Data, information and knowledge: have we got it right?

Max Boisot and Agustí Canals

Abstract. Economists make the unarticulated assumption that information is something that stands apart from and is independent of the processor of information and its internal characteristics. We argue that they need to revisit the distinctions they have drawn between data, information, and knowledge. Some associate information with data, and others associate information with knowledge. But since none of them readily conflates data with knowledge, this suggests too loose a conceptualisation of the term ‘information’. We argue that the difference between data, information, and knowledge is in fact crucial. Information theory and the physics of information provide us with useful insights with which to build an economics of information appropriate to the needs of the emerging information economy.

Keywords: Information – Knowledge – Economics of information – Information theory – Physics of information

JEL Classification: A12, D20, D80, M21

1 Introduction

Effective cryptography protects information as it flows around the world. Encryption, by developing algorithms that bury information deeply in data, provides “the lock and keys” of the information age (Singh, 1999, p. 293). Thus while the data itself can be made “public” and hence freely available, only those in possession of the “key” are in a position to extract information from it (Singh, 1999). Cryptography, in effect, exploits the deep differences between data and information.

Correspondence to: A. Canals, Av. Tibidabo, 39-43. 08035 Barcelona, Catalonia, Spain
Knowledge and information are not the same thing, either. Imagine, for example, receiving an encrypted message for which you possess the key and from which you extract the following information: “The cat is tired”. Unless you possess enough contextual background knowledge to realize that the message refers to something more than an exhausted cat – possibly a Mafia boss, for example – you may not be in a position to react in an adaptive way. To understand the sentence is not necessarily to understand the message. Only prior knowledge will allow a contextual understanding of the message itself, and the message, in turn will carry information that will modify that knowledge. Clearly, then, information and knowledge must also be distinguished from one another.

In everyday discourse, the distinction between data and information, on the one hand, and between information and knowledge, on the other, remains typically vague. At any given moment, the terms data and information will be used interchangeably; whereas at another, information will be conflated with knowledge. Although few people will argue that knowledge can ever be reduced to data, the two terms are unwittingly brought into a forced marriage by having the term information act as an informal go-between. The growing commercial interest in cryptography, however, suggests innumerable practical circumstances in which the need to distinguish between the three terms is becoming compelling. But if the distinction works in practice, does it work in theory? This is the question that our paper addresses.

Beginning with the second half of the twentieth century, a number of economists – Koopmans, Marschak, Hurwicz, and Arrow – began to concern themselves with the nature of the economic agent as a “rational information processor”. Since that time, information has become acknowledged as the key generator of wealth in post-industrial societies. We might therefore reasonably assume that, over the past fifty years, mainstream economists, concerned as they are with wealth creation, would have developed a conceptual approach to information that reflected its growing importance to their field.

In this paper, we shall argue that they have some way to go. Both Stiglitz and Lamberton have noted how, even at the end of the twentieth century, the economic profession’s conviction that there can be an ‘economics of information’ still has to reckon with the lack of any consensus as to what specifically it should cover (Stiglitz, 2000; Lamberton, 1998). As Arrow has commented, “It has proved difficult to frame a general theory of information as an economic commodity, because different kinds of information have no common unit that has yet been identified” (Arrow, 1973, p. iii). In fact, Arrow believed that such units were undefinable (Arrow, 1996).1 Economics, then, is still looking for a way of thinking about information that is adapted both to its own analytical needs as well as to the needs of the emerging information economy.

For this reason, we can support Lamberton’s plea that we abandon a unitary and all-purpose concept of information and develop instead a taxonomy based on significant characteristics of information (Lamberton, 1996, pp. xx-xxii). How-

---

1 James Boyle has analyzed the incoherence of information economics over a period of fifty years in his Shamans, Software and Spleens (1996).