Thomas Dittrich

INFORMATION DYNAMICS

In Classical and Quantum Systems

Springer

The Frontiers Collection

The books in this collection are devoted to challenging and open problems at the forefront of modern science and scholarship, including related philosophical debates. In contrast to typical research monographs, however, they strive to present their topics in a manner accessible also to scientifically literate non-specialists wishing to gain insight into the deeper implications and fascinating questions involved. Taken as a whole, the series reflects the need for a fundamental and interdisciplinary approach to modern science and research. Furthermore, it is intended to encourage active academics in all fields to ponder over important and perhaps controversial issues beyond their own speciality. Extending from quantum physics and relativity to entropy, consciousness, language and complex systems—the Frontiers Collection will inspire readers to push back the frontiers of their own knowledge.

More information about this series at https://link.springer.com/bookseries/5342

Thomas Dittrich

Information Dynamics

In Classical and Quantum Systems

 Springer

Thomas Dittrich
Departamento de Física
Universidad Nacional de Colombia
Bogotá D.C., Colombia

Preface

"Information" is one of the most common technical terms in our times, the market abounds with literature on the subject—do we need yet another book on this subject? The present work grew out of the demand for a text accompanying a topical lecture on the concept of information in physics, at the graduate to postgraduate levels. The course, in turn, was and still is an experiment in teaching physics, an attempt to introduce a novel kind of interdisciplinary subject in a traditional science curriculum.

The lectures offered by a university or college physics department are usually characterized by a high degree of specialization and compartmentalization: "Theoretical Solid-State Physics", "Quantum Optics", "Experimental High-Energy Physics" are standard titles reflecting an organization of knowledge motivated by phenomenological criteria or by mere historical development. Correspondingly, the formation of a professional physicist is perhaps intended to be broad, but not primarily to be comprehensive. At the same time, there is an increasing conscience that a wider perspective, the ability to think "across the Two Cultures", to work in interdisciplinary teams, is a highly desired competence in the academy as well as in administration and industry. The demand is growing for courses that contribute to this facet of science education, oriented towards transversal concepts and broad syntheses. However, they mostly appear as separate parts of the curriculum, say under the label *studium generale*, hardly interlocked with the technical courses proper, and do not form a bridge between specialized and holistic teaching. Achieving it requires a reorganization of traditional contents, presenting them as components of a larger endeavour, and stimulating students to construct a coherent whole of the diverse pieces of the puzzle they are confronted with during their studies.

The teaching project underlying this book [Dit14] is intended as a contribution to this general objective, focussing on a subject that appears particularly suitable for the purpose: The concept of information, even in its most specific, quantitative meaning, proves applicable not only in physics, but also in the adjacent disciplines, in mathematics, including foundations and logics, and in biology, from genetics through development through evolution. Beyond science, the concept is fruitful in the humanities, e.g., in philosophy, linguistics, and history, as well as in the social sciences, in particular in Sociology and in economy (Fig. 1).

Fig. 1 Information
dynamics overlaps
significantly with the
neighbouring sciences,
mathematics, physics, and
biology, as well with
humanities such as
philosophy and
hermeneutics, and of course
with computer science and
other branches of
engineering (not shown)
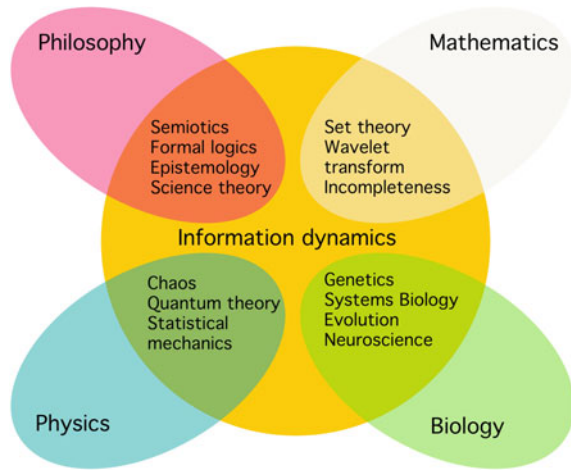
The concept of information is an extraordinarily powerful and versatile analytical tool. While in the sciences, information theoretical approaches are already well established and often even institutionalized as chairs, departments or entire faculties, they open a fresh novel access to many fields of the humanities, enabling a quantitative analysis where it was hitherto inconceivable and providing a firm common ground for a dialogue with the natural sciences.

A hypothesis that holds much of this book together is that these diverse disciplines do not merely use similar concepts referred to with the same term "information", but that they are actually dealing with manifestations of one and the same quantity, exchanged and transformed between different systems, places, and scales.

It cannot be the objective of this monograph to cover such a vast range of subjects. Rather, I would like to provide a conceptual basis, precise and solid enough to allow for mathematical reasoning in many cases, and at the same time sufficiently general to open the view towards a broad range of applications. This means that, unlike traditional physics lectures, a comprehensive idea plays the guiding role, in a wider sense than just a formalism or a set of axioms like Newton's laws or Maxwell's equations. The central aim is to demonstrate the unifying power of the concept of information. More specific topics are addressed selectively as case studies to illustrate and substantiate the general argumentation. The style of reasoning, adapted to an interdisciplinary content and a wider audience, also has to leave the traditional scheme of theoretical-physics lectures behind, occasionally replacing detailed derivations by qualitative arguments, often based on analogies.

With its level of rigour and completeness, the book occupies an intermediate position between semi-popular science [Llo06, Sei06, Flo10, Gle11, Sto15] and specialized monographs pretending completeness and technical accuracy [Bri56, Pie80, Ash90, Roe00, Mac03, CT06]. While detailed knowledge from molecular biology through mathematical logics obviously cannot be expected from a typical

reader, some tolerance in the face of advanced mathematics and the willingness to switch frequently between very diverse topics will be helpful.

The title "Information Dynamics" has also been chosen to indicate a particular emphasis on the flow of information in space and time, its exchange among parts of a system, and its ascent and descent between different scales. Indeed, this work is rooted in the author's academic background in chaos and quantum chaos and is even inspired by a specific article: In a seminal paper [Sha80], Robert Shaw interprets deterministic chaos in terms of a directed flow of information from invisible microscopic scales to macroscopic visibility. As an immediate consequence, this idea leads to the question of how information is processed at the smallest scales, that is, in Quantum Mechanics. There is increasing agreement that the fundamental limits quantum physics imposes on information provide the most appropriate approach even to quantum phenomena we perceive as enigmatic, such as entanglement and nonlocality. Towards the largest scales, basic questions of cosmology invite to being interpreted in terms of entropy. In this sense, the book might well be titled "The information cycle", referring by analogy to the hydrological cycle.

The theory of deterministic chaos has been appreciated as a step towards replacing randomness as an irreducible element in the foundations of physics. Precisely, quantum mechanics, otherwise considered as the last stronghold of insurmountable unpredictability, however, prevents truly chaotic behaviour. The analysis of this surprising outcome inspired a number of profound insights into the character and origin of random processes in physics. Indeed, clarifying concepts and phenomena around randomness, bearing on causality vs. Chance, predictability vs. randomness, symmetry vs. disorder, ..., is another aim of this work: A third possible title could read "Chance: classical stochasticity vs. quantum determinism".

The first eight chapters of this volume are dedicated to interpreting natural phenomena in terms of information processing. They span an arc from basic definitions and simple applications through epistemological aspects to the central part on information processing in physical systems. Following an introduction from a historical point of view, the subsequent two sections cover elementary material such as general definitions and ramifications of the concept of information. Chapter 2 applies them to illustrative examples from quite diverse fields. The three applications presented there, the genetic code, formal logics, and Fourier transformation, not only demonstrate the use, but also the enormous versatility of the concept. They also serve as a reference for later chapters, concerning, e.g., the notion of gates, to recur in the context of computation, and uncertainty relations, so central for quantum mechanics.

Chapters 3 and 4 examine more fundamental issues around information that share their epistemological nature. They analyze the concepts of causality and predictability and oppose them to the creation of novelty, epitomized by randomness. With Chaps. 5–7, attention returns to physics in a more direct sense and thus to a safer ground as concerns mathematical tools. They present a detailed account of how information is processed in classical physical systems, beginning in Chap. 5 on the microscopic level. Hamiltonian dynamics provides the most detailed description of how information is exchanged between scales and among different degrees

of freedom. The central tenet here is its strict conservation in all processes. From a rather more macroscopic viewpoint, Chap. 6 considers dissipative systems as sinks and chaotic systems as sources of entropy. Chapter 7 attempts a synthesis of these contrasting views, both relevant for systems far from equilibrium. It reveals their intimate relation, manifests, in particular, in fluctuation-dissipation theorems, and culminates in an interpretation of the time arrow implicit in the Second Law of Thermodynamics in terms of information flows.

This reasoning directs attention towards the bottom level of the information cycle, where the inconsistencies of classical physics in this respect become inescapable. Chapter 8 turns to the radical way quantum mechanics redraws our image of information processing in physical systems. The progress achieved in recent decades in quantum optics and related areas suggests a contemporary view on quantum mechanics based on information, which may eventually supersede the traditional Copenhagen interpretation. Forming a cornerstone of this book, Chapter 8 covers a wide range of topics, from information-related postulates through quantum measurement and quantum chaos.

The last two chapters take up the same metaphor of natural systems as information processors but read it backwards, considering computers as physical systems. They are dedicated to the conditions the laws of physics impose on information processing, in particular, in artificial systems. Also, here, quantum mechanics implies a radically different view. Chapter 9 introduces the concept of computation as a particular feature attributable to man-made systems, but by no means restricted to them. How far and in which sense computation in quantum systems is different from classical computation is explored in Chap. 10. Chapters 9 and 10 are definitely not nearly suited to replace monographs on conventional or quantum computation, such as Refs. [BEZ00, NC00, BCS04]. Quantum information and computation are among the buzzwords associated with the information era. However, rather than praising quantum high technology, these sections should provide a conceptual basis solid enough to enable a more balanced view towards these developments. Readers waiting for Alice and Bob to enter the stage will be disappointed.

Even with the broad scope pretended in this book, it is inevitable that some topics, intimately akin to its subject, are not given the space they deserve. This is the case for two areas in particular: Nonequilibrium statistical mechanics is at the heart of natural information processing. The fundamental role of the concept of entropy for statistical mechanics has been made popular notably by the seminal work of E. T. Jaynes [Jay57]. In the present book, nonequilibrium phenomena show up in various contexts. However, hardly a chapter—parts of Chap. 7—could be dedicated specifically to this vast subject. Readers are instead referred to the bibliography.

Talking of the particular way computers process information, it would have been tempting to compare it with Nature's solution to the problem: neural computing, specifically in the human brain. In brain research, neuro- and cognitive science are huge subjects in rapid development where information theory attains an increasingly crucial role. It could not be done any justice in this book. Again, readers are invited to follow citation paths.

This book does not conclude a mature subject. It is a snapshot of an active research area in rapid development. Like this, I hope that the reader will close it with the impression of a buzzing construction site, not of a proud venerable edifice. The intention is to inspire ideas that hopefully will contribute to the intellectual endeavour outlined in this work. Should the book provoke more questions than it can answer, it would have achieved its purpose.

The present work would not have been possible without the help and support from numerous sides. Gert-Ludwig Ingold gave the decisive initial impulse by asking me, "Why do you not consider complementing your topical lecture on the concept of information in physics by a textbook?". He accompanied the progress of the book patiently throughout the years and advised me, in particular, on issues related to quantum mechanics. Philippe Binder not only encouraged me untiringly from the very beginning, but became a real mentor of the book project. He was always eager to answer my questions on all topics around classical information and complexity, and he accompanied the writing with suggestions and constructive criticism. Angela Lahee, editor at Springer Nature, gave me the confidence and safety that was indispensable for me to enter into serious work after writing a few sketches of the first sections. Behind her editorial decisions and her patience with my repeated pleas to extend the deadline for completion, I always enjoyed her enthusiasm for my book project. I am owing a very different kind of encouragement to the students of my lectures on information dynamics at Universidad Nacional de Colombia, who motivated me permanently with their curiosity, their sincere interest in the subject, and countless bright, exploratory, and often enough critical questions. In the final phase of the project, Frank Großmann, Frank Jülicher, Leonardo Pachón, Holger Schanz, Walter Strunz, and Carlos Viviescas agreed to review different sections of the book with their particular expertise in the corresponding subjects. I am deeply indebted to all these colleagues and friends.

Besides persons, it was two institutions that made this book possible by their support. The first sketches of a pedagogical project covering the dynamics of information, in particular, in systems with complex dynamics, arose during my time as a postdoctoral researcher at the Max Planck Institute for the Physics of Complex Systems in Dresden. I am sure that the particularly stimulating atmosphere, the countless opportunities inviting to freely discuss even highly sophisticated questions with colleagues working on related subjects, contributed to those first ideas. Numerous short stays ever since at the MPIPKS consolidated my sense of belonging to this institute. As an indispensable practical aid, without which this book could never have completed, the institute granted me access to its electronic journal library throughout the writing process. Universidad Nacional de Colombia, the university I am affiliated with for more than twenty years, gave me the required freedom to dedicate a substantial part of my working time to this book. Above all, the Physics Department accepted and encouraged my proposal of a topical lecture on the concept of information and gave me the opportunity to hold it every few semesters as part

of my teaching duty. I am sure that the liberal and innovative spirit at this university
has facilitated these generous concessions.

Bogotá, spring 2022   Thomas Dittrich

Contents

Part I
Natural Systems as Information Processors

Chapter 1
The Concept of Information

1.1 Some History

Just as trilobites and hand axes characterize their respective paleontological and archaeological strata, there are concepts that can serve as "leitfossils" of certain eras of cultural history, in particular of the history of science and technology. The notion of information certainly has reached this level of prominence, as could easily be evidenced, say, by a word count including all classes of documents produced in our times.

In this respect, information has much in common with a related category, energy: Physical by nature, it applies to a huge variety of situations also outside the scientific realm. In fact, both concepts also share the dichotomy between a precise meaning in physics, including a quantitative definition, and a rather broad and sloppy usage in everyday language. Their adoption from and penetration back into common vocabulary can be traced back to similar historical processes: The history of science and technology on the one hand and the general political, social, and economic development on the other are in permanent interaction [Ber71]. Obviously, the evolution of science depends on economic and social boundary conditions, it reacts in its specific way, for instance, to periods of peace and of war. At the same time, major progress achieved in science and technology can have significant repercussions on common life, even on political decisions. While small steps usually go unnoticed by the general public, it is well possible to correlate grand epochs in the history of science with concomitant global changes in the society, often but not necessarily coinciding with scientific revolutions in the sense of Thomas Kuhn [Kuh70].

Table 1.1 is a crude attempt to identify such long-term parallel developments during the last half millennium, not attempting precision nor completeness or uniqueness. It also indicates how these last three eras have been accompanied by simultaneous technological progress, reflected in novel devices being invented and entering daily life. The central concepts like energy and information have played the role of universal metaphors, applied not only in their original field but to just about every subject that invites for scientific understanding and interpretation. While Laplace and

Table 1.1 Subdividing the last half millennium into three major epochs demonstrates strong correlations between scientific and technological progress, paradigms dominating science and philosophy, and social and economic development.

Centuries	Era	Characteristic quantity	Characteristic devices	Paradigms	Prevailing socio -eco - nomic model
16 - 17	Mechanistic	Force	Wedge, lever, pulley, tooth wheel, clock, printing press	The universe as clockwork, celestial mechanics, Laplace's demon	Agriculture, manufacture
18 - 19	Thermo - dynamic	Energy	Watermill, windmill, steam engine, combustion motor	Energy conservation, thermodynamic systems, Maxwell's demon	Industrial
20 - 21	Informatic	Entropy, information	Telegraph, telephone, computer	The world as computer/ network	Postindustrial knowledge society

his contemporaries gained deep insight interpreting nature as a deterministic mechanism, as "cosmic clockwork", we enjoy "Programming the Universe" [Llo06], seeing it as an all-embracing quantum computer.

In a similar way as Laplace and his contemporaries interpreted all of nature as a deterministic mechanism, as "cosmic clockwork", we tend to associate organs and their functions to engines or parts thereof, to compare our brains with computers and genetic control with algorithms.

The category of information penetrates our life from the TV news till the purchase of a computer or a cell phone. The present book is intended to contribute to a conscious critical use of the concept, based on solid knowledge of its meaning and its limitations.

The way this notion gained shape in the course of academic discussion is a particularly fascinating chapter of the recent history of science and technology. Suffice it to mention but a few of the most important protagonists [Sei06, Gle11]:

The title of founding father is, without doubt, due to Ludwig Boltzmann. The essence of his conception of entropy in the early nineteenth century has essentially survived all modifications and ramifications the notion suffered ever since. Before spreading further outside physics, the idea revolutionized statistical mechanics and inspired his great contemporaries: Maxwell's demon continues stimulating discussion and research till today. Gibbs' paradox may be considered as the first crack in the edifice of classical physics, paving the way towards quantum mechanics. Planck's conception of the quantum of action was inspired by profound reflections on the nature of entropy [Pla49]. While nineteenth century engineering was focused on machinery converting energy from one form into another (see Table 1.1), such as the Steam engine, it has already seen the first sophisticated apparatus dedicated to information processing, including the control unit of the Jacquard weaving loom (Fig. 1.1) and Babbage's "difference engine " and " analytical engine".

Fig. 1.1 The Jacquard weaving loom (1805) is probably the first man-made tool generating patterns from a digital Memory medium, astonishingly similar to the punch cards of the 20th cty. It continues to serve as a metaphor to explain reproducible pattern formation controlled by digital data, for example the genetic control of embryonic development [Nus04].

It was mathematicians who saw the enormous potential of the concept, gave it a rigorous form, sufficiently general to be applied all over science, and explored its implications till the most mind-boggling consequences. Shannon, in the course of quite practical work for the Bell Telephone Co., proposed the basic definition (see Eq. 1.3.9 below), valid till today. Charles Babbage, Ada Lovelace, and Alan Turing, three personalities with a somewhat tragic biography, pioneered the idea of a computer and the foundations of its mathematical description. Kurt Gödel did not refer to information explicitly, but it is undisputed by now that it is the key to understanding his Incompleteness Theorem, arguably the greatest intellectual achievement of the twentieth century [Cha75, Cha82].

Information gained impact again within physics with the work of theoreticians who recognized its relevance in quantum mechanics. John von Neumann not only "quantized" the concept, as the Von-Neumann entropy, but also applied it to discuss fundamental problems of quantum mechanics, such as the measurement process [Neu18]. Leon Brillouin, besides his direct contributions to the field. was the first to conceive a systematic account of the role of information in physics, with his classical textbook on the subject [Bri56]. Leo Szilard reduced Maxwell's demon to its very essence, a single-bit decision, bringing out for the first time the equivalence of entropy and information [Szi29].

From the 1960s onwards, the concept already became so commonplace that selecting the most relevant contributors is inevitably arbitrary. In physics, names come to mind such as Richard Feynman who was the first to contemplate quantum computing [Fey82]. Ilya Prigogine and Hermann Haken introduced concepts closely

related to information into statistical mechanics far from equilibrium: dissipative structures [PN77] and synergetics [Hak83], resp. Edwin T. Jaynes recast statistical mechanics, basing it on the principle of maximum entropy [Jay57]. Philip W. Anderson, in the shadow of his epochal work on solid-state physics, contributed important reflections on the concept of complexity [And72]. Rolf Landauer was the first to address systematically the question of fundamental physical limits to computation and coined the statement "Information is physical" [Lan91]. The last sections of this book are largely based on his work.

Among mathematicians of these generations excels, above all, Andrey Kolmogorov, who advanced mathematical physics in a broad spectrum of objects related to probability and information. Aleksandr Khintchin, closer to pure mathematics and probability theory, contributed in a similar direction to this field. With a rather unique profile merging mathematical physics with epistemology, Ray Solomonoff is known as the father of algorithmic complexity [Sol64], his work forming the nucleus of Sect. 4.2.2.

The contributions of contemporary philosophers, mathematicians, physicists will be accounted for in the sequel, in the context of the respective subject.

1.2 The "Three Dimensions" of Information

Before delving into physics and mathematics ,an approach to information should be addressed that originates in philosophy, anthropology, and linguistics. It is in many respects broader than and complements the scientific point of view, emphasizes other facets and allows us to embed the concept of information in a more general context.

Under the headline "semiotics", a theory of Communication has been pioneered by Ferdinand de Saussure [Sau77] and Charles Sanders Peirce [Pei34] and developed further by Thomas A. Sebeok [Seb76] and others. Its most well-known feature is the distinction of three categories, three "dimensions" of information, which however are not to be understood as a division into disjoint sectors, but rather as nested sets:

syntactic information

syntactic information refers to the quantity and formal structure of messages. It counts signs, circumscribes their repertoire and their mutual relations. Alluding to the general meaning of "syntax", it defines rules and restrictions how to compose legitimate sequences of signs, often in a hierarchical fashion. The motion of Messages in space and time, reduced to a mere Flow of information, pertains to the syntactic realm. The notions of sender and receiver reduce to mere sources and sinks. In the context of languages, it is orthography and grammar ("syntax") that stand for syntactic information.

The syntactic aspect of information is relevant for physics and chemistry, for structural sciences such as mathematics and informatics, and for technology related to communication and computation.

It is important to keep in mind that meaning is excluded in syntactic information. The meaning of signs, and more generally, communication as an exchange of messages, enter the scene only with:

semantic information

Semantics is dedicated to the relationship between symbols and their meaning. Obviously, to apply, semantic information requires a language to exist, be it in the most rudimentary form. It includes the relation between symbols and the elements of reality they are referring to (encyclopedias) as well as between equivalent symbols of different codes and languages (dictionaries). Semantic information is concerned with synonyms (different words within the same language with identical or overlapping meanings) and homonyms (words with two or more distinct meanings).

In the context of semantics, the concepts of sender and receiver acquire the additional aspect of *understanding*, not contained in syntactic information. Only on this level, it makes sense talking of *truth* (as opposed to falsehood, to error, lie, etc.). Truth has to do with the relationship between two levels of reality, objects and their symbols, and cannot be grasped within the symbolic level alone, as numerous futile attempts to "define" truth confirm (such as "The proposition 'It rains' is true if and only if it rains."). Representing reality in images, words, formulas, …, is a collective endeavour that cannot be reduced to a formal criterion.

Within the natural sciences, only biology deals with codes and languages and thus with sematic information. It becomes the central aspect in the humanities, in particular in linguistics, in hermeneutics and literature, and in the arts. Both the fine arts (painting and sculpture and their more recent ramifications) and the performing arts (music and theatre) typically do not use established codes but create meaning by inventing new symbols "in the making". The concept of truth is of obvious relevance for journalism.

Pragmatic information

Pragmatic information finally takes into account that senders may have intentions emitting messages, and receivers may or may not react as desired by these intentions. With pragmatics, norms come into the play, the dichotomy of true and false is complemented by that of good and evil.

Contexts where we can meaningfully talk of Pragmatic information are even more restricted than in the case of semantic information. It only applies to the communication of intelligent beings we can attribute intentions to. The disciplines where pragmatic information becomes relevant range from psychology through the Social sciences (particularly history, political science, and economy), through law.

Syntactic, semantic, and pragmatic information require increasingly higher degrees of organization and complexity in the systems they are ascribed to. This suggests a hierarchical relation between them, as sketched in Fig. 1.2.

It will become clear in the following sections that binding the concept of information too tightly to the triad "sender – message – receiver" as its fundamental elements, as is implicit in the semiotic approach and is frequently reproduced in recent literature on information, is in fact inappropriate and leads to an overly narrow outline of
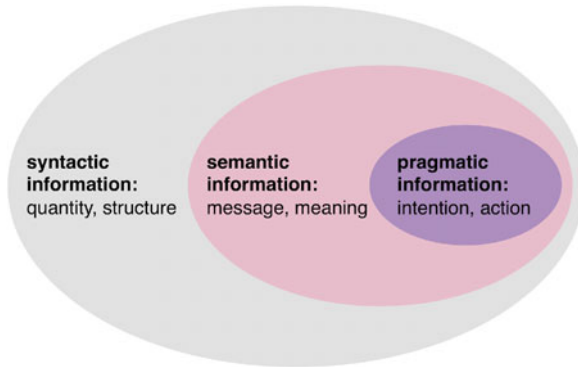
Fig. 1.2 Semiotics distinguishes three "dimensions" of the concept of information: Syntactic information deals with the sheer quantity of information and the internal structure of symbol sequences. With semantics, messages transmit meanings from senders to receivers, while pragmatic information analyzes intentions of senders and the actions their messages evoke in the receivers.

Fig. 1.3 René Magritte's painting "La trahison des images" (1928–1929) illustrates, from an artist's point of view, the distinction between an object and the symbol representing it: An image of a pipe is not a pipe; one cannot smoke it. The perplexing effect of the painting is due to Magritte's mixing three levels of meaning into one image, the object (the pipe), the symbol referring to it (the picture of the pipe) and a metalanguage statement "(Ceci n'est pas une pipe)".

the notion. There is a host of subjects in physics and mathematics that invite to be discussed in terms of information but do not allow nor require identifying anything as sender or receiver.

It is a fascinating question, to be addressed en passant, exactly where in the history of nature a phenomenon to be called "meaning" emerged for the first time. As far as we know till now, symbols ,that is, objects encoding for others, did not exist in any sense before the advent of life on Earth. With the genetic code, a systematic relationship—a dictionary—arose between one class of long-lived molecules (RNA or DNA), serving as memory, and another class of very reactive but short-lived molecules (proteins), capable of affecting and altering their environment, see Sect. 2.2 . The genetic code

is located on the borderline between natural phenomena that could still be described without reference to meaning and the onset of semantic information.

Further on in biological evolution, languages of very diverse type developed several times, from chemical cell signalling through olfactory, acoustical and optical signals in multicellular organisms through the formal object language in humans.

It is particularly illuminating to consider how objects, images, smells, sounds, gestures, … assume the role of symbols, that is, come to be understood as pointers towards specific phenomena as their meaning. An important factor in this process is certainly the existence of correlations, which at least initially relate symbol and meaning. As an illustrative example, take conditioning, epitomized in the paradigmatic experiment known as Pavlov's Dog [CM10]: Being confronted in a systematically correlated manner with the sound of a bell and the serving of meat, in this order, the dog learns to associate the bell with the food and drools already upon hearing the bell alone. For the dog, the sound of the bell has become an acoustic signal anticipating the meat, a symbol that represents the upcoming feeding. Indeed, in many languages we can still observe the origin of words in an onomatopoetic imitation of sounds correlated to their meaning, and in primordial scripts the origin of graphical symbols in pictogram-like representations of the objects they refer to is evident [Jac81].

In this context, a terminology should be introduced that will be become indispensable in the discussion of self-reference below: Symbols, irrespective of their acquired function in communication, continue forming part of objective reality, and as such can be referred to by other symbols (Fig. 1.3). Symbols whose meaning is itself a symbol can be associated to a distinct, higher layer of language, called *metalanguage*, each one forming a *metasymbol*. Symbols referring to *metasymbols*, in turn, are *metametasymbols* and belong to a *metametalanguage* etc., in an obvious sequence of nested recursive definitions: For example, ✄ is a pictogram, "Ceci n'est pas une pipe" is a sentence in French, the quotation marks indicating that this sentence belongs to the *object language*, relative to this English metalanguage sentence embedding it, while "'Word' n'est pas un mot français" adds one more level to the hierarchy, this time indicated by single apostrophes.

A comprehensive feature associated with information, both in its common and in its scientific usage, is its independence of the medium carrying it. It is an everyday experience that the audio signal of a telephone call is transferred back and forth between acoustic waves, electric impulses in cables, radio-frequency carrier waves, light pulses in optical fibres and other media, without being severely distorted. Marshall McLuhan's proverbial phrase "The medium is the message" is an ironic inversion of what in general we consider as a fundamental contrast: A message can be transmitted via any medium, a medium can carry whatever message.

More specifically, in physics, the same amount of information, now in its syntactic sense, can be stored in the shape of a solid object, in the magnetization pattern imprinted in a magnetizable material, or the absorbence of photographic emulsion, or even in the probability density distribution of an ensemble in phase space. It can be transmitted by acoustic waves in air or surface waves in water, by electromagnetic and even gravitational waves, and as a matter wave according to quantum mechanics.

Information shares this universality with the concept of energy, in its multiple manifestations in every context in physics, including sheer mass as implied by Special Relativity Theory.

The contrast of message and medium recurs here as the dichotomy of information vs. mass or energy, resembling the opposite pair form and matter, which pervades the philosophical discussion since the antiquity. In a radical view, every structure in space and time ultimately reduces to information, leaving only a featureless mass for the component of matter. It is taken to the extreme, complemented by the idea of the two-state system as minimum manifestation of a quantum system, in John Archibald Wheeler's succinct dictum "It from bit" [Whe90a,Whe90b], and has been worked out by Carl Friedrich von Weizsäcker as a metaphysical system, the theory of the "ur-alternative" [Wei71, Wei85]. The following subsection resumes proposals to define information quantitatively and their consequences.

1.3   From Boltzmann's Entropy to Shannon's Information

In the context of physics, information is distinct from all other physical quantities in that an operational definition, clear instructions how to measure it, can hardly be given. That it belongs to a separate category is also evident from the fact that it applies to the very concept of measurement, as a transfer of information. Basically, information is a mathematical concept, it pertains to the realm of structural sciences, not to the natural sciences. Yet it is so ubiquitous in physics and in particular, as thermodynamical entropy, becomes indirectly measurable, that the statement "Information is physical" [Lan91] appears well justified.

With this intermediate position between physics and mathematics, not unlike geometrical concepts, information is not only comparable with dimension but is even closely related to it in its meaning, see Sect. 6.4: another instance of a quantitative notion that is crucial in physics but cannot be measured with a "dimensiometer".

We owe it to Ludwig Boltzmann's ingenious intuition when the concept of entropy is introduced as counting the number N of "complexions" of a physical system, a huge step forward from previous attempts to define it as a thermodynamic quantity.

In conceiving this term, Boltzmann thought of microscopically distinct states of matter, states that can be distinguished by some microscopic physical quantity, but are not necessarily discernible by macroscopic observation. Anticipating later generalizations, in particular by Shannon, we can define complexions as *distinguishable states of a system*, thus avoiding a reference to physical terms. All the subtleties of this definition are hidden in the word "distinguishable". It implies a high degree of subjectivity: distinguishable by whom, according to which criterion, with respect to which quantity, measured to which accuracy? We shall see below that by analyzing this subjectivity, a rich structure emerges which in fact contributes to the versatility of the concept. A quantity that allows to interpret Boltzmann's complexions in modern terms is the probability density in phase space, see Sect. 5.2. It is remarkable that Boltzmann himself was already quite aware of this problem and in fact based his

original argument [Bol09] on the assumption of a *finite number of discrete states*, corresponding, e.g., to energies $E_n = n\Delta E/N$. He considered this, however, as an intermediate formal step, to be complemented by a limit process $N \to \infty$, $\Delta E \to 0$.

Quantifying the repertoire of states a system can assume was the radically new aspect Boltzmann introduced with his entropy. Otherwise the quantity should comply with all general conditions other thermodynamic potentials fulfill. In particular, he required it to be additive: Consider a physical system composed of two separate subsystems, 1 and 2, i.e., two parts not interacting with one another, not interchanging matter or energy. In this case, the total entropy should just be the sum of the entropies of the two subsystems.

If systems 1 and 2 comprise N_1 and N_2 states, resp., the total number of states is the product, $N = N_1 N_2$. The function replacing products by sums is the logarithm, hence the decision to define entropy S as the logarithm of this number,

$$S = c\log(N),\qquad\qquad(1.3.1)$$

c denoting an arbitrary constant that fixes the unit of entropy. In the context of statistical mechanics, the total number of distinguishable states is proportional to the volume Ω of the accessible phase space, the quotient being given by the size $\Delta\Omega$ of the smallest discernible phase-space cell,

$$S = c\ln\left(\frac{\Omega}{\Delta\Omega}\right).\qquad\qquad(1.3.2)$$

The denominator $\Delta\Omega$ is also required to render the argument of the logarithm dimensionless. This equation is equivalent to the emblematic formula $S = k\log W$ engraved on Boltzmann's tombstone.

As elsewhere in physics, appropriate units depend on context. In thermodynamics and statistical mechanics, the standard application is to macroscopic ensembles. In this case, manageable numbers for the entropy are achieved defining

$$S = k_B\ln(N),\qquad\qquad(1.3.3)$$

where $k_B = 1.38 \times 10^{-16}$erg/K is the Boltzmann constant, with dimension energy over absolute temperature, in units of erg over degrees Kelvin. In the context of symbol strings, the established procedure refers to the shortest sequence of yes–no questions that defines the sequence at hand. This amounts to

$$I = \mathrm{lb}(N) = c\ln(N),\qquad\qquad(1.3.4)$$

with $c = \mathrm{lb}(e) = 1/\ln(2)$ ("lb" denoting "binary logarithm"), defining the bit as unit of information.
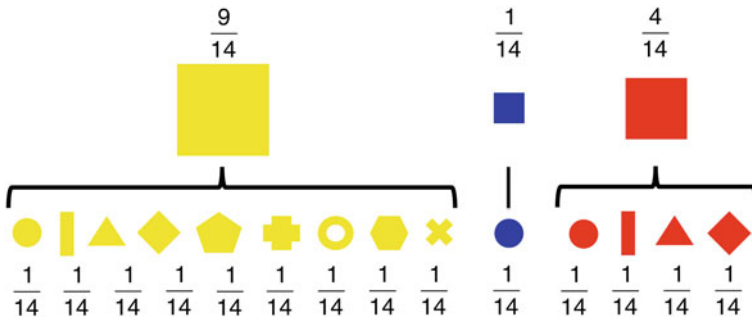
Fig. 1.4 Boltzmann's original entropy definition assigns the same weight to all distinct microstates. In later generalizations, sets of microstates, considered indistinguishable on a higher level, are lumped into single items, weighted according to the number of microstates they contain. In the figure, the corresponding weights would be 9/14, 1/14, and 4/14, respectively.

Boltzmann's original version, Eqs. (1.3.1) or (1.3.2), assumes that there are no a priori preferences for anyone of the N microstates, attributing democratically the same weight to all of them. In almost all applications of entropy, however, the states or symbols counted in measuring an information content already correspond to sets of microstates that are considered as equivalent or indistinguishable (Fig. 1.4). Depending on the size of these sets, the states no longer occur with the same frequency but can be assigned different probabilities. This freedom is indispensable for example in the context of abbreviations and substitution rules that replace sequences of more elementary signs by higher-level symbols.

An important instance and application of varied frequencies, quite unrelated to statistical physics, is the design of codes. In this context, the role of macrostates is played by high-level symbol sets or alphabets, with a large number of symbols, for example the Latin alphabet with around 26 characters, to be encoded by a more elementary symbol set, say the Morse code with only two signals (dot $\cdot$ and dash —), analogous to the microstates. Fixing the number of signals per Morse code symbol representing a single Latin character, this would require $5 = \lceil \log_2(26) \rceil$ signals per letter of the alphabet. However, admitting also shorter lengths, symbols of four signals per letter or less are sufficient. The code can then be optimized, concerning the total number of signals per average message, by assigning the shortest symbols (one signal) to the most frequent Latin letters (in a language of reference, say English) and the longest symbols (four signals) to the rarest letters. This is how the Morse code has been devised (Fig. 1.5). The appropriate criterion for an optimal symbol length is the Shannon information per symbol.

If M is the total number of microstates, M_j of which belong to the jth distinguishable class, $j = 1, \ldots, J$, $\sum_{j=1}^{J} M_j = M$, then probabilities

$$p_j = \frac{M_j}{M} \tag{1.3.5}$$

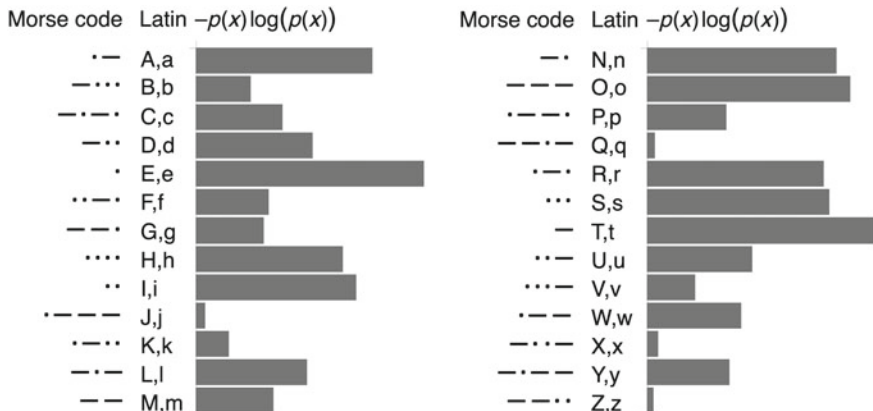| Morse code	Latin	$-p(x)\log(p(x))$		Morse code	Latin	$-p(x)\log(p(x))$
·—	A,a			—·	N,n	
—···	B,b			———	O,o	
—·—·	C,c			·——·	P,p	
—··	D,d			——·—	Q,q	
·	E,e			·—·	R,r	
··—·	F,f			···	S,s	
——·	G,g			—	T,t	
····	H,h			··—	U,u	
··	I,i			···—	V,v	
·———	J,j			·——	W,w	
—·—	K,k			—··—	X,x	
—·—	L,l			—·——	Y,y	
——	M,m			——··	Z,z	

Fig. 1.5 The Morse code assigns a combination of dots · and dashes – to each letter of the Latin alphabet. In order to optimize the code regarding the number of single signals per transmitted information, the length of each Morse code sign, from one to four signals, is adapted to the relative frequency $p(x)$ of each letter (in English). Frequent letters such as "e" and "t" transmit least information and are encoded by a single signal, rare letters such as "q" and "z" are encoded by four signals. An appropriate criterion to determine the length of each code symbol is the weighted logarithm $p(x)\log(p(x))$, see Eq. (1.3.9) (bar chart).

can be attributed to the class j, with $\sum_{j=1}^{J} p_j = 1$. The total number of distinct sequences consisting of M of these symbols is not $M!$, but has to be divided by all the numbers $M_j!$ of combinations of microstates that can be formed within each class j,

$$N = \frac{M!}{\prod\limits_{j=1}^{J} M_j!}. \tag{1.3.6}$$

The corresponding total information is [Bri56]

$$I_M = c \ln(N) = c \left(\ln(M!) - \sum_{j=1}^{J} \ln(M_j!) \right). \tag{1.3.7}$$

If the number of symbols is large, $M_j \gg 1$ for all classes j, this expression can be simplified using Stirling's formula, $\ln(x!) \approx x \ln(x) - x$, valid for $x \gg 1$:

$$I_M \approx c \left(M \ln(M) - M - \sum_{j=1}^{J} M_j \ln(M_j) + \sum_{j=1}^{J} M_j \right)$$

$$= cM \left(\ln(M) - \sum_{j=1}^{J} \frac{M_j}{M} \ln(M_j) \right)$$

$$= -cM \sum_{j=1}^{J} \frac{M_j}{M} \ln\left(\frac{M_j}{M} \right). \tag{1.3.8}$$

Comparing with Eq. (1.3.5), this implies for the information per symbol,

$$I_1 = \frac{I_M}{M} = -c \sum_{j=1}^{J} p_j \ln(p_j). \tag{1.3.9}$$

This is the definition proposed in 1948 by Claude E. Shannon [Sha48,SW49], following a similar expression suggested by J. W. Gibbs in the context of statistical mechanics [Gib02,Gib93]. It generalizes Eq. (1.3.1) but contains it as a special case; for equal probabilities $p_j = 1/J$, it reduces to Boltzmann's expression,

$$I = -c \sum_{j=1}^{J} \frac{1}{J} \ln\left(\frac{1}{J} \right) = c \ln(J). \tag{1.3.10}$$

The minus sign compensates for the fact that, with $0 \leq p_j \leq 1$, $j = 1, ..., J$, the logarithms in Eq. (1.3.9) are all negative. At the other extreme, opposite to equal probabilities, Shannon's definition also contains the deterministic case that only a single choice j_0 applies,

$$p_j = \delta_{j-j_0} = \begin{cases} 1 & j = j_0, \\ 0 & \text{else,} \end{cases} \tag{1.3.11}$$

so that

$$I = -c \sum_{j-1}^{J} \delta_{j-j_0} \ln(\delta_{j-j_0}) = 0. \tag{1.3.12}$$

At the same time, the two cases (1.3.10) and (1.3.12) mark the absolute maximum and minimum values, resp., the entropy can assume, a fact to be analyzed further in the sequel.

1.4 Sign: Entropy and Negentropy: Actual Versus Potential Information

Till here, the concepts of entropy and of information have been used indiscriminately, despite their highly divergent origins in thermodynamics and in communication, resp. Indeed, it is one of the objectives of this book to demonstrate that this is a legitimate manifestation of the enormous versatility of the concept. Notwithstanding, where it appears appropriate, the term "entropy" will be preferred in thermodynamic contexts and "information" where symbol strings and other discrete entities are concerned, always keeping in mind that this distinction is ambiguous and in fact unnecessary.

Calling the quantity defined in Eq. (1.3.9) "information" would appear inadequate also on the background of the colloquial use of the word. As the two limiting cases concluding the last subsection show, I reaches its maximum precisely when all states involved are equally probable, that is, if nothing is known about the system, while the minimum 0 is assumed when the state is exactly fixed. It could hardly be more counterintuitive.

Erwin Schrödinger and Leon Brillouin, contemporaries of Shannon, were quite aware of this problem and suggested the term "negative entropy" or "negentropy" instead, for $-S$ as the more legitimate candidate for a measure of information. As a result, however, an incurable confusion concerning the sign of this quantity has been created that continues to cause misunderstandings in the literature.

For the sake of consistency of the terminology to be used in this volume, I introduce another distinction that helps keeping these meanings apart. What is in fact measured by entropy is the magnitude of state space accessible to the system or the number of signs available to compose a message of. It is therefore adequate to denominate it *potential* information. Furthermore, to account also for ambiguities with respect to the state of knowledge, resolution, measurement accuracy etc., implicit in the term "distinguishable", it makes sense to define a *relative* information, with respect to some reference value. Together, this leads to the definition

$$\Delta I_{\text{pot}} = c\Big(\ln(N_{\text{final}}) - \ln(N_{\text{initial}})\Big), \qquad (1.4.1)$$

where "initial" and "final" refer to a dynamical time evolution, to a measurement, an observation, or whatever process causes a change in the number of available states. Potential information is equivalent to the lack of knowledge, to ignorance, to uncertainty about the state of the system.

As complementary quantity, introduce *actual* information as measuring what is already known about the system. This would amount to

$$\Delta I_{\text{act}} = c(\ln(N_{\text{initial}}) - \ln(N_{\text{final}})) = -\Delta I_{\text{pot}}, \qquad (1.4.2)$$

Actual information stands for what is known, for the knowledge an observer, a measurement, a theory, …, has on the state of the system. It measures the fraction