

*The rise of the Pragmatic Web:
implications for rethinking meaning and
interaction*

Book or Report Section

Accepted Version

Jones, R. (2020) The rise of the Pragmatic Web: implications for rethinking meaning and interaction. In: Tagg, C. and Evans, M. (eds.) *Historicising the Digital : English Language Practices in New and Old Media*. de Gruyter Mouton, Berlin. ISBN 3110620391 doi:
<https://doi.org/10.1515/9783110670837-003> Available at
<http://centaur.reading.ac.uk/87378/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1515/9783110670837-003>

Publisher: de Gruyter Mouton

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other

copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Rodney H. Jones

The rise of the Pragmatic Web: Implications for rethinking meaning and interaction

Abstract

Much attention has been paid to the rise of the “Semantic Web” – a set of protocols for meanings to be shared across different platforms and systems of encoding to allow our machines to “talk” to each other. Less attention has been paid to the development of what has been called the “Pragmatic Web”, which consists of networks of tools, protocols and agents (both human and algorithmic) that interact to make and interpret meanings through *actions*. In some ways, however, advances in creating a “Pragmatic Web” have outstripped efforts to create a “Semantic Web”, a fact seen most clearly in the various algorithms that monitor the actions of users and transform them into inferences about the kinds of content they wish to be fed or the kinds of products they want to buy – algorithms that are at the heart of search engines like Google, recommender systems like that used by Amazon.com, and filtering systems like Facebook’s EdgeRank. The rise of the Pragmatic Web introduces challenges to scholars of pragmatics and discourse analysis, forcing us to rethink fundamental principles of meaning and interaction, including how implicature is created, how inferences are formed, the role of context in assigning meaning to action, and what constitutes a “speech act” to begin with. This chapter explores the ways digital communication involving networks of human and non-human agents is changing how we study pragmatics, and the implications of this for how we understand the social and ethical aspects of computer mediated communication.

Rodney H. Jones
University of Reading
212 Edith Morley Building
Whiteknights Campus
University of Reading
Reading UK
r.h.jones@reading.ac.uk

1. “This is your digital life”

At the time I was writing this chapter, social media was going through a bit of a hard time. It had just been revealed that a firm called Cambridge Analytica had harvested the data of millions of Facebook users and their friends by convincing them to take a ‘personality test’ called “This is your digital life”. The quiz was like so many of those that circulate through social media, promising users to reveal to them what character they would be in Shakespeare or how they would die in *Game of Thrones* in exchange for granting access to their data. In the end, Facebook had to admit that the data of over 87 million users was used by Cambridge Analytica to support the political campaigns of right-wing American candidates, including Donald Trump, as well as to influence the British Brexit vote. The scandal ignited debates not just about online privacy and ethics, but about the fundamental business model of social media.

In most of these debates, however, the key focus was “information” in a very conventional sense — what the media referred to as a “data breach”. And most people assumed that “the data” that had been “breached” consisted chiefly of the content of their

profiles— information such where they lived, the university they had attended, their five favourite movies, and who their friends were. In truth, however, this “explicit” information was the least useful information for Cambridge Analytica. What they were really after was the *implicit* information, the record of seemingly trivial actions people took when interacting with their friends — what they were really after was the ‘likes’. By combining a person’s history of “likes” (or what were subsequently reframed as “reactions”) with their profile information and the kinds of people they interact with, Cambridge Analytica was able to infer a great deal about people without having to delve into things like personal messages, status updates, photos, or all of the other information Facebook holds (Cadwalladr and Graham-Harrison 2018; Grassegger and Krogerus 2017).

Obviously, the issue here was not just “information”, but also the kinds of meanings that could be *inferred* from the information collected. For this, Cambridge Analytica used a sophisticated algorithm which it developed based on the research of Michal Kosinski and his colleagues (Kosinski et al. 2013), who had come up with a model that was able to predict users’ personality traits based on the posts they had ‘liked’ on Facebook. They found that on the basis of only 68 ‘likes’ it was possible to predict, with a high degree of accuracy, a user’s skin colour, political affiliation, sexuality, religion, alcohol, cigarette and drug use, and even whether their parents were divorced. The predictive power of this algorithm was then combined with the communicative power of thousands of “bots”, which were able to react instantly to trending topics and produce targeted posts for particular “microcategories” of people.

There are at least four things about this incident that are of interest to scholars of English language and discourse analysts. The first is the fact that the most valuable source of information for Cambridge Analytica came not so much from things that people said about themselves or others, but from things that they *did*, specifically the “speech act” of clicking the “like” button underneath a post in their Newsfeed. The second is that the important ‘meaning’ of these “likes” was not the fact that someone liked something, but rather “deeper meanings” that could be inferred from these acts of liking. Of course, many of us have the experience of trying to infer what people mean when they “like” certain things on social media (see Maíz-Arévalo 2017). The difference here was that the inferences that were drawn by Cambridge Analytica’s powerful algorithm were not formed based on individual reactions, but on the *trail* of reactions users created over time, correlated with the reactions of thousands of other users. Finally, it is notable that the main thing these inferences were used to do was to generate “membership categories” (Antaki and Widdincombe 1998; Sacks 1992), similar to the kinds of membership categories people use to make sense of others, predict their behaviour and tailor their communication to them. But again, these membership categories were different from the recognizable “social” categories that people use to negotiate their identities and make judgements about others. They were sophisticated aggregates of “measureable types” (Cheney-Lippold 2017), “microcategories” (like “straight male smoking vegetarian heavy metal fan who had grown up in a single parent family and made less than \$50K a year”), identities that combined hundreds of data points to predict what sort of messages users would respond to.

From the point of view of linguistics, then, the “meaning” of the information harvested by Cambridge Analytica was not so much a matter of the meanings of the words people wrote on their profiles or in their status updates, but of the meanings that could be *inferred* based on people’s actions, the informational contexts in which those actions were produced, and the kinds of identities that were associated with them. In other words, “meaning” here was not so much a matter of *semantics* as a matter of *pragmatics*. But what Cambridge Analytica’s algorithm was engaged in was no ordinary brand of pragmatics. It was, we might say, “pragmatics on steroids”.

This chapter is about the relationship between meaning and action in digitally mediated communication in the first decades of the early twenty-first century, and how that relationship is increasingly governed by non-human participants in our interactions. It argues that understanding how people interact with technologies (and how technologies interact with people) requires that we look beyond notions of “data” and “information” to the pragmatic dimensions of digitally mediated communication. I am less interested in what Yus (2011) calls “cyberpragmatics”, the term he uses to refer to the ways “different applications for Internet communication (chatrooms, Messenger, e-mail, Web pages, etc.) affect the quality and quantity of contextual information accessed by users” and the “inferential strategies that people engage in while processing information exchanged with other people using digital technology”, and more interested in how meanings are made and inferred in the kinds of complex sociotechnical assemblages such as the one described above. These are environments in which communication is not just a matter of people talking to people “through” applications, but also of people talking to applications, and applications talking to one another, environments that don’t just involve multiple media, with different affordances for inferring context, or different platforms with different norms for how information is shared and interpreted, but environments which link together different platforms and media as information is shared and processed by algorithms which shape the kinds of information human actors are exposed to and the kinds of interactions they can have in real time. The question I wish to address then is to what extent the repertoire of concepts and tools developed in “analogue pragmatics” (the pragmatics of unmediated human-to-human communication) since the mid-twentieth century, or even the “cyberpragmatics” (the pragmatics of “computer-mediated communication”) of the twenty-first, are adequate for understanding the pragmatics of human-algorithm communication, what I have elsewhere referred to as “algorithmic pragmatics” (Jones 2016a).

2. Historicizing the “Pragmatic Web”

The history of the internet is in many ways a history of human engagement with different ways of understanding meaning — how meaning is created and what can be done with it. The earliest way of seeing the internet was in terms of syntax— a model in which information derives meaning from its relationship to other information. The “syntactic web” is basically a web of interrelated static information linked together with hyperlinks. The way information is connected to other information and the “grammar” of those connections is what makes it possible for us to navigate through it in a meaningful way. But the problem with this vision is that it offers a very narrow range of meaning making options and makes processes of information retrieval and interpretation sometimes cumbersome and time consuming. The syntactic web has improved significantly over the years, Google’s Page Rank algorithm, which exploits the hyperlinking structure of the web as a means of ranking some information as more important than other information, being a notable example. But the way we have come to experience Google (and, more importantly, the way Google *experiences us*) goes far beyond this initial syntactic vision of information.

A more ambitious way of approaching the web, and the one that has dominated most discourse around software development in the last decade, is the vision of the “Semantic Web”, originally articulated by Tim Berners-Lee (Berners-Lee & Hendler 2001). The idea of the Semantic Web sees the internet as a networks of concepts, relations, and rules which can be used by different platforms and agents to “talk to” each other. As Berners-Lee (2000; 191) put it, “The first step is putting data on the Web in a form that machines can naturally understand, or converting it to that form. This creates ... a Semantic Web – a web of data that can be processed directly or indirectly by machines”.

Rather than just a collection of linked documents, the Semantic Web is potentially a collection of linked *ontologies*, that is, sets of different concepts about reality and relations between them. Initial advances toward the vision of a Semantic Web were made possible through the development of ‘metadata’ (Greenberg et al. 2003), machine-readable descriptions of data and documents which has allowed static information resources to be transformed into dynamic resources that interact with and respond to one other. This is what allows, for example, highly customised web content to be delivered to different users (Cardoso 2007). The development of machine learning technologies, data mining, advanced statistics, and predictive analytics has given a boost to this vision of a Semantic Web, making the extraction of meaning from unstructured and semi-structured data, text, images and videos much easier. As Cabeda (2017) puts it, “instead of rewriting the web so that the computers could understand, we are making the computers smarter and are slowly empowering them to aggregat(e) information ...and learn what it means”.

For many, however, the vision of a Semantic Web, even a super-powered version fueled by big data, has proven unsatisfactory, particularly when it comes to getting computers to interact with humans in ways in which meaning might be less ontological and more experiential. For them, the key problem with the notion of the Semantic Web lies in its underlying assumptions about meaning itself, specifically the assumption that meaning is a matter of “context free facts and logical rules” (Schoop et al. 2006: 75, see also McCool 2005). It fails to take into account the way meaning is shaped by the way people and machines use information *together* in “the context of the informational moment” (Casgrande 1999:4). “The human factor in the Semantic Web”, wrote Di Maio in 2008 “is still a largely unresolved issue, prompting central questions that ... have been left behind in technical debates” (Di Maio 2008: 1).

This dissatisfaction with the model of meaning offered by the Semantic Web has led to what Casanovas and his colleagues (Casanovas et al 2017) describe as a “pragmatic turn” in web development, an increased interest in how user behaviour and the contexts in which information is used can be represented in ways that machines can understand and respond to. What has come to be called the “Pragmatic Web” is a vision of the internet as a network of tools, protocols and agents (both human and algorithmic) that interact to make and interpret meaning through actions. Whereas meaning in the Semantic Web is a matter of relations between signs, meaning in the Pragmatic Web is created through relations between *actions* and the *contexts* within which those actions take place.

For many computer scientists and digital strategists (see e.g. de Moor 2005; Leonard-Hansen 2009), the “real” Pragmatic Web is a vision for the future—what Weigand and Paschke (2012: 190) call “Web 4.0”. But in many ways the web is *already* implicated in a wide range of pragmatic processes, a fact seen most clearly in the ways algorithms regularly monitor the actions of users (such as “likes”) and transform them into inferences about the kinds of content they wish to be fed or the kinds of products they wish to buy. At the time of writing, search engines like Google, for example, serve users results not just based on the syntactic technology of the original PageRank algorithm, but on inferences its algorithm forms about what users “want” based on thousands of previous actions of searching and clicking on particular results, as well as where that user is and what they might be doing when they search. Recommender systems like that used by Amazon.com similarly are becoming better and better at inferring what kinds of products people might buy based on the “footprints” they leave as they transverse both digital and physical spaces. Finally, filtering systems like Facebook’s EdgeRank, which determines the order in which updates appear on users’ newsfeeds, are getting better at predicting what kinds of posts will trigger “reactions”, not just because Facebook the company wishes to do a better job at connecting people to the people they care about but because the more users react to posts, the better Facebook’s

algorithms get at making future inferences about them. In all of these systems, meaning is created not just through the syntactic relationship between different bits of data, or the semantic relationship between different concepts represented through metadata, but through the ways algorithms have become adept at inferring meaning from the actions people perform and in enticing them to act in particular ways that make their intentions, desire and goals more legible to these algorithms.

The Pragmatic Web, then, at least the way I am treating it in this chapter, is not so much a “thing” as it is a *perspective* on the way human-computer interaction has developed in the early twenty-first century. As Weigand and Paschke (2012: 183) put it:

The Pragmatic Web is a perspective. It is not a web besides the WWW or Semantic Web. Considering the Internet as a platform of communication and coordination means that messages are not viewed in isolation but in the context of an interaction. Although messages can be interpreted to some extent in a context-free manner (the “literal” meaning), this does never exhaust their meaning. We also need to know (or design) its relationship with other messages, and actions, in the context of use.

The question I would like to consider then, is not the degree to which the Pragmatic Web is achievable as a technological reality, but rather the degree to which the Pragmatic Web is achievable *as an analytical perspective* given our current understanding of pragmatics, or what I am calling “analogue pragmatics”, the set of tools developed to explain how humans create meaning and make inferences. We have already seen how the introduction of new technologies (such as the telephone) alter the way people create implicature and form inferences, requiring linguists to rethink the tools they use to understand these strategies (Hutchby 2001). What is different about the internet is that the medium does not just provide a new architecture within which humans exchange information, but that the medium itself has become a participant in the exchange of information, a participant which brings to interactions inferential processes that are very different than those used by humans. As Kitchin and Dodge (2011) argue, algorithms don’t just mediate but also function to *constitute* social interactions and associations. Hayles (2010: 96) similarly characterises the environment created by digital media as an environment “that senses the environment, creates a context for ... information, communicates internally among components, draws inferences from the data, and comes to conclusions that, in scope if not complexity, far exceed what an unaided human could achieve”.

Of course, when I speak of human-algorithm communication, I am not implying that algorithms are free-agents. They are created by humans and designed to advance the individual or institutional goals of their creators, whether those goals are to stoke the flames of political tribalism or to sell books. But, although the overall goals of algorithmic systems broadly reflect the intentions of their human programmers, the actual micro-mechanics of data-gathering, aggregation, inference and response are largely automated, and sometimes end up advancing broader goals in specific ways that programmers may not have anticipated. For example, YouTube’s recommender system favoured pro-Trump videos nearly 4 to 1 over pro-Clinton videos during the 2016 US elections, not because the creators of the algorithms favoured Trump, but because the algorithms were programmed to maximize engagement, and the algorithms figured out quite quickly that pro-Trump videos resulted in more responses (both negative and positive) from viewers (Lewis 2018). In another example, Microsoft’s machine-learning chatbot, Tay, had to be taken offline within 24 hours of its launch on Twitter after its conversation and language patterns became disturbingly racist, not because of the racist biases of its creators, but because that is how the algorithm learned to interact from human users of Twitter (Vincent 2016). In other words, what is programmed into

algorithms is the directive to pursue a goal as efficiently as possible, and with newer systems, enhanced with machine learning capacities, to learn how to increase efficiency using the data that is made available to them. Rather than seeing algorithms, then, as simple, unitary interactants, it is better to view them as socio-technical assemblages (Latour, 1990) that both embody the plans and goals of human creators and adapt to the behavior of the humans with which they interact.

Problems of algorithmic biases versus the intentions of their creators, as well as problems of privacy and surveillance, as illustrated in the Cambridge Analytical scandal discussed above, are not so much problems about information as about the algorithmic *processing* of information and the way inferences about information are put to use; not so much problems of texts as problems of context; not so much problems of the Semantic Web as problems of the Pragmatic Web.

3. Pragmatics: analogue and algorithmic

Just as a pragmatic approach to digital communication involves looking at the web in a new way, as a web of actions and inferences, it also involves approaching pragmatics in a new way. Analogue pragmatics gives us a toolkit for understanding how people manipulate features of text and context to design interactions in ways that influence the direction and outcomes of communication. Algorithmic pragmatics, rather than focusing on closed systems of usually dyadic communication, sees pragmatic norms as operating within open complex systems in which information circulates and inferences are formed on multiple levels, in multiple contexts, based on multiple logics that both exploit and defy the forms of reasoning characteristic of human-to-human communication. Understanding it requires asking how algorithms form inferences based on people's actions, how they create context and construct identities for people, and how they coerce behavioral change (Gorayska and Mey 2002). Whereas analogue pragmatics is about how people "do things with words", algorithmic pragmatics is about how algorithms "do things" with people.

I do not mean to imply that the insights of analogue pragmatics are obsolete when it comes to understanding the interaction between humans and algorithms. Humans are still humans, and the processes of meaning making and interpretation that they have developed with human interlocutors naturally affect their actions with and inferences about the workings of algorithms (Jones, forthcoming). In fact, many of the strategies internet companies use to entice people into making information about themselves available to algorithms by, for example, completing online quizzes or clicking on "Agree" in the dialogue boxes that pop up on their screens, are based on cleverly exploiting the principles and conventions of analogue pragmatics (such as the "Cooperative Principle" and expectations about the conditional relevance of sequential actions in communication; see Jones 2016b). Furthermore, the interfaces through which algorithms gather and respond to information are usually based on traditional assumptions about how human communication takes place (Galloway 2012), creating situations in which communication with an algorithm is made to resemble communication with a human, or in which certain types of human-to-human communication (such as "liking") is encouraged because it provides data that is more "legible" to algorithms and more easily processed by them.

Similarly, it might be argued that human pragmatic reasoning is also essentially "algorithmic", that the whole idea of pragmatics is that people (and machines) produce and infer meanings by following processual logics, "recipes" for figuring out what is going on. The main difference between analogue and algorithmic pragmatics has to do with the underlying epistemological bases for these processual logics and the relative affordances for

information processing of the human brain and the computer, two differences which I will elaborate on in the remainder of this chapter.

Although pragmatics, as it has evolved over the past half a century, comes in many flavours, from more socially informed approaches (e.g. Culpepper 2011; Jaspers et al. 2010; Mey 2001) to more cognitive approaches (e.g. Beaver & Clark 2008; Bara 2010; Nuyts 1991), what all scholars of pragmatics have in common is a concern with how people communicate and discern intentions below the level of explicit meaning. Much has been made of the “black box” nature of communicating with and through algorithms, what DeNicola (2012: 266) calls the “iceberg quality” of digital communication. For students of pragmatics, however, the “black box” nature of communication is nothing new, nor is it the sole provenance of technologically mediated communication. The central premise of pragmatics is that *all* communication is mediated through the “black box” of the human mind, and the only way communicators can discern the content of the minds of their interlocutors is through processes of inference formed through analyzing the relationship between “output” and the circumstances within which output is produced.

The topics that have traditionally been of interest to students of pragmatics are 1) deixis, the ways language “indexes” people, objects or concepts in the external world; 2) conversational implicature, how it is possible for people to mean more (or less) than what they say; 3) pragmatic inference, how people figure out what people mean; 4) speech acts, how words can be used to perform various actions and the social conventions that make this possible; 5) discourse structure, how the sequence of utterances can affect what they mean; 6) participation, the different kinds of roles speakers and hearers can play in conversation and how those roles affect what they can say, do and understand; and 7) identity, how who we are or who we are taken to be affects how we communicate and how people interpret what we say (Jones 2016c; Levinson 1983; Mey 2001). While all of these aspects of analogue pragmatics are somehow implicated in algorithmic pragmatics, I will limit my discussion here to four main areas where I see the need for students of pragmatics to refine their tools or redefine their concepts to cope with the new analytical challenges posed by digitally-mediated communication. These are: 1) speech acts, 2) inference and implicature, 3) context, and 4) identity.

3.1 *Speech acts*

According to Austin (1976: 76), whenever we communicate we are engaged in a “field of actions”, and we must distinguish what words “mean” from the way they operate within this field. Austin referred to this mode of operation as “force”, the effect words have on the situated experiences of interlocutors; in other words, what words “do”. In formulating an algorithmic pragmatics, then, the first question that must be asked is what is the difference between how people use language to do things and how algorithms do things?

The most obvious difference between people and algorithms is in the arena of what Austin called *performativity*. In natural language, certain utterances have the ability to materially change the contexts in which they are uttered, but only if they meet certain conditions. “I now pronounce you husband and wife”, for instance, only results in two people being married if it is uttered by a particular kind of person at a particular time and in a particular place. The language of algorithms, on the other hand, is *purely performative*; a performativity that presupposes its own felicity. If code can be considered a language, Galloway says, it is a special kind of language: it is “the only language that is *executable* . . . the first language that actually *does what it says*” (Galloway 2004: 165-6; emphasis mine).

How does this feature of code affect us when we interact with algorithms? The most important way is that it imposes a kind of “tyranny of performativity”. We cannot help but

“act” whenever we interact through digital media, because algorithms automatically turn our utterances into actions — when we type a word on our keyboards, we set into motion a chain of operations within our own devices and other devices connected to ours over a network; when we react to or comment on a friend’s post, we set into motion a series of algorithmically orchestrated operations that can affect the order of information on other people’s newsfeeds as well as our own and may, when aggregated with other actions, eventually be interpreted as a “request” to receive certain political advertisements, news stories, or “memes”, even if that was not our intention.

At the same time, just as algorithms automatically turn words into actions, they also automatically turn actions into “words” (or, rather, code), not only making those actions “legible” in particular ways to computers, but also preserving them for future reference. As Andrejevic (2009: 53) puts it, the internet is a “realm wherein every action, interaction, and transaction generates information about itself”. Clicks, steps, purchases, even the speed and rhythm with which one navigates websites or types on their keyboard (see e.g. Epp et al. 2011) become information that is relentlessly recorded. Algorithmically extexualised actions are then turned back on users in feedback loops that are often designed to reinforce the same actions or re-engineer these actions in ways more amiable to the algorithm’s agenda. Algorithmically coerced actions are coerced based on what we have already done, and what the algorithm wants us to do next can change moment by moment based on our responses. Perhaps this is what Google’s Michael Schmidt meant when he said, half jokingly, that people don’t want Google to give them information; they want Google “to tell them what they should be doing next” (Jenkins 2010).

The first step to formulating an algorithmic pragmatics, then, is to regard digital systems not as “information systems”, but as “action systems”, systems designed to act on our information and compel us to act in particular ways. Goldkuhl and Ågerfalk (2002: 85) argue that the key to understanding and designing computer systems is not “usability”, but “*actability*”, which they define as “an information system’s ability to perform actions, and to permit, promote and facilitate the performance of actions by users, both through the system and based on information from the system”. Of course, this view of information systems as action systems is, in some ways, not so different from Austin’s insistence that human communication takes place within a field of actions. The difference is the speed and efficiency with which computer systems transform information into actions and actions into information.

3.2 Inferences and implicature

The central focus of pragmatics is the “guesswork” involved in social interaction: how people interpret the intentions of others, and how people communicate meanings and perform actions without making their intentions explicit. Understanding language is more than just decoding words; it involves exquisitely sensitive processes of inferential reasoning. The only way the “guesswork” of communication works is because of a tacit agreement between speakers and hearers about the logical principles that should govern this “guesswork”. If nothing else, Grice (1989) argued, talking is a form of “rational behaviour”. The rational basis of human communication, however, is not just a matter of agreeing upon a set of principles or “maxims”; it is a matter of committing to a particular epistemology. The epistemological stance outlined by Grice and his followers (e.g. Horn 1984; Levinson 2000) is basically that knowledge is the result of deductive “if...then” reasoning: It is a “causal” epistemology in which hearers form hypothesis about meaning based on testing input against a mutually agreed upon “theory” of communication.

Algorithms, on the other hand, while also engaging in guesswork, do so based on a very different epistemology: knowledge is a matter not of causation but of *correlation*. No underlying set of rules or “theory” of communication is necessary. Meaning is constructed mathematically, probabilistically, based on correlations between pieces of input. That is why Google can translate languages without actually “knowing” the rules of grammar or the meanings of words, simply by calculating the probability of a particular word in a particular context having a particular meaning based on billions of other occurrences of the use of that word in context. The probabilistic induction of algorithmic inferencing can tell retailers what book you might buy by correlating your past purchase with that of other readers, and political operatives can tell what candidate you might vote for based on your supermarket purchases or whether or not you shared posts related to Hello Kitty on Facebook (Cadwalladr and Graham-Harrison 2018), as well as how these behaviors might be altered with just the right algorithmically-informed nudge at the right time.

There are risks related to all inferencing, whether algorithmic or analogue. One or both of the participants, for example, may be in possession of too little, too much, or the wrong kind of information with which to make inferences, or one of the participants may have access to resources for interpreting information not available to the other, making control of the personal information that is “given” and “given off” more difficult (Goffman 1959). When it comes to human-algorithm communication, these potential asymmetries are an inevitable feature of the interaction. The ability of algorithms to infer information from seemingly unrelated information makes it harder for people to control what can be known about them, and far exceeds humans’ ability to infer information about algorithms. Furthermore, algorithmic inferencing is much harder to challenge. A good example is the way police officers “profile” suspects. Sacks (1972) describes the analogue process by which police decide whether or not to stop and search a potential suspect, noting that the “propriety” of the inference is based on an agreed upon set of expectations about what sort of behaviour would rouse an “ordinary person” to suspicion, and it is only if the officer is able to convince a judge or jury that the inference has been properly arrived at that the person is convict-able. Nowadays, however, many police forces make use of what is known as “predictive policing”, which uses algorithms to, for example, deploy police officers to neighbourhoods where there is a higher probability of a crime being carried out. What arouses suspicion, then, is simply being present in the neighbourhood. This, of course, results in more crimes being discovered in these neighbourhoods, which can snowball into what Ensign and her colleagues (Ensign et al. 2017) call a “runaway feedback loop” that makes the algorithm’s prediction a self-fulfilling prophesy. In such cases, contrary to the well-known warning of scientists, correlation *is* causation, since it causes what it predicts (see also Brayne 2017; Krasmann 2007).

3.3 Context

Although people and algorithms use very different kinds of epistemologies when making inferences, they both depend on the same sorts of information, including the *content* of the language that has been produced, the *contexts* in which this language may later appear, and the *identities* of the speakers and hearers. But what “context” and “identity” mean to algorithms is dramatically different from what they mean for humans. In one sense, algorithms’ understanding of “context” and “identity” is extremely reductive, limited to information that can be digitised and made *legible* to computers. On the other hand, their ability to laminate different contexts onto one another, to rapidly create new contexts within which to interpret information, and to construct very precise “identity types” based on very

specific goals, allows them to use these kinds of information in ways that are not available to humans.

Understanding context and how we use it to communicate in the analogue world is already a rich and contested topic (see for example Duranti and Goodwin 1992; van Dijk 2008). Linguists and discourse analysts have debated whether context in communication should be approached as a micro-social phenomenon or a macro-social one, and how much context is fixed or emergent. One thing that most scholars concur on, however, is that human communication depends a great deal on the degree to which people are able to negotiate and, to some extent, agree upon the “definition” of the particular communicative situations in which they find themselves. Goffman (1972) argued that all situations have their own properties that follow from the fact of co-presence between two or more people, and that this co-presence and the “mutual monitoring possibilities” it affords makes it possible for people to negotiate “what’s going on here”. While pragmatic contexts naturally consist of individual contexts, communication requires that some common context be defined, often based on community norms and/or joint goals. In other words, there must be some alignment regarding what aspect of the situation “matters” when people create implicature and form inferences. More cognitively oriented linguists see this negotiation about what matters in terms of “relevance” (Sperber and Wilson 1986), arguing that communication is not so much a matter of shared social norms, but a matter of shared biologically determined cognitive imperatives to filter out potentially irrelevant information and to focus on information more likely to provide informational rewards.

The rise of technologies like social media and mobile communication has problematized both social and cognitive understandings of context, making it more difficult for people both to negotiate shared understandings of context with others and to determine what information is more likely to be relevant in particular situations. This is because, as Weigand and Paschke (2012) point out, the Pragmatic Web is not a web of texts, but a web of contexts, which are exceedingly difficult to keep separate from one another. For humans, this has resulted in new challenges around juggling multiple contexts online and negotiating what they wish their interlocutors to take as relevant (boyd 2010; Tagg et al 2017).

What is sometimes problematic for people is, for algorithms, rich with opportunities. For algorithms, the network of linked contexts that make up the internet is really a rich array of circumstances for forming inferences about our actions and utterances, and for seeing meaning in multiple text-context combinations *simultaneously*. Because algorithms have at their disposal informational horizons that exceed human perception, they are able to make relevant aspects of context about which people may not even be aware. While for humans, for example, the context of a web search might be a particular goal or activity, such as a school assignment, for Google, the context stretches back to all of the other searches this individual has performed, and may even include their email conversations about related topics.

Much has been made of algorithms’ ability to design the contexts in which people consume and share information, such as one’s Facebook Newsfeed, an ability that was particularly evident in the 2016 US election and the UK Brexit vote, and much has also been made of the ability of algorithms to *recontextualise* information in ways that create new contexts that affect the ways we interpret texts and what we do with them, as when information from that web search suddenly becomes part of the context of another site in the form of a “clickbait” headline related to what was being searched for. Maybe what is most challenging about communicating with algorithms, though, is not that their tools for determining what is *relevant* in a given context or set of contexts is so different from ours, but that the “reasoning” behind these determinations are opaque to us.

Humans still have some advantages over computers when it comes to contextualization in their ability to mutually monitor signals from interlocutors that are

illegible to computers, and to adapt to subtle changes in the physical and social environment, framing and reframing their understanding and “design” of contexts as they go along (Tagg et al. 2017). As digital media increasingly weave themselves into the fabrics of human physical lives, however, embedded in watches, refrigerators, thermostats, and even clothing, not to mention the mobile phones that are carried everywhere, algorithms will become more adept at sensing patterns in human behaviour: where we go, who we interact with, even what we say, and be able use this information to design more efficient contexts for interpreting our actions.

3.4 Identities

In interaction between humans, the social identities of interlocutors is an important ingredient for sense making. We use the social identities of interlocutors to interpret their words, and we use words to construct or alter aspects of our identities. In other words, like context, identity is both “fixed”, based on a certain set of “membership categories” available in our social worlds and relevant to whatever we happen to be doing, and “emergent”, negotiated and co-constructed with our interlocutors in accordance with certain strategic goals (Bucholtz and Hall 2005, Jones 2016c, Sacks 1992). The “fixed” and “emergent” aspects of identity are not separate, but rather intertwined: we use membership categories strategically, appropriating and discarding them, making some more relevant in certain instances than others, and mixing them together to form hybrid identities.

Similarly, when it comes to algorithms, much of the way they operate on us and affect our textual and interactional options is based on “who they think we are”. Much has been written about the “data doubles” (Haggerty and Ericson 2000) that algorithms cobble out of our incremental actions and utterances, and how these digital shadows follow us, both online and off, determining what sorts of discounts we are offered at shops and whether or not we are able to get loans from banks. In his book *We are Data*, Cheney-Lippold (2017) argues that algorithms make sense of us by turning us into “measurable types”. The worst thing about these “types”, he believes, is that they somehow obscure our “true selves”. “Who we are in the face of algorithmic interpretation”, he writes, “is who we are computationally calculated to be... and when our embodied individualities get ignored, we increasingly lose control not just over life but over how life itself is defined” (Cheney-Lippold 2017: 5). He suggests that “[w]hen data defines us, the complexities of our emotional and psychological lives online are flattened out for purposes of mass-scale, approximate data analysis” (Cheney-Lippold 2017: 53).

While this is no doubt true to some degree, the ways algorithms construct our identities and reflect them back to us is more complicated, and potentially more problematic. First of all, the processing of people into “recognizable types” that algorithms engage in is not entirely different from the way humans use “the recognizability of people as certain sorts of people or, more specifically, people as certain sorts of members of society” (Day n.d.: 1) as a resource for dealing with others. It is just that the categories available to algorithms are potentially more complex, and the ways they group different categories together (based on correlation rather than causation, see above), provides them with different kinds of resources for judging “who we are” and acting accordingly. What characterises algorithmically constructed identities is, first, their precision — we are not just men or gay or caucasian, but members of complex “micro-categories” consisting of bundles of inferences that algorithms have formed about us, and second, their flexibility; we are not the same from one moment to the next or from one site to the next. Algorithms adjust their assessments of us with every click or swipe, aggregating and processing these quotidian actions in ways that can reveal not just our personality traits, but our momentary moods. The problem might not be, as

Cheney-Lippold suggests, that algorithms ignore our emotional complexities, but just the opposite, that they get revealed and explained in ways that we ourselves are not even aware of. In other words, the problem with data doubles is not that they don't reflect our "true selves", but that they *do*, robbing us of the opportunity to engage in the kinds of self-presentation and impression management that is central to social life.

The most important thing to remember about algorithmic identity projects, though, is that, like all identity projects, they are "goal directed". The identities that algorithms imagine for us are configured towards certain "category specific activities" (Sacks, 1992), in this case, usually shopping. In studies of Membership Categorization Analysis, Jayusi (1984) has highlighted the *moral* and *normative* character of categories and categorization. In the famous example by Sacks (1992), for instance, the fact that the categories of mother and baby are inferred by the action of the mother picking the baby up is predicated on the idea that mothers *should* pick up their babies when they cry. The normalising power of algorithms is even more efficient. Algorithms don't just tell you "who they think you are" but also reward you for acting accordingly by offering you discounts, special offers, and other enticements, so that gradually you become the kind of person (or customer) you are expected to be.

4. Conclusion: pragmatic competence in the age of the algorithm

In pointing out some of the differences between the way humans do pragmatics, and the way algorithms do, I have not intended to demonise algorithms nor to suggest we have become slaves to them. We have not. In fact, as important as the ways algorithms have come to control us are the processes humans have been developing to figure out how to control algorithms and use them for their own purposes. Galloway (2006: 106) argues that video games are "allegories for our contemporary life under the protocological network of continuous informatic control". As much as the code of the game produces the "perimeters for discursive possibility (and) possible actions...and subjective experiences" made available to players (Cheney-Lippold 2017: 218), good players are still able to master the game. "To play the game", Galloway (2006: 91) writes, "means to play the code of the game. To win means to know the system. And thus to interpret a game means to interpret its algorithm".

Some have suggested that the way for us to get good at playing this game is to open up the "black box" (Pasquale 2015), that is, to force tech companies to tell us how their algorithms work, to lay bare their inferential processes. It is an approach that rests on assumptions that being able to *see* a system is the same as being able to know how it works and control it, a deep seated enlightenment notion that drives many of the attempts by governments to legislate the digital world. Insisting that companies like Facebook and Google tell us how their algorithms work, however, will probably have little effect, not just because big tech companies are not likely to part so easily with their intellectual property, but because even if algorithms *were* made transparent, their real workings would remain opaque to all but the most technologically savvy of us (Ananny and Crawford 2018). But most important, making algorithms transparent would compromise much of the efficiency and convenience we have come to expect from digital communication, just as asking people to make everything they mean explicit would render human communication not just incredibly inefficient but also frequently uncivil.

Coping with the communicative challenges posed by the pragmatic web will require more than just trying to make algorithms more "accountable", because they are just one component of the pragmatic web. As I said above, algorithms do not operate independent of their human interlocutors, nor do they exist independent of the institutional agendas, social practices, and relations of power through which they are created (Ananny and Crawford 2018; Crawford 2016). The way to approach the pitfalls of the pragmatic web, therefore, is

not just with political or technical solutions, but with a fuller engagement with *pragmatics*. Training in digital literacies must involve helping people to develop the new kinds of *pragmatic competences* that digital environments demand, helping them to become more sensitive to how algorithms process not just their “information” but also their actions, and helping them to reflect upon not just the inferential processes of algorithms, but also on the inferential processes they themselves develop as they interact with both human and non-human actors within the Pragmatic Web.

After the dust from the Cambridge Analytica scandal has settled, the kinds of algorithms implicated in it will continue their relentless processing of our information. What is perhaps good about the scandal is that it brought the issue of algorithms into the public conversation – it (aided by algorithms) elevated algorithmic pragmatics to the top of our Newsfeeds. And the more algorithms are included in our thinking about how our lives are organized, the better we will become at mastering the game of the Pragmatic Web.

References

- Ananny, Mike & Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media and Society*, 20 (3). 973–989.
- Andrejevic, Mark. 2009. Privacy, exploitation, and the digital enclosure. *Amsterdam Law Forum* 1 (4). 47–62.
- Antaki, Charles & Sue Widdicombe (eds.). 1998. *Identities in talk*. London: Sage.
- Austin, John L. 1976. *How to do things with words*. Oxford: Oxford University Press.
- Bara, Bruno G. 2010. *Cognitive pragmatics: The mental processes of communication*. Cambridge, MA: MIT Press.
- Beaver, David I. and Brady Z. Clark. 2008. *Sense and sensitivity: How focus determines meaning*. Malden, MA: Wiley-Blackwell.
- Berners-Lee, Tim 2000. *Weaving the web: The past, present and future of the World Wide Web by its inventor*. London, Texere.
- Berners-Lee, Tim, James Hendler & Ora Lassila. 2001. The Semantic Web. *Scientific American*. May. Retrieved 10 April 2018, from https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf
- boyd, danah. 2010. Social network sites as networked publics: Affordances, dynamics and implications. In Z. Papacharissi (ed.) *A networked self: Identity, community and culture on social network sites*, 39-58. New York: Routledge.
- Brayne, Sarah. 2017. Big data surveillance: The case of policing. *American Sociological Review*, 82 (5). 977–1008.
- Bucholtz, Mary & Kira Hall. 2005. Identity and interaction: a sociocultural linguistic approach. *Discourse Studies* 7(4–5). 585–614.

Cabeda, José. 2017, June 4. Semantic Web is Dead, Long live the AI!!! Retrieved 10 April 2018, from <https://hackernoon.com/semantic-web-is-dead-long-live-the-ai-2a5ea0cf6423>

Cardoso, Jorge. 2007. The syntactic and Semantic Web. In Jorge Cardoso (ed.) *Semantic Web services: Theory, tools and application*, 1-24. Hershey, PA: IGA Indertnaitonal.

Cadwalladr, Carole & Emma Graham-Harrison. 2018, March 17. How Cambridge Analytica turned Facebook 'likes' into a lucrative political tool. Retrieved 8 April 2018, from <http://www.theguardian.com/technology/2018/mar/17/facebook-cambridge-analytica-kogan-data-algorithm>

Casanovas P., Rodríguez-Doncel V., González-Conejero J. 2017. The role of pragmatics in the web of data. In: Francesca Poggi & Alessandro Capone (eds) *Pragmatics and law. Perspectives in pragmatics, philosophy & psychology*, vol 10. Cham, Switzerland: Springer.

Cheney-Lippold, John. 2017. *We are data: Algorithms and the making of our digital selves*. New York: NYU Press.

Crawford, Kate. 2016. Can an algorithm be agonistic? Ten scenes from life in calculated publics. *Science, Technology and Human Values* 41(1): 77–92.

Culpeper, Jonathan (ed.). 2011. *Historical sociopragmatics*. Amsterdam: John Benjamins Publishing.

Day, Dennis. n.d. Membership categorization analysis. ms. University of Southern Denmark. Retrieved 8 April, 2018 from bit.ly/2ujzWFn

de Moor, Aldo. 2005. Patterns for the Pragmatic Web. In Frithjof Dau, Marie -Laure Mugnier, & Gerd Stumme, *Conceptual structures: Common semantics for sharing knowledge Vol. 3596*, 1–18. Berlin, Heidelberg: Springer Berlin Heidelberg.

DeNicola, Lane. 2012. EULA, Codec, API: On the opacity of digital culture. In Pelle Snickars and Patrick Vonderau (eds.), *Moving data: The iPhone and the future of media*, 265-277. New York: Columbia University Press.

Di Maio, Paola. 2008. The missing pragmatic link in the Semantic Web. *Business Intelligence Advisory Service* 8 (7). 1-3.

Duranti, Alessandro, & Goodwin, Charles G. (eds.). 1992. *Rethinking context: Language as an interactive phenomenon*. Cambridge England & New York: Cambridge University Press.

Ensign, Danielle, Friedler, Sorelle A., Neville, Scott, Scheidegger, Carolos, & Venkatasubramanian, Suresh. 2017. Runaway feedback loops in predictive policing. *ArXiv:1706.09847 [Cs, Stat]*. Retrieved from <http://arxiv.org/abs/1706.09847>

Epp, Clayton, Michael Lippold, & Regan L. Mandryk. 2011. Identifying emotional states using keystroke dynamics'. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 715–724. New York: ACM.

- Galloway, Alexander R. 2004. *Protocol: How control exists after decentralization*. Cambridge: MA: MIT Press.
- Galloway, Alexander. 2006. *Gaming: Essays on algorithmic culture*. Minneapolis: University of Minnesota Press.
- Galloway, Alexander 2012. *The interface effect*. Cambridge, UK & Malden, MA: Polity Press.
- Goffman, Erving. 1959. *The presentation of self in everyday life*. New York: Doubleday.
- Goffman, Erving. 1972. *Relations in public: microstudies of the public order*. Harper & Row.
- Goldkuhl, Gören & Pär J. Ågerfalk. 2002. Actability: A way to understand information systems pragmatics. In Kecheng Liu, Rodney J. Clarke, Peter Bøgh Andersen, Ronald K. Stamper & Al-Sayed Abou-Zeid (eds.) *Coordination and communication using signs: Studies in Organisational Semiotics*, vol. 2. (pp. 85–113). Boston: Kluwer Academic Publishers.
- Gorayska, Barbara & Jacob L. Mey. 2002. Introduction: Pragmatics of Technology. *International Journal of Cognition and Technology* 1(1). 1–20.
- Grasseger, Hannes & Mikael Krogerus. Jan. 28, 2017. The data that turned the world upside down. *Motherboard* (Accessed on March 30, 2018 from https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win) (Originally published in
- Greenberg, Jane, Stuart Sutton, & D. Grant Campbell. 2003. Metadata: A fundamental component of the semantic web. *Bulletin of the American Society for Information Science and Technology* April-May. 16-18.
- Grice, Herbert Paul. 1989. *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Haggerty, Kevin D. and Ericson, Richard V. 2000. The surveillant assemblage. *British Journal of Sociology* 51(4). 605–622.
- Hayles, N. Katherine. 2010. RFID: Human agency and meaning in information intensive environments. In Jörgen Schäfer & Peter Gendolla, P. (eds.). *Beyond the screen: Transformations of literary structures, interfaces and genres*, 95-122. Bielefeld: Transcript-Verlag.
- Horn, Laurence R. 1984. Toward a new taxonomy for pragmatic inference: Q-based and Rbased implicature. In Deborah Schiffrin (ed.), *Georgetown University Round Table on Languages and Linguistics* (pp. 11-42), Washington, D.C.: Georgetown University Press.
- Hutchby, Ian. 2001. *Conversation and technology: From the telephone to the Internet*. New York: Wiley.
- Jayusi, Lena. 1984. *Categorization and the moral order*. London: Routledge and Kegan Paul.

Jaspers, Jürgen, Jan-Ola Östman, and Jef Verschueren (eds.). 2010. *Society and language use*. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Jenkins, Holman W. Jr. 2010, August 14. Google and the search for the future. *Wall Street Journal*. Retrieved 8 April 2018, from <https://www.wsj.com/articles/SB10001424052748704901104575423294099527212>

Jones, Rodney H. 2016a. Surveillance. In Alexandra Georgakopoulou and Tereza Spilloti (eds.), *The Routledge handbook of language and digital communication*, 408-411. London: Routledge.

Jones, Rodney H. 2016b. Linguistics and the study of online surveillance. A paper presented at the 17th Annual Conference of the Association of Internet Researchers (AoIR), Berlin, October 5-8.

Jones, Rodney H. 2016c. *Spoken discourse*. London: Bloomsbury.

Jones, Rodney H. 2016d. 'Have you swiped your nectar card?' Pretextuality and practices of surveillance. A lecture at the Lancaster Literacy Research Centre, University of Lancaster, January 25. <https://www.lancaster.ac.uk/linguistics/news-and-events/events-archive-to-aug-2016/5494>

Jones, Rodney H. forthcoming. 'Folk algorithmics': Reading and writing in the age of the algorithm. *Linguistics and Education*.

Kitchin, Rob and Dodge, Martin. 2011. *Code/Space: Software and Everyday Life*. Cambridge, Mass.: The MIT Press.

Kosinski, Michal, Stillwell, David and Graepel, Thore. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15). 5802–5805. <https://doi.org/10.1073/pnas.1218772110>

Krasmann, Suzanne, 2007. [The enemy on the border: Critique of a programme in favour of a preventive state](#), *Punishment & Society*, 9(3). 301–318.

Latour, Bruno. 1990. Technology is Society Made Durable. *The Sociological Review*, 38(1 suppl). 103–131.

Leonard-Hansen, Alisa. 2009, November 20. The future is all about context: The Pragmatic Web. Retrieved 11 February 2019, from https://readwrite.com/2009/11/20/future_all_about_context_the_pragmatic_web/

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, Stephen C. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Lewis, Paul. 2018, February 2. Fiction is outperforming reality: how YouTube's algorithm distorts truth. *The Guardian*. Retrieved from <https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth>

- Maíz-Arévalo, Carmen. 2017. Getting 'liked'. In Christian Hoffmann and Wolfram Bublitz (eds.). *Pragmatics of social media*, 575-606. Berlin & Boston: Mouton De Gruyter.
- McCool, Rob. 2005, December. Rethinking the Semantic Web, Part 1. *IEEE Internet Computing*, 86-88.
- Mey, Jacob L. 2001. *Pragmatics: An introduction* (2nd edition). Malden, MA: Wiley-Blackwell.
- Nuyts, Jan. 1991. *Aspects of a cognitive-pragmatic theory of language: On cognition, functionalism, and grammar*. Amsterdam: John Benjamins Publishing Company.
- Pasquale, Frank. 2015. *The black box society: The secret algorithms that control money and information*. Cambridge: Harvard University Press.
- Sacks, Harvey. 1972. Notes on police assessment of moral character. In D. Sudnow (ed.) *Studies in social interaction*, 280-293. New York: Free Press.
- Sacks, Harvey. 1992. *Lectures on conversation* (vol. 2), Gail Jefferson (ed.). Oxford: Blackwell.
- Schoop, Mareike, Aldo de Moor, & Jan L.G. Dietz, 2006. The Pragmatic Web: A manifesto. *Communication of the ACM* 49(5). 75-6.
- Sperber, Dan and Deirdre Wilson. 1986. *Relevance: Communication and cognition*. London: Blackwell.
- Tagg, Caroline, Philip Seargeant & Amy Aisha Brown. 2017. *Taking offence on social media: Conviviality and communication on Facebook*. London: Palgrave Macmillan.
- van Dijk, Tuen A. 2008. *Discourse and context: A sociocognitive approach*. Cambridge: Cambridge University Press.
- Vincent, James. 2016, March 24. Twitter taught Microsoft's friendly AI chatbot to be a racist asshole in less than a day. *The Verge*. Retrieved February 11, 2019, from <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>
- Weigand, Hans & Adrian Paschke. 2012. The Pragmatic Web: Putting rules in context. In *Rules on the Web: Research and Applications*, 182-192. Berlin & Heidelberg: Springer.
- Yus, Francisco. 2011. *Cyberpragmatics: Internet-mediated communication in context*. Amsterdam & Philadelphia: John Benjamins.