Preface

What is information? Information is an important concept that is studied extensively across a range of disciplines, from the physical sciences to genetics to psychology to epistemology. Information continues to increase in importance: the present has been referred to as the "Information Age," one of the most rapidly growing occupational categories is "Information Technology," and the subject of "Information Science" is seen by many as a rising academic discipline. Information has seldom been studied in a consistent way across different disciplines; this work presents a single understanding of information that can be used in each of these disciplines.

One may understand information in a variety of ways. For some, information is found in facts that were previously unknown. For others, a fact must have some economic value to be considered as information. Information may be something with meaning, words put together into sentences, brush strokes on an artist's canvas, or a crescendo in music. Other people emphasize the movement through a communication channel from one location to another when describing information.

In all of these instances, the information is the set of characteristics of the output of a process, the characteristics produced from the set of *possible* characteristics of the output. The information produced by the process is *about* the *process* itself and *about* its *input*. This *informative* output may be observed and analyzed, and the output may be captured by a statement describing the information and the relationships between the information containing variables. Such an informative statement may describe the state of nature at the output of the process.

Information is assumed a priori to come from processes when the processes produce output. While this can be argued philosophically consistent with many ideas of causation, rules about how processes operate or how fast they might operate, or the characteristics of information itself, may be addressed based upon other theoretical considerations. For example, using mathematical methods, Turing was able to show that certain outputs cannot be guaranteed to be produced by certain types of processes, providing a formal limitation on the types of information that can be produced. Other types of information are studied best scientifically, such as by empirical observation or by computer simulations. For example, the number of occurrences of different English language terms in a body of text may be studied and the reasons for the observed distributions may be suggested and laws governing the distribution of terms developed.

The characteristics of the output can be observed or inferred using tools that support the ability to observe. The nature of an observation is partially dependent on the type of characteristic, such as whether a variable is discrete or continuous. Because the output of a process is determined by both the process and the input to the process, the output of the process can be inferred given knowledge about the process and its input. For example, given that a process adds two numbers together and that the input numbers are 2 and 3, one can infer the output of 5, inferred information based on knowledge about the process and its input.

How information is understood as a concept is largely dependent on how one understands processes. It is probably not coincidental that the rigorous understanding of information theory in communication systems developed at the same time as computer science and the rapid advancement of the electrical and software implementation of processes. Claude Shannon, who deserves much of the credit for developing the study of information theory in communication circuits, performed pioneering work on digital electronics that serves as the foundation for computer circuits used today. Advances in Mathematics and Computer Science have shown that in some types of general processing, certain informative outputs cannot be produced. Imagine that certain types of information cannot be produced in our universe but might be producible in a different universe with different laws, or that information might be produced by my type of computer but not by your type of computer. These issues with information illustrate how important it is to understand the nature of information-producing processes when trying to understand information.

Several scholars have developed formal models of processing and its limits. Some models, like solving Diophantine (integer) equations, can be shown to have equivalent limiting properties to all processes meeting certain requirements. Other models, such as Universal Turing Machines, have the procedural characteristics of familiar computing devices, making their use attractive in procedural situations. In addition, Universal Turing machines can be shown to be equivalent to a large class of processes that accept input and then produce a single response, and the universal Turing machines' limitations, as well as the limitations in solving Diophantine equations, have proven to be useful in learning the circumstances in which information cannot be produced.

Lambda calculus similarly models processes, although its functional nature may not appeal to some who prefer to view a process as a sequence of steps. There are innumerable programing languages with differing means of expressing processing, as well as of describing the information used and produced by the process. Given the variety of ways that one can express the operation of a process, the variation in the ability to make statements about the program may prove useful. Anthony Hoare proposed a technique allowing one to develop proofs of computer program correctness. By using this, one can show that a process either does, or does not, produce the desired output from a given set of inputs. Computational complexity provides tools that can predict the type of execution speed that would be expected in executing a process. Learning such formal methods in disciplines such as computer science or electrical engineering provides many of the tools useful in modeling information processes, while learning programing languages may give one an appreciation of how information processes function.

The operation, inputs, and outputs of processes are inferred through a number of mechanisms. Knowledge about the type of internal structure of the operation, such as whether it is discrete or continuous, may be used to learn about the operation. Valiant developed a formal theory about how Boolean functions are learned, serving as the basis for a model of processes composed of Boolean components and the processes' Boolean output. Estimating the nature of the operation of a process increases difficulty when there is little prior knowledge about the type of operations that might be performed. Conceptually, the knowledge about the process is continuously or discretely changed to better reflect the process being learned. Processes may Darwinistically evolve so that those processes with the greatest likelihood of survival are used as the basis for further evolution, and those processes with a lower likelihood of survival are less likely to be used or reproduce and are thus less likely to have a chance to evolve.

The amount of information at a process's output is proportional to the number of possible characteristics that are available. There are a number of measures that could be used in calculating the amount of information, such as the number of characteristics possible, the square of the number of characteristics possible, the positive square root of the number of characteristics possible, the logarithm of the number of characteristics possible, and so forth. All these measures increase as the number of possible characteristics increases, and decrease as the number of possible characteristics decreases. However, they increase and decrease at different rates. An information measure that is proportional to the number of variable characteristics that might occur has advantages over other types of measures. Consider (1) a variable with four possible values, numbered 1, 2, 3, and 4, or (2) two other variables, an *even number* (yes or no) and the value *greater than* 2 (yes or no). It would be desirable that the amount of information observed in picking one of the 4 characteristics is equal to the sum of the amounts of information observed when using the second example with two variables. One can show mathematically that the only time this holds true for any number of possible variable characteristics is when the measure of information is related to the logarithm of the number of characteristics. This type of logarithmic measure, developed and used by physicists in the nineteenth century and by electrical engineers in the early twentieth century, is not the only possible measure of information, but this measure (and minor variations) are the only measures that have this desirable additive property.

In the following chapters, information concepts and phenomena are examined in an effort to understand them; given a hierarchy of information processes, where one process uses other processes in the hierarchy. For example, knowledge, belief, truth, and facts can be viewed as inter-related entities within a human information hierarchy, along with other types of processes, such as perception and observation, that occur at a lower level in the hierarchy. Intelligence exists when reasoning processes manipulate data, whether using perceptual data from the outside world or stored information already in the brain. The perception of information from the outside world and stored information are generally below the more intelligent reasoning processes that occur higher in the information hierarchy. The representation of knowledge is founded in the output of these higher level processes. Lower level processes provide data that are transformed into this knowledge, as are beliefs and other mental phenomena. Similarly, both natural and artificial reasoning can be understood in terms of processes. Processes can infer from several inputs to a generalizing output. These may be used to model probabilistic, fuzzy, or other formal descriptions of reasoning.

1 What Information from Processes Provides

A model of information consistent with an information-from-process approach provides a number of advantages over using some other models of information.

- Different academic fields have developed and used their own definitions of information. *Understanding information as the output of a process is discipline independent and those in any field can use this model of information for their own internal purposes*, as well as to use it to communicate with those in other fields. Since processes occur everywhere, the information-from-process model can be applied to any domain. Knowledge gained from studying information from processes in one discipline can be applied to other fields that use similar processes or have similar inputs to processes. The emphasis on domain independence allows us to focus on commonalities, rather than on the differences that often seem to be the focus of domain specific definitions. Different academic domains exist to solve specific classes of problems, but this does not mean that one should ignore common characteristics between one's own problems and other types of problems.
- In some academic domains, information exists almost completely within and between humans. We assume that *all domains are of equal interest and value to information scientists.* While we respect the beliefs of those who place humans or primates or other groups in a privileged position in scientific studies, we assume below that studies of information both within certain species and outside certain species are equally useful and relevant.
- *Understanding information apart from other concepts, such as computers and people, gives one the ability to develop a broader and more inclusive view of information.* Many choose to study information only in the context of human communication, while others view information science as the use of computer technology by people. This addresses information on its own. Once information is fully understood by itself and the informational characteristics of processes

are understood by themselves, then information can be better understood and more effectively applied to other domains, such as humans or computer systems.

- *This model of information is applied across a wide range of domains,* showing the strength of the model, instead of just arguing for the superiority of this approach over others. Other models of information may also be useful, and the additional examination of how they may be applied across a range of domains would be interesting.
- Using information as the output of processes, *one can describe information systems, broadly construed, as more than just communication systems and as more than just computing systems.*
- Some information cannot be produced from certain processes. While most people do not think of information from a process as being intrinsically limited, a number of types of processes have been extensively studied and have been shown to have *formally demonstrable limitations on the types of information that processes can always produce.*
- A given unit of information can be understood as being produced by a number of processes. One can study a human vocalization by examining a particular chemical interaction within a cell, the actions of a neuron, or the way speech is produced through vibrating human vocal chords, or one can study humans as a species. People themselves can be studied many different ways, socially, electrically, politically, and so forth, and all of these can be viewed in terms of processes. Clearly, information in any situation can be produced consistently with a number of different types of processes, and using *the concept of information-from-processes allows all the processes that one might conceive of to have their information studied.*
- *This model of information is more fine grained than the approaches taken by many in the library, archival, and computer sciences* where pre-existing packaging of groups of information is used. Books, articles, and computer files may contain large quantities of information, combined in complex ways, so that a significant analysis of the individual statements of information and knowledge becomes very difficult. While methods of decomposing information vessels have been proposed, such as through the application of data and text mining, much greater power and more accuracy is available when informative statements are originally made available for individual analysis, possibly through the labeling of the statements by an author when they are placed into larger information vessels, such as books. The units that we choose to study make a large difference in the level of rigor and the productiveness of any study of information.
- *A process often produces a physical result that is information, being informative about the process and about its input.* While one might think of information as an abstract entity, such as an idea that is produced by a sentient process, in many circumstances a process produces something physical or useful, information as a thing. Information as a thing, such as a book or musical recording, is itself informative about the producing process or its input, and serves as a special case of information-from-processes, where the output is long lasting.

- While communication models may emphasize coding, *the information-from-process model emphasizes representations, without the implicit agreement existing between sender and receiver coding systems that often exists in descriptions of communication systems.* Representations are often particular to the producing process, and may be interpreted differently by different processes that find the representation in their input. If the image of a tree exists on the retina, the brain receiving electrical impulses from the eye does not recreate an actual tree, instead of producing an internal representation in the brain that is believed to differ significantly from the image on the retina.
- Understanding processes and their operation allows one to *describe, predict and understand the informative output of processes, leading to rigorous studies of information phenomena.* A calculating device that is asked to add the numbers 2 and 3 together can be predicted to produce information representing 5, and counting on one's fingers provides a mathematical reason for this particular answer. Many scholarly fields seek to describe accurately what is occurring; occasionally they correctly predict what will happen, and ideally each academic discipline will develop an accurate understanding of what is occurring within the domain's processes. In information related academic fields, one may begin with the understanding of a small number of types of information producing processes, and then move forward to the understanding of more complex information phenomena.
- *Understanding information has many aspects, including the application of psychological, mathematical, philosophical, and economic techniques.* The science of information begins with more philosophical and a priori aspects, such as the idea that all information is produced by something. Studying information from processes uses mathematical methods to describe processes and what they can produce, and on many levels information phenomena are scientifically studied using a range of theoretical and empirical methods.

The applications of information to human affairs may be grouped into several types. These types of processes are found throughout various types of information systems. Knowing the nature of the processes and what they can and cannot produce allows one to better understand specific information phenomena. Below, our focus is on information and the basic types of processes that can produce information; there is relatively little concern in many cases with *how* each type of process is implemented, as almost any process can be implemented in numerous different ways with different types of technologies.

The most basic process is the *copying process*, which produces at its output, the input. In some respects, this process does little. However, producing an identical copy serves to move information spatially from the input of the process to the output, such as from one location to another, or from one time to another. Some processes are stable over time, such as the nature of a printed book that remains relatively stable day after day. Other processes, such as a video display on an electronic device, are designed to change output quickly as the image being represented by the input signal changes. For humans, information is often fixed or frozen at the output of a process. A printing press prints a book, and a machine produces a disk containing a video. This fixing process produces information-as-thing. Some observations inherently change what is being observed, making true copying processes, that leave the original entity as it was, impossible in such circumstances.

Representing input information in different forms at the output of a process occurs in a wide range of processes. Representations are somewhat arbitrary; all other factors being equal, as long as the recipient of a representation can determine what the representation refers to, the representation has been adequate for the recipient. Representations may have other characteristics, such as supporting the correction of errors that might occur in the representation. If there were an individual letter missing in a book, could you still read most or perhaps all of the book? Other representation processes may be designed to take up as little memory as possible. One might typographically represent a light glowing with a 1 and not glowing with a 0, or one could use statements like *the light is turned on* and *the light is turned off*. Note that if the prose representation is used rather than the numeric representation, the loss of a single typographic character still allows one to infer the state of the light, while the loss of a single typographic character in the numeric case may result in the complete loss of information about the state of the light.

Information may be transmitted though a *channel*, a copying process that moves the information through a distance. A channel is often spoken of as a set of copying processes that have the specific function of transmitting information with as little loss as possible.

Reasoning processes emphasize the production of new information. Logical operations may be used to produce complex outputs based on the inputs and the logical processing. Quantitative processing may allow one to infer probabilistic information in the output. Numerous processes have evolved that result in humans having the special and possibly unique characteristics that they exhibit. Beliefs are produced through perceptual and reasoning processes. Truth may be determined through several different types of processes, such as consistency with other, accepted facts. Knowledge is produced in processes that incorporate several types of inputs, with traditional definitions of knowledge often allowing for the incorporation of beliefs, their truth, and justification for the beliefs in the production of knowledge.

2 CHIPL Programming

A programing package is available at http://InformationFromProcesses.org that contains the programing language CHIPL (Chapel Hill Information Processing Language) and a set of examples program files, some of which have been used to produce some of the figures in this book. The default output from CHIPL appears on a computer screen, but output is also generated that may be used as input by the LATEX document formatter (that is used to produce this book) and the gnuplot graphing package which takes output from CHIPL in some cases and produces LATEX input (as well as other forms of input).

Springer

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