. Poincaré's account of chance

In the essay cited in section I, Poincaré divided chance events into three groups: those whose causal histories are extremely sensitive to small changes in the initial conditions; those whose causes are "small and numerous"; and events that lie at the intersection of unrelated causal histories (coincidences).

The first group is exemplified by a perfectly symmetric cone balanced on its tip. The slightest disturbance would cause it to topple over. If we had a complete description of the cone's environment, we could predict the direction in which it would fall. Since we don't have such a description, we represent the initial disturbance by a probability distribution. The classical laws of motion map this probability distribution onto a probability distribution of the possible orientations of the fallen cone's axis. In the context of Poincaré's deterministic picture of the physical world ("We have become absolute determinists, and even those who want to reserve the rights of human free will let determinism reign undividedly in the inorganic world at least."[20]), both the initial and final probability distributions represent incomplete states of knowledge.

An account of the same experiment that incorporates the view of initial conditions described in the last section leads to a different conclusion. Each of the cone's classical microstates is defined by the angular coordinates of its axis and their rates of change. We can represent such a state by a point in an abstract four-dimensional state space. Now, no experimental technique can produce a state represented by a single point in a continuum, even in principle. To represent the state of a cone initially balanced on its tip we need to use a probability distribution of classical microstates whose representative points are smoothly distributed within some neighborhood containing the point that represents the unstable equilibrium state. No matter how small that neighborhood may be, the classical laws of motion map the probability distribution of initial states onto a prob-

20 Poincaré, 1908, p. 1380

ability distribution of final states that assigns a finite probability to every finite range of final orientations. Because the initial probability distribution represents objective, historically determined indeterminacy (rather than incomplete knowledge), the cone's final orientation is unpredictable in principle. The probability distribution that characterizes the cone's initial macrostate evolves deterministically into a probability distribution that objectively characterizes a continuous range of distinct macrostates. Which of these macrostates actually occurs is a matter of pure, objective chance.

The cone's evolution is extremely sensitive to small changes in the initial conditions. Extreme sensitivity to initial conditions is the defining feature of what is now called deterministic chaos. Two decades before his essay on chance appeared, Poincaré had discovered that the orbits of particles in a system of mutually gravitating particles, such as the solar system, are of two kinds: "regular" orbits, for which a small change in the initial conditions produces a new orbit that deviates slowly from the original one; and "chaotic" orbits, for which a small change in the initial conditions produces a new orbit that deviates from the original orbit at an initially exponential rate. The mathematical literature now contains many other examples of dynamical systems that evolve deterministically but, from a practical standpoint, unpredictably.

Weather systems are a particularly striking example of such systems. In 1961 Edward Lorenz explained why it is impossible to predict tornadoes. He found that numerical solutions to a mathematical model of a weather system were extremely sensitive to round-off errors in the initial conditions. As Lorenz put it later, "The flap of a butterfly's wing can change the weather forever."

Accidental measurement errors exemplify Poincaré's second group of chance phenomena. Scientists assume that errors that cannot be assigned to known causes result from myriad small, unrelated causes of random sign and magnitude. Analogously, the position and velocity of a molecule in a gas sample or of a microscopic particle suspended in a liquid are resultants of innumerable random collisions with ambient molecules. In these and similar examples, the randomness that scientists posit may be genuine, objective randomness. For example, if a gas sample's history determines only its macroscopic properties (chemical composition, temperature, mass), as is generally the case, a complete description of the sample does not assign each molecule a definite position and a definite velocity.

Poincaré's third group of chance phenomena includes life-altering coincidences, like Pip's encounter with the escaped convict in the opening chapter of Dickens' *Great Expectations*. In a deterministic universe every event is a consequence of antecedent conditions. When an event results from the intersection of seemingly unrelated histories, we attribute it to chance because "our weakness forbids our considering the entire universe and makes us cut it up into slices."[21] In a description of the universe that comports with the strong cosmological principle, true coincidences abound – though, of course, not all events we customarily attribute to chance are objectively unpredictable. A separate historical argument is needed in each instance.

V.B. Chance in biology

Unlike other animals, we can imagine, evaluate, and decide between possible courses of action. Other species have evolved traits that help them to survive in particular environments, but we alone have been able to invent ways of making a living virtually anywhere on Earth. And we alone respond not only flexibly but also creatively to challenges posed by a seemingly unpredictable environment. This creative capacity has been shaped by evolution. I will argue that the biological processes on which it rests embody the same strategy as evolution itself.

V.B.1. Evolution

Ernst Mayr has emphasized the role of chance in evolution:[22]

> Evolutionary change in every generation is a two-step process: the production of genetically unique new individuals and the selection of the progenitors of the next generation. The important role of chance at the first step, the production of variability, is universally acknowledged (Mayr 1962[23]), but the second step, natural selection, is on the whole viewed rather deterministically: Selection is

21 Poincaré, 1908, p. 1386

22 Ernst Mayr, "How to carry out the adaptationist program?" in *Toward a New Philosophy of Biology*, Cambridge: Harvard University Press, 1988, p. 159

23 Ernst Mayr, "Accident or design, the paradox of evolution" in G.W. Leeper, ed., *The Evolution of Living Organisms*, Melbourne: Melbourne University Press, 1962

> a non-chance process. What is usually forgotten is the important role chance plays even during the process of selection. In a group of sibs it is by no means necessarily only those with the most superior genotypes that will reproduce. Predators mostly take weak or sick prey individuals but not exclusively, nor do localized natural catastrophes (storms, avalanches, floods) kill only inferior individuals. Every founder population is largely a chance aggregate of individuals, and the outcome of genetic revolutions, initiating new evolutionary departures, may depend on chance constellations of genetic factors. There is a large element of chance in every successful colonization. When multiple pathways toward the acquisition of a new adaptive trait are possible, it is often a matter of a momentary constellation of chance factors as to which one will be taken (Bock 1959).[24]

Mayr's view on the role of chance draws support from a body of observational evidence ranging from molecular biology to ecology. Within this vast domain, chance events are ubiquitous. They not only supply candidates for natural selection. They also determine its direction at crucial junctures.

V.B.2. The immune system

One of evolution's responses to the challenges posed by an unpredictable environment is the immune system. Plants and animals are susceptible to attack by a virtually unlimited variety of microorganisms. In vertebrates the immune system responds by deploying white-blood cells (lymphocytes) each of which makes and displays a single specific kind of antibody – a protein capable of binding to a specific molecular configuration on the surface of another cell. The immune system must be able to deploy a vast array of distinct antibodies – far more than can be encoded in an animal's genome; and all of these antibodies must fail to recognize the animal's own cells. To meet these requirements, the immune system uses a modified form of evolution's cyclic two-step strategy:

Step 1a: the generation of diversity by a random process. During a lymphocyte's development from a stem cell, the gene that encodes the antibody it will eventually display is randomly and imprecisely assembled from several widely separated segments of DNA. Thus different stem cells evolve into lymphocytes that display

24 Bock, W.J. *Evolution* 13, 1959, 194-211

different antibodies. Each stem cell is the progenitor of a clone of lymphocytes that make and display identical antibodies.

Step 2a: selection for failure to recognize an animal's own cells. Immature lymphocytes that bind to an animal's own cells die, leaving behind a population of lymphocytes that is still diverse enough to cope with invasions by foreign cells.

Step 2b: selection by a foreign cell. The binding of a circulating lymphocyte to a foreign cell causes it to proliferate, producing a clone of lymphocytes all of which make, display, and discharge into the blood the same antibody. This step concludes the primary immune response.

Step 1b: renewed generation of diversity. A few members of the clone selected by a particular foreign cell remain in the blood. When the same foreign cell reappears, these "memory cells" mutate very rapidly, producing new, more highly specialized candidates for selection.

Step 2c: renewed selection. Some of these candidates fit the invading cell even better than their precursors. Repeated many times, this cycle constitutes the secondary immune response, which is both quicker and more massive than the primary response.[25]

V.B.3. Visual perception

The visual system represents evolution's solution to a different problem posed by the unpredictability of the environment. We need to be able to distinguish between the faces of family members and strangers. More generally, we need to be able to distinguish between faces and other visual configurations. More generally still, we need to be able to pick out objects from a visual field that is not intrinsically differentiated. In short, we need to be able to convert a pair of retinal images into a three-dimensional scene populated by objects endowed with meaningful attributes. How does the visual system make sense of visual stimuli that lack intrinsic meaning?

The art historian E. H. Gombrich[26] arrived at an answer to this question by an indirect route. He began by asking himself: Why does

25 This is a very incomplete account of the vertebrate immune system. For a more complete account, see any recent graduate-level textbook on molecular and cell biology.

26 Gombrich, E.H. *Art and Illusion*, (Princeton: Princeton University Press 1961) eleventh printing of the second, revised, edition, with a new preface, 2000

representational art have a history? Why do visual artists represent the visible world differently in different periods? Gombrich argued that painters do not – indeed cannot – simply record what they see. They must master a complex technique for producing visual illusions. And perception, he argued, must work the same way. The history of representational art, he wrote, is "a secular experiment in the theory of perception" – an account of how "artists discovered some of [the] secrets of vision by 'making and matching.'"[27]

The "secret of vision" is that what we see is the result of a mental construction. Initially, and at every subsequent stage of this construction, a candidate percept, or schema, is put forward, tested, and altered in the light of the test's outcome. The conscious image is the final, stable schema:

> [T]he very process of perception is based on the same rhythm that we found governing the process of representation: the rhythm of schema and correction. It is a rhythm which presupposes constant activity on our part in making guesses and modifying them in the light of our experience. Whenever this test meets with an obstacle, we abandon the guess and try again ...[28]

The psychologist Ulrich Neisser has given a clear and comprehensive account of the cyclic theory of perception and cognition. [29]More recently, the neuropsychologist Chris Frith contrasts old-fashioned linear accounts of perception with accounts that model perception as a loop, as Gombrich did:

> In a linear version of perception, energy in the form of light or sound waves would strike the senses and these clues about the outside world would somehow be translated and classified by the brain into objects in certain positions in space. ... A brain that uses prediction works in almost the opposite way. When we perceive something, we actually start on the inside: a prior belief [i.e., schema], which is a model of the world in which there are objects in certain positions in space. Using this model, my brain can predict what signals my eyes and ears should be receiving. These predictions are compared with the actual signals and, of course, there will be errors. ... These errors teach my brain to perceive. [They] tell the brain that its model of the world is not good enough. The nature of the errors tells the brain how to make a better model of the world. And so we go round the loop again and again until the

27 ibid. p. 29

28 ibid. pp. 271-2

29 See Ulric Neisser, *Principles and Implications of Cognitive Psychology*, San Francisco: W.H. Freeman, 1976

> errors are too small to worry about. Usually only a few cycles of the loop are sufficient, which might take the brain only 100 milliseconds.[30] ... So, what we actually perceive are our brain's models of the world. ...You could say that our perceptions are fantasies that coincide with reality.[31]

Pre-existing schemata constrain the construction of visual percepts. Some of these are present at birth; we can only see what the architecture of the brain allows us to see. Visual illusions and ambiguous figures illustrate different aspects of our inability to overcome such genetic constraints. Other schemata, like those that enable a skilled painter or caricaturist, to create a convincing representation, are constructed during the artist's lifetime.

The processes that enable us to make sense of aural stimuli (spoken language, music), presumably involve cycles of making and matching, as well as hierarchical schemata analogous to those that underlie the perception of visual stimuli.

V.B.4. Perception and neurobiology

The neural underpinnings of perception are not yet fully understood. Gerald Edelman and his colleagues[32] and Jean-Pierre Changeux and his colleagues[33] have offered models that are similar in many respects but differ in others. Changeux summarizes his view of perception as follows:

> The key postulate of the theory is that the brain spontaneously generates crude, transient representations ... These particular mental objects, or *pre-representations* [Gombrich's and Neisser's "schemata," Frith's "beliefs"], exist *before* the interaction with the outside world. They arise from the recombination of existing sets of neurons or neuronal assemblies, and their diversity is thus great. On the other hand, they are labile and transient. Only a few of them are stored. This storage results from a *selection*! ... Selection

30 Frith, C, *Making Up the Mind*, Oxford: Blackwell, 2007, pp. 126-7

31 ibid. pp.134-5

32 Edelman, G.M. and V.B. Mountcastle, *The Mindful Brain* (Cambridge: MIT Press, 1978); Edelman, G.M., *Neural Darwinism*, (New York: Basic Books, 1987); Edelman, G.M.and Tononi, G., *A Universe of Consciousness: How Matter Becomes Imagination* (New York: Basic Books, 2000); Edelman, G.M. *Second Nature* (New Haven, Yale University Press, 2006) and other references cited there.

33 Changeux, J.P. *The Physiology of Truth: Neuroscience and Human Knowledge* translated from the French by M.B. DeBevoise, (Cambridge: Belknap Press/Harvard, 2004) and references cited there.

> follows from a *test of reality* [which] consists of the comparison of a percept ["perceptual activity aroused by sensory stimulation."[34]] with a pre-representation.[35] (Italics in original.)

Citing Edelman's work as well as an earlier book of his own, Changeux emphasizes that the generation of diversity (as well as selection) plays an essential role in learning:

> Knowledge acquisition is indirect and results from the selection of pre-representations. *Spontaneous activity plays a central role in all this, acting as a Darwinian-style generator of neuronal diversity.*[36] (emphasis added.)

V.B.5. The capacity for reflective choice

All animals learn from experience. They tend to repeat behaviors for which they have been rewarded in the past and to avoid behaviors for which they have been punished. Some animals, though, learn from experience in ways that allow for risk-taking, exploratory behavior, and delayed rewards. Economists, students of animal behavior, and cognitive neuroscientists have developed algorithms that seek to mimic such flexible learning strategies, and have constructed hypothetical neural networks that instantiate algorithms of this kind.[37] For example, learning algorithms usually use a "reward prediction error signal" (the difference between the actual reward and a suitably weighted average of previous rewards) to evaluate candidates for choice. Some investigators have found evidence that the activity of midbrain dopamine neurons functions as a reward prediction error signal.[38]

It may well be that much of human learning and decision-making is mediated by neural architecture that instantiates complex and sophisticated algorithms of this kind. In light of what ethologists have learned about the behavior of monkeys and apes, it would be surprising if this were not the case. But humans seem to have an extra, qualitatively different capacity for learning and decision-mak-

---

34 ibid. p. 6.

35 Changeux, J.-P. *Neuronal Man*, (Princeton: Princeton University Press, 1985) p. 139

36 Changeux, *The Physiology of Truth*, p. 58

37 For recent reviews, see *Nature Neuroscience*, Volume 111, Number 4, April 2008, pp. 387-416.

38 Bayer. H.M. and P.W. Glimcher, *Neuron* 47, Issue 1, 2005, 129 – 141 and references cited there

ing: a capacity for reflective choice. We are able construct mental representations of scenes and scenarios that are not directly coupled to external stimuli (as in perception) or to movement (as in reflexes). We call on this capacity when we imagine possible courses of action, and then go on to imagine the possible consequences of each of these invented candidates for choice. Our ancestors used it when they painted pictures on the walls of their caves and created the first human languages. We use it when we compose an original sentence or a tune or when we try to solve an abstract problem. It allows us to reconstruct the distant past from fragmentary evidence and to envision the distant future. It makes possible the life of the mind.

The brain of every animal contains a model of the world – a hierarchical set of schemata, some of which can be modified by the animal's experience. Our models of the world are different from those of other animals not just because they are more susceptible to modification through learning. They also contain a "theoretical" component that is missing from the world models of other animals: a set of *invented* beliefs about animals, gods, weather, the Sun, the stars, other people, and ourselves. The theoretical component reflects the distinctively human capacity for imagination and invention. What are the biological underpinnings of this capacity? How does the brain generate the patterns of neural structure and activity on which imagined scenes and scenarios supervene?

Presumably, imagination and invention rely on the same strategy as perception and learning: "making and matching" – the construction, testing, and stabilization of progressively more adequate schemata. This cyclic process is constrained at many levels. At the deepest level lie evolutionarily ancient "value systems"[39] – diffusely projecting centers in the brainstem that release a variety of neuromodulators and neurotransmitters. These systems evolved to reward actions that promote survival and reproduction and discourage actions that threaten survival and reproduction. At a higher level in the hierarchy are the ethical schemata we acquire in childhood through indoctrination and education. Near the top are the schemata that define the skills and values we acquire and hone during our lifetimes. These resident schemata constrain and regulate the

39 Edelman, G.M. *Second Nature* (New Haven, Yale University Press, 2006) and earlier references cited there.

candidates for selection the brain generates when it constructs a new schema.

The composition of a work of art illustrates the complementary roles of randomness and constraint. The constraints are partly of the artist's own making but also owe much to the tradition within which he or she is working. We recognize J. S. Bach's compositions as belonging to the Baroque tradition but also as being inimitably *his*. A work of art owes its style to the constraints. It owes its freshness and originality (without which it wouldn't be a work of art) to the richness and variety of the images that made their way into the artist's mind during the creation of the work. And because the process that generated these images has, as I suggest, a random component, its outcomes are unpredictable in principle.

What brain structures and patterns of brain activity might supply the neural substrate for the schemata we've been discussing? Changeux has suggested that the initial schemata, or pre-representations, are "dynamic, spontaneous, and transient states of activity arising from the recombination of preexisting populations of neurons."[40] He has also suggested that "variability may arise from the random behavior of neuronal oscillators, but it may also be caused by transmission failures across synapses." [41] And he has speculated about how evolving schemata might be evaluated and stabilized.[42]

V.B.6. Consciousness

Up until now I have not mentioned consciousness. Understanding the biological basis of freedom and understanding the biological basis of consciousness present related but distinct problems. Neuropsychologists agree that choices and decisions result from largely unconscious physical processes. These neural processes have two distinct outcomes. One is the willed action; the other is the awareness of willing the action. Is the awareness of willing an action also its cause?

From a thorough examination of the evidence bearing on this question the psychologist Daniel Wegner has concluded that the answer is no:

40 Changeux, *The Physiology of Truth*, p. 5l

41 loc. cit.

42 op. cit. pp. 61-62

> It usually seems that we consciously will our voluntary actions, but this is an illusion.[43] ... Conscious will arises from processes that are psychologically and anatomically distinct from the processes whereby mind creates action. (emphasis added)[44]

A famous experiment by Benjamin Libet and his co-workers[45] supports Wegner's thesis. Volunteers whose brain activity was being continuously monitored were instructed to lift a finger when they felt the urge to do so, and to report the time (indicated by a spot moving around a clock face) when they felt that urge. A sharp increase in brain activity heralded the lifting of a finger. *But this spike occurred a third of a second, on average, before the volunteers became aware of the urge.* Since a cause must precede its effect, the urge could not have been the cause of the action.

Although Libet's experiment provides a useful quantitative datum bearing on hypotheses about how brain states are related to consciousness,[46] it does not bear directly on the present account of free will. Libet himself endorsed this opinion:

> My conclusion about free will, one genuinely free in the nondetermined sense, is then that its existence is at least as good, if not a better, scientific option than is its denial by deterministic theory.[47]

The present account supports this view. It has been concerned not with idealized (and often trivial) choices between two given alternatives but with what I've called reflective choice, in which the alternatives may not be given beforehand, or not completely given, and in which one works out and evaluates the possible consequences of each imagined alternative. Much of the work involved in such processes undoubtedly does not rise to the level of consciousness. But consciousness accompanies the parts of the process that *seem to us* most crucial.

In this respect, consciousness seems to play the same role as it clearly does when we learn to execute complex tasks. A violinist

43 Wegner, D.M, *The Illusion of Conscious Will* (Cambridge: MIT Press, 2002) p. 1

44 ibid. p. 29

45 Libet, B., Gleason, C.A. Wright, E.W. and D. K. Pearl, "Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act," *Brain* 106, 623-42

46 For a plausible and detailed hypothesis about the neural correlate of consciousness, see Edelman, G.M. and G. Tononi, *A Universe of Consciousness* (New York: Basic Books, 2000). For more on Libet's experiment, see Kane, op. cit. Note 12, p. 232.

47 Libet, B. "Do We Have Free Will" in Kane, R. ed., *The Oxford Handbook of Free Will*, (Oxford: Oxford University Press, 2002)

masters a difficult passage by a more or less protracted conscious process. When she comes to play the passage in a performance with fellow-members of her string quartet, she is no longer conscious of the intricacies of the passage or of the technical difficulties she has mastered. Her mind is on shaping the passage to meet the complex and not completely predictable musical demands of the particular moment. She screens out everything else. This often-noted feature of consciousness irresistibly suggests that it has a (biological) function. T. H. Huxley's steam locomotive runs perfectly well without its whistle, but the whistle nevertheless has an essential warning function.

VI. Libertarian free will

I have argued that the traditional view of free will, which holds that we are the authors of actions that help shape the future, fits comfortably into the conceptual framework of biology, which in turn fits comfortably into an unconventional but conservative picture of the physical world. In conclusion, let me review the key points of the argument.

The picture of the physical world I advocate is conservative in that it assumes that all of our current well-established physical theories are valid in their respective domains. It is unconventional in the way it links these domains to one another and in the role it assigns to chance. These features of the new picture are consequences of the assumption that an exhaustive description of the state of the universe at very early times (a) does not privilege any position or direction in space (the strong cosmological principle) and (b) depends only on physical constants and (perhaps) a few additional parameters. The strong cosmological principle entails that an exhaustive description of the physical universe has a statistical character and that the probabilities that figure in it have a concrete physical interpretation: the probability that a certain condition obtains at a particular instant of cosmic time is the fractional volume of the set of points at which the condition obtains at that instant.

Like Gibbs's statistical mechanics and its quantum counterpart, an account of cosmic evolution that comports with these cosmological assumptions characterizes macroscopic systems and their environments by probability distributions of microstates. These probability distributions are conventionally viewed as incomplete descriptions of systems in definite though unknown, or even unknowable, microstates. The present account interprets them as complete descriptions. Because microstates evolve deterministically, the conventional interpretation implies that macroscopic systems evolve deterministically. In the present account, by contrast, a macroscopic system's initial state need not uniquely determine its subsequent states.

Macroscopic indeterminism is an essential feature of biology's conceptual framework. Theodosius Dobzhansky famously wrote that "nothing in biology makes sense except in the light of evolution," and Ernst Mayr emphasized that chance plays a key role in both genetic variation and natural selection. Yet, as Mayr often stressed, this role clashes with the very limited and specific role that physicists conventionally assign to chance. An account of cosmic evolution that incorporates the cosmological assumptions mentioned above is hospitable to the biological worldview of Dobzhansky and Mayr.

Living organisms and populations are constantly challenged by unpredictable events in their environment. At the same time, macroscopic unpredictability opens up evolutionary opportunities, as when a seed borne aloft by unpredictable air currents colonizes a new habitat. The cyclic two-step strategy described by Mayr, which uses chance to defeat chance, is embodied not only in evolution itself but also in some of its products, including the immune response and visual perception. I have argued that is such an adaptation.

Free Will As a Scientific Problem[1]

I. Philosophy and the "basic facts"

According to John Searle[2], "the overriding question in contemporary philosophy" is "How do we fit in?" How can we reconcile "a conception of ourselves as conscious, intentionalistic, rational, social, institutional, political, speech-act performing, ethical and free will possessing agents" with "the basic facts," our present "reasonably well-established conception of the basic structure of the universe." As an example of the tension between our self-conception and the basic facts, Searle cites the problem of freedom and determinism. He argues that free will is a fact of conscious experience and that consciousness is a biological phenomenon, a "higher-level biological feature of the brain" (p. 48). If free will is not an illusion, physical laws and the antecedent state of a deliberator's brain cannot determine the outcome of a deliberative process. It follows that "consciousness is a feature of nature that manifests indeterminism." Up to this point I agree with the argument. But now Searle appeals to what he considers to be one of the "basic facts": "[W]e know that quantum indeterminism is the only form of indeterminism that is indisputably established as a fact of nature" (p. 74). So, Searle concludes, "consciousness manifests quantum indeterminism" (p. 75).

Although quantum indeterminism is indeed a fact of nature – a feature of our present scientific description of the world that future advances seem highly unlikely to change – its origin remains con-

1 This unpublished manuscript was written in August 2011.

2 Searle, John R., 2004. *Freedom and Neurobiology*, Columbia University Press, New York

troversial. There is no settled view about how the world that classical macrophysics describes – a refined version of the world of experience – is related to quantum microphysics. This paper sketches a novel scientific approach to this issue and discusses its implications for determinism, the nature of free will, and the relations between physics and biology and between biology and consciousness. One of my conclusions will be that for reasons that have little to do with quantum indeterminism we have the capacity to shape the future through our choices, plans, and actions.

II. What is quantum indeterminism?

Quantum indeterminism is sometimes thought to imply that processes governed by the laws of quantum mechanics have indeterminate outcomes. In fact, quantum mechanics' law of change, Schrödinger's equation, is deterministic in exactly the same sense as Newton's second law of motion: both laws determine how isolated systems – idealized physical systems that do not interact with the outside world – change with time. Although quantum mechanics and classical mechanics describe physical states differently, they agree that the state of an isolated system at any given moment determines its state at any subsequent moment. Yet quantum mechanics also *predicts* that certain processes, such as radioactive decay, have indeterminate outcomes. How does quantum mechanics reconcile the deterministic character of its law of change with its prediction that the lifetime of an unstable atomic nucleus is unpredictable in principle?

The quantum description of the history of an isolated radium nucleus is indeed deterministic. But the decay of an isolated radium nucleus – an event in the history of an isolated radium nucleus – is unobservable. What *is* observable is the result of an interaction between one of the decay products (in this example, a radon nucleus and a helium nucleus, or alpha particle) and a macroscopic detector whose construction has been specified in the language of classical physics. Quantum mechanics predicts that the *measured* lifetime of a radium nucleus has different values on different occasions, and it predicts the probability distribution of measured lifetimes. Experiments confirm these predictions.

Quantum indeterminism manifests itself only in processes of this kind: processes in which a microscopic system initially in a definite quantum state interacts with a classical measuring apparatus designed to measure one of the microscopic system's properties. A quantum measurement leaves the measured system in one of several possible quantum states and leaves the apparatus in a correlated classical state. For example, each possible post-measurement state of the quantum system might be correlated with the position of a pointer. Quantum mechanics predicts both the possible outcomes of such a measurement and their probabilities.

Could measurement-like processes in the brain mediate exercises of free will? Free will, as defended by Immanuel Kant, William James, Robert Kane, and John Searle, is more than the ability to make unforced choices between given alternatives. It is the capacity to shape future events. Our present understanding of neurophysiology offers little support for the view that measurement-like processes in the brain underlie choices that are neither forced nor random. Must we then conclude that our felt capacity to shape the future through our free choices is an illusion? This dilemma invites a closer look at how physicists describe quantum indeterminism.

III. The standard account of quantum indeterminism

Physical theories are self-contained mathematical structures linked to the results of possible measurements by auxiliary rules. Newtonian mechanics represents a physical system's measurable properties by real variables (i.e., mathematical objects whose possible values are real numbers). The auxiliary rule that links these variables to the outcomes of possible measurements identifies the value of each variable either with the result of an idealized measurement or with the result of a calculation that expresses the result of the measurement in terms of measurements of other variables such as position and elapsed time. In this respect, Newtonian descriptions are refined versions of descriptions in ordinary language.

The mathematical structure of quantum mechanics is less closely tied to experience. Quantum mechanics represents an isolated system's possible physical states by "state vectors," vectors of unit length in an abstract multidimensional vector space. It represents the system's measurable properties by operators, mathematical objects that act on vectors in this space. The standard formulation of quantum mechanics[3] contains a rule that links the mathematical description of an isolated physical system to the results of possible measurements of the system's properties. This rule, the measurement postulate, equates the average of "a large number" of measurements of a given property to a quantity that depends on the operator that represents the measured property and the state vector that represents the state of the measured system. From this rule and quantum mechanics' mathematical formalism one deduces: (a) that a single measurement does not in general have a definite outcome; (b) the set of possible outcomes; and (c) the probability associated with each possible outcome (or range of possible outcomes). Thus quantum indeterminacy is a consequence not of the mathematical formalism alone but of the formalism plus a rule that links the mathematical formalism to the results of (ideal) measurements.

Experiments leave no room for doubt that the standard formulation of quantum mechanics is correct. As an instrument for making predictions about the outcomes of measurements, the standard formulation leaves nothing to be desired. But many physicists have sought a deeper account of the linkage between quantum physics and classical physics than that provided by the measurement postulate. This account would not just postulate that a composite system consisting of a quantum system coupled to a classical measuring apparatus evolves indeterministically. It would explain why. The account that emerges from the following discussion implies that contrary to conventional opinion, quantum indeterminism is not the only form of indeterminism. A variety of macroscopic processes, I will argue, have indeterminate outcomes; chance is endemic in the macroscopic domain.

IV. Quantum mechanics and classical physics

Quantum mechanics and classical mechanics are closely related. Not only do classical properties like position, momentum, and energy have quantum counterparts. At a deep formal level the laws that govern classical properties and their rates of change are identical with the laws that govern their quantum counterparts. Moreover, the

3 Dirac, P. A. M., T*he Principles of Quantum Mechanics*, fourth edition (revised), (Oxford, Clarendon Press, 1967)

domains in which classical mechanics and quantum mechanics are valid overlap. For example, the classical description of an electron circling a proton approximates the quantum description of an electron in a hydrogen atom when the electron's angular momentum (the product of its momentum and the radius of its circular orbit) greatly exceeds Planck's constant. The formal similarity between quantum and classical mechanics and the overlap between their domains have motivated many attempts over many years to formulate quantum mechanics in a way that includes classical mechanics as a limiting case. That these attempts have not yet been completely successful is due largely to a central feature of quantum mechanics: the principle of superposition.

To illustrate the principle, consider a free electron. Because its position coordinates are represented by operators in an abstract vector space, they do not have definite values at a given moment. The electron could, however, be in a state in which its *measured* position was almost certain to lie inside a small sphere S. It could also be in a different state, in which its measured position was almost certain to lie inside a different, nonoverlapping sphere S' arbitrarily distant from S. Call the state vectors that represent these two states U and V. The principle of superposition says that any linear combination of these state vectors, $aU + bV$, represents a possible state, where a and b are complex numbers whose squared magnitudes add up to 1. States represented by such state vectors have no classical analogue; a point-like particle cannot be in two non-overlapping regions at the same time. Of course, quantum mechanics (which includes the measurement postulate) does not claim that they can be. It predicts (and experiments confirm) that when an electron is in a superposition of the states U and V its measured position will either be a point in S or a point in S'. It also predicts the probabilities of these outcomes: they are the squared magnitudes of the coefficients a and b.

The domain of quantum mechanics has no clearly defined boundary. Nothing in the mathematical formalism indicates that it applies only to small or simple systems. Let us therefore assume, as most physicists do and as I will do, that the domain of quantum mechanics includes macroscopic systems. In other words, let us assume that an isolated macroscopic system has a set of possible quantum states and that it is in one of these states. Now assume that the measuring apparatus in a quantum measurement, which is necessarily a macroscopic system, is initially in a definite quantum state and that the combined system (measured system + measuring apparatus) remains isolated during the measurement. Then, as John von Neumann, showed in 1932, an ideal measurement would cause the state vector of the combined system to evolve into a superposition of state vectors each of which represents one of the measurement's possible outcomes as given by the measurement postulate; and the probability of each outcome, as given by the measurement postulate, would coincide with the squared magnitude of the coefficient of the corresponding state vector in the superposition.

But a quantum measurement does not, as this account predicts, produce a superposition of outcome states. It produces just one of them. To bring his account into agreement with the measurement postulate (and experiment), von Neumann postulated that the predicted superposition of outcome states no sooner forms than it collapses unpredictably onto one of them, with a probability given by the squared magnitude of the coefficient of the corresponding state vector in the superposition. (This hypothetical process is called the collapse, or reduction, of the state vector.)

Now, superpositions do not collapse in other physical contexts. Every quantum state of a composite system AB made up of interacting quantum systems A and B is a superposition of states in each of which A is in a definite quantum state and B is in a correlated quantum state; and such superpositions never collapse. They are a ubiquitous (as well as distinctive) feature of quantum descriptions. Their existence has been experimentally confirmed on innumerable occasions. What, then, distinguishes quantum measurements from other physical processes governed by quantum laws? Here are three answers that have been offered by physicists:

1. Eugene Wigner[4] once argued that what distinguishes quantum measurements from other physical processes is that a quantum measurement necessarily involves consciousness, for it is completed only when an observer becomes aware of its outcome. And it is at that point that our knowledge of the state of the combined system (mea-

---

4 Wigner, E. P., "Remarks on the Mind-Body Question" in *Symmetries and Reflections*, (Cambridge MIT Press, 1962)

sured system + measuring apparatus) changes in a way not governed by Schrödinger's deterministic law of change:

> In other words, the impression which one gains at an interaction [between an observer and a physical system], called also *the result of an observation*, modifies the wave function of the system. The modified wave function is, furthermore, in general unpredictable before the impression gained at the interaction has entered our consciousness: it is the entering of an impression into our consciousness which alters the wave function ... . It is at this point that consciousness enters the theory unavoidably and unalterably.[5]

Wigner regarded the argument summarized in the preceding passage as the weaker of two arguments "support[ing] the existence of an influence of … consciousness on the physical world" (p. 181)." The stronger argument "is based on the observation that we do not know of any phenomenon in which one subject is influence by another without exerting an influence thereupon." I will return to the question of how consciousness fits into a scientific description of the world later in this essay.

2. Wigner[6] later defended an instrumental interpretation of quantum mechanics.

> It appears that the statistical nature of the outcome of a measurement is a basic postulate, that the function of quantum mechanics is not to describe some "reality," whatever this term means, but only to furnish statistical correlations between [an observation and] subsequent observations. This assessment reduces the state vector to a calculational tool, an important and useful tool, but not a representation of "reality."

Many if not most contemporary physicists agree with this view. It provides a philosophical justification for the standard formulation, preempting the question "Why does the state vector collapse?"

3. Hugh Everett III[7] postulated that the domain of quantum mechanics includes all physical systems up to and including the physical universe. He argued that we should identify the measur-

5 ibid. pp. 175-6

6 Wigner, E. P., 1976. "Lecture Notes" in J. A. Wheeler and W. H. Zurek, eds., *Quantum Theory and Measurement*, (Princeton, Princeton University Press,1983)

7 Everett, Hugh, III, 1957. *Reviews of Modern Physics* 29, 454

ing apparatus in a quantum measurement with the universe minus the measured system. Von Neumann's account then predicts that a quantum measurement creates a superposition of quantum states of the universe. But since quantum mechanics is universally valid, this superposition never collapses. Its components represent macroscopically distinguishable, coexistent states of the universe, all equally real.

Several contemporary cosmological theories incorporate Everett's assumption that quantum mechanics is universally valid. All of these theories raise Searle's question "How do we fit in?" with a vengeance. For in that question "we" now means not just "we humans" but more broadly "the world that classical physics describes." So far as I know, that question has not yet been satisfactorily answered. We lack a physical theory that postulates the universal validity of quantum mechanics and contains classical physics and general relativity, the classical (and strongly confirmed) theory of the physical universe and of the structure of space-time, as limiting cases.

V. Microphysics and macrophysics: statistical mechanics

The measurement postulate and von Neumann's collapse postulate serve in different ways to bridge the gap between microphysics and macrophysics. The gap itself predates quantum mechanics by almost two and a half centuries. In the *Principia* Newton, a convinced atomist, proposed a molecular model of air to account for Robert Boyle's empirical law relating the pressure and the volume of an enclosed sample of air. He attributed the fact that an isolated sample of air expands to fill any enclosure, no matter how capacious, to a hypothetical repulsive force between neighboring air molecules.

In 1738 Daniel Bernoulli proposed a much simpler molecular model. He assumed that the hypothetical air molecules travel freely between relatively short-lived collisions. When a molecule bounces off a wall it transfers momentum to the wall, in accordance with Newton's second and third laws of motion. Averaged over many molecular impacts and over a macroscopic time interval, the result is a steady pressure. This model, like Newton's, predicted that the pressure a gas sample exerts on the walls of its container is inversely proportional to its volume. But it does more. If one identifies the average kinetic energy of a gas molecule with the gas temperature, as was done a century later, Bernoulli's formula for gas pressure includes the remaining empirical gas laws as well as Avogadro's hypothesis.

These successes were gratifying – but also, in one respect, surprising. A macroscopic sample of air has a vast number of microscopic degrees of freedom – six for every molecule in the sample. Yet it seems to have only two macroscopic degrees of freedom: experiments show that the values of two macroscopic properties of the sample, such as its temperature and its pressure, determine all of its measurable properties. In two papers, published in 1860 and 1866, James Clerk Maxwell explained why. He argued that molecular collisions, governed by Newton's laws of motion, cause the distribution of molecular velocities in an isolated gas sample to evolve toward a unique equilibrium distribution that depends on a single parameter, the average molecular kinetic energy.

Rudolf Clausius and others had earlier identified this quantity with the gas temperature. Experiments show that a gas sample that is well insulated from its surroundings does indeed quickly relax into a state whose macroscopic properties are averages of appropriate molecular properties over the Maxwell distribution of molecular velocities.

In 1872 Ludwig Boltzmann extended Maxwell's theory. Maxwell had studied the distribution of molecular velocity in a uniform gas sample; Boltzmann studied the joint distribution of molecular position and momentum in (possibly) nonuniform gas samples, and derived a mathematical law – his transport equation – that governs changes in this distribution produced by molecular collisions. Boltzmann also discovered a statistical counterpart of entropy[8]and proved that in a uniform gas sample it increases monotonically until it reaches the largest value that is consistent with the combined energy of the molecules in the sample. The distribution of molecular velocities then becomes the Maxwell distribution. This counterpart to the law of entropy growth is known as Boltzmann's *H* theorem. Its proof depends in part on the fact that a Newtonian description of an

8 The statistical entropy of a discrete probability distribution is the mean value of the negative logarithm of the probability. The statistical entropy of a continuous probability distribution is the mean value of the negative logarithm of the probability density.

encounter between two molecules doesn't change when one reverses the direction of the time axis.

The Maxwell-Boltzmann theory applies to samples of an ideal gas. It predicts that molecular collisions cause an isolated sample of an ideal gas to relax into a state in which its molecules are uniformly distributed within the enclosure and have a Maxwell velocity distribution. This state is the counterpart of thermodynamic equilibrium. Relations between appropriate molecular properties, averaged over the distribution of molecular position and velocity, mirror relations between thermodynamic quantities that prevail in thermodynamic equilibrium.

In 1901 Josiah Willard Gibbs published a more general statistical theory of thermodynamic equilibrium. It applies to any system of N particles whose motions and interactions are governed by Newton's laws. Gibbs characterized the macrostates of such a system by probability distributions of its microstates, each specified by a set of $6N$ real numbers, the three position coordinates and three momentum components of each of the system's N particles. As Maxwell and Boltzmann had done, he identified the system's macroscopic properties with mean values of appropriate microscopic properties. In particular, he identified entropy with statistical entropy as given by Boltzmann's formula. He proved that the statistical entropy of his "canonical distribution" – a generalization of the Maxwell molecular-velocity distribution – exceeds that of any other distribution with the same mean micro-energy. He also proved that the statistical entropy of any N-particle distribution of microstates of an isolated system is constant in time.

Gibbs's theory, which he called statistical mechanics, requires only small formal changes[9] when one uses quantum mechanics rather than classical mechanics to describe microstates. So modified, it reproduces all the laws of equilibrium thermodynamics and goes far beyond it. Like the standard formulation of quantum mechanics it leaves nothing to be desired as an instrument for making predictions about measurement outcomes. Also like the standard formulation of quantum mechanics it raises questions that fall in an area where

9 The statistical entropy of a discrete probability distribution is the mean value of the negative logarithm of the probability. The statistical entropy of a continuous probability distribution is the mean value of the negative logarithm of the probability density.

physics and philosophy overlap. One of these questions concerns the interpretation of probability.

The probability distributions that figure in Maxwell's and Boltzmann's statistical theories represent relative frequencies: we can identify the probability that the position coordinates and momentum components of a gas molecule lie in given ranges with the fraction, or proportion, of gas molecules in a macroscopic sample whose position coordinates and momentum components lie in these ranges at a given moment. Since a macroscopic sample contains a vast number of molecules, this interpretation is virtually exact. But how are we to interpret the probability distributions of microstates that characterize macrostates in Gibbs's theory?

Gibbs imagined a large or infinite collection of replicas of the macroscopic system, each in a definite microstate. He identified the probability associated with a given range of microstates with the fraction of the imaginary replicas whose microstates lie in that range. Recognizing that relative frequencies in an imaginary collection are just as abstract as the probabilities they represent, Gibbs referred to his statistical descriptions as analogues of thermodynamics. But because, as I have mentioned, quantum statistical mechanics not only duplicates the predictions of thermodynamics but also goes well beyond them, it is presumably the more fundamental theory. And if that is the case, the probability distributions that represent macrostates in quantum statistical mechanics need a physical interpretation.

The most obvious possibility is to suppose that the isolated macroscopic system whose equilibrium macrostates are characterized by a probability distribution of microstates is actually in one of these microstates. The probability distribution then represents an observer's limited knowledge of the system's microstate:

> [M]acroscopic observers, such as we are, are under no circumstances capable of observing, let alone measuring, the microscopic dynamic state of a system which involves the determination of an enormous number of parameters, of the order of 10^{23}. ... [Thus] a whole ensemble of possible dynamical states corresponds to the same macroscopic state, compatible with our knowledge.[10]

E. T. Jaynes carried this interpretation a step further. He argued that "statistical mechanics [is] a form of statistical inference rather

10 Jancel, R., *Foundations of Classical and Quantum Statistical Mechanics*, (Oxford, Pergamon, 1963), p. xvii

than a physical theory." Its "computational rules are an immediate consequence of the maximum-entropy principle," which yields "the best estimates that could have been made on the basis of the information available" [11]On this view, the statistical entropy of a probability distribution that represents an isolated system's equilibrium macrostate represents physicists' lack of information about the system's microstate; and Gibbs's theorem that the statistical entropy of the canonical distribution exceeds that of any other distribution subject to the same constraints exemplifies the principle of maximum-entropy inference.

Werner Heisenberg interpreted statistical mechanics in much the same way. He also linked physicists' incomplete knowledge of the microstructure of macroscopic systems (and, more generally, of the world) to their inability to predict the outcomes of quantum measurements:

> [The interaction between a measured quantum system and a measuring device] introduces a new element of uncertainty, since the measuring device is necessarily described in terms of classical physics; such a description contains all the uncertainties concerning the microscopic structure of the device which we know from thermodynamics, and since the device is connected with the rest of the world, it contains in fact the uncertainties of the microscopic structure of the whole world. These uncertainties may be called objective in so far as they are simply a consequence of the description in terms of classical physics and do not depend on any observer. They may be called subjective in so far as they refer to our incomplete knowledge of the world.[12]

Another version of the epistemic interpretation of probability distributions in equilibrium statistical mechanics begins with the remark that every macroscopic system interacts weakly with its surroundings. This interaction causes the system to visit a range of microstates with nearly the same energy. One then identifies the probability that the system's actual (but unknown) microstate lies in a given range of microstates with the relative frequency with which the system's

11 Jaynes, E.T., 1957, *Phys. Rev*. 106, 620; 108, 171

12 Heisenberg, W. *Physics and Philosophy*, (London, Allen and Unwin, 1958), pp. 53-54

microstate visits this range; see, for example, Schrödinger,[13] Landau and Lifshitz,[14] and Feynman.[15]

In contrast with these epistemic interpretations, I will argue that probability distributions of microstates, viewed in a particular cosmological context, characterize macrostates completely. This view does not assume that macroscopic systems are "really" in definite – however short-lived – microstates. Thus it is a variety of what Lawrence Sklar[16] in his insightful account of the foundations of equilibrium and non-equilibrium statistical mechanics calls "tychism."

VI. Microphysics and macrophysics: time's arrow

Whereas equilibrium statistical mechanics fully reproduce the mathematical laws of equilibrium thermodynamics, statistical theories that describe how systems initially in non-equilibrium states relax into equilibrium run into two problems. These are exemplified by Boltzmann's transport equation, which governs changes in the joint probability distribution of molecular position and momentum resulting from molecular collisions in an isolated gas sample. Boltzmann proved that these changes have a one-way character. They cause the statistical entropy of this probability distribution to increase monotonically toward the largest value that is consistent with the sample's mean energy per molecule; and this value characterizes the equilibrium distribution (Boltzmann's H theorem). Two questions now arise. 1. How can molecular interactions governed by Newton's time-reversal-invariant laws of motion give rise to one-way macroscopic behavior? 2. How can the growth of molecular statistical entropy be reconciled with Gibbs's proof that the statistical entropy of the joint N- particle probability distribution (for an isolated system of N particles) is constant in time?

The source of directionality in Boltzmann's derivation of his transport equation isn't hard to spot. Following Maxwell, Boltzmann assumed that the *incoming* velocities of colliding molecules are sta-

13 Schrödinger, E. *Statistical Mechanics*, (Cambridge, Cambridge University Press, 1948), p. 3

14 Landau, L. and Lifshitz, E.M., *Statistical Physics*, Part 1, revised, translated from the Russian by J.B. Sykes and M.J. Kearsley, (Reading, MA, Addison-Wesley 1980),

15 Feynman, R., *Statistical Mechanics, a Set of Lectures*, (Reading, MA , W.A.Benjamin, 1972)

16 Sklar, L., *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*, (New York , Cambridge University Press, 1993)

tistically uncorrelated. That is, he assumed that the joint probability distribution of the incoming velocities of the collision partners is the product of the individual probability distributions. Now, Newton's laws of motion imply that the combined energy and the combined momentum of colliding molecules have the same values before and after a collision. So if the incoming velocities of the collision partners are uncorrelated, their outgoing velocities must be correlated. Boltzmann's derivation assumes, however, that molecular correlations are *permanently* absent. This assumption cannot be true for an isolated gas sample. Even if molecular correlations were absent initially, they would subsequently be produced by molecular collisions.

Yet Boltzmann's equation enjoys strong experimental support. Not only does it include as special cases the phenomenological laws that govern such irreversible processes as heat flow, molecular diffusion, and viscous dissipation of relative fluid motions, it also enables one to express the coefficients that figure in these laws in terms of quantities that characterize molecular properties, molecular motions, and molecular interactions.[17] Predictions based on Boltzmann's equation have passed all experimental tests with flying colors. Boltzmann's statistical theory belongs to a large class of statistical theories that describe irreversible processes and that contain counterparts to his H theorem. Van Kampen[18] has pointed out that these theories all depend on "repeated randomness assumptions," analogous to Boltzmann's assumption that molecular correlations are permanently absent; Sklar, in the book cited in Note 13, calls them "rerandomization posits." What justifies them? Prigogine has argued that subjective justifications are implausible:

[In Boltzmann's theory] irreversibility comes from supplementary phenomenological or subjectivist assumptions, from 'mistakes.' But how can we account for the wealth of important results and concepts that derive from the second law? In a sense living things, we ourselves, are then 'mistakes.'[19]

17 Clausius, in 1857, derived these phenomenological laws from Bernoulli's model, supplemented by an additional molecular parameter: the average distance a molecule travels between collisions. And Maxwell, in the two papers mentioned above, improved Clausius's theory by constructing a detailed account of molecular collisions and their effects.

18 Van Kampen, N.G., *Stochastic Processes in Physics and Chemistry*, 3rd ed., (Amsterdam Elsevier, 2007)

19 Prigogine, I., *From Being to Becoming*, (New York, W. H. Freeman, 1980), p. 157

Prigogine and his collaborators have argued that macroscopic irreversibility must be rooted in irreversible microscopic laws that underlie quantum mechanics in its present form.

A more modest suggestion justifies repeated-randomness assumptions by an appeal to environmental interactions. Experimental physicists can effectively prevent an enclosed gas sample from exchanging matter or energy with the outside world. But they cannot prevent the leakage of information associated with molecular correlations. For if the molecules that make up the walls of the enclosure have a maximally random probability distribution (i.e., a probability distribution whose statistical entropy is as large as possible), collisions between gas molecules and wall molecules create statistical correlations between wall molecules and gas molecules and attenuate correlations between gas molecules, thus preventing them from building up to a point where they invalidate the assumption that the incoming velocities of colliding molecules are statistically uncorrelated.

The assumptions that enclosed gas samples initially lack molecular correlations and are embedded in random environments, which wick away correlation information, exemplify an approach that J. M. Blatt[20] and others have called "interventionism." Blatt constructed a mathematical model that allowed him to estimate the rate at which random interactions between an enclosed gas sample and the walls of its container destroys correlation information. He noted that interventionism had not been a popular approach:

> There is a common feeling that it should not be necessary to introduce the wall of the system in so explicit a fashion. ... Furthermore, it is considered unacceptable philosophically, and somewhat "unsporting," to introduce an explicit source of randomness and stochastic behavior directly into the basic equations. Statistical mechanics is felt to be a part of mechanics, and as such one should be able to start from purely causal behavior[21] .

Sklar, in the book cited in Note 13, gave a more detailed critique of interventionism. Shenker[22] responded to this critique and offered a qualified defense of interventionism.

20 Blatt, J. M. 1959, *Prog. Theor. Phys.* 22, 745

21 ibid. p. 747

22 Shenker, O.R., 2000, "Interventionism in Statistical Mechanics: Some Philosophical Remarks" (preprint)

The *physical* problems broached by interventionism are (a) to supply objective definitions of "randomness" and "correlation information" and (b) to justify the assumption that macroscopic systems are initially deficient in correlation information and are embedded in random environments. I will address these problems in due course.

Interventionist theories of irreversibility in statistical mechanics are analogous to decoherence theories of quantum measurement.[23] As emphasized by Niels Bohr, the measuring apparatus in a quantum measurement is necessarily a macroscopic system, and the registration of a measurement outcome is an irreversible macroscopic process. Decoherence calculations show how interaction between the combined system in a quantum measurement and a random environment, such as a dilute gas or a radiation field, effectively randomizes the relative phases of the coefficients in the superposition predicted by von Neumann's account of an ideal measurement. Decoherence calculations explain why the superposition of macroscopically distinguishable quantum states predicted by that account cannot exhibit effects analogous to interference between light waves in diffraction experiments.[24] But as Erich Joos and Hans-Dieter Zeh emphasized in a classic paper[25] on decoherence and quantum measurement, decoherence alone does not explain why quantum measurements have definite outcomes.

Interventionism resolves the apparent contradiction between Gibbs's theorem that the statistical entropy of the N-particle distribution is constant in time and Boltzmann's theorem that the statistical entropy of an initially non-equilibrium one-particle distribution increases monotonically with time:

If (and only if) molecular correlations are absent, the N-particle distribution reduces to a product of identical one-particle distributions,

23 For a clear and comprehensive review of decoherence theories, see Schlosshauer, M., *Decoherence and the Quantum-to-Classical Transition*, corrected 2d printing (Berlin Springer, 2008)

24 How two light waves with the same wavelength interfere at a given point depends on their relative phases at that point: they interfere constructively if they are in phase destructively if they are out of phase. Interference between state vectors in a superposition depends on the relative phase of their coefficients in the superposition. The interaction between the combined system in a quantum measurement and a random environment, such as a dilute gas or a radiation field, effectively randomizes the relative phases of the coefficients in the superposition predicted by von Neumann's account of an ideal measurement.

25 Joos, E. and Zeh, H. D., 1985. *Zeitschrift für Physik* B 59, 223

and the statistical entropy of the N-particle distribution becomes N times the statistical entropy of the one-particle distribution.

It is easy to prove that if molecular correlations are present (so that the N-particle distribution does not reduce to a product of identical one-particle distributions), the statistical entropy of the N-particle distribution (call it S_N) is less than N times the statistical entropy S_1 of the one-particle distribution. The difference represents *correlation information*: $I_{correlation} = NS_1 - S_N$. Since S_N is constant in time, the growth of correlation information requires that the one-particle statistical entropy S_1 to increase with time, as Boltzmann inferred from his transport equation.

If correlation information is initially absent, it is created by molecular encounters, and the one-particle statistical entropy increases. This state of affairs continues to prevail if the gas sample under consideration interacts weakly with a random environment that wicks away correlation information and disperses it to the wider universe.

VII. The relevance of cosmology

Much of physics treats systems whose interaction with the rest of the universe is either negligible or can be described in a simple way. As Heisenberg reminded his readers in the essay I have been quoting,

> [I]t is important to remember that in natural science we are not interested in the universe as a whole, including ourselves, but we direct our attention to some part of the universe and make that the object of our study [p. 52]

But as Zeh[26] has pointed out, we cannot consistently assume that macroscopic systems are truly isolated; we must allow for their interaction with their surroundings. And once we do that, we find ourselves on a slippery slope.

Zeh's argument is straightforward. An isolated system is in a definite quantum state. But the possible quantum states of a macroscopic system are so closely spaced in energy that they must be "entangled" with the quantum states of the part of the environment with which the system interacts.[27] The same argument applies to every

26 Zeh, H-D, 1970/ *Foundations of Physics* 1, 69

27 Interaction between two systems entangles their quantum states in the same sense as the interaction between a measured system and a measuring apparatus entangles the quantum states of the system and the apparatus, or the interaction between the electrons in a helium atom entangles the quantum states of the individual electrons.

bounded part of the environment. So – this is the slippery slope – if quantum mechanics applies on all scales, the quantum states of any macroscopic system must be entangled with quantum states of the rest of the universe. This conclusion forms the starting point of many-worlds interpretations of quantum mechanics, beginning with Everett's "relative state" interpretation.

But if we accept the conclusion that the universe is in a definite quantum state, we face the (unsolved) problem of explaining how the world of classical physics, which includes the world of experience, fits in. We also face the problem of explaining how Einstein's theory of space, time, and gravitation – general relativity – fits in. A theory that included quantum mechanics and general relativity as limiting cases would, of course, solve that problem; but such a theory doesn't yet exist. Nevertheless, many physicists postulate that quantum mechanics does apply at all scales and that the universe is in a definite quantum state.

The relation between quantum mechanics and classical physics is problematic in another respect. Quantum mechanics and general relativity enjoy overwhelming observational and experimental support in their respective domains. But their domains overlap. And this poses a problem. General relativity is a classical, deterministic theory; but quantum measurements can produce unpredictable macroscopic changes in the structure of space-time. How can these apparently contradictory features of our two most fundamental theories be reconciled?

As I have discussed, many physicists postulate that quantum mechanics is universally valid. They hope and expect that general relativity will one day be shown to be a limiting case of a quantum theory of gravity. Einstein, by contrast, hoped that quantum mechanics would one day be found to be a limiting case of a deeper deterministic field theory – a hope shared by few contemporary physicists.

I will suggest a third way. We do not yet have a unified set of mathematical laws that includes the laws of quantum mechanics and the field equations of general relativity as limiting cases. But I will sketch a theory of initial and boundary conditions that makes quantum mechanics and general relativity compatible in their shared domain. At the same time it offers solutions to the quantum measurement problem and the problem of time's arrow. The framework of the proposed theory of initial and boundary conditions is a version of relativistic cosmology. It unites but does not unify quantum mechanics and general relativity, showing that they coexist peaceably in their common domain.

VIII. Relativistic cosmology: the cosmological principle

Newton speculated that the stars were distant suns uniformly distributed throughout an infinite Euclidean space. But because his theory of gravitation does not apply to an infinite, unbounded distribution of mass, he was unable to formulate a mathematical theory based on this idea. Before formulating his generalization of Newton's theory, general relativity, Einstein had hoped that it would fill this gap, and soon after completing the theory, in 1915, he tried to apply it to an idealized model of the universe: a uniform, unbounded, pressure-free, static medium. He found that his field equations had no solution that satisfied these conditions, and in 1917 he suggested a modification of his 1915 field equations that allowed them to have a static solution. Five years later Alexander Friedmann showed that while the original field equations do not have static solutions, they do have non-static solutions, in which space and its contents undergo a uniform expansion from (or towards) a singular state of infinite mass density.

In 1929 Edwin Hubble announced that the most distant galaxies whose distances and line-of-sight velocities could then be measured were systematically[28] receding from Earth at speeds proportional to their distances; they were taking part in a uniform expansion, just as Friedmann, unbeknownst to Hubble, had predicted. Modern observations of galaxies and of the cosmic microwave background, discovered in 1965, support this conclusion.

To estimate the distances of distant galaxies, Hubble assumed that they and their stellar populations have the same statistical properties as nearby galaxies and stellar populations. (This enabled him to use distance criteria calibrated on objects close enough to have measurable parallaxes.) The *cosmological principle*, the starting point for conventional cosmological theories, is a generalization of Hub-

28 Galaxies also have "peculiar" velocities associated with local deviations from uniformity. Until Hubble's observations, these masked the systematic component associated with the expansion of space.

ble's assumption (which he called the principle of uniformity). It says that there is a system of spacetime coordinates relative to which no statistical property of the universe at a given moment serves to define a preferred position or direction in space.

Physicists usually study idealized models of real physical systems. They take it for granted that the initial and boundary conditions that characterize these models hold only approximately. Galileo assumed that the effects of air resistance on the motions of falling and sliding objects masked simple and exact mathematical laws, and in his experiments he took pains to minimize these effects. Astrophysicists know that stars rotate and are chemically inhomogeneous; but they begin by idealizing them as chemically homogeneous, nonrotating gas spheres. In the same spirit one might – and many physicists do – regard the cosmological principle as characterizing a class of simplified models of the universe. By contrast, the following considerations take as their starting point the assumption that the cosmological principle is an exact symmetry of the initial conditions that characterize the universe (as it is of all our present physical laws). I assume further that a statistical description that enjoys this symmetry cannot be augmented by nonstatistical information. In this sense a statistical description that comports with the assumption (which I will refer to as the strong cosmological principle) is complete.

A Newtonian universe cannot satisfy the strong cosmological principle. Consider, for example, a statistically uniform distribution of free particles. A complete Newtonian description of such a distribution at a given moment would specify the distance between every particle and its nearest neighbor. Thus it would assign every particle a real number – the instantaneous distance of its nearest neighboring particle. But the number of particles (and pairs of nearest neighbors) is at most countably infinite. So if the distribution is random, there is zero probability that two of these real numbers coincide. Every particle is uniquely situated with respect to its neighbors.

In contrast, it follows from Heisenberg's indeterminacy principle that quantum mechanics assigns any bounded region of a uniform distribution of free particles a finite number of quantum states, provided the particles' momenta (or energies) are also bounded. Suppose the distribution is infinitely extended, as comparisons between astronomical observations and refined versions of Friedmann's cosmological models indicate, and that all its statistical properties are uniform. Then with probability one, any given bounded region will have infinitely many replicas in the same quantum state. It follows that any two realizations of the same (uniform) statistical description are *finitely indistinguishable* in the sense that any bounded region of one realization has infinitely many exact matches in any other realization. This conclusion and its supporting argument can easily be extended to *any* statistical description of an infinite universe that satisfies the cosmological principle. *In effect, a statistical description of an infinite universe that does not privilege any point or direction in space has a single realization.*

IX. The growth of order and the growth of entropy

Extrapolating the present state of the (observable) universe backward in time, one arrives at an era when, at each moment, the cosmic medium closely approximates a mixture of free particles in thermal and chemical equilibrium.[29] The relative concentrations of particle kinds in thermodynamic equilibrium depend on the mass density and the temperature. As the medium expands, its mass density and its temperature decrease. At sufficiently early times the rates of equilibrium-maintaining particle reactions greatly exceed the rate at which the mass density and the temperature are changing, so particle reactions are able to maintain equilibrium at the instantaneous values of the mass density and the temperature. Now, particle reaction rates and the rate at which space is expanding both decrease with decreasing mass density; but the expansion rate decreases more slowly. Eventually the particle reactions tasked with maintaining the relative concentrations of particle kinds appropriate to chemical equilibrium at the instantaneous mass density and temperature become unable to do so, and the relative abundances of helium and some other light elements become frozen in.[30]

If we define the statistical information of a probability distribution as the amount by which the distribution's statistical entropy falls short of its largest allowed value, then the process just described –

29 Weinberg, S., *Gravitation and Cosmology*, (New York, Wiley, 1972)

30 Weinberg (footnote 26) gives a detailed account of nucleogenesis in an initially hot universe; Anthony Aguirre (*Astrophysical Journal*, 521.1 (1999): 17.) discusses nucleogenesis in an initially cold universe.

nucleogenesis – creates information – specifically, chemical information. Much later, thermonuclear reactions in the core of the Sun degrade some of this information when they burn hydrogen into helium. Some of the energy released by these reactions is converted into sunlight, which drives the biological processes that sustain life on Earth.

The expansion also creates structural information. Self-gravitating astronomical systems could not have existed at the high mass densities that prevailed when helium and light nuclei were formed. They must have come into being later in the cosmic expansion.

There is no consensus about how this happened. On one scenario[31] an initially cold cosmic medium solidifies as metallic hydrogen.[32]As the expansion continues, the medium breaks up into fragments whose cohesion energies are approximately equal to their gravitational binding energies. These fragments, the first self-gravitating systems, are less massive by one or two orders of magnitude than the giant planets. At this stage the cosmic medium is a cold "gas" whose "particles" are solid-hydrogen fragments.

Because the "particles" are randomly distributed, the gas's internal energy contains a negative contribution associated with the fluctuating part of their local gravitational interactions as well as a positive contribution due to the particle motions relative to the expanding background produced by the fluctuating local gravitational field. Initially these contributions are equal, but the expansion attenuates the positive contribution faster than the negative contribution, so that eventually small self-gravitating clusters of "particles" separate out as self-gravitating systems. These newly formed self-gravitating systems now take over the role of particles, and the process – gravitational clustering – continues, giving rise to self-gravitating systems on progressively larger scales.

This scenario predicts that the initial binding energy per unit mass of a self-gravitating system is proportional to the one-third power of the system's mass. Astronomical measurements are consistent with the predicted relation over a range of masses that extends from giant

31 Layzer, D. *Constructing the Universe*, (New York, W. H. Freeman, 1984); *Cosmogenesis*,(New York, Oxford University Press, 1990)

32 Layzer, D. and Hively, R., 1973, *Astrophysical Journal* 179, 361

planets and their satellites to rich galaxy clusters – eighteen powers of ten.

The assumption that the early universe was cold, first suggested by Zel'dovich in 1962, conflicts with the standard interpretation of the cosmic microwave background as a relic of a primordial radiation-dominated phase of cosmic evolution, the hot big bang.

The standard interpretation accounts for some observed features of the cosmic microwave background. The cold-universe scenario, in contrast, interprets the cosmic microwave background as thermalized radiation from an early generation of supermassive stars.[33]

Self-gravitating systems evolve toward states of dynamical equilibrium, in which the cohesive effect of gravity balance the disruptive effect of internal motions. But these states differ radically from states of *thermodynamic* equilibrium. Consider, for example, a self-gravitating gas cloud of nearly uniform temperature. As the cloud loses energy by radiation, it contracts *and its temperature increases*. Thus a self-gravitating gas cloud in dynamical equilibrium has negative heat capacity. (By contrast, a system in thermodynamic equilibrium necessarily has positive heat capacity.) As the cloud evolves it departs progressively further from the featureless state of thermodynamic equilibrium: a radial temperature gradient develops and heat flows outward from the center. If the core temperature becomes high enough, thermonuclear reactions produce a radial gradient of chemical composition.

Of course, the local macroscopic processes that take place in a self-gravitating gas cloud – the transfer of heat from the cloud to its cooler surroundings, the flow of heat down the steepening radial temperature gradient, the thermonuclear reactions in the cloud's core – all generate entropy. But these entropy-generating processes drive the cloud and its surroundings away from global thermodynamic equilibrium.

Thus in bounded self-gravitating systems, as in the expanding cosmic medium, gravity opposes the macroscopic processes that seek to establish thermodynamic equilibrium.

---

33 Aguirre, A., 1999. *Astrophysical Journal* 521, 17 (1999)

X. Entropy and the law of entropy growth

If thermodynamic equilibrium prevails locally in a self-gravitating system, one can define the system's entropy as the sum of the entropies of its infinitesimal parts. The law of entropy growth then applies to an expanding universe composed of self-gravitating systems and radiation. But one cannot infer from this extended law of entropy growth that the universe is tending toward the unchanging, featureless state of global thermodynamic equilibrium ("heat death") envisioned by Clausius and Kelvin in the mid-nineteenth century, because as discussed above, the expansion of the cosmic medium and the contraction of bounded self-gravitating systems drive local conditions away from thermodynamic equilibrium.

Clausius extrapolated the law of energy conservation and the law of entropy growth (the first and second laws of thermodynamics) from macroscopic systems and processes to the universe as a whole. Neither extrapolation is valid.

The law of energy conservation does not apply in the uniformly expanding space predicted by Friedmann's cosmological solutions to Einstein's field equations. For example, if the cosmic medium is a uniform ideal gas, the theory predicts that every particle slows down relative to its local standard of rest; its momentum and its kinetic energy both decrease as the medium expands. The energy of an ideal-gas sample likewise decreases with time, though it does no work on its surroundings.

The thermodynamic law of entropy growth applies in a much narrower domain than the law of energy conservation: Clausius's definition of entropy applies only to systems in local thermodynamic equilibrium. Boltzmann's definition of statistical entropy is far more general. Statistical entropy is a property of the probability distribution of microstates that characterizes a macrostate. And if such probability distributions have an objective character, as I argue below, statistical entropy is just as objective as thermodynamic entropy. Yet Boltzmann's H theorem cannot be viewed as an instance of a universal law, because its derivation depends on the assumption that information associated with molecular correlations is permanently absent. This assumption in turn follows from an initial condition (that correlation information is absent), a boundary condition (that nominally isolated gas samples actually interact with random environments), and a plausible but not rigorous physical argument (that correlation information flows from a sample to its random environment). As mentioned above, analogues of Boltzmann's *H* theorem for other macroscopic systems rely on analogous initial and boundary conditions and an analogous argument about the role of the environment.

Because experiments confirm the predictions of Boltzmann's *H* theorem and its analogues, we can infer that the initial and boundary conditions on which the derivations of these theorems rest are ordinarily satisfied: the probability distributions that characterize macrostates of newly formed – or newly prepared – macroscopic systems ordinarily lack correlation information; and these systems *ordinarily* have random surroundings. These initial and boundary conditions have a quasi-universal character. They are ordinarily but not *necessarily* satisfied.

> E. L. Hahn's spin echo experiment[34] shows that when appropriate kinds of correlation information are present initially and are sufficiently resistant to degradation by random interactions, a random distribution of microstates can evolve into a highly nonrandom distribution. The microstates in question are orientations of magnetic moments of nuclei in a macroscopic liquid sample. The sample is in an applied magnetic field whose direction is the same throughout the sample but whose magnitude has a small position-dependent random component. The state of the collection of magnetic moments is characterized by the joint distribution of their orientations and positions. The collection is prepared in a state in which the magnetic moments are all accurately parallel (or antiparallel) to a direction perpendicular to the direction of the applied magnetic field. The joint distribution of orientations and positions then contains a large quantity of orientation information and virtually no information associated with orientation-position correlations. The magnetic field exerts a torque on each magnetic moment, causing it to rotate in a plane perpendicular to the di-

34 Hahn, E. L., 1950, *Phys. Rev*. 80, 580

rection of the field. Owing to the random component of the applied field, the magnetic moments rotate at slightly different rates, gradually getting out of alignment. During this part of the experiment orientation information is converted into correlation information. Eventually the orientations of the magnetic moments are randomly distributed in directions perpendicular to the direction of the applied magnetic field; the orientation information that was present initially has all been converted into correlation information. In a gas sample, correlation information produced by the decay of single-particle information is quickly dispersed by molecular interactions. In the spin echo experiment it remains localized and is amenable to experimental manipulation. An ingenious experimental intervention now reverses the flow of information, converting the correlation information back into orientation information. The process just described – the conversion of orientation information into correlation information and back again into orientation information – is accompanied by the ordinary entropic decay of single-particle information through particle-particle interactions, but on a time scale significantly longer than that of the "anti-entropic" process.

To sum up, I have argued that the thermodynamic law of entropy growth does not apply beyond its original domain: isolated macroscopic systems in local (or global) equilibrium. In particular, it does not apply to self-gravitating systems. Boltzmann's transport equation, his H theorem, and their generalizations (master equations, generalized H theorems) apply to macroscopic systems that are not in local thermodynamic equilibrium, but they, too, are not laws. They rest on initial and boundary conditions that are ordinarily, but not necessarily, satisfied by both natural and prepared systems. This, I will now argue, is a consequence of the simplest account of the structure and evolution of the universe that is consistent with our most fundamental and most highly confirmed physical laws.

XI. Initial and boundary conditions; the prevalence of chance

The initial and boundary conditions that characterize physical systems are products of historical processes. We can think of these processes as episodes in a history of the physical universe. Of course, we are not yet able to construct, or even sketch, a complete history of the physical universe. The fragmentary history I propose rests on two assumptions: the strong cosmological principle; and the assumption that at some early time the cosmic medium closely approximated a uniform, uniformly expanding distribution of free particles in local thermodynamic equilibrium.

A full history would ground the second assumption in antecedent initial conditions and in physical laws that contain fewer unexplained constants than our current laws and cosmological models. But if both assumptions should turn out to be correct, a full account would preserve the distinctive features of the present account:

1. The classical variables that figure in Einstein's description of the structure and contents of spacetime are to be interpreted as random variables – mathematical objects characterized not by a definite value at each point of space-time but by a set of possible values and corresponding probabilities. We can interpret these probabilities as relative frequencies, or proportions, in infinite samples whose members are randomly distributed throughout space. For example, the probability that the mass density at a point lies in a given range of values is the fraction of points in a uniformly and randomly distributed sample of points at which the mass density lies in that range; the joint probability that the mass densities at two points with a given separation lie in given ranges is the fraction of a sample of pairs of points that have the given separation in which the mass densities at the two points lie in the given ranges, and so on.

As discussed below, this interpretation of Einstein's description of spacetime and its contents resolves the prima facie conflict between the deterministic character of Einstein's field equations and the fact that quantum measurements alter the macroscopic structure of spacetime unpredictably.

2. The probability distributions of microstates that characterize early states of the universe contain little or no statistical information per unit mass. As the universe expands, macroscopic processes create

information or change its qualitative character (through processes that always generate statistical entropy). But the quantity of information per unit mass remains far smaller than its largest allowed value. Thus randomness prevails.

3. The initial and boundary conditions that characterize macroscopic systems and processes are expressed by probability distributions of microstates, which in turn are determined by their history.

4. Such histories usually determine the values of macroscopic mechanical and thermodynamic variables but do not usually create information associated with persistent micro-level information (though as the spin echo experiment illustrates, they can do so).

Theories that describe irreversible macroscopic processes rest on instances of this generalization. This remark explains why the arrow of time defined by varied macroscopic processes in nominally isolated macroscopic systems coincides with the arrow defined by the cosmic expansion.

5. As discussed below, many macroscopic processes other than quantum measurements have indeterminate outcomes. The present account of chance resembles in important ways an account given a century ago by Henri Poincaré[35] in a popular essay. Poincaré asked why the outcomes of certain deterministic processes seem to be correctly predicted by "the laws of chance." As his first example Poincaré considered an ideal cone initially balanced on its tip. Imprecision in its initial positioning and tiny uncontrollable external disturbances cause the cone to topple in an unpredictable direction. But if the experiment is repeated many times, the final azimuth of the cone's axis will be smoothly (though not necessarily uniformly) distributed between 0 and 2π radians. In this example a deterministic law maps small differences between initial values of the azimuth of the cone's axis onto large differences between its final values. The smooth distribution of final azimuths requires only that the initial azimuths be smoothly distributed over a narrow subrange of their possible values.

In the 1880s Poincaré discovered the phenomenon now called deterministic chaos. The outcomes of chaotic processes depend sensitively on their initial conditions. In the discovery context small dif-

35 Poincaré, H. "Chance" in in *Science and Method* (New York, Dover, 2003)

ferences between the initial conditions of test particles in a gravitating system may cause their orbits to diverge at an exponential rate. If the initial values of the parameters that define an orbit are smoothly distributed over a small subrange of their possible values, the possible values of these parameters at a later time will be smoothly distributed over the entire range. Examples of chaotic processes are legion, ranging from meteorology to biology.

Poincaré argued that the initial conditions that characterize the cone balanced on its tip as well as those that characterize chaotic orbits in the solar system are in fact smoothly distributed on very small scales because historical processes have smoothed out irregularities on the smallest scales. The present historical account of initial and boundary conditions suggests a closely related but somewhat simpler explanation: The experimental setup that creates the initial state of Poincaré's cone specifies a probability distribution of initial conditions that does not contain enough information to specify the cone's final azimuth. Similarly, the historically determined probability distribution that characterizes the initial position and velocity of an asteroid in a chaotic orbit does not contain enough information to specify the asteroid's position after a lapse of 4.5 billion years.

We can define a classical microstate of Poincaré's cone, in part, by the azimuth of its axis. We can define the cone's macrostates, in part, by the precision of a given measuring apparatus. Initially the cone is in a macrostate in which the azimuth of its axis doesn't have a definite value, but as the cone's angle of tilt increases, the number of experimentally distinguishable azimuths – and hence the number of distinguishable macrostates – increases. Analogously, the orbit of an asteroid may be sensitive to small changes in its initial position and velocity. A historical account characterizes the initial state by a joint probability distribution of positions and velocities, which evolves into a distribution that characterizes a multitude of observationally distinguishable orbits.

To accommodate such situations we need to modify the rule that links probability distributions of (classical or quantum) microstates to classical macrostates. The standard rule equates the value of a macroscopic variable in a given macrostate to the result of averaging the corresponding microscopic variable over the probability distribution of microstates that represents the given macrostate. We modify it in three ways.

First, we characterize macrostates by experimentally distinguishable ranges (or aggregates) of microstates, as in the above examples. A probability distribution of microstates may then represent two or more experimentally distinguishable macrostates.

Second, we equate the result of averaging a microscopic variable over such a probability distribution to the result of averaging the measured value of the corresponding macroscopic variable over a "large number" of replicas of the measurement.

Finally, to incorporate into our rule the fact that neither physical laws nor initial and boundary conditions that comply with the strong cosmological principle serve to define a particular position, we interpret the set of replicas mentioned in the preceding paragraph as a "cosmological ensemble" – a set of replicas randomly and uniformly distributed throughout an infinite space. (Like Gibbs's ensembles, a cosmological ensemble is made up of imaginary replicas. But each replica in a cosmological ensemble is in a definite macrostate. And cosmological ensembles have a physical interpretation: they allow us to express the assumption that physics cannot make unconditional predictions about where in the universe given measurement outcomes are realized.)

These rules enable us to calculate the probabilities of experimentally distinguishable measurement outcomes from measurements of mean values: Following an argument given by Dirac[36] in a related context, let V denote a macroscopic property whose possible values are real numbers. Let the index k label the possible outcomes of a measurement of V and let the index r label replicas in a cosmological ensemble. Let $I(V, k, r)$ be the function of V, k, and r that is equal to 1 if a measurement of V at the rth replica has the kth outcome and is equal to 0 otherwise. The value of $I(V, k, r)$ averaged over the members of a cosmological ensemble is the fraction $f(V, k)$ of replicas for which a measurement of V has the outcome k. We can think of the set $\{k\}$ of outcomes as a sample space and the set of fractions $\{f(V, k)\}$ as a set of probabilities on this sample space.

36 Dirac, P.A.M., Reference cited in Note 2, p. 47

XII. Quantum measurement

The preceding rule for linking a probability distribution of (classical or quantum) microstates to the possible outcomes of a measurement and their probabilities applies to quantum measurements. The isolated macroscopic system now consists of a quantum system one of whose properties we wish to measure, a macroscopic measuring apparatus that interacts with the quantum system, and a bounded random environment[37] that interacts with the measuring apparatus. We assume, as in decoherence calculations, that this system has quantum states that evolve in accordance with Schrödinger's equation. But we do not make the customary assumption that the system is initially in one or another of its microstates. We assume instead that it is in a macrostate characterized by a probability distribution of its microstates. Application of the preceding rule then reproduces the measuring postulate of the standard formulation without further ado: it predicts that ideal measurements have definite outcomes given, along with their relative frequencies in a cosmological ensemble, by the measuring postulate.

XIII. QM and GR

As mentioned earlier, general relativity's deterministic description of the evolution of space-time structure clashes with the fact that quantum measurements affect the local structure of space-time in unpredictable ways. The present account dissolves this contradiction. From a macroscopic standpoint the unpredictability of the post-measurement position of a pointer in a quantum measurement is no more problematic than the unpredictability of the final orientation of Poincaré's cone. In both cases macroscopic unpredictability results from an objective absence of information in the probability distribution of microstates that characterizes the initial state of an isolated system. In both cases a deterministic law – Schrödinger's equation in the first case, Newton's laws of motion and gravitation in the second – governs the evolution of the system's microstates.

37 The random environment could consist of the microscopic degrees of freedom of the macroscopic measuring apparatus.

XIV. The irreducibility of macrophysics and the unity of physics

The holy grail of physics is a Theory of Everything. Such a theory would include as limiting cases our present strongly confirmed laws and would contain far fewer adjustable physical constants than figure in these laws. As I have already emphasized, accounts of physical systems and processes depend on initial and boundary conditions as well as laws; and it has long been understood that laws and initial/boundary conditions are not entirely distinct categories. To derive macrophysical laws such as Boltzmann's *H* theorem and its generalizations from more fundamental microscopic laws one needs to impose appropriate initial and boundary conditions. In this essay I have argued that these conditions are products of a historical process whose description rests on simple cosmological initial conditions and a strong version of the cosmological principle.

The account I have sketched of this historical process knits together our present laws in other ways as well. It shows how initial and boundary conditions link the temporal direction of macroscopic processes to the direction of the cosmic expansion, it offers a simple and direct answer to the question of why quantum measurements have definite but unpredictable outcomes, and it reconciles the unpredictability of quantum measurement outcomes with the deterministic character of Einstein's field equations.

Besides joining these loose ends, a historical account of initial conditions offers a new view of the role of chance in macroscopic processes. Physicists have conventionally held that the outcomes of macroscopic processes other than quantum measurements are predictable in principle. Some, though not all, evolutionary biologists have taken issue with this doctrine, which also seems to be at odds with judgments based on ordinary experience. But physics as conventionally interpreted assures us that to a contemporary version of the omniscient mind posited by Laplace in his essay on chance, nothing except quantum measurement outcomes would be unpredictable. The historical account of initial conditions sketched in this essay supports the contrary view suggested by evolutionary biology and experience: much of what we observe in the world around us is influenced by chance. This generalization applies not just to aspects of our physical environment, like weather. As discussed in a little more detail below, randomness plays an essential role in the biological world.

XV. Is biology a part of physics?

Some physicists consider physical theories to be nothing more than devices for linking measurements to other measurements. Others – realists – regard our present theories as descriptions, perhaps partial or approximate, of a unified mathematical structure behind experience. The second view, which was held by Einstein, draws support from the history of physics. Strongly confirmed theories have not been overturned by their successors. They have remained in place as limiting cases, valid in circumscribed domains, of the successor theories.[38] And as the scope of physical theories and the accuracy of their predictions has increased, the fundamental theories have become fewer, more comprehensive, and more abstract. History thus supports the view that our present physical theories capture, or at least approximate mathematical regularities behind experience and that these regularities belong to a unified mathematical structure.

What characterizes the objects and processes that belong to the world that physics describes and physicists try to understand? Consider atoms. Two and a half millennia ago Leucippus and Democritus tried to link the sizes and shapes of hypothetical atoms to observed properties of bulk matter. Newton in the *Principia* tried to account for Robert Boyle's empirical law relating the pressure and the volume of an enclosed sample of air by positing air atoms moving and interacting in ways governed by his laws of motion. Half a century later, Daniel Bernoulli introduced a much simpler atomic model of air, and in the nineteenth century Rudolf Clausius and James Clerk Maxwell significantly extended Bernoulli's model. But Maxwell realized that Newtonian physics and his own theory of electricity and magnetism could not explain the observation that atoms always absorb and radi-

---

38 Thomas Kuhn, argued in *The Structure of Scientific Revolutions* (1959) that a successor theory overturns its predecessor because meanings of scientific terms common to the two theories are incommensurable. But physical theories are mathematical constructs; they are not verbal-conceptual constructs that have been made more precise through the use of mathematical "language." The axioms of Newtonian mechanics are neither inconsistent nor incommensurable with the axioms of relativistic mechanics; in a precise and completely describable sense they approximate those axioms.

ate light at a fixed set of frequencies. Meanwhile Ernst Mach argued that the atomic hypothesis was methodologically unsound because it invoked unseen entities. Physical theories, in his view, should seek to represent, rather than explain, experience.

In his 1905 paper on Brownian motion Einstein invoked a different criterion for physical hypotheses: falsifiability (as Popper later called it). If his predicted relation between the motions of a liquid's hypothetical molecules and the observable motions of microscopic particles suspended in the liquid should be shown to be incorrect, he wrote, this would "provide … a weighty argument … against the molecular-kinetic conception of heat [i.e., the atomic hypothesis]." Experimental confirmation of Einstein's law, which came a few years later, strengthened the case not only for the atomic hypothesis but for a way of doing theoretical physics that relies more heavily on mathematical invention and the testing of theoretical predictions than on the analysis of facts.

At the other end of the size scale, the physical universe crossed the boundary that separates physics from metaphysics in two steps. In 1915 Einstein published a theory of gravitation that applies to an unbounded, statistically uniform distribution of mass, and in the 1920s Edwin Hubble supplied observational evidence that the astronomical universe is indeed unbounded and statistically uniform.

In short, a combination of mathematical invention, experiment, and observation shapes a physical realist's conception of the physical world. The question "Is X a constituent of the physical world" can be rephrased as "Does X figure in a mathematical theory that is tightly linked to fundamental physical theories and is strongly confirmed by experiment or observation?" According to this criterion, quarks are constituents of the physical world, while "dark energy" is not – or at least not yet.

What about living organisms and biological processes? Living organisms are physical systems, because they are made up entirely of atoms and molecules drawn from the nonliving environment; and biological processes are physical processes, because they obey the same physical and chemical laws as nonliving systems. Yet living organisms and biological processes are not just physical systems and processes. They have a distinctive character, which they owe entirely to their distinctive initial and boundary conditions.

Like the initial and boundary conditions that characterize nonliving systems, those that characterize living organisms and biological processes were shaped by history. Although the opening chapter of the history of life exists only in rough, competing drafts. the authors of these drafts agree that life arose by chance in a nonliving environment through processes governed by well-understood physical and chemical laws. Can we then conclude that biology is at bottom a branch of physics, like condensed-matter physics and astrophysics? Do biological systems and processes belong to the world that physics describes or could describe? Or, as Ernst Mayr[39] and other biologists have argued, is biology an autonomous science?

Biology has a number of terms that do not appear in the physical sciences, such as *function, fitness, purpose, adaptation*. Can such terms be explained, however clumsily, in the language of physics and chemistry, augmented if necessary by explicit definitions?

Take *function*. Molecular physics and chemistry supply detailed accounts of the physical structure and chemical properties of hemoglobin. For example, they explain its capacity to bind oxygen molecules. But molecular physics and chemistry alone cannot tell us that in vertebrates the biological role of hemoglobin depends on its affinity for oxygen. *Bio*chemistry continues the chemical story. It seeks to understand not only how hemoglobin performs its biological function but also how the molecule and its function have evolved from simpler precursors. Can this continuation of the chemical story be recast in the language of physics and chemistry?

Part of it is already in that language. The chemical processes that involve or depend on hemoglobin belong to the common subject matter of chemistry and biochemistry. The other part of the story involves the notion of fitness. Changes in the structure of hemoglobin that affect its ability to bind and release oxygen molecules under specific environmental conditions affect an animal's prospects for survival and reproduction. Fitness is a measure of these prospects.

---

39 Mayr, Ernst, "How To Carry Out The Adaptationist Program?" in *Toward a New Philosophy of Biology*, (Cambridge, Harvard University Press, 1988) p. 159

And there's the rub. Population geneticists have defined fitness in various ways, but all the definitions are prospective; they all refer to the future.

In a given population, genetic changes that have a significant random component give rise to variants of hemoglobin. *If the population's environment doesn't change for a sufficiently long period of time*, heritable variants whose oxygen affinity is optimal for the given, unchanging conditions will gradually come to dominate the population's gene pool: the possessors of sub-optimal variants will have fewer descendants than the descendants of possessors of optimal variants.

This example illustrates an essential aspect of evolution: the emergence and subsequent fixation of new or modified traits through random genetic variation and natural selection *in an unchanging environment*.

The example of hemoglobin evolution in an unchanging environment resembles the "evolution" of a cone initially balanced, imperfectly, on its tip. The fate of any given molecular variant is predictable; so is the path of a cone whose initial orientation and angular momentum have been specified. The cone's initial state is characterized by a probability distribution of micro-conditions; analogously, one might perhaps be able to assign a probability per unit time to the appearance of each possible variant of the hemoglobin molecule and then go on to predict the relative frequencies of variants after many generations. So the story of hemoglobin evolution in an unchanging environment can perhaps be recast in language familiar to physicists and chemists.

The emergence of evolutionary novelties poses a more formidable challenge to the view that evolutionary stories can be recast as physic-chemical stories. Evolutionary novelties are the results of a creative process:

> Evolution is a creative process, in exactly the same sense in which composing a poem or a symphony, carving a statue, or painting a picture are creative acts. An artwork is novel, unique, and unrepeatable; ... The evolution of every phyletic line yields a novelty that never existed before and is a unique, unrepeatable, and irreversible proceeding.[40]

40 Dobzhansky, Theodosius, *Genetics of the Evolutionary Process* (New York, Columbia University Press, 1970)

Evolutionary novelties are also unpredictable:

> Evolutionary change in every generation is a two-step process, the production of genetically unique new individuals and the selection of the progenitors of the next generation. The important role of chance at the first step, the production of variability, is universally acknowledged (Mayr 1962), but the second step, natural selection, is on the whole viewed rather deterministically: Selection is a non-chance process. What is usually forgotten is the important role chance plays even during the process of selection. In a group of sibs it is by no means necessarily only those with the most superior genotypes that will reproduce. Predators mostly take weak or sick prey individuals but not exclusively, nor do localized natural catastrophes (storms, avalanches, floods) kill only inferior individuals. Every founder population is largely a chance aggregate of individuals, and the outcome of genetic revolutions, initiating new evolutionary departures, may depend on chance constellations of genetic factors.[41]

According to the central argument of the present essay, the random element in both genetic variability and natural selection is objective and irreducible. In physical contexts the randomness inherent in initial conditions entails that many kinds of macroscopic processes (besides quantum measurement) have objectively unpredictable outcomes; but the *possible* outcomes of such processes are predictable, along with their probabilities; and we may interpret the probability that attaches to a particular outcome or range of outcomes as its relative frequency in a cosmological ensemble. In the context of evolution, the possible outcomes – evolutionary novelties – are themselves unpredictable. Even if these outcomes and their probabilities could be predicted – a feat that would perhaps require the expenditure of more free energy than the Sun could supply in its lifetime – the prediction would be useless, for it would assign an infinitesimal probability to the novelties that have actually been produced in the course of evolution.

To sum up, although living organisms are physical systems and biological processes obey physical laws, life and its history are not part of the world current physical theories describe. Life is a natural

41 For a more detailed version of the following argument see Layzer, "Naturalizing Libertarian Free Will" (previous chapter)

phenomenon, firmly anchored in the physical world and its laws; but the initial and boundary conditions that characterize living systems and biological processes ensure that the history of life is creative and hence, in a way that transcends physical indeterminacy, unpredictable.

XVI. The problem of free will

Defenders of libertarian free will usually grant at the outset that events other than the outcomes of quantum measurements are determined by universal physical laws and antecedent conditions. They must then explain how it can be that we are able to shape the future through our choices and decisions. In this essay I have argued that the premise is false: Events in the macroscopic world are not determined by universal physical laws and antecedent conditions; a wide class of macroscopic processes have indeterminate outcomes. And if the processes involved in reflective choice belong to this class, there is no scientific reason why we should not accept the proposition that we shape the future through our choices and decisions.

Biology supplies a strong positive case for libertarian free will.38 As mentioned earlier, Mayr and other evolutionary biologists have stressed the central role of chance in evolution. Genetic variation has a random component, but if genetic variation were entirely random, complex adaptations could never have evolved. The evolution of complexity requires genetic regulation of the ways in which chance manifests itself in genetic variation.

It is easy to see why. A central and strongly confirmed tenet of evolutionary theory is that complex organs such as eyes evolved from less complex but fully functional predecessors. These predecessors themselves evolved from less complex but also functional predecessors, and so on until we arrive at “a simple light sensor for circadian (daily) and seasonal rhythms around 600 million years ago” At each stage of this multistage process genetic modifications that improved the eye’s function emerged through the usual combination of genetic variation and selection. But for the process to work, the genetic variations that occur at each stage must not significantly impair the function achieved at that stage. This means that at each stage the genes and gene combinations that encode the eye’s developmental program must be held safe from harmful variation – not from all variation, just harmful variation. And indeed experiments show that genetic variation is suppressed to varying degrees at different genetic loci. For example, proofreading and error-correction processes suppress transcription errors (which when not suppressed are a source of variability); exchanges of genetic material between homologous chromosomes during meiosis are nonrandom in ways that preserve gene combinations whose disruption would lower fitness while allowing others. Molecular mechanisms that regulate and channel genetic variation are themselves products of evolution's two-stage process.

Open behavioral programs enable animals, including single-celled animals, to thrive in environments that change in unpredictable ways. All animals learn from experience. They tend to repeat behaviors for which they have been rewarded in the past and to avoid behaviors for which they have been punished. Some animals also learn from experience in ways that allow for risk-taking, exploratory behavior, and delayed rewards. Economists, students of animal behavior, and cognitive neuroscientists have developed algorithms that seek to mimic such flexible learning strategies, and have constructed hypothetical neural networks that instantiate algorithms of this kind.[42]

In light of what ethologists have learned about the behavior of monkeys and apes, we can plausibly conjecture that much of human learning and decision-making is mediated by neural architecture that embodies complex and sophisticated algorithms of this kind. But humans have an extra, qualitatively different capacity for learning and decision-making:a capacity for reflective choice. We are able construct mental representations ofscenes and scenarios that are not directly coupled to external stimuli (as in perception) or to movement (as in reflexes). We call on this capacity when we imagine possible courses of action and then go on to imagine the possible consequences of each of these invented candidates for choice. Our ancestors used it when they painted pictures on the walls of their caves and created the first human languages. We use it when we compose an original sentence or a tune or when we try to solve an abstract problem. It allows us to reconstruct the distant past from

42 Lamb, T. D. "Evolution of the Eye," *Scientific American*, July 2011

fragmentary evidence and to envision the distant future. It makes possible the life of the mind.

The brain of every animal contains a model (or "theory") of the world – a hierarchical set of schemata that regulate the animal's behavior. Some of these schemata can be modified by the animal's experience. According to the psychologist Thomas Suddendorf,[43] children exhibit "the ability to entertain and collate offline mental models (e.g., about past, future, or imaginary situations) in addition to the primary reality model" around the age of two. He cites behavioral evidence that great apes, *but not monkeys*, also have this capacity. Around age four, children demonstrate a new capacity: they become able to understand representations *as* representations. Their world models begin to include a "theoretical" component: a set of beliefs about animals, gods, weather, the Sun, the stars, other people, and ourselves. Although great apes have "offline" mental models, Suddendorf writes, they show no behavioral evidence for this second capacity – the capacity to understand that mental representations are products of their own imagining. Creative thought requires the second capacity. We have it; according to Suddendorf, nonhuman primates do not. So free will in the strong or libertarian sense is a distinctively human biological capacity. Like evolution itself, it harnesses chance in the service of creativity.

XVII. Free will, consciousness, and the brain

How does free will fit into a scientific picture of the world? The preceding account of free will rests on psychology and, more broadly, on biology. I have argued that the capacity to invent and evaluate possible courses of action is a distinctively human biological adaptation whose precursors are found in other animals. I have also argued that biology is not only different from physics in the ways that Mayr and other evolutionary biologists have discussed; it is also not reducible to physics. Fitness, for example, is a property of genes or gene combinations that encode particular developmental programs in members of a particular population; but for the reasons I have discussed, it is not a *physical* property. Talk about the *functions* of biomolecules, organs, and behavioral traits, is a necessary (and perfectly

43 For recent reviews, see *Nature Neuroscience*, Volume 111, Number 4, April 2008,

scientific) part of biological discourse, but it cannot be paraphrased in the vocabulary of the physical sciences.

Nor can talk about mental models and the conscious aspects of creative decision-making be translated into talk about neural circuitry and neural processes. Neuroscience seeks to understand the biological underpinnings of mental states and of conscious and unconscious mental processes, but as Max Bennett and P.M.S. Hacker[44] have argued in considerable detail and with great clarity, mental states and processes are not reducible to brain states and processes. Just as biology rests on but cannot be reduced to physics, psychology rests on but cannot be reduced to neuroscience. Nevertheless mental states and processes can be studied and at least partially understood by the methods of psychology and comparative psychology.

Do conscious acts of will cause our voluntary actions? From a thorough examination of the evidence bearing on this question the psychologist Daniel Wegner[45] has concluded that the answer is no. "Conscious will arises from processes that are psychologically and anatomically distinct from the processes whereby mind creates action [p. 29]." This conclusion accords well with the arguments and conclusions of the present essay. I have argued, as Henri Bergson did a century ago, that we act most freely when we act most creatively. Whether conscious acts of will are essential features of the extended mental processes involved in reflective decision-making is an empirical question. My own experience, for what it is worth, suggests that conscious acts of will play at most a minor part in reflective decision-making. What seems undeniable is that we believe we can alter the course of events through our plans and projects. This essay has argued that that is not an illusion.

17 August 2011

44 Bennett, M. R. and Hacker, P.M.S., *Philosophical Foundations of Neuroscience*, (Oxford, Blackwell, 2003)

45 Wegner, Daniel M., *The Illusion of Conscious Will* (Cambridge, MIT Press, 2002)

Afterword

As two researchers who knew well and were deeply influenced by David Layzer's life and work, we are honored to have been able to help his family in assembling, editing, and communicating his final book.

David Layzer believed in and argued for the existence of true novelty, creativity, and freedom in the physical world, and against a worldview of *physicalism* that he saw as regarding all of these as convenient fictions. In that worldview everything we see around us today was "already contained" within the beginning state of the Universe, bound to that early state by an inevitable and barren unfolding via mathematical physical laws.

In this work and others he argued that the flaw in this physicalist view is *not* in the mathematical laws of nature that partly underly it, but rather in the historical and cosmological context in which those laws operate. His postulated Strong Cosmological Principle implies that chance is inevitable: a fully-detailed description of the Universe simply does not exist – or perhaps more accurately, the most-detailed description is *statistical* in the sense that no matter how much data is specified about some region of the Universe, there will be predictive questions with only probabilistic answers. This applies at all times, and only statistical statements about the current universe can possibly follow from its early state. Importantly, while it is widely believed that quantum mechanics implies (some sense of) objective chance, Layzer saw the objective chance inherent in the Strong Cosmological Principle to be the *basis* rather than the *result* of quantum uncertainty, forming a completely original way of viewing the quantum measurement problem that has been taken up by other researchers including one of us (AA).

If the forms and structures inhabiting the current Universe were not (implicitly) present at its beginning, how did this information arise? In this question Layzer has provided insight after insight. He appears to be among the first who crisply connected non-equilibrium processes in cosmology to the formation of chemical and other order – which is now a somewhat textbook view. But he also pointed out how this can be viewed as a competition between entropy generation in a system versus the increase in the maximal possible entropy of that system; an expanding universe provides a widening gap between the two that is cosmological order. This insight has still not been fully concretized in thinking about the statistical mechanics of the Universe, though work in that direction is happening, including by one of us (AA). The cosmic store of order is the grist for the creation of higher levels of order, through astrophysical structure formation, chemical evolution, biological evolution and then neural systems; these combine to transform some of that cosmologically-generated reservoir of order into all of the artifacts we value.

How this creative transformation takes place in the evolutionary and neural domains occupied Layzer through much of his later life, but *Why We are Free* provides only a very compact view, so it is worth here bringing forward a bit more material here, pointing to further sources.

For biology, he approvingly quotes Theodosius Dobzhansky:

> "A living organism resembles a work of art, and the evolutionary process resembles the creation of a work of art."[1]

Layzer's insight into biological evolution endorses the role of chance in the creation of random genetic variations, but he sees subsequent natural selection as a non-chance process. He quotes evolutionary biologist Ernst Mayr:.

> Evolutionary change in every generation is a two-step process: the production of genetically new individuals and the selection of the progenitors of the next generation. The important role of chance at the first step, the production of variability is universally acknowledged, but the second step, natural selection, is on the whole viewed rather deterministically: Selection is a non-chance process.[2]

1 Dobzhansky, T., *Proc. Ninth Int. Congress Genetics* (1954), pp. 435-449

2 Mayr, E., *Toward a New Philosophy of Biology*. (Cambridge, Harvard University Press, 1988) p. 21

And Layzer writes

> Randomness is the raw material from which processes governed by deterministic mathematical laws fashion myriad novel varieties of physical and – especially – biological order.[3]

For Layzer, chance again plays an essential role in freedom of the will, but it does not make our choices and decisions themselves random, any more than biological evolution created random life forms. Just as randomness seeds genetic variation that is then selected for fitness, random chance simply generates multiple alternative possibilities for thoughts and actions.

In this way Layzer draws a parallel between biological evolution, a two-step process as Ernst Mayr described it, and free will, which has also been described as a two-step process by many thinkers since at least William James in the 1880's.[4] One of us (BD) has identified nearly two dozen philosophers and scientists who have proposed or endorsed two-stage models of free will since James.[5] James was also the first thinker to draw the parallel between Darwinian evolution (then a new idea), freedom of the will, and chance.

In an unpublished manuscript, "Naturalizing Libertarian Free Will," Layzer explained...

> It entails a picture of the physical universe in which chance prevails in the macroscopic domain (and hence in the world of experience). Because chance plays a key role in the production of genetic variation and in natural selection itself, evolutionary biologists have long advocated such a picture. Chance also plays a key role in other biological processes, including the immune response and visual perception. I argue that reflective choice and deliberation, like these processes and evolution itself, is a creative process mediated by indeterminate macroscopic processes, and that through our choices we help to shape the future.[6]

But Layzer makes clear that the ultimate decision or choice between alternative possibilities is in no way itself random, but an act of self-determination:

> To be fully human is to be able to make deliberate choices. Other animals sometimes have, or seem to have, conflicting desires, but we alone are able to reflect on the possible consequences of differ-

3 See chapter VII - Chance and Order.
4 See Doyle, R.O. "Jamesian Free Will," in *William James Studies* (2010) Vol. 5, p. 1
5 See https://informationphilosopher.com/books/scandal/Two-Stage_Models.pdf.
6 See https://informationphilosopher.com/solutions/scientists/layzer/Naturalizing_Libertarian_Free_Will.doc. (2010) p.2.

> ent actions and to choose among them in the light of broader goals and values. Because we have this capacity we can be held responsible for our actions; we can deserve praise and blame, reward and punishment. Values, ethical systems, and legal codes all presuppose freedom of the will...
>
> A decision is free to the extent that it results from deliberation. Absence of coercion isn't enough. Someone who bases an important decision on the toss of a coin seems to be acting less freely than someone who tries to assess its consequences and to evaluate them in light of larger goals, values, and ethical precepts.11[7]

This makes clear that in Layzer's view not all of the Universe is free — that is a capability enjoyed by only a very special set of systems — like human minds — far up the hieheirarchy of complexity. But in his view all of the Universe is creative. From the creation of chemical order in the early universe to the creation of biological and mental novelty, new structures are continually coming into being, and allowing others to come into being. This notion extends even to mental creations such as the mathematical laws of physics that we use to describe the Universe itself. Layzer often quoted Einstein's famous observation that scientific theories are "free creations of the human mind." When one of us (AA) once asked him how to reconcile this with the notion that mathematical statements are true (or not) long before being discovered by people, he gave a beautiful response: many, many, many statements in mathematics follow from a given set of axioms. But this set of "all possible true statements" is *information free* — just like a physical system in equilibrium. Human mathematicians and physicists *select* particular ones from that giant set. In doing so, we create information that did not exist before.

Layzer's work was deeply novel itself, and in many ways unconventional. We hope that this volume might help others to discover and build on some of his many insights, continuing to unfold the creative process he began.

Anthony Aguirre, UC Santa Cruz

Bob Doyle, Harvard University

January, 2021

7 See https://informationphilosopher.com/solutions/scientists/layzer/Free_Will_As_A_Scientific_Problem.pdf. (2011) p. 31

Bibliography

Aguirre, A. (1999) *Astrophysical Journal*, 521, 17-29.

Aguirre, A., and Tegmark, M. (2011) "Born in an Infinite Universe: a Cosmological Interpretation of Quantum Mechanics." *Phys. Rev. D* 84.10: 105002.

Ayala, F.J. and T. Dobzhansky., *Studies in the Philosophy of Biology*, Berkeley: University of California Press

Bergmann, P. J. and J. L. Lebowitz (1955) *Phys. Rev.* 99, 578

Blatt, J. M., (1959) *Prog. Theor. Phys.* 22, 745

Boltzmann, L. (1895) *Lectures on Gas Theory*, translated by Stephen G. Brush, Dover Publications

Cohen, E. G. D. (1962) *Proceedings of the NUFFIC International Summer Course in Science at Nijenrode Castle*, The Netherlands, August, 1961. Amsterdam: North-Holland Press

Crick, F. (1994) *The Astonishing Hypothesis*, New York: Simon and Schuster

Dirac, P.A.M. (1967) *The Principles of Quantum Mechanics*, fourth edition (revised), Oxford: Clarendon Press

Dobzhansky, T. (1954) *Proc. Ninth Int. Congress Genetics*, pp. 435-449

Dobzhansky, T. (1974) "Chance and Creativity in Evolution" in Ayala and Dobzhansky, p. 329.

Doyle, Robert O. (2010) "Jamesian Free Will," *William James Studies*, vol.5, p1.

Doyle, Bob (2011) *Free Will: The Scandal in Philosophy*, Cambridge: I-Phi Press

Eddington, A. S. (1958), *The Nature of the Physical World*, Ann Arbor: University of Michigan Press

Einstein, A. (1952) *Relativity, the Special and the General Theory*, 15th edition, New York: Crown Publishers, Einstein, A. (1923) "The Foundations of the General Theory of Relativity," in *The Principle of Relativity*, Methuen and Company, reprinted by Dover Publications.

Einstein, A. (1953) *The Meaning of Relativity*, 5th edition, Princeton: Princeton University Press

Fermi, E. (1937) *Thermodynamics*, New York: Prentice-Hall

Gerhart, J. and Kirschner, M. (2007) *Proceedings of the National Academy of Sciences*, May 15, 2007. 104 (suppl 1) 8582-8589

Gibbs, J. W. (1902) *Elementary Principles in Statistical Mathematics, Developed with Special Reference to the Rational Foundations of Thermodynamics*, New Haven: Yale University Press

Greenblatt, S. (2009) *The Swerve*, New York: Norton

Hahn, E. L. (1950) *Phys. Rev.*, 80, 580

Heisenberg, W. (1958) *Physics and Philosophy: The Revolution in Modern Science*, London: Allen & Unwin

Home, D. and A. Robinson, (1995) "Einstein and Tagore: Man, nature and mysticism", *Journal of Consciousness Studies* 2

Hubble, E. (1936) *The Realm of the Nebulae*, Yale University Press

Kane, R. (2002) *The Oxford Handbook of Free Will*, New York: Oxford University Press

Kirschner, M. W. and Gerhart, J. C. (2005) *The Plausibility of Life*, Yale University Press

Kirschner, M. W. and Gerhart, J. C. (2007) *Proc. National Academy of Sciences* 104, supp. 1, May 15, 2007

Laplace, Pierre Simon, (1951) *A Philosophical Essay on Probabilities*, translated into English from the French 6th ed. by Truscott, F.W. and Emory, F.L. New York: Dover Publications

Layzer, D. (1980) *The American Naturalist*, 115, 6

Layzer, D. (1990) *Cosmogenesis: The Growth of Order in the Universe*, New York: Oxford University Press

Lebowitz, J. L and P. J. Bergmann (1959) *Annals of Physics* 1, 1

Mayr, E. (1983) *The American Naturalist* 121: 324-33; reprinted in Mayr, 1988,

Mayr, E. (1988) *Toward a New Philosophy of Biology*. Cambridge: Harvard University Press

Mermin, N. D. (2014) "QBism puts the scientist back into science", *Nature*, 26 March 2014,

Nagel, T. (2012) *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature Is Almost Certainly False*, Cambridge: Belknap Press

Nagel, T. (2013) "The Core of 'Mind and Cosmos," *New York Times*, August 18, 2013

Norton, J. D. (2015) "The Hole Argument", *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2015/entries/spacetime-holearg/>

Penrose, R. (1989) *The Emperor's New Mind*. Oxford: Oxford University Press

Planck, Max. (1945) *Treatise on Thermodynamics*, trans. Alexander Ogg. Third edition translated from the seventh German edition. New York: Dover Publications

Ridderbos, T. M. and M. L. G. Redhead, *Foundations of Physics*. 28, 1237 (1988)

Schlosshauer, M. (2008) *Decoherence and the Quantum-to-Classical Transition* (corrected 2d printing) Berlin: Springer

Schlosshauer, M. (2014) "The quantum-to-classical transition and decoherence," *arXiv*:1404.2635v1 [quant.phys]

Schrödinger, E. (1948) *Statistical Thermodynamics*, Cambridge: Cambridge University Press

Searle, J. R. (2004) *Freedom and Neurobiology*, New York: Columbia University Press

Shannon, C.E. (1948) *The Bell System Technical Journal*, Vol. 27, pp. 379–423, 623–656, July, October, 1948

Thomson, W. (1848) *Philosophical Magazine*, October, 1848

Van Kampen, N. G. (2007) *Stochastic Processes in Physics and Chemistry*, 3d edition. Amsterdam: Elsevier

Wegner, D.M. (2002) *The Illusion of Conscious Will*. Cambridge: MIT Press

Weinberg, S. (1972) *Gravitation and Cosmology.* New York: Wiley

Weinberg, S. (2017) *New York Review of Books*, January 19, 2017

Weinberg, S. (1995) *The Quantum Theory of Fields*, Volume 1, New York: Cambridge University Press

Wigner, E. (1970) "Remarks on the Mind-Body Question" in *Symmetries and Reflections*, Cambridge: MIT Press

Wilson, E.O. (1998) *Consilience*. New York: Knopf

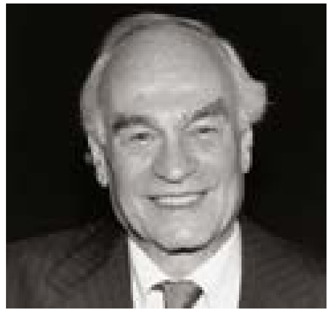
Index

J

K

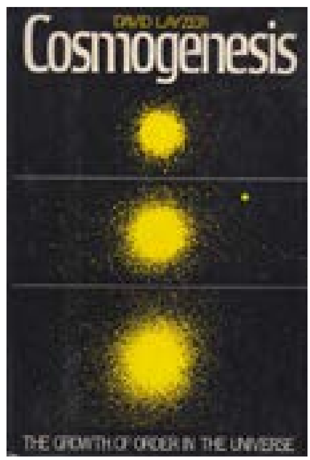
L

M



David Layzer was the first Donald H. Menzel Professor of Astrophysics at Harvard, and a pioneering interdisciplinary thinker in cosmology, astrophysics, information theory and biology. A wide-ranging and foundational thinker, his 1990 book *Cosmogenesis, The Growth of Order in the Universe*, laid out a compelling holistic understanding of cosmic evolution, structure formation, and the growth of information in a big-bang universe.

In *Why We Are Free*, his third and final book, Layzer distills a lifetime of thinking about the fundamental nature of order and cosmic evolution, to ultimately address one of the most basic questions: what is our place and nature, in a world governed by physical law? *Why We Are Free* crisply presents and rigorously defends a view "in which our joys and our sorrows, our memories and our ambitions, our sense of personal identity and free will, are just as real as the objects and relations of the world physics describes."



In *Cosmogenesis* Layzer described the *initial conditions* of the universe and how orderly structures like planets, stars, and galaxies could form despite the entropy increase of the second law of thermodynamics. In this book Layzer began his study of free will, connecting it to a "primordial randomness." He saw this randomness as an alternative to quantum indeterminism and as the source of freedom from philosophical determinism.

In *Why We Are Free*, Layzer writes, "the strong cosmological principle and the assumption of primordial randomness explain ...the irreversibility of the universe." They are also his basis for freedom of the will.

ISBN 978-0-9835802-5-6
51995
9 780983 580256

David Layzer on the web
www.davidlayzer.com
en.wikipedia.org/wiki/David_Layzer
informationphilosopher.com/solutions/scientists/layzer/