Complexity and Information

– A Critical Evaluation of Algorithmic Information Theory

Panu Raatikainen

"Complexity" is a catchword of certain extremely popular and rapidly developing interdisciplinary new sciences, often called accordingly the sciences of complexity[1]. It is often closely associated with another notably popular but ambiguous word, "information"; information, in turn, may be justly called the central new concept in the whole 20th century science. Moreover, the notion of information is regularly coupled with a key concept of thermodynamics, viz. *entropy*. And like this was not enough, it is quite usual to add one more, at present extraordinarily popular notion, namely *chaos*, and wed it with the above-mentioned concepts.[2]

It is my aim in this paper to critically analyse this conceptual mess from a logical and philosophical point of view, concentrating on the concepts of complexity and information, and the question concerning the true relation between them. I shall focus especially on the so-called algorithmic information theory, which has lately become extraordinarily popular, especially in theoretical computer science and in the above-mentioned new sciences of complexity; indeed, today one can hardly open a popular science book without meeting this theory[3]. It is notable that this approach simply equates information content with complexity; and it is furthermore regularly offered as the preferable explication of the notions of entropy[4] and chaos[5]. This theory has also been a source of various rather fantastic philosophical speculations.[6]

1 "Information" in ordinary language

Before taking a closer look to the various technical approaches called – with better or worse justification – information theories, let us for a moment consider what one usually actually means by the word "information" in ordinary language. This is because I aim to evaluate how well various theories of information that have been put forward actually capture this pretheoretical notion, i.e. how apt explications of the informal concept of information they really are.

The Oxford English Dictionary (OED) explains the root of the verb "inform" as follows: "**inform**. ... L. *informāre*, to give form to, shape, fashion, form an idea of, describe". The primary sense of the word had undergone various developments in ancient and mediaeval Latin, and in French, before it appeared in English; the Latin word had a very restricted use. The earliest uses of "information" (and "inform") in English, in the middle-ages, were related to teaching, training, instructing and disciplining (usually the mind, or the character): "**I.1.a.** The action of informing ... ; formation or moulding of the mind or character, training, instruction, teaching, communication of instructive knowledge. Now *rare* or *Obs.* ... **b.** with an and pl. An item of training; an instruction. *Obs.* ... **c.** Divine instruction, inspiration. *Obs.* ... **d.** Capacity of informing; instructiveness. *rare*." (OED)

Note that all the above senses are classified as rare or obsolete. The relevant basic senses of "information", according to the OED, are the following: "**2.** The action of informing (in sense **5** of the verb); communication of the knowledge or 'news' of some fact or occurrence; the action of telling or fact of being told of something." Sense **5** of the verb "inform" referred above is: "To impart knowledge of some particular fact or occurrence to (a person); to tell (one) of or acquaint (one) with something; to appraise." And the OED adds: "The prevailing modern sense." Further, the OED continues: "**3.a.** Knowledge communicated concerning some particular fact, subject or event; that of which one is appraised or told; intelligence, news. *spec.* contrasted with *data*. ...... **b.** with *an* and *pl.* An item of information or intelligence; a fact or circumstance of which one is told. ..."

Although the OED finds in total 18 different meanings for the word "information", what is left as the basic modern ordinary senses, if one ignores the above-mentioned rare or obsolete senses, various technical juridical meanings and, of course, the definition of the term in the mathematical information theory, are indeed the above **2.** and **3.a.–b.** In various information theories, "information" is used primarily to refer to what is communicated rather than to the action of communication; hence it is the sense **3.a.–b.** that is most pertinent here. Further, the interest of information theories is not so much in the concept of information as such but rather in measuring the amount of information in a message, i.e. its informativeness. And this is explained in the OED as follows (ignoring again obsolete and juridical senses): "**informative** ... Having the quality of informing ... Having the quality of imparting knowledge or communicating information; instructive"; and further: "**informativeness** ... The quality or condition of being informative".

I think one can justly conclude from the above that in the standard modern sense of the word, information belongs essentially to the context of communication. Moreover, most of the senses under **3.a.** and **b.** seem to imply that information is, in some sense, the content of the communicated message ("that of which one is appraised or told", or "a fact or circumstance of which one is told"; see above) rather than just the message itself as a syntactical object, i.e. as a string of uninterpreted symbols (although "An item of information or intelligence" above is somewhat ambiguous in this respect) – there will be much more of this difference below, in section 3.2. Accordingly, how informative a message is corresponds to the property of being instructive, to the intelligence it contains, to its news value.

Evidently information in its ordinary sense, as something that is communicated, is primarily a language-mediated issue, i.e. what bears information is some sort of linguistics entity, "a message" consisting of symbols. It is rather clearly assumed that in the communication situation there are both a sender – likely a conscious and intentional agent – who informs, and a receiver, someone who is informed. What bears information is thus primarily linguistic expressions, or symbols. Naturally, one can use pictures, smoke signals, sticks and stones, or whatsoever; but they convey information only insofar as there is an agreed interpretation of these symbols between a sender and a receiver. Indeed, one of the essential characteristics of language in general is its conventionality and arbitrariness; usually linguistic symbols have no intrinsic or necessary connection with their referents. Consequently, whatever bears information does so solely in virtue of the conventional interpretation attached to it by the relevant linguistic community (minimally, by the sender and the receiver).

Although admittedly one speaks often as though one received information directly from the physical reality e.g. in perception, or in an experiment or a measurement, it is important to see that one is then using merely a figure of speech: there is no intentional sender, no mediating language, and thus no message; actually this is a case of causal interaction rather than of information transmitting or communication. Even if the "receiver" consequently forms a belief with a propositional content, this is only as a result of conceptualization on his part: the physical "situation" just does not have this propositional content, nor is this content transmitted in the causal interaction between the physical reality and the observer. Strictly speaking, it is only the observation sentence reporting the observation that can convey any information. To claim otherwise is to illegitimately personify the physical reality, and to project to the physical reality properties of a communication situation that simply are not out there.[7]

Nothing I have said above is particularly new or original, nor should it be too controversial. The general recognition of the conventionality of (at least a part of) natural language is an age-old philosophical truism; the theme was discussed vividly already in Plato's *Kratylos.* The above issues are closely related to the well known distinction, due to Paul Grice in his classical article "Meaning" [1957], between natural and non-natural meaning. By the former, he means a case like 'foetid breath means tooth decay', i.e. a case where 'means' is roughly synonymous with 'is a reliable sign of'. This should be clearly distinguished from the cases of meaning which are restricted to language or other symbolic system, i.e. what Grice called non-natural meaning. Similar distinction can be observed also already in Charles Sanders Peirce's [1867] famous thricotomy of indices, icons and symbols. These denote the different ways in which signs may be connected to their referents: (1) *indices* have some sort of factual, e.g. causal relation to what they signify; this connection holds independently of any interpretation; (2) *icons* resemble in some respect their referents; they have shared properties with what they represent; (3) finally, *symbols* can act as sign only because there is a general practice to use them to represent their referents, because they are interpreted in this particular way. Clearly the difference between indices and symbols in analogous to Grice's distinction between natural and non-natural meaning.

But uncontroversial, and even trivial, as the above observations on the notion of information may appear, they are worthy of stating here, for they appear to be ignored repeatedly in motivation, application and interpretation of certain popular theories of information, as we will shortly see.

2 Some logical distinctions

At this point, let us make a digression and pay attention to certain basic logical distinction that will play an essential role in what follows.

2.1 Use and mention

Confusing an object and its name is a classical philosophical fallacy. The risk of such confusion is much larger when one is considering language; one regularly conflates the case of *using* an expression to say something, and the case of *mentioning* the expression i.e. speaking about the expression. But easy as this confusion is to make, it can indeed lead to terrible philosophical conclusions.

The clear distinction between *use* and *mention* of a linguistic expression is indeed entirely fundamental. Already Frege, in his *Grundgesetze* ([1893],

p. 4), recognized the importance of this distinction. After him, it was more or less ignored by the logicians for some thirty years. The issue was discussed again clearly in Carnap's *Logische Syntax* ([1934], §§41–42). The contemporary philosopher who has especially warned against confusing use and mention is W. V. Quine (see e.g. Quine [1940], §4; [1987b]).

Let me quote Quine's brilliant and already classical exposition at some length:

> If the object is a man or a city, physical circumstances prevent the error of using it instead of its name; when the object is a name or other expression in turn, however, the error is easily committed. As an illustration of the essential distinction, consider these following three statements:

> (1) Boston is populous,
> (2) Boston is disyllabic,
> (3) ‘Boston’ is disyllabic.
> The first two are incompatible, and indeed (1) is true and (2) false. Boston is a city rather than a word, and whereas a city may be populous, only a word is disyllabic. To say that the place-name in question is disyllabic we must use, not that name itself, but a name of it. The name of a name or other expression is commonly formed by putting the named expression in single quotation marks; the whole, called *quotation*, denotes its interior. This device is used in (3), which, like (1), is true. (3) contains a name of the disyllabic word, just as (1) contains a name of the populous city in question. (3) is *about* a word which (1) *contains*; and (1) is about no word at all, but a city. In (1) the place-name is *used*, and in this way the city is *mentioned*; in (3) a quotation is used, and in this way the place-name is mentioned. We mention x by using a name of x; and a statement about x contains a name of x....
>
> ...‘Boston is populous’ is about Boston and contains ‘Boston’; ‘‘Boston’ is disyllabic’ is about ‘Boston’ and contains “Boston”. “Boston” designates ‘Boston’, which in turn designates Boston. To mention Boston we use ‘Boston’ or a synonym, and to mention ‘Boston’ we use “Boston” or a synonym. “Boston” contains six letters and just one pair of quotation marks; ‘Boston’ contains six letters and no quotation marks; and Boston contains some 800 000 peoples. (Quine [1940], pp. 23–24)

Further, it is worth noting that quotation is by no means the only way to mention an expression:

> Scrupulous use of quotation marks is the main practical measure against confusing objects with their names. But... this particular method of naming expressions is not theoretically essential. E.g. using elaborately descriptive names of ‘Boston’, we might paraphrase (3) in either of the following ways:
>
> The word composed successively of the second, fifteenth, nineteenth, twentieth, fifteenth, and fourteenth letters of the alphabet is disyllabic.
>
> The 4354th word of Chants Democratic is disyllabic.

(Quine [1940], p. 26)

2.2 Object language and metalanguage

The distinction between *use* and *mention* is closely related to another distinction with a vital importance, viz. that between the *object language* and the *metalanguage*, i.e. the language about which one is speaking, and the language *in* which one is speaking about the object language. This distinction was first suggested by Russell in his introduction to Wittgenstein's *Tractatus* [1922] in order to avoid the radical conclusions of the work. It was made unavoidable in logic by Tarski, who in his epoch-making work on the concept of truth (Tarski [1933], cf. [1944]) demonstrated that the distinction is, in a definite sense, necessary for avoiding the contradictions caused by the semantic paradoxes like "The Liar" (i.e. "This sentence is false"). Today the distinction is fully standard in logic and philosophy.

There is, of course, a contrary tendency in logic, due to Gödel's justly famous method of the arithmetization of syntax (introduced in Gödel [1931]) which, in a sense, enables one to speak about the language in the very same language. Obviously, there is no actual contradiction here, for Tarski's work focuses on semantics, whereas Gödel's method is applied only to syntax. And even if one is able, via coding – or Gödel numbering, as it is standardly called – to discuss the syntax of a language in the very same language, it is no less necessary to make a clear distinction between the expression one mentions, and the expression one uses to speak about the former expression (and thus the use of the experssions as a part of object language, and as a part of metalanguage). The distinction is just as essential in the Gödelian approach: e.g. there is all the difference between a numeral $\overline{n}$ and the numeral $\overline{\ulcorner n \urcorner}$ which names the Gödel number of the former numeral (cf. Quine [1987a]).

3 The varieties of information theories

I next give a very short review of the various suggested theories of information.

3.1 The statistical theory of information

The statistical theory of information emerged from the practical engineering problem of efficient signal transmission e.g. in telegraphy and telephony. The basic ideas of the theory were suggested already in the 1920s, by H. Nyquist [1924] and especially by R. Hartley [1928]. However, it was Claude Shannon's classical treatment "A Mathematical Theory of Communication" [1948] that really established a systematic theory and received extraordinary attention in the scientific community. It has become the standard mathematical theory of information.[8]

In the statistical theory of information, one considers a fixed ensemble of possible messages m_i which occur in a communication channel with a certain probability p_i – where probability is standardly understood according to the frequency interpretation. A few natural requirements (positivity, additivity) on a possible information measure naturally lead to the following definition (first suggested by Wiener [1948]): The *amount of information* provided by a single message $m_i, I(m_i) = -log_2\ p_i$.

Further, assume that X is a discrete random variable, which may take the values $x_1, \ldots, x_n$, and that $x_i (1 \leq i \leq n)$ has the probability p_i. Then the average uncertainty relative to X, i.e. the *entropy*[9] (or Shannon-entropy) of X, $H(X)$ is (again, this is the unique function that satisfies certain natural requirements):

$$H(X) = H(p_1 \ldots . p_n) = -\Sigma p_i\ log_2\ p_i.$$

One has, on the basis of the above definitions, developed an enormously rich mathematical theory, dealing e.g. with conditional entropy, optimal coding, "redundancy" in the messages and the effects of "noise". These further – certainly important – developments of the theory do not, however, have any bearing on my present philosophical study.

3.2 The semantic theory of information

Although Shannon himself consistently emphasized that his theory has absolutely nothing to do with the content, or meaning, of messages ("... semantic aspects of communication are irrelevant to the engineering problem"; Shannon [1948]), very few of his followers have been able to resist the temptation to derive from the theory illicit semantical conclusion concerning the information *content* of a message, so strongly suggested by the ordinary non-technical meaning of "information" and "informativeness" (cf. Bar-Hillel [1955]). As Bar-Hillel writes, "it is psychologically almost impossible not to make the shift from the one sense of information, ... i.e. information = signal sequence, to the other sense, information = what is expressed by the signal sequence ... ([1955], p. 284)". Consequently, the development of the statistical theory of information, wed with the inability to clearly separate the above two senses of "information", has led to various untenable speculations.[10]

Basically, the semantic theory of information is a theory that aims to take seriously this distinction, which is in fact just a special case of the distinction between use and mention, and explicitly theorize about what is expressed by messages, i.e. about their information content[11]. The basic ideas of semantical theory of information were suggested, informally, by Popper already in his *Logik der Forschung* [1934]. As a systematic theory it was initiated by Carnap and Bar-Hillel in the early fifties (apparently reacting to misuse of Shannonian theory); and it has been since then importantly developed and generalized by Hintikka.[12]

The intuitive point of departure of this theory is the idea that the more possibilities (possible states of affairs) a sentence rules out, the more informative it is, i.e. information is the elimination of uncertainty. For example, the sentence "Apples are not blue" may very well be, in the sense of the statistical theory (when mentioned), less probable and hence more "informative" than the sentence "Apples are red"; but certainly the latter sentence is more specific and contains more information (when used). Hence one must clearly distinguish between the probability of a string of symbols, and the probability of the state of affairs that the strings expresses.

In semantic information theory, one considers a particular language L. Let $W_1, \ldots, W_n$ be, in some definite sense, the mutually exclusive maximal consistent sentences of L (Carnap's "state descriptions", or Hintikka's "constituents", depending on the logical setting and the expressive recources). Now every consistent sentence is compatible with some of these and excludes all the others. Next, assume that the probability p_i of the state of affairs expressed by W_i ($1 \leq i \leq n$) such that $p_1 + \ldots + p_n = 1$ is given, for all i. As a consequence, a definite probability is thus attached to every sentence. (In the semantic theory of information, probability is usually understood either according to the logical interpretation of probability or according to some epistemic interpretation of probability.)

Carnap and Bar-Hillel have suggested the following two measures of information (where $p(H)$ is the probability of the state of affairs expressed by H):

$$cont(H) = 1 - p(H)$$

$$inf(H) = -log\ p(H)$$

Let me emphasize that $p(H)$ here is the probability of the state of affairs expressed by H, *not* the probability of 'H' in some communication channel, i.e. H is used, not mentioned. Bar-Hillel suggests that *cont*(H) measures the substantive information content of sentence H, whereas *inf*(H) measures the surprise value, or the unexpectedness, of the sentence H. It can be argued that the pre-theoretical notion of information is somewhat ambiguous, sometimes used more in the sense of *cont*, and sometimes rather in the sense of *inf*. It is worth of noting that, if one focuses on *inf*, the statistical and the semantical theories thus share the same "information calculus": they differ in how this calculus is interpreted.

There is a feature in the classical theory of semantic information that one may find somewhat counterintuitive: according to it a logical inference cannot provide new information. However, Hintikka has later developed measures of information which do not have this property: he has defined the notion of "surface information" which may increase in course of logical inference, in contradistinction to "depth information" which is zero for all logical truths (Hintikka [1970b]). Such more fine-grained notions are, in my mind, philosophically very important and interesting refinements of the theory of semantic information. Nevertheless, in the present study I shall not lean on them but utilize only the most general features of the semantic information theory (that is, I shall lean solely on its explication of the notion of information content, not on any particular measure of information).

3.3 Algorithmic information theory

Algorithmic information theory is a new-comer in the field. It is notable that unlike the above two theories, according to it information is not to be defined in terms of probability. The basic idea of this theory, i.e. the notion of program-size complexity, was suggested in the 1960s independently by Solomonoff [1964], Kolmogorov [1965] and Chaitin [1966]. However, its original purpose was not to function as a definition of informativeness. Solomonoff used it in his work on the inductive inference, Kolmogorov aimed initially to give a satisfactory definition for the problematic notion of random sequence in probability theory, and Chaitin was studying just the program-size complexity of Turing machines. It was Kolmogorov who first (in 1965) suggested that this notion provides an explication of the concept of information content of a particular string of symbols. Later Chaitin followed him in this interpretation.

The intuitive idea behind this theory is that the more difficult an object is to specify or describe, the more complex it is. One then explicates this intuition by considering binary strings outputted by a Turing machine T, and defining the *complexity* of a binary string s as the size of the minimal program that, when given to T, prints s and halts. Finally, one simply equates the *information content* of a string with its complexity.[13]

As I must later in my critical discussion dig rather deep in the technical details of algorithmic information theory, I shall next give a slightly more detailed survey of the theory.[14]

One considers the set of finite binary strings $B^* = \{\Lambda, 0, 1, 00, 01, 10, 11, 000, \ldots\}$, where Λ is the empty string. The length of a string s is deonted by $l(s)$. $\langle x, y \rangle$ is any standard pairing function (which is fixed below). Assume that one is considering objects from a countable domain of objects S. To each object x in S, one attaches a code $n(x)$, that is a natural number or the corresponding binary string. Let us assume that this coding is one-one.

Next one fixes a way to present algorithms. For most cases, any explication of computable functions, e.g. partial recursive functions, or Turing machines, would do. However, for certain technical reasons, it is convenient to assume that one uses Turing machines that accept only so-called prefix-free, or instantaneous, binary programs (for details, see e.g. Chaitin [1975]). It is clear that one can easily compute all partial recursive functions by such computers.

DEFINITION. *The algorithmic complexity* of a string s relative to a Turing machine T, $K_T(s)$, is $min\ \{l(p) : T(p) = s\}$.

Through coding, it is customary to generalize this definition to apply to any countable domain of objects. One simply equates the objects with their codes.

Let us assume that one has fixed an effective one-one coding of all Turing machines to N, or to B^*, such that one can numerate Turing machines: $T_1, T_2, T_3, \ldots$, or equally, when coded and given to a Universal Turing machine, programs $p_1, p_2, p_3, \ldots$. So, let me emphasize this from the beginning, there are two very different codings in use here: the coding of objects of some domain, $n(x)$, and the coding of Turing machines. This fact will play a crucial role below.

Also fundamental for this theory is the notion of *conditional* or *relative* complexity:

DEFINITION. The complexity K_T of x *conditional* to y is defined by $K_T(x/y) =$

$min\{l(p) : T(\langle p, y\rangle) = x\}$, and $K_T(x/y) = \infty$ if there is no such p.

From the existence of Universal Turing machines (Turing [1936-7]) follows almost immediately the following key result of algorithmic information theory:

THE INVARIANCE THEOREM. There is an optimal Turing machine U such that for any Turing machine T there is a constant $c < \infty$ such that for all x, y we have $K_U(x/y) \leq K_T(x/y) + c$.

Note that c here is indeed independent of x and y, and depends only on T; it may be aptly called the cost of simulating T by U; in favourable cases (i.e. with a suitable U) it is simply the code number of T. Because of this theorem, it is customary in algorithmic information theory to fix some optimal machine U, and omit the subscripts and ignore the finite differences that may result.

The rather unelegant behaviour of the above-defined conditional complexity in the case of additivity led Chaitin to revise the definition of relative or conditional complexity (in Chaitin [1975]); since then, he has defined the relative or conditional complexity $C(x/y)$ of x given y to be the size of of the smallest program to calculate x from the *minimal program* for y[15] – whereas in the traditional concept of relative complexity $K(x/y)$ one is directly given y. The difference between these two definitions will play a certain role below.

4 A critical evaluation of algorithmic information theory

As was noted above, the confusions and illicit semantical conclusions with respect to the statistical theory of information have received a comprehensive critical analysis from the logicians and philosophers. Algorithmic information theory has, on the other hand, so far been left largely in peace. Although it has been developed into a formally rich theory with numerous ramifications, there has been very little analytic reflection concerning its motivation and soundness of its basic assumptions.

Unfortunately, algorithmic information theory shares many of the limitations of the statistical theory of information, and is apparently even less conscious about them. Particularly, it has developed, as far as I can see, in complete ignorance of the fundamental criticism directed by logicians and philosophers towards the classical statistical theory of information and of the development of the semantic theory of information. Understandably, then, it continues to contain many of the confusions that were clearly pointed out, in the case of Shannonian theory, already in the 1950's – in addition to several several confusions of its own. Let us finally start our critical examination of algorithmic information theory. My aim is to show that the theory is – the impressive mathematical developments notwithstanding – wed with serious logical confusions.

For a start, recall that one intuitively motivates algorithmic information theory by considering the problem of describing or specifying certain objects, and concluding that an object may be reasonably called "complicated" if its description is necessarily long. I grant this agrees quite well with common sense.

However, in my mind it is a little bit odd to explicate the intuitive idea of specifying an object by algorithms, Turing machines, or programs. For specifications are a sort of descriptive expressions. Algorithms, on the other hand, are rather sets of instructions or commands; a human computer may or may not follow them; a deterministic computing machine has no choice. As a result, one finds, or constructs, an object. But certainly it is one thing to specify an object, and another thing to give instructions sufficient for finding one – not to mention for constructing one; think of "the murderer of Olof Palme", or "the first manned spaceship to Mars" (assuming there will someday be one).

Nevertheless, in mathematical contexts, at least, this may not be that serious a problem, so I shall bypass it and assume below – at least for the sake of an argument – that one can conveniently explicate the intuitive idea of describing, specifying, or defining, an object with the help of an algorithm, or a Turing machine, that computes it. Thus I do not aim here to question the theory as an explication of the notion of complexity.

The crucial question for algorithmic information theory, qua a theory of information, is however the following: is it really a reasonable explication of the notion of information content? My conclusion will be thoroughly negative. Let us begin to study what are my reasons for this strong conclusion.

4.1 Preparatory considerations on complexity and information content

To begin with, I shall discuss the question of plausibility of algorithmic information theory, and especially its basic claim that one should consider the algorithmic complexity of a string to be its information content, on the common-sense level and base my consideration on certain fundamental intuitions concerning the notions of information and informativeness. These general remarks serve as a preparation for the more technical and detailed considerations that follow.

It is illuminating to compare the ideas suggested by algorithmic information theory with certain highly plausible ideas that are the basis of the semantic theory of information. Let us recall, in particular, that according to the latter theory, the more possibilities an expression excludes, the more informative it is, and that consequently a logical tautology (which does not rule out any possible states of affairs but allows them all) has a minimal information content.

Consider then the infinite set of tautologies of propositional logic, all minimally informative. Assume next that one has fixed any suitable coding of a formalized language to the binary strings. Obviously there are then, when mentioned, i.e. considered as syntactical objects, arbitrarily (algorithmically) complex propositional tautologies. However, is there any reason to think that the information content of a complex tautology is high, i.e. that it is highly informative? No; it does not exclude any possible states of affairs; it does not tell anything about the world. Actually, a simple sentence p_0 is much more informative (it says at least something, viz. rules out $\neg p_0$) than a more complex but totally uninformative tautology such as

$$\neg p_{27} \vee p_0 \vee \neg p_7 \vee p_3 \vee p_7 \vee \neg p_0 \vee \neg p_3 \vee p_{27};$$

and evidently, one could construct arbitrarily complex such examples. But such complex trivialities tell us nothing; they are in no natural sense highly informative. That certain tautologies require lots of space to describe (or, require a long program to print), when *mentioned* in the *metalanguage*, has no consequences concerning their information content, when *used* themselves. Likewise, turning to first order languages, a simple sentence $(\forall x)P(x)$ is clearly much more informative than a complex sentence

$$P(a_3) \vee P(a_8) \vee P(a_{17}) \vee P(a_{41}) \vee P(a_{93}) \vee P(a_{126}) \vee P(a_{457});$$

the former logically implies the latter (but not *vice versa*), and therefore it is, in any plausible sense of the word, more informative. In fact, a sentence highly informative for its content may be very simple for its form: think of "$E = mc^2$", for example.

According to common sense uncorrupted by any "information theory", if a sentence S logically implies another sentence S', but not *vice versa*, S is clearly more informative, i.e. contains more information, than S'. However, neither the statistical nor the algorithmic theory of information respect this basic idea. They ascribe a high information content to a logically weak sentence (e.g. to a tautology of propositional logic) if it happens to be algorithmically complex, or "improbable" – when *mentioned*. They pay no attention to what the sentence says when *used*, i.e. to its content. Also, one would expect that two messages that are logically equivalent (at least, if this equivalence is in addition of a trivial sort) would have the same information content, or would be equally informative. But again, neither the statistical nor the algorithmic approach satisfy this intuitive requirement; for any sentence has infinitely many logical equivalents, which may have arbitrarily large Shannonian or algorithmic "information content".

That is, neither of these approaches cares whether a message is a bare contradiction, trivial tautology, or a genuinely informative statement; they focus solely on the "probability" or the complexity of the syntactical form of sentences when mentioned. It is exactly this sort of counter-intuitive aspects that have motivated the development of the semantic theory of information. By the way, one may note that although the statistical information theory is often classified (and one could likewise classify the algorithmic information theory) as a syntactic theory of information, in contradistinction to the semantic theory of information, the above problems have nothing to do with the meanings of non-logical terms of the language; being a contradiction or a tautology, the properties of implying something or being equivalent with something, are all completely syntactical (proof theoretical) properties; yet neither the statistical theory nor algorithmic theory is able to take them to the account.

Already at this general level of consideration, one begins to see how the algorithmic information theory ignores the distinction between use and mention of an expression. This becomes even more evident below.

4.2 The question of the bearer of information

The above preparatory critical discussion concerned the basic idea of algorithmic information theory, according to which the information content of an object is to be equated with its complexity, i.e. the size of its minimal specification. In fact, however, the literature on algorithmic information theory is rather confused in the question of the bearer of information: Is it the object specified, or the specification? For consider first the following representative quotations:

> The entropy, or *information content*, or complexity, of a string is defined to be the number of bits needed to specify it . . . (Chaitin [1974b], p. 495; my emphasis)

> Intuitively, *the amount of information* in a finite string is the size

(number of binary digits or *bits*) of the shortest program that, without additional data, computes the string and terminates. ... Thus, a long sequence of 1's such as

$$\overbrace{11111\ldots1}^{10{,}000\text{ times}}$$

contains little information because a program of size about log 10,000 bits outputs it:

`for i:= 1 to 10,000`

`print 1`. (Li and Vitányi [1993], Preface, p. v; my emphasis)

> We identify the length of the description of x with respect to a fixed specification function D_0 with the 'algorithmic (descriptional) complexity' of x. ... This complexity can be viewed as '*absolute information content*': the amount of information which needs to be transmitted between all senders and receivers when they communicate the message in absence of any other*a priori* knowledge which restricts the domain of the message. (Li and Vitányi [1993], p. 2; my emphasis)

> A way to measure the *information content* of some text is to determine the size of the smallest string (code, input) from which it can be reproduced by some computer (decoder, interpreter). (Calude [1994], p. 25)

In all the above quotations, it is clearly assumed that it is the object specified (or, more exactly, the string outputted by a Turing machine) that bears information, i.e. has certain information content. Nevertheless, in other contexts it is as clearly assumed that it is rather the specification (or, the program inputted to a Turing machine) that contains the information:

> If any object is "simply" constructed, then for its description a small quantity of information is sufficient; but if it is "complicated", then its description must contain much information. ... it is convenient to call the quantity thus introduced the "complexity". (Kolmogorov [1970/1983], p. 32)

> The complexity of a binary string is the minimum quantity of information needed to define the string...... the complexity of a binary string is the information needed to define it ... (Chaitin [1974a])

> We define the complexity of a text as the length of the shortest binary word containing all the information that is necessary

> for recovering the text in question with the help of some fixed method of coding. (Zvonkin and Levin [1970], p. 88)

Li and Vitányi even have both ideas in the same paragraph:

> One interpretation of the complexity $C(x)$ is as the quantity of information needed for the recovery of an object x from scratch. ... Hence the complexity is 'absolute information' in an object. (Li and Vitányi [1993], p. 140.)

In the light of my earlier discussion on the notion of information, it should be clear that it is only the specification that can bear information, not an arbitrary object that has been specified – I find the latter idea totally preposterous (although, obviously it is possible that the objects described are themselves linguistic entities and thus bear some information; I shall discuss this special case below).

Note also that one is here dangerously close to defining the information content of a string to be the information content of the (minimal) string that specifies the former string, and thus of giving a circular definition. One breaks the circle only by understanding the latter, i.e. the information content of the specification, to be just its length. But if, at the end of the day, the information contained in a binary string is defined to be simply as its length, one may wonder why this definition has not been used from the beginning; why introduce partial recursive functions and open the door for uncomputability? As most of the strings are anyway incompressible[16], one would lose a little. But be that as it may, for plainly the length and the information content – in any plausible sense – of an expression are in no way related.

Prima facie, it is thus very surprising that one would like to conclude that the (algorithmic) complexity of an object is to be equated with the information content of the object; after all, all sorts of objects may be justly called complex, but only linguistic entities can have an information content. Explicit arguments justifying this step are surprisingly sparse in the literature. Most often, one tends to take the mere formal analogues with the traditional statistical theory of information as a sufficient justification. But this is hardly convincing; for the whole approach has been motivated by questioning the adequacy of the classical approach. And in any case, the mere formal similarity (and there are clear differences as well) of the resulting, rather general calculi just cannot be a sufficient reason for equating the two concepts. For one should recall e.g. that the statistical theory of information and the semantical theory of information share, after all, exactly the same caluculus, but concern yet wholly different issues.

There is, nevertheless, in Kolmogorov's writings a sort of argument for taking the complexity to be the information content. Let us turn to this line of reasoning.

4.3 Kolmogorov on the information content

Kolmogorov's line of reasoning, although somewhat implicitly present also in various other papers including his pioneering one [1965], is most explicit in his mature presentation [1970/1983][17]:

> The complexity of specifying any object can be facilitated when any other object is already specified. This fact reflects the following definition of the relative complexity of an object x, given an object y:
>
> $$K_S(x/y) = \min_{S(n(y),p)=n(x)} l(p)$$
>
> Here the method S of relative determinations is a function of two arguments, the number of the object y and the number p of the programme for computing the number $n(x)$ when y is given. ...
>
> ... If the relative complexity $K(x/y)$ is much smaller than the unconditional complexity $K(x)$, then it is natural to interpret it as an indication that the object y contains some "information" about x. It is, therefore, natural to regard the difference
>
> $$I_S(x/y) = K_S(x) - K_S(x/y)$$
>
> as a quantitative measure of the information about x contained in y. As a value of the second argument of the function $S(n, p)$ we admit the number 0, and we put
>
> $$S(n, 0) = n$$
>
> (the zero programme from n produces n). Then
>
> $$K_S(x/x) = 0, \quad I_S(x/x) = K_S(x).$$
>
> Thus, the complexity $K_S(x)$ can be called the information contained in an object about itself. (Kolmogorov [1970/1983], p. 37)

Kolmogorov's argument is as ingenious as it is problematic. What actually happens in this argument is that one takes the name (code) of an object y, i.e. $n(y)$, gives it to a computer which, according to an entirely different coding (of Turing machines, or programs) interprets it as a program (or rather, a subprogram) which together with another (sub-) program p, outputs the name of an object x, i.e. $n(x)$. But surely it is completely accidental whether the binary code (according to one coding system) of an object y is, according to another totally unrelated coding system, a code of a program that can somehow facilitate the computation of the name of another object x. It is quite bizarre to conclude that in such cases the object y somehow contains information about the object x. In other words, when interpreted as a program inputted for the relevant Turing machine, '$n(y)$' has absolutely nothing to do with y.

Thus one illegitimately conflates two codings, two very different interpretations of as such meaningless binary strings. The case is completely analogous to the following, more transparent case: "Pain", in English, means pain; "Pain", in French, means bread; one can satisfy one's hunger by bread; however, one cannot conclude that thus one can satisfy one's hunger by pain. That is, serious confusions are likely to arise if one wildly changes the interpretation of one's language in the midst of an argument.

Likewise, it is simply absurd to think that an object always contains information *about* itself. This is obviously false when one thinks about physical objects, e.g. tables and stones, which just don't contain information about anything. It is likewise false about most linguistic objects, or "texts": e.g. the sentence "It will snow tomorrow" contains information about tomorrows weather, not about the sentence itself; "Turing" names Turing, a man and a genius, not the word; and the expression "the present president of U.S.A." specifies a certain very powerful man, not the definite description. It is one thing – a platitude – that any object is identical with itself and that any sentence is logically equivalent with itself, and a wholly different thing that an expression describes itself, or that an object contains information *about* itself – which is in most cases simply false, if not a total nonsense. Again, only confusing totally use and mention can one end up with such conclusion.

Besides, it is noteworthy that the trick in the final part of Kolmogorov's reasoning is done by the stipulation that $S(n,0) = n$.[18] It then follows immediately that $K_S(x/x) = 0$ and $I_S(x/x) = K_S(x)$. Of course, technically there is nothing wrong in such a stipulation. But it is important to see that it is a rather arbitrary decision. Here n as an argument of S is interpreted as a subprogram, and the outputted n is then interpreted as a member of whatever domain one is considering. All this by no means justifies the conclusion that the complexity $K_S(x)$ is – in any reasonable sense of the word – the information contained in the object x about itself.

There is also a further confusion, present in the writings of Kolmogorov and recurring in the literature of this field, between the complexity of an object and the complexity of the name of the object – a clear-cut case of confusion between use and mention. To illustrate its absurdity, let us again consider a concrete example: "Chlamydomonas angulosa" is clearly a more complex string of symbols than "Bubo bubo"; but may one conclude that *Chlamydomonas angulosa*, a certain flagellate, is a more complex object than *Bubo bubo*, the eagle owl? And, what would be even more absurd, how on earth could one conclude that a flagellate thus contains more information than an eagle owl? (Neither of them contain any information, in any natural sense of the word.)

Similar confusions are repeated in the case of the notion of relative information, i.e. in the idea that an object may facilitate the specification of another object. One more illustrative example: A full description of the edge of a certain knife may considerably facilitate the description of the whole knife. However, it is the description of the edge that contains the useful information, not the edge (which just doesn't contain information) – one should keep the distinction between an object and its name or description sharp as a knife.

Note that Chaitin's revised definition of relative or conditional complexity (see above 3.3.) appears to be in preferable this respect – it compares programs to programs, or "specifications" with "specification", and not "specifications" and objects, and does not thus confuse two different interpretations of strings (it will be discussed in more detail below).

In sum, there is a double confusion here; first, one confuses two very different and totally unrelated codings of binary strings, and concludes that the complexity of a sequence of symbols has some intrinsic connection to its information content; and second, it is oddly assumed that complexity of a sequence of symbols used conventionally and completely arbitrarily to denote an object somehow reveals the complexity of the object itself, or even (following from the above confusion) the information content of the object – whatever that could mean.

4.3.1 Texts as objects

One clear source of confusion is that Kolmogorov and his associates sometimes (but not always!) think that the objects "described", or "specified", are themselves some sort of texts:

> One of the central concepts in this article is the concept of the complexity of a certain text (communication). We define the complexity of a text as the length of the shortest binary word containing all the information that is necessary for recovering the text in question with the help of some fixed method of coding. (Zvonkin and Levin [1970], p. 88)

Accordingly, Kolmogorov discussed the possibility of applying different information measures 'to an estimate of the quantity of information contained in a novel' (Kolmogorov [1970/1983], p. 37), and the meaning of 'asking how much information is contained in "War and Peace"?' (Kolmogorov [1965], p. 3).

Nevertheless, although this makes the resulting view slightly less absurd – for in this case the objects that are described may indeed bear some information – the approach is yet guilty of the disastrous confusion of object language (the text described) and the metalanguage (the language in which the description is given; in this case, the programming language in question), and of use and mention.

Let us thus consider, as an example, the following case that is directly relevant to the present issue: let the language in question, i.e. the object language, be the language of arithmetic $L(A)$, and take as the "text" in question e.g. the numeral '3' (officially, $0'''$). Given some coding of this language, this numeral has a binary code s. Assume then that the chosen coding of Turing machines happens to attach the code s to a Turing machine which prints m and halts, where m in turn is the code (in the former coding of $L(A)$) of numeral naming, say, 78498 (certainly such codings are possible, and indeed effectively fixable). Now it is indeed hard to see how on earth the numeral '3' or the natural number 3 contains information about the number 78498 or the corresponding numeral (I am a little bit uncertain how exactly to state the "conclusion", given the total ignorance of the distinction between use and mention, or object and its name, by the theory). And yet it is claimed, by the algorithmic information theory, that this is the case.

Moreover, as noted above, the restriction of these considerations to texts is rather an exception than a rule, as the following quotations clearly demonstrate:

> Consider an "indexed domain of objects", i.e., a countable set $X = x$, with a finite sequence $n(x)$ of zeros and ones ... associated with each element as its index. (Kolmogorov [1965], pp. 4–5)

> Suppose that we are dealing with some domain D of objects in which there is already some standard numbering of objects by numbers $n(x)$. (Kolmogorov [1970/1983], p. 32)

> Identify an object x from a countable infinite sample space S with its index $n(x)$. (Li and Vitányi [1993], p. 90)

Here one is clearly allowed to apply algorithmic information theory to any sort of objects. And apparently all sorts of objects may consequently contain information. Indeed, the literature is loaded with such implausible speculations on the information contained in all sorts of non-linguistic physical objects.

It might have helped if one had clearly distinguished the input language and the output language, which may in general be wholly different languages, although may, of course, coincide. This latter possibility has apparently had the unfortunate consequence that researchers in this field – presumably not very familiar with the crucial philosophical and logical distinctions that matter here (i.e. between use and mention, and the object language and the metalanguage) – have got so deeply confused.

4.4 Chaitin on complexity and information

In certain writings by Chaitin, there occurs also another line of heuristic reasoning that may appear to justify the identification of complexity and information content, although its actual bearing turns out to be rather unclear. Let us thus study Chaitin's approach and the evolution of his view.

To begin with, it is interesting to note that in his early papers from the 1960s, Chaitin does not at all consider the task of measuring the information content of outputted strings. Although he says that he studies "... [t]he use of Turing machines ... from the point of view of information theory" [Chaitin 1966], this means in practice only that he considers the number of bits of the minimal program with a given output. Moreover, he simply notes (in 1.9.) that "There is some connection between the present subject and that of Shannon in [1948]". There is yet no hint of the idea that one could thus measure the information content of the outputted string.

The information-theoretic aspect becomes more visible later, but Chaitin's way of viewing the issue differs considerably from that of Kolmogorov. In his [1970], Chaitin wrote:

> Here we are interested in examining the viewpoint of information theory concerning the efficient transmission of information.

> An information source may be redundant, and information theory teaches us to code or compress messages so that what is redundant is eliminated and communications equipment is optimally employed. ...The receiver must decode the messages; that is, to expand them into their original form. In summary, information theory teaches us that messages from an information source that is not completely random (that is, which does not have maximum entropy) can be compressed. The definition of randomness is merely the converse of this fundamental theorem of information theory; if lack of randomness in a message allows it to be coded into a shorter sequence, then the random messages must be those that cannot be coded into shorter messages. A computing machine is clearly the most general possible decoder for compressed messages. We thus consider that this definition of randomness is in perfect agreement and indeed strongly suggested by the coding theorem for a noiseless channel of information theory.

Thus Chaitin does *not* here consider programs as descriptions *of* outputs, i.e. somehow speaking *about* them, as metalanguage speaks about object-language, but a program is rather considered to be the output – the message – itself, just in coded form. (Although a code can certainly in some occasions be viewed as a name, thus mentioning what it codes, I would argue that in the communication model it is much more natural to consider coded and uncoded message as different *tokens* of the same *type*; or, to use a different image, their relation is more like that between an expression in spoken and written English than that between an expression and its name or specification in the metalanguage. Think of the Morse code, for example.) Thus Chaitin's heuristic model is *essentially different* from that of Kolmogorov.

The picture on which Chaitin leans is, however, rather strange. What is the purported coding method is this picture? Is there a coding function at all? On the one hand, there is no unique "code" of a message; for, by the Padding Lemma[19], there are infinitely many alternative "codes" of any "message"; further, the property of being a "code" of a given "message", in this sense, is undecidable[20]. On the other hand, if one intends here that the minimal program is the code to be used (it appears that this is Chaitin's actual intention), one is again in deep trouble: for, although the decoding function is here at least a partial recursive function, the imagined *coding function* is not even partial recursive, but is strongly non-computable[21]; that is, there is no effective, mechanical procedure for compressing messages. So who on earth is the *sender* in this model – God? An Oracle? I find this setting very implausible as a model of communication. One is given – as fallen from the Heaven – coded messages without there being any imaginable sender or any realistic coding method. Certainly one usually thinks that in a communication system there is a sender who has at hand a computable coding method that gives for each message a unique code – that both coding and decoding are effective, mechanical procedures. (In fact, I think it is reasonable to require that both coding and decoding methods are primitive recursive (and thus total), i.e. that neither of them contain unbounded searches.)

Note that the situation is not even the one of deciphering a code, for even in that case there is a computable coding method – it is just not known. The situation in the above case is even worse: there is no computable coding method, neither known nor unknown. Moreover, some of the apparent "codes" are not real codes, for once given to "the decoder", i.e. to a Universal Turing machine, they result an infinite loop, and the machine never halts; and the "receiver" has no effective means to separate such only apparent codes from the genuine ones.

Further, as the theory of algorithmic complexity clearly tells us, only a very small minority of all "messages" are compressible[22] (of the type that could be "coded" by a program such as "Print 1000 times 'I love you'!"), so in most cases there is no possibility (i.e. not even a non-computable function) to compress the message considerably anyway, and they must be transmitted as they are. So – I am repeating my earlier query – why introduce the idea of only partial recursive decoding function, and thus implicitly assume a strongly non-computable coding method, only because of those very few untypical compressible messages? Why not require total primitive recursive coding and decoding functions? That would, at least, make the model more realistic as a model of communication, and keep it manageable. Anyway, I think one may justly conclude that Chaitin's view that one has here a reasonable model of efficient communication is quite implausible.

But be that as it may, Chaitin did not originally use this communication model to justify the equation of complexity and information content. Indeed, as far as I have been able to find, before 1974 Chaitin does not at all consider the information content of outputted strings, and much less equates it with their algorithmic complexity. That is, before 1974 the information-theoretic aspect of the theory is for Chaitin restricted to the amount of information needed to specify a string, which is measured by him simply by the number of bits (the length) of the specification, i.e. the inputted program.

However, in [1974b] Chaitin's perspective changed, with no further explanation: "The entropy, or *information content*, or complexity, of a string is defined to be the number of bits needed to specify it." One may guess that Chaitin simply borrowed the idea of equating complexity and information content from Kolmogorov. In any case, Chaitin's [1975] paper "A Theory of Program Size Identical to Information Theory" meant a big departure in more than one respect: not only did he introduce prefix-free programs to the theory (this had been anticipated by others, e.g. by Levin) and revise the definition of the conditional or relative complexity; in this paper Chaitin also begun to move more explicitly towards Kolmogorov's way of viewing the issue.

The justification of this fundamental change is, nevertheless, surprisingly dim. Chaitin clearly gives great weight to the formal similarity between the classical information calculus and formulas resulting from his revised approach to the algorithmic complexity. He writes: "There is a persuasive analogy between the entropy concept of information theory and the size of programs"; and: "H and I satisfy the fundamental inequalities of information theory to within error terms of the order of unity." (Chaitin [1975]) It should be clear, though, that this is in itself a rather weak justification – to say the least – for identifying the two subject matters.

Besides this, Chaitin refers to the above-quoted paragraph from his [1970] and writes: "... think of a computer as decoding equipment at the receiving end of a noiseless binary communication channel. Think of its programs as code words, and of the result of the computation as the decoded message." (Chaitin [1975]) Now Chaitin apparently thinks that his above-discussed, rather problematic communication picture somehow justifies his new position.

Later in the paper, Chaitin defines – without any explanation or justification – "the information in one tuple of strings about another", as follows (I have simplified the notation inessentially):

$$I_U(s : t) = K_U(t) - K_U(t/s),$$

where $K_U(t/s)$ is Chaitin's revised conditional complexity (see 3.3., and note 15). Although this differs a little from Kolmogorov's setting, it does not avoid the basic difficulties. Let it suffice to say here that again the relevant partial information about the second string is in the specification of the first string, not in the first string itself (the above knife example and the example about 3 and 78498 are again relevant). Thus the definition does no really capture the notion it is presumed to define.

Since then, Chaitin has unhesitatingly equated information content and algorithmic complexity. Thus, e.g. in a joint paper with Schwartz from

1978 one can read: "$I(X)$, the algorithmic *information content* of X, is defined to be the size in bits of the smallest programs for U to compute X." (Chaitin and Schwartz [1978]); and a couple of years later, Chaitin speaks about "the *information content* of an individual object, which is a measure of how difficult it is to specify or describe how to construct or calculate that object." (Chaitin [1982a], my emphasis).

Chaitin gives, in the above-mentioned papers of in any other publication, absolutely no argument or clear reason for this identification of complexity and information content. In particular, Chaitin never uses explicitly a line of reasoning like that of Kolmogorov. And, because of the problems with his definition of $I_U(s:t)$ mentioned above, the Kolmogorovian argument would not work in the case of Chaitin's slightly revised setting either.

One may thus conclude that Chaitin has not provided any good justification for his later-day identification of information content and complexity. And yet, possibly following Chaitin, this simple but highly problematic identification has unfortunately become the received view in the literature of this field.

4.5 Conclusions

In sum, Kolmogorov's line of reasoning that leads to the idea that the algorithmic complexity is a good measure of the information content of an object is wed to serious confusions. Nor does this argument work with Chaitin's revised definition of conditional complexity. Further, the heuristic communication model that Chaitin has used is rather implausible, and certainly does not justify the equation of algorithmic complexity and information content. And independently of these problems, it has become entirely evident that the complexity and the information content of a linguistic expression have no real connection.

My general conclusion from my above analysis is quite destructive: I claim that "the algorithmic information theory" is in all respects inappropriate as a theory of information. Obviously, the popular name of this rich field of study is accordingly a misnomer. I suggest that one should rather use the name "the algorithmic theory of complexity" – for the theory is after all, at least in its own limits, a reasonably plausible formal explication of the informal notion of complexity. And certainly the notion of the program-size complexity is in itself a notion worth studying. In addition, the algorithmic complexity of an expression as a syntactical object may well provide a reasonable explication for the notion of the simplicity of a theory, which has been for a long time a subject of intensive discussion in the philosophy of science.

Notes

[1] See e.g. Casti [1994], Pagels [1988], Gell-Mann [1994], Zurek [1990b].

[2] Ruelle [1991], Ford [1983], [1989], Casti [1994], Gell-Mann [1994]. For information and entropy (but not chaos), see the useful (although quite uncritical) survey Schumacher [1993]; for a critical discussion, see Wicken [1987].

[3] E.g. Rucker [1987], Pagels [1988], Ruelle [1991], Stewart [1991], Casti [1994], Gell-Mann [1994], Barrows [1992] – to mention but some.

[4] E.g. Bennett [1982], Zurek [1989], [1990a].

[5] See e.g. Ford [1983], [1989]; Shuster [1988]. For a critical discussion, see Winnie [1992] and Batterman [1993].

[6] See e.g. the final chapter of Calude [1994] and the numerous references given there.

[7] I thus strongly disagree with Fred Dretske, in his *Knowledge and the Flow of Information* [1981]. For a detailed and thoroughgoing criticism of Dretske, see Loewer [1982], [1987]; cf. Putnam [1994].

[8] As this theory, which is a sort of standard theory of information, is very well known, and I am not actually discussing it on the later parts of this paper, I'll give only a very short review. For details of the theory, see e.g. Ash [1965], Feinstein [1958], Gallagher [1968]; cf. Cherry [1952], [1957].

[9] As it happens, the idea of calling the quantity thus defined "entropy" was not Shannon's own idea but suggested to him by John von Neumann; Shannon had thought of using "uncertainty". I agree completely with Wicken [1987] that this was a very unhappy christening that has only caused confusion.

[10] For a detailed and penetrating critical examination, see Bar-Hillel [1955].

[11] Although, no one has ever, to my knowledge, expressed the difference between these two theories in terms of use and mention; I find this both surprising and regrettable, for I think this basic distinction greatly helps to clarify the issue.

[12] Carnap and Bar-Hillel [1952]; Bar-Hillel [1952], [1955]; Hintikka [1968], [1970a]. See also Hintikka's later philosophical reflection in his [1993].

[13] This is indeed the general practise in this field nowadays. But I must grant that this statement is unfairly simplifying with respect to Kolmogorov's original reasoning via conditional complexity that led him to equate the two concepts. I shall discuss this issue thoroughly in 4.2.

[14] An useful concise survey is Cover et al. [1989]. For comprehensive surveys, see Li and Vitányi [1993] (this textbook is certainly the standard exposition of this field), Calude [1994]; cf. Chaitin [1987].

[15] More formally, Chaitin's revised definition goes as follows:

$$H_C(t/s) =_{df} \mathit{min}\ l(p) : (C(p, s^*) = t),$$

where in turn

$$s^* =_{df} \mathit{min}\ p : (U(p, \Lambda) = s.$$

[16] That is, most binary strings are "random" i.e. have the algorithmic complexity approximately equal to their lengths. For further details on compressiblity and incompressibility, see e.g. Li and Vitányi [1993], sections 2.2. and 3.2.

[17] Very similar line of reasoning can be found, e.g., in Li and Vitányi [1993], p. 140.

[18] Note that Kolmogorov's notation is slightly nonstandard; his $S(n, p)$ corresponds to the more usual $U(p, n)$, i.e. Kolmogorov writes input before program.

[19] The Padding Lemma is a basic fact of recursive function theory (see e.g. Odifreddi [1987], Proposition II.1.6, p. 131): it says that given one index of a partial recursive function, one can effectively generate infinitely may other indices of the same function. It follows e.g. from the possibility of adding to a Turing machine program arbitrarily many redundant commands with no effect on the computation.

[20] This fact follows easily from Turing's [1936-7] classical undecidability result on the halting problem.

[21] It is well known that there is no effective general method for finding the minimal program of a given string (in fact, also this can be derived easily from the undecidability of the halting problem).

[22] Cf. note 16.

References

Ash, Robert (1965) *Information Theory*, John Wiley & Sons, Inc., New York.

Bar-Hillel, Yehoshua (1952) "Semantic Information and its Measures", *Transactions of the Tenth Conference on Cybernetics*, Josiah Macy Jr. Foundation, New York, 33–48. Reprinted in Bar-Hillel (1964), chapter 17.

—— (1955) "An Examination of Information Theory", *Philosophy of Science* 22, 86–105. Reprinted in Bar-Hillel (1964), chapter 16.

—— (1964) *Language and Information, Selected Essays on Their Theory and Application*, Addison Wesley and The Jerusalem Academic Press, Reading, Mass. & Jerusalem.

Barrow, J. D. (1992) *Pi in the Sky*, Oxford U. P, Oxford.

Batterman, Robert W. (1993) "Defining Chaos", *Philosophy of Science* 60, 43–66.

Bennett, Charles H. (1982) "The Thermodynamics of Computation – A Review", *International Journal of Theoretical Physics* 12, 905–940.

Brillouin, Leon (1956) *Science and Information Theory*, Academic Press, New York.

Calude, Christian (1994) *Information and Randomness*, Springer-Verlag.

Carnap, Rudolf (1934) *Logische Syntax der Sprache*, Springer, Wien.

Carnap, Rudolf and Yehoshua Bar-Hillel (1952) "An Outline of a Theory of Sematnic Information"; *Technical Report No. 247*, M.I.T. Research Laboratory in Electronics. Reprinted in Bar-Hillel (1964), chapter 15.

Casti, John L. (1994) *Complexification*, HarperCollins, New York.

Chaitin, Gregory J. (1966) "On the length of programs for computing binary sequences", *Journal of the ACM* 13, 54–569.

—— (1970) "On the difficulty of computations", *IEEE Transactions on Information Theory* IT–16, 5–9.

—— (1974a) "Information-theoretic computational complexity", *IEEE Transactions on Information Theory* IT–20, 10–15.
—— (1974b) "Information-theoretic limitations of formal systems", *Journal of the ACM* 21, 403–424.
—— (1975) "A Theory of Program Size Identical to Information Theory", *Journal of the ACM* 22, 329–340.
—— (1982a) "Algorithmic information theory", *Encyclopedia of Statistical Sciences*, volume 1, Wiley, New York, 38–41.
—— (1982b) "Gödel's theorem and Information", *International Journal of Theoretical Physics* 22, 941–954.
Chaitin, Gregory J. and Jacob T. Schwartz (1978) "A note on Mote-Carlo primality tests and algorithmic information theory", *Communications on Pure and Applied Mathematics* 31, 521–527.
Cherry, Colin (1952) "The Communication of Information", *American Scientist* 40, 640–664.
—— (1957) *On Human Communication*, M.I.T. Press, Cambridge, Mass.
Cover, Thomas M, Peter Gacs and Robet Gray (1989) "Kolmogorov's Contributions to Information Theory and Algorithmic Complexity", *The Annals of Probability*, Vol. 17, No. 3, 840–865.
Dretske, Fred (1981) *Knowledge and the Flow of Information*, Blackwell, Oxford.
Feinstein, A. (1958) *Foundations of Information Theory*, McGrav-Hill Book Co., New York.
Ford, Joseph (1983) "How Random is a Coin Toss", *Physics Today* 36, 40–47.
—— (1989) "What is Chaos that We Should be Mindful of It?", in P. Davies (ed.) *The New Physics*, Cambridge, Cambridge University Press, 348–371.
Frege, Gotlob (1893) *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*, I. Band, Verlag von H. Pohle, Jena.
Gallagher, Robert G. (1968) *Information Theory and Reliable Communication.*
Gell-Mann, Murray (1994) *The Quark and the Jaguar*, W. H. Freeman and Company, New York.
Grice, Paul (1957) "Meaning", *Philosophical Review* 66, 377–388.
Gödel, Kurt (1931) "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I", *Monatshefte für Mathematik Physik* 38:173–198.
Hartley, R.V.L. (1928) "Transmission of Information", *Bell System Technical Journal* 7, 535–563.
Hintikka, Jaakko (1968) "The Varieties of Information and Scientific Explanation", in B. van Rootselaar and J.F. Staal (eds.) *Logic, Methodology and Philosophy of Science III*, North-Holland, Amsterdam.
—— (1970a) "On Semantic Information", in Hintikka and Suppes (1970), 3–27.
—— (1970b) "Surface Information and Depth Information", in Hintikka and Suppes (1970), 263–297.
—— (1993) "On proper (popper?) and improper uses of information in epistemology", *Theoria* LIX, Parts 1–3, 158–165.
Hintikka, Jaakko and Patrick Suppes (eds.) (1970) *Information and Inference*, D. Reidel, Dorecht.

Kolmogorov, Andrei N. (1965) "Three approaches for defining the concept of information quantity", *Problems of Information Transmission* 1(1), 1–7.
—— (1968) "Logical basis for information theory and probability theory", *IEEE Transactions on Information Theory* IT–14, 662–664.
—— (1970/1983) "Combinatorial foundations of information theory and the calculus of probabilities", *Russian Mathematical Surveys* 38, no. 4, 29–40. (the text was written in 1970)
Li, Ming and Paul Vitányi (1993) *An Introduction to Kolmogorov Complexity and Its Applications*, Springer-Verlag, New York.
Loewer, Barry (1982) "Review of Dretske", *Philosophy of Science* 49, 297–300.
—— (1987) "From Information to Intentionality", *Synthese* 70, 287–317.
Nyquist, H. (1924) "Certain Factors Affecting Telegraph Speed", *Bell System Technical Journal* 3.
Odifreddi, P. (1987) *Classical Recursion Theory*, North-Holland, Amsterdam.
Pagels, Heinz R. (1988) *The Dreams of Reason*, Simon and Schuster, New York.
Peirce, Charles S. (1867) "On a New List of Categories", in *Collected Papers*, Vol. I, Chapter 6.
Popper, Karl R. (1934) *Logic der Forschung*, Springer, Wien.
Putnam, Hilary (1994) "Probability and the Mental", in H. Putnam: *Words and Life*, Harvard University Press, Cambridge, Mass, Chapter 19, 376–388.
Quine, W.V. O (1940) *Mathematical Logic*, Norton, New York.
—— (1987a) "Gödel's Theorem", in W,V. Quine: *Quiddities*, The Belknap Press of Harvard University Press, Cambridge, Mass., 83–85.
—— (1987b) "Use versus Mention", in W,V. Quine: *Quiddities*, The Belknap Press of Harvard University Press, Cambridge, Mass., 231–235.
Rucker, Rudy (1987) *Mind Tools – The Mathematics of Information*, Houghton Mifflin, Boston.
Ruelle, David (1991) *Chance and Chaos*, Princeton U. P., Princeton.
Russell, Bertrand (1922) "Introduction", in Wittgenstein [1922].
Schumacher, J. M. (1993) "Information and Entropy", *CWI Quarterly*, Vol. 6, No. 2, 97–120.
Shannon, Claude E. (1948) "A Mathematical Theory of Communication" *Bell System Techical Journal* 27, 379–423, 623–656.
Shannon, Claude E. and Warren Weaver (1949) *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana.
Shuster, H. G. (1988) *Deterministc Chaos: An Introduction*, 2nd revised edition, VCH Publishers, New York.
Solomonoff, Ray J. (1964) "A Formal Theory of Inductive Inference", *Information and Control* 7, 1–22, 224–254.
Stewart, Ian (1992) *The Problems of Mathematics*, Oxford U. P, Oxford.
Svozil, Karl (1993) *Randomness and Undecidability in Physics*, World Scientific, Singapore.
Tarski, Alfred (1933) "The Concept of Truth in Formalized Languages", in Alfred Tarski: *Logic, Semantics, Metamathematics: Papers form 1923 to 1938*, Clarendon Press, Oxford, 152–278.

—— (1944) “The Semantic Conception of Truth”, *Philosophy and Phenomenological Research* 4, 341–375.
Turing, Alan M. (1936–7) “On computable numbers, with an application to the Entscheidungsproblem”, *Proceedings of the London Mathematical Society* (2) 42, 230–265; correction, *ibid.* 43, 544–546.
Wicken, Jeffrey S. (1987) “Entropy and Information: Suggestions for Common Language”, *Philosophy of Science*, 54, 176–193.
Wiener, Norbert (1948) *Cybernetics, or Control and Communication in the Animal and the Machine*, M.I.T. Press, Cambridge, Mass.
Winnie, John A. (1992) “Computable Chaos”, *Philosophy of Science* 59, 263–275.
Wittgenstein, Ludvig (1922) *Tractatus Logico-Philosophicus*, Routlege and Kegan Paul, London.
Zurek, W. H. (1989) Thermodynamic cost of computation, algorithmic compexity and the information metric”, *Nature*, Vol. 341, 14. September 1989, 119–124.
—— (1990a) “Algorithmic Information Content, Church–Turing Thesis, Physical Entropy, and Maxwell's Demon”, in Zurek (ed.) (1990b), 73–89.
—— (ed.) (1990b) *Complexity, Entropy, and the Physics of Information*, SFI Studies in the Sciences of Complexity, Vol VIII, Addison-Wesley.
Zvonkin, A. K. and Levin, L. A. (1970) “The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms”, *Russian Mathematical Surveys* 25, 83–124.