

*Information Disorder Online is an Issue of Information  
Quality*

Uyiosa Omoregie

---

Issues of information disorder online could be reduced to the issue of information quality: not just about distinguishing truth from falsehood but highlighting, describing legitimate and credible information. “Trustworthiness” is the buzzword in info-quality circles, trending content online should have a signal of the level of ‘trust’ we can place on such content. Social media platforms have recognized the importance of signals of trust, for low-quality content to be de-prioritized and high-quality content amplified. According to Nick Clegg, VP of Global Affairs and Communications at Facebook:

“Facebook is also in the relatively early stages of exploring whether and how to rank some important categories of content differently — like news, politics, or health — in order to make it easier to find posts that are valuable and *informative*.” (emphasis mine). [1]

The algorithms built into a social media platform play an important role in the virality of content online and appear more focused on the quantity of user engagement with content than the quality of the content and the engagement [2] [3] [4]. Life in the “information age” today requires that the issues of information disorder (mis-information, dis-information and mal-information) online must be taken seriously: quality of information must be as important as quantity of information. This current paper introduces the concepts “off-information” and “non-information” and can be viewed as an attempt to construct a tentative theory of information quality for written online content. Shannon, the father of information theory, described a quantitative (mathematical) theory of communication [5]. In Shannon’s theory, the fundamental challenge of communication (hence information), is the need for reproduction at

one point, the exact or approximate message produced at another point. Shannon's theory led to technologies that encode, transmit, decode and store information. This is central to the information age. Efficiency and accuracy were of paramount importance to Shannon: the 'meaning' of the information was a secondary issue. This mathematical theory of information is a *quantitative* theory more about engineering than linguistics:

"Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages" [5].

If, according to Shannon, the "semantic aspects" are irrelevant, then what really is 'information'? Information scientists have adopted a general definition of information (GDI). According to the GDI, information is data that is well-formed and meaningful [6]. Misinformation and disinformation are not genuine information because they are false, although they may have semantic content. Semantic *factual* content is what distinguishes authentic information from false information [6] [7] (in declarative language).

Information *quality* is an issue primarily about meaning: the semantic aspects are more relevant than engineering efficiency or grammatical accuracy. A sentence can be grammatically correct but have no meaning and an engineering solution can transmit more data *quantity* at the expense of information *quality*. Language functions properly only when it expresses meaning. Chomsky, however, showed that the grammar of a language is really the theory of a language [8]. Grammar as theory is "autonomous and independent of meaning" [8]. Chomsky's famous example:

- (1) Colorless green ideas sleep furiously.
- (2) Furious sleep ideas green colorless.

Both sentences are meaningless. Sentence (1) is grammatical but none of these two sentences would pass the information quality test. These two sentences have no semantic content, we propose to label such content "non-information". Likewise, in the examples below, all sentences are grammatical but only sentence (5) contains authentic information. Sentences (3) and (4) are meaningless (non-information) and (6) is a hypothesis:

- (3) That man, the human next door, he is actually a reptile.

(4) He murdered that innocent man, but because he didn't really kill him, the victim is dead.

(5) That tomato you are holding in your hand is a fruit.

(6) The sun will rise tomorrow.

Unless further evidence is given (or the presuppositions made clear[9] to support the claim made in (3), an organism cannot simultaneously be human and a reptile. It is possible that the writer of (3) has some secret knowledge about humans and reptiles and the belief expressed could be justified. Knowledge can only convey information when it is made concrete and articulated through the mind[10], and background information is required for comprehension of most sentences [17]. If there is tacit knowledge hidden in the mind of the writer of (3), it is not expressed coherently: no information is communicated. If belief in (3) could be justified, that would not make the statement necessarily true[11]. Sentence (4) could be further clarified but as it stands is meaningless. The claim in (5) may appear strange to many people (a tomato is generally thought of as a vegetable), but, even without support (clarification in the sentence), the statement is authentic because it is scientifically correct. Stating that a tomato is a fruit is different from stating that the sun will rise tomorrow. The claim in (6) is not fact but a hypothesis because what is asserted is really speculation based on historical antecedent [12]. Information can be regarded as the product from data processed through a filter of facts, logic and semantics (see figure 1 below).

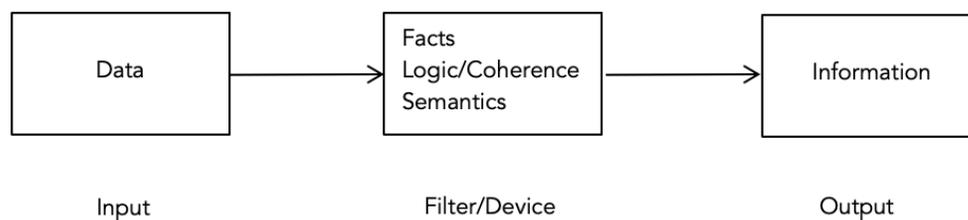


Figure 1: A device/filter separating information from data

Imperative (or instructional) content contains no propositions. Such content differs from declarative content because it cannot be analysed as true or false. Truth and falsity can only be found in propositions [13] [14]. In figure 1 above, instructional content would only be checked for logic/coherence and semantic issues but cannot be fact-checked. We propose that instructional content be analysed using the 'harm principle' [15] and by the legal system operating where the content is being analysed. Moral norms are important but distinct from

legal systems [15]. For content that is instructional and meaningful but illegal or potentially harmful we propose the label “off-information”. Of the four sentences below, only sentence (8) would be considered off-information: it is an instructional meaningful sentence with potential to cause harm if obeyed. Sentence (7) is meaningful but not realistic, (9) is wrong advice but not misinformation and (10) is meaningless hence non-information (no evidence of extraterrestrial visitors on Earth yet.)

- (7) Let’s eat all the food in every restaurant in this city today.
- (8) Swallow arsenic when you feel depressed.
- (9) Do not believe anything a teacher or professor tells you.
- (10) You should listen to extraterrestrials from Venus living in the apartment next to ours.

## References

- [1] Clegg, N. You and the Algorithm: It Takes Two to Tango. *Facebook*, <https://about.fb.com/news/2021/03/you-and-the-algorithm-it-takes-two-to-tango/>
- [2] Zhang, A, Ranganathan, A, Metz, S, et al. A structured response to misinformation: defining and annotating credibility indicators in news articles. *The 2018 Web Conference Companion*. <https://dl.acm.org/doi/10.1145/3184558.3188731>
- [3] Hao, K. How Facebook got addicted to spreading misinformation. *MIT Technology Review*, [www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/](http://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/)
- [4] DiResta, R. Free speech is not the same as free reach. *Wired*, <https://www.wired.com/story/free-speech-is-not-the-same-as-free-reach/>
- [5] Shannon, C. A mathematical theory of communication. *Bell Sys Tech J* 1948; 27: 379-423, 623-656.
- [6] Floridi, L. Semantic conceptions of information. *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/information-semantic/>
- [7] Levitin, DJ. *A Field Guide to Lies*. Toronto: Penguin, 2020, p. xv.
- [8] Chomsky, N. *Syntactic Structures*. Berlin: Mouton de Guyter, 1957, p.49.
- [9] Yablo S 24.251 *Introduction to philosophy of language*. <https://ocw.mit.edu>.

- [10] Stenmark D. Information vs. knowledge: the role of intranets in knowledge management. *Proceedings of the 35th Hawaii International Conference on Systems Sciences*, <https://www.computer.org/csdl/pds/api/csdl/proceedings/download-article/12OmNymjN0R/pdf>
- [11] Gettier, E. Is justified true belief knowledge? *Analysis* 1966; 23: 121-3.
- [12] Wittgenstein L. *Tractatus Logico-Philosophicus*. London: Paul, Trench, Trubner & Co., 1922. P.14
- [13] Lycan, WG. *Philosophy of Language*. New York: Routledge, 2019, p.73.
- [14] Ramsey FP. Critical notices, *Tractatus Logico-Philosophicus*. *Mind* 1923; XXXII: 465-478
- [15] Mill, JS. *On Liberty*. London: Penguin Classics, 1859/2010.
- [16] Nussbaum, MC. Minimum core obligations: toward a deeper philosophical enquiry. *jamesgstewart.com*, <http://jamesgstewart.com/minimum-core-obligations-toward-a-deeper-philosophical-inquiry/>
- [17] Searle, JR. 'Sneaked' or 'snuck'. *The New York Review of Books*, <https://www.nybooks.com/articles/2002/03/14/sneaked-or-snuck/>