STANFORD ENCYCLOPEDIA
OF PHILOSOPHY

Information

First published Fri Oct 26, 2012

Philosophy of Information deals with the philosophical analysis of the notion of information both from a historical and a systematic perspective. With the emergence of the empiricist theory of knowledge in early modern philosophy, the development of various mathematical theories of information in the twentieth century and the rise of information technology, the concept of "information" has conquered a central place in the sciences and in society. This interest also led to the emergence of a separate branch of philosophy that analyzes information in all its guises (Adriaans & van Benthem 2008a,b; Lenski 2010; Floridi 2002, 2011). Information has become a central category in both the sciences and the humanities and the reflection on information influences a broad range of philosophical disciplines varying from logic (Dretske 1981; van Benthem & van Rooij 2003; van Benthem 2006, see the entry on logic and information), epistemology (Simondon 1989) to ethics (Floridi 1999) and esthetics (Schmidhuber 1997a; Adriaans 2008) to ontology (Zuse 1969; Wheeler 1990; Schmidhuber 1997b; Wolfram 2002; Hutter 2010).

There is no consensus about the exact nature of the field of philosophy of information. Several authors have proposed a more or less coherent philosophy of information as an attempt to rethink philosophy from a new perspective: e.g., quantum physics (Mugur-Schächter 2002, see the entry on semantic conceptions of information), logic (Brenner 2008), semantic information (Floridi 2011; Adams & de Moraes 2016), communication and message systems (Capurro & Holgate 2011) and meta-philosophy (Wu 2010, 2016). Others (Adriaans & van Benthem 2008a; Lenski 2010) see it more as a technical discipline with deep roots in the history of philosophy and consequences for various disciplines like methodology, epistemology and ethics. Whatever one's interpretation of the nature of

philosophy of information is, it seems to imply an ambitious research program consisting of many sub-projects varying from the reinterpretation of the history of philosophy in the context of modern theories of information, to an in depth analysis of the role of information in science, the humanities and society as a whole.

The term "information" in colloquial speech is currently predominantly used as an abstract mass-noun used to denote any amount of data, code or text that is stored, sent, received or manipulated in any medium. The detailed history of both the term "information" and the various concepts that come with it is complex and for the larger part still has to be written (Seiffert 1968; Schnelle 1976; Capurro 1978, 2009; Capurro & Hjørland 2003). The exact meaning of the term "information" varies in different philosophical traditions and its colloquial use varies geographically and over different pragmatic contexts. Although an analysis of the notion of information has been a theme in Western philosophy from its early inception, the explicit analysis of information as a philosophical concept is recent, and dates back to the second half of the twentieth century. At this moment it is clear that information is a pivotal concept in the sciences and humanities and in our every day life. Everything we know about the world is based on information we received or gathered and every science in principle deals with information. There is a network of related concepts of information, with roots in various disciplines like physics, mathematics, logic, biology, economy and epistemology. All these notions cluster around two central properties:

> **Information is extensive.** Central is the concept of *additivity*: the combination of two independent datasets with the same amount of information contains *twice* as much information as the separate individual datasets. The notion of extensiveness emerges naturally in our interactions with the world around us when we count and measure objects and structures. Basic conceptions of more abstract

mathematical entities, like sets, multisets and sequences, were developed early in history on the basis of structural rules for the manipulation of symbols (Schmandt-Besserat 1992). The mathematical formalisation of extensiveness in terms of the log function took place in the context of research in to thermodynamics in the nineteenth (Boltzmann 1866) and early twentieth century (Gibbs 1906). When coded in terms of more advanced multi-dimensional numbers systems (complex numbers, quaternions, octonions) the concept of extensiveness generalizes in to more subtle notions of additivity that do not meet our everyday intuitions. Yet they play an important role in recent developments of information theory based on quantum physics (Von Neumann 1932; Redei & Stöltzner 2001, see entry on quantum entanglement and information).

Information reduces uncertainty. The amount of information we get grows linearly with the amount by which it reduces our uncertainty until the moment that we have received all possible information and the amount of uncertainty is zero. The relation between uncertainty and information was probably first formulated by the empiricists (Locke 1689; Hume 1748). Hume explicitly observes that a choice from a larger selection of possibilities gives more information. This observation reached its canonical mathematical formulation in the function proposed by Hartley (1928) that defines the amount of information we get when we select an element from a finite set. The only mathematical function that unifies these two intuitions about extensiveness and probability is the one that defines the information in terms of the negative log of the probability: $I(A) = -\log P(A)$ (Shannon 1948; Shannon & Weaver 1949, Rényi 1961).

The elegance of this formula however does not shield us from the conceptual problems it harbors. In the twentieth century various proposals for formalization of concepts of information were made:

- **Qualitative Theories of Information**
    1. **Semantic Information:** Bar-Hillel and Carnap developed a theory of semantic Information (1953). Floridi (2002, 2003, 2011) defines semantic information as well-formed, meaningful and truthful data. Formal entropy based definitions of information (Fisher, Shannon, Quantum, Kolmogorov) work on a more general level and do not necessarily measure information in meaningful truthful datasets, although one might defend the view that in order to be measurable the data must be well-formed (for a discussion see section 6.6 on Semantic Information). Semantic information is close to our everyday naive notion of information as something that is conveyed by true statements about the world.
    2. **Information as a state of an agent:** the formal logical treatment of notions like knowledge and belief was initiated by Hintikka (1962, 1973). Dretske (1981) and van Benthem & van Rooij (2003) studied these notions in the context of information theory, cf. van Rooij (2004***not in bib*) on questions and answers, or Parikh & Ramanujam (2003) on general messaging. Also Dunn seems to have this notion in mind when he defines information as "what is left of knowledge when one takes away believe, justification and truth" (Dunn 2001: 423; 2008). Vigo proposed a Structure-Sensitive Theory of Information based on the complexity of concept acquisition by agents (Vigo 2011, 2012).
- **Quantitative Theories of Information**
    1. **Nyquist's function:** Nyquist (1924) was probably the first to express the amount of "intelligence" that could be transmitted given a certain line speed of a telegraph systems in terms of a log function: $W = k \log m$, where W is the speed of transmission, K is a constant, and m are the different voltage levels one can choose from.
    2. **Fisher information:** the amount of information that an observable random variable X carries about an unknown parameter θ upon which the probability of X depends (Fisher 1925).
    3. **The Hartley function:** (Hartley 1928, Rényi 1961, Vigo 2012). The amount of information we get when we select an element from a finite set S under uniform distribution is the logarithm of the cardinality of that set.
    4. **Shannon information:** the entropy, H, of a discrete random variable X is a measure of the amount of uncertainty associated with the value of X (Shannon 1948; Shannon & Weaver 1949).
    5. **Kolmogorov complexity:** the information in a binary string x is the length of the shortest program p that produces x on a reference universal Turing machine U (Turing 1937; Solomonoff 1960, 1964a,b, 1997; Kolmogorov 1965; Chaitin 1969, 1987).
    6. **Entropy measures in Physics:** Although they are not in all cases strictly measures of information, the different notions of entropy defined in physics are closely related to corresponding concepts of information. We mention *Boltzmann Entropy* (Boltzmann, 1866) closely related to the Hartley Function (Hartley 1928), *Gibbs Entropy* (Gibbs 1906) formally equivalent to Shannon entropy and various generalizations like *Tsallis Entropy* (Tsallis 1988) and *Rényi Entropy* (Rényi 1961).
    7. **Quantum Information:** The qubit is a generalization of the classical bit and is described by a quantum state in a two-state quantum-mechanical system, which is formally equivalent to a two-dimensional vector space over the complex numbers (Von Neumann 1932; Redei & Stöltzner 2001).

Until recently the possibility of a unification of these theories was generally doubted (Adriaans & van Benthem 2008a), but after two

decades of research, perspectives for unification seem better.

The contours of a unified concept of information emerges along the following lines:

- Philosophy of information is a sub-discipline of philosophy, intricately related to the philosophy of logic and mathematics. Philosophy of **semantic information** (Floridi 2011, D'Alfonso 2012, Adams & de Moraes, 2016) again is a sub-discipline of philosophy of information (see the informational map in the entry on semantic conceptions of information). From this perspective philosophy of information is interested in the investigation of the subject at the most general level: data, well-formed data, environmental data etc. Philosophy of semantic information adds the dimensions of *meaning* and *truthfulness*. It is possible to interpret quantitative theories of information in the framework of a philosophy of semantic information (see section 6.5 for an in-depth discussion).
- Various **quantitative concepts of information** are associated with different narratives (counting, receiving messages, gathering information, computing) rooted in the same basic mathematical framework. Many problems in philosophy of information center around related problems in philosophy of mathematics. Conversions and reductions between various formal models have been studied (Cover & Thomas 2006; Grünwald & Vitányi 2008; Bais & Farmer 2008). The situation that seems to emerge is not unlike the concept of energy: there are various formal sub-theories about energy (kinetic, potential, electrical, chemical, nuclear) with well-defined transformations between them. Apart from that, the term "energy" is used loosely in colloquial speech.
- **Agent based** concepts of information emerge naturally when we extend our interest from simple measurement and symbol manipulation to the more complex paradigm of an agent with

knowledge, beliefs, intentions and freedom of choice. They are associated with the deployment of other concepts of information.

The emergence of a coherent theory to measure information quantitatively in the twentieth century is closely related to the development of the theory of computing. Central in this context are the notions of **Universality**, **Turing equivalence** and **Invariance:** because the concept of a Turing system defines the notion of a universal programmable computer, all universal models of computation seem to have the same power. This implies that all possible measures of information definable for universal models of computation (Recursive Functions, Turing Machine, Lambda Calculus etc.) are asymptotically *invariant*. This gives a perspective on a unified theory of information that might dominate the research program for the years to come.

1. Information in Colloquial Speech

The lack of preciseness and the universal usefulness of the term "information" go hand in hand. In our society, in which we explore reality by means of instruments and installations of ever increasing complexity (telescopes, cyclotrons) and communicate via more advanced media (newspapers, radio, television, SMS, the Internet), it is useful to have an abstract mass-noun for the "stuff" that is created by the instruments and that "flows" through these media. Historically this general meaning emerged rather late and seems to be associated with the rise of mass media and intelligence agencies (Devlin & Rosenberg 2008; Adriaans & van Benthem 2008b).

In present colloquial speech the term information is used in various loosely defined and often even conflicting ways. Most people, for instance, would consider the following inference *prima facie* to be valid:

> If I get the information that *p* then I know that *p*.

The same people would probably have no problems with the statement that "Secret services sometimes distribute false information", or with the sentence "The information provided by the witnesses of the accident was vague and conflicting". The first statement implies that information necessarily is true, while the other statements allow for the possibility that information is false, conflicting and vague. In everyday communication these inconsistencies do not seem to create great trouble and in general it is clear from the pragmatic context what type of information is designated. These examples suffice to argue that references to our intuitions as speakers of the English language are of little help in the development of a rigorous philosophical theory of information. There seems to be no

pragmatic pressure in everyday communication to converge to a more exact definition of the notion of information.

2. History of the Term and the Concept of Information

Until the second half of the twentieth century almost no modern philosopher considered "information" to be an important philosophical concept. The term has no lemma in the well-known encyclopedia of Edwards (1967) and is not mentioned in Windelband (1903). In this context the interest in "Philosophy of Information" is a recent development. Yet, with hindsight from the perspective of a history of ideas, reflection on the notion of "information" has been a predominant theme in the history of philosophy. The reconstruction of this history is relevant for the study of information.

A problem with any "history of ideas" approach is the validation of the underlying assumption that the concept one is studying has indeed continuity over the history of philosophy. In the case of the historical analysis of information one might ask whether the concept of "*informatio*" discussed by Augustine has any connection to Shannon information, other than a resemblance of the terms. At the same time one might ask whether Locke's "historical, plain method" is an important contribution to the emergence of the modern concept of information although in his writings Locke hardly uses the term "information" in a technical sense. As is shown below, there is a conglomerate of ideas involving a notion of information that has developed from antiquity till recent times, but further study of the history of the concept of information is necessary.

An important recurring theme in the early philosophical analysis of knowledge is the paradigm of manipulating a piece of wax: either by simply deforming it, by imprinting a signet ring in it or by writing characters on it. The fact that wax can take different shapes and secondary qualities (temperature, smell, touch) while the volume (extension) stays the same, make it a rich source of analogies, natural to Greek, Roman and medieval culture, where wax was used both for sculpture, writing (wax tablets) and encaustic painting. One finds this topic in writings of such diverse authors as Democritus, Plato, Aristotle, Theophrastus, Cicero, Augustine, Avicenna, Duns Scotus, Aquinas, Descartes and Locke.

2.1 Classical Philosophy

In classical philosophy "information" was a technical notion associated with a theory of knowledge and ontology that originated in Plato's (427–347 BCE) theory of forms, developed in a number of his dialogues (*Phaedo, Phaedrus, Symposium, Timaeus, Republic*). Various imperfect individual horses in the physical world could be identified as horses, because they participated in the static atemporal and aspatial idea of "horseness" in the world of ideas or forms. When later authors like Cicero (106–43 BCE) and Augustine (354–430 CE) discussed Platonic concepts in Latin they used the terms *informare* and *informatio* as a translation for technical Greek terms like *eidos* (essence), *idea* (idea), *typos* (type), *morphe* (form) and *prolepsis* (representation). The root "form" still is recognizable in the word *in-form-ation* (Capurro & Hjørland 2003). Plato's theory of forms was an attempt to formulate a solution for various philosophical problems: the theory of forms mediates between a static (Parmenides, ca. 450 BCE) and a dynamic (Herakleitos, ca. 535–475 BCE) ontological conception of reality and it offers a model to the study of the theory of human knowledge. According to Theophrastus (371–287 BCE) the analogy of the wax tablet goes back to Democritos (ca. 460–380/370 BCE) (*De Sensibus* 50). In the *Theaetetus* (191c,d) Plato compares the function of our memory with a wax tablet in which our perceptions and thoughts are imprinted like a signet ring stamps impressions in wax. Note that the metaphor of imprinting symbols in wax

is essentially spatial (extensive) and can not easily be reconciled with the aspatial interpretation of ideas supported by Plato.

One gets a picture of the role the notion of "form" plays in classical methodology if one considers Aristotle's (384–322 BCE) doctrine of the four causes. In Aristotelian methodology understanding an object implied understanding four different aspects of it:

Material Cause:: that as the result of whose presence something comes into being—e.g., the bronze of a statue and the silver of a cup, and the classes which contain these

Formal Cause:: the form or pattern; that is, the essential formula and the classes which contain it—e.g., the ratio 2:1 and number in general is the cause of the octave-and the parts of the formula.

Efficient Cause:: the source of the first beginning of change or rest; e.g., the man who plans is a cause, and the father is the cause of the child, and in general that which produces is the cause of that which is produced, and that which changes of that which is changed.

Final Cause:: the same as "end"; i.e., the final cause; e.g., as the "end" of walking is health. For why does a man walk? "To be healthy", we say, and by saying this we consider that we have supplied the cause. (Aristotle, *Metaphysics* 1013a)

Note that Aristotle, who rejects Plato's theory of forms as atemporal aspatial entities, still uses "form" as a technical concept. This passage states that knowing the form or structure of an object, i.e., the *information*, is a necessary condition for understanding it. In this sense information is a crucial aspect of classical epistemology.

The fact that the ratio 2:1 is cited as an example also illustrates the deep connection between the notion of forms and the idea that the world was governed by mathematical principles. Plato believed under influence of an older Pythagorean (Pythagoras 572–ca. 500 BCE) tradition that "everything that emerges and happens in the world" could be measured by means of numbers (*Politicus* 285a). On various occasions Aristotle mentions the fact that Plato associated ideas with numbers (Vogel 1968: 139***1974 a separate edition?*). Although formal mathematical theories about information only emerged in the twentieth century, and one has to be careful not to interpret the Greek notion of a number in any modern sense, the idea that information was essentially a mathematical notion, dates back to classical philosophy: the form of an entity was conceived as a structure or pattern that could be described in terms of numbers. Such a form had both an ontological and an epistemological aspect: it explains the essence as well as the understandability of the object. The concept of information thus from the very start of philosophical reflection was already associated with epistemology, ontology and mathematics.

Two fundamental problems that are not explained by the classical theory of ideas or forms are 1) the actual act of knowing an object (i.e., if I see a horse in what way is the idea of a horse activated in my mind) and 2) the process of thinking as manipulation of ideas. Aristotle treats these issues in *De Anime*, invoking the signet-ring-impression-in-wax analogy:

By a "sense" is meant what has the power of receiving into itself the sensible forms of things without the matter. This must be conceived of as taking place in the way in which a piece of wax takes on the impress of a signet-ring without the iron or gold; we say that what produces the impression is a signet of bronze or gold, but its particular metallic constitution makes no difference: in a similar way the sense is affected by what is coloured or flavoured or sounding, but it is indifferent what in each case the substance is;

what alone matters is what quality it has, i.e., in what ratio its constituents are combined. (*De Anime*, Book II, Chp. 12)

Have not we already disposed of the difficulty about interaction involving a common element, when we said that mind is in a sense potentially whatever is thinkable, though actually it is nothing until it has thought? What it thinks must be in it just as characters may be said to be on a writing-tablet on which as yet nothing actually stands written: this is exactly what happens with mind. (*De Anime*, Book III, Chp. 4)

These passages are rich in influential ideas and can with hindsight be read as programmatic for a philosophy of information: the process of *informatio* can be conceived as the imprint of characters on a wax tablet (*tabula rasa*), thinking can be analyzed in terms of manipulation of symbols.

2.2 Medieval Philosophy

Throughout the Middle Ages the reflection on the concept of *informatio* is taken up by successive thinkers. Illustrative for the Aristotelian influence is the passage of Augustine in *De Trinitate* book XI. Here he analyzes vision as an analogy for the understanding of the Trinity. There are three aspects: the corporeal form in the outside world, the *informatio* by the sense of vision, and the resulting form in the mind. For this process of information Augustine uses the image of a signet ring making an impression in wax (*De Trinitate*, XI Cap 2 par 3). Capurro (2009) observes that this analysis can be interpreted as an early version of the technical concept of "sending a message" in modern information theory, but the idea is older and is a common topic in Greek thought (Plato *Theaetetus* 191c,d; Aristotle *De Anime*, Book II, Chp. 12, Book III, Chp. 4; Theophrastus *De Sensibus* 50).

The *tabula rasa* notion was later further developed in the theory of knowledge of Avicenna (c. 980–1037 CE):

The human intellect at birth is rather like a *tabula rasa*, a pure potentiality that is actualized through education and comes to know. Knowledge is attained through empirical familiarity with objects in this world from which one abstracts universal concepts. (Sajjad 2006 [Other Internet Resources [hereafter OIR]])

The idea of a *tabula rasa* development of the human mind was the topic of a novel Hayy ibn Yaqdhan by the Arabic Andalusian philosopher Ibn Tufail (1105–1185 CE, known as "Abubacer" or "Ebn Tophail" in the West). This novel describes the development of an isolated child on a deserted island. A later translation in Latin under the title *Philosophus Autodidactus* (1761) influenced the empiricist John Locke in the formulation of his *tabula rasa* doctrine.

Apart from the permanent creative tension between theology and philosophy, medieval thought, after the rediscovery of Aristotle's *Metaphysics* in the twelfth century inspired by Arabic scholars, can be characterized as an elaborate and subtle interpretation and development of, mainly Aristotelian, classical theory. Reflection on the notion of *informatio* is taken up, under influence of Avicenna, by thinkers like Aquinas (1225–1274 CE) and Duns Scotus (1265/66–1308 CE). When Aquinas discusses the question whether angels can interact with matter he refers to the Aristotelian doctrine of hylomorphism (i.e., the theory that substance consists of matter (*hylo* (wood), matter) and form (*morphè*)). Here Aquinas translates this as the in-formation of matter (*informatio materiae*) (*Summa Theologiae*, 1a 110 2; Capurro 2009). Duns Scotus refers to *informatio* in the technical sense when he discusses Augustine's theory of vision in *De Trinitate*, XI Cap 2 par 3 (Duns Scotus, 1639, "De imagine", *Ordinatio*, I, d.3, p.3).

The tension that already existed in classical philosophy between Platonic idealism(*universalia ante res*) and Aristotelian realism (*universalia in rebus*) is recaptured as the problem of universals: do universal qualities like "humanity" or the idea of a horse exist apart from the individual entities that instantiate them? It is in the context of his rejection of universals that Ockham (c. 1287–1347 CE) introduces his well-known razor: entities should not be multiplied beyond necessity. Throughout their writings Aquinas and Scotus use the Latin terms *informatio* and *informare* in a technical sense, although this terminology is not used by Ockham.

2.3 Modern Philosophy

The history of the concept of information in modern philosophy is complicated. Probably starting in the fourteenth century the term "information" emerged in various developing European languages in the general meaning of "education" and "inquiry". The French historical dictionary by Godefroy (1881) gives *action de former, instruction, enquête, science, talent* as early meanings of "information". The term was also used explicitly for legal inquiries (*Dictionnaire du Moyen Français (1330–1500)* 2015). Because of this colloquial use the term "information" loses its association with the concept of "form" gradually and appears less and less in a formal sense in philosophical texts.

At the end of the Middle Ages society and science are changing fundamentally (Hazard 1935; Ong 1958; Dijksterhuis 1986). In a long complex process the Aristotelian methodology of the four causes was transformed to serve the needs of experimental science:

1. The Material Cause developed in to the modern notion of matter.
2. The Formal Cause was reinterpreted as geometric form in space.
3. The Efficient Cause was redefined as direct mechanical interaction between material bodies.

4. The Final Cause was dismissed as unscientific. Because of this, Newton's contemporaries had difficulty with the concept of the force of gravity in his theory. Gravity as action at a distance seemed to be a reintroduction of final causes.

In this changing context the analogy of the wax-impression is reinterpreted. A proto-version of the modern concept of information as the structure of a set or sequence of simple ideas is developed by the empiricists, but since the technical meaning of the term "information" is lost, this theory of knowledge is never identified as a new "theory of information".

The consequence of this shift in methodology is that only phenomena that can be explained in terms of mechanical interaction between material bodies can be studied scientifically. This implies in a modern sense: the reduction of intensive properties to measurable extensive properties. For Galileo this insight is programmatic:

> To excite in us tastes, odors, and sounds I believe that nothing is required in external bodies except shapes, numbers, and slow or rapid movements. (Galileo 1623 [1960: ***page*])

These insights later led to the doctrine of the difference between primary qualities (space, shape, velocity) and secondary qualities (heat, taste, color etc.). In the context of philosophy of information Galileo's observations on the secondary quality of "heat" is of particular importance since they lay the foundations for the study of thermodynamics in the nineteenth century:

> Having shown that many sensations which are supposed to be qualities residing in external objects have no real existence save in us, and outside ourselves are mere names, I now say that I am inclined to believe heat to be of this character. Those materials which produce heat in us and make us feel warmth, which are

known by the general name of "fire," would then be a multitude of minute particles having certain shapes and moving with certain velocities. (Galileo 1623 [1960: ***page*])

A pivotal thinker in this transformation is René Descartes (1596–1650 CE). In his *Meditationes*, after "proving" that the matter (*res extensa*) and mind (*res cogitans*) are different substances (i.e., forms of being existing independently), the question of the interaction between these substances becomes an issue. The malleability of wax is for Descartes an explicit argument against influence of the *res extensa* on the *res cogitans* (*Meditationes* II, 15). The fact that a piece of wax loses its form and other qualities easily when heated, implies that the senses are not adequate for the identification of objects in the world. True knowledge thus can only be reached via "inspection of the mind". Here the wax metaphor that for more than 1500 years was used to *explain* sensory impression is used to argue *against* the possibility to reach knowledge via the senses. Since the essence of the *res extensa* is extension, thinking fundamentally can not be understood as a spatial process. Descartes still uses the terms "form" and "idea" in the original scholastic non-geometric (atemporal, aspatial) sense. An example is the short formal proof of God's existence in the second answer to Mersenne in the *Meditationes de Prima Philosophia*

> I use the term idea to refer to the *form* of any given thought, immediate perception of which makes me aware of the thought.
> (*Idea nomine intelligo cujuslibet cogitationis formam illam, per cujus immediatam perceptionem ipsius ejusdem cogitationis conscious sum*)

I call them "ideas" says Descartes

> only in so far as they make a difference to the mind itself when they *inform* that part of the brain.

(*sed tantum quatenus mentem ipsam in illam cerebri partem conversam* **informant**). (Descartes, 1641, *Ad Secundas Objections, Rationes, Dei existentiam & anime distinctionem probantes, more Geometrico dispositae*.)

Because the *res extensa* and the *res cogitans* are different substances, the act of thinking can never be emulated in space: machines can not have the universal faculty of reason. Descartes gives two separate motivations:

> Of these the first is that they could never use words or other signs arranged in such a manner as is competent to us in order to declare our thoughts to others: (…) The second test is, that although such machines might execute many things with equal or perhaps greater perfection than any of us, they would, without doubt, fail in certain others from which it could be discovered that they did not act from knowledge, but solely from the disposition of their organs: for while reason is an universal instrument that is alike available on every occasion, these organs, on the contrary, need a particular arrangement for each particular action; whence it must be morally impossible that there should exist in any machine a diversity of organs sufficient to enable it to act in all the occurrences of life, in the way in which our reason enables us to act. (*Discourse de la méthode*, 1647)

The passage is relevant since it directly argues against the possibility of artificial intelligence and it even might be interpreted as arguing against the possibility of a universal Turing machine: reason as a universal instrument can never be emulated in space. This conception is in opposition to the modern concept of information which as a measurable quantity is essentially spatial, i.e., extensive (but in a sense different from that of Descartes).

Descartes does not present a new interpretation of the notions of form and idea, but he sets the stage for a debate about the nature of ideas that evolves around two opposite positions:

Rationalism: The Cartesian notion that ideas are innate and thus *a priori*. This form of rationalism implies an interpretation of the notion of ideas and forms as atemporal, aspatial, but complex structures i.e., the idea of "a horse" (i.e., with a head, body and legs). It also matches well with the interpretation of the knowing subject as a created being (*ens creatu*). God created man after his own image and thus provided the human mind with an adequate set of ideas to understand his creation. In this theory growth, of knowledge is *a priori* limited. Creation of new ideas *ex nihilo* is impossible. This view is difficult to reconcile with the concept of experimental science.

Empiricism: Concepts are constructed in the mind *a posteriori* on the basis of ideas associated with sensory impressions. This doctrine implies a new interpretation of the concept of idea as:

> whatsoever is the object of understanding when a man thinks … whatever is meant by phantasm, notion, species, or whatever it is which the mind can be employed about when thinking. (Locke 1689, bk I, ch 1, para 8)

Here ideas are conceived as elementary building blocks of human knowledge and reflection. This fits well with the demands of experimental science. The downside is that the mind can never formulate apodeictic truths about cause and effects and the essence of observed entities, including its own identity. Human knowledge becomes essentially probabilistic (Locke 1689: bk I, ch. 4, para 25).

Locke's reinterpretation of the notion of idea as a "structural placeholder" for any entity present in the mind is an essential step in the emergence of the modern concept of information. Since these ideas are not involved in the justification of apodeictic knowledge, the necessity to stress the atemporal and aspatial nature of ideas vanishes. The construction of concepts on the basis of a *collection of elementary ideas* based in sensorial experience opens the gate to a reconstruction of *knowledge as an extensive property of an agent*: more ideas implies more probable knowledge.

In the second half of the seventeenth century formal theory of probability is developed by researchers like Pascal (1623–1662), Fermat (1601 or 1606–1665) and Christiaan Huygens (1629–1695). The work *De ratiociniis in ludo aleae* of Huygens was translated in to English by John Arbuthnot (1692). For these authors, the world was essentially mechanistic and thus deterministic, probability was a quality of human knowledge caused by its imperfection:

> It is impossible for a Die, with such determin'd force and direction, not to fall on such determin'd side, only I don't know the force and direction which makes it fall on such determin'd side, and therefore I call it Chance, wich is nothing but the want of art;… (John Arbuthnot *Of the Laws of Chance* (1692), preface)

This text probably influenced Hume, who was the first to marry formal probability theory with theory of knowledge:

> Though there be no such thing as Chance in the world; our ignorance of the real cause of any event has the same influence on the understanding, and begets a like species of belief or opinion. (…) If a dye were marked with one figure or number of spots on four sides, and with another figure or number of spots on the two remaining sides, it would be more probable, that the former would turn up than the latter; though, if it had a thousand sides marked in the same manner, and only one side different, the probability

would be much higher, and our belief or expectation of the event more steady and secure. This process of the thought or reasoning may seem trivial and obvious; but to those who consider it more narrowly, it may, perhaps, afford matter for curious speculation. (Hume 1748: Section VI, "On probability" 1)

Here knowledge about the future as a degree of belief is measured in terms of probability, which in its turn is explained in terms of the number of configurations a deterministic system in the world can have. The basic building blocks of a modern theory of information are in place. With this new concept of knowledge empiricists laid the foundation for the later development of thermodynamics as a reduction of the secondary quality of heat to the primary qualities of bodies.

At the same time the term "information" seems to have lost much of its technical meaning in the writings of the empiricists so this new development is not designated as a new interpretation of the notion of "information". Locke sometimes uses the phrase that our senses "inform" us about the world and occasionally uses the word "information".

> For what information, what knowledge, carries this proposition in it, viz. "Lead is a metal" to a man who knows the complex idea the name lead stands for? (Locke 1689: bk IV, ch 8, para 4)

Hume seems to use information in the same casual way when he observes:

> Two objects, though perfectly resembling each other, and even appearing in the same place at different times, may be numerically different: And as the power, by which one object produces another, is never discoverable merely from their idea, it is evident cause and effect are relations, of which we receive information from experience, and not from any abstract reasoning or reflection. (Hume 1739: Part III, section 1)

The empiricists methodology is not without problems. The biggest issue is that all knowledge becomes probabilistic and *a posteriori*. Immanuel Kant (1724–1804) was one of the first to point out that the human mind has a grasp of the meta-concepts of space, time and causality that itself can never be understood as the result of a mere combination of "ideas". What is more, these intuitions allow us to formulate scientific insights with certainty: i.e., the fact that the sum of the angles of a triangle in Euclidean space is 180 degrees. This issue cannot be explained in the empirical framework. If knowledge is created by means of combination of ideas then there must exist an *a priori* synthesis of ideas in the human mind. According to Kant, this implies that the human mind can evaluate its own capability to formulate scientific judgments. In his *Kritik der reinen Vernunft* (1781) Kant developed transcendental philosophy as an investigation of the necessary conditions of human knowledge. Although Kant's transcendental program did not contribute directly to the development of the concept of information, he did influence research in to the foundations of mathematics and knowledge relevant for this subject in the nineteenth and twentieth century: e.g., the work of Frege, Husserl, Russell, Brouwer, L. Wittgenstein, Gödel, Carnap, Popper and Quine.

2.4 Historical Development of the Meaning of the Term "Information"

The history of the term "information" is intricately related to the study of central problems in epistemology and ontology in Western philosophy. After a start as a technical term in classical and medieval texts the term "information" almost vanished from the philosophical discourse in modern philosophy, but gained popularity in colloquial speech. Gradually the term obtained the status of an abstract mass-noun, a meaning that is orthogonal to the classical process-oriented meaning. In this form it was picked up by several researchers (Fisher 1925; Shannon 1948) in the twentieth century who introduced formal methods to measure "information". This, in its

turn, lead to a revival of the philosophical interest in the concept of information. This complex history seems to be one of the main reasons for the difficulties in formulating a definition of a unified concept of information that satisfies all our intuitions. At least three different meanings of the word "information" are historically relevant:

"Information" as the process of being informed.
This is the oldest meaning one finds in the writings of authors like Cicero (106–43 BCE) and Augustine (354–430 CE) and it is lost in the modern discourse, although the association of information with processes (i.e., computing, flowing or sending a message) still exists. In classical philosophy one could say that when I recognize a horse as such, then the "form" of a horse is planted in my mind. This process is my "information" of the nature of the horse. Also the act of teaching could be referred to as the "information" of a pupil. In the same sense one could say that a sculptor creates a sculpture by "informing" a piece of marble. The task of the sculptor is the "information" of the statue (Capurro & Hjørland 2003). This process-oriented meaning survived quite long in western European discourse: even in the eighteenth century Robinson Crusoe could refer to the education of his servant Friday as his "information" (Defoe 1719: 261). It is also used in this sense by Berkeley: "I love information upon all subjects that come in my way, and especially upon those that are most important" (*Alciphron* Dialogue 1, Section 5, Paragraph 6/10, see Berkeley 1732).

"Information" as a state of an agent,
i.e., as the result of the process of being informed. If one teaches a pupil the theorem of Pythagoras then, after this process is completed, the student can be said to "have the information about the theorem of Pythagoras". In this sense the term "information" is the result of the same suspect form of substantiation of a verb (*informare* > *informatio*) as many other technical terms in philosophy (substance,

consciousness, subject, object). This sort of term-formation is notorious for the conceptual difficulties it generates. Can one derive the fact that I "have" consciousness from the fact that I am conscious? Can one derive the fact that I "have" information from the fact that I have been informed? The transformation to this modern substantiated meaning seems to have been gradual and seems to have been general in Western Europe at least from the middle of the fifteenth century. In the renaissance a scholar could be referred to as "a man of information", much in the same way as we now could say that someone received an education (Adriaans & van Benthem 2008b; Capurro & Hjørland 2003). In "Emma" by Jane Austen one can read: "Mr. Martin, I suppose, is not a man of information beyond the line of his own business. He does not read" (Austen 1815: 21).

"Information" as the disposition to inform,
i.e., as a capacity of an object to inform an agent. When the act of teaching me Pythagoras' theorem leaves me with information about this theorem, it is only natural to assume that a text in which the theorem is explained actually "contains" this information. The text has the capacity to inform me when I read it. In the same sense, when I have received information from a teacher, I am capable of transmitting this information to another student. Thus information becomes something that can be stored and measured. This last concept of information as an abstract mass-noun has gathered wide acceptance in modern society and has found its definitive form in the nineteenth century, allowing Sherlock Homes to make the following observation: "… friend Lestrade held information in his hands the value of which he did not himself know" ("The Adventure of the Noble Bachelor", Conan Doyle 1892). The association with the technical philosophical notions like "form" and "informing" has vanished from the general consciousness although the association between information and processes like storing, gathering, computing and teaching still exist.

3. Building Blocks of Modern Theories of Information

With hindsight many notions that have to do with optimal code systems, ideal languages and the association between computing and processing language have been recurrent themes in the philosophical reflection since the seventeenth century.

3.1 Languages

One of the most elaborate proposals for a universal "philosophical" language was made by bishop John Wilkins: "An Essay towards a Real Character, and a Philosophical Language" (1668). Wilkins' project consisted of an elaborate system of symbols that supposedly were associated with unambiguous concepts in reality. Proposals such as these made philosophers sensitive to the deep connections between language and thought. The empiricist methodology made it possible to conceive the development of language as a system of conventional signs in terms of associations between ideas in the human mind. The issue that currently is known as the *symbol grounding problem* (how do arbitrary signs acquire their inter-subjective meaning) was one of the most heavily debated questions in the eighteenth century in the context of the problem of the origin of languages. Diverse thinkers as Vico, Condillac, Rousseau, Diderot, Herder and Haman made contributions. The central question was whether language was given *a priori* (by God) or whether it was constructed and hence an invention of man himself. Typical was the contest issued by the Royal Prussian Academy of Sciences in 1769:

> *En supposant les hommes abandonnés à leurs facultés naturelles, sont-ils en état d'inventer le langage? Et par quels moyens parviendront-ils d'eux-mêmes à cette invention?*

Assuming men abandoned to their natural faculties, are they able to invent language and by what means will they come to this invention?[1]

The controversy raged on for over a century without any conclusion and in 1866 the Linguistic Society of Paris (*Société de Linguistique de Paris*) banished the issue from its arena. [2]

Philosophically more relevant is the work of Leibniz (1646–1716) on a so-called *characteristica universalis*: the notion of a universal logical calculus that would be the perfect vehicle for scientific reasoning. A central presupposition in Leibniz' philosophy is that such a perfect language of science is in principle possible because of the perfect nature of the world as God's creation (*ratio essendi = ration cognoscendi*, the origin of being is the origin of knowing). This principle was rejected by Wolff (1679–1754) who suggested more heuristically oriented *characteristica combinatoria* (van Peursen 1987). These ideas had to wait for thinkers like Boole (1854, *An Investigation of the Laws of Thought*), Frege (1879, *Begriffsschrift*), Peirce (who in 1886 already suggested that electrical circuits could be used to process logical operations) and Whitehead and Russell (1910–1913, *Principia Mathematica*) to find a more fruitful treatment.

3.2 Optimal Codes

The fact that frequencies of letters vary in a language was known since the invention of book printing. Printers needed many more "e"s and "t"s than "x"s or "q"s to typeset an English text. This knowledge was used extensively to decode ciphers since the seventeenth century (Kahn 1967; Singh 1999). In 1844 an assistant of Samuel Morse, Alfred Vail, determined the frequency of letters used in a local newspaper in Morristown, New Jersey, and used them to optimize Morse code. Thus the

core of theory of optimal codes was already established long before Shannon developed its mathematical foundation (Shannon 1948; Shannon & Weaver 1949). Historically important but philosophically less relevant are the efforts of Charles Babbage to construct computing machines (Difference Engine in 1821, and the Analytical Engine 1834–1871) and the attempt of Ada Lovelace (1815–1852) to design what is considered to be the first programming language for the Analytical Engine.

3.3 Numbers

The simplest way of representing numbers is via a *unary system*. Here the length of the representation of a number is equal to the size of the number itself, i.e., the number "ten" is represented as "\\\\\\\\\\". The classical Roman number system is an improvement since it contains different symbols for different orders of magnitude (one = I, ten = X, hundred = C, thousand = M). This system has enormous drawbacks since in principle one needs an infinite amount of symbols to code the natural numbers and because of this the same mathematical operations (adding, multiplication etc.) take different forms at different orders of magnitude. Around 500 CE the number zero was invented in India. Using zero as a placeholder we can code an infinity of numbers with a finite set of symbols (one = I, ten = 10, hundred = 100, thousand = 1000 etc.). From a modern perspective an infinite number of position systems is possible as long as we have 0 as a placeholder and a finite number of other symbols. Our normal decimal number system has ten digits "0, 1, 2, 3, 4, 5, 6, 7, 8, 9" and represents the number two-hundred-and-fifty-five as "255". In a binary number system we only have the symbols "0" and "1". Here two-hundred-and-fifty-five is represented as "11111111". In a hexadecimal system with 16 symbols (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, a, b, c, d, e, f) the same number can be written as "ff". Note that the length of these representations differs considerable. Using this representation, mathematical operations can be standardized irrespective of the order of magnitude of numbers we are dealing with, i.e.,

the possibility of a uniform algorithmic treatment of mathematical functions (addition, subtraction, multiplication and division etc.) is associated with such a position system.

The concept of a positional number system was brought to Europe by the Persian mathematician al-Khwarizmi (ca. 780–ca. 850 CE). His main work on numbers (ca. 820 CE) was translated into Latin as *Liber Algebrae et Almucabola* in the twelfth century, which gave us amongst other things the term "algebra". Our word "algorithm" is derived from *Algoritmi*, the Latin form of his name. Positional number systems simplified commercial and scientific calculations.

In 1544 Michael Stifel introduced the concept of the exponent of a number in *Arithmetica integra* (1544). Thus 8 can be written as 2^3 and 25 as 5^2. The notion of an exponent immediately suggests the notion of a logarithm as its inverse function: $\log_b b^a) = a$. Stifel compared the arithmetic sequence:

$$-3, -2, -1, 0, 1, 2, 3$$

in which the term 1 have a difference of 1 with the geometric sequence:

$$\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8$$

in which the terms have a ratio of 2. The exponent notation allowed him to rewrite the values of the second table as:

$$2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3$$

which combines the two tables. This arguably was the first logarithmic table. A more definitive and practical theory of logarithms is developed by John Napier (1550–1617) in his main work (Napier 1614). He coined the term logarithm (logos + arithmetic: ratio of numbers). As is clear from the

match between arithmetic and geometric progressions, logarithms reduce products to sums:

$$\log_b(xy) = \log_b(x) + \log_b(y)$$

They also reduce divisions to differences:

$$\log_b(x/y) = \log_b(x) - \log_b(y)$$

and powers to products:

$$\log_b(x^p) = p \log_b(x)$$

After publication of the logarithmic tables by Briggs (1624) this new technique of facilitating complex calculations rapidly gained popularity.

3.4 Physics

Galileo (1623) already had suggested that the analysis of phenomena like heat and pressure could be reduced to the study of movements of elementary particles. Within the empirical methodology this could be conceived as the question how the sensory experience of the secondary quality of heat of an object or a gas could be reduced to movements of particles. Bernoulli (*Hydrodynamica* published in 1738) was the first to develop a kinetic theory of gases in which macroscopically observable phenomena are described in terms of microstates of systems of particles that obey the laws of Newtonian mechanics, but it was quite an intellectual effort to come up with an adequate mathematical treatment. Clausius (1850) made a conclusive step when he introduced the notion of the mean free path of a particle between two collisions. This opened the way for a statistical treatment by Maxwell who formulated his distribution in 1857, which was the first statistical law in physics. The definitive formula that

tied all notions together (and that is engraved on his tombstone, though the actual formula is due to Planck) was developed by Boltzmann:

$$S = k \log W$$

It describes the entropy S of a system in terms of the logarithm of the number of possible microstates W, consistent with the observable macroscopic states of the system, where k is the well-known Boltzmann constant. In all its simplicity the value of this formula for modern science can hardly be overestimated. The expression "$\log W$" can, from the perspective of information theory, be interpreted in various ways:

- As the amount of *entropy* in the system.
- As the length of the *number* needed to count all possible microstates consistent with macroscopic observations.
- As the length of an optimal *index* we need to identify the specific current unknown microstate of the system, i.e., it is a measure of our "lack of information".
- As a measure for the *probability* of any typical specific microstate of the system consistent with macroscopic observations.

Thus it connects the additive nature of logarithm with the extensive qualities of entropy, probability, typicality and information and it is a fundamental step in the use of mathematics to analyze nature. Later Gibbs (1906) refined the formula:

$$S = - \sum_i p_i \ln p_i,$$

where p_i is the probability that the system is in the ith microstate. This formula was adopted by Shannon (1948; Shannon & Weaver 1949) to characterize the communication entropy of a system of messages. Although there is a close connection between the mathematical treatment

of entropy and information, the exact interpretation of this fact has been a source of controversy ever since (Harremoës & Topsøe 2008; Bais & Farmer 2008).

4. Developments in Philosophy of Information

The modern theories of information emerged in the middle of the twentieth century in a specific intellectual climate in which the distance between the sciences and parts of academic philosophy was quite big. Some philosophers displayed a specific anti-scientific attitude: Heidegger, "*Die Wissenschaft denkt nicht*." On the other hand the philosophers from the Wiener Kreis overtly discredited traditional philosophy as dealing with illusionary problems (Carnap 1928). The research program of logical positivism was a rigorous reconstruction of philosophy based on a combination of empiricism and the recent advances in logic. It is perhaps because of this intellectual climate that early important developments in the theory of information took place in isolation from mainstream philosophical reflection. A landmark is the work of Dretske in the early eighties (Dretske 1981). Since the turn of the century, interest in Philosophy of Information has grown considerably, largely under the influence of the work of Luciano Floridi on semantic information. Also the rapid theoretical development of quantum computing and the associated notion of quantum information have had it repercussions on philosophical reflection.

4.1 Popper: Information as Degree of Falsifiability

The research program of logical positivism of the Wiener Kreis in the first half of the twentieth century revitalized the older project of empiricism. Its ambition was to reconstruct scientific knowledge on the basis of direct observations and logical relation between statements about those observations. The old criticism of Kant on empiricism was revitalized by

Quine (1951***which one*). Within the framework of logical positivism induction was invalid and causation could never be established objectively. In his *Logik der Forschung* (1934) Popper formulates his well-known demarcation criterion and he positions this explicitly as a solution to Hume's problem of induction (Popper 1934 [1977: 42]). Scientific theories formulated as general laws can never be verified definitively, but they can be falsified by only one observation. This implies that a theory is "more" scientific if it is richer and provides more opportunity to be falsified:

> Thus it can be said that the amount of empirical information conveyed by a theory, or its *empirical content*, increases with its degree of falsifiability. (Popper 1934 [1977: 113], emphasis in original)

This quote, in the context of Popper's research program, shows that the ambition to measure the *amount of empirical information in scientific theory conceived as a set of logical statements* was already recognized as a philosophical problem more than a decade before Shannon formulated his theory of information. Popper is aware of the fact that the empirical content of a theory is related to its falsifiability and that this in its turn has a relation with the probability of the statements in the theory. Theories with more empirical information are less probable. Popper distinguishes *logical probability* from *numerical probability* ("which is employed in the theory of games and chance, and in statistics"; Popper 1934 [1977: 119]). In a passage that is programmatic for the later development of the concept of information he defines the notion of logical probability:

> *The logical probability of a statement is complementary to its falsifiability:* it increases with decreasing degree of falsifiability. The logical probability 1 corresponds to the degree 0 of

falsifiability and *vice versa*. (Popper 1934 [1977: 119], emphasis in original)

It is possible to interpret numerical probability as applying to a subsequence (picked out from the logical probability relation) for which a *system of measurement* can be defined, on the basis of frequency estimates. (Popper 1934 [1977: 119], emphasis in original)

Popper never succeeded in formulating a good formal theory to measure this amount of information although in later writings he suggests that Shannon's theory of information might be useful (Popper 1934 [1977], 404 [Appendix IX, from 1954]). These issues were later developed in philosophy of science. Theory of conformation studies induction theory and the way in which evidence "supports" a certain theory (Huber 2007 [OIR]). Although the work of Carnap motivated important developments in both philosophy of science and philosophy of information the connection between the two disciplines seems to have been lost. There is no mention of information theory or any of the more foundational work in philosophy of information in Kuipers (2007a), but the two disciplines certainly have overlapping domains. (See, e.g., the discussion of the so-called Black Ravens Paradox by Kuipers (2007b) and Rathmanner & Hutter (2011).)

4.2 Shannon: Information Defined in Terms of Probability

In two landmark papers Shannon (1948; Shannon & Weaver 1949) characterized the communication entropy of a system of messages *A*:

$$H(P) = -\sum_{i \in A} p_i \log_2 p_i$$

Here p_i is the probability of message *i* in *A*. This is exactly the formula for Gibb's entropy in physics. The use of base-2 logarithms ensures that the code length is measured in bits (binary digits). It is easily seen that the communication entropy of a system is maximal when all the messages have equal probability and thus are typical.

The amount of information *I* in an individual message *x* is given by:

$$I(x) = -\log p_x$$

This formula, that can be interpreted as the inverse of the Boltzmann entropy, covers a number of our basic intuitions about information:

- A message *x* has a certain probability p_x between 0 and 1 of occurring.
- If $p_x = 1$ then $I(x) = 0$. If we are certain to get a message it literally contains no "news" at al. The lower the probability of the message is, the more information it contains. A message like "The sun will rise tomorrow" seems to contain less information than the message "Jesus was Caesar" exactly because the second statement is much less likely to be defended by anyone (although it can be found on the web).
- If two messages *x* and *y* are unrelated then $I(x \text{ and } y) = I(x) + I(y)$. Information is *extensive*. The amount of information in two combined messages is equal to the sum of the amount of information in the individual messages.

Information as the negative log of the probability is the only mathematical function that exactly fulfills these constraints (Cover & Thomas 2006). Shannon offers a theoretical framework in which binary strings can be interpreted as words in a (programming) language containing a certain amount of information (see 3.1 Languages). The expression $-\log p_x$ exactly gives the length of an optimal code for message *x* and as such formalizes the old intuition that codes are more efficient when frequent

letters get shorter representations (see 3.2 Optimal codes). Logarithms as a reduction of multiplication to addition (see 3.3 Numbers) are a natural representation of extensive properties of systems and already as such had been used by physicists in the nineteenth century (see 3.4 Physics).

One aspect of information that Shannon's definition explicitly does not cover is the actual content of the messages interpreted as propositions. So the statement "Jesus was Caesar" and "The moon is made of green cheese" may carry the same amount of information while their meaning is totally different. A large part of the effort in philosophy of information has been directed to the formulation of more semantic theories of information (Bar-Hillel & Carnap 1953; Floridi 2002, 2003, 2011). Although Shannon's proposals at first were almost completely ignored by philosophers it has in the past decennia become apparent that their impact on philosophical issues is big. Dretske (1981) was one of the first to analyze the philosophical implications of Shannon's theory, but the exact relation between various systems of logic and theory of information are still unclear (see 3.5 Logic***not present*).

4.3 Solomonoff, Kolmogorov, Chaitin: Information as the Length of a Program

This problem of relating a set of statements to a set of observations and defining the corresponding probability was taken up by Carnap (1945, 1950). He distinguished two forms of probability: Probability$_1$ or "degree of confirmation" $P_1(h;e)$ is a *logical* relation between two sentences, a hypothesis h and a sentence e reporting a series of observations. Statements of this type are either analytical or contradictory. The second form, Probability$_2$ or "relative frequency", is the statistical concept. In the words of his student Solomonoff (1997):

Carnap's model of probability started with a long sequence of symbols that was a description of the entire universe. Through his own formal linguistic analysis, he was able to assign *a priori* probabilities to any possible string of symbols that might represent the universe.

The method for assigning probabilities Carnap used, was not universal and depended heavily on the code systems used. A general theory of induction using Bayes' rule can only be developed when we can assign a universal probability to "any possible string" of symbols. In a paper in 1960 Solomonoff (1960, 1964a,b) was the first to sketch an outline of a solution for this problem. He formulated the notion of a *universal distribution*:

consider the set of all possible finite strings to be programs for a universal Turing machine U and define the probability of a string x of symbols in terms of the length of the shortest program p that outputs x on U.***is this a quote?*

This notion of Algorithmic Information Theory was invented independently somewhat later separately by Kolmogorov (1965) and Chaitin (1969). Levin (1974) developed a mathematical expression of the universal *a priori* probability as a universal (that is, maximal) lower semicomputable semimeasure M, and showed that the negative logarithm of $M(x)$ coincides with the Kolmogorov complexity of x up to an additive logarithmic term. The actual definition of the complexity measure is:

Kolmogorov complexity The algorithmic complexity of a string x is the length $l(p)$ of the smallest program p that produces x when it runs on a universal Turing machine U, noted as $U(p) = x$:

$$K(x) := \min_{p}\{l(p), U(p) = x\}$$

Algorithmic Information Theory (a.k.a. Kolmogorov complexity theory) has developed into a rich field of research with a wide range of domains of applications many of which are philosophically relevant (Li & Vitányi 1997):

- It provides us with a general theory of induction. The use of Bayes' rule allows for a modern reformulation of Ockham's razor in terms of Minimum Description Length (Rissanen 1978, 1989; Barron, Rissanen, & Yu 1998; Grünwald 2007) and minimum message length (Wallace 2005). Note that Domingos (1998) has argued against the general validity of these principles.
- It allows us to formulate probabilities and information content for individual objects. Even individual natural numbers.
- It lays the foundation for a theory of learning as data compression (Adriaans 2007).
- It gives a definition of randomness of a string in terms of incompressibility. This in itself has led to a whole new domain of research (Niess 2009; Downey & Hirschfeld 2010).
- It allows us to formulate an objective *a priori* measure of the predictive value of a theory in terms of its randomness deficiency: i.e., the best theory is the shortest theory that makes the data look random conditional to the theory. (Vereshchagin & Vitányi 2004).

There are also down-sides:

- Algorithmic complexity is uncomputable, although it can in a lot of practical cases be approximated and commercial compression programs in some cases come close to the theoretical optimum (Cilibrasi & Vitányi 2005).
- Algorithmic complexity is an asymptotic measure (i.e., it gives a value that is correct up to a constant). In some cases the value of this constant is prohibitive for use in practical purposes.

- Although the shortest theory is always the best one in terms of randomness deficiency, incremental compression of data-sets is in general not a good learning strategy since the randomness deficiency does not decrease monotonically with the compression rate (Adriaans & Vitányi 2009).
- The generality of the definitions provided by Algorithmic Information Theory depends on the generality of the concept of a universal Turing machine and thus ultimately on the interpretation of the Church-Turing-Thesis.
- The Kolmogorov complexity of an object does not take in to account the amount of time it takes to actually compute the object. In this context Levin proposed a variant of Kolmogorov complexity that penalizes the computation time (Levin 1973, 1984):

  Levin complexity The Levin complexity of a string x is the sum of the length $l(p)$ and the logarithm of the computation time of the smallest program p that produces x when it runs on a universal Turing machine U, noted as $U(p) = x$:

  $$Kt(x) := \min_p \{l(p) + \log(time(p)), U(p) = x\}$$

Algorithmic Information Theory has gained rapid acceptance as a fundamental theory of information. The well-known introduction in *Information Theory* by Cover and Thomas (2006) states: "… we consider Kolmogorov complexity (i.e., AIT) to be more fundamental than Shannon entropy" (2006: 3).

The idea that algorithmic complexity theory is a foundation for a general theory of artificial intelligence (and theory of knowledge) has already been suggested by Solomonoff (1997) and Chaitin (1987). Several authors have defended that data compression is a general principle that governs human cognition (Chater & Vitányi 2003; Wolff 2004 [OIR]). Hutter (2005,

2007a,b) argues that Solomonoff's formal and complete theory essentially solves the induction problem. Hutter (2007a) and Rathmanner & Hutter (2011) enumerate a plethora of classical philosophical and statistical problems around induction and claim that Solomonoff's theory solves or avoids all these problems. Probably because of its technical nature, the theory has been largely ignored by the philosophical community. Yet, it stands out as one of the most fundamental contributions to information theory in the twentieth century and it is clearly relevant for a number of philosophical issues, such as the problem of induction.

5. Systematic Considerations

In a mathematical sense information is associated with measuring extensive properties of classes of systems with finite but unlimited dimensions (systems of particles, texts, codes, networks, graphs, games etc.). This suggests that a uniform treatment of various theories of information is possible. In the Handbook of Philosophy of Information three different forms of information are distinguished (Adriaans & van Benthem 2008b):

Information-A:
Knowledge, logic, what is conveyed in informative answers

Information-B:
Probabilistic, information-theoretic, measured quantitatively

Information-C:
Algorithmic, code compression, measured quantitatively

Because of recent development the connections between Information-B (Shannon) and Information-C (Kolmogorov) are reasonably well understood (Cover & Thomas 2006). The historical material presented in this article suggests that reflection on Information-A (logic, knowledge) is

historically much more interwoven than was generally known up till now. The research program of logical positivism can with hindsight be characterized as the attempt to marry a possible worlds interpretation of logic with probabilistic reasoning (Carnap 1945, 1950; Popper 1934; for a recent approach see Hutter et al. 2013). Modern attempt to design a Bayesian epistemology (Bovens & Hartmann 2003) do not seem to be aware of the work done in the first half of the twentieth century. However, an attempt to unify Information-A and Information-B seems a viable exercise. Also the connection between thermodynamics and information theory have become much closer, amongst others, due to the work of Gell-Mann & Lloyd (2003) (see also: Bais and Farmer 2008). Verlinde (2011, 2017) even presented a reduction of gravity to information (see the entry on information processing and thermodynamic entropy).

5.1 Philosophy of Information as An Extension of Philosophy of Mathematics

With respect to the main definitions of the concept of information, like Shannon Information, Kolmogorov complexity, semantic information and quantum information, a unifying approach to a philosophy of information is possible, when we interpret it as an extension to the philosophy of mathematics. The answer to questions like "What is data?" and "What is information?" then evolves from one's answer to the related questions like "What is a set?" and "What is a number?" With hindsight one can observe that many open problems in the philosophy of mathematics revolve around the notion of information.

If we look at the foundations of information and computation there are two notions that are crucial: the concept of a data set and the concept of an algorithm. Once we accept these notions as fundamental the rest of the theory data and computation unfolds quite naturally. One can "plug in" one's favorite epistemological or metaphysical stance here, but this does

not really affect foundational issues in the philosophy of computation and information. One might sustain a Formalist, Platonic or intuitionistic view of the mathematical universe (see entry on philosophy of mathematics) and still agree on the basic notion of what effective computation is. The theory of computing, because of its finitistic and constructivist nature, seems to live more or less on the common ground in which these theories overlap.

5.1.1 Information as a natural phenomenon

Information as a scientific concept emerges naturally in the context of our every day dealing with nature when we measure things. Examples are ordinary actions like measuring the size of an object with a stick, counting using our fingers, drawing a straight line using a piece of rope. These processes are the anchor points of abstract concepts like length, distance, number, straight line that form the building blocks of science. The fact that these concepts are rooted in our concrete experience of reality guarantees their applicability and usefulness. The earliest traces of information processing evolved around the notions of counting, administration and accountancy.

Example: Tally sticks
One of the most elementary information measuring devices is *unary counting* using a tally stick. Tally sticks were already used around 20,000 years ago. When a hypothetical prehistoric hunter killed a deer he could have registered this fact by making a scratch "|" on a piece of wood. Every stroke on such a stick represents an object/item/event. The process of unary counting is based on the elementary operation of *catenation of symbols* into *sequences*. This measuring method illustrates a primitive version of the concept of *extensiveness* of information: the length of the sequences is a measure for the amount of items counted. Note that such a sequential process of counting is

non-commutative and non-associative. If "|" is our basic symbol and ⊕ our concatenation operator then a sequence of signs has the form:

$$((\dots(| \oplus |)\dots) \oplus |) \oplus |)$$

A new symbol is always concatenated at the end of the sequence.

This example helps to understand the importance of *context* in the analysis of information. In itself a scratch on a stick may have no meaning at all, but as soon as we decide that such a scratch *represents* another object or event it becomes a *meaningful symbol*. When we manipulate it in such a context we process information. In principle a simple scratch can represent any event or object we like: symbols are conventional.

Definition: A *symbol* is a mark, sign or word that indicates, signifies, or is understood as representing an idea, object, or relationship.

Symbols are the semantic anchors by which symbol manipulating systems are tied to the world. Observe that the meta-statement:

The symbol "|" signifies object *y*.

if true, specifies semantic information:

- It is *wellformed*: the statement has a specific syntax.
- It is *meaningful*: Only in the context where the scratch "|" is actually made deliberately on, e.g., a tally stick or in a rock to mark a well defined occurrence it has a meaning.
- It is *truthful*.

Symbol manipulation can take many forms and is not restricted to sequences. Many examples of different forms of information processing can be found in prehistoric times.

Example: Counting sheep in Mesopotamia

With the process of urbanization, early accounting systems emerged in Mesopotamia around 8000 BCE using clay tokens to administer cattle (Schmandt-Besserat 1992). Different shaped tokens were used for different types of animals, e.g., sheep and goats. After the registration the tokens were packed in a globular clay container, with marks representing their content on the outside. The container was baked to make the registration permanent. Thus early forms of writing emerged. After 4000 BCE the tokens were mounted on a string to preserve the order.

The historical transformation from sets to strings is important. It is a more sophisticated form of coding of information. Formally we can distinguish several levels of complexity of token combination:

- An *unordered* collection of *similar* tokens in a container. This represents a *set*. The tokens can move freely in the container. The volume of the tokens is the only relevant quality.
- An *unordered* collection of tokens of *different types* in a container. This represents a so-called *multiset*. Both volume and frequency are relevant.
- An *ordered* collection of *typed* tokens on a string. This represents a *sequence* of symbols. In this case the length of the string is a relevant quality.

5.1.2 Symbol manipulation and extensiveness: sets, multisets and strings

Sequences of symbols code more information than multisets and multisets are more expressive than sets. Thus the emergence of writing itself can be seen as a quest to find the most expressive representation of administrative data. When measuring information in sequences of messages it is important to distinguish the aspects of *repetition*, *order* and *grouping*. The

extensive aspects of information can be studied in terms of such structural operations (see entry on substructural logics). We can study sets of messages in terms of operators defined on sequences of symbols.

Definition: Suppose m, n, o, p, … are symbols and $\oplus$ is a *tensor* or *concatentation* operator. We define the class of sequences:

1. Any symbol is a sequence
2. If α and β are sequences then $(\alpha \oplus \beta)$ is a sequence

For sequences we define the following basic properties on the level of symbol concatenation:

1. **Contraction:**

$$(m \oplus m) = m.$$

   Contraction destroys information about *frequency* in the sequence. Physical interpretation: two occurrences of the same symbol can collapse to one occurrence when they are concatenated.

2. **Commutativity:**

$$(m \oplus n) = (n \oplus m)$$

   Commutativity destroys information about *order* in the sequence. Physical interpretation: symbols may swap places when they are concatenated.

3. **Associativity:**

$$(p \oplus (q \oplus r)) = ((p \oplus q) \oplus r)$$

   Associativity destroys information about *nesting* in the sequence. Physical interpretation: symbols may be regrouped when they are

concatenated.

Observation: Systems of sequences with contraction, commutativity and associativity behave like sets. Consider the equation:

$$\{p,q\} \cup \{p,r\} = \{p,q,r\}$$

When we model the sets as two sequences $(p \oplus q)$ and $(p \oplus r)$, the corresponding implication is:

$$(p \oplus q), (p \oplus r) \vdash ((p \oplus q) \oplus r)$$

Proof:

$((p \oplus q) \oplus (p \oplus r))$	Concatenation
$((q \oplus p) \oplus (p \oplus r))$	Commutativity
$(((q \oplus p) \oplus p) \oplus r)$	Associativity
$((q \oplus (p \oplus p)) \oplus r)$	Associativity
$((q \oplus p) \oplus r)$	Contraction
$((p \oplus q) \oplus r)$	Commutativity

The structural aspects of sets, multisets and strings can be formulated in terms of these properties:

Sets: Sequences of messages collapse into sets under *contraction*, *commutativity* and *associativity*. A set is a collection of objects in which each element occurs only once:

$$\{a,b,c\} \cup \{b,c,d\} = \{a,b,c,d\}$$

and for which order is not relevant:

$$\{a,b,c\} = \{b,c,a\}.$$

Sets are associated with our normal everyday *naive concept of information* as new, previously unknown, information. We only update our set if we get a message we have not seen previously. This notion of information is *forgetful* both with respect to sequence and frequency. The set of messages cannot be reconstructed. This behavior is associated with the notion of *extensionality* of sets: we are only interested in equality of elements, not in frequency.

Multisets: Sequences of messages collapse into multisets under *commutativity* and *associativity*. A multiset is a collection of objects in which the same element can occur multiple times

$$\{a,b,c\} \cup \{b,c,d\} = \{a,b,b,c,c,d\}$$

and for which order is not relevant:

$$\{a,b,a\} = \{b,a,a\}.$$

Multisets are associated with a *resource sensitive concept of information* defined in *Shannon Information*. We are interested in the *frequency* of the messages. This concept is *forgetful* with regards to sequence. We update our set every time we get a message, but we forget the structure of the sequence. This behavior is associated with the notion of *extensiveness* of information: we are both interested in equality of elements, and in frequency.

Sequences: Sequences are associative. Sequences are ordered multisets: $aba \neq baa$. The whole structure of the sequence of a message is stored. Sequences are associated with *Kolmogorov complexity* defined as the length of a sequence of symbols.

Sets may be interpreted as spaces in which objects can move freely. When the same objects are in each others vicinity they collapse in to one object.

Multisets can be interpreted as spaces in which objects can move freely, with the constraint that the total number of objects stays constant. This is the standard notion of extensiveness: the total volume of a space stays constant, but the internal structure may differ. Sequences may be interpreted as spaces in which objects have a fixed location. In general a sequence contains more information than the derived multiset, which contains more information than the associated set.

> *Observation*: The interplay between the notion of sequences and multisets can be interpreted as a formalisation of the *malleability of a piece of wax* that pervades history of philosophy as the paradigm of information. Different sequences (forms) are representations of the same multiset (matter). The volume of the piece of wax (length of the string) is constant and thus a measure for the amount of information that can be represented in the wax (i.e. in the sequence of symbols). In terms of quantum physics the stability of the piece of wax seems to be an emergent property: the statistical instability of objects on an atomic level seem to even out when large quantities of them are manipulated.

5.1.3 Sets and numbers

The notion of a set in mathematics is considered to be fundamental. Any identifiable collection of discrete objects can be considered to be a set. The relation between theory of sets and the concept of information becomes clear when we analyze the basic statement:

$$e \in A$$

Which reads the object e is an element of the set A. Observe that this statement, if true, represents a piece of semantic information. It is wellformed, meaningful and truthful. (see entry on semantic conceptions of information) The concept of information is already at play in the basic

building blocks of mathematics. The philosophical question "What are sets?" the answer to the *ti esti* question, is determined *implicitly* by the Zermelo-Fraenkel axioms (see entry on set theory), the first of which is that of *extensionality*:

> Two sets are equal if they have the same elements.

The idea that mathematical concepts are defined implicitly by a set of axioms was proposed by Hilbert but is not uncontroversial (see entry on the Frege-Hilbert controversy). The fact that the definition is implicit entails that we only have *examples* of what sets are without the possibility to formulate any positive predicate that defines them. Elements of a set are not necessarily physical, nor abstract, nor spatial or temporal, nor simple, nor real. The only prerequisite is the possibility to formulate clear judgments about membership. This implicit definition of the notion of a set is not unproblematic. We might define objects that at first glance seem to be proper sets, which after scrutiny appear to be internally inconsistent. This is the basis for:

> **Russell's paradox**: This paradox, which motivated a lot of research into the foundations of mathematics, is a variant of the liars paradox attributed to the Cretan philosopher Epeimenides (ca. 6 BCE) who apparently stated that Cretans always lie. The crux of these paradoxes lies in the combination of the notions of: *Universality*, *Negation*, and *Self-reference*.

Any person who is not Cretan can state that all Cretans always lie. For a Cretan this is not possible because of the universal negative self-referential nature of the statement. If the statement is true, he is not lying which makes the statement untrue: a real paradox based on self contradiction. Along the same lines Russel coined the concept of the *set of all sets that are not member of themselves*, for which

membership cannot be determined. Apparently the *set of all sets* is an inadmissible object within set theory. In general there is in philosophy and mathematics a limit to the extent in which a system can verify statements about itself within the system. (For further discussion, see the entry on Russell's paradox.)

The implicit definition of the concepts of sets, entails that the class is essentially *open* itself. There are mathematical definitions of objects of which it is unclear or highly controversial whether they define a set or not.

Modern philosophy of mathematics starts with the Frege-Russell theory of numbers (Frege 1879, 1892, Goodstein 1957, see entry on alternative axiomatic set theories) in terms of sets. If we accept the notion of a class of objects as valid and fundamental, together with the notion of a one-to-one correspondence between classes of objects, then we can define numbers as sets of equinumerous classes.

> **Definition:** Two sets A and B are *equinumerous*, $A \sim B$, if there exists a one-to-one correspondence between them, i.e., a function $f : A \to B$ such that for every $a \in A$ there is exactly one $f(a) \in B$.

Any set of, say four, objects then becomes a representation of the number 4 and for any other set of objects we can establish membership to the equivalence class defining the number 4 by defining a one to one correspondence to our example set.

> **Definition:** If A is a finite set, then $S_A = \{X \mid X \sim A\}$ is the class of all sets equinumerous with A. The associated *generalization operation* is the *cardinality function*: $|A| = S_A = \{X \mid X \sim A\} = n$. This defines a *natural number* $|A| = n \in \mathbb{N}$ associated with the set A.

We can reconstruct large parts of the mathematical universe by selecting appropriate mathematical example objects to populate it, beginning with

the assumption that there is a single unique empty set $\emptyset$ which represents the number 0. This gives us the existence of a set with only one member $\{\emptyset\}$ to represent the number 1 and repeating this construction, $\{\emptyset, \{\emptyset\}\}$ for 2, the whole set of natural numbers $\mathbb{N}$ emerges. Elementary arithmetic then is defined on the basis of Peano's axioms:

1. Zero is a number.
2. If a is a number, the successor of a is a number.
3. Zero is not the successor of a number.
4. Two numbers of which the successors are equal are themselves equal.
5. (induction axiom.) If a set S of numbers contains zero and also the successor of every number in S, then every number is in S.

The fragment of the mathematical universe that emerges is relatively uncontroversial and both Platonists and constructivists might agree on its basic merits. On the basis of Peano's axioms we can define more complex functions like addition and multiplication which are closed on $\mathbb{N}$ and the inverse functions, subtraction and division, which are not closed and lead to the set of whole numbers $\mathbb{Z}$ and the rational numbers $\mathbb{Q}$.

5.1.4 Measuring information in numbers

We can define the concept of information for a number n by means of an unspecified function $I(n)$. We observe that addition and multiplication specify *multisets*: both are *non-contractive* and *commutative* and *associative*. Suppose we interpret the tensor operator $\oplus$ as multiplication $\times$. It is natural to define the *semantics* for $I(m \times n)$ in terms of addition. If we get both messages m and n, the total amount of information in the combined messages is the sum of the amount of information in the individual messages. This leads to the following constraints:

> **Definition:** *Additivity Constraint*:

$$I(m \times n) = I(m) + I(n)$$

Furthermore we want bigger numbers to contain more information than smaller ones, which gives a:

Definition: *Monotonicity Constraint*:

$$I(m) \leq I(m+1)$$

We also want to select a certain number a as our *basic unit of measurement*:

Definition: *Normalization Constraint*:

$$I(a) = 1$$

The following theorem is due to Rényi (1961):

Theorem: The Logarithm is the only mathematical operation that satisfies Additivity, Monotonicity and Normalisation.

Observation: The logarithm $\log_a n$ of a number n characterizes our intuitions about the concept of information in a number n *exactly*. When we decide that 1) multisets are the right formalisation of the notion of extensiveness, and 2) multiplication is the right operation to express additivity, then the logarithm is the only measurement function that satisfies our constraints.

We define:

Definition: For all natural numbers $n \in \mathbb{N}^+$

$$I(n) = \log_a n.$$

- For $a = 2$ our unit of measurement is the *bit*

- For $a = e$ (i.e., Euler's number) our unit of measurement is the *gnat*
- For $a = 10$ our unit of measurement is the *Hartley*

5.1.5 Measuring information and probabilities in sets of numbers

For finite sets we can now specify the amount of information we get when we know a certain element of a set conditional to knowing the set as a whole.

Definition: Suppose S is a finite set and we have:

$$e \in S$$

then,

$$I(e \mid S) = \log_a |S|$$

i.e., the log of the cardinality of the set.

The bigger the set, the harder the search is, the more information we get when we find what we are looking for. Conversely, without any further information the *probability* of selecting a certain element of S is $p_S(x) = \frac{1}{|S|}$. The associated function is the so-called Hartley function:

Definition: If a sample from a finite set S uniformly at random is picked, the information revealed after the outcome is known is given by the *Hartley function* (Hartley 1928):

$$H_0(S) = \log_a |S|$$

The combination of these definitions gives a theorem that ties together the notions of conditional information and probability:

Unification Theorem: If S is a finite set then,

$$I(x \mid S) = H_0(S)$$

The information about an element x of a set S conditional to the set is equal to the probability that we select this element x under uniform distribution, which is a measure of our *ignorance* if we know the set but not which element of the set is to be selected.

Observation: Note that the Hartley function unifies the concepts of *entropy* defined by Boltzmann $S = k \log W$, where W is the cardinality of the set of micro states of system S, with the concept of *Shannon information* $I_S(x) = -\log p(x)$. If we consider S to be a set of messages, then the probability that we select an element x from the set (i.e., get a message form S) under uniform distribution p is $\frac{1}{|S|}$. $H_0(S)$ is also known as the *Hartley Entropy* of S.

Using these results we define the *conditional amount of information* in a subset of a finite set as:

Definition: If A is a finite set and B is an arbitrary subset $B \subset A$, with $|A| = n$ and $|B| = k$ we have:

$$I(B \mid A) = \log_a \binom{n}{k}$$

This is just an application of our basic definition of information: the cardinality of the class of subsets of A with size k is $\binom{n}{k}$.

The formal properties of the concept of probability are specified by the Kolmogorov Axioms of Probability:

Definition: $P(E)$ is the probability P that some event E occurs. (Ω, F, P), with $P(\Omega) = 1$, is a *probability space*, with sample space Ω, *event space* and *probability measure*.

Let $P(E)$ be the probability P that some event E occurs. Let (Ω, F, P), with $P(\Omega) = 1$, be a *probability space*, with sample space Ω, *event space* F and *probability measure* P.

1. The probability of an event is a *non-negative real number*
2. There is a *unit of measure*. The probability that one of the events in the event space will occur is 1: $P(\Omega = 1)$
3. Probability is *additive over sets*:

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

One of the consequences is *monotonicity*: if $A \subseteq B$ implies $P(A) \leq P(B)$. Note that this is the same notion of additivity as defined for the concept of information. At subatomic level the Kolmogorov Axiom of additivity loses its validity in favor of a more subtle notion (see section 5.3).

5.1.6 Perspectives for unification

From a philosophical point of view the importance of this construction lies in the fact that it leads to an ontologically neutral concept of information based on a very limited robust base of axiomatic assumptions:

- It is **reductionist** in the sense that once one accepts the concepts like classes and mappings, the definition of the concept of *Information* in the context of more complex mathematical concepts naturally emerges.
- It is **universal** in the sense that the notion of a set is universal and open.
- It is **semantic** in the sense that the notion of a set itself is a semantic concept.

- It **unifies** a variety of notions (sets, cardinality, numbers, probability, extensiveness, entropy and information) in one coherent conceptual framework.
- It is **ontologically neutral** in the sense that the notion of a set or class does not imply any ontological constraint on its possible members.

This shows how Shannon's theory of information and Boltzmann's notion of entropy are rooted in more fundamental mathematical concepts. The notions of a *set of messages* or a *set of micro states* are specializations of the more general mathematical concept of a *set*. The concept of information already exists on this more fundamental level. Although many open questions still remain, specifically in the context of the relation between information theory and physics perspectives on a unified theory of information now look better than at the beginning of the twenty-first century.

5.1.7 Information processing and the flow of information

The definition of the amount of information in a number in therms of logarithms allows us to classify other mathematical functions in terms of their capacity to process information. The *Information Efficiency* of a function is the difference between the amount of information in the input of a function and the amount of information in the output (Adriaans 2016 [OIR]). It allows us to measure *how information flows* through a set of functions. We use the shorthand $f(\overline{x})$ for $f(x_1, x_2, \ldots, x_k)$:

Definition: *Information Efficiency of a Function*: Let $f : \mathbb{N}^k \to \mathbb{N}$ be a function of k variables. We have:

- the *input information* $I(\overline{x})$ and
- the *output information* $I(f(\overline{x}))$.
- The information efficiency of the expression $f(\overline{x})$ is

$$\delta(f(\overline{x})) = I(f(\overline{x})) - I(\overline{x})$$

- A function f is *information conserving* if $\delta(f(\overline{x})) = 0$ i.e., it contains exactly the amount of information in its input parameters,
- it is *information discarding* if $\delta(f(\overline{x})) < 0$ and
- it has *constant information* if $\delta(f(\overline{x})) = c$.
- it is *information expanding* if $\delta(f(\overline{x})) > 0$.

In general deterministic information processing systems do not *create* new information. They only *process* it. The following *fundamental theorem about the interaction between information and computation* is due to Adriaans and Van Emde Boas (2011):

Theorem: Deterministic programs do not expand information.

This is in line with both Shannon's theory and Kolmogorov complexity. The outcome of a deterministic program is always the same, so the probability of the outcome is 1 which gives under Shannon's theory, 0 bits of *new* information. Likewise for Kolmogorov complexity, the output of a program can never be more complex than the length of the program itself, plus a constant. This is analyzed in depth in Adriaans and Van Emde Boas (2011). In a deterministic world it is the case that if:

        program(input)=output

then

$$I(\text{output}) \leq I(\text{program}) + I(\text{input})$$

The essence of information is uncertainty and a message that occurs with probability "1" contains no information. The fact that it might take a long time to compute the number is irrelevant as long as the computation halts.

Infinite computations are studied in the theory of Scott domains (Abramsky & Jung 1994).

Estimating the information efficiency of elementary functions is not trivial. The primitive recursive functions (see entry on recursive functions) have one information expanding operation, *counting*, one information discarding operation, *choosing*, all the others are information neutral. The information efficiency of more complex operations is defined by a combination of counting and choosing. From an information efficiency point of view the elementary arithmetical functions are complex families of functions that describe computations with the same outcome, but with different computational histories.

Some arithmetical operations expand information, some have constant information and some discard information. During the execution of deterministic programs expansion of information may take place, but, if the program is effective, the descriptive complexity of the output is limited. The flow of information is determined by the succession of types of operations, and by the balance between the complexity of the operations and the number of variables.

We briefly discuss the information efficiency of the two basic recursive functions on two variables and their coding possibilities:

Addition When computing sums of natural numbers we make choices in the order of the computations that influence the information efficiency of the resulting computation. Depending on the order of the underlying operations a variety of values may emerge for the information efficiency: e.g., the conditional information in the number 200 given the set $\{2, 47, 53, 98\}$ varies with the method of computation:

$$(2 + 98) + (47 + 53) = (2 + 47) + (53 + 98) = 200$$

but:

$$\delta(2 + 98) + \delta(47 + 53) + \delta(100 + 100) \approx -1.08 >$$

$$\delta(2 + 47) + \delta(53 + 98) + \delta(49 + 151) \approx -1.74$$

Even from a cognitive point of view we experience this difference in difficulty when computing the same sum in different ways:

Observation: Information efficiency is non-associative for addition. if S is a finite set of natural numbers then the expression:

$$\delta\left(\sum_{i \in S} i\right)$$

is not defined.

This explains, in part, why the so-called subset Sum Problem (see section 6.3) is hard. The information efficiency of expressions describing sums of subsets of numbers is not uniquely defined, a fact that influences the possibilities to search these sets.

Addition is associated with information storage in terms of *sequences* or strings of symbols. It is *information discarding* for natural numbers bigger than 1. We have $\delta(a + b) < 0$ since $\log(a + b) < \log a + \log b$. Still, addition has information preserving qualities. If we add numbers with different log units we can reconstruct the frequency of the units from the resulting number:

$$\begin{aligned} 232 &= 200 + 30 + 2 \\ &= (2 \times 10^2) + (3 \times 10^1) + (2 \times 10^0) \\ &= 100 + 100 + 10 + 10 + 10 + 1 + 1 \end{aligned}$$

Since the information in the building blocks, 100, 10 and 1, is given the number representation can still be reconstructed. This implies that natural numbers code in terms of *addition of powers of k* in principle two types of information: value *and* frequency. We can use this insight to code *complex typed* information in *single* natural numbers. Basically it allows us to code any natural numbers in a string of symbols of length $\lceil \log_k n \rceil$, which specifies a quantitative measure for the amount of information in a number in terms of the length of its code. See section 3.3 for a historical analysis of the importance of the discovery of position systems for information theory.

Multiplication is by definition *information conserving*. We have: $\delta(a \times b) = 0$, since $\log(a \times b) = \log a + \log b$. Still multiplication does not preserve all information in its input: the order of the operation is lost. This is exactly what we want from an operator that characterizes an extensive measure: only the *extensive* qualities of the numbers are preserved. If we multiply two numbers 3×4, then the result, 12, allows us to reconstruct the original computation, in so far as we can reduce all its components to their most elementary values: $2 \times 2 \times 3 = 12$. This leads to the observation that some numbers act as *information building blocks* of other numbers, which gives us the concept of a *prime number*:

> **Definition:** A *prime number* is a number that is only divisible by itself or 1.

The concept of a prime number gives rise to the *Fundamental Theorem of Arithmetic*:

> **Theorem:** Every natural number n greater than 1 is either prime or a product of a multiset A_p of primes, and this multiset is unique for n.

The Fundamental Theorem of Arithmetic can be seen as a theorem about conservation of information: for every natural number there is a set of natural numbers that contains exactly that same amount of information. The factors of a number form a so-called *multiset*: a set that may contain multiple copies of the same element: e.g., the number 12 defines the multiset $\{2, 2, 3\}$ in which the number 2 occurs twice. This makes multisets a powerful device for coding information since it codes qualitative information (i.e., the numbers 2 and 3) as well as quantitative information (i.e., the fact that the number 2 occurs twice and the number 3 only once). This implies that natural numbers in terms of *multiplication of primes* also code two types of information: value *and* frequency. Again we can use this insight to code *complex typed* information in *single* natural numbers.

5.1.8 Information, primes, and factors

Position based number representations using addition of powers are straightforward and easy to handle and form the basis of most of our mathematical functions. This is not the case for coding systems based on multiplication. Many of the open questions in the philosophy of mathematics and information arise in the context of the concepts of the Fundamental Theorem of Arithmetic and Primes. We give a short overview:

> **(Ir)regularity of the set of primes.**
> Since antiquity it is known that there is an infinite number of primes. The proof is simple. Suppose the set of primes P is finite. Now multiply all elements of P and add 1. The resulting number cannot be divided by any member of P, so P is incomplete. An estimation of the density of the prime numbers given by the *Prime Number Theorem* (see entry in *Encyclopaedia Britannica* on Prime Number Theorem [OIR]). It states that the gaps between primes in the set of natural numbers of size n is roughly $\ln n$, where $\ln$ is the natural logarithm based on Euler's number e. A refinement of the density estimation is

given by the so-called *Riemannn hypothesis*, formulated by him in 1859 (Weisstein 2002 [OIR]), which is commonly regarded as deepest unsolved problems in mathematics, although most mathematicians consider the hypothesis to be true.

(In)efficiency of Factorization.

Since multiplication conserves information the function is, to an extent, reversible. The process of finding the unique set of primes for a certain natural number *n* is called *factorization*. Observe that the use of the term "only" in the definition of a prime number implies that this is in fact a *negative* characterization: a number *n* is prime if there exists no number between 1 and *n* that divides it. This gives us an *effective* procedure for factorization of a number *n* (simply try to divide *n* by all numbers between 1 and *n*), but such techniques are not *efficient*.

If we use a position system to represent the number *n* then the process of identifying factors of *n* by trial and error will take a deterministic computer program at most *n* trials which gives a computation time exponential in the length of the representation of the number which is $\lceil \log n \rceil$. Factorization by trial and error of a relatively simple number, of, say, two hundred digits, which codes a rather small message, could easily take a computer of the size of our whole universe longer than the time passed since the big bang. So, although theoretically feasible, such algorithms are completely unpractical.

Factorization is possibly an example of so-called *trapdoor* one-to-one function which is easy to compute from one side but very difficult in its inverse. Whether factorization is really difficult, remains an open question, although most mathematicians believe the problem to be hard. Note that factorization in this context can be seen as the process of decoding a message. If factorization is hard it can be used as an

encryption technique. Classical encryption is based on multiplying codes with large prime numbers. Suppose Alice has a message encoded as a large number *m* and she knows Bob has access to a large prime *p*. She sends the number $p \times m = n$ to Bob. Since Bob knows *p* he can easily reconstruct *m* by computing $m = n/p$. Since factorization is difficult any other person that receives the message *n* will have a hard time reconstructing *m*.

Primality testing versus Factorization.

Although at this moment efficient techniques for factorization on classical computers are not known to exist, there is an efficient algorithm that decides for us whether a number is prime or not: the so-called AKS primality test (Agrawal et al. 2004). So, we might know a number is not prime, while we still do not have access to its set of factors.

Classical- versus Quantum Computing.

Theoretically factorization is efficient on quantum computers using Shor's algorithm (Shor 1997). This algorithm has a non-classical quantum subroutine, embedded in a deterministic classical program. Collections of quantum bits can be modeled in terms of complex higher dimensional vector-spaces, that, in principle, allow us to analyze an exponential number 2^n of correlations between collections of *n* objects. Currently it is not clear whether larger quantum computers will be stable enough to facilitate practical applications, but that the world at quantum level has relevant computational possibilities can not be doubted anymore, e.g., quantum random generators are available as a commercial product (see *Wikipedia* entry on Hardware random number generator [OIR]). As soon as viable quantum computers become available almost all of the current encryption techniques become useless, although they can be replaced

by quantum versions of encryption techniques (see the entry on Quantum Computiong).

Observation: The existence of an infinite set of prime numbers is an indication that, although the set of natural numbers ℕ is *defined* by the Peano axioms, this characterization is fundamentally *incomplete*.

There is an infinite number of observations we can make about the set ℕ that are not implied directly by the axioms, but involve a considerable amount of computation.

5.1.9 Incompleteness of arithmetic

In a landmark paper in 1931 Kurt Gödel proved that any consistent formal system that contains elementary arithmetic is fundamentally incomplete in the sense that it contains true statements that cannot be proved within the system. In a philosophical context this implies that the semantics of a formal system rich enough to contain elementary mathematics cannot be defined in terms of mathematical functions within the system, i.e., there are statements that contain semantic information about the system in the sense of being *well-formed*, *meaningful* and *truthful* without being *computable*.

Central is the concept of a *Recursive Function*. (see entry on recursive functions). Such functions are defined on numbers. Gödel's notion of a recursive function is closest to what we would associate with computation in every day life. Basically they are elementary arithmetical functions operating on natural numbers like addition, subtraction, multiplication and division and all other functions that can be defined on top of these.

We give the basic structure of the proof. Suppose F is a formal system, with the following components:

- It has a finite set of symbols
- It has a syntax that enables us to combine the symbols in to well-formed formulas
- It has a set of deterministic rules that allows us to derive new statements from given statements
- It contains elementary arithmetic as specified by Peano's axioms (see section 5.1.3 above).

Assume furthermore that F is consistent, i.e., it will never derive false statements form true ones. In his proof Gödel used the coding possibilities of multiplication to construct an image of the system (see the discussion of Gödel numbering from the entry on Gödel's Incompleteness Theorems). According to the fundamental theorem of arithmetic any number can be uniquely factored in to its primes. This defines a one-to-one relationship between multisets of numbers and numbers: the number 12 can be constructed on the basis of the multiset $\{2, 2, 3\}$ as $12 = 2 \times 2 \times 3$ and vice versa. This allows us to code any sequence of symbols as a specific individual number in the following way:

- A unique number is assigned to every symbol
- Prime numbers locate the position of the symbol in a string
- The actual number of the same primes in the set of prime factors defines the symbol

On the basis of this we can code any sequence of symbols as a so-called Gödel number, e.g., the number:

$$2 \times 3 \times 3 \times 5 \times 5 \times 7 = 3150$$

codes the multiset $\{2, 3, 3, 5, 5, 7\}$, which represents the string "abba" under the assumption $a = 1, b = 2$. With this observation conditions close to those that lead to the paradox of Russel are satisfied: elementary

arithmetic itself is rich enough to express: *Universality*, *Negation*, and *Self-reference*.

Since arithmetic is consistent this does not lead to paradoxes, but to incompleteness. By a construction related to the liars paradox Gödel proved that such a system must contain statements that are true but not provable: there are true sentences of the form "I am not provable".

> **Theorem:** Any formal system that contains elementary arithmetic is fundamentally *incomplete*. It contains statements that are *true* but not *provable*.

In the context of philosophy of information the incompleteness of mathematics is a direct consequence of the rich possibilities of the natural numbers to code information. In principle any deterministic formal system can be represented in terms of elementary arithmetical functions. Consequently, If such a system itself contains arithmetic as a sub system, it contains a infinite chain of endomorphisms (i.e., images of itself). Such a system is capable of reasoning about its own functions and proofs but since it is consistent (and thus the construction of paradoxes is not possible within the system) it is by necessity incomplete.

5.2 Information and Symbolic Computation

Recursive functions are abstract relations defined on natural numbers. In principle they can be defined without any reference to space and time. Such *functions* must be distinguished from the *operations* that we use to compute them. These operations mainly depend on the type of *symbolic representations* that we choose for them. We can represent the number seven as unary number ||||||| , binary number 111, Roman number VII, or Arabic number 7 and depending on our choice other types of *sequential symbol manipulation* can be used to compute the addition two plus five is seven, which can be represented as:

$$|| + ||||| = |||||||$$
$$10 + 101 = 111$$
$$II + V = VII$$
$$2 + 5 = 7$$

Consequently we can read these four sentences as four statements of the *same* mathematical truth, or as statements specifying the results of four *different* operations.

> *Observation*: There are (at least) two different perspectives from which we can study the notion of computation. The semantics of the symbols is different under these interpretations.

- The **Recursive Function Paradigm** studies computation in terms of *abstract functions* on *natural numbers* outside space and time. When interpreted as a mathematical fact, the $+$ sign in $10 + 101 = 111$ signifies the *mathematical function called addition* and the $=$ sign specifies *equality*.
- The **Symbol Manipulation Paradigm** studies computation in terms of *sequential operations* on *spatial representations of strings of symbols*. When interpreted as an operation the $+$ sign in $10 + 101 = 111$ signifies the *input for a sequential process of symbol manipulation* and the $=$ sign specifies the *result of that operation* or *output*. Such an algorithm could have the following form:

$$\begin{array}{r} 10 \\ +101 \\ \hline 111 \end{array}$$

This leads to the following tentative definition:

Definition: Deterministic Computing on a Macroscopic Scale can be defined as the local, sequential, manipulation of discrete objects according to deterministic rules.

In nature there are many other ways to perform such computations. One could use an abacus, study chemical processes or simply manipulate sequences of pebbles on a beach. The fact that the objects we manipulate are discrete together with the observation that the dataset is self-referential implies that the data domain is in principle Dedekind Infinite:

Definition: A set S is Dedekind Infinite if it has a bijection $f : S \rightarrow S'$ to a proper subset $S' \subset S$.

Since the data elements are discrete and finite the data domain will be countable infinite and therefore isomorphic to the set of natural numbers.

Definition: An infinite set S is *countable* if there exists a bijection with the set of natural numbers $\mathbb{N}$.

For infinite countable sets the notion of information is defined as follows:

Definition: Suppose S is countable and infinite and the function $f : S \rightarrow \mathbb{N}$ defines a one-to-one correspondence, then:

$$I(a \mid S, f) = \log f(a)$$

i.e., the amount of information in an index of a in S given f.

Note that the correspondence f is specified explicitly. As soon as such an index function is defined for a class of objects in the real world, the manipulation of these objects can be interpreted a form of computing.

5.2.1 Turing machines

Once we choose our symbols and our operational rules the system starts to produce statements about the world.

Observation: The meta-sentence:

The sign "0" is the symbol for zero.

specifies *semantic information* in the same sense as the statement $e \in A$ does for sets (see section 6.6). The statement is *wellformed*, *meaningful* and *truthful*.

We can study symbol manipulation in general on an abstract level, without any semantic implications. Such a theory was published by Alan Turing (1912–1954). Turing developed a general theory of computing focusing on the actual operations on symbols a mathematician performs (Turing 1936). For him a computer was an abstraction of a real mathematician sitting behind a desk, receiving problems written down on an in-tray (the inut), solving them according to fixed rules (the process) and leaving them to be picked up in an out-tray (the output).

Turing first formulated the notion of a general theory of computing along these lines. He proposed abstract machines that operate on infinite tapes with three symbols: blank (b), zero (0) and one (1). Consequently the data domain for Turing machines is the set of relevant tape configurations, which can be associated with the set of binary strings, consisting of zero's and one's. The machines can read and write symbols on the tape and they have a transition function that determines their actions under various conditions. On an abstract level Turing machines operate like functions.

Definition: If T_i is a *Turing machine* with index i and x is a string of zero's and one's on the tape that function as the *input* then $T_i(x)$ indicates the tape configuration after the machine has stopped, i.e., its *output*.

There is an infinite number of Turing machines. Turing discovered that there are so-called universal Turing machines U_j that can emulate any other Turing machine T_i.

> **Definition:** The expression $U_j(\overline{T_i}x)$ denotes the result of the emulation of the computation $T_i(x)$ by U_j after reading the self-delimiting description $\overline{T_i}$ of machine T_j.

The self-delimiting code is necessary because the input for U_j is coded as one string $\overline{T_i}x$. The universal machine U_j separates the input string $\overline{T_i}x$ in to its two constituent parts: the description of the machine $\overline{T_i}$ and the input for this machine x.

The self-referential nature of general computational systems allows us to construct machines that emulate other machines. This suggests the possible existence of a 'super machine' that emulates all possible computations on all possible machines and predicts their outcome. Using a technique called diagonalization, where one analyzes an enumeration of all possible machines running on descriptions of all possible machines, Turing proved that such a machine can not exist. More formally:

> **Theorem:** There is no Turing machine that predicts for any other Turing machine whether it stops on a certain input or not.

This implies that for a certain universal machine U_i the set of inputs on which it stops in finite time, is uncomputable. Not every machine will stop on every input.

> **Definition:** The *Halting set* is the set of combinations of Turing machines T_i and inputs x such that the computation $T_i(x)$ stops.

The existence of universal Turing machines indicates that the class embodies a notion of *universal computing*: any computation that can be performed on a specific Turing machine can also be performed on any other universal Turing machine. This is the mathematical foundation of the concept of a general programmable computer. These observations have bearing on the theory of information: certain measures of information, like Kolmogorov complexity, are defined, but not computable.

The proof of the existence uncomputable functions in the class of Turing machines is similar to the incompleteness result of Gödel for elementary arithmetic. Since Turing machines were defined to study the notion of computation and thus contain elementary arithmetic. The class of Turing machines is in itself rich enough to express: *Universality*, *Negation* and *Self-reference*. Consequently Turing machines can model universal negative statements about themselves. Turing's uncomputability proof is also motivated by the liars paradox, and the notion of a machine that stops on a certain input is similar to the notion of a proof that exists for a certain statement. At the same time Turing machines satisfy the conditions of Gödel's theorem: they can be modeled as a formal system F that contains elementary Peano arithmetic.

> *Observation*: Since they can emulate each other, the *Recursive Function Paradigm* and the *Symbol Manipulation Paradigm* have the *same computational strength*. Any function that can be computed in one paradigm can also by definition be computed in the other.

This insight can be generalized:

> **Definition:** An infinite set of computational functions is *Turing complete* if it has the same computational power as the general class of Turing machines. In this case it is called Turing equivalent. Such a system is, like the class of Turing machines, universal: it can emulate any computable function.

The philosophical implications of this observation are strong and rich, not only for the theory of computing but also for our understanding of the concept of information.

5.2.2 Universality and invariance

There is an intricate ration between the notion of universal computing and that of information. Precisely the fact that Turing Systems are universal allows us to say that they process information, because their universality entails invariance:

> **Small Invariance Theorem:** The concept of information in a string x measured as the length of the smallest string of symbols s of a program for a universal Turing machine U such that $U(s) = x$ is asymptotically invariant under selection of different universal Turing machines

Proof: The proof is simple and relevant for philosophy of information. Let $l(x)$ be the length of the string of symbols x. Suppose we have two different universal Turing machines U_j and U_k. Since they are universal they can both emulate the computation $T_i(x)$ of Turing machine T_i on input x:

$$U_j(\overline{T}_i^j x)$$

$$U_k(\overline{T}_i^k x)$$

Here $l(\overline{T}_i^j)$ is the length of the code for T_i on U_j and $l(\overline{T}_i^k)$ is the length of the code for T_i on U_k. Suppose $l(\overline{T}_i^j x) \ll l(\overline{T}_i^k x)$, i.e., the code for T_i on U_k is much less efficient that on U_j. Observe that the code for U_j has constant length, i.e., $l(\overline{U}_j^k) = c$. Since U_k is universal we can compute:

$$U_k(\overline{U}_j^k \, \overline{T}_i^j x)$$

The length of the input for this computation is:

$$l(\overline{U}_j^k \, \overline{T}_i^j x) = c + l(\overline{T}_i^j x)$$

Consequently the specification of the input for the computation $T_i(x)$ on the universal machine U_k never needs to longer than a constant. $\square$

This proof forms the basis of the theory of Kolmogorov complexity and is originally due to Solomonoff (1964***a or b?*) and discovered independently by Kolmogorov (1965) and Chaitin (1969). Note that this notion of invariance can be generalized over the class of Turing Complete Systems:

> **Big Invariance Theorem:** The concept of information measured in terms of the length of the input of a computation is asymptotically invariant for for Turing Complete systems.

Proof: Suppose we have a Turing Complete system F. By Definition any computation $T_i(x)$ on a Turing machine can be emulated in F and vice versa. There will be a special universal Turing machine U_F that emulates the computation $T_i(x)$ in F: $U_F(\overline{T}_i^F x)$. In principle $\overline{T}_i^F$ might use a very inefficient way to code programs such that $\overline{T}_i^F$ can have any length. Observe that the code for any other universal machine U_j emulated by U_F has constant length, i.e., $l(\overline{U}_j^F) = c$. Since U_F is universal we can also compute:

$$U_F(\overline{U}_j^F \, \overline{T}_i^j x)$$

The length of the input for this computation is:

$$l(\overline{U}_j^F \, \overline{T}_i^j x) = c + l(\overline{T}_i^j x)$$

Consequently the specification of the input for the computation $T_i(x)$ on the universal machine U_F never needs to be longer than a constant.

$\square$

How strong this result is becomes clear when we analyze the class of Turing complete systems in more detail. In the first half of the twentieth century three fundamentally different proposals for a general theory of computation were formulated: Gödel's recursive functions (Gödel 1931), Turing's automata (Turing 1937) and Church's Lambda Calculus (Church 1936). Each of these proposals in its own way clarifies aspects of the notion of computing. Later much more examples followed. The class of Turing equivalent systems is diverse. Apart from obvious candidates like all general purpose programming languages (C, Fortran, Prolog, etc.) it also contains some unexpected elements like various games (e.g., Magic: The Gathering [Churchill 2012 OIR]). The table below gives an overview of some conceptually interesting systems:

An overview of some Turing Complete systems

System	Data Domain
General Recursive Functions	Natural Numbers
Turing machines and their generalizations	Strings of symbols
Diophantine Equations	Integers
Lambda calculus	Terms
Type-0 languages	Sentences
Billiard Ball Computing	Ideal Billiard Balls
Cellular automata	Cells in one dimension
Conway's game of life	Cells in two dimensions

We make the following:

Observation: The class of Turing equivalent systems is open, because it is defined in terms of purely operational mappings between computations.

A direct consequence of this observation is:

Observation: The general theory of computation and information defined by the class of Complete Turing machines is ontologically neutral.

It is not possible to derive any necessary qualities of computational systems and data domains beyond the fact that they are general mathematical operations and structures. Data domains on which Turing equivalent systems are defined are not necessarily physical, nor temporal, nor spatial, not binary or digital. At any moment a new member for the class can be introduced. We know that there are computational systems that are weaker than the class of Turing machines (e.g., regular languages). We cannot rule out the possibility that one-day we come across a system that is stronger. The thesis that such a system does not exist is known as the Church-Turing thesis (see entry on Church-Turing thesis):

Church-Turing Thesis: The class of Turing machines characterizes the notion of algorithmic computing exactly.

We give an overview of the arguments for and against the thesis:

Arguments in favor of the thesis: The theory of Turing machines seems to be the most general theory possible that we can formulate since it is based on a very limited set of assumptions about what computing is. The fact that it is universal also points in the direction of its generality. It is difficult to conceive in what sense a more powerful system could be "more" universal. Even if we could think of such a more powerful system, the in- and output for such a system would have to be finite and discrete and the

computation time also finite. So, in the end, any computation would have the form of a finite function between finite data sets, and, in principle, all such relations can be modeled on Turing machines. The fact that all known systems of computation we have defined so far have the same power also corroborates the thesis.

Arguments against the thesis: The thesis is, in its present form, unprovable. The class of Turing Complete systems is open. It is defined on the basis of the existence of equivalence relations between systems. In this sense it does not define the notion of computing intrinsically. It doesn't not provide us with a philosophical theory that defines what computing *exactly is*. Consequently it does not allow us to exclude any system from the class *a priori*. At any time a proposal for a notion of computation might emerge that is fundamentally stronger. What is more, nature provides us with stronger notions of computing in the form of quantum computing. Quantum bits are really a generalization of the normal concept of bits that is associated with symbol manipulation, although in the end quantum computing does not seem to necessitate us to redefine the notion of computing so far. We can never rule out that research in physics, biology or chemistry will define systems that will force us to do so. Indeed various authors have suggested such systems but there is currently no consensus on convincing candidates (Davis 2006). Dershowitz and Gurevich (2008) claim to have vindicated the hypothesis, but this result is not generally accepted (see the discussion on "Computability – What would it mean to disprove the Church-Turing thesis", in the Other Internet Resources [OIR]).

Being Turing complete seems to be quite a natural condition for a (formal) system. Any system that is sufficiently rich to represent the natural numbers and elementary arithmetical operations is Turing complete. What is needed is a finite set of operations defined on a set of discrete finite data elements that is rich enough to make the system self-referential: its

operations can be described by its data elements. This explains, in part, why we can use mathematics to describe our world. The abstract notion of computation defined as functions on numbers in the abstract world mathematics and the concrete notion of computing by manipulation objects in our every day world around us coincide. The concepts of information end computation implied by the *Recursive Function Paradigm* and the *Symbol Manipulation Paradigm* are the same.

> *Observation*: If one accepts the fact that the Church-Turing thesis is open, this implies that the question about the existence of a universal notion of information is also open. At this stage of the research it is not possible to specify the *a priori* conditions for such a general theory.

5.3 Quantum Information and Beyond

We have a reasonable understanding of the concept of classical computing, but the implications of quantum physics for computing and information may determine the philosophical research agenda for decades to come if not longer. Still it is already clear that the research has repercussions for traditional philosophical positions: the Laplacian view (Laplace 1814 [1902]) that the universe is essentially deterministic seems to be falsified by empirical observations. Quantum random generators are commercially available (see *Wikipedia* entry on Hardware random number generator [OIR]) and quantum fluctuations do affect neurological, biological and physical processes at a macroscopic scale (Albrecht & Phillips 2014). Our universe is effectively a process that generates information permanently. Classical deterministic computing seems to be too weak a concept to understand its structure.

Standard computing on a macroscopic scale can be defined as *local, sequential, manipulation of discrete objects according to deterministic*

rules. Is has a natural interpretation in operations on the set of natural numbers *N* and a natural measurement function in the log operation $\log : \mathbb{N} \to \mathbb{R}$ associating a real number to every natural number. The definition gives us an adequate information measure for countable infinite sets, including number classes like the integers $\mathbb{Z}$, closed under *subtraction*, and the rational numbers $\mathbb{Q}$, closed under *division*.

The operation of *multiplication* with the associated *logarithmic* function characterizes our intuitions about additivity of the concept of information exactly. It leads to a natural bijection between the set of natural numbers $\mathbb{N}$ and the set of multisets of numbers (i.e., sets of prime factors). The notion of a multiset is associated with the properties of *commutativity* and *associativity*. This program can be extended to other classes of numbers when we study division algebras in higher dimensions. The following table gives an overview of some relevant number classes together with the *properties of the operation of multiplication* for these classes:

Number Class	Symbol	Dimensions	Countable	Linear	Commutative	Associative
Natural numbers	$\mathbb{N}$	1	Yes	Yes	Yes	Yes
Integers	$\mathbb{Z}$	1	Yes	Yes	Yes	Yes
Rational numbers	$\mathbb{Q}$	1	Yes	Yes	Yes	Yes
Real numbers	$\mathbb{R}$	1	No	Yes	Yes	Yes
Complex numbers	$\mathbb{C}$	2	No	No	Yes	Yes
Quaternions	$\mathbb{H}$	4	No	No	No	Yes
Octonions	$\mathbb{O}$	8	No	No	No	No

The table is ordered in terms of increasing generality. Starting from the set of natural numbers $\mathbb{N}$, various extensions are possible taking into account closure under subtraction, $\mathbb{Z}$, and division, $\mathbb{Q}$. This are the number classes

for which we have adequate finite symbolic representations on a macroscopic scale. For elements of the reel numbers $\mathbb{R}$ such a representations are not available. The real numbers $\mathbb{R}$ introduce the aspect of manipulation of infinite amounts of information in one operation.

Observation: For *almost all* $e \in \mathbb{R}$ we have $I(e) = \infty$.

More complex division algebras can be defined when we introduce imaginary numbers as negative squares $i^2 = -1$. We can now define complex numbers: $a + bi$, where a is the real part and bi the imaginary part. Complex numbers can be interpreted as vectors in a two dimensional plane. Consequently they lack the notion of a strict *linear order* between symbols. Addition is quite straightforward:

$$(a + bi) + (c + di) = (a + b) + (c + d)i$$

Multiplication follows the normal distribution rule but the result is less intuitive since it involves a negative term generated by i^2:

$$(a + bi)(c + di) = (ac - bd) + (bc + ad)i$$

In this context multiplication ceases to be a purely extensive operation:

Observation: Multiplication is not information efficient for imaginary numbers. Specifically the equality $\sqrt{xy} = \sqrt{x}\sqrt{y}$ does not hold, as is clear from the following example:

$$1 = \sqrt{1} = \sqrt{-1 \times -1} \neq \sqrt{-1}\sqrt{-1} = i \times i = -1$$

More complicated numbers systems with generalizations of this type of multiplication in 4 and 8 dimensions can be defined. Kervaire (1958) and Bott & Milnor (1958) independently proved that the only four division algebras built on the reals are $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$, so the table gives a comprehensive view of all possible algebra's that define a notion of

extensiveness. For each of the number classes in the table a separate theory of information measurement, based on the properties of multiplication, can be developed. For the countable classes $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ these theories ware equivalent to the standard concept of information implied by the notion of Turing equivalence. Up to the real numbers these theories satisfy our intuitive notions of extensiveness of information. For complex numbers the notion of *information efficiency* of multiplication is destroyed. The quaternions lack the property of *commutativity* and the octonions that of *associativity*. These models are not just abstract constructions since the algebras play an important role in our descriptions of nature:

1. Complex numbers are used to specify the mathematical models of quantum physics (Nielsen & Chuang 2000).
2. Quaternions do the same for Einstein's special theory of relativity (De Leo 1996).
3. Some physicists believe octonions form a theoretical basis for a unified theory of strong and electromagnetic forces (e.g., Furey 2015).

We briefly discuss the application of vector spaces in quantum physics. Classical information is measured in bits. Implementation of bits in nature involves macroscopic physical systems with at least two different stable states and a low energy reversible transition process (i.e., switches, relays, transistors). The most fundamental way to store information in nature on an atomic level involves qubits. The qubit is described by a state vector in a two-level quantum-mechanical system, which is formally equivalent to a two-dimensional vector space over the complex numbers (Von Neumann 1932; Nielsen & Chuang 2000). Quantum algorithms have, in some cases, a fundamentally lower complexity (e.g., Shor's algorithm for factorization of integers (Shor 1997)).

Definition: The *quantum bit*, or *qubit*, is a generalization of the classical bit. The quantum state of qubit is represented as the linear superposition of two orthonormal basis vectors:

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Here the so-called Dirac or "bra-ket" notion is used: where $|0\rangle$ and $|1\rangle$ are pronounced as "ket 0" and "ket 1". The two vectors together form the *computational basis* $\{|0\rangle, |1\rangle\}$, which defines a vector in a two-dimensional *Hilbert space*. A combination of n qubits is represented by a *superposition vector* in a 2^n dimensional Hilbert space, e.g.:

$$|00\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, |01\rangle = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, |10\rangle = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, |11\rangle = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

A pure qubit is a coherent *superposition* of the basis states:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

where α and β are complex numbers, with the constraint:

$$|\alpha|^2 + |\beta|^2 = 1$$

In this way the values can be interpreted as probabilities: $|\alpha|^2$ is the probability that the qubit has value 0 and $|\beta|^2$ is the probability that the qubit has value 1.

Under this mathematical model our intuitions about computing as local, sequential, manipulation of discrete objects according to deterministic rules evolve in to a much richer paradigm:

1. **Infinite information** The introduction of *real numbers* facilitates the manipulation of objects of infinite descriptive complexity, although there is currently no indication that this expressivity is actually necessary in quantum physics.

2. **Non-classical probability** *Complex numbers* facilitate a richer notion of extensiveness in which probabilities cease to be classical. The third axiom of Kolmogorov loses its validity in favor of probabilities that enhance or suppress each other, consequently extensiveness of information is lost.

3. **Superposition and Entanglement** The representation of qubits in terms of complex high dimensional vector spaces implies that qubits cease to be isolated discrete objects. Quantum bits can be in superposition, a situation in which they are in two discrete states at the same time. Quantum bits fluctuate and consequently they *generate* information. Moreover quantum states of qubits can be correlated even when the information bearers are separated by a long distance in space. This phenomenon, known as *entanglement* destroys the property of *locality* of classical computing (see the entry on quantum entanglement and information).

From this analysis it is clear that the description of our universe at very small (and very large) scales involves mathematical models that are alien to our experience of reality in everyday life. The properties that allow us to understand the world (the existence of stable, discrete objects that preserve their identity in space and time) seem to be *emergent* aspects of a much more complex reality that is incomprehensible to us outside its mathematical formulation. Yet, at a macroscopic level, the universe facilitates elementary processes, like counting, measuring lengths, and the manipulation of symbols, that allow us to develop a consistent hierarchy of mathematical models some of which seems to describe the deeper underlying structure of reality.

In a sense the same mathematical properties that drove the development of elementary accounting systems in Mesopotamia four thousand years ago, still help us to penetrate in to the world of subatomic structures. In the past decennia information seems to have become a vital concept in physics. Seth Lloyd and others (Zuse 1969; Wheeler 1990; Schmidhuber 1997b; Wolfram 2002; Hutter 2010) have analyzed computational models of various physical systems. The notion of information seems to play a major role in the analysis of black holes (Lloyd & Ng 2004; Bekenstein 1994 [OIR]). Erik Verlinde (2011, 2017) has proposed a theory in which gravity is analyzed in terms of information. For the moment these models seem to be purely descriptive without any possibility of empirical verification.

6. Anomalies, Paradoxes, and Problems

Some of the fundamental issues in philosophy of Information are closely related to existing philosophical problems, others seem to be new. In this paragraph we discuss a number of observations that may determine the future research agenda. Some relevant questions are:

1. Are there uniquely identifying descriptions that do not contain all information about the object they refer to?
2. Does computation create new information?
3. Is there a difference between construction and systematic search?

Since Frege most mathematicians seem to believe that the answer to the first question is positive (Frege 1879, 1892). The descriptions "The morning star" and "The evening star" are associated with *procedures* to identify the planet Venus, but they do not give access to all information about the object itself. If this were so the discovery that the evening star is in fact also the morning star would be uninformative. If we want to maintain this position we get into conflict, because in terms of information theory the answer to the second question is negative (see section 5.1.7).

Yet this observation is highly counter intuitive, because it implies that we never can construct new information on the basis of deterministic computation, which leads to the third question. These issues cluster around one of the fundamental open problems of Philosophy of Information:

Open problem What is the interaction between Information and Computation?

Why would we compute at all, if according to our known information measures, deterministic computing does not produce new information? The question could be rephrased as: should we use Kolmogorov or Levin complexity (Levin 1973, 1974, 1984) as our basic information measure? In fact both choices lead to relevant, but fundamentally different, theories of information. When using the Levin measure, computing generates information and the answer to the three questions above is a "yes", when using Kolmogorov this is not the case. The questions are related to many problems both in mathematics and computer science. Related issues like approximation, computability and partial information are also studied in the context of Scott domains (Abramsky & Jung 1994). Below we discuss some relevant observations.

6.1 The Paradox of Systematic Search

The essence of information is the fact that it reduces uncertainty. This observation leads to problems in opaque contexts, for instance, when we search an object. This is illustrated by Meno's paradox (see entry on epistemic paradoxes):

And how will you enquire, Socrates, into that which you do not know? What will you put forth as the subject of enquiry? And if you find what you want, how will you ever know that this is the thing which you did not know? (Plato, Meno, 80d1-4)

The paradox is related to other open problems in computer science and philosophy. Suppose that John is looking for a unicorn. It is very unlikely that unicorns exist, so, in terms of Shannon's theory, John gets a lot of information if he finds one. Yet from a descriptive Kolmogorov point of view, John does not get new information, since he already knows what unicorns are. The related *paradox of systematic search* might be formulated as follows:

Any information that can be found by means of systematic search has no value, since we are certain to find it, given enough time. Consequently information only has value as long as we are uncertain about its existence, but then, since we already know what we are looking for, we get no new information when we find out that it exists.

Example: Goldbach conjectured in 1742 that every even number bigger than 2 could be written as the sum of two primes. Until today this conjecture remains unproved. Consider the term "The first number that violates Goldbach's conjecture". It does not give us all information about the number, since the number might not exist. The prefix "the first" ensures the description, if it exists, is unique, and it gives us an algorithm to find the number. It is a *partial uniquely identifying* description. This algorithm is only *effective* if the number really exists, otherwise it will run forever. If we find the number this will be great news, but from the perspective of descriptive complexity the number itself will be totally uninteresting, since we already know the relevant properties to find it. Observe that, even if we have a number n that is a counter example to Goldbach's conjecture, it might be difficult to verify this: we might have to check almost all primes $\leq n$. This can be done *effectively* (we will always get a result) but not, as far as

we know, *efficiently* (it might take "close" to *n* different computations) .

A possible solution is to specify the constraint that it is *illegal* to measure the information content of an object in terms of partial descriptions, but this would destroy our theory of descriptive complexity. Note that the complexity of an object is the length of the shortest program that produces an object on a universal Turing machine. In this sense the phrase "the first number that violates Goldbach's conjecture" is a perfect description of a program, and it adequately measures the descriptive complexity of such a number. The short description reflects the fact that the number, if it exists, is very special, and thus it has a high possibility to occur in some mathematical context.

There are relations which well-studied philosophical problems like the Anselm's ontological argument for God's existence and the Kantian counter claim that existence is not a predicate. In order to avoid similar problems Russell proposed to interpret unique descriptions existentially (Russell 1905): A sentence like "The king of France is bold" would have the following logical structure:

$$\exists(x)(KF(x) \wedge \forall(y)(KF(y) \rightarrow x = y) \wedge B(x))$$

This interpretation does not help us to analyze decision problems that deal with existence. Suppose the predicate *L* is true of *x* if I'm looking for *x*, then the logical structure of the phrase "I'm looking for the king of France" would be:

$$\exists(x)(KF(x) \wedge \forall(y)(KF(y) \rightarrow x = y) \wedge L(x)),$$

i.e., if the king of France does not exist it cannot be true that I am looking for him, which is unsatisfactory. Kripke (1971) criticized Russell's solution and proposed his so-called causal theory of reference in which a

name get its reference by an initial act of "baptism". It then becomes a *rigid designator* (see entry on rigid designators) that can be followed back to that original act via causal chains. In this way *ad hoc* descriptions like "John was the fourth person to step out of the elevator this morning" can establish a semantics for a name.

In the context of mathematics and information theory the corresponding concept is that of names, constructive predicates and ad-hoc predicates of numbers. For any number there will be in principle an infinite number of true statements about that number. Since elementary arithmetic is incomplete there will be statements about numbers that are true but unprovable. In the limit a vanishing fragment of numbers will have true predicates that actually compress their description. Consider the following statements:

1. The symbol "8" is the name for the number eight.
2. The number *x* is the 1000th Fibonacci number.
3. The number *x* is the first number that violates the Goldbach conjecture.

The first statement simply specifies a name for a number. The second statement gives a partial description that is constructive, information compressing and unique. The 1000th Fibonacci number has 209 digits, so the description "the 1000th Fibonacci number" is much more efficient than the actual name of the number. Moreover, we have an algorithm to construct the number. This might not be that case for the description in the third statement. We do not know whether the first number that violates the Goldbach conjecture exists, but if it does, the description might well be *ad hoc* and thus gives us no clue to construct the number. This rise to the conjecture that there are *data compressing effective ad hoc descriptions*:

Conjecture: There exist numbers that are compressed by non-constructive unique effective descriptions, i.e., the validity of the description can be checked effectively given the number, but the number cannot be constructed effectively from the description, except by means of systematic search.

The conjecture is a more general variant of the so-called P vs. NP thesis (see section 6.3). If one replaces the term "effective" with the term "efficient" one gets a formulation of the P $\neq$ NP thesis.

6.2 Effective Search in Finite Sets

When we restrict ourselves to effective search in finite sets, the problem of partial descriptions, and construction versus search remain. It seems natural to assume that when one has a definition of a set of numbers, then one also has all the information about the members of the set and about its subsets, but this is not true. In general the computation of the amount of information in a set of numbers is a highly non-trivial issue. We give some results:

> **Lemma** A subset $A \subset S$ of a Sset can contain more information conditional to the set than the set itself.

Proof: Consider the set S of all natural numbers smaller than n. The descriptive complexity of this set in bits is $\log_2 n + c$. Now construct A by selecting half of the elements of S randomly. Observe that:

$$I(A \mid S) = \log_2 \binom{n}{n/2}$$

We have:

$$\lim_{n \to \infty} \frac{I(A \mid S)}{n} = \lim_{n \to \infty} \frac{\log_2 \binom{n}{n/2}}{n} = 1$$

The conditional descriptive complexity of this set will be: $I(A \mid S) \approx n + c \gg \log n + c$. $\square$

A direct consequence is that we can lose information when we merge two sets. An even stronger result is:

> **Lemma:** An element of a set can contain more information than the set itself.

Proof: Consider the set S of natural numbers smaller then 2^n. The cardinality of S is 2^n. The descriptive complexity of this set is $\log n + c$ bits, but for half of the elements of S we need n bits to describe them. $\square$

In this case the description of the set itself is highly compressible, but it still contains non-compressible elements. When we merge or split sets of numbers, or add or remove elements, the effects on the amount of information are in general hard to predict and might even be uncomputable:

> **Theorem:** Information is not monotone under set theoretical operations

Proof: Immediate consequence of the lemmas above. $\square$

This shows how the notion of information pervades our everyday life. When John has two apples in his pocket it seems that he can do whatever he wants with them, but, in fact, as soon as he chooses one of the two, he has created (new) information. The consequences for search problems are clear: we can always effectively perform bounded search on the elements and the set of subsets of a set. Consequently when we search for such a set of subsets by means of partial descriptions then the result generates (new) information. This analysis prima facie appears to force us to accept that in mathematics there are simple descriptions that allow us to identify

complex objects by means of systematic search. When we look for the object we have only little information about it, when we finally find it our information increases to the set of full facts about the object searched. This is in conflict with our current theories of information (Shannon and Kolmogorov): any description that allows us to identify an object effectively by deterministic search contains all relevant information about the object. The time complexity of the search process then is irrelevant.

6.3 The P versus NP Problem, Descriptive Complexity Versus Time Complexity

In the past decennia mathematicians have been pondering about a related question: suppose it *would* be easy to check whether I have found what I'm looking for, how hard can it be to find such an object? In mathematics and computer science there seems to be a considerable class of decision problems that cannot be solved constructively in polynomial time, $t(x) = x^c$, where c is a constant and x is the length of the input), but only through systematic search of a large part of the solution space, which might take exponential time, $t(x) = c^x$. This difference roughly coincides with the separation of problems that are computationally feasible from those that are not.

The issue of the existence of such problems has been framed as the possible equivalence of the class P of decision problems that can be solved in time polynomial to the input to the class NP of problems for which the solution can be checked in time polynomial to the input. (Garey & Johnson 1979; see also Cook 2000 [OIR] for a good introduction.)

Example: A well-known example in the class NP is the so-called *subset sum* problem: given a finite set of natural numbers S, is there a subset $S' \subseteq S$ that sums up to some number k? It is clear that when someone proposes a solution $X \subseteq S$ to this problem we can easily check whether the elements of X add up to k, but we might have to check almost all subsets of S in order to find such a solution ourselves.

This is an example of a so-called decision problem. The answer is a simple "yes" or "no", but it might be hard to find the answer. Observe that the formulation of the question conditional to S has descriptive complexity $\log k + c$, whereas most random subsets of S have a conditional descriptive complexity of $|S|$. So any subset S' that adds up to k might have a descriptive complexity that is bigger then the formulation of the search problem. In this sense search seems to generate information. The problem is that if such a set exists the search process is bounded, and thus effective, which means that the phrase "the first subset of S that adds up to k" is an adequate description. If P = NP then the Kolmogorov complexity and the Levin complexity of the set S' we find roughly coincide, if $P \neq NP$ then in some cases $Kt(S') \gg K(S')$. Both positions, the theory that search generates new information and the theory that it does not, are counterintuitive from different perspectives.

The P vs. NP problem, that appears to be very hard, has been a rich source of research in computer science and mathematics although relatively little has been published on its philosophical relevance. That a solution might have profound philosophical impact is illustrated by a quote from Scott Aaronson:

> If P = NP, then the world would be a profoundly different place than we usually assume it to be. There would be no special value in "creative leaps," no fundamental gap between solving a problem and recognizing the solution once it's found. Everyone who could appreciate a symphony would be Mozart; everyone who could follow a step-by-step argument would be Gauss.... (Aaronson 2006 – in the Other Internet Resources)

In fact, if P = NP then every object that is easy to describe and easy to check is easy to find.

6.4 Model Selection and Data Compression

In current scientific methodology the sequential aspects of the scientific process are formalized in terms of the empirical cycle, which according to de Groot (1969) has the following stages:

1. Observation: The observation of a phenomenon and inquiry concerning its causes.
2. Induction: The formulation of hypotheses—generalized explanations for the phenomenon.
3. Deduction: The formulation of experiments that will test the hypotheses (i.e., confirm them if true, refute them if false).
4. Testing: The procedures by which the hypotheses are tested and data are collected.
5. Evaluation: The interpretation of the data and the formulation of a theory—an abductive argument that presents the results of the experiment as the most reasonable explanation for the phenomenon.

In the context of information theory the set of observations will be a data set and we can construct models by observing regularities in this data set. Science aims at the construction of true models of our reality. It is in this sense a semantical venture. In the 21-st century the process of theory formation and testing will for the largest part be done automatically by computers working on large databases with observations. Turing award winner Jim Grey framed the emerging discipline of e-science as the fourth data-driven paradigm of science. The others are empirical, theoretical and computational. As such the process of automatic theory construction on the basis of data is part of the methodology of science and consequently of philosophy of information (Adriaans & Zantinge 1996; Bell, Hey, &

Szalay 2009; Hey, Tansley, and Tolle 2009). Many well-known learning algorithms, like decision tree induction, support vector machines, normalized information distance and neural networks, use entropy based information measures to extract meaningful and useful models out of large data bases. The very name of the discipline Knowledge Discovery in Databases (KDD) is witness to the ambition of the Big Data research program. We quote:

> At an abstract level, the KDD field is concerned with the development of methods and techniques for making sense of data. The basic problem addressed by the KDD process is one of mapping low-level data (which are typically too voluminous to understand and digest easily) into other forms that might be more compact (for example, a short report), more abstract (for example, a descriptive approximation or model of the process that generated the data), or more useful (for example, a predictive model for estimating the value of future cases). At the core of the process is the application of specific data-mining methods for pattern discovery and extraction. (Fayyad, Piatetsky-Shapiro, & Smyth 1996: 37)

Much of the current research focuses on the issue of selecting an optimal computational model for a data set. The theory of Kolmogorov complexity is an interesting methodological foundation to study learning and theory construction as a form of data compression. The intuition is that the shortest theory that still explains the data is also the best model for generalization of the observations. A crucial distinction in this context is the one between *one- and two-part code optimization*:

- **One-part Code Optimization:** The methodological aspects of the theory of Kolmogorov complexity become clear if we follow its definition. We begin with a well-formed dataset y and select an

appropriate universal machine U_j. The expression $U_j(\overline{T_i}x) = y$ is a true sentence that gives us information about y. The first move in the development of a theory of measurement is to force all expressiveness to the instructional or procedural part of the sentence by a restriction to sentences that describe computations on empty input:

$$U_j(\overline{T_i}\emptyset) = y$$

This restriction is vital for the proof of invariance. From this, in principle infinite, class of sentences we can measure the length when represented as a program. We select the ones (there might be more than one) of the form $\overline{T_i}$ that are shortest. The length $l(\overline{T_i})$ of such a shortest description is a measure for the information content of y. It is asymptotic in the sense that, when the data set y grows to an infinite length, the information content assigned by the choice of another Turing machine will never vary by more than a constant in the limit. Kolmogorov complexity measures the information content of a data set in terms of the shortest description of the set of instructions that produces the data set on a universal computing device.

- **Two-part Code Optimization:** Note that by restricting ourselves to programs with empty input and the focus on the *length* of programs instead of their *content* we gain the quality of invariance for our measure, but we also lose a lot of expressiveness. The information in the actual program that produces the data set is neglected. Subsequent research therefore has focused on techniques to make the explanatory power, hidden in the Kolmogorov complexity measure, explicit.

A possible approach is suggested by an interpretation of Bayes' law. If we combine Shannon's notion of an optimal code with Bayes' law, we get a rough theory about optimal model selection. Let $\mathcal{H}$ be a set of hypotheses

and let x be a data set. Using Bayes' law, the optimal computational model under this distribution would be:

$$M_{map}(x) = argmax_{M \in \mathcal{H}} \frac{P(M)P(x \mid M)}{P(x)}$$

This is equivalent to optimizing:

$$argmin_{M \in \mathcal{H}} - \log P(M) - \log P(x \mid M)$$

Here $-\log P(M)$ can be interpreted as the length of the optimal *model code* in Shannon's sense and $-\log P(x \mid M)$ as the length of the optimal *data-to-model code*; i.e., the data interpreted with help of the model. This insight is canonized in the so-called:

> **Minimum Description Length (MDL) Principle:** The best theory to explain a data set is the one that minimizes the sum in bits of a description of the theory (model code) and of the data set encoded with the theory (the data to model code).

The MDL principle is often referred to as a modern version of Ockham's razor (see entry on William of Ockham), although in its original form Ockham's razor is an ontological principle and has little to do with data compression. In many cases MDL is a valid heuristic tool and the mathematical properties of the theory have been studied extensively (Grünwald 2007). Still MDL, Ockham's razor and two-part code optimization have been the subject of considerable debate in the past decennia (e.g., Domingos 1998; McAllister 2003).

The philosophical implications of the work initiated by Solomonoff, Kolmogorov and Chaitin in the sixties of the 20-th century are fundamental and diverse. The universal distribution m proposed by Solomonoff, for instance, codifies all possible mathematical knowledge and when updated on the basis of empirical observations would in

principle converge to an optimal scientific model of our world. In this sense the choice of a universal Turing machine as basis for our theory of information measurement has philosophical importance, specifically for methodology of science. A choice for a universal Turing machine can be seen as a choice of *a set of bias* for our methodology. There are roughly two schools:

- **Poor machine:** choose a small universal Turing machine. If the machine is small it is also general and universal, since there is no room to encode any bias in to the machine. Moreover a restriction to small machines gives small overhead when emulating one machine on the other so the version of Kolmogorov complexity you get gives a measurement with a smaller asymptotic margin. Hutter explicitly defends the choice of "natural" small machines (Hutter 2005; Rathmanner & Hutter 2011), but also Li and Vitányi (2008) seem to suggest the use of small models.
- **Rich machine:** choose a big machine that explicitly reflects what you already know about the world. For Solomonoff, the inventor of algorithmic complexity, the choice of a universal Turing machine is the choice for a universal prior. He defends an evolutionary approach to learning in which an agent constantly adapts the prior to what he already has discovered. The selection of your reference Turing machine uniquely characterizes your *a priori* information (Solomonoff 1997).

Both approaches have their value. For rigid mathematical proofs the poor machine approach is often best. For practical applications on finite data sets the rich model strategy often gets much better results, since a poor machine would have to "re-invent the wheel" every time it compresses a data set. This leads to the conclusion that Kolmogorov complexity inherently contains a theory about scientific bias and as such implies a methodology in which the class of admissible universal models should be explicitly formulated and motivated *a priori*. In the past decennia there have been a number of proposals to define a formal unit of measurement of the amount of structural (or model-) information in a data set.

- Aesthetic measure (Birkhoff 1950)
- Sophistication (Koppel 1987; Antunes et al. 2006; Antunes & Fortnow 2003)
- Logical Depth (Bennet 1988)
- Effective complexity (Gell-Mann, Lloyd 2003)
- Meaningful Information (Vitányi 2006)
- Self-dissimilarity (Wolpert & Macready 2007)
- Computational Depth (Antunes et al. 2006)
- Facticity (Adriaans 2008)

Three intuitions dominate the research. A string is "interesting" when …

- a certain amount of computation is involved in its creation (Sophistication, Computational Depth);
- there is a balance between the model-code and the data-code under two-part code optimization (effective complexity, facticity);
- it has internal phase transitions (self-dissimilarity).

Such models penalize both maximal entropy and low information content. The exact relationship between these intuitions is unclear. The problem of meaningful information has been researched extensively in the past years, but the ambition to formulate a universal method for model selection based on compression techniques seems to be misguided:

Observation: A measure of meaningful information based on two-part code optimization can never be *invariant* in the sense of Kolmogorov complexity (Bloem et al. 2015), even if we restrict the underlying theory to total primitive recursive functions (Adriaans forthcoming***not in bib*).

There seems to be no *a priori* mathematical justification for the approach, although two-part code optimization continues to be a valid approach in an empirical setting of data sets that have been created on the basis of repeated observations. Phenomena that might be related to a theory of structural information and that currently are ill-understood are: phase transitions in the hardness of satisfiability problems related to their complexity (Simon & Dubois 1989; Crawford & Auton 1993) and phase transitions in the expressiveness of Turing machines related to their complexity (Crutchfield & Young 1989, 1990; Langton 1990; Dufort & Lumsden 1994).

6.5 Determinism and Thermodynamics

Many basic concepts of information theory were developed in the nineteenth century in the context of the emerging science of thermodynamics. There is a reasonable understanding of the relationship between Kolmogorov Complexity and Shannon information (Li & Vitányi 2008; Grünwald & Vitányi 2008; Cover & Thomas 2006), but the unification between the notion of entropy in thermodynamics and Shannon-Kolmogorov information is very incomplete apart from some very *ad hoc* insights (Harremoës & Topsøe 2008; Bais & Farmer 2008). Fredkin and Toffoli (1982) have proposed so-called billiard ball computers to study reversible systems in thermodynamics (Durand-Lose 2002) (see the entry on information processing and thermodynamic entropy). Possible theoretical models could with high probability be corroborated with feasible experiments (e.g., Joule's adiabatic expansion, see Adriaans 2008).

Questions that emerge are:

- What is a computational process from a thermodynamical point of view?

- Can a thermodynamic theory of computing serve as a theory of non-equilibrium dynamics?
- Is the expressiveness of real numbers necessary for a physical description of our universe?

These problems seem to be hard because 150 years of research in thermodynamics still leaves us with a lot of conceptual unclarities in the heart of the theory of thermodynamics itself (see entry on thermodynamic asymmetry in time).

Real numbers are not accessible to us in finite computational processes yet they do play a role in our analysis of thermodynamic processes. The most elegant models of physical systems are based on functions in continuous spaces. In such models almost all points in space carry an infinite amount of information. Yet, the cornerstone of thermodynamics is that a finite amount of space has finite entropy. There is, on the basis of the theory of quantum information, no fundamental reason to assume that the expressiveness of real numbers is never used in nature itself on this level. This problem is related to questions studied in philosophy of mathematics (an intuitionistic versus a more platonic view). The issue is central in some of the more philosophical discussions on the nature of computation and information (Putnam 1988; Searle 1990). The problem is also related to the notion of phase transitions in the description of nature (e.g., thermodynamics versus statistical mechanics) and to the idea of levels of abstraction (Floridi 2002).

In the past decade some progress has been made in the analysis of these questions. A basic insight is that the interaction between time and computational processes can be understood at an abstract mathematical level, without the burden of some intended physical application (Adriaans & van Emde Boas 2011). Central is the insight that deterministic programs do not generate new information. Consequently deterministic

computational models of physical systems can never give an account of the growth of information or entropy in nature:

> *Observation*: The Laplacian assumption that the universe can be described as a deterministic computer is, given the fundamental theorem of Adriaans and van Emde Boas (2011) and the assumption that quantum physics as a essentially stochastic description of the structure of our reality, incorrect.

A statistical reduction of thermodynamics to a deterministic theory like Newtonian physics leads to a notion of entropy that is *fundamentally different* from the information processed by deterministic computers. From this perspective the mathematical models of thermodynamics, which are basically differential equations on spaces of real numbers, seem to operate on a level that is not expressive enough. More advanced mathematical models, taking in to account quantum effects, might resolve some of the conceptual difficulties. At a subatomic level nature seems to be inherently probabilistic. If probabilistic quantum effects play a role in the behavior of real billiard balls, then the debate whether entropy increases in an abstract gas, made out of ideal balls, seems a bit academic. There is reason to assume that stochastic phenomena at quantum level are a source of probability at a macroscopic scale (Albrecht & Phillips 2014). From this perspective the universe is a constant source of, literally, astronomical amounts of information at any scale.

6.6 Logic and Semantic Information

Logical and computational approaches to the understanding of information both have their roots in the "linguistic turn" that characterized the philosophical research in the beginning of the twentieth century and the elementary research questions originate from the work of Frege (1879, 1892, see the entry on logic and information). The ambition to quantify information in *sets of true sentences*, as apparent in the work of researchers like Popper, Carnap, Solomonoff, Kolmogorov, Chaitin, Rissanen, Koppel, Schmidthuber, Li, Vitányi and Hutter is an inherently semantic research program. In fact, Shannon's theory of information is the only modern approach that explicitly claims to be non-semantic. More recent quantitative information measures like Kolmogorov complexity (with its ambition to codify all scientific knowledge in terms of a universal distribution) and quantum information (with its concept of *observation* of physical systems) inherently assume a semantic component. At the same time it is possible to develop quantitative versions of semantic theories (see entry on semantic conceptions of information).

The central intuition of algorithmic complexity theory that an intension or meaning of an object can be a computation, was originally formulated by Frege (1879, 1892). The expressions "1 + 4" and "2 + 3" have the same extension (*Bedeutung*) "5", but a different intension (*Sinn*). In this sense one mathematical object can have an infinity of different meanings. There are opaque contexts in which such a distinction is necessary. Consider the sentence "John knows that $\log_2 2^2 = 2$". Clearly the fact that $\log_2 2^2$ represents a specific computation is relevant here. The sentence "John knows that $2 = 2$" seems to have a different meaning.

Dunn (2001, 2008) has pointed out that the analysis of information in logic is intricately related to the notions of intension and extension. The distinction between intension and extension is already anticipated in the *Port Royal Logic* (1662) and the writings of Mill (1843), Boole (1847) and Peirce (1868) but was systematically introduced in logic by Frege (1879, 1892). In a modern sense the extension of a predicate, say "*X* is a bachelor", is simply the set of bachelors in our domain. The intension is associated with the meaning of the predicate and allows us to derive from the fact that "John is a bachelor" the facts that "John is male" and "John is unmarried". It is clear that this phenomenon has a relation with both the

possible world interpretation of modal operators and the notion of information. A bachelor is by necessity also male, i.e., in every possible world in which John is a bachelor he is also male, consequently: If someone gives me the information that John is a bachelor I get the information that he is male and unmarried for free.

The possible world interpretation of modal operators (Kripke 1959) is related to the notion of "state description" introduced by Carnap (1947). A state description is a conjunction that contains exactly one of each atomic sentence or its negation (see section 4.3). The ambition to define a good probability measure for state descriptions was one of the motivations for Solomonoff (1960, 1997) to develop algorithmic information theory. From this perspective Kolmogorov complexity, with its separation of data types (programs, data, machines) and its focus on true sentences describing effects of processes is basically a semantic theory. This is immediately clear if we evaluate the expression:

$$U_j(\overline{T_i}x) = y$$

As is explained in section 5.2.1 the expression $U_j(\overline{T_i}x)$ denotes the result of the emulation of the computation $T_i(x)$ by U_j after reading the self-delimiting description $\overline{T_i}$ of machine T_j. This expression can be interpreted as a piece of *semantic information* in the context of the *informational map* (See entry on semantic conceptions of information) as follows:

- The universal Turing machine U_j is a **context** is which the computation takes place. It can be interpreted as a **possible computational world** in a modal interpretation of computational semantics.
- The sequences of symbols $\overline{T_i}x$ and y are **well-formed data**.
- The sequence $\overline{T_i}$ is a self-delimiting **description of a program** and it can be interpreted as a piece of well-formed **instructional data**.

- The sequence $\overline{T_i}x$ is an **intension**. The sequence y is the corresponding **extension**.
- The expression $U_j(\overline{T_i}x) = y$ states the result of the program $\overline{T_i}x$ in world U_j is y. It is a **true sentence**.

The logical structure of the sentence $U_j(\overline{T_i}x) = y$ is comparable to a true sentence like:

> In the context of empirical observations on planet earth, the bright star you can see in the morning in the eastern sky is Venus

Mutatis mutandis one could develop the following interpretation: U_j can be seen as a context that, for instance, codifies a *bias* for scientific observations on earth, y is the *extension* Venus, $\overline{T_i}x$ is the *intension* "the bright star you can see in the morning in the eastern sky". The intension consists of T_i, which can be interpreted as some general astronomical observation routine (e.g., instructional data), and x provides the well-formed data that tells one where to look (bright star in the morning in the eastern sky).

This suggests a possible unification between more truth oriented theories of information and computational approaches in terms of the informational map presented in the entry of semantic conceptions of information. We delineate some research questions:

- What is a good logical system (or set of systems) that formalizes our intuitions of the relation between concepts like "knowing", "believing" and "being informed of". There are proposals by: Dretske (1981), van Benthem (2006; van Benthem & de Rooij 2003), Floridi (2003, 2011) and others. A careful mapping of these concepts onto our current landscape of known logics (structural, modal) might clarify the strengths and weaknesses of different proposals.

- It is unclear what the *specific difference* (in the Aristotelian sense) is that separates *environmental data* from other data, e.g., if one uses pebbles on a beach to count the number of dolphins one has observed, then it might be impossible for the uninformed passer by to judge whether this collection of stones is environmental data or not.
- The category of *instructional data* seems to be too narrow since it pins us down on a specific interpretation of what computing is. For the most part Turing equivalent computational paradigms are not instructional, although one might defend the view that programs for Turing machines are such data.
- It is unclear how we can cope with the *ontological duality* that is inherent to the self referential aspects of Turing complete systems: Turing machines operate on data that *at the same time* act as representations of programs, i.e., instructional *and* non-instructional.
- It is unclear how a theory that defines information exclusively in terms of true statements can deal with fundamental issues in quantum physics. How can an inconsistent logical model in which Schrodinger's cat is at the same time dead and alive contain any information in such a theory?

6.7 Meaning and Computation

Ever since Descartes, the idea that the meaningful world, we perceive around us, can be reduced to physical processes has been a predominant theme in western philosophy. The corresponding philosophical self-reflection in history neatly follows the technical developments from: Is the human mind an automaton, to is the mind a Turing machine and, eventually, is the mind a quantum computer? It is not the place here to discuss these matters extensively, but the corresponding problem in philosophy of information is relevant:

Open problem: Can meaning be reduced to computation?

The question is interwoven with more general issues in philosophy and its answer directly forces a choice between a more *positivistic* or a more *hermeneutical* approach to philosophy, with consequences for theory of knowledge, metaphysics, aesthetics and ethics. It also effects direct practical decisions we take on a daily basis. Should the actions of a medical doctor be guided by evidence based medicine or by the notion of *caritas*? Is a patient a conscious human being that wants to lead a meaningful life, or is he ultimately just a system that needs to be repaired?

The idea that meaning is essentially a computational phenomenon may seem extreme, but here are many discussions and theories in science, philosophy and culture that implicitly assume such a view. In popular culture, e.g., there is a remarkable collection of movies and books in which we find evil computers that are conscious of themselves (2001, *A Space Odyssey*), individuals that upload their consciousness to a computer (1992, *The Lawnmower Man*), and fight battles in virtual realities (1999, *The Matrix*). In philosophy the position of Bostrom (2003), who defends the view that it is very likely that we already live in a computer simulation, is illustrative. There are many ways to argue the pros and cons of the reduction of meaning to computation. We give an overview of possible arguments for the two extreme positions:

- *Meaning is an emergent aspect of computation*: Science is our best effort to develop a valid objective theoretical description of the universe based on intersubjectively verifiable repeated observations. Science tells us that our reality at a small scale consists of elementary particles whose behavior is described by exact mathematical models. At an elementary level these particles interact and exchange information. These processes are essentially computational. At this most basic level of description there is no room for a subjective notion of meaning. There is no reason to deny that we as human being experience a meaningful world, but as such this must be an

emergent aspect of nature. At a fundamental level it does not exist. We can describe our universe as a big quantum computer. We can estimate the information storage content of our universe to be 10^{92} bits and the number of computational steps it made since the big bang as 10^{123} (Lloyd 2000; Lloyd & Ng 2004). As human beings we are just subsystems of the universe with an estimated complexity of roughly 10^{30} bits. It might be technically impossible, but there seems to be no theoretical objection against the idea that we can in principle construct an exact copy of a human being, either as a direct physical copy or as a simulation in a computer. Such an "artificial" person will experience a meaningful world, but the experience will be emergent.

- *Meaning is ontologically rooted in our individual experience of the world and thus irreducible*: The reason scientific theories eliminate most semantic aspects of our world, is caused by the very nature of methodology of science itself. The essence of meaning and the associated emotions is that they are rooted in our individual experience of the world. By focusing on repeated observations of similar events by different observers scientific methodology excludes the possibility of an analysis of the concept of meaning *a priori*. Empirical scientific methodology is valuable in the sense that it allows us to abstract from the individual differences of conscious observers, but there is no reason to reduce our ontology to the phenomena studied by empirical science. Isolated individual events and observations are by definition not open to experimental analysis and this seems to be the point of demarcation between science and the humanities. In disciplines like history, literature, visual art and ethics we predominantly analyze individual events and individual objects. The closer these are to our individual existence, the more meaning they have for us. There is no reason to doubt the fact that sentences like "Guernica is a masterpiece that shows the atrocities of war" or "McEnroe played such an inspired match that he deserved to

win" uttered in the right context convey meaningful information. The view that this information content ultimately should be understood in terms of computational processes seems too extreme to be viable.

Apart from that, a discipline like physics, that until recently overlooked about 68% of the energy in the universe and 27% of the matter, that has no unified theory of elementary forces and only explains the fundamental aspects of our world in terms of mathematical models that lack any intuitive foundation, for the moment does not seem to converge to a model that could be an adequate basis for a reductionistic metaphysics.

As soon as one defines information in terms of true statements, some meanings become computational and others lack that feature. In the context of empirical science we can study groups of researchers that aim at the construction of theories generalizing structural information in data sets of repeated observations. Such processes of theory construction and *intersubjective verification and falsification* have an inherent computational component. In fact, this notion of intersubjective verification seems an essential element of mathematics. This is the main cause of the fact that central questions of humanities are not open for quantitative analysis: We can disagree on the question whether one painting is more beautiful than the other, but not on the fact that there are two paintings.

It is clear that computation as a conceptual model pays a role in many scientific disciplines varying from cognition (Chater & Vitányi 2003), to biology (see entry on biological information) and physics (Lloyd & Ng 2004; Verlinde 2011, 2017). Extracting meaningful models out of data sets by means of computation is the driving force behind the Big Data revolution (Adriaans & Zantinge 1996; Bell, Hey, & Szalay 2009; Hey, Tansley, & Tolle 2009). Everything that multinationals like Google and Facebook "know" about individuals is extracted from large data bases by

means of computational processes, and it cannot be denied that this kind of "knowledge" has a considerable amount of impact on society. The research question "How can we construct meaningful data out of large data sets by means of computation?" is a fundamental meta-problem of science in the twenty-first century and as such part of philosophy of information, but there is no strict necessity for a reductionistic view.

7. Conclusion

The first domain that could benefit from philosophy of information is of course philosophy itself. The concept of information potentially has an impact on almost all philosophical main disciplines, ranging from logic, theory of knowledge, to ontology and even ethics and esthetics (see introduction above). Philosophy of science and philosophy of information, with their interest in the problem of induction and theory formation, probably both could benefit from closer cooperation (see 4.1 Popper: Information as degree of falsifiability). The concept of information plays an important role in the history of philosophy that is not completely understood (see 2. History of the term and the concept of information).

As information has become a central issue in almost all of the sciences and humanities this development will also impact philosophical reflection in these areas. Archaeologists, linguists, physicists, astronomers all deal with information. The first thing a scientist has to do before he can formulate a theory is gathering information. The application possibilities are abundant. Datamining and the handling of extremely large data sets seems to be an essential for almost every empirical discipline in the twenty-first century.

In biology we have found out that information is essential for the organization of life itself and for the propagation of complex organisms (see entry on biological information). One of the main problems is that current models do not explain the complexity of life well. Valiant has started a research program that studies evolution as a form of computational learning (Valiant 2009) in order to explain this discrepancy. Aaronson (2013) has argued explicitly for a closer cooperation between complexity theory and philosophy.

Until recently the general opinion was that the various notions of information were more or less isolated but in recent years considerable progress has been made in the understanding of the relationship between these concepts. Cover and Thomas (2006), for instance, see a perfect match between Kolmogorov complexity and Shannon information. Similar observations have been made by Grünwald and Vitányi (2008). Also the connections that exist between the theory of thermodynamics and information theory have been studied (Bais & Farmer 2008; Harremoës & Topsøe 2008) and it is clear that the connections between physics and information theory are much more elaborate than a mere *ad hoc* similarity between the formal treatment of entropy and information suggests (Gell-Mann & Lloyd 2003; Verlinde (2011, 2017). Quantum computing is at this moment not developed to a point where it is effectively more powerful than classical computing, but this threshold might be passed in the coming years. From the point of view of philosophy many conceptual problems of quantum physics and information theory seem to merge into one field of related questions:

- What is the relation between information and computation?
- Is computation in the real world fundamentally non-deterministic?
- What is the relation between symbol manipulation on a macroscopic scale and the world of quantum physics?
- What is a good model of quantum computing and how do we control its power?
- Is there information beyond the world of quanta?

The notion of information has become central in both our society and in the sciences. Information technology plays a pivotal role in the way we organize our lives. It also has become a basic category in the sciences and the humanities. Philosophy of information, both as a historical and a systematic discipline, offers a new perspective on old philosophical problems and also suggests new research domains.

Bibliography

Aaronson, Scott, 2013, "Why Philosophers Should Care About Computational Complexity", in *Computability: Turing, Gödel, Church, and Beyond*, Brian Jack Copeland, Carl J. Posy, and Oron Shagrir (eds.), Cambridge, MA: The MIT Press. [Aaronson 2013 preprint available online]

Abramsky, Samson and Achim Jung, 1994, "Domain theory", in *Handbook of Logic in Computer Science (vol. 3): Semantic Structure*, Samson Abramsky, Dov M. Gabbay, and Thomas S. E. Maibaum (eds.),. Oxford University Press. pp. 1–168.

Adams, Fred and João Antonio de Moraes, 2016, "Is There a Philosophy of Information?", *Topoi*, 35(1): 161–171. doi:10.1007/s11245-014-9252-9

Adriaans, Pieter, 2007, "Learning as Data Compression", in *Computation and Logic in the Real World*, S. Barry Cooper, Benedikt Löwe, and Andrea Sorbi (eds.), (Lecture Notes in Computer Science: Volume 4497), Berlin, Heidelberg: Springer Berlin Heidelberg, 11–24. doi:10.1007/978-3-540-73001-9_2

——, 2008, "Between Order and Chaos: The Quest for Meaningful Information", *Theory of Computing Systems* (Special Issue: Computation and Logic in the Real World; Guest Editors: S. Barry Cooper, Elvira Mayordomo and Andrea Sorbi), 45(4): 650–674. doi:10.1007/s00224-009-9173-y

Adriaans, Pieter and Peter van Emde Boas, 2011, "Computation, Information, and the Arrow of Time", in *Computability in Context: Computation and Logic in the Real World*, by S Barry Cooper and Andrea Sorbi (eds), London: Imperial College Press, 1–17. doi:10.1142/9781848162778_0001

Adriaans, Pieter and Johan van Benthem, 2008a, "Introduction: Information Is What Information Does", in Adriaans & van Benthem 2008b: 3–26. doi:10.1016/B978-0-444-51726-5.50006-6

—— (eds.), 2008b, *Philosophy of Information*, (Handbook of the Philosophy of Science 8), Amsterdam: Elsevier. doi:10.1016/C2009-0-16481-4

Adriaans, Pieter and Paul M.B. Vitányi, 2009, "Approximation of the Two-Part MDL Code", *IEEE Transactions on Information Theory*, 55(1): 444–457. doi:10.1109/TIT.2008.2008152

Adriaans, Pieter and Dolf Zantinge, 1996, *Data Mining*, Harlow, England: Addison-Wesley.

Agrawal, Manindra, Neeraj Kayal, and Nitin Saxena, 2004, "PRIMES Is in P", *Annals of Mathematics*, 160(2): 781–793. doi:10.4007/annals.2004.160.781

Albrecht, Andreas and Daniel Phillips, 2014, "Origin of Probabilities and Their Application to the Multiverse", *Physical Review D*, 90(12): 123514. doi:10.1103/PhysRevD.90. 123514

Antunes, Luís and Lance Fortnow, 2003, "Sophistication Revisited", in *Proceedings of the 30th International Colloquium on Automata, Languages and Programming* (Lecture Notes in Computer Science: Volume 2719), Jos C. M. Baeten, Jan Karel Lenstra, Joachim Parrow, and Gerhard J. Woeginger (eds.), Berlin: Springer, pp. 267–277. doi:10.1007/3-540-45061-0_23

Antunes, Luis, Lance Fortnow, Dieter van Melkebeek, and N.V. Vinodchandran, 2006, "Computational Depth: Concept and Applications", *Theoretical Computer Science*, 354(3): 391–404.

doi:10.1016/j.tcs.2005.11.033

Aquinas, St. Thomas, 1265–1274, *Summa Theologiae*.

Arbuthnot, John, 1692, *Of the Laws of Chance, or, a method of Calculation of the Hazards of Game, Plainly demonstrated, And applied to Games as present most in Use*, translation of Huygens' *De Ratiociniis in Ludo Aleae*, 1657.

Aristotle. *Aristotle in 23 Volumes*, Vols. 17, 18, translated by Hugh Tredennick, Cambridge, MA: Harvard University Press; London, William Heinemann Ltd. 1933, 1989.

Austen, Jane, 1815, *Emma*, London: Richard Bentley and Son.

Bar-Hillel, Yehoshua and Rudolf Carnap, 1953, "Semantic Information", *The British Journal for the Philosophy of Science*, 4(14): 147–157. doi:10.1093/bjps/IV.14.147

Bais, F. Alexander and J. Doyne Farmer, 2008, "The Physics of Information", Adriaans and van Benthem 2008b: 609–683. doi:10.1016/B978-0-444-51726-5.50020-0

Barron, Andrew, Jorma Rissanen, and Bin Yu, 1998, "The Minimum Description Length Principle in Coding and Modeling", *IEEE Transactions on Information Theory*, 44(6): 2743–2760. doi:10.1109/18.720554

Barwise, Jon and John Perry, 1983, *Situations and Attitudes*, Cambridge, MA: MIT Press.

Bell, Gordon, Tony Hey, and Alex Szalay, 2009, "Computer Science: Beyond the Data Deluge", *Science*, 323(5919): 1297–1298. doi:10.1126/science.1170411

Bennett, C. H., 1988, "Logical Depth and Physical Complexity", in Rolf Herken (ed.), *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press, pp. 227–257.

Berkeley, George, 1732, *Alciphron: Or the Minute Philosopher*, Edinburgh: Thomas Nelson, 1948–57.

Bernoulli, Danielis, 1738, *Hydrodynamica*, Argentorati: sumptibus Johannis Reinholdi Dulseckeri. [Bernoulli 1738 available online]

Birkhoff, George David, 1950, *Collected Mathematical Papers*, New York: American Mathematical Society.

Bloem, Peter, Steven de Rooij, and Pieter Adriaans, 2015, "Two Problems for Sophistication", in *Algorithmic Learning Theory*, (Lecture Notes in Computer Science 9355), Kamalika Chaudhuri, Claudio Gentile, and Sandra Zilles (eds.), Cham: Springer International Publishing, 379–394. doi:10.1007/978-3-319-24486-0_25

Boltzmann, Ludwig, 1866, "Über die Mechanische Bedeutung des Zweiten Hauptsatzes der Wärmetheorie", *Wiener Berichte*, 53: 195–220.

Boole, George, 1847, *Mathematical Analysis of Logic: Being an Essay towards a Calculus of Deductive Reasoning*, Cambridge: Macmillan, Barclay, & Macmillan. [Boole 1847 available online].

–––, 1854, *An Investigation of the Laws of Thought: On which are Founded the Mathematical Theories of Logic and Probabilities*, London: Walton and Maberly.

Bostrom, Nick, 2003, "Are We Living in a Computer Simulation?", *The Philosophical Quarterly*, 53(211): 243–255. doi:10.1111/1467-9213.00309

Bott, R. and J. Milnor, 1958, "On the Parallelizability of the Spheres", *Bulletin of the American Mathematical Society*, 64(3): 87–89. doi:10.1090/S0002-9904-1958-10166-4

Bovens, Luc and Stephan Hartmann, 2003, *Bayesian Epistemology*, Oxford: Oxford University Press. doi:10.1093/0199269750.001.0001

Brenner, Joseph E., 2008, *Logic in Reality*, Dordrecht: Springer Netherlands. doi:10.1007/978-1-4020-8375-4

Briggs, Henry, 1624, *Arithmetica Logarithmica*, London: Gulielmus Iones.

Capurro, Rafael, 1978, *Information. Ein Beitrag zur etymologischen und ideengeschichtlichen Begründung des Informationsbegriffs* (Information: A contribution to the foundation of the concept of

information based on its etymology and in the history of ideas), Munich, Germany: Saur. [Capurro 1978 available online].

–––, 2009, "Past, Present, and Future of the Concept of Information", *TripleC: Communication, Capitalism & Critique*, 7(2): 125–141. doi:10.31269/triplec.v7i2.113

Capurro, Rafael and Birger Hjørland, 2003, "The Concept of Information", in Blaise Cronin (ed.), *Annual Review of Information Science and Technology (ARIST)*, 37: 343–411 (Chapter 8). doi:10.1002/aris.1440370109

Capurro, Rafael and John Holgate (eds.), 2011, *Messages and Messengers: Angeletics as an Approach to the Phenomenology of Communication* (*Von Boten Und Botschaften*, (Schriftenreihe Des International Center for Information Ethics 5), München: Fink.

Carnap, Rudolf, 1928, *Scheinprobleme in der Philosophie* (Pseudoproblems of Philosophy), Berlin: Weltkreis-Verlag.

–––, 1945, "The Two Concepts of Probability: The Problem of Probability", *Philosophy and Phenomenological Research*, 5(4): 513–532. doi:10.2307/2102817

–––, 1947, *Meaning and Necessity*, Chicago: The University of Chicago Press.

–––, 1950, *Logical Foundations of Probability*, Chicago: The University of Chicago Press.

Chaitin, Gregory J., 1969, "On the Length of Programs for Computing Finite Binary Sequences: Statistical Considerations", *Journal of the ACM*, 16(1): 145–159. doi:10.1145/321495.321506

–––, 1987, *Algorithmic Information Theory*, Cambridge: Cambridge University Press. doi:10.1017/CBO9780511608858

Chater, Nick and Paul Vitányi, 2003, "Simplicity: A Unifying Principle in Cognitive Science?", *Trends in Cognitive Sciences*, 7(1): 19–22. doi:10.1016/S1364-6613(02)00005-0

Church, Alonzo, 1936, "An Unsolvable Problem of Elementary Number

Theory", *American Journal of Mathematics* 58(2): 345–363. doi:10.2307/2371045

Cilibrasi, Rudi and Paul M.B. Vitanyi, 2005, "Clustering by Compression", *IEEE Transactions on Information Theory*, 51(4): 1523–1545. doi:10.1109/TIT.2005.844059

Clausius, R., 1850, "Ueber die bewegende Kraft der Wärme und die Gesetze, welche sich daraus für die Wärmelehre selbst ableiten lassen", *Annalen der Physik und Chemie*, 155(3): 368–397. doi:10.1002/andp.18501550306

Conan Doyle, Arthur, 1892, "The Adventure of the Noble Bachelor", in *The Adventures of Sherlock Holmes*, London: George Newnes Ltd, story 10.

Cover, Thomas M. and Joy A. Thomas, 2006, *Elements of Information Theory*, second edition, New York: John Wiley & Sons.

Crawford, James M. and Larry D. Auton, 1993, "Experimental Results on the Crossover Point in Satisfiability Problems", *Proceedings of the Eleventh National Conference on Artificial Intelligence*, AAAI Press, pp. 21–27. [Crawford & Auton 1993 available online]

Crutchfield, James P. and Karl Young, 1989, "Inferring Statistical Complexity", *Physical Review Letters*, 63(2): 105–108. doi:10.1103/PhysRevLett.63.105

–––, 1990, "Computation at the Onset of Chaos", in *Entropy, Complexity, and the Physics of Information*, W. Zurek, editor, SFI Studies in the Sciences of Complexity, VIII, Reading, MA: Addison-Wesley, pp. 223–269. [Crutchfield & Young 1990 available online]

D'Alfonso, Simon, 2012, "Towards a Framework for Semantic Information", Ph.D. Thesis, Department of Philosophy, School of Historical and Philosophical Studies, The University of Melbourne. D'Alfonso 2012 available online

Davis, Martin, 2006, "Why There Is No Such Discipline as Hypercomputation", *Applied Mathematics and Computation*, 178(1):

4–7. doi:10.1016/j.amc.2005.09.066

Defoe, Daniel, 1719, *The Life and Strange Surprising Adventures of Robinson Crusoe of York, Mariner: who lived Eight and Twenty Years, all alone in an uninhabited Island on the coast of America, near the Mouth of the Great River of Oroonoque; Having been cast on Shore by Shipwreck, wherein all the Men perished but himself. With An Account how he was at last as strangely deliver'd by Pirates. Written by Himself*, London: W. Taylor. [Defoe 1719 available online]

De Leo, Stefano, 1996, "Quaternions and Special Relativity", *Journal of Mathematical Physics*, 37(6): 2955–2968. doi:10.1063/1.531548

Dershowitz, Nachum and Yuri Gurevich, 2008, "A Natural Axiomatization of Computability and Proof of Church's Thesis", *Bulletin of Symbolic Logic*, 14(3): 299–350. doi:10.2178/bsl/1231081370

Descartes, René, 1641, *Meditationes de Prima Philosophia* (Meditations on First Philosophy), Paris.

——, 1647, *Discours de la Méthode* (Discourse on Method), Leiden.

Devlin, Keith and Duska Rosenberg, 2008, "Information in the Study of Human Interaction", Adriaans and van Benthem 2008b: 685–709. doi:10.1016/B978-0-444-51726-5.50021-2

Dictionnaire du Moyen Français (1330–1500), 2015, "Information", in *Dictionnaire du Moyen Français (1330–1500)*, volume 16, 313–315. [Dictionnaire du Moyen Français available online]

Domingos, Pedro, 1998, "Occam's Two Razors: The Sharp and the Blunt", in *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining* (KDD–98), New York: AAAI Press, pp. 37–43. [Domingos 1998 available online]

Downey, Rodney G. and Denis R. Hirschfeldt, 2010, *Algorithmic Randomness and Complexity*, (Theory and Applications of Computability), New York: Springer New York. doi:10.1007/978-0-387-68441-3

Dretske, Fred, 1981, *Knowledge and the Flow of Information*, Cambridge,

MA: The MIT Press.

Dufort, Paul A. and Charles J. Lumsden, 1994, "The Complexity and Entropy of Turing Machines", in *Proceedings Workshop on Physics and Computation*. PhysComp '94, Dallas, TX: IEEE Computer Society Press, 227–232. doi:10.1109/PHYCMP.1994.363677

Dunn, Jon Michael, 2001, "The Concept of Information and the Development of Modern Logic", in *Zwischen traditioneller und moderner Logik: Nichtklassiche Ansatze* (*Non-classical Approaches in the Transition from Traditional to Modern Logic*), Werner Stelzner and Manfred Stöckler (eds.), Paderborn: Mentis, 423–447.

——, 2008, "Information in Computer Science", in Adriaans and van Benthem 2008b: 581–608. doi:10.1016/B978-0-444-51726-5.50019-4

Dijksterhuis, E. J., 1986, *The Mechanization of the World Picture: Pythagoras to Newton*, Princeton, NJ: Princeton University Press.

Duns Scotus, John [1265/66–1308 CE], *Opera Omnia* (The Wadding edition), Luke Wadding (ed.), Lyon, 1639; reprinted Hildesheim: Georg Olms Verlagsbuchhandlung, 1968.

Durand-Lose, Jérôme, 2002, "Computing Inside the Billiard Ball Model", in *Collision-Based Computing*, Andrew Adamatzky (ed.), London: Springer London, 135–160. doi:10.1007/978-1-4471-0129-1_6

Edwards, Paul, 1967, *The Encyclopedia of Philosophy*, 8 volumes, New York: Macmillan Publishing Company.

Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth, 1996, "From Data Mining to Knowledge Discovery in Databases", *AI Magazine*, 17(3): 37–37.

Fisher, R. A., 1925, "Theory of Statistical Estimation", *Mathematical Proceedings of the Cambridge Philosophical Society*, 22(05): 700–725. doi:10.1017/S0305004100009580

Floridi, Luciano, 1999, "Information Ethics: On the Philosophical Foundation of Computer Ethics", *Ethics and Information Technology*,

1(1): 33–52. doi:10.1023/A:1010018611096

——, 2002, "What Is the Philosophy of Information?" *Metaphilosophy*, 33(1–2): 123–145. doi:10.1111/1467-9973.00221

—— (ed.), 2003, *The Blackwell Guide to the Philosophy of Computing and Information*, Oxford: Blackwell. doi:10.1002/9780470757017

——, 2010, "The Philosophy of Information as a Conceptual Framework", *Knowledge, Technology & Policy*, 23(1–2): 253–281. doi:10.1007/s12130-010-9112-x

——, 2011, *The Philosophy of Information*, Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199232383.001.0001

Fredkin, Edward and Tommaso Toffoli, 1982, "Conservative Logic", *International Journal of Theoretical Physics*, 21(3–4): 219–253. doi:10.1007/BF01857727

Frege, Gottlob, 1879, *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle.

——, 1892, "Über Sinn und Bedeutung", *Zeitschrift für Philosophie und philosophische Kritik*, NF 100.

Furey, C., 2015, "Charge Quantization from a Number Operator", *Physics Letters B*, 742(March): 195–199. doi:10.1016/j.physletb.2015.01.023

Galileo Galilei, 1623 [1960], *Il Saggiatore* (in Italian), Rome; translated as *The Assayer*, by Stillman Drake and C. D. O'Malley, in *The Controversy on the Comets of 1618*, Philadelphia: University of Pennsylvania Press, 1960, 151–336.

Garey, Michael R. and David S. Johnson, 1979, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, (A Series of Books in the Mathematical Sciences), San Francisco: W. H. Freeman.

Gell-Mann, Murray and Seth Lloyd, 2003, "Effective Computing". SFI Working Paper 03-12-068, Santa Fe, NM: Santa Fe Institute. [Gell-Mann & Lloyd 2003 available online]

Gibbs, J. Willard, 1906, *The Scientific Papers of J. Willard Gibbs in Two Volumes*, 1. Longmans, Green, and Co.

Godefroy, Frédéric G., 1881, *Dictionnaire de l'ancienne langue française et de tous ses dialectes du 9e au 15e siècle*, Paris: F. Vieweg.

Gödel, Kurt, 1931, "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I", *Monatshefte für Mathematik und Physik*, 38–38(1): 173–198. doi:10.1007/BF01700692

Goodstein, R. L., 1957, "The Definition of Number", *The Mathematical Gazette*, 41(337): 180–186. doi:10.2307/3609188

Grünwald, Peter D., 2007, *The Minimum Description Length Principle*, Cambridge, MA: MIT Press.

Grünwald, Peter D. and Paul M.B. Vitányi, 2008, "Algorithmic Information Theory", in Adriaans and van Benthem 2008b: 281–317. doi:10.1016/B978-0-444-51726-5.50013-3

Groot, Adrianus Dingeman de, 1961 [1969], *Methodology: Foundations of Inference and Research in the Behavioral Sciences* (*Methodologie: grondslagen van onderzoek en denken in de gedragswetenschappen*), The Hague: Mouton.

Harremoës, Peter and Flemming Topsøe, 2008, "The Quantitative Theory of Information", in Adriaans and van Benthem 2008b: 171–216. doi:10.1016/B978-0-444-51726-5.50011-X

Hartley, R.V.L., 1928, "Transmission of Information", *Bell System Technical Journal*, 7(3): 535–563. doi:10.1002/j.1538-7305.1928.tb01236.x

Hazard, Paul, 1935, *La Crise de La Conscience Européenne (1680–1715)*, Paris: Boivin.

Hey, Anthony J. G., Stewart Tansley, and Kristin Tolle (eds.), 2009, *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Redmond, WA: Microsoft Research. [Hey et al. 2009 available online]

Hintikka, Jaakko, 1962, *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, (Contemporary Philosophy), Ithaca, NY:

Cornell University Press.

——, 1973, *Logic, Language Games and Information: Kantian Themes in the Philosophy of Logic*, Oxford: Clarendon Press.

Hume, David, 1739–40, *A Treatise of Human Nature*. Reprinted, L.A. Selby-Bigge (ed.), Oxford: Clarendon Press, 1896. [Hume 1739–40 [1896] available online]

——, 1748, *An Enquiry concerning Human Understanding*. Reprinted in *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, 1777 which was reprinted, L.A. Selby-Bigge (ed.), Oxford: Clarendon Press, 1888 (second edition 1902). [Hume 1748 [1902] available online]

Hutter, Marcus, 2005, *Universal Artificial Intellegence: Sequential Decisions Based on Algorithmic Probability*, (Texts in Theoretical Computer Science, an EATCS Series), Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/b138233

——, 2007a, "On Universal Prediction and Bayesian Confirmation", *Theoretical Computer Science*, 384(1): 33–48. doi:10.1016/j.tcs.2007.05.016

——, 2007b, "Algorithmic Information Theory: a brief non-technical guide to the field", *Scholarpedia*, 2(3): art. 2519. doi:10.4249/scholarpedia.2519

——, 2010, "A Complete Theory of Everything (will be subjective)", *Algorithms*, 3(4): 329–350. doi:10.3390/a3040329

Hutter, Marcus, John W. Lloyd, Kee Siong Ng, and William T.B. Uther, 2013, "Probabilities on Sentences in an Expressive Logic", *Journal of Applied Logic*, special issue: *Combining Probability and Logic: Papers from Progic 2011*, Jeffrey Helzner (ed.), 11(4): 386–420. doi:10.1016/j.jal.2013.03.003.

Ibn Tufail, *Hayy ibn Yaqdhan*, translated as *Philosophus Autodidactus*, published by Edward Pococke the Younger in 1671.

Kahn, David, 1967, *The Code-Breakers, The Comprehensive History of Secret Communication from Ancient Times to the Internet*, New York: Scribner.

Kant, Immanuel, 1781, *Kritik der reinen Vernunft* (Critique of Pure Reason), Germany.

Kervaire, Michel A., 1958, "Non-Parallelizability of the n-Sphere for n > 7", *Proceedings of the National Academy of Sciences of the United States of America*, 44(3): 280–283. doi:10.1073/pnas.44.3.280

al-Khwārizmī, Muḥammad ibn Mūsā, ca. 820 CE, *Hisab al-jabr w'al-muqabala, Kitab al-Jabr wa-l-Muqabala* (*The Compendious Book on Calculation by Completion and Balancing*), Translated by Frederic Rosen, London: Murray, 1831. [al-Khwarizmi translation available online]

Kolmogorov, A.N., 1965, "Three Approaches to the Quantitative Definition of Information", *Problems of Information Transmission*, 1(1): 1–7. Reprinted 1968 in *International Journal of Computer Mathematics*, 2(1–4): 157–168. doi:10.1080/00207166808803030

Koppel, Moshe, 1987, "Complexity, Depth, and Sophistication", *Complex Systems*, 1(6): 1087–1091. [Koppel 1987 available online]

Kripke, Saul A., 1959, "A Completeness Theorem in Modal Logic", *The Journal of Symbolic Logic*, 24(1): 1–14. doi:10.2307/2964568

——, 1971, "Identity and Necessity", in Milton K. Munitz (ed.), *Identity and Individuation*, New York: New York University Press, pp. 135-164.

Kuipers, Theo A.F. (ed.), 2007a, *General Philosophy of Science: Focal Issues*, Amsterdam: Elsevier Science Publishers.

——, 2007b, "Explanation in Philosophy of Science", in Kuipers 2007a.

Laplace, Pierre Simon, Marquis de, 1814 [1902], *A Philosophical Essay on Probabilities*, F.W. Truscott and F.L. Emory (trans.), New York: J. Wiley; London: Chapman & Hall.

Langton, Chris G., 1990, "Computation at the Edge of Chaos: Phase Transitions and Emergent Computation", *Physica D: Nonlinear*

Phenomena, 42(1–3): 12–37. doi:10.1016/0167-2789(90)90064-V

Lenski, Wolfgang, 2010, "Information: A Conceptual Investigation", *Information 2010*, 1(2): 74–118. doi:10.3390/info1020074

Levin, Leonid A., 1973, "Universal Sequential Search Problems", *Problems of Information Transmission*, 9(3): 265–266.

––––,1974, "Laws of Information Conservation (Non-Growth) and Aspects of the Foundation of Probability Theory", *Problems of Information Transmission*, 10(3): 206–210.

––––, 1984, "Randomness Conservation Inequalities; Information and Independence in Mathematical Theories", *Information and Control*, 61(1): 15–37. doi:10.1016/S0019-9958(84)80060-1

Li, Ming and Paul Vitányi, 2008, *An Introduction to Kolmogorov Complexity and Its Applications*, (Texts in Computer Science), New York: Springer New York. doi:10.1007/978-0-387-49820-1

Lloyd, Seth, 2000, "Ultimate Physical Limits to Computation", *Nature*, 406(6799): 1047–1054. doi:10.1038/35023282

Lloyd, Seth and Y. Jack Ng, 2004, "Black Hole Computers", *Scientific American*, 291(5): 52–61. doi:10.1038/scientificamerican1104-52

Locke, John, 1689, *An Essay Concerning Human Understanding*, J. W. Yolton (ed.), London: Dent; New York: Dutton, 1961. [Locke 1689 available online]

McAllister, James W., 2003, "Effective Complexity as a Measure of Information Content", *Philosophy of Science*, 70(2): 302–307. doi:10.1086/375469

Mill, John Stuart, 1843, *A System of Logic*, London.

Montague, Richard, 2008, "Universal Grammar", *Theoria*, 36(3): 373–398. doi:10.1111/j.1755-2567.1970.tb00434.x

Mugur-Schächter, Mioara, 2003, "Quantum Mechanics Versus a Method of Relativized Conceptualization", in *Quantum Mechanics, Mathematics, Cognition and Action*, Mioara Mugur-Schächter and Alwyn van der Merwe (eds.), Dordrecht: Springer Netherlands, 109–

307. doi:10.1007/0-306-48144-8_7

Napier, John, 1614, *Mirifici Logarithmorum Canonis Descriptio* (*The Description of the Wonderful Canon of Logarithms*), Edinburgh: Andre Hart. Translated and annotated by Ian Bruce, www.17centurymaths.com. [Napier 1614 [Bruce translation] available online].

Nielsen, Michael A. and Isaac L. Chuang, 2000, *Quantum Computation and Quantum Information*, Cambridge: Cambridge University Press.

Nies, André, 2009, *Computability and Randomness*, Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199230761.001.0001

Nyquist, H., 1924, "Certain Factors Affecting Telegraph Speed", *Bell System Technical Journal*, 3(2): 324–346. doi:10.1002/j.1538-7305.1924.tb01361.x

Ong, Walter J., 1958, *Ramus, Method, and the Decay of Dialogue, From the Art of Discourse to the Art of Reason*, Cambridge MA: Harvard University Press.

Parikh, Rohit and Ramaswamy Ramanujam, 2003, "A Knowledge Based Semantics of Messages", *Journal of Logic, Language and Information*, 12(4): 453–467. doi:10.1023/A:1025007018583

Peirce, Charles S., 1868, "Upon Logical Comprehension and Extension", *Proceedings of the American Academy of Arts and Sciences*, 7: 416–432. doi:10.2307/20179572

––––, 1886 [1993], " Letter Peirce to A. Marquand", Reprinted in *Writings of Charles S. Peirce: A Chronological Edition, Volume 5: 1884–1886*, Indianapolis: Indiana University Press, pp. 424–427. See also Arthur W. Burks, 1978, "Book Review: 'The New Elements of Mathematics' by Charles S. Peirce, Carolyn Eisele (editor)", *Bulletin of the American Mathematical Society*, 84(5): 913–919. doi:10.1090/S0002-9904-1978-14533-9

Popper, Karl, 1934, *The Logic of Scientific Discovery*, (*Logik der Forschung*), English translation 1959, London: Hutchison. Reprinted

1977.

Putnam, Hilary, 1988, *Representation and reality*, Cambridge, MA: The MIT Press.

Quine, W.V.O., 1951a, "Two Dogmas of Empiricism", *The Philosophical Review*, 60: 20–43.

––––, 1951b, "Main Trends in Recent Philosophy: Two Dogmas of Empiricism", *The Philosophical Review*, 60(1): 20–43. Reprinted in his 1953 *From a Logical Point of View*, Cambridge, MA: Harvard University Press. doi:10.2307/2181906

Rathmanner, Samuel and Marcus Hutter, 2011, "A Philosophical Treatise of Universal Induction", *Entropy*, 13(6): 1076–1136. doi:10.3390/e13061076

Rédei, Miklós and Michael Stöltzner (eds.), 2001, *John von Neumann and the Foundations of Quantum Physics*, (Vienna Circle Institute Yearbook, 8), Dordrecht: Kluwer.

Rényi, Alfréd, 1961, "On Measures of Entropy and Information", in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, Berkeley, CA: The Regents of the University of California, pp. 547–561. [Rényi 1961 available online]

Rissanen, J., 1978, "Modeling by Shortest Data Description", *Automatica*, 14(5): 465–471. doi:10.1016/0005-1098(78)90005-5

––––, 1989, *Stochastic Complexity in Statistical Inquiry*, (World Scientific Series in Computer Science, 15), Singapore: World Scientific.

Rooy, Robert van, 2004, "Signalling Games Select Horn Strategies", *Linguistics and Philosophy*, 27(4): 493–527. doi:10.1023/B:LING.0000024403.88733.3f

Russell, Bertrand, 1905, "On Denoting", *Mind*, new series, 14(4): 479–493. doi:10.1093/mind/XIV.4.479

Schmandt-Besserat, Denise, 1992, *Before Writing* (Volume I: From Counting to Cuneiform), Austin, TX: University of Texas Press.

Schmidhuber, Jüurgen, 1997a, "Low-Complexity Art", *Leonardo*, 30(2): 97–103. doi:10.2307/1576418

––––, 1997b, "A Computer Scientist's View of Life, the Universe, and Everything", in *Foundations of Computer Science*, (Lecture Notes in Computer Science, 1337), Christian Freksa, Matthias Jantzen, and Rüdiger Valk (eds.), Berlin, Heidelberg: Springer Berlin Heidelberg, 201–208. doi:10.1007/BFb0052088

Schnelle, H., 1976, "Information", in Joachim Ritter (ed.), *Historisches Wörterbuch der Philosophie*, IV [Historical dictionary of philosophy, IV] (pp. 116–117). Stuttgart, Germany: Schwabe.

Searle, John R., 1990, "Is the Brain a Digital Computer?", *Proceedings and Addresses of the American Philosophical Association*, 64(3): 21–37. doi:10.2307/3130074

Seiffert, Helmut, 1968, *Information über die Information* [Information about information] Munich: Beck.

Shannon, Claude E., 1948, "A Mathematical Theory of Communication", *Bell System Technical Journal*, 27(3): 379–423 & 27(4): 623–656. doi:10.1002/j.1538-7305.1948.tb01338.x & doi:10.1002/j.1538-7305.1948.tb00917.x

Shannon, Claude E. and Warren Weaver, 1949, *The Mathematical Theory of Communication*, Urbana, IL: University of Illinois Press.

Shor, Peter W., 1997, "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer", *SIAM Journal on Computing*, 26(5): 1484–1509. doi:10.1137/S0097539795293172

Simon, J.C. and Olivier Dubois, 1989, "Number of Solutions of Satisfiability Instances – Applications to Knowledge Bases", *International Journal of Pattern Recognition and Artificial Intelligence*, 3(1): 53–65. doi:10.1142/S0218001489000061

Simondon, Gilbert, 1989, *L'individuation Psychique et Collective: À La Lumière des Notions de Forme, Information, Potentiel et*

Métastabilité (L'Invention Philosophique), Paris: Aubier.

Singh, Simon, 1999, *The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography*, New York: Anchor Books.

Solomonoff, R. J., 1960, "A Preliminary Report on a General Theory of Inductive Inference". Report ZTB-138, Cambridge, MA: Zator. [Solomonoff 1960 available online]

–––, 1964a, "A Formal Theory of Inductive Inference. Part I", *Information and Control*, 7(1): 1–22. doi:10.1016/S0019-9958(64)90223-2

–––, 1964b, "A Formal Theory of Inductive Inference. Part II", *Information and Control*, 7(2): 224–254. doi:10.1016/S0019-9958(64)90131-7

–––, 1997, "The Discovery of Algorithmic Probability", *Journal of Computer and System Sciences*, 55(1): 73–88. doi:10.1006/jcss.1997.1500

Stalnaker, Richard, 1984, *Inquiry*, Cambridge, MA: MIT Press.

Stifel, Michael, 1544, *Arithmetica integra*, Nuremberg: Johan Petreium.

Tarski, Alfred, 1944, "The Semantic Conception of Truth: And the Foundations of Semantics", *Philosophy and Phenomenological Research*, 4(3): 341–376. doi:10.2307/2102968

Tsallis, Constantino, 1988, "Possible Generalization of Boltzmann-Gibbs Statistics", *Journal of Statistical Physics*, 52(1–2): 479–487. doi:10.1007/BF01016429

Turing, A. M., 1937, "On Computable Numbers, with an Application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society*, s2-42(1): 230–265. doi:10.1112/plms/s2-42.1.230

Valiant, Leslie G., 2009, "Evolvability", *Journal of the ACM*, 56(1): Article 3. doi:10.1145/1462153.1462156

van Benthem, Johan F.A.K., 1990, "Kunstmatige Intelligentie: Een Voortzetting van de Filosofie met Andere Middelen", *Algemeen Nederlands Tijdschrift voor Wijsbegeerte*, 82: 83–100.

–––, 2006, "Epistemic Logic and Epistemology: The State of Their Affairs", *Philosophical Studies*, 128(1): 49–76. doi:10.1007/s11098-005-4052-0

van Benthem, Johan and Robert van Rooy, 2003, "Connecting the Different Faces of Information", *Journal of Logic, Language and Information*, 12(4): 375–379. doi:10.1023/A:1025026116766

van Peursen, Cornelis Anthonie, 1987, "Christian Wolff's Philosophy of Contingent Reality", *Journal of the History of Philosophy*, 25(1): 69–82. doi:10.1353/hph.1987.0005

Vereshchagin, Nikolai K. and Paul M.B. Vitányi, 2004, "Kolmogorov's Structure Functions and Model Selection", *IEEE Transactions on Information Theory*, 50(12): 3265–3290. doi:10.1109/TIT.2004.838346

Verlinde, Erik, 2011, "On the Origin of Gravity and the Laws of Newton", *Journal of High Energy Physics*, 2011(4). doi:10.1007/JHEP04(2011)029

–––, 2017, "Emergent Gravity and the Dark Universe", *SciPost Physics*, 2(3): 016. doi:10.21468/SciPostPhys.2.3.016

Vigo, Ronaldo, 2011, "Representational Information: A New General Notion and Measure of Information", *Information Sciences*, 181(21): 4847–4859. doi:10.1016/j.ins.2011.05.020

–––, 2012, "Complexity over Uncertainty in Generalized Representational Information Theory (GRIT): A Structure-Sensitive General Theory of Information", *Information*, 4(1): 1–30. doi:10.3390/info4010001

Vitányi, Paul M., 2006, "Meaningful Information", *IEEE Transactions on Information Theory*, 52(10): 4617–4626. doi:10.1109/TIT.2006.881729 [Vitányi 2006 available online].

Vogel, Cornelia Johanna de, 1968, *Plato: De filosoof van het transcendente*, Baarn: Het Wereldvenster.

Von Neumann, John, 1932, *Mathematische Grundlagen der Quantenmechanik*, Berlin: Springer.

Wallace, C. S., 2005, *Statistical and Inductive Inference by Minimum Message Length*, Berlin: Springer. doi:10.1007/0-387-27656-4

Wheeler, John Archibald, 1990, "Information, Physics, Quantum: The Search for Links", in *Complexity, Entropy and the Physics of Information*, Wojciech H. Zurek (ed.), Boulder, CO: Westview Press, 309–336. [Wheeler 1990 available online]

Whitehead, Alfred and Bertrand Russell, 1910, 1912, 1913, *Principia Mathematica*, 3 vols, Cambridge: Cambridge University Press; 2nd edn, 1925 (Vol. 1), 1927 (Vols 2, 3).

Wilkins, John, 1668, "An Essay towards a Real Character, and a Philosophical Language", London. [Wilkins 1668 available online]

Windelband, Wilhelm, 1903, *Lehrbuch der Geschichte der Philosophie*, Tübingen: J.C.B. Mohr.

Wolfram, Stephen, 2002, *A New Kind of Science*, Champaign, IL: Wolfram Media.

Wolpert, David H. and William Macready, 2007, "Using Self-Dissimilarity to Quantify Complexity", *Complexity*, 12(3): 77–85. doi:10.1002/cplx.20165

Wu, Kun, 2010, "The Basic Theory of the Philosophy of Information", in *Proceedings of the 4th International Conference on the Foundations of Information Science*, Beijing, China, Pp. 21–24.

–––, 2016, "The Interaction and Convergence of the Philosophy and Science of Information", *Philosophies*, 1(3): 228–244. doi:10.3390/philosophies1030228

Zuse, Konrad, 1969, *Rechnender Raum*, Braunschweig: Friedrich Vieweg & Sohn. Translated as *Calculating Space*, MIT Technical Translation AZT-70-164-GEMIT, MIT (Proj. MAC), Cambridge, MA, Feb. 1970. English revised by A. German and H. Zenil 2012. [Zuse 1969 [2012] available online]

Academic Tools

⚘ How to cite this entry.
⚘ Preview the PDF version of this entry at the Friends of the SEP Society.
⚘ Look up this entry topic at the Indiana Philosophy Ontology Project (InPhO).
PP Enhanced bibliography for this entry at PhilPapers, with links to its database.

Other Internet Resources

Aaronson, Scott, 2006, Reasons to Believe, *Shtetl-Optimized* blog post, September 4, 2006.

Adriaans, Pieter W., 2016, "A General Theory of Information and Computation", unpublished manuscript, November 2016, arXiv:1611.07829.

Bekenstein, Jacob D., 1994, "Do We Understand Black Hole Entropy?", Plenary talk at Seventh Marcel Grossman meeting at Stanford University., arXiv:gr-qc/9409015.

Churchill, Alex, 2012, Magic: the Gathering is Turing Complete.

Cook, Stephen, 2000, The P versus NP Problem, Clay Mathematical Institute; The Millennium Prize Problem.

Huber, Franz, 2007, Confirmation and Induction, entry in the *Internet Encyclopedia of Philosophy*.

Sajjad, H. Rizvi, 2006, "Avicenna/Ibn Sina", entry in the *Internet Encyclopedia of Philosophy*.

Weisstein, E.W., 2002, "The Riemann Hypothesis", *Wolfram Mathworld*.

Computability – What would it mean to disprove Church-Turing thesis?, discussion on Theoretical Computer Science StackExchange.

Prime Number Theorem, *Encyclopedia Britannica*, December 20, 2010.

Hardware random number generator, *Wikipedia* entry, November 2018.

Wolff, J. Gerard, 2004, "Unifying Computing and Cognition: The SP Theory and Its Applications", ArXiv:Cs/0401009, January,.***check

whether same*

Related Entries

Aristotle, Special Topics: causality | Church-Turing Thesis | epistemic paradoxes | Frege, Gottlob: controversy with Hilbert | Frege, Gottlob: theorem and foundations for arithmetic | function: recursive | Gödel, Kurt: incompleteness theorems | information: biological | information: semantic conceptions of | information processing: and thermodynamic entropy | logic: and information | logic: substructural | mathematics, philosophy of | Ockham [Occam], William | Plato: middle period metaphysics and epistemology | Port Royal Logic | properties | quantum theory: quantum entanglement and information | rationalism vs. empiricism | rigid designators | Russell's paradox | set theory | set theory: alternative axiomatic theories | set theory: continuum hypothesis | time: thermodynamic asymmetry in

Notes to Information

1. The Royal Prussian Academy of Sciences (*Königlich Preussische Academie der Wissenschaften* or *Académie Royale des Sciences et des Belles-Lettres* of Berlin) had annual essay contests starting in 1744. This one was issued in 1769 and the prize awarded in 1771 to Johann Gottfried Herder (see entry).

2. From article 2 of the regulations issued in 1866:

La Société n'admet aucune communication concernant, soit l'origine du langage soit la création d'une langue universelle.

The Society does not accept any communication concerning either the origin of language or the creation of a universal language.