Information for Cognitive Agents

Nir Fresco

The Edelstein Centre,
The Hebrew University of
Jerusalem, Israel
fresco.nir@gmail.com

Abstract. Humans use information in everyday activities, including learning, planning, reasoning and decision-making. There is broad agreement that, in some sense, human cognition involves the processing of information, and, indeed, many psychological and neuroscientific theories explain cognitive phenomena in information-theoretic terms. However, it is not always clear which of the many concepts of 'information' is the one relevant to understanding the nature of human cognition. Here, I suggest that information should be understood pragmatically. Whatever the criteria for information are, what makes some *x* informational has to do with how an agent either processes or can process *x*. Information is defined as meaningful structured representations of perceptual data. Their meaningfulness is determined by their behavioural effect on the agent.

Keywords: Cognition; Information; Behavioural effect; Data; Cognitive Science.

1 Introduction

There is broad agreement that, in some sense, human cognition involves the processing of information. Humans regularly use information in learning, planning, reasoning and decision-making. Many theories in cognitive science explain cognitive phenomena in information-theoretic terms. Yet, 'information' means many things to many people. So, it is not always clear which of the many concepts of 'information' is the one relevant to understanding the nature of human cognition. C. Shannon and W. Weaver defined information-content as the probability of a message being selected from a finite set of messages with any selection being equally probable [1]. R. V. L. Hartley before them had developed measures for the capacities of different types of information systems to transmit information [2]. More recently, Kolmogorov Complexity has defined the information-content in a binary string s as the length of the shortest program that produces s on a universal Turing machine [3, 4].

However, all these offer quantitative analyses of information for measuring the information-content in a message, rather than a theory of information as the *thing* that is to be measured. As noted by Hartley, Shannon and Weaver, their theories focused on physical features of signal communication, rather than the psychological or semantic features of information. Whilst quantitative aspects of information-content are clearly of importance to an information-theoretic analysis of cognition, it seems crucial to fix the concept of semantic 'information' that is used by information theories of cognition in artificial intelligence and cognitive science broadly. In the next section, I survey a few of the well-known theories of semantic information and point out their deficiencies as the basis for informational theories of cognition.

In this paper, I suggest that information should be understood pragmatically first and foremost, if we are to understand human cognition information-theoretically. Whatever the criteria for information are, what makes data informational (for an agent) has to do with how the agent either processes or can process these data. (Here, I adopt L. Floridi's *data-oriented* definition of information [5] with important modifications as is discussed below.) Information should be best understood as meaningful structured representations of perceptual data as is discussed in Section 3. The meaningfulness of perceived data is determined by their behavioural effect on the agent as a triadic, rather than dyadic, relation

involving a physical object (or event or property or state of affairs), the agent's neural state and the behavioural effect on the agent. The account sketched here resembles other neo-Peircean analyses of representation [6, 7] as well as more recent accounts of information [8, 9]. The relationships between the present account and other neo-Peircean analyses are discussed in Section 4. Section 5 concludes the paper with some general reflections.

2 A brief survey of accounts of semantic information

An important principle underlying many probabilistic accounts of semantic information had been originally formulated by K. Popper. "[T]he amount of empirical information conveyed by a [set of sentences...] increases with its degree of falsifiability" [10]. This principle was later coined the *Inverse Relationship Principle* (IRP): the less likely a message is, the more informative (or rather *informational*) it is [11]. The first systematic theory of semantic information based on IRP was formulated by Y. Bar-Hillel and R. Carnap [12]. According to this theory, the thing that carries information or has informational content is *sentences*. The *meaningfulness* of information is relative to some logical probability space. Information is assigned to messages about events and the selected information measure depends on the logical probability of events or some properties of an object the message is about. Logical probability is defined in this context as a function of the set of possible worlds a sentence rules out.

Some have argued that this theory (and any other IRP-based theory) leads to a paradoxical result [5, 13]. If all the consequences of known sentences are known, any *logically* true sentence (that is, a tautology) does not increase knowledge and, hence, does not contain information. A tautology excludes no possible worlds and its logical probability is 1. At the same time, a self-contradictory sentence excludes all possible worlds and its logical probability is 0. Counter intuitively it contains *maximal* information. I return to this so-called paradox below, but for now, it should be noted that the Bar-Hillel/Carnap theory cannot serve as a basis for human cognition broadly. For it is defined in terms of sentences, and the domain of cognition is broader than language processing alone.

A more recent theory of information was offered by F. Dretske [14]. His theory is premised on the idea that information can be used as part of a reductive analysis of knowledge and cognition. On his view, the information carried by a message is relative to the epistemic state of the agent receiving that message. He was motivated by the central observation in the Shannon/Weaver theory that the receipt of information should reduce the agent's uncertainty. By applying the underlying communication model in the Shannon/Weaver theory to knowledge, the source of messages is the physical world and the receiver is a would-be knower. For Dretske, perceptual knowledge can (and should) be understood in terms of information. "K knows that s is F = K's belief that s is F is caused (or causally sustained) by the information that s is F" [14]. The information that s is F affects K's belief in such a way that the information suffices for the formation of the belief absent other contributing (or conflicting) factors. K must discern physical events in the world that carry the particular information, and those events have to cause (or causally sustain) K's belief that s is F. Moreover, the informational content of a message is also conditional on what K already knows when receiving the message. Importantly, Dretske maintained that information must be truthful. "Information is what is capable of yielding knowledge, and since knowledge requires truth, information requires it also" [14]. Other supporters of the idea that information must be truthful include P. Grice [15], J. Barwise [11] and P. Allo [13].

Floridi has adopted some of Dretske's main ideas (including the idea that information cannot be false), whilst rejecting IRP and insisting on a stronger constraint on semantic content. His two main motivations for adopting the Veridicality Thesis (i.e., that information must be truthful) are (a) to provide a link between information and *knowledge*, and (b) to avoid the Bar-Hillel/Carnap paradox concerning the alleged informativeness of contradictions [5]. The first motivation is similar in spirit to Dretske's in establishing a close link between knowledge and information. The second motivation – being that tautologies contain *no* information, whereas contradictions contain *maximum* information (an underlying principle of classical

logic) – has led him to deny IRP and suggest a stronger constraint that is based on closeness to truth. According to Floridi, "the amount of informativeness of each [message] can be evaluated *absolutely*, as a function of (a) [...] the alethic value possessed by [the message] and (b) the degree of discrepancy [...] between [the message] and a given state of the world" [5]. (Note the difference from Dretske's approach where information is conditional on the epistemic state of the receiver.)

Yet, besides the veridicality constraint, he proposes to understand information as meaningful and structured data. Unlike the Bar-Hillel/Carnap theory, information carriers are understood as data rather than sentences only. What is a datum? In its simplest form, it is the lack of uniformity in the real world. Examples of a datum include a black dot on a white page, the presence of some noise, a light in the dark or a logical 0 as opposed to a 1. A datum is defined as two distinct uninterpreted variables in a domain that is left open to further interpretation [5]. Data are structured when they are "rightly put together, according to the rules (syntax) that govern the chosen system, code or language being used. Syntax here must be understood broadly, not just linguistically" [16]. That they are meaningful means that the data "must comply with the meanings (semantics) of the chosen system, code or language in question. [...] The data constituting information can be meaningful independently of an informee [and need not be] necessarily linguistic" [16].

There are clearly other important theories of information that are worth exploring, but this exceeds the scope of this paper. For example, D. MacKay offered a quantitative theory of semantic information based on the receiver's increase in knowledge. "[W]e have gained information when we know something now that we didn't know before; when 'what we know' has changed" [17]. Another example is B. Skyrms' analysis of information – grounded in signalling games – where senders of signals observe states of the world and communicate with receivers that in turn choose an act in response to receiving signals [18]. For him, information is correlated with states of the world as well as with actions.

3 Towards a theory of semantic information as meaningful structured representations of data

Space only permits a few, brief remarks regarding the adequacy of the theories of information outlined in Section 2. (This is discussed elsewhere [19].) The Bar-Hillel/Carnap theory of information is defined in terms of sentences, and, thus, is unable to account for many non-linguistic informational aspects of cognition. Dretske and Floridi's accounts of information aim specifically at explaining knowledge. Yet, that objective has led them to adopt the Veridicality thesis that restricts the applicability of information to other cognitive phenomena. Cognitive agents cannot always ascertain the veracity of the information they process and one of the most important methods of learning is by trial and error that clearly involves making mistakes (or false information). The processing of information in cognitive agents is insensitive to the veridicality of the information. Belief change models, for example, explain rationality is terms of *justified* doxastic commitments that are *consistent*. These models are underpinned by the principle that all information, even veridical information, is defeasible and subject to revision under the right conditions. Besides, on standard frameworks of belief change, false perceptual information can actually lead to truth approximation via belief revision and increase the agent's overall knowledge base.

To underscore the *pragmatic* value of information for the receiving agent consider a simple example. Suppose that the *same* message is sent twice by the same information source. The two messages clearly carry the *same* information-content. Nevertheless, only the message that is *successfully received* by the receiver *first* is informative. Of course, receiving the second message – with the very same information content – can still be useful, for example, in the presence of noise: the first message could have been distorted during transmission. Moreover, in some contexts, each of the messages, arguably, carries additional meta-information that is its temporal indexing: message one was sent (or rather

received) at T_x and the second at T_y . This temporal indexing might also be pragmatically significant: it may tell the receiver that some state of the information source has remained unchanged. Nevertheless, all this is meta-information in addition to the information-content of each of the individual messages (e.g., if each message includes a timestamp as part of its content, the information-content of the two messages is different).

Crucial to the new theory sketched herein is the triadic basis of information. Rather than taking information to be a dyadic relation that obtains between signs and objects (or states of affairs) in the world, information requires a third element: its receiver. On Floridi's theory, for example, some information (i.e., environmental information) supposedly exists in the world independently of any receivers (e.g., concentric rings in the trunk of a tree that can be used to calculate the tree's age qualify as information even in the absence of any perceiver) [5]. But as argued by Dretske, the informativeness of a message is relative to the epistemic state of the receiving agent. Smoke in the forest (usually reliably) signifies there being fire to receivers of information that interpret the signals (smoke particles or combustion aerosols) as a potential imminent danger nearby. This triadic relation can already be found in the works of C. S. Peirce: something is a sign (also "representamen") only if it signifies an object with respect to an "interpretant" (i.e., a mediating representation in the mind of some agent) [20]. Whilst there is a causal correlation between smoke and fire based on natural regularities, the receiver of the signals (smoke particles) plays a key role in the formation of the information (there being fire in the forest). The receiver may know that smoke machines are used in the forest (for some bizarre reason) and, consequently, may not interpret the signals received as there being fire in the forest.

The theory proposed here uses Floridi's data-oriented definition of information with some important modifications. Objects, events or states of affairs in the world are sources of physical signals or data with which they are causally correlated. Physical data as discontinuities in the world exist "out there" unstructured. Their structuring is an ongoing dynamic interaction between the receiving agent and her environment. But data need not always originate externally to the receiver. An organism, for example, can receive pain signals from one of its limbs. Further, the structure of the data in the wild is determined by an agent-environment function. If either of these two contributing factors is missing, there is no information just data. In that sense, the physical data "out there" constrain the information that can be formed by the receiver on their basis. Unless the agent is hallucinating in a void or dreaming, her perceptions are formed on the basis of stimuli (understood as data) from the world to which she is sensitive. Our cognitive apparatus only allows us to discriminate some, but not all, physical discontinuities and nomic regularities in the world. (Whilst elephants, for example, are sensitive to infrasound, humans are not readily sensitive to infrasound signals.) Only those data to which we are cognitively sensitive can give rise to the formation of information. Any perceived physical data "out there" are encoded, or represented, as some form of neural patterns (e.g., as action potentials or activation patterns). The precise form of representation is a further empirical

The *meaningfulness* of the perceptually structured data is determined by their behavioural effect (either positive or negative) on the receiver. Such behavioural effect is broadly construed to encompass more than just *observable* behaviour. It amounts to, roughly, the change produced in the receiver's action(s), belief(s) or goal(s) resulting from the data perceived (e.g., leaving the forest immediately when smelling or seeing smoke on a very hot day). In that sense, the state of the world – as signified by the perceived data – and the receiver are connected. This change implies, as argued in [21], that there exists a requisite flexibility of behaviour in the receiver, such that the perceived data *can* yield some change in the receiver. It makes little or no sense to describe a rigid system S as being *informed* by something if S cannot somehow behave differently upon receiving these data. Further, any consequence of the perceived data is the result of how the receiver interprets the data and behaves in the world accordingly [22]. However, for the perceived data to

be *meaningful* there need not be any necessary dependence on a kind of coordination system amongst *senders* and receivers. Data need not be communicated amongst agents in order to *be meaningful*, and can flow directly from the world to the receiver [21]. Indeed, the world does not *communicate* with agents. It is rather the sensitivity of the receiver to particular regularities or physical discontinuities in the world that "flow" to the receiver.

Moreover, the effect concerned need not be necessarily positive (e.g., the receiver being informed about a nearby reservoir of water); it can often be negative (e.g., drawing a false conclusion regarding the distance of the reservoir). The distinction between negative and positive effects is what determines the relevance of information, as argued by D. Wilson and D. Sperber [23], not whether the meaningful structured data qualify as information. On their view, information is relevant to the agent when it (1) relates to her background information to derive conclusions that matter for her beliefs or actions, and (2) requires less processing effort by the agent. Others define the relevance of information relative to goals. A piece of information is relevant (for a goal) iff "it is a candidate for a belief that supports the processing of that goal" [24]. But either way, the relevance of information can only be determined once we have established what qualifies as information. The meaningfulness of the perceptual data is a prerequisite for the information being relevant. Understood this way, there is clearly room for mistakes (as a negative effect) in the agent forming information. An agent may mistake smoke particles for indicators of fire nearby, where, as a matter of fact, that smoke may be produced by smoke machines. Her escape from the forest would be rationally justified absent other overriding factors, despite there being no fire or imminent danger.

The theory proposed herein postulates that there is an important distinction to be made between *information-that* and *information-how* on the basis of the role information plays in cognitive processing. Information-how (e.g., 'In case of fire, break the glass and press the button') is prescriptive and informs an agent about which action has to be performed to achieve a particular result. As such, for cognitive agents it expresses an expectation for some goal-directed action on the part of the receiver in a given context. Information-that (e.g., 'Not all birds can fly') is descriptive and is about events, objects and states of affairs in the world. Cognitive agents use information-that to represent and form beliefs about, rather than merely externally react to, their environment. Both types of information play an important role in the way cognitive agents negotiate with their environment in terms of acting and believing. Neither information-how nor information-that need be restricted to sentences.

Lastly, why is this particular view of information considered apt to capture the kind of information processing often invoked in cognitive science? First, understanding information as being carried by data allows a broader applicability of the theory beyond linguistic aspects of cognition alone. To understand cognitive agency, what we want is a theory that focuses on *physical* information, and in that regard data-centred theories fare better. Sentences convey information, but so do sunlight and smoke, for example. Yet, unlike the Floridian data-centred theory of information, the present theory does not insist on the Veridicality thesis. Cognitive agents all too often make mistakes in interpreting perceptual data. Such mistakes should also be accounted for in explaining cognition. Second, information in cognitive science provides a naturalistic foundation for the explanation of cognition and behaviour. Humans and other organisms survive and reproduce by tuning themselves to reliable but imperfect cues that represent correlations between internal variables and environmental stimuli as well as between environmental stimuli and opportunities and threats [25]. The *meaningfulness* of perceived data described above is determined precisely by such "reliable but imperfect cues" the agent is sensitive to.

Third, the theory is neither too narrow nor too broad for our purposes. It is not too narrow in either imposing strict conditions that only few cognitive processes satisfy (e.g., the veridicality of the data for knowledge) or being limited to a subset of cognitive phenomena (e.g., language processing). It is compatible with the contemporary cognitive

scientific view that "the brain reveals itself proactive in its interface with external reality" being an *interpreter* rather than a mirror of that reality [26]. "[R]esearch [...] has shown how signals coding predictions about [...] simple features of relevant events can influence several stages of neural processing" [26]. The proposed theory is equally compatible, for example, with a recent, and contentious, view of the brain as an hypothesis-testing mechanism that attempts to minimise the error of its predictions about perceptual data from the world [27]. Both "bottom-up" signals (perceptual input data) and "top-down" signals embodying predictions about the probable causes of the perceptual input data can qualify as information according to our theory. At the same time, the theory is not too broad so as to make information vacuous. Information can come at degrees. Some data do not give rise to information, since the receiver is not sensitive to them. Other data are simply not meaningful to their receiver. And although both a tautology and a contradiction, for example, can be informational, they are less or more *useful* and/or *relevant* in a given context.

4 A comparison with other neo-Peircean theories

In this section, the relationships between the proposed theory and other neo-Peircean analyses of representation and information are discussed. To begin with, consider B. von Eckardt's analysis of non-mental representation. In [6] she adapts Peirce's triadic relation that obtains amongst the represented object, the representing vehicle (representamen) and the mental effect in the mind of the interpreter of the sign (interpretant). The represented object could be a physical object, a relation, a state of affairs or a property. The representing vehicle - what she calls the representation bearer - such as a map, a photo or a spoken word, can be individuated in terms of its nonrepresentational (or material) properties. Both the represented object and the representation bearer are, at least in principle, objectively verifiable. von Eckardt claims that in order for R to be an actual - rather than merely a possible - representation there must currently exist an actual interpreter bearing the right relation to R. The resemblance to the proposed theory of information should be clear. Information is understood pragmatically and in a manner that requires an actual consumer of physical data (that can be upgraded to information under the right conditions). On the other hand, data need not be communicated by senders. Physical data "out there" can at best be classified as *potential* information in the absence of consumers.

G. O'Brien and J. Opie build on von Eckardt analysis of non-mental representation and add that the vehicles of mental representation should be understood as some kind of neural states [7]. Given their commitment to a naturalistic account of cognition, they seek to explain the act of interpretation in *naturalistic* terms in order to avoid a vicious circle. They claim that the only viable alternative is treating interpretation in terms of some modification of the cognitive agent's behavioural dispositions towards the represented object. Here, too, the similarity is clear. The proposed theory of information suggests that the meaningfulness of perceived data (and, therefore, their being *informational*) is determined by their behavioural effect on the agent. It is suggested that on receiving new information some effect in the agent triggers an action or a response (e.g., forming/changing a belief-state).

On E. Jablonka's functional-evolutionary analysis of semantic information, the distinction suggested above between information-that and information-how becomes very blurry. That is the case, for example, when 'functional' means that signals received by either a human- or natural-selection designed system play a causal role that "usually contributes to the goal-oriented behavior of this system" [9]. An apple pie recipe and a piece of software are instances of functional-evolutionary information for a cook and a computer, respectively, in a manner akin to the appearance of black cloudy sky leading to the shelter-seeking action of an observing ape. Nevertheless, insofar as we seek to understand the role information processing plays in cognitive tasks in the *lifetime* of an agent, rather than over evolutionary time, the information-how/information-that distinction

seems worth preserving.

Lastly, on J. Queiroz, et al. neo-Peircean theory, information has the nature of a process of communicating a "form" to the interpretant [8]. That process constrains the possible patterns of behaviour of the interpreter. Information is taken typically as an interpreter-dependent "objective" process. Accordingly, it cannot be dissociated from a situated agent. On their view, it is only as a result of the interpretation process that information triadically connects the sign, object(s), and an effect on the interpreter. A sign (somehow) effectively communicates a form from the (represented) object to the interpretant, whilst changing the state of the interpreter. This account raises some interesting questions, which are not tackled here, about the objectivity of this process when it is dependent on a particular agent and about the communication of the form of an object to the interpreter (the world does not talk to us...). Nevertheless, it can be seen again that information is not simply "out there" in the world independently of a perceiver. Information is a dynamic construct that results from an ongoing interaction between the agent and its environment.

5 Concluding remarks

This short paper contributes to a long-standing and much-debated question of what concept of 'information' is suitable for understanding human cognition in terms of information processing. It is often argued, in cognitive science, that cognition is an information processing system. The literature contains many diverse theories of information (of which I have surveyed but a few here) pulling in different directions, thereby leading to disparate definitions of 'information'. Information, so I have suggested whilst adapting a neo-Peircean approach, should be understood pragmatically. Whatever the criteria for information are, what makes *x* a piece of information has to do with the way the agent either processes or can process *x* in actively engaging with her environment. Of course, it does not follow that a unified theory of information is either forthcoming or even possible. In different contexts, such as game theory or economics, information may be defined differently. The theory proposed herein is motivated by doing justice to the cognitive sciences. However, much more work is required to fully develop it.

Acknowledgements. This paper has benefited from the insightful comments and suggestions of two anonymous referees for the 2014 International Workshop on Artificial Intelligence and Cognition. Thanks are due to Aditya Ghose, Patrick McGivern, Michaelis Michael and Joel Pearson for useful discussions on topics related to the paper. This research has been supported by a fellowship from the Edelstein Centre for the History and Philosophy of Science, Technology and Medicine and a research grant from the Israeli Ministry of Aliyah and Immigrant Absorption. My attendance at the workshop was funded in part by the Rosselli Foundation.

References

- 1. Shannon, C.E., Weaver, W.: The mathematical theory of communication. University of Illinois Press, Urbana (1949).
- 2. Hartley, R.V.L.: Transmission of Information. Bell Syst. Tech. J. 7, 535–563 (1928).
- 3. Kolmogorov, A.N.: Three approaches to the quantitative definition of information. Probl. Inf. Transm. 1, 1-7 (1965).
- 4. Chaitin, G.J.: Algorithmic information theory. Cambridge University Press, Cambridge, UK (2004).
- 5. Floridi, L.: The philosophy of information. Oxford University Press, Oxford (2011).
- 6. Von Eckardt, B.: What is cognitive science? MIT Press, Cambridge, Mass. (1993).
- 7. O'Brien, G., Opie, J.: Notes toward a structuralist theory of mental representation. In: Staines, P.J., Clapin, H., and Slezak, P.P. (eds.) Representation in mind: new approaches to mental representation. pp. 1–20. Elsevier: Morgan Kaufmann, Amsterdam (2004).
- 8. Queiroz, J., Emmeche, C., El-Hani, C.N.: A Peircean Approach to "Information" and

- its Relationship with Bateson's and Jablonka's Ideas. Am. J. Semiot. 24, 75 (2008).
- 9. Jablonka, E.: Information: Its Interpretation, Its Inheritance, and Its Sharing. Philos. Sci. 69, 578-605 (2002).
- 10. Popper, K.R.: The Logic Of Scientific Discovery. Routledge, London (2002).
- 11. Barwise, J.: Information flow: the logic of distributed systems. Cambridge University Press, Cambridge (1997).
- 12. Bar-Hillel, Y., Carnap, R.: Semantic Information. Br. J. Philos. Sci. 4, 147-157 (1953).
- 13. Allo, P.: A Classical Prejudice? Knowl. Technol. Policy. 23, 25–40 (2010).
- 14. Dretske, F.I.: Knowledge & the flow of information. MIT Press, Cambridge, Mass. (1981).
- 15. Grice, H.P.: Studies in the way of words. Harvard University Press, Cambridge, Mass. (1989).
- 16. Floridi, L.: Information: a very short introduction. Oxford University Press, Oxford; New York (2010).
- 17. MacKay, D.M.: Information, mechanism and meaning. MIT Press, Cambridge, MA (1969).
- 18. Skyrms, B.: Signals: evolution, learning, & information. Oxford University Press, Oxford (2010).
- 19. Fresco, N., Pearson, J.: How theories of cognition define information. (Unpublished).
- 20. Peirce, C.S.: On a New List of Categories. Proc. Am. Acad. Arts Sci. 7, 287–298 (1868).
- 21. Cao, R.: A teleosemantic approach to information in the brain. Biol. Philos. 27, 49–71 (2012).
- 22. Millikan, R.G.: Biosemantics. J. Philos. 86, 281–297 (1989).
- 23. Wilson, D., Sperber, D.: Relevance Theory. In: Horn, L.R. and Ward, G.L. (eds.) The handbook of pragmatics. pp. 607–632. Blackwell, Malden, MA (2005).
- 24. Paglieri, F., Cristiano Castelfranchi: Trust in Relevance. In: Ossowski, S., Toni, F., and Vouros, G. (eds.) Proceedings of AT 2012 First International Conference on Agreement Technologies. pp. 332 346. CEUR-WS.org, Dubrovnik (2012).
- 25. Scarantino, A., Piccinini, G.: Information without truth. Metaphilosophy. 41, 313–330 (2010).
- 26. Nobre, A.C., Correa, A., Coull, J.: The hazards of time. Curr. Opin. Neurobiol. 17, 465–470 (2007).
- 27. Hohwy, J.: The predictive mind. Oxford University Press, Oxford, United Kingdom (2013).