Chapter 1
Information

1.1 Introduction

Generations of writers have referred to *information*, but individuals in various fields have used the term differently. As the electrical engineer Ralph Hartley noted in a 1928 article [91], "information is a very elastic term." While working at Bell Laboratories, a leading research institute at that time, he developed an early mathematical measure of the amount of information passing through telephone circuits, making it "possible to set up a definite quantitative measure of information based on physical considerations alone," ignoring the psychological factors that were often present in earlier discussions of information. Hartley considered how different methods of representing information could be used to send information at different rates of speed, depending on the representational system used. By developing measures of information that were both objective and independent of psychological factors, Hartley was one of the earliest scholars on the path toward developing what is now referred to as *information theory* by scientists.

Almost three quarters of a century later, a Nobel Prize committee declared George Akerlof's essay "The Market for Lemons" "the single most important study in the literature on the economics of information." When purchasing a previously-owned car, how much should one pay, given that the car might have important flaws? Is the fact that it is for sale an indicator that there is something wrong with it and the seller no longer considers it worth keeping? When a car has been treated well by the original owner and the car is felt by the seller to be a good car, not a "lemon," how does the seller convey this information in a way that would establish the relatively high worth for this car, as opposed to the worth of a car that truly is a lemon? The information possessed by one party to a transaction, information that might be lacking by another person, is *asymmetric information*. When presented with the opportunity to purchase a high quality vehicle, the potential purchaser might be unwilling to pay the true value of the car because the potential buyer does not know that it is a high quality automobile, and thus the seller will often ask for a lower price than the car is worth because the buyer does not have the information the seller has about

R. M. Losee, *Information from Processes*, DOI: 10.1007/978-3-642-31190-1_1,

the quality of the car. The buyer might overpay for a lemon that the seller knows is worth little but where the asking price for this lemon is the average price for this model, based on a mix of cars in good condition and cars that are lemons. Ackeroff describes different information being held by each participant in a transaction, with the information being used by participants to maximize their own economic benefit.

Fred Dretske, a philosopher, wrote about information in the context of knowledge and the nature of meaning [60]. By expanding on ideas from Shannon's *Theory of Communication* and philosophical studies of knowledge, Dretske was able to define information in a manner that captured the human use and acquisition of information. Developed from earlier probabilistic concepts of information, Dretske's ideas placed the information in a message in the context of the knowledge present in the message's receiver. By providing a rigorous set of ideas about information consistent with the interests of those studying beliefs and ideas, Dretske moved philosophy forward by formally linking information with ideas about knowledge.

John Achibald Wheeler, arguably one of the leading physicists of the twentieth century, saw information at the core of understanding physics. For Wheeler, "every physical quantity, every it, derives its ultimate significance from bits, binary yes-or-no indications, a conclusion which we epitomize in the phrase, *it from bit*" [199, p. 3]. As the person who coined the astronomical phrase *black hole*, he examined parts of the universe, such as areas outside a black hole, that might not be able to obtain information about other parts of the universe, such as information about events occurring within a black hole. Wheeler envisioned a future for physics where "we will have learned to understand and express *all* physics in the language of information" [199, p. 8]. The concepts of information are so basic for Wheeler that understanding information is akin to, and as essential as, the understanding of mass or energy.

These writers all focused on information in a way that advanced their arguments. Is there is a single phenomenon called information that was being studied by these people when addressing different problems? Clearly, people in different groups use different vocabularies and may use different meanings for the same term, such as *red* representing a political orientation in the vocabulary of political science, the notion of being in debt for accountants, or a color for artists. For the term *information*, different disciplines may have different nuances, but there are also common underlying phenomena that tie together the different uses. What are these underlying phenomena, and how do we understand the characteristics and the rules associated with these phenomena? Different disciplines may address their field-specific interests by describing and understanding their issues consistent with a discipline-independent information phenomena.

A unitary science of information is needed for a variety of reasons. Foremost is the practical need for a science that will allow one to describe, predict, and understand the increasing amount of potentially useful information available, as well as to provide tools to study systems that capture this rapidly increasing amount of information. The types of problems being encountered are increasingly complex and interdisciplinary in nature, necessitating more sophisticated and yet unifying models of information. Being able to predict the performance of information phenomena is also important; no reasonable person would determine how far they could travel with a car at a fixed speed for a certain period by driving it repeatedly. Instead, one might note that traveling 50 miles per hour for 2 h would result in traveling 100 miles, using a widely accepted and frequently used predictive model. The nature of the smallest phenomena, when combined, determine the nature of the larger, combined phenomena, and describing these smallest phenomena in a rigorous discipline-independent manner allows for the broader phenomena to be based upon firm bases. General tools for studying information, allowing one to describe, predict, and understand actions, are at the core of any serious study of information. In fact, the rigorous study of information may serve as the basis for the rigorous study of most academic disciplines. Information is a fundamental phenomena in the universe.

1.2 Information

The functions and operations studied by these scholars are all processes that generate output. The characteristics of this output constitute the *information* produced by that process. This information is *about* both the process and any inputs to the process. One might define information in an abstract manner as *that which is about something*. The information in the output of a process *represents* the process's input. This general model always describes the fundamentals of information-like phenomena.

Other information from outside the process is assumed to occur at the output only if the information is carried by the input through the process in question; the information in the output of a process is only about the information that is contained in the input to the process or in the process itself. The output characteristics can become the input information to successor processes, which in turn produce additional output with characteristics, with informative processes continuing to make available their output for input into subsequent informative processes. The output characteristics may be embedded in an electronic message, with information moving from one physical location to another, or located in physically or electro-magnetically recorded information, moving from one time period to another. A process may be understood as a communication channel between a past and a future, passing through the present [49]. The output of a process may be treated as an object, such as a pre-historic painting on a cave wall, or a recently printed best-selling novel that one can purchase, read, or place on a bookshelf.

Information is produced in various ways by living creatures. A kitten's purr provides information that one of several purr-producing emotional states exists within the cat. Scholars have proposed models of information moving through communication channels such as the nerves in a cat, and these information models can be used to explain and measure the amount of information moving between a cat's brain and the vibrating vocal cords. Other researchers have examined the meaning contained in utterances, such as a cat purring or a person speaking. Some economists focus on information as having a value related to its use supporting decision making by the listener, such as whether a cat should be stroked by its human, or whether the cat should be left alone by the human to avoid being scratched. What is the value of hearing that it is going to rain before going on a picnic, compared to the cost associated with not knowing that it is going to rain and then leaving for the picnic unprepared for the raining weather?

The economic value of information is associated with a context. A picture showing the damage to a car from a collision might be very valuable in seeking an insurance settlement. A photograph of the car showing how nice the other side looks might be useful for selling the car on the Internet. Switching the information representations, or pictures, of the same entity from their assigned roles would likely result in receiving less money selling the car or by receiving a smaller payment from the insurance company.

In our world, information associated with the output of a process can be viewed on many scales. In the output of microscopic physical processes, there might be chemical reactions producing light or heat. Social groups of primates function as a unit, waging war and conquest, producing babies and caring for them, as well as developing social structures. For multi-celled organisms, impulses move through neurons, carrying information, while larger scale processes, such as a human, might sense the meaning in the purring sounds in a happy cat or the printed word. Describing information as the characteristics in the processes' output gives one a general statement about information that can be applied to all information-related phenomena. The precision and generality of this view of information allows for its application to a wide range of phenomena, providing a powerful tool for developing and understanding both precise, useful models of information, and the application of these models.

As an example of an informative process, consider a portable music playing device that plays recorded music or speech. Whether the player is a stand-alone music playback device or has more sophisticated communication or computational capabilities along with playing the recorded audio material, information is produced by the device, and possibly noticed by the user. The music player extracts bits of information from its memory, converts them to analog signals, and then amplifies the signals, eventually causing earphones to vibrate in such a way that the user hears an accurate representation of the music or speech stored in the music player's memory. The information present in the earphones' vibration is about the state of the components in the player. For example, if the player is broken and distorts music, the information at the output is not only about the music but is also descriptive about the condition of the music player.

Existing systems may be viewed as a single large system or several smaller, connected systems. A radio receives a signal sent from a distant antenna and then transforms it into sounds. The operation of the receiving and amplifying process is partially determined by the available inputs. The system that provides the vibration of the radio's speaker encompasses both the radio itself, the transmitter and its antenna, and possibly the electrical system supplying power to the radio. A larger view of the process, including the radio receiver, transmitter, and program production allows one to conceive of a relatively self-contained system. The radio may also be viewed as an entire small process, with several inputs to the process, including the received signal and the resulting audio output for the listener.

Using a cellular telephone, a person may communicate with a friend, transferring facts and ideas. The recipient of the message is presented with information about the operation of the telephone process and the originator of the information, the person speaking to them. The entire telephone network may be understood as a single process, or the person talking and the individual telephone system they are using may be viewed as a process whose output is the sound generated by the speaker of the message recipient's telephone.

Hearing a cellphone make a noise indicating the presence of an incoming call often has an economic value to the telephone's owner. There may be good news, bad news, or the telephone call may be the result of a telephone number incorrectly entered into another telephone by a stranger. There is clearly a large economic value to hearing that if you take a certain action in the next 30 min, you will be able to buy a company that you believe will likely make you wealthy [93]. Similarly, there is value in hearing from a real estate agent that your business competition is considering purchasing an existing factory that is twice as large as their current facility. The information in the output of a process often has an economic value to decision makers.

What do we gain by examining the nature of the information contained in outputs produced by a process? The greatest strength of a process based information model is that it is field independent. Processes exist everywhere in the universe, with some processes being microscopic and some gigantic. Whether domains are mechanical or probabilistic or best described using quantum models, processes can be described that capture these operations. Subjective phenomena, like knowledge or observations, can be described by a process, as can the most easily describable objective phenomenon from a physics or chemistry textbook.

Information processes are often studied by information system analysts, who examine existing systems of processes and describe the processes based on criteria such as naturalness, ease of implementation of the processes with computer software, and usefulness of the output from each process for end users of the system. An ideal systems analyst can develop knowledge about the different types of processes that are useful when modeling the system, finding those that will be most effective for both the potential user community and the organization that is implementing and maintaining the system. Understanding the professional and personal needs of potential users is often achieved by studying the users of current systems, as well as knowledge of how humans interact with systems. At the same time, detailed knowledge about an environment, such as an organization, ranging from knowing about plans for the future, accounting details, and a variety of business functions, is necessary for effective analysis of a business information system.

1.3 Processes

Understanding the nature of processes can lead to a fuller appreciation of the information produced by processes, including what a process can and cannot produce and how it produces output. Systems may accept input information and then operate on

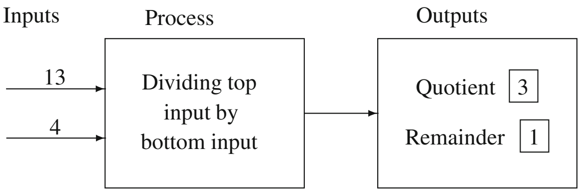

Fig. 1.1 Input being manipulated by the process, followed by information being produced in the outputs of the process. The two variables in the output could have any of a range of values, but the process produces 3 for the quotient and 1 for the remainder. The values taken on by these output variables are the information about the process and the inputs

the input, with information appearing at the output about the process, including how the process operated on the input. If a sketch artist produces a drawing of a cube, the reader might conclude that the artist has some knowledge of perspective and geometry, while a sketch by the author of children playing on sand dunes would probably generate comparisons of the author's artistic skills with the artistic skills of many five year old children. What an artist views is the input to an artistic process that eventually produces a piece of art. The input to a process, such as the input shown in Fig. 1.1, comes from another information-producing process. The output in Fig. 1.1 may serve as the input to successive processes. The sand dunes and children that the author sketches were themselves produced by other processes, just as the author as an artistic processor was himself produced.

The universe may be understood as a large network of information-producing processes. The size and scope of these processes allows one to have a hierarchy of processes, with the processes consisting of other processes of different sizes at different levels. The size or scope of a process effects the information produced by the process. A large process, such as a self-contained entity like a cruise ship, has output that is informative about the ship and its functioning. Within the ship, one would likely observe people speaking, and thus producing information, as well as computers producing bits of information. Each process produces information at its output, and as processes are nested, one wholly within another, information is being produced completely inside processes that themselves are producing information completely within other processes that produce information, and so forth. Note that these processes are nested within each other and exist within a hierarchy; both nested and hierarchical models are similar, but may emphasize different aspects of information and processes and their interactions.

The size and scope of processes that humans manipulate and build has changed as humanity has progressed, as in Fig. 1.2. Through developments in science and engineering, technology has increasingly harnessed microscopic and even atomic level processes, while at the same time larger and larger objects and processes are being built and used. The growing range of malleable processes, and the characteristics of their output, are at the core of civilization's progress.

Processes are everywhere and exist in everything. As mechanisms, a process can be viewed as a device that transfers physical forces through pressure and the movement

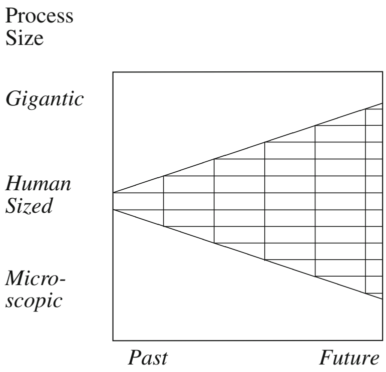

Fig. 1.2 The size and complexity of processes developed and used by humans (*shaded area*) spreads as time advances

of physical components of the system. Likewise, electro-magnetic systems produce output through the action of electronic and magnetic forces and phenomena. Clearly, processes are ubiquitous, and the information they produce is everywhere. The nature of this information reflects on the nature of different types of processes and inputs.

A process may be thought of as a black box that produces an output, based in part on the nature of the input to the system. One may expand on this model, to suggest that a process is any delimited area of the universe, with those forces or object impinging on this area being the input to the process, and all characteristics at the boundary of this process being the output. The nature of a process does not need to be fully understood to make use of its output, and while some processes are well understood, many processes are mysteries to those who routinely use their informative outputs. In the industrial world, people often use reverse engineering to determine the functionality of a black-box-like product, so that the processes' production capabilities can be reproduced. One might be able to explain the production of information in the output of a process given both the nature of the process and the input. One may also be able to explain the nature of a process given both the input and the output of the process. Similarly, the input to a process may often be inferred by knowing the process and the output of the process.

Processes and their outputs lay at the core of an understanding of information. When a process processes, and the output of the process has a value that is informative about the process and its input, information has been produced.For example, a calculator that takes two inputs, the numbers 3 and 2, along with a command to add, produces a 5 in the output window. This output reflects both the inputs and the nature of the process occurring in the calculator. At the same time, one can view the numbers entering the calculator as leaving the human who enters the data, with output information from one process, the human, becoming input information to another process, the calculator.

Processes operate through physical methods and other phenomena that can transmit forces. For example, gravity can cause the movement of objects, effecting a process. Electrical forces can attract or repel objects, depending on their electrical charge. These, and other forces detectable with the proper instrumentation, function as the action initiators for processes. Depending on the forces and the items being acted upon, processes may be continuous, where the output is observed as a smoothly varying value, or the outputs may be discrete, whole values. Some phenomena may be discrete in some senses and continuous in others.

Light is a form of energy transmitting process that has both discrete and continuous aspects to it. The discrete aspect of this is a photon emitting process that produces individual particle-like photons. These are detected by photon detectors, which can count the number of photons that have been registered by the photo-detector. Light also has a continuous, wave-like aspect. A beam of light that moves through two slits a short distance apart produces a wavelike interference pattern on a flat surface behind the slits. If one drops an object into water, waves move away from the point where the object entered the water. If two objects are dropped near each other in the water, the ripples from the two objects meet each other and where there is a high point in each of the two intersecting waves, there is a higher resulting wave. Interference occurs when the inputs are wavelike; particles without a wavelike component that move through two slits continue in a straight line and will not exhibit an interference phenomenon.

The operation of a process using force acts in such a way that the acceptance of the input occurs at the same time or before the output is produced. Observing the actions of the forces, in conjunction with the temporal order of occurrences, may allow the causes, the inputs, to be separated from the actions at the output.

Processes are often unique, but some processes are essentially the same as other processes. If one considers addition as a process, clearly $3 + 2$ produces output information that is equivalent to the output produced by the process adding $2 + 3$. One arbitrary rule would be to order all numbers being listed in an addition problem in increasing numeric order, so that $2 + 4 + 3$ and $4 + 3 + 2$ and $4 + 2 + 3$ will be represented, consistent with this ascending number rule, as $2 + 3 + 4$. Clearly, this order makes no difference in addition, but the order is quite significant in subtraction and division, and this canonical representation becomes important when determining the equivalence of processes. Normalized descriptions of these processes can take a variety of forms. One of the most useful ways to capture the informational output of processes is to describe all processes with the same descriptive approach. With a standardized or normalized form of a description, one can make statements more easily comparing two different processes, allowing one to make claims about the relationships between two processes, such as that the two processes are equal or that one is more efficient or simpler than the other.

Processes that produce information may be simple or complex. A simple device might be something such as a box, which turns on a single output light when an input button is pressed and turns off when the button is released. Electronic calculators are far more complex, being able to take a single digit number as input and display as output the square root of the input. How much more complex is this *square root* device than a simple device where one can press a button to turn on a light? If one were to implement a *square root* device using gears, springs, and other physical devices, it would take an individual a long time to design such a device and a long time to build it. A light switch is intuitively simple, and an electro-mechanical implementation could probably be designed faster than designing the *square root* device and might similarly be built much quicker. Generalizing from these methods of evaluating a system, the complexity of a system might be measured by determining the amount of time it takes to design the system, or the amount of time it takes to build the system or the amount of time necessary to process a specific type of input information. If one can imagine a machine that makes other devices, the characteristics of the machine that is produced and its operation may be measured in part from the performance of these device producing machines.

The complexity of processes is measured by considering the size of a computer program that would produce this type of output. By considering the sophistication of the program, as run on a specific standard machine, one can compare the complexity of different processes. A process that produces the same output, for example, the number "7", no matter what the input, would be a simple program, with a single statement such as *Print* "7". If we measure the information inherent in this program, we might count the number of statements or the size of the program. A program that calculates the square root of the input, on the other hand, will be much larger, with mathematical routines, numerous looping algorithms that attempt to move closer and closer to the correct value, and so forth. This program is far bigger, and carries far more information, than the simpler *Print* "7" program above.

The relationship between the information in the input to a process and the information in its output may be understood in different ways. While one may envision a computer program or a mathematical function as an analog for a process, the input and output relationship may also be understood in terms of other models. For example, as a mathematical mapping between the input and the output, the set theoretic understandings of mapping may be brought to bear on understanding the input and output relationships. Similarly, if one begins with a linguistic statement of the input values and one considers a statement of the output, the transformation is viewed in terms of language processing or, more formally, term rewriting systems. While other models are useful in the understanding of output information, the process models have been extensively studied and clearly provide useful tools for studying information.

1.4 Process Output

Information is carried in the variables and the characteristics of the output that are produced through the operation of the process. The output is determined partially by the inputs to the process. The inputs and outputs may be represented using algebraic variables such as x and y. When referring to the quantity x *plus* 2, x can be any number. These algebraic variables may be understood as symbolic names or classes that can take on a range of values, the characteristics. Using the variables and ranges of variables enables one to make statements about groups of items, such as that *Apples grow on trees*, without forcing one to iterate through the entire set of existing apples to say that each individual apple grows on a specific tree.

Variables may be viewed as specific containers that may hold any of several possible objects or representations. In a computer, a *byte* of computer storage might be designed to hold 8 bits of data, whether this is a number from 0 to 255 or one of 256 representations of natural language characters (e.g., the uppercase and lowercase letters used in the English language, punctuation marks, digits, and some additional control characters) or graphic characters used in video displays. When one electronically places new representations in this container, future operations "reading" this location will find the new contents and operate based upon these contents (until they are, in turn, replaced with other contents).

The information containing variables may be physical containers, such as a bucket, or conceptual containers, such as algebraic variables. When one refers to a variable in a language, such as an algebraic language, or a natural language, such as English, a copy or representation of what is in the container often is produced for further use. Thus, depending on the context, referring to a container may produce a copy of the contents of the container, or it may refer to the container itself. The expression x plus 2 gives one a copy of the value that is in the x variable or container and adds 2 to the value, without modifying the contents of the container. Information is what is in the container; the operation associated with observing the contents of the container is informative about the container's contents.

Output characteristics represent features that are about something, and thus in some way help identify the members of a class. People are often identified by their height, weight, and hair color, but there are other features we could use to identify the individuals. Alice could be identified as having 25 kg of weight per meter of height, and her height as one decimeter for every cat she has owned. These latter descriptions and characteristics are unusual. Features that are simple, easy to describe, and commonly used are often best at characterizing something, and the values that these features take on is the information about the object they characterize. Based on the philosophical principal of Occam's Razor, preferring simplicity to complexity, the simplest explanation for a set of variables is assumed to be the most desirable explanation. The information contained in a set of variables is best interpreted in a simple and succinct manner. Prose descriptions provide a language that refers to the values taken on by the characteristics. The statement "The blond guy over there is holding the frying pan" refers to one male, identifying the person in a unique manner with the hair color and a description of what he is holding.

The output of a process that is the input to a second process can be said to be capable of being *observed* when the second process incorporates this information as its input. This second process is an *observer*. System characteristics that are *not observable* are variables internal to the process. When adding two numbers together, for example, there might be a "carry" from one column of numbers, whose sum exceeds 9, over to the next column. Similarly, when considering whether to carry an umbrella today, people consider many possibilities; all this processing uses internal variables.

The information produced at the output of a process can be *represented* in different ways. Consider a process that mixes cyan and magenta inputs to produce a blue output. Objects of this color might be produced. Numbers representing the degree to which each of red, green, or blue are included could be produced. A descriptive English language statement could be produced.The term for the color blue is written several different ways using several different natural languages, from *blue* in English to *bleu* in French to *blau* in German. Different natural languages may use different orthography in recording natural language information, with Roman characters, Arabic, Dravidian, or Japanese Kanji scripts or symbols. Similarly, the number 2 can be represented by the Arabic "2" or when using Roman numerals by "II." One could represent 2 and 3 by "11" and "111" or by "1111" and "111111," doubling or tripling the number of "1"s used to represent the value in question.

If we represent a binary number composed of several ones and zeros, such as 1010, using an exact copy of the original number, that is, 1010, then having one of the bits randomly changed due to an error or noise produces an error in the resulting representation. The recipient or observer of the number might not realize that an error had occurred and might make bad decisions or arithmetic errors. However, if each bit were repeated five times, so that 1010 was represented as 11111 00000 11111 00000 and the observer knew the representation rules, then a simple single bit error could be easily corrected by noting what the majority of bits are for each of the 5 bit groupings. The observed value 11011 00000 11111 00000 would be understood as having mostly 1's in the first group of 5 bits, suggesting that it should be entirely 1's, correcting the error. When corrected, this would then suggest that the original bit pattern was 1010.

Representations can also be produced so that the original value is hidden to maintain secrecy. Imagine we develop a secret key, such as 1100, and a system that tells one that the original representation of information to be kept secret should be flipped (an original 1 becomes 0 and an original 0 becomes 1) wherever there is a 1 in the key. The original data 1010 with the key 1100 would produce the encrypted or secret representation 0110, with the leftmost 2 bits "flipped" because the leftmost two bits of the key are 1's. Using an opposite technique, the encrypted representation may be "decoded."

Representations exist within a specific physical and force-based environment, with the output of a process usually thought of as being of the same type of energy transmitting phenomena as the process itself. Thus, a mechanical process would be expected to produce a mechanical output, while an electrical process would similarly produce electrical output. The output for the process in question is the result of a very small, final sub-process linking part of the operational components of the process to the distinguishing features in output of the process. These resulting phenomena are observable by other processes, and these observations carry information from a source to a destination.

These processes can be understood as existing in almost any granularity or scale. When observing signals conveyed by lights from an electrical or optical process, one might observe hundreds of small locations on the surface of a light emitting device, or one may treat the device as a unit, and observe whether the entire unit is turned *on* or *off*. Individuals may learn the social conventions associated with observations. For example, children may describe a cat, rather than a large quantity of fur, growing from skin, that grows on flesh, and so forth.

As processes operate, their outputs carry information in the characteristics of their output. These characteristics are about a process and the input to the process, and the characteristics are the information in the output variables.

1.5 Communication

In the past century, sciences have been developed that provide room for many new applications of information. As a phenomena studied in many academic disciplines, information has been widely discussed, with field-specific discussions about information often being conducted in terms reflecting primarily the interests of that specific field.

As an example of field-specific views of problems, consider the following joke from a familiar genre of stories: an engineer, a physicist, and a mathematician are staying in rooms in a hotel and each is confronted with a fire outside their hotel room. The engineer wakes up smelling smoke, opens the door, sees a fire outside her door, and, after filling a wastebasket with water, dumps the water on the fire outside her door and puts out the fire. The physicist similarly smells smoke and opens his door. He sees a fire hose, and after calculating the water pressure and the angle at which the hose must be held, he uses the hose to put out the fire. The mathematician smells smoke and opens her hotel room door. She sees the fire and notices the fire hose. She then exclaims "Ah, a solution exists" and then closes her door [155]. People in different cultures and career paths view the world in their own ways; many of us could make-up a humorous story about the same topic but reflecting the interests of our own type of work and disciplines at which we wish to poke fun.

Before the last few centuries, the idea of knowledge existed in many cultures, but the examination of the concept of information was less common, when discussed at all. More recently, as more mechanisms have been developed that could produce information that was of direct use to humans, people began to notice that something was coming out of processes that could be described, measured, and formalized. For example, one of the earliest motivations for the examination of information may have been phenomena associated with light. Light, fire, and other forces were often seen as productions of the deities, and the characteristics of individuals, including the knowledge they have and that they gain, was often viewed across cultures as something special, often divine.

As scholars continued to describe phenomena as rigorously as possible over the centuries, increased examination of information-related phenomena resulted in developments leading to modern descriptions of information. Information theory began as a coherent and precise discipline largely due to the ideas developed by Claude Shannon, an electrical engineer and mathematician. Most active during the middle of the twentieth century,Shannon developed a theory of communication that provided elegant techniques, allowing others to analyze the representation and transmission of information during the rapid advances in electronics and computers that occurred in the decades after World War II. Shannon has been compared to Albert

Einstein and some found the comparison “unfair to Shannon” [147, p. 15]. Shannon contributed a rigorous framework to the study of information and communication that is mathematically elegant. His ideas lead to the development and expansion of the academic discipline of *information theory*. At the same time, people became aware of what Shannon’s model did not provide for those interested in understanding information.

If Shannon can be said to have developed a castle to hold the idea of information, there were certainly people before Shannon who built sheds and small houses and who Shannon acknowledges contributed significantly to his understanding of information and communication. From within the research laboratories of American Telephone and Telegraph over a decade before Shannon became active, there were early discussions of ideas about information. In the early 1920s, Harry Nyquist described the factors affecting communication speed in telephone circuits. Interestingly, Nyquist noted at the beginning of his most famous work that “this paper considers two fundamental factors entering into the *maximum speed* of transmission of intelligence...” [137, p. 324]. Discovering that there is a maximum speed for transmitting intelligence was a significant insight. Learning that there is a limit to how much “intelligence” can be transmitted, and that this limit is a function of other specific factors, significantly advanced the scientific and social scientific understanding of information.

The capability of a circuit to transmit at a given rate, referred to as the *bandwidth*, can be used to determine how much intelligence (to employ Nyquist’s terminology) may be transmitted per unit of time. For example, one might measure the capacity of an Internet connection going into a private residence as one million bits per second. In a second article, Nyquist showed that discrete signals could be transmitted at twice the bit rate of the bandwidth. He also showed in his original work that the shape of a signal effects how much can be transmitted, and that data transmission is most efficient given a specific set of characteristics for the waveform used to represent intelligence.

The intelligence that Nyquist understood to be transmitted consisted of anything transmittable over telephone cables using an arbitrarily chosen representation scheme. Nyquist understood that the way information is represented affects the amount of intelligence transmitted. He then was able to describe the “relative efficiency of various codes in transmitting intelligence” [137, p. 345], comparing several different Morse code systems. He goes on to propose an *ideal* code, which, while not optimal using current standards, shows the development of considerations used in such an endeavor. By developing a code for transmitting text that has the lowest number of signal elements per letter, producing the shortest average coded representation per letter of intelligence, the speed of transmission and throughput is maximized, holding all other signal and coding characteristics constant.

Ideal performance of a communication circuit allows the transmission of as much intelligence as possible during a given period. This statement can be further qualified so that, for example, when desiring error free transmission, one can seek as much error-free intelligence transmitted as possible during a given period. When error free transmission is not required, the goal may again be modified to provide the maximum rate with a specified error rate.

By the 1920s, Nyquist had shown that communication over a telephone circuit can be improved by varying several features, including the representation system, the bandwidth, the amount of power, and the amount of noise induced into the system from outside the system. Writing a few years after Nyquist's major work was published, Ralph Hartley wrote about the "Transmission of Information." More conceptual than Nyquist's work, Hartley advanced both the mathematics of information theory and, probably more importantly for history, conceptualized *information* as something that was represented, transmitted, and received. "Information is a very elastic term" [91, p. 536] Hartley noted, and he attempted to "set up for it a more specific meaning" in his work [91].

When informational signals are received, the recipient can usually estimate, based on what was received, what message was sent by the transmitter. If I transmit a picture of my cat, this is not a picture of my computer or a picture of a tree or a picture of my automobile. Hartley notes that as the transmission of information progresses, "we can say that the information becomes more precise" [91] as more options are excluded. Similarly, the statement that "the apples are red" eliminates many possible grammatical subjects for the sentence by noting that the topic of the sentence is "apples" and not pears, oranges, or peaches. Stating that the color of the apples is "red" eliminates other apple colors, such as "green" or "yellow." By successively eliminating possibilities, the precision of the information in a statement increases [91, p. 536].

How much information is added when options are excluded, such as when one receives a message that provides information and removes doubts about the state of nature? The amount of information that is provided by excluding options grows at a rate that is exponential with the number of options excluded. Through counting the number of different words that could be inserted into a particular position in a sentence, one can compute the probability that a particular word occurs in a sentence at that point. The more uncommon a term is, or the more options for terms exist in a particular spot in a sentence, the more information is present by having a particular term. Suppose that we have a fair coin that is tossed in the air twice, both times landing *heads* rather than *tails*. With the first toss of a coin, we may exclude the option that the first coin tossed might land as *tails*.[1] If we toss the second coin and it too lands *heads*, we have excluded the following pairs of options: *tails* followed by *tails*, *tails* followed by *heads*, and *heads* followed by *tails*. In this case, we have more than doubled the number of options excluded when we have only doubled the number of coin tosses. This example can be generalized, suggesting that the number of options excluded by coin tosses will always be 1 less than 2 raised to the power of the number of coin tosses, or 1 less than one of these numbers: 1, 2, 4, 8, 16, 32, 64, 128, and so forth. The position of each number in this series represents the number of coin tosses, starting with 0 (no tosses). A single coin toss (using the second number

[1] For purposes of this work, coin tosses are assumed to be fair, although the nature of the nonrandomness in coin tosses is clear [55].

in the series) excludes a single option, while using the third number in the series for two coin tosses suggests that 3 sets of options are excluded.

When there are 4 options present, then there are 2 bits of information. A bit here represents a *binary digit*, a unit of storage for either a 0 or a 1, possibly representing the absence or presence of a binary feature or characteristic. When this doubles to 8 options, there are 3 bits of information, with 16 options, 4 bits, and so forth. A single bit can take either of two values, such as 0 and 1. The relationship between the sequence of options 1, 2, 4, 8, 16, 32 and the number of bits, 0, 1, 2, 3, 4, 5 has the second sequence as the logarithm to base 2 of the components of the first. The logarithmic operation has the effect of transforming the number of options into the number of bits that exist, with the bits being additive. When there are the 4 options associated with 2 bits and 8 options with 3 bits, adding the $4 + 8 = 12$ options give the wrong number of actual options present (which would be 32), but adding 2 bits + 3 bits = 5 bits is correct.

Building on the earlier work of Nyquist and Hartley, Shannon developed his own measure of information. While the earlier engineers proposed measures based directly on the number of occurrences or absences of symbols, Shannon changed the problem and measured information associated with the probability of an event. He suggested that the negation of the logarithm of the probability of an event would provide a useful measure of information and that taking the logarithm of the probability provides the measure with desirable characteristics.[2] Logarithms can be used with probabilities as they were with options. The following probabilities produce the corresponding information, measured in bits:

Probability	Information (in bits)
1/2	1
1/4	2
1/8	3
1/16	4
1/32	5

Probabilities are logarithmically transformed to bits, as can the number of options, as was shown above.

Shannon noted that several variables, such as time, the bandwidth of a telephone or radio circuit, and the number of switching devices used to manipulate data "tend to vary linearly with the logarithm of the number of possibilities" [168, p. 4]. The use of a logarithmic measure also made the development of Shannon's measure mathematically simpler and more elegant, important factors for an active researcher spending months or years manipulating a single formula.

Hartley and others working at this time treated different options as though they were equally likely. While explicit about this assumption, there appears to have been little work in the information field assuming different probabilities of events

[2] Logarithms here are computed to base 2 unless otherwise stated, measuring the amount of information in bits.

until Shannon's work a decade later, which was able to address situations where different events would often have different chances of occurrence.The decision to use probabilities, as Shannon did, instead of choosing counts, as did Nyquist and Hartley, represents a choice faced by many modelers of systems. Models can address phenomena by counting events and objects, using what are referred to as combinatoric methods, or models can incorporate the probabilities of these same events and objects. Combinatoric models often relate various sets, considering the relative size of the different sets. Probabilistic models are based not on directly counting but instead on chances that events will occur. There are several different types of probabilities and different foundations for probability theory, leading to different interpretations for models and for subtle differences in explanations of information.

Nyquist and Hartley paved the way for Shannon by providing measures and a framework for the study of information. Shannon acknowledges in the second sentence of his most famous essay that "a basis for [a general theory of communication] is contained in the important papers of Nyquist and Hartley on this subject" [168, p. 31]. Besides the factors mentioned above, Hartley explicitly pointed out the necessity for the "elimination of psychological factors" [91, p. 536] in such a model of information. Providing for a non-psychological model of information set the stage for the development of a true science of information. This does not exclude meaning from the study of information: we will return to this later when we consider the meaning of "meaning" within information studies and the notion of subjective economic value of information to one or more individuals.

Claude Shannon published an essay titled "A Mathematical Theory of Communication" after World War II because of pressure from his colleagues to assemble and publish details of the work he had been conducting for several years [147, p. 26]. It was published first as an article in the *Bell System Technical Journal* in 1948 and then in 1949 as one of two essays in a book titled *The Mathematical Theory of Communication* that also contains a more philosophical and less mathematical essay by Warren Weaver on general applications of this model of communication. While Bell Telephone Laboratories patented many of the developments of that time period, Shannon's work on information theory was not patented, possibly because of its theoretical nature; the lack of a patent on this work certainly assisted in its diffusion [197].

For Shannon, communication produces a precise or approximate copy of a message at one location, the destination, that was originally at another location, the source; it is these messages that are *information*. These messages and this information may have semantic (or other) meaning and may have economic value, but these are explicitly not addressed by Shannon in his seminal work. For Shannon, information continues to be something that is encoded for transmission through a communicative process referred to as a channel.The output of the communication process exists at the input to the receiver, and the message is decoded at the receiver, providing information about the communication process and its input.

The communication process shown in Fig. 1.3 may be understood as a *source* communicating to a *destination*. The source provides its message to a *transmitter* through a loss-less connection. The transmitter communicates through a channel to

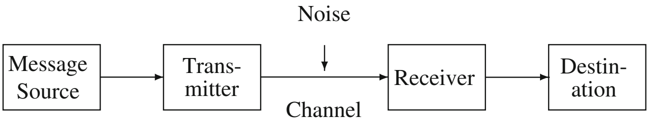

Fig. 1.3 Shannon's channel model

the *receiver*, which receives the message and gives it in a loss-less manner to the destination.

Communication begins with a message that the sender intends for the recipient to possess. The received message is similar, or identical, to the original message that was sent. Given a perfect channel with no interference of any sort, the received message will be a perfect copy of the original message. In more realistic circumstances, the received message will be similar to the message sent, with modifications being induced by noise from outside the system and components that modify the signal within the system.

Communications travel through a channel, a medium or process that can carry the message. A channel may accept noise, fluctuations induced from outside the system, changing the message being transmitted, and thus possibly transforming the message being sent from the source to the destination. Messages may be transmitted over a physical distance or through time. A mechanical device transfers a physical force, such as when one presses on one end of a piece of wood and the force is transferred to the other end. A force may be transmitted by electromagnetic forces, such as light or radio signals. A voice spoken at a radio station or at one Internet site is received as a physical motion of the material in a loudspeaker at the message's receiver. These messages can move information through time, such as a book that is printed and then read years later.

When the originator of the message produces the message to be sent, it is provided to a transmitter that is at one end of a communication channel leading to the receiver associated with the message's destination (Fig. 1.3). Electronic hardware on both ends of a physical telephone line are designed to use the same electronic protocol for communication. Both the transmitter and the receiver in these electronic systems are designed to use the same electronic waveforms for a given message, with a specific signal being represented for transmission (and then reception) by a combination of signal voltages, frequencies, and other electronic characteristics. A person speaking produces sound waves; a listener wishing to hear the message must have the physical ability to detect the same type of sound waves that the speaker produced.

A receiver acts at the end of the channel to capture the transmitted message and place it into a form suitable for the recipient. Filters in a receiver act to modify signals, sometimes to enhance some frequencies and to decrease the strength of signals at other frequencies. Some *band-pass* filters only allow signals within a certain frequency range to pass. Other filters are skewed to increase the ability of a channel to successfully transmit and receive a message by progressively blocking signals of lower frequencies with a *high-pass filter* or material of higher frequencies with a *low-pass filter*. Text filters similarly may pass or remove email containing certain terms or phrases or originating from certain addresses.

Shannon published his work on communication just as a revolution began in computational devices. Early computers were being described in the popular press as the United States experienced a growth spurt after World War II, accompanied by an increase in optimism regarding technology. Vannevar Bush, a U.S. science policy adviser during World War II and the president of the Massachusetts Institute of Technology (MIT), proposed a "memory extender" that was a precursor of the computers to be developed over the next decades [206]. Shortly after Shannon published his foundational work, transistors began to be developed. Computers would soon begin their decrease in size that has lead to the small portable computational devices of today, instead of the room-size computers used when Shannon was conducting his research.

Shannon's written work on information in the context of communication systems is beautifully written, if one is comfortable with the level of mathematics. Shannon is one of the best technical writers of his era, with an originality, simplicity, and precision that contributed to making his work powerful and popular. Shannon's work has limitations, and these limits were not obvious to everyone. The model of communication is not inclusive of semantic and economic values, and those interested in such human phenomena as *intent* or *meaning* or *dishonesty* became discouraged with the use of Shannon's model as it became increasingly obvious that it was limited to communication circuits as defined by engineers, or communication systems similar to such circuits.

The source and the destination may have different coding systems. When the word *love* is used by a source and a destination person, it is quite common that the concepts being generated and received are different: however, the sounds for the spoken word *love* are similar enough at both transmitter and receiver for both to agree on the "message" sent and received. When the recipient has a different context for the received message than that held by the sender, the recipient might be said to receive a different message than was originally sent. If Alan is a native speaker of Japanese and he speaks to Bill in Japanese, with Alan assuming that Bill understands Japanese, but where Bill understands virtually no Japanese, Bill will usually interpret what was said very differently than what Alan intended. While Bill's ear has received the sound waves carrying the message, his brain probably has not received the meaning that was intended by Alan.

Warren Weaver, who in the next decade became president of the American Association for the Advancement of Science, was a scientific generalist whose breadth enabled him to easily transfer ideas from one field to another. In addition to his essay that accompanies Shannon's work, he may be best known for his work on automatic translation of natural languages, work that crosses disciplinary boundaries. Weaver begins his essay on information with the sentence, "The word *communication* will be used here in a very broad sense to include all of the procedures by which one mind may affect another" [195, p. 3]. While Shannon describes the characteristics of telephone circuits in a rigorous and general manner, Weaver moved this work forward to examine the communication between one mind and another. Here he was perilously close to crossing the line into portraying Shannon's work as applying to meaning, a component of what many consider exists in communication between one person's mind and another person's mind. Shannon and his predecessors in the communication engineering area clearly reject the inclusion of "meaning" (Shannon) or "psychological factors" (Hartley) in their communication and information models. Shannon noted that "the subject of information theory has been sold, if not oversold"[167, p. 462]. How far Weaver intended to move toward meaning and more subjective components of communication remains the subject of debate. However, Weaver did attempt to keep his notion of information consistent with that of Shannon and the earlier work by Nyquist, when he said:

> Information is, we must steadily remember, a measure of one's freedom of choice in selecting a message. The greater this freedom of choice, and hence the greater the information, the greater is the uncertainty that the message selected is some particular one. Thus, greater freedom of choice, greater uncertainty, greater information go hand in hand [168, p. 109].

Clearly, whatever Weaver and Shannon intended, some people took information theory further than what Shannon and Weaver actually proposed. For example, before Shannon began developing his ideas, Hartley began viewing information as a thing and possibly as intelligence, moving beyond a simple mathematical model of electronic signals, and may be responsible for moving information into a fuzzier area than some were happy with in later decades.

It may be difficult for one now to place Shannon's ideas about information and its measurement in the context in which they were developed. Engineers had developed working radio and telephone systems that were rapidly improving over the course of the twentieth century. Most of the emphasis was on the transmitting and receiving devices and their individual components. Before beginning with the central problems in information theory in an encyclopedia article about information theory in 1968, Shannon noted that "this theory... is quite different from classical communication engineering theory which addresses the devices employed but not with that which is communicated" [166, p. 212]. His revolution started by developing a science of what was communicated, expanding beyond the science of that which communicates. In some cases, that which is communicated may have meaning, but in other cases, the information may have no meaning. Weaver notes that

> The word *information* in this theory, is used in a special sense that must not be confused with its ordinary usage. In particular, information must not be confused with meaning. In fact, two messages, one of which is heavily loaded with meaning and the other of which is pure nonsense, can be exactly equivalent, from the present viewpoint, as regards information [195, p. 8].

While meaning cannot be placed easily into the communication-based model of information that Shannon proposed, meaning, knowledge, and other phenomena can be viewed as the informational output of processes in the mind, given a more process-based understanding of information.

Shannon provides a deeper and more elaborate model of information and communication than the simpler models proposed by Nyquist and Hartley when he describes a mechanism used for communication, with a difference being drawn between the

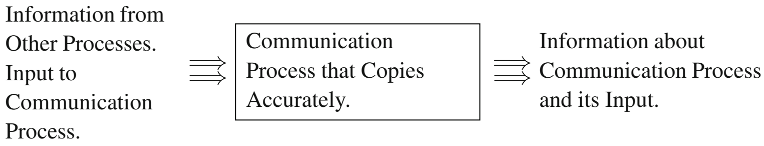

Fig. 1.4 Communication process

source and the transmitter and between the receiver and the destination. Shannon also provides a model of information content with a probabilistic model of information, based on the probabilities of individual messages that might occur. By developing coding models that serve as the basis for much of modern cryptography and error detection and correction, Shannon advanced information studies far beyond what was available through the periods of Nyquist and Hartley. Shannon and his colleagues developed probabilistic models that placed an emphasis on the science of studying uncertain characteristics. One of the many existing foundations upon which the present work rests is that of Shannon and others who developed and supported what has become known as *information theory*.

The Shannon model of information is consistent with a process-based model of information (Fig. 1.4). Anywhere that there exists a variable that can take on several values, there exists one or more processes that may be viewed as producing the values in the variable.The Shannon model may be viewed as a type of process based model and the amount of information present in the output of a process may be measured using Shannon's logarithmic measure applied to the probability of the output of a process.

1.6 The Physical World and Entropy

As information theory blossomed after World War II, some of the discussions about information developed links between the new areas of information theory and computation and the previously existing models of entropy from the field of thermodynamics. These models from physics provided the creators of information theory with possible relationships between physical mechanisms and various forces that carry information, although some of the newly developed connections between information theory and thermodynamics may have been strained.

The relationships between items in a system are the *structure* of the system. People observe structures present as observable patterns so often that they may not be conscious of the structure. However, differentiating between content and structure, what is there compared to how it is organized, may be a useful distinction in many cases. The structure of an object or system carries information.For example, when one is considering information about an object's structure, one may consider the end points to lines or curves used in drawing the outline for an object, the drawing's

Fig. 1.5 Diffusion of dye. Information (as structure) is lost as the diffusing process operates, increasing entropy as viewed by physicists

complexity, line or curve types, and the number of points connected to a specific point. The more complex the structure, the further it is from random and the more information is carried by the structure about how it was produced. A quick way of measuring the complexity of a structure would be to precisely describe the object and then ask yourself how long is the description of the structure. The macroscopic description of a house is almost always longer than the description of a child's wooden block, suggesting that the house is macroscopically more complex.

Information associated with the structure and the location of physical entities grows as systems become increasingly complex. Think of what happens when a drop of concentrated red food coloring is placed into a glass of clear water that has been sitting still for several minutes. The coloring will enter that water as a red drop, and then grow and disperse until after several minutes, all the water will have taken on a red tinge (Fig. 1.5). The description of the location of the food color is relatively simple when the food coloring is first placed into the water: it is all within a conceptual drop shaped area centered at the location where the drop was first placed. Later, after the food coloring has diffused the description of the location of each red molecule increases in complexity, each red molecule taking a random path through the water.

Accidentally spilling a set of organized items on the floor results in them being arranged in a somewhat random pattern. It usually takes longer to place them back into their original arrangement than it did to spill them on the floor in the first place; once structure is gone, it is very difficult to recover. Removing the structure takes less energy in most cases than creating the structure.

When a system of objects is mixed, the randomness of its structure, as well as its dissipated energy may be said to have increased. The second law of thermodynamics suggests that as entropy, as described by physicists, increases in a physical system, the useful energy decreases over time and the used energy increases, keeping the total system energy constant over time. When the balls on a pool table are placed together, or a drop of food coloring is placed into a glass of water, there is a regularity to the locations of the placed objects. If the pool balls are struck by a cue ball or the dye in the water is allowed to disperse, the objects' locations become less regular and increasingly random. An increase in physical entropy introduces an element of irreversibility to the system, as things that are mixed are not easily unmixed, just as red food color left in water is not easily reconcentrated back into its original form. The useful energy available in the one structure may be irrevocably dissipated.

Observing the state of a variable or a system also expends energy. More precisely, it is the resetting of the observing system to its initial state after every observation that expends energy [17, p. 235]. Observing the state of a variable is essentially making a copy of the object observed at the site where the observation is taking place. In many circumstances, the location where the original is to be copied already has other information. The act of observing therefore may require an initial erasure of the previous contents, an act that requires energy in all cases. Observing has an energy cost to it, and this cost is often proportional to the amount of information observed (and copied).

The term *entropy* has been used by many in information theory to describe the average information in a set of messages. However, the average information, in an information theoretic sense, may be viewed as the negative of the entropy as computed by a physicist. While placing a negative sign in front of the entropy formula in information theory may address this problem, as does using the term *negentropy* instead of entropy when discussing information theory, both disciplines often refer to entropy in a way that may conflict with some of the definitions used by the other. While an increase in entropy is seen when dye disperses in a cup of water, an increase in negentropy is seen when living beings consume material from outside themselves with less structure and convert the material into greater structured material, such as more cells, within themselves. The difference between the two forms of entropy may also be understood as the amount of randomness in a system, compared to the amount of information in a message or knowledge about the randomness, with both measures increasing as the randomness increases, although the message decreases the randomness in the recipient.

Historically, information theory used ideas that were popular in physics, such as the idea of entropy, but more recently physics has begun using ideas that have matured within the field of information theory. For several decades after Shannon's early articles were published, one would find general works on information theory often spending time linking information theory and thermodynamics. In part, this was due to a true link between them, and in part because of a desire to place information theory on firm, physical grounds. As information theory and physics have advanced, information has been placed in a more fundamental role within physics. Physicist John Wheeler [199] claimed *it from bit*, and the expression's increasing rate of occurrence in the physics literature indicates the growing importance of viewing physics and the study of physical phenomenon with information theoretic and computational approaches.

1.7 People and Information

Information is often described as being contained in, and in some ways related to natural language, ideas, and knowledge. When people speak, they convey information, interpreted by those who hear the speech as information. The information in a listener's head can be turned into beliefs and sometimes knowledge, based upon the relationships between the beliefs, what is already believed or known, and the observed world.

The information in the statement, "Let's get some coffee," may be viewed as being in the meaning of the statement, contained in the acoustic representations of the words that arrive at the listener's ears. One often hears the statement "I received the information" when a letter arrives, or a person is described as *informed* about a topic if they are knowledgeable about the topic in question. Information is carried by a speech on the television news, the written text on a web page, or the diagrams in a textbook. A person becomes informed through observation, a process that takes available sensory information and produces information in the brain, which, along with reasoning, produces knowledge within the mind.

There are many ways that information can be used by humans and there are many different ways of measuring information being used by humans. The idea that the amount of information is proportional to the rarity of a signal has utility for engineers but seems to provide a weak tool for those studying many aspects of natural language. The information in natural language might be better captured by counting meanings in a statement or measuring the economic good to which receiving a statement might be put. Discipline-specific ideas about information may prove inadequate when taken outside of their discipline, while working with discipline-independent ideas of information allows one to work with the same rigorously defined and understood fundamental concepts across disciplines.

Information about how to obtain a glass of water, or searching for tomorrow's weather using an Internet connected computer, has a value to many of us. The economic value of the information is often independent of the number of bits of information in a message. The value of information is calculated as the expected value of a human's actions if they did not have the information compared to the expected value of the human's actions with the information.

Useful information begins with a process producing output, with a second process assigning a value to the output or potential output from the first process. The second, valuing process includes cognitive or economic processes that may *use* the information from the first process and assigns a value to its output. The valuing process produces output, which may be one or more traditional output variables about the internal states of the valuing process and the output of the first process.

Information often takes on a value through explicit decision making processes that use the information to improve the expected quality of any decisions made. One might decide whether to enroll at a school full-time, working on an advanced educational degree, or to accept an employment offer, allowing one to do what seems to be the type of work one has always dreamed of doing. Advanced knowledge that the job will turn out to be boring, or that the education will turn out to be more exciting than one ever imagined, would obviously affect the likelihood of accepting the employment offer.

In interpersonal relationships, the value of information will differ from one individual to another and will vary depending on whether one would like to compete or to cooperate with the other person. Knowing what a friend's preferences are, or whether the friend is feeling a bit ill, or whether they are concerned about something at work, might allow one to make better decisions. Decisions may improve the friend's happiness, while using information can improve one's own situation at the expense of a friend, or at the expense of an opponent.

Information processes may be viewed as incorporating language, beliefs, or knowledge passing through them or being generated by them. These process may also have value based on other processes that can use the information the produced. These informative outputs will be examined in more detail in later chapters.

1.8 Hierarchies of Processes

Information is traditionally understood as carrying messages or facts about something. These facts are encoded or represented by information when transmitted and then decoded by the recipient. The image of a flower that you perceive with your eye is represented by electromagnetic patterns among the photons moving from the flower to your eyes. These representations serve as references to an object, a physical or conceptual entity, based on the characteristics of the referent.Every reference serves as a name or identifier. "Bob" serves as a non-unique name for the author of this book, while "Bob, the author of this book" serves as a unique identifier. Both of these names serve as representations for the author, as might a drawing by a three year old child or a photograph on a passport.

How do these representations or encodings occur? One way is through the application of a list of processes, each generating information that serves as the representation for progressively more complex processes as one moves up a hierarchy of processes [47,140]. The encoding (and eventual decoding) occurs within a process, and the representation process and the information it produces can be studied as one studies any other process or set of processes. A set of processes that *communicate* or *observe* can be viewed as a hierarchy of processes and representations. This is evidenced by the series of visual processes existing in the brain, where signals from the basic perceptual processes eventually reach processes generating social and emotional responses [161]. Transmitting at one level uses levels below it in the hierarchy to represent the message from a higher, initiating level. Similarly, the higher level may serve as a representation of a lower level, specifically, of the characteristics of the level below it. For simplicity, we always assume that the "lower" level represents the smaller and "more physical" processes, while those "higher-up" are assumed to be more complex and more sophisticated processes.

The processes used for transmitting and receiving often have a single output or single set of outputs; a single output process is referred to as a *function*. Many arithmetic operations that we learned as a child are functions. Addition, for example, may take two numbers and produce a third number that contains the sum as an output, containing information about both the inputs and the addition process. Functions might produce a complex object, such as when a medical database system takes a medical record number and returns that person's medical record.

Key components within hierarchical processes are inverse processes and inverse functions. Consider the *addition* function, which accepts as input two numbers, producing a sum. One might add 3 and 14 to produce the sum 17. The inverse of the operation is subtraction. Here we take the 3 away from 17 to produce 14. An addition function may be written with the function name followed by the data (listed within parentheses) upon which the operation operates. Common notation would have one write an addition problem as something like this: $14 + 3 = 17$, but we might also denote addition as $Add(14, 3) = 17$. Rather than having the addition operator (sign) or textual command written between the two numbers being added, such as in $3\ Add\ 2 = 5$, the operation is placed in front with its inputs following it in parentheses, e.g., $Add(2, 3) = 5$. These functions can be included within other functions, so that

$$Add(14, Add(2, 1)) = 17.$$

Here the inner addition of 2 plus 1 is computed first, producing 3, and this is then added to 14, yielding 17. Functions that accept other functions as inputs are sometimes referred to as second order functions.

Assume we have an inverse Add function, denoted as $InverseAdd$, that computes the first parameter in the parentheses minus the value of the second parameter. This function is the subtraction operation, although for our purposes it helps one to consider this operation as the inverse of addition. Consider the mathematical problem

$$InverseAdd(Add(17, 3), 3),$$

which adds $17+3$ together and then subtracts 3 from this, producing 17. The addition function has an inverse function, $InverseAdd$, which "undoes" what the addition function does.

Another arithmetic function is the multiplication function, $Multiply$. For example, multiplying 4 times 7 would be written as $Multiply(4, 7)$. An inverse form of this function, $InverseMultiply$, divides the value of the first parameter by the value of the second parameter. Thus,

$$InverseMultiply(100, 25)$$

divides 100 by 25, yielding 4. Consider the expression

$$InverseMultiply(Multiply(17, 5), 5).$$

Here we have 17 multiplied by 5, yielding 85, which is then divided by 5, yielding 17. One could replace both occurrences of 5 with any other single number for both occurrences, and the result would still be 17.

Consider now a non-arithmetic case where language is sent and received through a sonic medium, as in the bottom portion of Fig. 1.6. The information in the language is transmitted from sender to receiver through the encoding of the linguistic statement in a sonic form. This produces two forms of representation for an idea: Language and Sound. This process may be represented by

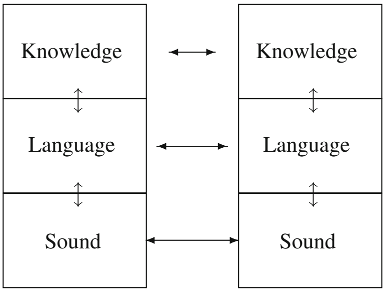

Fig. 1.6 Hierarchical model of human communication. Communication moves from left to right or right to left

$$Language(Sound(InverseSound(InverseLanguage(x)))),$$

where $Language(x)$ is the language process of the listener. If I desire to transmit (speak) a word to you, the $InverseLanguage()$ function encodes the word, producing the input to the $InverseSound()$ function that produces speech. This function places the coded sound into the atmosphere, where it is picked up and decoded by the listener's $Sound()$ function, which decodes the message into a form acceptable to the $Language()$ function, which decodes its input, producing the original word when there is no interference with the processes or their input.

When adding a layer to this hierarchical model of human thought and speech, one might add both encoding and decoding functions for the added layer, e.g. $InverseSound()$ and $Sound()$. For example, we may add both "knowledge" and "inverseknowledge" layers to the top of Fig. 1.6, incorporating a more sophisticated aspect of human thought, as follows:

$$Knowledge(Language(Sound(InverseSound(InverseLanguage(InverseKnowledge(x)))))).$$

Language and knowledge manipulating processes are discussed in Chap. 5.

Each function in Fig. 1.6 can be thought of as a black box that accepts communications from above (on the left side) and processes the input. The output (from the bottom of these devices) *indirectly* feeds into the inverse of the function (flowing upwards on the same level on the right side) but goes *directly* into another function (on the left) below it. Lower level functions are necessary if communication is to occur; additional functions must be added until the function at the bottom can provide a physical linkage between the bottom layer on one hierarchy and the corresponding functional layer on another hierarchy [47]. This bottom layer is where the innermost processes produce information along one or more dimensions.

A hierarchy can be used to model the assignment of value to information. A lower level process produces information that may go into a higher level process that then uses the information. The use of a lower level's output by a process above it may provide a value to the process above it. The nature of the processing in this higher level

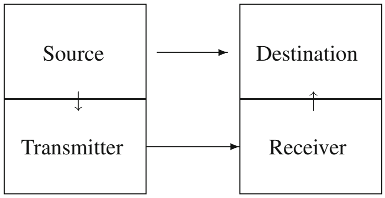

Fig. 1.7 Shannon's model of communication viewed in a hierarchical context

process and the input to it from the process below may produce a value, an indicator of what trade-offs might be made to produce an equivalent level of satisfaction.

A given level may have below it two or more lower hierarchies, or "legs." For example, a given human mind may communicate through several different physical processes, such as speech, gesture, and the written word. The same knowledge process is working in the brain, but the transmitted message can travel to its destination through several different physical media.

Information is "transmitted" by the actions of these hierarchical processes. As the processes operate, characteristics that are essential to communication are passed through by the processes. Characteristics incidental to the communication often remain unused. For example, although each of us pronounces words in a measurably different manner, most can understand the vast majority of words received that are spoken by native speakers of the languages we learned as a child. Many phonetic features such as minor speech variations are ignored.

What should constitute a layer receiving and producing information within a hierarchy of processes consists of several layers? Each layer is defined by the interfaces with the layers above and below as well as by the process producing the layers above or below or itself. Any process can be continually broken down until very small physical phenomenon are reached. For the study of macroscopic environments, the arbitrary choice of layers is required. They are most beneficially selected based upon naturalness considerations, processes and layers being described so that the processes are easily understood. The defining limits of a function are somewhat arbitrary, and hierarchies can usually be decomposed further. Decomposing processes is reductionist, and while there are arguments that in some cases the phenomena in micro-processes is intrinsically different than the phenomena found in macro-processes [22, 88], there is no clear evidence that in many cases the physical phenomena in the sum of the processes is qualitatively different than the phenomena in its parts. At the bottom of each hierarchy is a layer that contains the physical mechanism that allows communication to occur. This bottom layer is referred to as the "physical layer," no matter what size the function. The physical layer, as with all the other layers, can be defined such that it is very large, so that one layer performs a particular large, complex task, or it may be very small, with several different smaller layers performing different parts of a larger task.

Most communication uses inverse functions. Communication exists if and only if there is information produced by a process and then this encoding of the input is reversed through an inverse, or an approximation of the inverse, of the original process. The producer of the information that is presented to the original process communicates with the recipient at the output of the inverse process. Figure 1.7 shows a hierarchy of processes that allow a process to take what is transmitted, as input to a communication process, with the receiver at the output, and with the source communicating with the destination at a higher level of the hierarchy. However, the functions may not be inverses on all levels. If one speaks using one set of processes, and someone else listens using different processes, or reads lips through still other processes, there are not exact physical inverses for speech production and speech hearing or perception. The listener, on one scale, has to convert what is heard or seen on the lips back into meaning in a large scale inverse of what the speaker originally did. Similarly, a radio transmitter may function differently from a radio receiver, with the transmitter, for example, often using much more power to operate than the receiver. They are inverses on a functional level, but may physically be implemented in a different way.

The characteristics of one level may be passed on to the inputs for other levels, or the values may be dropped. Loss of a characteristic may be irrecoverable; the information will be permanently lost if the characteristic is independently valued and cannot be inferred from other characteristics. For example, a lost value for the author's gender can be easily recovered from the author's name or knowledge about the presence of a beard, while the loss of the value for the author's first name may be a permanent loss. It cannot be easily deduced from other characteristics, such as gender or the presence of a beard. For many people, the study of information examines the transmission of these characteristics' values from one level to the corresponding level at the destination. One can study knowledge moving from one person to another. Similarly, one can study beliefs moving from individual to individual.

Let us assume that a hierarchy exists with knowledge as the top layer and belief as the layer below it. When knowledge is to be transmitted (assuming that it can be transmitted), it is encoded in terms of beliefs, which are then further encoded into lower level objects. When these are received, the perceptual input is eventually transformed into belief. The characteristics of these beliefs then become the characteristics of the received knowledge.

The use of the hierarchical model allows the student of information to focus on the level in the hierarchy that is of greatest interest, rather than descending into a debate about whether information is of one nature or another, whether it is located at one level in the hierarchy or another. Information is produced at the level of interest, and the processes at that level *are* worthy of intellectual discussion, as are the information-producing processes at all the other levels in the hierarchy.

The hierarchical view of communication and information movement can also be applied to social science domains, such as the study of information transfer between members of a society. Information appears to spread from individual to individual in a manner sometimes referred to as information or technology diffusion [38, 160]. The number of people with a certain unit of "knowledge" grows exponentially, with the growth slowing when a sizable percentage of the people already know the information that is being diffused. Information that a national leader has been assassinated moves thorough a large part of the population in minutes, while urban myths or popular jokes spread at a much slower rate, and information theory spreads at a snail's pace.

Each individual may be modeled as his or her own set of hierarchies, with various physical communication links running between one person's hierarchy and another's. The hierarchies of most individuals will be squid-like, with tentacle-like structures sticking downwards to the different physical layers used by the one individual. These layers can communicate with other beings or media. All physical communication methods may be assumed to have similar upper levels in comparable people, and thus similar knowledge level functions.The lower layers differ from medium to medium. One layer may involve reading or writing literature. A second is based on conference presentations, as well as the informal conversation that takes place among older friends and new acquaintances. The physical movement of humans when employment shifts from one organization to another results in information being transferred, albeit by the physical movement of the employee from one site to another.

The transfer of new technology often uses linguistic or pictorial means. A lower level layer used for technology transfer will often be the same for many differing technologies, simplifying our understanding of the process. Some forms of technology transfer, such as changes in employment, result in different connections to the physical layers at the bottom of the hierarchy. While Shannon's model of information describes some of what goes on lower in the information hierarchies, many models describing technology transfer aim at higher levels of the information hierarchy.

The relationships between information hierarchies and understanding information as the characteristics of a process's output allows one to study the commonalities in many of the information phenomena that have been proposed. Information transfer occurs when the legs for these hierarchies connect. The diffusion of information thus depends on the nature of these hierarchical legs, and as information spreads from one hierarchy to another, information diffuses. The number of interconnections between each hierarchy partially determines the rate of diffusion. Information hierarchies serve as a valuable tool when examining a variety of information phenomena in a variety of domains.

A different perspective on understanding multiple information processes is by looking at the *interface* between one process and another, with the process interfaces providing information in both directions across a conceptual surface. Imagine using an information retrieval system or search engine on a computer. One first looks at the light emanating from the display surface, and then one possibly enters commands or requests. After checking that data has been entered properly, one might click on a "Search" or "OK" button. The displayed results are then seen by the searcher, and the user may further interact with the displayed results. There are numerous processes crossing from the searcher to the system and other processes that take input from the system and produce input for the searcher. Such an interface serves as a useful grouping of information producing processes where communication between processes occurs. Interfaces often have numerous information producing processes operating in parallel or "side-by-side," whereas a hierarchy of processes often consists of nested processes or one process being a "meta-level" process above another process.

1.9 Defining Information

Information may be defined as the characteristics of the output of any process. Over time, numerous other approaches to defining information have been used successfully in the description of phenomena across a range of disciplines. Understanding information as *ideas* has been widely held, as has information being defined as *useful* data [26, p. 47]. Some ideas about information are specific to an academic discipline, such as entropy is to physicists and chemists, bit rates are to electrical engineers, linguistic intent is to linguists as knowledge is to epistemologists. Some ideas are much broader than a discipline, with the notion of information having value being useful far beyond the boundaries of economics.

When developing a model describing a phenomenon, one captures the characteristics of the phenomenon, as well as relationships between the characteristics. These models can then be used to capture the underlying nature of a phenomenon, such as information. By making a claim about the nature of what is occurring and often why it is occurring, a model may be more explanatory than a simple definition. Scientific models ideally describe what is occurring, predict future occurrences, and explain why things occur as they do.

A definition of information may take several forms, focusing on different characteristics of information. For example, a definition of information may describe what is essential, what must be present for information to exist. Shannon's communication channel has essential components such as the transmitter, receiver, source, and destination. A definition of information may also be based, in part, upon requirements imposed from outside, such as moral values or field specific values, as well as relevant essentials for the field. It is difficult to imagine an economic model of information that would not contain the concept of economic value. Similarly, one cannot imagine a religious definition of information not containing a reference to actions or intervention by a deity or universal force consistent with the particular religion's perspective.

Common definitions of information have similarities to more precise and operational academic definitions. The definition that was used on the children's television program *Sesame Street* was that information is "news or facts about something." For an academic definition of information to be useful, the definition needs to have enough of the elements of a common definition to allow the user to see an obvious relationship between the common and academic definitions and the academic definition needs to allow one to address many of the same problems that are addressed when using a common, informal definition. For example, the process-based model of information has information always being "about" something.

Understanding what is information or informative, what value there is in information, or how much information there is in a situation or environment can lead to an overarching view of the different facets of information. While there can be such qualitative definitions for information, one can also use a quantitative perspective to emphasize that measuring information is a key aspect of understanding the phenomena, along with the characteristics included in definitions. Measuring characteristic values may use one of the following relations: *equal*, *less than*, or *greater than*, or the negation of one of these. For example, knowledge that one data communication rate exceeds another provides a value useful in describing an information relation.

Information phenomenon may be usefully defined and measured in an objective sense. Many academics now analyze texts by viewing what the reader takes from the text and what the reader brings to the text, rather than pursuing the meaning that the author may have intended when producing the text. While literary critics benefit from using this methodology, it can lead science to the abyss: if definitions and measures are only subjective or immeasurable, being only relative to the individual and what they bring with them to the science or measuring environment, can a scientist produce laws or theories that have any meaning for anyone else? Although academic theories may be interpreted somewhat differently by different individuals, we believe that accepting and trying to work with such measures is more profitable and more likely to lead to societally beneficial results than choosing to work through definitions and measurements that are completely subjective.

1.10 Characteristics of Information Phenomena

Examining the commonalities that occur in different aspects of information production, use, and analysis can be enhanced with a general view of information focusing on the outputs of informative processes. Understanding the regularities that occur across a wide range of informative processes lets one manipulate these processes to achieve desired goals, from printing documents to living a better life.A science of information that captures these regularities allows one to better understand information phenomena and to better and more efficiently use information by describing, predicting, and understanding it. A maxim captures the value of understanding a rigorous model of information: "there is nothing more *practical* than a good *theory*." A science provides a set of principles for a discipline, with rules governing the relationships between variables of interest to the field in question, whether the variables are subjectively or objectively determined. By conducting the science consistently and rigorously, observations of behavior and physical phenomena can be used in disproving hypotheses and in predicting future actions. Given our hierarchical "stacking" of processes, our science of information is consistent both with well understood low level physical processes, such as electrical communication through a wire, and with higher level, more abstract processes, such as meaning or knowledge.

Information is produced by processes. The approach taken here assumes *a priori* that processes must precede information, and that information cannot exist without a producing process. For example, one can speculate that information about the beginning of the universe is only available from the processes that operated then. When presented with a question such as "what came first, the chicken or the egg?" one must treat processes as predating the first information.

Understanding the nature of processes is essential to the understanding of information, with forces producing information at the output of the process. Static objects and entities within processes are manipulated by forces; the objects may be large, or too small to be observed with our best instruments, or they may be a complex of neurons in the brain having a given set of characteristics. The forces may be mechanical, such as the physical pressure exerted on mechanical gears in a clock. Physicists often describe four fundamental forces: gravitational, weak, electromagnetic, and strong forces. All forces act upon entities within a space, and the interactions of these forces and entities together form the operations in processes.

Describing a process is a key component in the study of producing information. One of the most popular methods for describing computer programs and many other types of processes is to model the process as a program on a universal Turing machine, an abstract model of a processing device. Because the Turing machine is a simple conceptual machine, numerous mathematical proofs have been developed about it and its permutations, making it a useful and powerful model of a processor. Because of its simplicity, simulations of Turing machines have also been widely implemented on real computers. Programs for the Turing machine can emulate a range of processes, and the description of a process is often best studied by analyzing its implementation on a Turing-like machine.

Processes also can be described by noting functional characteristics of a process. The size or complexity of a process is measured as proportional to the size of the smallest version of the process. Once the process has been placed into a standard form, one can then measure the information content or complexity, allowing for the comparison of the size or complexity of one process with corresponding characteristics of another process. The size of the universal Turing machine is constant, and thus one only needs to compare the varying size of the program that emulates the process, without regard to the size of the Turing machine. A process that could be written most compactly as a short program is a simpler process than a program representing a more complex process that can only be written in its shortest form as a longer program.

Processes can operate, be modified, and modify themselves over shorter or longer periods of time. One may view each changed process as a different process, since, in fact, they are different, or one can view a single process as stable but changing in some describable fashion. The changes might be random, or change may occur in a regular way.

The speed of operation of processes varies, depending on how the fundamental changes in the process take place. For example, automobiles and airplanes can change their speed of operation under operator control, and many engines have their speed governed by the amount of fuel provided to the engine. Electronic components often switch from a 1 to a 0, for example, at a rate related to the voltage, the force behind the electronic actions. Light travels at one speed in a "vacuum" and at a much slower speed through matter, sometimes at less than half the speed it travels in a vacuum. The process that transmits light clearly varies in its speed of operation. Information is produced at different speeds, depending on the speed of the information production process.

The output of a process is a set of one or more variables, each variable being capable of taking on one or more values in a given period. These variables are composed of entities and forces that allow the variable to hold this state or value. For example, an electronic clock might display the same hour digits for an entire hour, with the digits being displayed being held in this state through electronic components and the interactions between electrical forces that move through the circuitry.The values taken on by the output determine the *state* of the output. This state constitutes the information produced by the process. The set of all the possible states determines the set of all possible information that may be produced by process.The values held by output variables may be discrete, the variable holding either one value or another, or the variable may hold values that have a continuous range, such as the output for a music player as it varies in volume. Variables may also hold a complex of values. For example, a multi-course meal can be viewed as a variable, with a number of characteristics.

The characteristics of the variables must be observable for the information produced by the process to be the input to a second process. The values of variables are information, but we treat variables that are unobservable as producing unobservable information. Clearly, if variables are observable then we have the more commonly discussed information: observable information.

Variables providing information may take a brief amount of time to "settle" and become stable.In the simple case of a computer memory bit being set, enough force, such as electricity, must be supplied to turn a switch "on" or "off." It may take a few thousandths of a second to turn on some switches, while other devices might be orders of magnitude faster, such as those switches found in electronic computers.The value of a variable may shift, taking on one value and then another. This variation may be due to several phenomena. There may be random processes present that shift the variable from one value or state to another in unanticipated ways. One common reason for this is that there is not much force used in representing the variable. The greater the amount of force present, the more stable will be the variable. Once the information in the variables stabilizes, useful and constant observations may take place. Observing the value of the variable takes time; an observational process must occur that makes a copy or surrogate of the original data for movement to the observer.

The amount of information present in the output is proportional to the number of characteristics that each variable can exhibit. Some models are proportional to the number of different states associated with different characteristics, while other models treat the rarity of a state to be positively related to the information carried by the output being in that state.

The input to a process can be the output from another process, and the discussion above of the characteristics of the output captures the nature of inputs as they are produced by other processes.

The internal characteristics of a process, as well as the nature and functioning of inputs and outputs to a process, can be used when describing information in real world situations.

1.11 Studying and Using Information

Information means many things to many people, and it has become increasingly popular to discuss information and to associate oneself with information. Some refer to the present as *The Information Age*. Many universities have added *Information Science* and *Information Studies* programs to their curricula. Understanding information has become increasingly important, but much of the discussion is polarizing, arguing using one or another disciplinary view of information rather than trying to move outward toward an inclusive and precise model of information. For example, the common language definitions often assume that the recipient of information is learning something new; if it is not new, it is not information. Or they may assume that to be information, the material must be useful.

Specialized groups, such as academics, have their own concepts that they label as information. Those in the physical sciences have examined the structure and entropy of systems in terms of information. Electrical engineers and computer scientists examine information moving over a channel and being stored for manipulation. Language is often understood by linguists and philosophers as carrying information, with meaning in language, with knowledge having a large information component.

The operation of processes and their output characteristics are consistent with a range of philosophical views. Entities and processes may be understood as abstract entities that anyone can reason about, or one can treat entities and processes as things we can learn about only through sensation. The equation $2 + 2 = 4$ is true in any universe for those who view parts of the universe as abstract, such as Platonists do. Empiricists, on the other hand, would treat something as true only when the results are sensed, with various positions existing between the Platonists and empiricists. Described in the next chapter, Godel's models of processes and the solution of Diophantine (integer) equations may be understood as abstract formal models of producers of information, independent of the universe in which they exist. Empiricists view the world through their senses, understanding the information as the sensed value of observed output characteristics. Information is viewed as occurrences produced by physical or forceful processes. Those who rigorously model the nature of knowledge may move further to describe the relationships present between observed input characteristics, the processes, and the nature of the output.

Regardless of a problem's domain, each process that produces output by expending energy produces *information about* the process and its input.Information may be understood as the characteristics of the output of a process. This approach to information can be applied to any situation where there are processes, that is to say, everywhere, and to understand information, one should have an appreciation for processes.We now turn to a fuller examination of ways that processes may be described and modeled as they produce information.

http://www.springer.com/978-3-642-31189-5

Information from Processes
About the Nature of Information Creation, Use, and Representation
Losee, R.M.
2012, XVIII, 242 p., Hardcover
ISBN: 978-3-642-31189-5