Algorithmic detection of misinformation and disinformation: Gricean perspectives

Sille Obelitz Søe
Department of Information Studies, University of Copenhagen, Copenhagen, Denmark

Received 24 May 2017
Revised 6 October 2017
Accepted 17 October 2017

Abstract

Purpose – With the outset of automatic detection of information, misinformation, and disinformation, the purpose of this paper is to examine and discuss various conceptions of information, misinformation, and disinformation within philosophy of information.

Design/methodology/approach – The examinations are conducted within a Gricean framework in order to account for the communicative aspects of information, misinformation, and disinformation as well as the detection enterprise.

Findings – While there often is an exclusive focus on truth and falsity as that which distinguish information from misinformation and disinformation, this paper finds that the distinguishing features are actually intention/intentionality and non-misleadingness/misleadingness – with non-misleadingness/misleadingness as the primary feature. Further, the paper rehearses the argument in favor of a true variety of disinformation and extends this argument to include true misinformation.

Originality/value – The findings are novel and pose a challenge to the possibility of automatic detection of misinformation and disinformation. Especially the notions of true disinformation and true misinformation, as varieties of disinformation and misinformation, which force the true/false dichotomy for information vs mis-/disinformation to collapse.

Keywords Algorithms, Communication, Information, Philosophy, Misinformation, Disinformation

Paper type Conceptual paper

1. Introduction

The internet is full of communication. Some of this communication consists of false, inaccurate, and untrue information. As a reaction to this we have lately witnessed an increased interest in automatic detection (through algorithms) of misinformation and disinformation. Examples of such detecting-projects are the PHEME-project (2014), Kumar and Geethakumari's "Twitter algorithm" (2014), Karlova and Fisher's diffusion model (Karlova and Fisher, 2013), and the Hoaxy platform (Shao *et al.*, 2016) – to name a few.

The interest in detection of misinformation and disinformation follows an ancient philosophical quest for "the truth." The hope is to be able to single out misinformation and disinformation in order to prevent it from spreading, thereby enabling the spread of proper information – "the truth" – instead.

It is a new task for the algorithmic moderators of social media and the internet. The assumption is that false content online should be flagged – maybe even removed – in order to secure the best conditions for true content, thereby, helping people make the

At the author's PhD defence Professor David Bawden asked the author: "If I am on a deserted island and your thesis is all I have with me, what would you like me to gain from it? What should I remember?" This paper provides the author's answer. Thus, it is based on the author's dissertation (Søe, 2016). The author would like to thank Jens-Erik Mai, Erik J. Olsson, Jack Andersen, and Laura Skouvig for valuable comments and suggestions along the way.

Journal of Documentation
Vol. 74 No. 2, 2018
pp. 309-332
Emerald Publishing Limited
0022-0418
DOI 10.1108/JD-05-2017-0075

right decisions. That is, in crude terms, to decide what people should believe. Further, the assumption is that the detection of truth and falsity is sufficient for the detection of misinformation and disinformation. However, the full story is more complicated than mere detection of truth and falsity – which might actually prove complicated enough.

The mechanism of social media is communication. To post and share stories. To react and comment. To write statuses about oneself that friends, family, colleagues, and others can comment upon and share. To arrange and coordinate public and private events and invite people to attend. All these acts – both the verbal and non-verbal – are acts of communication. They are carried out at a specific time, within a specific context, and for a specific purpose – guided by belief, intention, and meaning. When a story is shared the original purpose of posting it, might change for another purpose in sharing it. The context changes as well and, most likely, also the belief and intention and maybe even the meaning. Thus, to determine whether something is misinformation or disinformation requires evaluative judgments of content, context, purpose, etc. and the question is whether such judgments can be automated. Further, the main question is what algorithms should look for in order to detect misinformation and disinformation – i.e. what misinformation and disinformation actually is in connection to one another and in connection to information. That is, what are the distinct and distinguishing features of information, misinformation, and disinformation, conceptually?

The paper is structured as follows. Section 2 sets the scene by dealing with four different detecting-projects. In Section 3 a Gricean framework of meaning, cooperation, and communication is laid out. Four different accounts of information, misinformation, and disinformation, all influenced by Grice, are presented and briefly discussed in Section 4. Discussions of information, misinformation, and disinformation, their nature, and their implications for automatic detection are carried out in Section 5 and Section 6 provides concluding remarks.

2. Automatic detection

The PHEME-project sets out to algorithmically detect and categorize rumors in social network structures (such as Twitter and Facebook) in near real time. The rumors are mapped according to four categories, which include "misinformation, where something untrue is spread unwittingly; and disinformation, where it's done with malicious intent." (Sheffield News, 2014). The purpose is to help journalists by developing a platform where stories and rumors can be fact-checked before a story is posted online or sent to print (PHEME, 2014). Kumar and Geethakumari (2014) propose an algorithm which can detect and flag whether a tweet is misinformation or disinformation. In their framework "Misinformation is false or inaccurate information, especially that which is deliberately intended to deceive [and] Disinformation is false information that is intended to mislead, especially propaganda issued by a government organization to a rival power or the media." (Kumar and Geethakumari, 2014, p. 3). The purpose of the algorithm is to improve decision making for individual users by letting the algorithm tell them whether a given tweet is information, misinformation, or disinformation, and thereby indirectly tell them whether they should retweet or not. In Karlova and Fisher's (2013) diffusion model misinformation and disinformation "extend the concept of information through their informativeness" and misinforming and disinforming function as types of information behavior. Thus, Karlova and Fisher (2013) define misinformation as inaccurate information and disinformation as deceptive information and their goal is to better understand how information spreads and diffuses in online social networks. Hoaxy (Shao *et al.*, 2016) is "a platform for the collection, detection, and analysis of online misinformation and its related fact-checking efforts." (Shao *et al.*, 2016, p. 745). Hoaxy solely deals with misinformation defined as "false or inaccurate information" (Shao *et al.*, 2016, p. 745) with examples such as rumors, false news, hoaxes and elaborate conspiracy theories (Shao *et al.*, 2016).

In these projects alone, four different understandings of misinformation, as well as three different understandings of disinformation are presented. Thus, depending on which algorithm is employed different answers will be given as to whether a particular tweet or Facebook post is misinformation, disinformation, or information. The four different algorithms detect different things under the same headings thereby risking to complicate decision-making for individuals rather than enable and facilitate it. As the algorithms potentially will judge what people should believe and how they should make decisions – i.e. supersede or be an addition to peoples' critical assessment – it is not sufficient or satisfactory to opt for the simplest definitions.

Therefore, a more developed and nuanced understanding of the notions information, misinformation, and disinformation – and especially their interconnections – is needed. That is, in order to develop algorithms which can detect misinformation and disinformation one must at least understand which features these algorithms should detect. In order to tell whether something is information, misinformation, or disinformation one must know what characterizes information, misinformation, and disinformation, respectively. One must know what to look for to tell the notions apart – i.e. the features which distinguish the notions from one another. These characteristics – these distinguishing features – are what the algorithms have to detect. Otherwise, how should the algorithms be able to discriminate between the three notions – in a satisfactory and sufficient way – and aid decision-making?

Currently, the different detecting-projects point to different features as those which characterizes information, misinformation, and disinformation, respectively. They employ features such as inaccuracy, untruth, falsity, and deception for misinformation and falsity, intended misleadingness, intended untruth, deception, and propaganda for disinformation. Further, both misinformation and disinformation are defined in terms of information within all four projects, wherefore, it is also necessary to know what counts as information. If, for instance, both misinformation and disinformation have deception as one of their features, then deception is not a sufficient feature for the algorithms to detect – the algorithms should also look for something else. If, however, deception is only a feature of disinformation then the algorithms have to detect for deception. In this case the question is how the algorithms can detect deception – i.e. what characterizes deception as opposed to non-deception?

If one were to pursue the task of singling out these distinguishing features two steps would have to be taken. First, in order to determine which features are unique for each notion in question (that be, unique by themselves or due to a specific combination) these notions would have to be examined and analyzed in connection to one another. Second, it would be necessary to further specify these distinguishing features. That is, if deceptiveness is a distinguishing feature then how is deception detected? Which features should let the algorithm categorize something as deception? Should the algorithms for instance detect for falsity as opposed to truth? Should they detect for trust and credibility of the source as a sign of non-deceptiveness? Or should they detect for something else entirely?

However, such a task seems impossible without crystal clear notions of information, misinformation, and disinformation. Unfortunately these three notions are not very clear – especially not when taken together.

Therefore, this paper explores and analyzes the very nature of information, misinformation, and disinformation – and their interconnections – in order to understand the challenges these notions pose to the possibility of automatic detection. In other words, the purpose of this paper is more conceptual than practical. The purpose is to get the fullest and deepest conceptual understanding of the notions information, misinformation, and disinformation, their interconnections, and distinguishing features – with an eye to algorithmic detection. The task of deriving operational specifications for algorithmic detection, from this conceptual clarification, will be left for others to pursue.

3. Gricean framework

Automatic detection of misinformation and disinformation are actions conducted within communicative structures. The various detecting-projects offer algorithmic detection of misinformation and disinformation in tweets and Facebook post, as well as other types of online communication. Thus, the notions of information, misinformation, and disinformation and their interconnections must be analyzed and discussed within a communicative framework (cf. the eye to algorithmic detection). That is, the algorithms have to work within communicative structures and upon communicative actions.

Such a communicative framework is provided by aspects of Grice's ordinary language philosophy. In order to grasp the implications of such a framework, both for the purpose of automatic detection and for the analyses of the notions of information, misinformation, and disinformation brief introductions to two of Grice's most famous theories are necessary.

Grice's (1967) communication theory (i.e. the Cooperative Principle and its maxims) and Grice's (1957) theory of meaning offer a fruitful framework for understanding the notions information, misinformation, and disinformation and their interconnections within philosophy of information. Grice's communicative aspects and insights regarding the differences between sentence-meaning and utterer's meaning, as well as natural meaning and nonnatural meaning, enable the development of a unified conceptualization of information, misinformation, and disinformation.

Grice is most famously known for his realization that what words, sentences, and utterances (including gestures)[1] literally or specifically mean is only half the story about meaning. The other very important part of the story is what people mean by their utterances (including gestures). Thus, Grice introduces a distinction between sentence-meaning and utterer's meaning (speaker-meaning) and thus paves the way for a pragmatic notion of meaning (utterer's meaning) as a supplement to a semantic notion of meaning (word-/sentence-meaning). The distinction between sentence-meaning and utterer's meaning and thereby the distinction between semantics and pragmatics is most fully and explicitly developed within Grice's (1967) William James lecture "Logic and Conversation" but the foundation for the distinction was already laid in Grice's (1957) "Meaning" almost 20 years prior to "Logic and Conversation[2]."

Grice's work has previously been used within information studies for various purposes. For instance, it has been used: to describe different forms of communication in connection to information retrieval – more specifically to analyze and explain information retrieval as a communicative process (Blair, 1992, 2003); to define a notion of information in terms of meaning – in order to develop a conversational notion of information quality (Mai, 2013); as heuristics in analyses of information as communication (Fox, 1983); and as the basis for work on misleadingness (Fallis, 2010, 2014). As is often the case, it is Grice's communication theory and theory of meaning which are at the heart of these examples.

3.1 Meaning as intention

In the essay "Meaning" (1957) Grice introduces and develops a distinction between, what he calls, natural meaning and nonnatural meaning. Natural meaning is characterized by the formula "x meant that p and x means that p entail p" (Grice, 1957, p. 213). The entailment secures that if some x means p, then p must be the case – p must be true or must obtain. In Grice's words if "Those spots mean (meant) measles." (Grice, 1957, p. 213) then it must be the case that the human with those spots has measles.

In contrast nonnatural meaning is characterized by the formula "x means that p and x meant that p do not entail p." (Grice, 1957, p. 214). Thus, in the case of nonnatural meaning there is no requirement that specific events obtain or that something specific is true or is the case. Nonnatural meaning is based on intentions and conventions and Grice's (1957) example of nonnatural meaning is "Those three rings on the bell (of the bus) mean that the bus is full." (p. 214). In this case the three rings on the bell could mean

almost anything – they only mean that the bus is full because this is what someone has collectively decided at some point.

One of Grice's key insights is that meaning is determined by how humans use language and gestures and this use is determined by the intentions which humans have when they communicate. Thus, for A (a human agent) to mean something (*p*) by x (an utterance), A must have three interdependent intentions. A must intend that:

(1) the hearer will be convinced that *p*;

(2) that the hearer recognizes A's intention in (1); and

(3) that it is because of the recognition in (2) that the hearer is convinced that *p*. (Grice, 1957).

For Grice these three intentions are necessary and sufficient for nonnatural meaning and thus a way of reducing nonnatural meaning to intentions. When the chauffeur rings the bell three times he adheres to the convention by expressing his intention to let the passengers know that the bus is full. Further, the chauffeur intends that the passengers recognize the intention and thereby – qua the recognition – are convinced that the bus is full. However, nonnatural meaning as expression of conventions and intentions leaves room for mistakes – the chauffeur can be wrong about whether the bus is full. But even if the chauffeur is wrong when he rings the bell three times, the ringing still keeps its nonnatural meaning that the bus is full. The nonnatural meaning of the convention is in place regardless of whether people make mistakes. Further, Grice notes that gestural acts (e.g. to make communicative signs with the hands, to draw a picture of something) can also be instances of nonnatural meaning as long as the three communicative intentions are in place. Thus, nonnatural meaning can be exerted non-verbally. This is the reason why Grice in parentheses remarks "I use 'utterance' as a neutral word to apply to any candidate for meaning$_{N[on]N[atural]}$; it has a convenient act-object ambiguity" (Grice, 1957, p. 216).

3.2 Communication as cooperation

The reduction of nonnatural meaning to the intentions people have when they utter something is crucial for the distinction between context-invariant meaning (word-/sentence-meaning) and context-dependent meaning (utterer's meaning) fully developed in the essay "Logic and Conversation" (1967). Grice observes that:

> [o]ur talk exchanges do not normally consist of a succession of disconnected remarks, and would not be rational if they did. They are characteristically, to some degree at least, cooperative efforts; and each participant recognizes in them, to some extent, a common purpose or set of purposes, or at least a mutually accepted direction (Grice, 1967, p. 26).

From this observation he formulates a general principle – the Cooperative Principle – which governs all normal conversation:

> Cooperative Principle: "Make your conversational contribution such as is required, at the state at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" (Grice, 1967, p. 26).

Although the Cooperative Principle is stated in imperative, it is supposed to explain how normal conversation proceeds and why this procedure is rational – the principle does not prescribe how a conversation should proceed.

In connection to the Cooperative Principle Grice describes four maxims (presented in the list below) – again stated in imperative but supposed to depict rules which people actually adhere to in conversation.

MAXIMS (source Grice, 1967, pp. 26-27).

Quantity (how much information is to be provided):

(1) make your contribution as informative as is required (for the current purposes of the exchange); and

(2) do not make your contribution more informative than is required.

Quality (of the information provided).

Supermaxim – try to make your contribution one that is true:

(1) do not say what you believe to be false; and

(2) do not say that for which you lack adequate evidence.

Relation:

(1) be relevant.

Manner (the form of the utterance).

Supermaxim – be perspicuous:

(1) avoid obscurity of expression;

(2) avoid ambiguity;

(3) be brief (avoid unnecessary prolixity); and

(4) be orderly.

When utterers exploit, violate, or flout one or more of the maxims while still observing the Cooperative Principle – that is without opting out of the cooperation of the conversation – conversational implicatures are generated.

The depiction of the notion "implicature" is perhaps the most interesting part of Grice's conversation theory. "Implicature" is a pragmatic notion in the sense that it denotes what an utterer means or implicates by an utterance beyond what he literally says – they are instances of utterer's meaning. They arise when the utterer means something more or something else than the literal (word-/sentence) meaning of what he says. Implicatures can be of two different kinds; either what is implicated can lie within the utterance itself and be derived solely from the conventional meaning of the words used (conventional implicature), or what is implicated can lie beyond the utterance and is determined by intentions and context (conversational implicature). In this way conversational implicatures are determined by what the utterer means and they can only be understood when the context is taken into account. This means that dependent on the context and the specific conversation the same utterance can be used to generate different conversational implicatures and thereby be used to mean different things. Therefore, Grice has it as a requirement for conversational implicatures that they have to be worked out by the hearer. Take Grice's (1967) own example:

> A is writing a testimonial about a pupil who is a candidate for a philosophy job, and his letter reads as follows: "Dear Sir, Mr. X's command of English is excellent, and his attendance at tutorials has been regular. Yours, etc." (Gloss: A cannot be opting out, since if he wished to be uncooperative, why write at all? He cannot be unable, through ignorance, to say more, since the man is his pupil; moreover, he knows that more information than this is wanted. He must therefore, be wishing to impart information that he is reluctant to write down. This supposition is tenable only if he thinks Mr. X is no good at Philosophy. This, then, is what he is implicating) (p. 33).

After the introduction of conventional and conversational implicatures, Grice leaves conventional implicatures behind and solely focuses on conversational implicatures, which he then refers to simply as implicatures. As mentioned conversational implicatures are generated when one or more of the maxims are violated or exploited. As examples of implicature-generating violations, Grice mentions metaphors, irony, and ambiguity, among others.

With the feature that implicatures are meaning beyond what is literally said the notion "implicature" is an explicit introduction of Grice's famous distinction between sentence-meaning

(i.e. what is literally said) and utterer's meaning (i.e. what is implicated) – the distinction which was alluded to in "Meaning" (1957).

It is Grice's various insights regarding "meaning" together with his emphasis on cooperation and communication, which comprise the Gricean framework for the analyses of information, misinformation, and disinformation in connection to automatic detection. What is special about Grice's work on "meaning" is that it paves the way for pragmatics. Utterer's meaning as well as nonnatural meaning are pragmatic notions where context, intentions, and beliefs make all the difference. When it comes to algorithmic detection of specific features within peoples' communication it is important with a framework which can account for pragmatics – i.e. context, intentions, etc. – and not just operates on the semantic level. This is the Gricean framework – a pragmatic communicative framework – that can deal with what people mean with their utterances beyond what these utterances literally or specifically mean.

4. The inevitable: Dretske, Floridi, Fox, and Fallis

Four philosophers and information studies scholars offer themselves as inevitable in regard to accounts of information, misinformation, and disinformation: Luciano Floridi, Fred Dretske, Christopher Fox, and Don Fallis. Their accounts all fall within philosophy of information – an intersection between analytical philosophy and information studies (cf. Søe, 2016).

In information studies it has been discussed whether information is a thing (Buckland, 1991), a sign plus meaning (Mai, 2013), a notion which is not needed (Furner, 2004), the quality of being informed (Day, 2001/2008), or "the pattern of organization of matter and energy" (Bates, 2006, p. 1044), to mention a few proposals. It has also been discussed whether "information" is objective (Bates, 2006) or subjective/situational (Hjørland, 2007), as well as how it is related to truth and meaning (Budd, 2011). Further, there is the question of the relation between so-called information theory – i.e. Shannon's (1948) *Mathematical Theory of Communication* – and information studies. Thus, Cornelius (2002) offers an overview of information theory in information studies from Shannon's statistical information to semantic information to the cognitive turn with Brookes' (1980) "fundamental equation of information science" (Cornelius, 2002, p. 407). According to Cornelius (2002) much of the discussions stems from the influence of Shannon's idea within information studies. Thus, Cornelius (2002) argues, "[p] roblems connected with this model [Shannon's model] have remained with us. Some of the concepts are ambiguous; the identification of information with a process has spancelled the debate; the problems of measuring the amount of information, the relation of information to meaning, and questions about the truth value of information have remained" (p. 394). These various conceptions of information and the accompanying discussions are but a fraction of what can be found both within information studies and within other disciplines working with information on a theoretical, conceptual, or technical level.

However, none of these conceptions takes misinformation and disinformation into account. This is where Luciano Floridi, Fred Dretske, Christopher Fox, and Don Fallis enter the picture as inevitable: besides accounts of information, they all in some way or other deal with misinformation and/or disinformation. Grice serves as foundation for general theories about information, misinformation, and disinformation due to his insights regarding "meaning" and cooperation in communication and as such already lies underneath many of the accounts of these notions.

Luciano Floridi is one of the most notable philosophers of information and his account of semantic information acts as the main reference within the field. Floridi's work is mainly centered on information, however, he does offer brief definitions of misinformation and disinformation, which makes him one of the few who address all three notions (although not to an extent as to count as a single unified treatment). Floridi's (2005a, b, 2007, 2011) work on

semantic information is partly developed from Fred Dretske's (1981) account of semantic information. Dretske's account has been, and continues to be, hugely influential within philosophy of information as elsewhere. As Cornelius (2002) writes "[m]uch of the recent literature on information outside information science centers on Dretske's (1981) work" (Cornelius, 2002, p. 416). This alone makes Dretske inevitable. Further, Dretske (1981) offers an account of misinformation in terms of misrepresentation. Floridi's notion of misinformation is a reference to Christopher Fox (1983); one of the few who have dealt with misinformation explicitly. Thus, Fox's account of misinformation together with his account of information is also inevitable. For the notion of disinformation, not much is said in Dretske, Floridi, and Fox's work. Dretske (1983) merely mentions the notion in a summary of his account of semantic information; Floridi (2005b, 2011) offers brief accounts of disinformation as a by-product of his account of semantic information; and Fox (1983) does not mention the notion at all. Thus, accounts of disinformation must be found elsewhere. Don Fallis (2009, 2011, 2014, 2015) has developed the most elaborate and thorough accounts of disinformation based on the literature on lying, misleading, and deceiving. James Fetzer (2004a, b) has also analyzed the notion of disinformation – and he argues against Floridi's notion of semantic information. However, Fallis' accounts of disinformation are more extensive than Fetzer's as Fallis explains more varieties of disinformation than Fetzer accounts for. Further, Fallis' analyses of disinformation are developed as direct replies to Fetzer (2004a, b) and Floridi (2005b, 2011) and comes with brief definitions of misinformation and information. Fallis is also one of the few who address all three notions.

For these reasons – and due to the fact that Grice plays an increasingly important role within their various definitions of information, misinformation, and disinformation – the main focus in this paper will be the accounts put forth by Dretske, Floridi, Fox, and Fallis as the foundation for analyses of information, misinformation, and disinformation.

4.1 Dretske: information is objective

Within philosophy of information – understood as the intersection between analytical philosophy and information studies – the multiple accounts of information roughly falls within two categories of research. Either the notions of information has the form of technical terms developed for specific uses and purposes; or the notions have the form of overall terms developed to capture "the ordinary sense" of information (for an overview cf. Adriaans, 2012; Adams and de Moraes, 2016; Capurro and Hjørland, 2003; Floridi, 2005b; Furner, 2010, 2014; Robinson and Bawden, 2014).

An exception from this general picture is philosopher Fred Dretske's (1981) account of semantic information as Dretske tries to do both at once. Dretske's account of information is developed as part of an information-theoretic account of knowledge with the ultimate purpose of explaining behavior – i.e. a technical notion of information for a specific purpose. However, at the same time Dretske tries to capture the ordinary concept of information which, in his view, is connected to knowledge, news, learning, and intelligence. Dretske's (1981) seminal work *Knowledge and the Flow of Information* in which he develops the semantic notion of information has proved rather influential within philosophy of information, where it serves as the reference – almost the origin – of a truth-requirement for information (e.g. Adams and de Moraes, 2016; Adriaans and van Benthem, 2008; Budd, 2011; Capurro and Hjørland, 2003; Cornelius, 2002; Fallis, 2009; Floridi, 2005a, b, 2011; Furner, 2010; Godfrey-Smith, 1989; Mai, 2013; Robinson and Bawden, 2014; Scarantino and Piccinini, 2010; The Π Research Network, 2013).

Dretske's (1981) account of semantic information, which is based on Shannon's (1948) *Mathematical Theory of Communication*[3], came with a turn in analytic philosophy when the interest in formal modeling was renewed. Formal modeling[4] had been suppressed since the late 1940s because of the ordinary language turn in analytic philosophy. Due to the

ordinary language turn earlier proposals for an account of semantic information – such as Carnap and Bar-Hillel's (1952) – had been more or less ignored (The Π Research Network, 2013). Thus, Dretske's account of semantic information came at a time when analytic philosophy as a field was ready to appreciate it and respond to it, which made it hugely influential.

On Dretske's account information is an objective and truthful commodity. It is objective in the sense that it exists in the world independent of agents. Information is true and objective because it consists of signals about events or states of affairs – i.e. signals that carry information that some condition has obtained. Information carrying signals is the legacy from Shannon's (1948) *Mathematical Theory of Communication*, but where Shannon was only concerned with the amount of information which traveled from sender to receiver, Dretske's addition is the semantic aspect, the informational content. In other words, Dretske is interested in which information, **which** the signal carries. In his efforts to describe information as an objective commodity, Dretske relies on Grice's (1957) distinction between natural and nonnatural meaning (cf. Section 3). Thus, Dretske stresses that information is equivalent to natural meaning, whereas it must be distinguished from "Grice's nonnatural meaning, the sense of meaning that is relevant to language and semantic studies" (Dretske, 1981, p. 242, footnote 9). Nonnatural meaning is ascribed to information by agents when they interact with it. It is in this ascription that misrepresentation, mistakes, and thereby falsity can occur. Mistakes are the consequence of misrepresentation. Misrepresentation occurs when the informational link between an information source and a type (e.g. a semantic structure) is not preserved in a token (e.g. a belief) of that type, which means that the belief can be false (Dretske, 1981). Misrepresentation as the generator of, for example, false beliefs is the source of misinformation on Dretske's account[5].

To put it short, in Dretske's (1988) view, information is an "objective, mind-independent, indicator relation. […] Talking about information is yet a third way of talking about the fundamentally important relation of indication or natural meaning" (pp. 58-59). Thus, information is equivalent to natural meaning in Grice's sense. It is the signals about sources.

4.2 Floridi: information is true

Dretske's account of semantic information as an inherently truthful concept has also influenced the most notable philosopher of information, Luciano Floridi. Floridi's notion of semantic information is the outset of most discussion within philosophy of information independent of whether one agrees with his definition or not.

Floridi (2005a, 2008) stresses that information is a polymorph and polysemantic concept and thus limits his account of information to the specific kind of declarative, objective, semantic information, so-called DOS-information. With Dretske (1981) and Grice (1989/1991)[6] in hand Floridi (2005a) argues that the Standard Definition of (semantic) Information (SDI), which stipulates that information is well-formed, meaningful data, needs revision. "According to SDI, alethic values are not embedded in, but supervene on semantic information […] meaningful and well-formed data qualify as information, no matter whether they represent or convey a truth or a falsehood or have no alethic value at all" (Floridi, 2005a, p. 359). It follows that tautologies are information and that "false information (including contradictions), i.e. misinformation, is a genuine type of DOS-information, not pseudo-information" (Floridi, 2005a, p. 359). For Floridi, this is an intolerable and indefensible result. In his view it causes semantic erosion and he argues that the adjective "false" in "false information" functions attributively and negates the noun "information." For Floridi (2005a) DOS-information has to be inherently truthful wherefore SDI cannot be a standard definition of DOS-information in his view. Thus, calling it the Dretske-Grice approach, Floridi adds a truth-condition to SDI and arrives at a

revised standard definition of semantic information (RSDI) which states that semantic information is "well-formed, meaningful and truthful data" (Floridi, 2005a, p. 367, 2011, p. 80).

On Floridi's account, "false information" is an oxymoron with the more proper name "misinformation" which is not a kind of information due to its falsity. More specifically Floridi (2005b) argues that "[w]hen semantic content is false, this is a case of misinformation (Fox, 1983). And if the source of misinformation is aware of its nature, one may speak of disinformation" (Floridi, 2005b, Section 3.2.3). In some of his newer writings, Floridi has slightly changed the definitions of misinformation and disinformation such that "misinformation is "well-formed and meaningful data (i.e. semantic content) that is false." "Disinformation" is simply misinformation purposefully conveyed to mislead the receiver into believing that it is information" (Floridi, 2011, p. 260). However, what is unchanged between the different definitions is that Floridi distinguishes between information and mis-/disinformation in terms of truth and falsity.

In other words, Floridi defines information, misinformation, and disinformation as semantic content where information is the truthful[7] part, misinformation is the false part, and disinformation is a subcategory of misinformation denoting the purposefully misleading part of false semantic content. The truth requirement for information is partly derived from Grice.

4.3 Fox: information is communication

Floridi refers to Fox (1983) in his definition of misinformation. In contrast to the notion of information, where more accounts have been developed than one can ascertain, only a few accounts of misinformation and disinformation have been put forth. Thus, Christopher Fox (1983) is one of the few researchers who have developed a genuine account of misinformation. Fox's account is developed in parallel to an account of information – or, in fact, it is the notion of information that is analyzed and tested, and then the conclusions are extended to include misinformation as well. Fox's idea is that information and misinformation behave in the same way with the sole difference that misinformation is necessarily false, whereas, information is alethically neutral[8]. For Fox (1983) information and misinformation are bound by language as they are defined as propositions expressed by sentences. In order for information or misinformation to be present, the agent expressing some proposition by a sentence must be in a position to know whether the proposition is true and the sentence must be heard and understood by another agent. These last two requirements (that the informer is "in a position to know" and that the informee "hears and understands") are derived from analyses of the acts of informing and misinforming. For Fox, informing and misinforming are kinds of telling which – in contrast to saying – requires that the one being told hears and understands what is being said[9]. "Informing" is then that specific kind of "telling" which requires that the informer "is in a position to know that P." That the agent has to "be in a position to know that P" does not mean that he necessarily has to know whether P. It simply means that it in some way should be possible for the agent to know whether the proposition is true or not – i.e. that he has "adequate justification to support a belief concerning whether p is the case" (Fox, 1983, p. 179). In order to specify the relation between information (misinformation) and the act of informing (misinforming) Fox (1983) works with, what he calls, a "truism." The "truism" says that "Information is that which is conveyed when X informs Y that P" (p. 190).

Throughout his philosophical analyses of information and misinformation Fox adheres to Grice's cooperative principle and its maxims (cf. Section 3) in order to explain why information, misinformation, inform, and misinform function as they do. For instance, Fox (1983) analyses whether "inform" (and thereby "information") entails truth and whether "misinform" (and thereby "misinformation") entails falsity. These analyses are conducted

through a consistency test that is based on Grice's cooperative principle and its maxims. The idea is that if P entails Q then the statement "P and Q" exhibits a performance oddity – i.e. the resulting sentence sounds odd – whereas there is no performance oddity in "P and Q" if P does not entail Q. It is the oddity, which Fox explains in terms of violations of Gricean maxims – i.e. it is maxim violations within specific sentences that give rise to the oddities. Further, Fox uses the cooperative principle and its maxims to explain why there is a default association between information and truth although he concludes that "inform" and "information" do not entail truth. The idea is that "if "inform" were used in lieu of "misinform" when P is believed to be false, then either one would not be as informative as required by maxim (1), or one would be obliged to state further that P is false, which violates maxim (4)." (Fox, 1983, p. 160).

It is important to note that Fox's account of information and misinformation is bound by language. For Fox information and misinformation are expressed linguistically. Non-verbal information and misinformation seem impossible on Fox's account as information and misinformation are defined as propositions expressed through sentences. The linguistic constraint for information and misinformation is not shared by Dretske, Floridi, and Fallis.

Just as Grice's cooperative principle and its maxims has an influence on the notions of information and misinformation – on Fox's account – the cooperative principle and especially the notion of implicature has an influence on the notion of disinformation.

4.4 Fallis: information is representation

The most extensive, thorough, and detailed accounts of disinformation have been developed by Don Fallis. Through multiple conceptual analyses Fallis (2009, 2010, 2011, 2014, 2015) works his way through the various definitions of disinformation, lying, misleading, and deceiving. Fallis' goal is to develop a conceptual account of disinformation, which encapsulates all the intuitive senses and different types of disinformation – such as visual disinformation, side effect disinformation, true disinformation, and adaptive disinformation – while excluding honest mistakes, satire, and jokes.

Fallis (2011) evaluates the three different formulations of disinformation, which Floridi offers in his writings. At a first glance Floridi's formulations look like a single account in different phrasings, however, according to Fallis (2011), they constitute three separate accounts – accounts which are either too broad or too narrow – or both at once. Where Floridi's (2005a) account of disinformation is formulated in contrast to his semantic notion of information, Fallis (2014) chooses a different approach. Based on Grice's (1957) account of meaning and his Cooperative Principle (Grice, 1967) (cf. Section 3) – as well as the literature on lying, misleading, and deceiving – Fallis (2014) arrives at an account where "disinformation is information that is intentionally misleading. That is, it is information that – just as the source of the information intended – is likely to cause people to hold false beliefs" (Fallis, 2014, p. 137). In order to capture adaptive (or evolutionary) disinformation Fallis (2015) broadens the definition such that "disinformation is misleading information that has the function of misleading someone" (Fallis, 2015, p. 413). Fallis (2015) specifies that a function can be acquired in two different ways either by evolution or by design – in both cases the misleading is non-accidental:

> Disinformation can acquire the function of misleading people in either of these two ways. Most forms of disinformation, such as lies and propaganda, are misleading because the source intends the information to be misleading [designed function]. But other forms of disinformation, such as conspiracy theories and fake alarm calls, are misleading simply because the source systematically benefits from their being misleading [evolved function] (Fallis, 2015, p. 413).

In both formulations of disinformation, Fallis defines information as "something that has representational content" (Fallis, 2014, p. 137) regardless of its truth value. Fallis (2015)

elaborates "information is something that represents some part of the world as being a certain way. In other words, it is something that has semantic (or representational) content […]. For instance, the text "The cat is on the mat" represents the cat as actually being on the mat; also a photograph of the cat can represent it as being on the mat." (Fallis, 2015, pp. 404-405). However, the cat need not actually be on the mat in order for the text to represent it. As Fallis puts it "I take the term information to refer to representational content that is false, as well as to representational content that is true" (Fallis, 2015, p. 406). Thus, on Fallis' account information is alethically neutral in the sense that "meaningful and well-formed data qualify as information, no matter whether they represent or convey a truth or a falsehood or have no alethic value at all." (Primiero, 2016, p. 101) – to put it in Floridian terms.

The Gricean implicature lies underneath both formulations of disinformation as "misleadingness" – in the literature on lying, misleading, and deceiving – is defined in terms of Gricean implicatures. Thus, something can become misleading due to the implicatum of the utterance. That is, it is the meaning of the implicature, which can render the utterance misleading. It is important to note that according to Mahon (2008) "misleadingness" is simply a feature of causing someone to hold a false belief independently of the means used to cause this false belief. Thus, in addition to implicatures, "misleadingness" can result from inaccuracies and mistakes, and can be generated by language, by pictures, or by gestures and other non-verbal acts.

Although the main focus is on disinformation, Fallis (2014, 2015) also provides some brief definitions of misinformation. Roughly put, misinformation is simply inaccurate information – that is, inaccurate representational content – of various kinds. Thus, in a footnote Fallis (2015) writes: "We should probably say that misinformation is information that is inaccurate and misleading; in fact, according to Skyrms (2010, p. 80), misinformation simply is misleading information" (Fallis, 2015, p. 423, note 7). On Fallis' accounts it is the intentional or non-accidental misleading which distinguishes disinformation from misinformation. "Inaccurate information (or misinformation) can mislead people whether it results from an honest mistake, negligence, unconscious bias, or (as in the case of disinformation) intentional deception" (Fallis, 2014, p. 136).

Thus, Fallis defines information, misinformation, and disinformation as representational content, where it is further specified that misinformation is inaccurate and misleading and disinformation is intentionally or non-accidentally misleading. The misleadingness is often generated through Gricean implicatures – especially in instances of disinformation.

5. Discussion

Each of these four philosophical accounts have different implications – and complications – for the possibility of automatic detection of misinformation or disinformation on the internet and they have different problems with regard to the definitions of information, misinformation, and disinformation. The interconnections between these three notions – as well as the definitions of each notion – are not very clear. However, the four philosophical accounts provide some insights with regard to the underlying assumptions guiding the goal of automatic detection. For instance, the PHEME-project's adherence to detection of truth and falsity as determinant for information vs mis- and disinformation can be found in the accounts by Dretske and Floridi and as such follows a tradition within philosophy of information. However, the accounts by Fox and Fallis question this assumption as they operate with alethic neutrality for information. Thus, further discussions of information, misinformation, and disinformation are required in order to get a clearer picture of the interconnections of these notions and thereby to get a clearer picture of how to define information, misinformation, and disinformation, respectively.

5.1 Misinformation vs disinformation

If we, for a moment, return to the detecting-projects one of the main challenges is the distinction between misinformation and disinformation. Recall that it is the definitions of misinformation, which vary the most between the four projects examined (cf. Section 2). The variations regard the question of intentions – is misinformation intended or unintended inaccuracy or falsity? Both the detecting-projects and the philosophical accounts acknowledge that there is a difference between misinformation and disinformation, yet it is not very clear what that difference actually is.

Within common dictionaries[10] and the journalistic literature[11] on misinformation and disinformation – i.e. the sources the detecting-projects adhere to – two different tracks are present. Either, misinformation and disinformation can be treated as synonyms, or they can be distinguished in terms of intentions and deception – that is, to define misinformation as unintended false, inaccurate, or misleading information and to define disinformation as false, inaccurate, or misleading information intended to deceive and/or mislead. The common trend within journalism seems to be to treat the two notions as synonyms and generally to stick with the notion of misinformation to denote all kinds of false, misleading, inaccurate, and deceptive information (e.g. Thorson, 2016; Wardle, 2017). The use of "misinformation" in lieu of all false or inaccurate content (i.e. intended, unintended, misleading, deceiving, and the like) underpins an understanding of the difference between information and misinformation in terms of truth and falsity. Information is the true part that shall be preserved, guarded, enhanced, and spread. Misinformation is the false part that shall be avoided, combated, suppressed, and stopped. When misinformation and disinformation are treated as synonyms there is no differentiation between intentional and purposeful misleading and unintended misleading such as honest mistakes, inaccuracies due to ignorance, etc. Thus, all kinds of misleadingness are treated equally and the goal becomes to guard against them all.

Within the theoretical and philosophical accounts of information, misinformation, and disinformation it is more common to treat misinformation and disinformation as two distinct notions instead of treating them as synonyms (e.g. Dretske, 1983; Floridi, 2005b, 2011; Fallis, 2009, 2011, 2014, 2015). The treatment of misinformation and disinformation as two distinct notions is in line with the general conception – within the literature on lying, misleading, and deceiving – that it is necessary to distinguish between lies (believed-false statements), misleadingness (based on inaccuracies or implicatures both verbal and gestural), and deception (successful and intentional misleading and lying) (cf. Mahon, 2008). The distinction between misinformation and disinformation is cast in terms of intentions and possible deception: misinformation is defined as false or inaccurate content in general and then disinformation is defined as that part of misinformation which is purposefully false, inaccurate, or misleading (and possibly deceptive) – i.e. the intended or intentionally/non-accidentally misleading part. Note that when disinformation is defined as the purposeful misleading part of misinformation, then there are no requirements for misinformation in terms of intentions and intentionality. For instance, it cannot be specified that misinformation is unintended misleading when disinformation as intentional misleading is a part of misinformation. Intentional misleading cannot be a subset of unintended misleading.

However, as misinformation is often referred in terms of honest mistakes, bias, unknown inaccuracies, ignorance, and the like[12] and a distinction between misinformation and disinformation is upheld it is reasonable to define the two notions as fully distinct concepts, where disinformation is not a part of misinformation. To single out disinformation as a fully distinct concept from misinformation enables a specification of misinformation as unintended misleading due to inaccurate or false content (cf. Søe, 2016). More specifically, I define misinformation as unintended misleadingness, inaccuracy, or falsity, whereas

disinformation is defined as intentional misleadingness, inaccuracy, or falsity. Thus, the distinguishing features between misinformation and disinformation are intentions and intentionality: The unintended vs the intentional (non-accidental) misleadingness, inaccuracy, and/or falsity[13]. "Misleadingness" is understood as the quality of being misleading. "Misleading" is the ability to lead others astray – to lead them in the wrong direction – for instance by having a propensity to cause false beliefs. It can be further emphasized that misleadingness is not a success term (Fallis, 2015; Mahon, 2008). This means that something can be misleading independent of whether someone is actually misled – as long as it has a propensity to mislead. Thus, misinformation and disinformation as defined in terms of misleadingness are not success terms either. According to Mahon (2008) deception is an achievement or success term. Thus, disinformation is not always deceptive. Disinformation can be deceptive, and is deceptive when it works and someone is actually misled, but it is not necessarily deceptive. Therefore, deception is not a necessary characteristic of disinformation.

5.2 True disinformation

Throughout his writings on disinformation, lies, and deception, Fallis (2011, 2014, 2015) argues in favor of true disinformation as a specific verbal or linguistic variety of disinformation. True disinformation obtains its misleadingness due to a false Gricean implicature. False implicatures are implicatures where the implicatum – i.e. what is implicated – is false. The specific feature of true disinformation is that the false implicature is made by saying something true. That is, what is literally said (or written) is true but what is implicated (the implicature) is false, whereby the utterance (or post/tweet) as a whole becomes misleading. A classic example of a false implicature is Adler's (1997) Nevada-example, which is rehearsed in Fallis (2014): "For instance, if a villain who means to harm my friend asks me where he is, I might truthfully reply, 'He's been hanging around the Nevada a lot' intending the villain to draw the false conclusion that my friend could be at this diner now (see Adler 1997, pp. 437-438)." (Fallis, 2014, p. 138). Thus, if the implicature works the villain is misled into believing something false by an utterance of something literally true.

"Falsely implicating" is widely accepted as a vehicle for misleading (Adler, 1997; Grice, 1967; Mahon, 2008; Stokke, 2013; Webber, 2013). To falsely implicate – i.e. to say something literally true and implicate something false – is the mechanism which Fallis (2011, 2014, 2015) terms true disinformation when it is intentional or non-accidental. When true disinformation spreads, for instance through an online network, the intention to mislead – or the foreseeability that something is misleading – might disappear. If someone is misled by true disinformation and chooses to pass it on with the belief that it is accurate, genuinely true, and non-misleading (or simply without realizing that it is misleading) then it is a case of unintended misleading. However, that the intention to mislead is gone does not alter that it is a case of falsely implicating while saying something literally true. Thus, Fallis' argument in favor of true disinformation can be extended in order to account for true misinformation as a new variety of misinformation – i.e. the unintended misleading by a false implicature in saying something which is literally true (cf. Søe, 2016 for the full argument of this extension). As already mentioned, implicatures are not the only means for generating misleadingness. According to Fallis (2015) and Mahon (2008) it is possible to deceive (i.e. intentionally and successfully mislead) through acts of omission. In these cases the deception is caused by the leaving-out of some information, which, if it was provided, would have prevented the misleading. In the same way true disinformation and true misinformation can be generated through acts of omission or negligence – where such acts might not count as false implicatures (e.g. if they are non-verbal). If something (an utterance, a picture, etc.) in itself is truthful but becomes misleading due to omission of

vital details then it is a case of true disinformation. If the omission is the result of a mistake or negligence, etc. then it is a case of true misinformation.

True disinformation and the extension to true misinformation is a problem for the detecting-projects because it challenges the true/false dichotomy for information vs mis- and disinformation. If only truth and falsity are detected then true misinformation and true disinformation will not be detected as misinformation and disinformation. To borrow the terminology from Floridi, true mis- and disinformation will count as well-formed, meaningful data, which is truthful. That is, due to the literal truth, yet misleading character, of what is written true misinformation and true disinformation enter the domain of information and will be detected as such. Otherwise, the algorithms should be able to detect the falsity of the Gricean implicatures. That means that the algorithms should be able to work out the implicature and recognize the discrepancy between the literal meaning and the meaning of the implicatum and further be able to recognize that the meaning of the implicatum is misleading within the specific context. Further, in order to capture cases of omission or neglect, the algorithms should be able to "know" what has been left out – i.e. what has been omitted or neglected.

In order to emphasize that the true/false dichotomy does collapse due to the possibility of true mis- and disinformation it is worth considering a possible objection to the argument. The objection is that the true/false dichotomy is not actually challenged by true mis- and disinformation and cannot easily be abandoned. At the core of the objection lies the argument that something false will always be present in cases of misleadingness – e.g. the false implicatum of an implicature – thus, in the end it will always be a matter of truth-values. However, in cases of misleadingness by omission, neglect, or ignorance it is not clear that anything false needs to be present besides the false beliefs obtained by those who are misled. The true/false dichotomy for information vs mis- and disinformation is formulated in connection to semantic content only. Thus, it can neither account for pragmatic meaning generated by implicatures nor the beliefs obtained based on any semantic content – misleading or not. Further, it is not clear how a false belief caused by misleadingness due to inaccuracy, neglect, omission, and the like could be regarded as part of the content of the misinformation or disinformation. If the false belief is somehow included as part of the content then every proposition, utterance, picture, gesture, etc. – i.e. all semantic content – run the risk of including something false. Such a result seems to preclude a distinction between misleadingness and non-misleadingness in the first place as everything becomes potentially misleading in case anyone obtains a false belief.

Implicatures can also work in the other direction in the sense that it is possible to implicate something true by saying something, which is literally false. This is for instance the case with satire and irony. Most often satire and irony are not misleading because the implicature is true and most people will understand the implicature – they will work it out. However, this does not change the fact, that what is literally said is false. This means, that if only truth and falsity are detected for then satire and irony would be detected as misinformation due to their false semantic content. If the purposefulness of the falsity – i.e. that satire and irony are made intentionally – is taken into account by the algorithms then satire and irony would be detected as disinformation.

As already mentioned, the implication of true misinformation and true disinformation for the purpose of automatic detection is that the mere detection of truth and falsity on a semantic level will not provide the correct classification of information, misinformation, and disinformation in all instances. True misinformation and true disinformation will end up in the information category and will not be detected as misinformation and disinformation. As the possibilities of true misinformation and true disinformation are based on the notion of false implicature and acts of omission which are widely accepted in the literature on lying, misleading, and deceiving, as well as foreseen by Grice himself, it cannot easily be argued

that these possibilities should not be accounted for in automatic detection. Thus, the detection of truth as an indicator of information cannot be upheld.

The collapse of the true/false dichotomy for distinguishing between information and mis-/disinformation has the further implication that something else must be the distinguishing feature of these notions. As implicatures can work in various ways and in both directions – false implicatures by saying something true, true implicatures by saying something false, true implicatures by saying something true, false implicatures by saying something false – it seems that in the end, what matters is whether what is said, written, or posted, etc. is misleading or not. Thus, it is possible to distinguish between information and mis-/disinformation in terms of non-misleadingness and misleadingness. Together with the distinguishing feature of intention/intentionality it is possible to speak of information as intentional non-misleading, misinformation as unintended misleading, and disinformation as intentional (non-accidental) misleading, where all three notions are alethically neutral, which means that they do not have fixed truth-values. All three notions can be referred to collectively as representational content in general.

5.3 False vs misleading

The notion of implicature as a vehicle for misleadingness (as well as part of language in general) implies that misinformation and disinformation (as well as information) are pragmatic notions. (Bear in mind that implicatures are not the only vehicles for misleadingness, but they are the specific vehicles that together with acts of omission enable true disinformation and true misinformation causing the true/false dichotomy to collapse.) Some semantic content can become misleading because it – besides its semantic and literal meaning – has a pragmatic meaning which lies beyond what is literally said and which implicates something that is inaccurate. Thus, pragmatic meaning is not necessarily misleading in itself. The misleadingness is determined by the connection of, or interplay between, various factors: the context of the utterance, the semantic meaning, and the pragmatic meaning. In order to detect misleadingness, for instance, as a feature of Gricean implicatures, one must be able to assess the context, the semantic meaning, and the pragmatic meaning. In Gricean words, one must be able to "work out" the implicature – that is to be able to make the right inferences – and then be able to recognize that the implicature is misleading. This means that the detecting-algorithms have to be able to work on a pragmatic level (which includes context, utterer's meaning, intentions, and beliefs, etc.) and not just on a semantic level. Otherwise, the algorithms will not be able to detect misinformation and disinformation. In other words, misleadingness is a pragmatic feature of meaning. Also in cases of omission – especially if these are intentional or non-accidental – as pragmatic meaning has a part to play here as well (recall nonnatural meaning as a pragmatic notion defined in terms of utterers' intentions). Thus, in order to detect information as intentional non-misleading, misinformation as unintended misleading, and disinformation as intentional (non-accidental) misleading the algorithms will have to employ two distinguishing features: intentions/intentionality in order to discriminate between something which is unintended and something which is intentional; and pragmatic meaning in order to discriminate between non-misleadingness and misleadingness.

Thus, there is shift from detection of falsity to detection of misleadingness as pointing to misinformation and disinformation. As well as a shift from detection of truth to detection of non-misleadingness as pointing to information.

Intentions and misleadingness (as a pragmatic feature of meaning) point to Grice's distinction between natural and nonnatural meaning. The algorithms have to work within the domain of nonnatural meaning – i.e. the kind of meaning which is determined by conventions, language, and intentions. In short, communication and the cooperative principle. Every tweet, Facebook post, meme, etc. is made within a specific context.

Behind them lie specific intentions, beliefs, or other mental states, as to why the tweet is made, the post is written, the picture is shared, etc. If the algorithms deploy a notion of information as fact, states of affairs, necessarily true, or the like – i.e. deploy a Dretskian notion of information as natural meaning – all these features, context, meaning, intentionality, etc. are lost within the detection.

For instance, in Grice's example of Mr X who is requiring a testimonial for a job application (cf. Section 3.2) the implicatum of:

(1) "Mr X's command of English is excellent, and his attendance at tutorials has been regular" (Grice, 1967, p. 33);

can only be worked out in connection to the context of Mr X applying for a philosophy job. If this context is not taken into account the intention behind the statement, the utterer's meaning, is lost and the statement is reduced to its semantic and literal meaning. The implicatum that Mr X is no good at philosophy is lost. In Grice's original example, the implicature about Mr X is a true implicature – that is, it is correct that Mr X is no good at philosophy. However, statement (1) could be used to make a false implicature in a case where Mr X was in fact good at philosophy but the writer of the testimonial for some reason wanted to mislead the receiver about this. In such a case, the true statement that Mr X is good at English, etc. will be true disinformation about Mr X because it falsely implicates that he is no good at philosophy while truly stating that his command of English is excellent, etc. Thus, statement (1) is literally true in both instances but dependent on the context it is either intentional non-misleading information implicating a lack of philosophy skills or it is intentionally misleading and thereby true disinformation falsely implicating a lack of philosophy skills. In yet other contexts, the implicatum might be something completely different.

Truth and falsity are not very useful categories by themselves when it comes to distinguishing between information, misinformation, and disinformation in any form and variety – in natural language, through pictures or gestures – and especially not when it comes to actions which in themselves cannot be true or false (they can be right or wrong) but can still truly or falsely implicate all various sorts of things in different contexts.

For instance, if Ben buys a book for his friend online, then the company selling the book algorithmically collects and stores Ben's purchase as a piece of information. If the algorithms deploy a Dretskian notion of information and are assumed to operate within a domain of natural meaning, then the underlying assumption is that the purchase is information about Ben – in some sense. Thus, Ben's purchase feeds into the pile of information connected to him. It is not reflected within the algorithmic collection of the purchase that the book is for Ben's friend and therefore does not necessarily say anything about Ben's reading habits and preferences – it only says something about Ben's actual purchase. Actually, it might be that Ben's friend loves Sci-Fi novels while Ben himself is more into biographies. Thus, if the purchase of a Sci-Fi novel for Ben's friend is algorithmically collected as information in the Dretskian sense (i.e. as natural meaning) it follows that the purchase is collected as (truthful) information related to Ben. The default assumption seems to be that Ben buys the Sci-Fi novel because he likes it or likes books of that type – at least it seems to be the default assumption guiding the recommender systems which will tell Ben about "other books you might like." When the algorithmic data collection does not reflect that the purchase is for Ben's friend, the purchase becomes misleading in regard to Ben's preferences – the data are fed into the recommendation for Ben. The collected data correctly reflects that Ben bought the Sci-Fi novel, however, it is not truthful information about Ben's preferences in books. This Ben example is an illustration of how algorithms can enable misleadingness through default assumptions about the truthful nature of links between people and their actions. It is an illustration of non-verbal

misleadingness executed by algorithms based on the assumptions the developers of these algorithms build into them. The problem for Ben is that the context of his purchase – that the book is for a friend – is not taken into account. These examples (i.e. "Ben" and "Mr. X") further illustrate that the mere detection of truth and falsity – which seems to be the current strategy for the algorithms designed to detect misinformation and disinformation – will not yield the appropriate results on various occasions as to what is actually misinformation and disinformation. As already stated what counts in order to distinguish between information, misinformation, and disinformation is whether some semantic content is non-misleading or misleading and whether the misleadingness is intentional (non-accidental) or unintended. These distinctions hold for all the varieties of disinformation and misinformation. Adaptive disinformation, visual disinformation, true disinformation, disinformation by omission, and side-effect disinformation are all instances of intentionally/non-accidentally misleading information. True misinformation, honest mistakes, and misinformation by neglect are all instances of unintendedly misleading information.

5.4 Algorithms as communicative actions

To detect whether something is misleading or not is a difficult task (if not impossible). Misleadingness as a pragmatic feature of meaning is generated through a mixture of context and content. Whether some content (i.e. representational content of any kind) in a specific context is misleading for a given person is further determined by what that person knows in advance and therefore what he is willing to believe. What people know in advance and thereby whether some content is misleading for them or not is very difficult to take into account when assessing whether some tweet, picture, or Facebook post is misleading or not. It varies with each individual. Furthermore, as already mentioned misleadingness is not a success term (Mahon, 2008). This means that nobody actually has to be mislead. Thus, some content can be misleading without actually misleading anybody. Therefore, Fallis (2015) argues that for something to be disinformation it has to have the propensity to mislead – that is, it has to be more likely to create false beliefs than not. "It should be noted that while a piece of disinformation must have the propensity to mislead, it does not have to actually mislead someone on any given occasion. Just like lying, disinformation is not a success term" (Fallis, 2015, p. 406). Thus, it should be possible to assess whether or not something has the propensity to mislead (i.e. to assess misleadingness) without having to know what every possible receiver knows in advance.

Whether automatic detection of misinformation and disinformation is possible is a question of whether automatic detection of misleadingness/non-misleadingness and intentions is possible. Blair (1992) makes an important point – in connection to information retrieval – which bears some influence on the possibility of automatic detection. Blair (1992) argues that one of the main reasons why Grice (1967) is significant for information retrieval is that "Grice showed that the decoding theories could not account for, among other things, how a hearer could tell when a speaker mis-spoke: for example, when the speaker made a mistake or lied (i.e. simple decoding theories would always take language literally, something that is clearly not the case in natural discourse)." (Blair, 1992, p. 204). When natural language, and especially implicatures, are involved the question for automatic detection is a question of pragmatics – that is, whether algorithms will always take language literally or whether they can detect the pragmatic features of natural discourse (communication). Add to this all the examples of non-linguistic communication – pictures, gestures, etc. – which can also generate misleadingness and thereby be instances of misinformation or disinformation. In the non-linguistic cases the misleadingness is also a pragmatic feature of meaning generated through a mixture of context and content. Misleadingness operates in the domain of nonnatural meaning – the kind of meaning explained in terms of people and their intentions to communicate certain things.

Goffey (2008) makes a similar point in his chapter on algorithms from *Software Studies: A Lexicon*: "Because pragmatics connects language to extrinsic factors, it becomes impossible to conceptualize a language as a self-sufficient system closed in on itself." (Goffey, 2008, p. 17). The challenge is that "with algorithms, formalization comes first, the express aim being to divorce (formal) expression from (material) content completely." (Goffey, 2008, p. 17). Goffey's point is that the pragmatics have been neglected as syntax and semantics have been seen as sufficient for algorithmic operations causing "the leap from the theoretical world to the practical world a difficult one to accomplish. Always the trivia of implementation details." (Goffey, 2008, p. 17). These insights marks the beginning of a bursting literature on algorithms as cultural phenomena and as such question the algorithm as a neutral and formal entity[14].

Blair (1992) and Goffey (2008) explicitly deal with natural language and the problems which arise when natural language has to be translated into algorithms and code. Although this paper deals with both linguistic and non-linguistic instances of representational content the points Blair and Goffey make provide some insights as to why pragmatics are not satisfactorily accounted for in the concepts of information, misinformation, and disinformation as described by the detecting-projects.

In some sense the detecting-algorithms are themselves communicative actions. The algorithms are defined and coded in a specific language using if-then rules – sometimes based on yes/no questions. Thus, within the algorithms lie instructions of the sort "if A then B", "if C and D then E", etc. – i.e. "if false then misinformation", "if false and intentionally so, then disinformation" (if the definitions from the PHEME-project are adhered to). Note that these "linguistic operations" are also exerted over content which is not linguistic in nature (i.e. picture, emoticons, etc.). The idea of the algorithm as a communicative action bears some resemblance to Blair's arguments for information retrieval as "communication between indexers and searchers" (Blair, 1992, p. 204) – i.e. asking questions and getting answers. The detecting-algorithms are, metaphorically speaking, designed to ask "What is this? Is it information, Y/N? Is it misinformation, Y/N? Is it disinformation, Y/N?" then getting the answers from the tweet, picture, or post itself. Algorithms as communicative actions further underpins the point that algorithms operate in the domain of nonnatural meaning (in Grice's sense). Actually the formalization, which, as Goffey (2008) points out, comes first in the case of algorithms, already marks the shift to the domain of nonnatural meaning because formalization itself is interpretation. Formalization is interpretation and representation and thus, the mere act of formalizing is already defined by meaning and intentions. Thus, within the formalization meaning and intentions are present and these become part of the algorithms when these are coded using the formalizations. Further, the coding itself is based on purpose and functionality of the algorithm (i.e. what the algorithm is for) together with meanings, ideas, and intentions of the coder – for instance, to detect disinformation together with what the coder depicts disinformation to be.

6. Conclusion

People communicate. They communicate in various ways – online, offline, through language, gestures, and pictures. Even algorithms are communicative actions based on formalizations and assumptions from their human developers. Sometimes communication can be misleading. The misleading or erroneous aspects of communication have prompted various attempts to develop algorithms for automatic detection of misinformation and disinformation in online social network structures. However, the main focus in these attempts has not been on communication but instead on truth-values of what is written or posted.

In the four detecting-projects examined in this paper it is truth (as pointing to information) and falsity (as pointing to mis-/disinformation) which are detected for in order to identify misinformation and disinformation as opposed to information. The underlying

assumption is that truth and falsity are the distinguishing features between information and mis-/disinformation, wherefore, truth and falsity are sufficient and necessary features of information, misinformation, and disinformation, respectively.

However, the picture is more complicated. The dominant focus on truth and falsity disregards the communicative aspects of online sharing, posting, liking, etc. For communication to work context, intention, belief, and meaning makes all the difference. Whether some post is misleading (intentionally or by mistake) is dependent upon its meaning – and the meaning (the nonnatural meaning) is determined by the context in which it is posted and the intentions with which it is posted (recall that nonnatural meaning in Grice's theory is reduced to three intentions on behalf of the utterer). This is the insight provided by the Gricean framework of communication, cooperation, and meaning. Communication is about meaning in context, and meaning is about intentions (Grice, 1957, 1967).

Within the Gricean insight regarding communication lies the insight that misleadingness – and not falsity as such – is the vehicle of misinformation and disinformation. The misleadingness of misinformation and disinformation can be generated by (false) Gricean implicatures (Fallis, 2014; Mahon, 2008), by acts of omission, by inaccuracies, etc. What is literally true can implicate something false enabling true misinformation and true disinformation as varieties of mis- and disinformation. What is literally false can implicate something true enabling literally false yet non-misleading information. I argue that these varieties of information, misinformation, and disinformation – where true misinformation is my development from Fallis' notion of true disinformation – challenge the true/false dichotomy for information vs misinformation and disinformation leaving truth and falsity insufficient for detection of the three notions. This means, that if truth and falsity are all the algorithms are developed to detect, then the algorithmic classification of information, misinformation, and disinformation will not yield the appropriate results on various occasions. For instance, true mis- and disinformation will be categorized as information, whereas false yet non-misleading information will be categorized as mis- or disinformation. Therefore, the real distinguishing features between information, misinformation, and disinformation are non-misleadingness/misleadingness and intentions/intentionality (where "intentions" also cover the notion of "unintended").

Information, misinformation, and disinformation are all instances of representational content in general[15] and are as such alethically neutral – i.e. they can be either true or false or have no alethic value. In detection of information, misinformation, and disinformation the primary features to detect are non-misleadingness and misleadingness. Second comes intention/intentionality in order to distinguish between intentional (non-accidental) and unintended misleading. This leaves truth and falsity tertiary features, which do not determine whether something is information, misinformation, or disinformation – although the representational content in question will often have truth values.

More specifically, the distinction between information, on the one hand, and mis- and disinformation, on the other hand, is that information is non-misleading (and intentionally so), whereas misinformation and disinformation are misleading. The distinction between misinformation and disinformation is then that misinformation is unintended misleading, whereas disinformation is intentionally (non-accidentally) misleading:

- information: intentionally non-misleading representational content;
- misinformation: unintended misleading representational content; and
- disinformation: intentionally (non-accidentally) misleading representational content.

The "hierarchy" of the distinguishing features – i.e. first non-misleadingness/misleadingness, then intention/intentionality, then truth-values – is in play whether the detection is algorithmic or done by a human browsing through a newsfeed.

Based on the conceptual clarification of information, misinformation, and disinformation the question is whether automatic detection of non-misleadingness and misleadingness, intentions and intentionality is feasible. For instance, non-misleadingness and misleadingness are pragmatic features of meaning. Especially misleadingness is on the pragmatic side of language because it can be generated through implicatures which works because of the differences between literal meaning and utterer's meaning. Thus, to detect non-misleadingness and misleadingness requires assessment of content, context, literal meaning, intentions, and the like in order to determine the utterer's meaning, hence the implicature (if any is present) and to work it out. Also in cases of non-linguistic representational content (i.e. pictures, etc.) detection of non-misleadingness and misleadingness requires assessment of content, context, meaning, and intentions. Therefore, to automatically detect misleadingness and non-misleadingness requires that algorithms are capable of working out implicatures and in general determine and "understand" pragmatic meaning. The writings of Blair (1992) and Goffey (2008) suggest that algorithms do not have this capacity as they are not capable of dealing with natural language – they are not capable of dealing with pragmatic meaning. This further suggests that automatic detection of misinformation and disinformation is not possible. As Goffey points out "with algorithms, formalization comes first, the express aim being to divorce (formal) expression from (material) content completely." (Goffey, 2008, p. 17). However, this seems to conflict with a Gricean understanding of language where pragmatics are determinant for whether something is misleading or not.

Notes

1. Grice uses the word "utterance" in a neutral way that includes both verbal and non-verbal actions (cf. section 3.1). Grice's usage is followed throughout this paper.
2. *Meaning* is Grice's second official publication. It was published in *The Philosophical Review*, 66, in July 1957. In *Studies in the Way of Words* (1989) Grice has added the year 1948 to *Meaning*, which indicates that it was written nine years prior to its publication.
3. By Dretske referred to as "The Mathematical Theory of Information" or "Communication Theory" (Dretske, 1981, p. 3 and p. 237, note 1).
4. Fomal modeling is a method for representing the world through mathematically defined models based on specific rules and equations.
5. "[…], false information and mis-information are not kinds of information – any more than decoy ducks and rubber ducks are kinds of ducks" (Dretske, 1981, p. 45).
6. In relation to some further points about the Quality Maxim (cf. section 3) Grice states that "[f]alse information is not an inferior kind of information; it just is not information" (1989, p. 371). However, Grice never offers an argument as to why it is so.
7. Floridi uses the word "truthful" rather than "truth" in his definition of semantic information in order to be able to capture instances of non-linguistic information which are not normally considered to be true or false. As examples Floridi mentions pictures, maps, gestures, and the like as instances of non-linguistic semantic information which can be said to be truthful rather than true.
8. Alethic neutrality means that alethic values (i.e. truth-values such as "truth" and "falsity", or no alethic value) are not embedded in but supervene on the concept in question (here information). On Fox's account misinformation has a fixed alethic value, i.e. "false," whereas information is neutral regarding alethic values and thus can take both the value "true" and the value "false."
9. If an utterance is not being heard and understood by anybody, then it is still an act of saying, but it is not an act of telling (Fox, 1983).
10. For example, Dictionary.com; Merriam-Webster, Learner's Dictionary; Oxford Dictionaries; Oxford English Dictionary; Vocabulary.com; collinsdictionary.com; dictionary.cambridge.org; and macmillandictionary.com.

11. Cf. Wardle (2016) for an ongoing reading list of journalistic articles on mis-/disinformation.
12. For example, Fallis (2009, 2011, 2014, 2015), PHEME (2014), Karlova and Fisher (2013), and Shao *et al.* (2016), as well as various dictionaries such as Dictionary.com; Merriam-Webster, Learner's Dictionary; Oxford Dictionaries; Oxford English Dictionary; and Vocabulary.com.
13. Fallis (2009, 2011, 2014, 2015) argues in favor of defining disinformation as intentional misleading rather than intended misleading in order to capture instances of side effect disinformation. "Intentional" here means a directedness or foreseeability that something is misleading. That is, even though it is not the intention to mislead it can be foreseen that what is said or written is misleading and therefore might actually mislead someone. Due to this foreseeability side effect disinformation is still disinformation even though it is not intended to mislead. Fallis emphasizes that in most cases disinformation is also intended to mislead.
14. For example, Gillespie (2014), Sandvig (2015), Striphas (2015), Bucher (2016), Burrell (2016), and Crawford (2016) – to mention a few.
15. This includes all the different varieties of representational content – text, pictures, gestures, etc.

References

Adams, F. and de Moraes, J.A. (2016), "Is there a philosophy of information?", *Topoi*, Vol. 35 No. 1, pp. 161-171.

Adler, J. (1997), "Lying, deceiving, or falsely implicating", *Journal of Philosophy*, Vol. 94 No. 9, pp. 435-452.

Adriaans, P. (2012), "Information", in Zalta, E.N. (Ed.), *Stanford Encyclopedia of Philosophy*, Fall 2013 ed., Stanford University, Palo Alto, 35pp., available at: http://plato.stanford.edu/archives/fall2013/entries/information/

Adriaans, P. and van Benthem, J. (Eds) (2008), *Philosophy of Information*, Elsevier B.V., Amsterdam and Oxford.

Bates, M.J. (2006), "Fundamental forms of information", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 8, pp. 1033-1045.

Blair, D.C. (1992), "Information retrieval and the philosophy of language", *The Computer Journal*, Vol. 35 No. 3, pp. 200-207.

Blair, D.C. (2003), "Information retrieval and the philosophy of language", *Annual Review of Information Science and Technology*, Vol. 37 No. 1, pp. 3-50.

Brookes, B.C. (1980), "The foundations for information science: part I. Philosophical aspects", *Journal of Information Science*, Vol. 2, pp. 125-133.

Bucher, T. (2016), "The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms", *Information, Communication & Society*, pp. 1-16, doi: 10.1080/1369118X.2016.1154086.

Buckland, M.K. (1991), "Information as thing", *Journal of the American Society for Information Science*, Vol. 42 No. 5, pp. 351-360.

Budd, J.M. (2011), "Meaning, truth, and information: prolegomena to a theory", *Journal of Documentation*, Vol. 67 No. 1, pp. 56-74.

Burrell, J. (2016), "How the machine 'thinks': understanding opacity in machine learning algorithms", *Big Data & Society*, pp. 1-12, doi: 10.1177/2053951715622512.

Capurro, R. and Hjørland, B. (2003), "The concept of information", *Annual Review of Information Science and Technology*, Vol. 37 No. 1, pp. 343-411.

Carnap, R. and Bar-Hillel, Y. (1952), "An outline of a theory of semantic information", Technical Report No. 247, Research Laboratory of Electronics, MIT, Cambridge, MA.

Cornelius, I. (2002), "Theorizing information for information science", *Annual Review of Information Science and Technology*, Vol. 36 No. 1, pp. 393-425.

Crawford, K. (2016), "Can an algorithm be agonistic? Ten scenes from life in calculated publics", *Science, Technology, & Human Values*, Vol. 71 No. 1, pp. 77-92.

Day, R.E. (2001/2008), *The Modern Invention of Information. Discourse, History, and Power*, Paperback edition, Southern Illinois University Press, Carbondale.

Dretske, F. (1981), *Knowledge and the Flow of Information*, MIT Press, Cambridge, MA.

Dretske, F. (1983), "Précis of knowledge and the flow of information", *The Behavioral and Brain Sciences*, Vol. 6, pp. 55-63.

Dretske, F. (1988), "Explaining Behavior", A Bradford Book, MIT Press, Cambridge, MA.

Fallis, D. (2009), A Conceptual Analysis of Disinformation, Preprint from iConference, Tucson, AZ, available at: www.ideals.illinois.edu/bitstream/handle/2142/15205/fallis_disinfo1.pdf?sequence=2

Fallis, D. (2010), "Lying and deception", *Philosophers' Imprint*, Vol. 10 No. 11, 42pp.

Fallis, D. (2011), "Floridi on disinformation", *Etica & Politica*, Vol. 13 No. 2, pp. 201-214.

Fallis, D. (2014), "The varieties of disinformation", in Floridi, L. and Illari, P. (Eds), *The Philosophy of Information Quality*, Springer, Cham, Heidelberg, New York, NY, Dordrecht and London, pp. 135-161.

Fallis, D. (2015), "What is disinformation?", *Library Trends*, Vol. 63 No. 3, pp. 401-426.

Fetzer, J.H. (2004a), "Information: does it have to be true?", *Minds and Machines*, Vol. 14, pp. 223-229.

Fetzer, J.H. (2004b), "Disinformation: the use of false information", *Minds and Machines*, Vol. 14, pp. 231-240.

Floridi, L. (2005a), "Is semantic information meaningful data?", *Philosophy and Phenomological Research*, Vol. LXX No. 2, pp. 351-370.

Floridi, L. (2005b), "Semantic conceptions of information", in Zalta, E.N. (Ed.), *Stanford Encyclopedia of Philosophy*, Spring 2013 ed., Stanford University, Palo Alto, 46pp., available at: http://plato.stanford.edu/archives/spr2013/entries/information-semantic/

Floridi, L. (2007), "In defence of the veridical nature of semantic information", *European Journal of Analytic Philosophy (EUJAP)*, Vol. 3 No. 1, pp. 31-41.

Floridi, L. (2008), "Trends in the philosophy of information", in Adriaans, P. and van Benthem, J. (Eds), *Philosophy of Information*, Elsevier B.V., Amsterdam and Oxford, pp. 113-131.

Floridi, L. (2011), *The Philosophy of Information*, Oxford Scholarship Online, Oxford, available at: www.oxfordscholarship.com

Fox, C.J. (1983), *Information and Misinformation. An Investigation of the Notions of Information, Misinformation, Informing, and Misinforming*, Greenwood Press, Westport, CT and London.

Furner, J. (2004), "Information studies without information", *Library Trends*, Vol. 52 No. 3, pp. 427-446.

Furner, J. (2010), "Philosophy and information studies", *Annual Review of Information Science and Technology*, Vol. 44 No. 1, pp. 154-200.

Furner, J. (2014), "Information without information studies", in Ibekwe-SanJuan, F. and Dousa, T.M. (Eds), *Theoreis of Information, Communication and Knowledge*, Springer Science + Business Media B.V., Dordrecht, Heidelberg, New York, NY and London, pp. 143-179.

Gillespie, T. (2014), "The relevance of algorithms", in Gillespie, T., Boczkowski, P.J. and Foot, K.A. (Eds), *Media Technologies: Essays on Communication, Materiality, and Society*, The MIT Press, Cambridge, MA, pp. 167-193.

Godfrey-Smith, P. (1989), "Misinformation", *Canadian Journal of Philosophy*, Vol. 19 No. 4, pp. 533-550.

Goffey, A. (2008), "Algorithm", in Fuller, M. (Ed.), *Software Studies: A Lexicon*, MIT Press, Cambridge MA, pp. 15-20.

Grice, H.P. (1957/1989/1991), "Meaning", *Studies in the Way of Words*, Paperback edition, First Harvard University Press, Cambridge, MA and London, pp. 213-223.

Grice, H.P. (1967/1989/1991), "Logic and conversation", *Studies in the Way of Words*, Paperback edition, First Harvard University Press, Cambridge, MA and London, pp. 22-40.

Grice, H.P. (1989/1991), *Studies in the Way of Words*, Paperback edition, First Harvard University Press, Cambridge, MA, and London.

Hjørland, B. (2007), "Information: objective or subjective/situational?", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 10, pp. 1448-1456.

Karlova, N.A. and Fisher, K.E. (2013), "A social diffusion model of misinformation and disinformation for understanding human information behavior", *Information Research*, Vol. 18 No. 1.

Kumar, K.P.K. and Geethakumari, G. (2014), "Detecting misinformation in online social networks using cognitive psychology", *Human-centric Computing and Information Sciences*, Vol. 4 No. 1, 22pp.

Mahon, J. (2008), "The definition of lying and deception", in Zalta, E.N. (Ed.), *Stanford Encyclopedia of Philosophy*, Fall 2019 ed., Stanford University, Palo Alto, 16pp., available at: http://plato.stanford.edu/archives/fall2009/entries/lying-definition/

Mai, J.-E. (2013), "The quality and qualities of information", *Journal of the American Society for Information Science and Technology*, Vol. 64 No. 4, pp. 675-688.

PHEME (2014), "About Pheme", available at: www.pheme.eu (accessed March 3, 2014).

Primiero, G. (2016), "Information in the philosophy of computer science. Ch. 10", in Floridi, L. (Ed.), *The Routledge Handbook of Philosophy of Information*, Routledge, London.

Robinson, L. and Bawden, D. (2014), "Mind the gab: transitions between concepts of information in varied domains", in Ibekwe-SanJuan, F. and Dousa, T.M. (Eds), *Theories of Information, Communication and Knowledge*, Springer Science + Business Media B.V., Dordrecht, Heidelberg, New York, NY and London, pp. 121-141.

Sandvig, C. (2015), "Seeing the sort: the aesthetic and industrial defense of 'the algorithm' ", *Journal of the New Media Caucus*, 18pp., available at: http://median.newmediacaucus.org/art-infrastructures-information/seeing-the-sort-the-aesthetic-and-industrial-defense-of-the-algorithm/

Søe, S.O. (2016), "The urge to detect, the need to clarify. Gricean perspectives on information, misinformation, and disinformation", PhD thesis, Faculty of Humanities, University of Copenhagen.

Scarantino, A. and Piccinini, G. (2010), "Information without truth", *Metaphilosophy*, Vol. 41 No. 3, pp. 313-330.

Shannon, C.E. (1948), "A mathematical theory of communication", *The Bell System Technical Journal*, Vol. 27, pp. 379-423.

Shao, C., Ciampaglia, G.L., Flammini, A. and Menczer, F. (2016), "Hoaxy: A platform for tracking online misinformation", *WWW'16 Companion*, Montréal and Québec, April 11-15, pp. 745-750, available at: http://dx.doi.org/10.1145/2872518.2890098

Sheffield News (2014), "EU project to build lie detector for social media", February 18, available at: www.sheffield.ac.uk/news/nr/lie-detector-social-media-sheffield-twitter-facebook-1.354715

Skyrms, B. (2010), *Signals. Evolution, Learning, & Information*, Oxford University Press.

Stokke, A. (2013), "Lying, deceiving, and misleading", *Philosophy Compass*, Vol. 8 No. 4, pp. 348-359.

Striphas, T. (2015), "Algorithmic culture", *European Journal of Cultural Studies*, Vol. 18 Nos 4-5, pp. 395-412.

The Π Research Network (2013), "The philosophy of information – an introduction", Version 1.0, The Society for the Philosophy of Information, available at: http://socphilinfo.org/sites/default/files/i2pi_2013.pdf

Thorson, E. (2016), "Belief echoes: the persistent effects of corrected misinformation", *Political Communication*, Vol. 33 No. 3, pp. 460-480.

Wardle, C. (2016), "(M|D)isinformation reading list", *First Draft News*, Cambridge, MA, available at: https://firstdraftnews.com/misinformation-reading-list/

Wardle, C. (2017), "Fake news. It's complicated", *First Draft News*, Cambridge, MA, available at: https://firstdraftnews.com/fake-news-complicated/ (accessed February 16, 2017).

Webber, J. (2013), "Liar!", *Analysis*, Vol. 73 No. 4, pp. 651-659.

Corresponding author
Sille Obelitz Søe can be contacted at: sille.obelitz@hum.ku.dk