The cooked and the raw*

Robert Hammarberg
School of Library Science, The University of Michigan, Ann Arbor, MI 48109, U.S.A.

Received 10 September 1981

The notion of data and information being different in kind is based on the Foundationalist thesis that there exist raw, brute, facts which constitute the data that form the basis for information, and, ultimately, of knowledge. Foundationalism fails, however, because if these data are really different in kind from information and knowledge, then no comparisons are possible between the former and the latter, and the notion of data (in the sense of raw, brute, fact) becomes useless. If, on the other hand, comparisons *can* be made between data and information or knowledge, the data would have to be of the same *kind* as the latter. Given that an Information Processing System (IPS) cannot process data except in terms of whatever representational language is inherent to it, data could not even be apprehended by an IPS without becoming representational in nature, and thus losing their status of being raw, brute, facts.

The representational language of the IPS provides the categories in terms of which the IPS 'views' reality, and thus this language will define what constitutes reality for the IPS in question. Consequently, this language will define what constitutes signals, signs, and information for the IPS, as well. Any definition of information must therefore be relative to a given IPS, and, in the case of a human IPS, what is regarded in common sense terms as linguistically encoded information cannot be independently characterized in purely physical terms.

1. Definability of information

Every so often, it seems, information scientists are beset by an urge to take stock of their enterprise and begin asking themselves "What is information?" The question is always asked with some trepidation, for in the asking is implied that the information scientist, whose business it is to speak knowledgeably about information, might not really know what it is he is talking about. And, when it then turns out that a good answer to the question is awfully hard to come by, there is despair and doubt about Information Science (IS) ever amounting to a 'real' science [1–9].

Now it may well be the case that IS is in a preparadigmatic stage, and thus not quite deserving of the exalted label 'science' [10]. But the reason for this ought not be that IS has not been able to define exactly what information is. After all, the life sciences have not managed to define life, and physics – defined as the science of matter and energy – has not been able to give us definitive definitions of matter and energy. So in this regard IS is certainly in good company. It seems likely that it is the search for the definition that constitutes science, and that definitive definitions, like absolute truths, would (if they exist at all) be the terminal points of science, not the beginnings.

Yet it is clear that if we are to investigate information we can not start from a null definition either. We shall have to have some notions about the nature of the beast in order to know where and how to look for it. In cases such as this, it is usually best to start out from some common sense definition, something that will help us get onto the track, though it may prove wrong in the long run [11]. When one solicits such common sense definitions – even those of sophisticated common sense – one is likely to encounter two kinds. The first would be definitions of a statistical nature, involving uncertainty and entropy, derived from Shannon and Weaver [12]. However, this type of characterization of information is not much in vogue anymore, since it leaves out what most ISists now see as a necessary ingredient in information, namely meaning. Shannon's and Weaver's is a signalling theory. ISists would be interested in what is conveyed by the signal. Therefore, so long as one does not (*pace* McLuhan) regard the medium as the message, it might turn out to be a serious case of conceptual confusion to equate the signal with what it conveys [1,3,13–14]. On the other hand, one may well question (and I shall, later) whether anything can be considered to be a signal without having to refer to meaning. Doing so would, obviously, imply a concept of 'signal' different from Shannon's and Weaver's.

By taking meaning to be a necessary ingredient of information one drastically narrows the scope of the search. Meaning is something inextricably connected with human beings, and information thus comes to be something that necessarily involves a human ingre-

* With perfunctory apologies to Claude Lévi-Strauss.

North-Holland Publishing Company
Journal of Information Science 3 (1981) 261–267.

0165-5515/81/0000–0000/$02.75 © 1981 North-Holland

dient. The second kind of definition of information that we encounter is human-oriented in this respect, for it relates information to another human phenomenon, namely knowledge. But it also brings in what is claimed to be a human-independent factor, namely data. This second view, which is likely to be the majority view, would look something like the following: Data are unprocessed, raw facts. Information, on the other hand, is something more refined. If data are raw, then information has been cooked. And, if we continue the culinary metaphor, knowledge would be something akin to the nourishment that is derived from the concoction. (For common sense views along these lines, see Letters to the Editor in *Online* 3(3) (1979).)

Two things are to be noted about this second view. First, if this is what IS is concerned with, then its subject matter overlaps to a considerable degree with that of cognitive psychology and epistemology in that all three are concerned with the question of how knowledge is attained. Second, both information and knowledge are seen as derivative of data, and presumably also definable only in relation to data. Data, however, should be intrinsically definable. In taking this approach to the question of knowledge attainment, IS would be travelling a route that is very familiar to psychology and epistemology, namely the route of Empiricist Foundationalism.

2. Foundationalism

Foundationalism gets its name from the thesis that there exists a rock-bottom foundation upon which knowledge is based. Empiricist Foundationalism holds this foundation to consist of raw, brute, factual data about the world [15]. Data thus come to occupy a very special position: data are foundational in virtue of being unprocessed givens, and thus different *in kind* from either information or knowledge. It is this latter thesis, that data are different in *kind* from information and knowledge, that is troublesome.

At first blush, Foundationalism seems an extremely attractive hypothesis, and the denial of its central thesis, the existence of foundational data, would seem to fly in the face of common sense. The existence of foundation data is, it is claimed, necessary for the concepts of objectivity and truth (and all that goes therewith) to be meaningful. If we are faced with contradictory information, the contradiction should be resolvable by appeal to the data, 'the facts'. And how do we attain factual information about the world if not by apprehending matters of fact as they exist, i.e., by apprehending the data? And last, but not least, Foundationalism holds out the promise that knowledge might be attained from data by means of an algorithm. If that were the case, the attainment of knowledge could possibly be automated and made more efficient than it is now when we are largely dependent on fallible human processing. This latter aspect of Foundationalism is one that is cherished by organizations such as business and government who view knowledge in terms of power. Once this idea takes root, it follows naturally that data, the raw material for information and knowledge, should be regarded as a valuable resource to be gathered and carefully managed.

This suggests a functional definition of data: 'data' is that which serves as raw material for a process that yields information [16]. We would, in other words, have something that could be represented by the conventional view of an Information Processing System (IPS) which accepts data as input which then is processed into information. But a functional definition will tell us nothing about the intrinsic properties of data that would allow us to recognize a datum as being different in kind from information. The Foundationalist would tend to regard an IPS as an interface with reality, and reality as an aggregate of data.

It is at this juncture that the sands begin to shift under Foundationalism. The Foundationalist can now opt for Naive Realism, according to which the entities, properties, and relations of the world (i.e., the data) somehow enter the IPS *directly, in their intrinsic form.* Under these circumstances the data would undoubtedly be different in kind from anything processed by the IPS. However, it should not require more than a moment's reflection to realize that this option is absurd. But the alternative would have to be some form of indirect Realism (provided, of course, that one is committed to *some* form of Realism), and then the contention that there could be a categorical distinction in kind between data and their supposed derivatives becomes quite dubious.

Imagine being an IPS. What, now, would be your view of the data, i.e., the input? Would you have any way of transcending your perceptions? What you are aware of appears to be simply given [17], in the sense that you would be conceiving of it in some kind of terms, and you could not choose between alternative terms from some term-free point of view. In other words, no IPS could have access to matters

not reprensentable in the 'language' in terms of which the IPS functions. Matters not representable are not accessible, and matters accessible are so only in virtue of being presented in the language of the IPS. Thus from the point of view of any IPS, *its* data are going to appear ultimate to *it* – not because of any inherent qualities of its 'perceptions', but simply because it cannot 'see things' in any other way.

The fact of these matters seems to be that an IPS – any IPS, including one that is a sentient being – is a prisoner of its own representational processes: We can never escape a point of view. This, of course, was Kuhn's [18] message: we are paradigm-bound – and not only in doing science, but in all our cognitive-perceptual activities. Now, consider again the basic Foundationlist tenet, that the data are the rock-bottom facts. They would be rock-bottom in virtue of not having been transmuted into any kind of representational form. Thus the truth of a proposition, a piece of information, could be judged by its degree of correspondence with the data. But, in order for the two (the data and the information) to be at all comparable, they shall have to be of the same general kind. They would both have to be compared within one and the same representational framework, otherwise there could be no basis for comparison (this is the old apples-and-oranges problem). But this, of course, defeats the assumption that data and information would be different in kind.

The Foundationalist's dilemma is this: in order to be foundational, data must be raw and untainted by any processing. But if they are, they cannot be apprehended by an IPS, for it can apprehend only in terms of its representational language. However, if the data are represented in the IPS, they are no longer unprocessed and raw. Once the data – taken as the brute facts of 'outside' reality – are brought 'inside' the IPS, they would have to be represented in terms of whatever representational language the IPS employs. This makes comparisons between information and 'data' possible, but 'data' has now lost its essential property of rawness.

A Foundationalist position on the nature of data seems untenable, and with it goes the notion of absolute objectivity. Data are not different in kind from information, not raw, not "inherently inviolate and ··· informatively committal about the nature of the world" [19]. What we take to be data cannot be singled out as possessing any special characteristics that set them apart from information. Thus they can not be characterized intrinsically, but only extrinsically as a starting point for some cognitive venture (and among such ventures should be included all scientific activities such as theory formation and hypothesis testing). But being a starting point is not the same as being a foundation, for a starting point is not sacrosanct – the edifice can stand even if the starting point is discredited. Not so in the case of foundations.

This seems to be the way the modelling of reality actually works (as opposed to how the Foundationalist thinks it *ought* to work). What is considered to be given (data) at the beginning of an investigation may be altered and perhaps even completely abandoned as the investigation progresses. Under these circumstances, what is regarded as truth and reality cannot be a matter of correspondence, as under Foundationalism, for there is nothing foundational to correspond to. Instead, truth becomes a matter of coherence: the representations must be internally coherent. In the case of human beings, the coherence should ideally apply to the totality of one's beliefs. (For more on these ideas, see [20]).

3. Communication

Considerations of how data, information, and knowledge could interact thus lead to the conclusion that all three should be referred to the category of the cooked. But then, all three would be purely intra-personal in nature. This is, of course, problematic for those who regard IS as concerned with communication, with the inter-personal transmission and extra-personal storage of something they call 'information'. The existence of communication is, of course, not in doubt, and communication does not proceed by means of magic or telepathy, but via a physical bridge, some extra-personal medium. We speak of information and knowledge being stored in various forms, occurring in print, being found in journals, and so on. Information and knowledge are also said to be transmitted orally from one person to another by means of speech. The question is whether such common sense claims can be backed up by any coherent characterization of this physical bridge – whether information and/or knowledge can be given meaningful physical definition, even in principle.

For something to be communication, it must be intentional in nature. When I see certain tracks in the snow, I may come into possession of the

information that a wolf has passed by my house. But this is not something that the snow (nor the wolf, for that matter) is communicating to me [21]. In dealing with written or oral language the question of intentionality seldom becomes acute. When confronted with something we take to be linguistic in nature, we assume that it is intentional and meaningful (even in the cases where nobody understands it, as in the case of Etruscan writings). The assumption is that a meaningful 'message' has been intentionally encoded in a physical medium, and that it can be decoded by anyone who knows the code.

But how do we know that something is linguistic in nature; how do we know that some physical configuration or event constitutes an encoded message? It is widely assumed there is something in the physical phenomenon itself that announces its informational nature. A further assumption is usually made as well, namely that the physical phenomena that have these informational features stand in a causal relationship to the cognitive phenomena of information and knowledge (c.f. [2]). If this is the case, then it should be possible to isolate those features that define informational content and study them by the methods of physics. From here the step is not great to the claim that information in all its forms can be reduced to physics. By now we seem to have slid over into a signalling theory, where certain physical features (or feature complexes) would serve as foundational data.

4. The proposition of information as a physical surrogate of knowledge

In a series of recent articles, Farradane has proposed that information be defined as "a physical surrogate of knowledge (e.g. language) used for communication" [22–23]. In fact, he holds a physical definition to be necessary if IS is ever to become an experimental (and by implication, a 'real') science:

> If information science is to be at all an experimental science, one must have some observable elements or phenomena which can be isolated for initial study, and proceed from these to the more complex and difficult phenomena related to them [22].

There is a clear Foundationalist bias in evidence here. Farradane wants "information" in the form of "the written or spoken surrogate of knowledge" to be the "invariant starting point" for further investigations. The invariance of the starting point, and hence its foundational nature, is to be guaranteed by its being physically defined. The appeal to physics is typical of Foundationalism. Physics is taken to be the ultimate science to which all natural phenomena are reducible, dealing as it (supposedly) does with the absolute rock-bottom, raw, brute, facts of the world.

There is a distinction (one that Farradane is aware of [6]) that must be made here for the sake of making clear what it is we are dealing with. There is an ambiguity in the word 'information', in that it may signify, on the one hand, a cognitive phenomenon, and, on the other hand, certain physical phenomena that are somehow related to the cognitive ones. Analogous ambiguities are found in the case of 'color' and 'sound'. A color may be regarded as electromagnetic radiation of a certain wavelength, or it may be regarded as a phenomenal quality, something that is experienced. Similarly, a sound may be an acoustic waveform, or something that is heard (phenomenal sound). In either case, it is obvious that experienced colors and sounds could not exist outside the human mind. The difference in kind between the phenomenal and the physical can be made vivid by the realization that experienced pains are nothing like that which supposedly causes them, e.g., the sharp tip of a needle pressed against your skin. Information can hardly be said to be a phenomenal quality, but information as 'cognized' could not exist outside the human mind either, and would clearly be different in kind from anything outside the mind that one might try to correlate with it.

For a definition such as Farradane's to work, one should have to postulate the existence of a biunique relationship (if and only if) between phenomenal and physical events, categories, and entities. Moreover, the relationships of type-identity and similarity expressed in terms of phenomenal natural classes (e.g., same color, similar color) should be explicable in terms of their correlations with natural classes of physics. Only under such conditions of biuniqueness would it make sense to study and experiment with Farradane's "observable elements or phenomena" and expect them to have any bearing on the study of (supposedly) corresponding psychological events. This is the essence of physicalistic reductionism: that all psychological events and categories could be shown to be coextensive with physical events and entities [24].

Consider the role of meaning in information. Meaning is a necessary ingredient of information, but what would be a physical surrogate of meaning? Meaning is not to be found in any purely physical phenomenon. Rather, meaning is *given* to, ascribed to, certain events, and it is only in virtue of such ascriptions that they become signs or signals. Nothing is a sign or symbol in and of itself; certainly a mere variation in some physical magnitude is not sufficient for something to be a signal. It must be kept in mind that perception is not a passive process of receiving impressions from without. Rather, perception is a constitutive process of individuation and categorization, involving cognitive aspects such as judgement. It is this active nature of perception that makes correlations between cognitive-perceptual categories and 'the outside world' so intractable.

5. Language

Let us now take a look at language, since this is what Farradane suggests as an example of what he has in mind, and since meaning, of course, belongs in the domain of the linguistic. Farradane is clearly wrong in taking language to be a surrogate for knowledge. First, language is not a surrogate for anything, and second, all information is certainly not knowledge. Farradane has apparently fallen victim to another instance of the kind of ambiguity we just discussed above. Language – the real thing – is a cognitive phenomenon, and does not exist outside the human mind [25]. But language can be given physical representation, graphically (writing) or acoustically (speech). It is these physical representations (or surrogates, if you wish) that Farradane wants to regard as information.

What, then, is the relationship between (cognitive) language and its 'physical surrogates'? The answer is, that no one has been able to establish any constant correlations between any linguistic categories or classes and any physical events or entities. There is nothing in the acoustic stream of speech that corresponds to perceived entities such as words or speech sounds (phonemes), not to mention such higher level entities as sentences. And, as for graphic surrogates for language, just consider what you are looking at right now. Physically, what is there is a collection of streaks of ink on paper. Now, would it be possible to establish any coherent correlations between a piece of information such as, say, "My dog is bald", and streaks of ink? Just consider how many different kinds of graphical representation this particular piece of information could have. It could be represented in different languages, in different alphabets (Latin, Cyrillic, Arabic, etc.), syllabaries, ideograms, in Morse code, in different handwritings, in ink, pencil, different type fonts, upper case or lower, etc., etc., etc. Are there any natural classes of physics that could be established among these variations that could be shown to correlate with the information that "My dog is bald"? Not likely! And could it be shown that the relation between this information and the information that "My cat is bald" would be due to a systematic alternation between two classes of physical events? Less likely still!

To be scientifically useful, the thesis of physical definability of information would have to mean that occurrences of the same information should be correlatable with an invariant set of physical defining features, such that pieces of different information would be different exactly in virtue of being correlated with different features, whereas, on the other hand, tokens of the same piece of information would be correlated with the same set of features. Thus the essential taxonomic function of a science, to show how the sameness and difference between tokens of information could be accounted for, would be fulfilled. No such sets have ever been found, and there is little reason to believe that any every will. The reason for this is that the properties constituting such a set would not form a natural class of physics. The set would, in the case of the information that "My dog is bald", be disjoint in physical terms, in that it would contain acoustic features (for the spoken word) and features pertaining to streaks of ink on paper (for the written word), and so on. And the only criterion for separating these acoustic features and streaks of ink from other acoustic features and streaks of ink would not be physical at all, but rather semantic. Thus the set of physical features would constitute a set *ONLY* in virtue of having been correlated with the same information (i.e., the semantic content of the sentence), not on the basis of any physical criteria. In other words, the natural classes of information (what makes for informational sameness or difference) are not coextensive with any natural classes of physics. Thus a physical definition of information could not stand alone, but would always have to be created on the basis of meaning [26–27].

In the final analysis, the problem lies in the assumption that physics is epistemically neutral, that it deals with the raw, brute facts of the world. But

physics is, of course, just as theory-bound as any other human endeavour, and its statements are cast in representational form which always reflects a point of view. And the forms that have been found to be useful in physics just do not seem to dovetail too well with the forms that are useful in describing language and phenomenal qualities. Physics is not a "binary science", as Pearson [28] would have it. A physical relation such as 'longer than' is not simply a relation between two objects, but must (as we have learned from Special Relativity) specify the frame of reference in which the observation of length takes place. Length is a relative concept that only reveals a relation between the observer and an object. And as to the existence of objects, those quintessential entities of Foundationalist speculation, they too might simply be creatures of our point of view. This, at least, is the view of the quantum physicists:

> The doctrine that the world is made up of objects whose existence is independent of human consciousness turns out to be in conflict with facts established by experiment [29].

Thus the Kantian claim that objects conform to our modes of cognition has re-emerged in modern science:

> The 'furniture of the world' does not come prepackaged in the form of individuals with properties, apart from human intervention: Either the analysis provided by the cognitive system that we might call 'common sense understanding' or the more self-conscious idealizations of the scientist seeking to comprehend some aspect of physical or mental reality [30]. (C.f. also [31].)

To insist on proceeding from raw things-in-themselves (alias 'data') toward information and knowledge would thus seem rather futile.

6. Knowledge

In conclusion, let me say a few words about knowledge. Strictly speaking, what is characteristic of knowledge is that it is always true. Consider the sentence

Fritz knows that the earth is flat.

This sentence is clearly anomalous, and the anomaly lies in the violation of the requirement that the verb 'to know' have as its object a proposition that is true [32]. This is a view shared by Foundationalist and non-Foundationalist alike. The difference between the two lies in their differing criteria for truth: Is it a matter of coherence or correspondence? In any case, a clear difference can be made between information and knowledge: We may have false information, but to speak of false knowledge is like speaking of false truths – it is a *contradictio in adjecto*. At any given time we entertain numerous propositions that may be true or false. These are *beliefs*. Beliefs, too, may turn out to be false, but those that are true and justified constitute knowledge. Thus 'knowledge' would be a proper subset of 'belief'. This distinction may not be of any immediate relevance to IS, since physical definition of 'belief' is as remote as any other. Nevertheless, since we are groping about for definitions, this is a distinction that should at least be noted.

Notes

[1] H. Wellisch, From information science to informatics: A terminological investigation, J. Librarianship 4(3) (1972) 157–187.
[2] V. Rosenberg, The scientific premises of information science, J. Amer. Soc. Inform. Sci. 25(4) (1974) 263–279.
[3] B.C. Brookes, The fundamental problem of information science, Informatics 2. Proceedings of a conference held by the Aslib Co-ordinate Indexing Group on 25–27 March, 1974, at New College, Oxford. (Aslib, London, 1975) pp. 50–56.
[4] G. Wersig and U. Neveling, The phenomena of interest to information science, Inform. Scientist 9(4) (1975) 127–140.
[5] C.I. Barnes and J.H. Petrie, Definitions of information science, Inform. Scientist 10(3) (1976) 122–124.
[6] J. Farradane, Towards a true information science, Inform. Scientist 10(3) (1976) 91–101.
[7] N.J. Belkin, Information concepts for information science, J. Documentation 34(1) (1978) 55–85.
[8] J. Pemberton, The inverted file, Online 3(2) (1979) 6–7.
[9] Information science? (Letters to the Editor), J. Inform. Sci. 2(6) (1980) 313.
[10] N.J. Belkin, Towards a definition of information for informatics, Informatics 2 (c.f. [3]) pp. 42–49.
[11] B.C. Brookes, The foundations of information science, Part I: Philosophical aspects, J. Inform. Sci. 2(3/4) (1980) 125–133.
[12] C.E. Shannon and W. Weaver, The Mathematical Theory of Communication (University of Illinois Press, Urbana, IL, 1949).
[13] J. Hintikka, On semantic information, in: J. Hintikka and P. Suppes, Eds., Information and Inference (Reidel, Dordrecht, 1970) pp. 3–27.
[14] B. Langefors and K. Samuelson, Information and Data in Systems (Petrocelli/Charter, New York, 1976) p. 93.

[15] There are several different kinds of Foundationalism, depending on what is held to be foundational. Some hold the foundations of knowledge to be divine inspiration, others believe them to be "clear and distinct ideas". Empiricist Foundationalism holds experience, derived from or via sense data, to be foundational, and that is the only kind of Foundationalism that shall concern us here.
[16] B. Langefors and K. Samuelson, op cit. p. 94.
[17] The word 'datum' (plural 'data') is the participial form of Latin 'dare' = to give.
[18] T. Kuhn, The Structure of Scientific Revolutions, 2nd edition (The University of Chicago Press, Chicago, 1970).
[19] N. Rescher, Dialectics (State University of New York Press, New York, 1977) pp. 33–34.
[20] N. Rescher, The Coherence Theory of Truth (Clarendon, Oxford, 1973).
[21] Here the Foundationalist will suggest that the tracks in the snow constitute raw sensations, data from which information can be derived. But seeing something as a track (let alone a wolf-track), and seeing something as snow requires cognitive acts of classification. And attempts to eliminate everything of a cognitive nature will arrive at nothing left to fill the role of 'data'. "Elementary sensations exist only by virtue of abstraction from the perceptual world, and therefore ... any propopsed division into sensation and perception must be artificial and misleading". See G. A. Miller and P.N. Johnson-Laird, Language and Perception (Harvard University Press, Cambridge, 1976) p. 12.
[22] J. Farradane, The nature of information, J. Inform. Sci. 1(1) (1979) 13–17.
[23] J. Farradane, Knowledge, information, and information science, J. Inform. Sci. 2(2) (1980) 75–80.
[24] J.A. Fodor, The Language of Thought (Crowell, New York, 1975) Ch. 1.
[25] N. Chomsky, Language and Mind, enlarged edition (Jovanovich, New York, 1972) p. 95.
[26] J.A. Fodor, Psychological Explanation (Random House, New York, 1968) Ch. 3.
[27] R. Harré, The Principles of Scientific Thinking (The University of Chicago Press, Chicago, 1970) Ch. 8.
[28] C. Pearson, The basic concept of the sign, in: A.R. Benenfeld and E. J. Kazlauskas, Eds., Communicating Information, Proc. 43rd ASIS Annual Meeting, 1980. (Knowledge Industries Publications, White Plains, NY, 1980) pp. 367–369.
[29] B. d'Espagnat, The quantum theory and reality, Sci. Amer. (1979) 158–181.
[30] N. Chomsky, Rules and Representations (Columbia University Press, New York, 1980) pp. 218–219.
[31] N. Chomsky, Reflections on Language (Pantheon, New York, 1975) Ch. 1.
[32] We are dealing here with propositional knowledge ("knowing that"), and not knowledge of the kind expressed in "I know English" or "I know how to ride a bike".