

Algorithmic memory and the right to be forgotten on the web

Big Data & Society
January–June 2017: 1–11
© The Author(s) 2017
DOI: 10.1177/2053951717703996
journals.sagepub.com/home/bds



Elena Esposito

Abstract

The debate on the right to be forgotten on Google involves the relationship between human information processing and digital processing by algorithms. The specificity of digital memory is not so much its often discussed inability to forget. What distinguishes digital memory is, instead, its ability to process information without understanding. Algorithms only work with data (i.e. with differences) without remembering or forgetting. Merely calculating, algorithms manage to produce significant results not because they operate in an intelligent way, but because they “parasitically” exploit the intelligence, the memory, and the attribution of meaning by human actors. The specificity of algorithmic processing makes it possible to bypass the paradox of remembering to forget, which up to now blocked any human-based forgetting technique. If you decide to forget some memory, the most immediate effect is drawing attention to it, thereby activating remembering. Working differently from human intelligence, however, algorithms can implement, for the first time, the classical insight that it might be possible to reinforce forgetting not by erasing memories but by multiplying them. After discussing several projects on the web which implicitly adopt this approach, the article concludes by raising some deeper problems posed when algorithms use data and metadata to produce information that cannot be attributed to any human being.

Keywords

Right to be forgotten, algorithms, memory, artificial intelligence, privacy, Big Data

Introduction

The debate raised by the so-called right to be forgotten on Google shows the difficulties and opportunities associated with the social consequences of algorithms in the particularly complex issue of the management of memory.

The spread of Big Data and the increasingly active role of digital procedures in many areas of social life confront us with situations in which algorithms collect data, identify correlations, analyze patterns, and autonomously produce additional information. When a user does a search on Google, a situation arises in which neither those who built the machine nor those who programmed it nor those who entered the data knew of the output or could have predicted its particularities. The information the user gets cannot properly be attributed to any of these human agents. If it can be attributed to anyone or anything, this could only be the algorithm that accomplishes the search.

This condition produces unprecedented attribution problems of epistemological, moral, and legal character

(Floridi and Sanders, 2004; Koops et al., 2010; Simon, 2012). Who should be held accountable for the operation of algorithms and its consequences? The company that produces the algorithms and the programmers who design them do not know the data on which they work and do not control the actual output of the process. This does not mean of course that the algorithms work with “raw data” resulting directly from the world (boyd and Crawford, 2012; Gillespie, 2014; Gitelman, 2013; Mittelstadt et al., 2016) nor that they work in a neutral way.¹ The algorithms themselves, however, work without an intention and without knowing or understanding the materials they process.

University Bielefeld, Germany

Corresponding author:

Elena Esposito, Faculty of Sociology, Universitätsstraße 25, 33615 Bielefeld, Germany.
Email: elena.esposito@uni-bielefeld.de



The difficulty with these questions is not primarily the possible attribution of legal status to nonhuman entities, since this already happens with companies and organizations.² The difficulty is, instead, related to the fact that we are dealing with decisions that are taken independently by algorithms. And even if algorithms are and remain fully determined (Etzioni, 2016), the latest developments in programming techniques show that they are particularly useful when their procedures do not reproduce the processes of human rationality (Burrell, 2016: 7). Both in theory and in practice, therefore, we have to consider the difference between algorithmic and human data processing, with its potential implications for the accountability of the outcomes.

My analysis of these issues will take its point of departure from the judgment of the European Court of Justice on 13 March 2014 (causa C-131/12)³ on the right of citizens to request the removal from web search results of the links associated with their name, understood as the “right to be forgotten.” The ruling, which directly addresses the role of algorithms in the processing of social information, raised a lively debate around the consequences of digitalization for memory. In the discourse about web memory, algorithms are often associated with an unprecedented extension of remembering, expressed by the widespread idea that “the Internet never forgets,”⁴ because data stored in digital archives can automatically be made available through search engines. But the situation is, of course, more complex, since storage and accessibility are, in fact, two separate issues requiring different tools and different decisions. Digital memory remembers a lot but also forgets a lot, in new and articulated ways. Information can be lost because it is not stored, because its support is damaged or because it cannot be accessed with the available tools.

In web memory, remembering and forgetting are not two opposing components that negate each other. The availability of memories can increase together with the loss of memory (forgetting). To deal with this condition, I argue, we need a concept of memory more complex than the common idea of an accumulation of memories. In particular, we need to reevaluate the active role of forgetting as a necessary component. I take the problems raised by the legislation to implement the right to be forgotten on the web as an opportunity to test a different way of observing social memory and its technological support.

The paper opens by presenting the ruling of the European Court of Justice, its motivations, and the theoretical and practical problems of its implementation. Many of these problems stem from the difficulty of dealing with the active role of algorithms in the management of information. These difficulties, I argue in

the subsequent section, are connected with the fact that algorithms today do not try to reproduce the human form of information processing. Algorithms, as we shall see, do not refer to meanings but directly to data, i.e. to the underlying differences, which by themselves are not meaningful. The fact that algorithms process data and not meanings—now in enormous quantities and with extraordinary speed—gives rise to the impression that algorithms can remember everything. But if algorithmic processing is a form of memory, it faces a task quite different from that confronting human memory: whereas for the latter the challenge is to remember enough, the challenge for algorithmic memory is to be able to forget enough and in a controlled way. Toward the goal of developing a concept of social memory appropriate to our web society, in the next section my analysis aims to show that algorithms, precisely because they operate autonomously and differently from human intelligence, can also provide the tools to manage this problem. Algorithms can implement, for the first time, the classical insight that it might be possible to reinforce forgetting not by erasing memories but by multiplying them. In the final part of the paper, I briefly discuss several projects already available on the web that deal with the management of forgetting from this perspective. I conclude by addressing the control problems that remain open when dealing with systems that work with data produced every moment, often unconsciously, by users connected with the web and with one another.

The right to be forgotten on the web

The ruling of the European Court of Justice reacts to a complaint lodged by a Spanish citizen against Google. The company was accused of infringing his privacy rights because its search engine made his personal data accessible to everyone on the web, even if the event they referred to had been resolved for a number of years and the reference had become irrelevant. The Court was asked to judge whether individual citizens should have the right to make their personal information untraceable (the right to be forgotten: §20) after a certain time simply because they wish it (“without it being necessary... that the inclusion of the information in question... causes prejudice to the data subject”: ruling C-131/12, §100). The Court also had to decide whether the company Google should be held responsible for the processing of personal data and forced to suppress the links to web pages containing information on the person in question, even if the information remains available on the web pages containing it and its publication is lawful.

The problem to which the European Court reacted with its judgment is related to the unprecedented role of

algorithms in the production of social memory. On the web, data processing uses algorithms, which act on enormous amount of data, apparently with no limit in their processing and storage capability. Making information accessible to everyone with an internet connection, the web intensifies the problem of the “*droit à l’oubli*,” which has a long legal tradition starting from French law. This right protects the desire of a citizen who has been convicted of a criminal act and has paid his debt to justice to be no longer remembered for those past facts and to be able to build a new life and a new public image. The right to be forgotten is directly connected with the ability to keep the future open—a “right to reinvention” (Solove, 2011) that protects the future of the person from a colonization by the past. Nietzsche knew it very well, when he spoke of the “need of oblivion for life,”⁵ even more important than the ability to remember—because without forgetting one would remain bound to an eternal presence of the past, which does not allow to build a different future. Without forgetting you cannot plan nor can you hope.

This is certainly plausible. The judgment of the European Court recognizes this right for European citizens, and forces Google to remove the links to the personal data of the ones who request it—unless that information has public relevance. In a condition in which search engines give access to the countless amount of data available on the web, however, the right to be forgotten protected by the European Court becomes much more extensive than the classic “*droit à l’oubli*,” both materially and socially: it concerns any act (especially those inconsequential on the penal level but relevant for image and reputation: the private episodes that Google should make untraceable) and includes any person (not only criminals but each of us, particularly teenagers).

This extension of the right inevitably produces social coordination problems. The forgetting of anyone also affects the forgetting of others—those who are involved in the same event and maybe do not want to be forgotten, but also those who may be involved in the future or interested in similar events, and would like to preserve the ability to access the relevant information. The protection of individual forgetting collides with the right to information and with the creation of a reliable shared public sphere (Nabi, 2014; Toobin, 2014).

The ruling of the European Court states that the right to privacy overrules public interest in finding personal information, unless the person holds a public role (§ 97). The issue is extremely controversial and fits into the open debate about the definition and limits of privacy in the web society (Nissenbaum, 2004; Solove, 2007b, 2011).

The solution proposed by the European Court, however, also raises practical implementation problems,

due to the active role of algorithms. The judgment considers Google accountable and responsible for the excess of memory in our digital world,⁶ on the basis of the principle according to which is responsible “the natural or legal person (...) which determines the purposes and means of the processing of personal data (...) whether or not by automatic means” (§4). Google itself, on the contrary, claims that it cannot be held responsible, because the processing of data is performed by the search engine and the company “has no knowledge of those data and does not exercise control over the data” (§22). Can the autonomy of the operation of algorithms relieve the company from the responsibility for data management?

The European Court denies it, although it distinguishes the processing of data by Google from the processing by publishers and journalists. Even if Google as a company does not decide on data processing, search engine activity makes data, in principle, accessible to internet users, including those who otherwise would not have found the original page (§36). It also allows users to get a “structured overview” of the information relating to a person, “enabling them to establish a more or less detailed profile” (§37). This affects the privacy of the persons concerned in different and more incisive ways than publishing the information. The processing of data by Google is more subtle but more dangerous than the one by publishers and journalists, therefore the company is charged with suppressing the links to people who require it, even if the publication is lawful and the information remains available.⁷

This decision implies, without making it explicit, a specific definition of social memory and forgetting. Is memory the ability to store information in an archive, even if it is inaccessible? Or does it depend on the ability to find the information when you need it? Is memory storage or is it remembering?⁸ Ascribing to Google the management of the right to oblivion implies a clear choice: data are considered forgotten if they are made difficult to trace (acting on search engines), while social memory should be preserved by the storage of data in the pages of newspapers and in other archives.

David Drummond, general counsel of Google, commenting on the judgment of the European Court complained that it puts Google in a sort of no man’s land,⁹ without any of the protections that legislation provides to media, archives, and other communication tools.¹⁰ The ruling does not consider the specificity of the company and does not comment on its claims regarding the unprecedented autonomy of the operation of algorithms. Google acts on the data without knowing and without controlling them, therefore it is neither a library, a catalog, a newspaper, a newsstand, nor a service provider. Google is a search engine.

Search engines are not active as newspapers, publishers, and libraries, which select and organize the information to be disclosed, but are not even passive as pure intermediaries, which merely provide access to materials they did not choose and do not know. The information users receive in response to their requests is organized, selected, and ranked in a way that had not been previously decided by anyone and cannot be attributed to none other than the search engine. Search engines give access to information they produced themselves (Toobin, 2014). But how do algorithms produce it, and do they understand it?

Data-driven agency

The legislation collides with the difficulties related to new forms of agency in the digital world (Hildebrandt, 2015). The actor that selected and produced the additional information in Google is an algorithm (PageRank or similar), which uses the available signals to produce information that was foreseen neither by the programmers nor by the authors nor by the user. The produced information, if it was known to someone, was only known to the algorithm – but does it make sense to say that the algorithm knows it? And does it make sense to hold it accountable?

Algorithms process data (and manage information) in a different way than human information processing and understanding – and this is the root of the success of Big Data. The recent approach of Big Data is actually quite distant from the models of Artificial Intelligence (AI) from the 1970s and 1980s, that aimed, by imitation or by analogy (“strong” and “weak” AI) at reproducing with a machine the processes of human intelligence (Nilsson, 2010). Now this is no longer what the systems do, and some designers declare it explicitly: “We do not try and copy intelligence” (Solon, 2012), it would be too heavy a burden. Examples can be multiplied from all areas in which algorithms are most successful. Translation programs do not try to understand the documents and their designers do not rely on any theory of language (Boellstorff, 2013). Algorithms translate texts from Chinese without knowing Chinese, and their programmers do not know it either. Spell checkers can correct typographical errors in any language because they do not know the languages or their (always different) spelling rules. Digital assistants operate with words without understanding what words mean and text-producing algorithms “don’t reason like people in order to write like people” (Hammond, 2015: 7). Algorithms competing with human players in chess, poker, and Go do not have any knowledge of the games nor of the subtleties of human strategies (Silver and Hassabis, 2016). Recommendation programs using collaborative

filtering know absolutely nothing about the movies, songs, or books they suggest and can operate as reliable tastemakers (Grossman, 2010; Kitchin, 2014: 4).

Just as human beings first became able to fly when they abandoned the idea of building machines that flap their wings like birds,¹¹ digital information processing only managed to achieve the results that we see today when it abandoned the ambition to reproduce in digital form the processes of the human mind. Since they do not try to resemble our consciousness, algorithms become more and more able to act as competent communication partners, responding appropriately to our requests and providing information that no human mind ever developed and that no human mind could reconstruct.

Practices like machine learning and knowledge discovery in databases allow algorithms to produce information that does not start from meaningful elements—they do not process information. Algorithms only process data. Data by themselves are not meaningful. They are just numbers and figures, that only become significant when processed and presented in a context, producing information. Information requires data, but data are not enough to have information. The same data can be or not be informative for different people and in different contexts, for example when a communication is repeated. That there is a train strike or a stock market crash is not informative any more when you read the news for the second time on a different newspaper, although the data remain the same. The same news can also be informative for someone but not for others, who maybe are not interested in finance. Referring to Bateson’s definition of information as a “difference that makes a difference” (Bateson, 1972: 582), we can say that data are differences (strike/no strike) which become informative when they make a difference for someone in a given moment (who, for example, decides to go by car rather than by train, or to stay at home).

Algorithms only process differences, from whatever source and with whatever meaning. They only need data that they get from the web, deriving them not only from what we think but also from what we do without thinking and without being aware of. Digital machines are able to identify in the materials circulating on the web patterns and correlations that no human being identified, processing them in such a way as to be informative for the users. Human beings, however, need information. When communicated to users, the results of algorithmic processing generate information and have consequences (Agrawal, 2005; Hammond, 2015), but outgoing information does not need incoming information. The revolutionary communicative meaning of Big Data is the ability to do without information while producing information. In Mireille

Hildebrandt's words: "We have moved from an information society to a data-driven society" (Hildebrandt, 2015: 46).

The memory of the web society

When it uses algorithms, social memory shows this kind of data-driven agency, working in a way different from our familiar forms of memorizing and, thus, creating different problems. Whereas in the past the problem of memory was the inability to remember, now the problem of social memory is the inability to forget.¹² Especially since the spread of the Web 2.0, with its virtually unlimited capacity to store and process data, the web seems to allow for a form of perfect remembering. Our society seems to be able to remember everything.¹³ The default value automatically attained, if you don't decide otherwise, and which demands neither energy nor attention, is now remembering—not forgetting (Blanchette and Johnson, 2002). To remember has become much easier and cheaper—remembering has become the norm. Only as an exception, if it becomes necessary, we do decide to forget.

Think of our everyday praxis on the web while dealing with texts, pictures, and e-mails. We lack the time to choose and to forget. Without deciding to preserve anything, as habit we preserve everything, as the machine invites us to do. To choose and to decide requires more attention and time. Usually there is no need to do so, due to the availability also of very effective techniques to search out interesting information in the mass of data, as and when the need arises: for example to find a particular message among the saved e-mails. Then we remember everything, recording it in the spaces (in the cloud) of a web which by itself does not have any procedure to forget.¹⁴ The judgment of the European Court reflects this approach: the problem is the accessibility for internet users of citizens' data in the indelible archives of the web, and the law wants to protect the ability of the web to forget (and the possibility for the citizens to be forgotten).¹⁵

Thereby the web succeeds in managing and making available an enormous amount of information. But does it make sense to say that the web has unlimited memory or even that it has a memory? Memory is not just storage, and an efficient memory is not made of unlimited data. Memory requires the ability to focus and select data, and to produce information referring to a meaningful context, which implies both the ability to remember and the ability to forget.

When it comes to memory, not only everyday speech but also a large part of scientific reflection basically refer to the management of *remembering*. Increasing memory is understood as increasing memories or strengthening the ability to remember. In this view

forgetting appears only as the negation of memory (Ricoeur, 2004: 412): if forgetting increases, remembering decreases and vice versa. The opposite idea, that *forgetting* is a key component of memory, required for abstraction and reflection, is not new, but always remained in the shadows. From Themistocles (Cicero, *de Oratore* 2.74.299) on there have always been voices claiming that the ability to forget is even more important than the ability to remember (Weinrich, 1996). Remembering and forgetting are the two sides of memory, both essential for its functioning (Esposito, 2002).

Forgetting is not the simple erasure of data. Instead, it is an active mechanism that inhibits the memorization of all stimuli except a few ones, enabling to focus attention and to autonomously organize information in one's own processes (Anderson, 2003; Hulbert and Anderson, 2008). Forgetting is needed to focus on something and use past experience to act in a flexible, context appropriate manner, not starting from scratch each time but also not doing always the same whenever the same situation occurs. One must be able to distinguish the present moment from an eternal presence of the past. Forgetting then is also needed to be able to remember in a proper sense, building an internal horizon of references and recursions to face the present. The act of remembering produces and requires a parallel forgetting (Hulbert and Anderson, 2008: 8).

The web, which stores all data in a kind of eternal present (Lepore, 2015), is not able to forget but is not even able to properly remember. The processing of data is entrusted to algorithms, which do not use abstraction and do not need it. Therefore, they can develop the amazing efficiency that we observe in our everyday use of the web. Algorithms such as PageRank are so fast and powerful because they do not need to "understand" the information on which they work in order to classify and connect it in a way that becomes meaningful for the users.¹⁶ The machine works without abstraction and without reference to meaning. Merely calculating, algorithms manage to produce intelligent and significant results not because they operate in an intelligent way, but because they "parasitically" exploit the intelligence and the attribution of meaning by the users of the web, in a process that continuously feeds on itself (Esposito, 2014). All successful web projects use, in one way or another, "googlization" practices (Rogers, 2013; Vaidhyanathan, 2011), harvesting, copying, aggregating, and processing data derived from user behavior—as PageRank does to produce an updated and efficient ranking of websites. Google uses links to learn how important a page is, but also to learn what it is about and to direct its own internal organization, which is continuously renewed depending on the

connections and affinities “discovered” in the operations of users (Langville and Meyer, 2006).

Dealing with data, algorithms that cannot abstract behave like the memorists studied by Luria or like the patients of hypermnestic syndrome (Erdelyi, 1996; Luria, 1987; Parker et al., 2006)¹⁷ who cannot forget. Like these patients, they are not able to activate the mechanism that distinguishes what they are interested in remembering. They are not able to build their own abstract context that guides selection and forgetting. Abstracting is actually remembering *and* forgetting. Algorithms do not abstract, they merely calculate. They do not properly remember and do not properly forget.

When algorithms allow us to contextualize/forget (and they do it: we get from Google selective lists of links to sites that may be of interest to us), they can do it not because they learn to select, but because they “import” in their procedures the selections made by users and use them to guide their own behavior. The criteria for deciding which sites are relevant and should appear first in the list of search results are not produced by the algorithm and are not even decided from the beginning by programmers but are derived from the choices of the users. A website is considered relevant by the algorithm if many web users connected to it many times.¹⁸ The context of the selection is derived from previous contextualizations. The algorithm forgets what had been forgotten by the users.¹⁹

Forgetting without remembering

How can we deal with a social memory driven by algorithms that do not understand and do not abstract, as the memory of our digital society that derives from the web the information circulating in communication? How can we ensure at the same time the preservation of the past and the openness of the future, when the agents that manage data move in an eternal present, without remembering and without forgetting?

The most evident aspect of digital media is a shift of problem from analog memory. Our society was always concerned with protecting the ability to remember (storing and retrieving data), while today we are primarily concerned with protecting the ability to forget (contextualizing). But the two sides of memory have an interesting asymmetry, known since ancient times. You can decide to enhance remembering, and with *ars memoriae* (Yates, 1966) we have for thousands of years elaborate techniques to do so. But we do not have an *ars oblivionalis* as an effective technique to enhance forgetting (Lachmann, 1991: 11; Weinrich, 1997: 9ff). If you want to forget and decide to do it, the most immediate effect is the opposite of the intended one, because you draw attention to the content at stake, increasing first

remembering.²⁰ Remembering to forget is paradoxical, and deciding to forget almost impossible.

Also on the web this kind of “boomerang effect” has been observed. Reputation management sites on the web (cf. reputation.com) warn that attempts to remove content are often counterproductive (Woodruff, 2014). Once the request of “forgetting” has been accepted by Google, when someone does a search on a particular person, among the results appears a warning that some of the contents have been removed in the name of the right to be forgotten. The obvious consequence is to increase curiosity and interest in that content. Sites have been immediately produced (like hiddenfromgoogle.com) that collect the links removed because of the right to oblivion. Wikipedia also released a list of links to articles that Google had removed from its search engine in accordance to the “right to be forgotten.”²¹ Ironically, these “reminders” of the contents that the law requires to forget are perfectly legal, because the ruling prohibits only to keep the links to the pages, not the content of the pages themselves, which continue to be available on the websites of the newspapers or of the other sources that diffused them.

Hindering remembering is not enough to produce forgetting. You have to circumvent the paradox of remembering to forget in an indirect, more complex way. Mnemotechnics itself recognized that in order to reinforce forgetting you should rather multiply the range of available memories (Weinrich, 1996). If you increase memories, every piece of information is lost in the mass and becomes difficult to find, so in fact it is lost as if it were forgotten. This practice never produced an authentic technique (an *ars oblivionalis*) because of human’s limited capacity to store and process data (to remember) that would be overloaded by an unmanageable mass of memories. To be able to forget we would have to give up the ability to remember. Algorithms, however, do not have this problem because of their virtually unlimited capability to manage data, which is the basis of their excessive remembering but can also be used to reinforce forgetting.

Thus, to control forgetting on the web in a manner specific to algorithmic memory, one could adopt a procedure directly opposed to the practice of deleting contents or making them unavailable. This is the direction in which some recent techniques for protection of privacy are going, which is often understood as protection of forgetting. Strategies of “obfuscation” (Brunton and Nissenbaum, 2015) have been designed to produce misleading, false, or ambiguous data parallel to each transaction on the web—in practice multiplying the production of information to hinder a meaningful contextualization. If together with every search for information in the web, or together with any input of

information on social media like FaceBook, a dedicated software produces a mass of other entirely irrelevant operations, it will be difficult to select and focus on relevant information, i.e. to remember.²²

These techniques, however, require an a priori selection of the memories you want to forget, for which you activate the obfuscation process. But in many cases you want to forget memories that you never thought to have to forget, and these are the cases targeted by the legislation on the right to be forgotten.²³ There are proposals, partly effective, that adopt the same approach to produce a posteriori an equivalent of forgetting. They act directly on Google's search results through the multiplication of information. When a person has been publicly shamed on the web, they produce artificially fake sites with other independent information, with the explicit purpose of pushing the shaming information so far down the search result that it effectively vanishes (Ronson, 2015: 214ff.). The service ReputationDefender²⁴ starts from the assumption that "deleting is impossible." To combat negative or undesired items about a person, they generate a wide range of unique, high quality positive content about that person and push it up in the search results. As a result, "negative material gets dumped down to pages where nobody will see it."

The idea is not to erase memories but to enhance forgetting. When the algorithm multiplies data, it does not pay attention to this process—it doesn't "remember" it. The multiplication of memories goes on in the machine without meaning and without understanding. This proliferation makes each datum more marginal, lost in the mass. As in forgetting, it becomes increasingly difficult to find and to use, thereby fulfilling the right to oblivion. The factual conditions of forgetting are carried out without having to activate remembering, bypassing in a sense the paradox of *ars oblivionalis*.

But artificial memory, as remembering and as forgetting as well, requires constant maintenance. Mnemotechnics work if you go on taking care of the palaces and caves of memory (Bolzoni, 1995). Memory athletes should not stop training (Foer, 2011). Similarly, an effective artificial forgetting must always be renewed, because Google constantly changes its algorithms and its targets (Ronson, 2015: 267ff.; Woodruff, 2014: 157). Forgetting does not happen once and for all, as an erasure of memories. You must reverse engineer Google and continue to renew forgetting as an active process, producing more and different memories with different strategies.

Data-driven memory

These forgetting strategies are ingenious but address the issue of forgetting from the point of view of

information management: how is it possible to forget the information available to search engines. They adopt the same approach as the European Court of Justice. But algorithms do not work with information. They work with data, creating different problems.

The legislation on the right to be forgotten addresses the indexing of pages in the search engine. When the request of a citizen is accepted, the indexing is blocked and Google is not allowed to provide a link when a search is made, even if the data remain available in their original location (e.g., the digital archive of a newspaper). Google cannot deliver the information to the users answering their query. It is like blocking the use of the catalog of a library, while at the same time preserving the books and other materials. This solution corresponds to the legislator's attempt to combine the protection of forgetting with the parallel need to protect memory. As Viviane Reding, the European Commission's vice-president, said: "It is clear that the right to be forgotten cannot amount to a right of the total erasure of history." To preserve the openness of the future one would not want to lose the past. All data are still stored at the respective sites, but the "forgotten" items are no longer accessible via Google search. The ruling acts on remembering, not on memory. This of course leaves the users exposed to the boomerang effect of forgetting, since the original pages continue to be available on the web and can become accessible (can be remembered) with different search tools (or even with google.com or any of its sites outside Europe).

But there are deeper, more fundamental problems. Google's indexing, as the catalog of a library, delivers information. The algorithm itself, however, as all algorithms operating on the web, "feeds" on data, which are much more diffused and much more extensive than the information understood and thought by someone at some time.²⁵ Algorithms derive data from the information available in materials on the web (texts, documents, videos, blogs, files of all types) and from the information provided by users: their requests, recommendations, comments, chats. They are also able to extract data from information on information: the metadata that describe content and properties of each document, such as title, creator, subject, description, publisher, contributors, type, format, identifier, source, language, and much more. Each of this data refer to a different context than the original information, of which the author is usually not aware, and that he does not explicitly intend to communicate. The Internet of things and other forms of ambient intelligence also produce a multitude of data individuals do not realize, monitoring their behavior, their location, their movements, and their relationships.

Moreover, and most importantly, algorithms are able to use all these data for a variety of secondary

uses largely independent of the intent or the original context for which they were produced (Mayer-Schönberger and Cukier, 2013: 103), processing them to find correlations and patterns with calculations that the human mind could not realize nor understand, but which become informative. These include curious examples such as the fact that vegetarians miss fewer flights (Siegel, 2016), or that the divorce rate in Maine correlates with per capita consumption of margarine (<http://www.tylervigen.com/spurious-correlations>). But such secondary use of data also makes it possible to gain information relevant for the profiling and surveillance of citizens.

In these processes algorithms use the “data exhaust” (Mayer-Schönberger and Cukier, 2013: 6) or the “data shadows” (Koops, 2011) generated as by-product by people’s activities on the web and, increasingly, in the world at large. It is a sort of “data afterlife” (Adkins and Lury, 2012: 6) which goes far beyond the representational quality of numbers and of information and depends on the autonomous activity of algorithms. Each difference makes a difference in many different ways, more and more independent from the original difference. Algorithms use data to produce information that cannot be attributed to any human being. In a way, algorithms remember memories that had never been thought by anyone.

This is a great opportunity for the social management of information, but also a most serious threat to the freedom of self-determination of individuals and to the possibility of an open future. Information can be inaccessible to the indexing in accordance with the right to be forgotten, while data continue to be remembered and used by the algorithms to produce different information (Amoore and Piotukh, 2015: 355).²⁶ Moreover, the implementation of the right to be forgotten involves itself the collection of a lot of metadata on which personal data can be used on what purpose, revealing personal preferences that, albeit anonymized, can be exploited for profiling (Custers, 2016).

Conclusion

The current legislation on the right to be forgotten on Google addresses the problems of digital memory in ways that still refer to human processes of memory and communication, supported by analog media.²⁷ The difficulties that arise signal that to handle and regulate digital forms of “second-order memory” we need a more radical approach (Koops, 2011: 250). Digital algorithms-based memory is not made of information, but uses the information recorded in communication (primary memory) to get its data, that are often different from those on which the original information was based, and partly uncontrollable. This is not to say that

the production and circulation of memories cannot be controlled, because the need for control rather increases with the power of algorithms. As the most effective proposals to manage the excess of data show, however, the problems are different. We must face algorithms directly as autonomous agents, with processes, procedures, and problems that cannot be traced back to our familiar forms of attribution and accountability. This is what Google claimed in the European Court case on the right to be forgotten, and what the ruling left unanswered.

Acknowledgements

Research for this paper was supported by fellowships at the Italian Academy at Columbia University and at the Institute for Advanced Study at the University of Warwick. For comments, criticisms, and suggestions on an earlier draft of this paper, I would like to thank David Stark, Jonathan Bach, Robin Wagner-Pacifici and the anonymous referees of *Big Data & Society*.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. The dataset on which algorithms work is always related to human intervention: it includes only the data that have been provided, could be different if it were set in another way, and many of the data result from the behavior and decisions of individuals. The procedures of algorithms, moreover, are clearly the outcome of human design with its references and its biases, also in the cases in which the designers themselves do not know the details of the working of the machines.
2. The responsibility of algorithms could also be extended to the organizations or persons using and programming them (Floridi, 2017).
3. http://curia.europa.eu/juris/document/document.jsf?jsessionid=9ea7d0f130d5379d711f29d04ed998979558ddf6668b.e34KaxiLc3eQc40LaxqMbN4Pa3qOe0?text=&docid=152065&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=687556#Footnote*
4. Cf., for example: <http://www.scientificamerican.com/article/the-internet-never-forgets/>.
5. “Zu allem Handeln gehört Vergessen (...); es ist möglich fast ohne Erinnerung zu leben, ja glücklich zu leben, wie das Tier zeigt; es ist aber ganz und gar unmöglich, ohne Vergessen überhaupt zu leben” (Nietzsche, 1874: 116)—“all action requires forgetting (...); it is possible to live almost without remembering, it is even possible to live

- happily, as animals show. But it is absolutely impossible to live without forgetting.”
6. On the distinction between responsibility and accountability, cf. Simon (2012). According to Simon, responsibility would require intentionality, which cannot be attributed to technical artifacts. Algorithms may be accountable, but should not be made responsible – which of course leaves the issue of responsibility open.
 7. Google not only has the task of suppressing the links, but also of deciding whether to accept the requests of “being forgotten,” balancing them with the public right to information. This raises an important question of legitimacy (Ambrose, 2015). With which right can a private entity take decisions of public importance, without having being elected or appointed with a transparent procedure, and even without specification of the criteria to be followed in the decision (if and how long data can be considered of public interest, who are the private individuals to protect, when should the right to privacy of the individual prevail over public access)? Citizens still have the possibility of appealing to the judicial authority (§82), but this is a second step.
 8. The extensive debate on cultural memory is focused on the definition of social memory referring to this question. Starting with Halbwachs’ (1952) classic text, cf. Assmann (1992), Assmann and Hölscher (1988), and Esposito (2002).
 9. Google tried to get directives and guidelines in a series of meetings and discussions organized in 2014 in a tour of European capitals. As it could be expected, the results were more media than content – but this does not make the move less meaningful from the point of view of the company and of the management of its decision-making responsibility.
 10. The legislation on the right to be forgotten includes a “media exception” for the processing of personal data “carried out solely for journalistic purposes” (§9; Rosen, 2012). But also archives and catalogs are protected: there is a further exception for processing data for historical, statistical, and scientific research purposes (§7; Van Hoboken, 2013: 20). Intermediaries, on the other hand, are not held responsible if they do not know the information to which they provide access (van Hoboken, 2013: 26)—as stated in the EU directive on electronic commerce that establishes a “safe harbor” for Internet service providers that operate as “mere conduit.”
 11. The metaphor was used by Hans Blumenberg (Blumenberg, 1957).
 12. This is one of the points emphasized by Mayer-Schönberger (2009) in his book with the evocative title *Delete*.
 13. It even remembers the future, working as an “anticipation machine” which also answers questions not yet asked. Google’s function “autocomplete” finishes writing the question while you are typing the first few words. On the emphasis about the predictive power of algorithms, see Siegel (2016).
 14. As many observers remark, however, the web also forgets a lot, but again in different ways than our familiar memory. Web content is ephemeral on many dimensions, most of them new (Ambrose, 2012; Barone et al., 2015; Chun, 2011; Lepore, 2015). Besides the classic problems of physical rot (natural causes such as fire or flood) there are technical problems like hardware, software or network failures, viruses, accidental file deletion, changes of media and protocols.) Moreover, there are all the difficulties of reference rot and link rot: contents that become impossible to access, with the link producing the infamous error messages 404: Page Not Found, or pages that still exist but with a different URL (Davis, 2010). The average life of web pages is less than 100 days, and in many cases is better in measured in hours rather than days (<http://blogs.loc.gov/digitalpreservation/2011/11/the-average-lifespan-of-a-webpage/>). Reacting to these problems, specialized tools for Web Preservation have been developed, as permalinks (<https://perma.cc>), the Wayback Machine (<https://archive.org/web/>) or Google’s Cache.
 15. The two aspects of forgetting are obviously different, as highlighted by Rouvroy (2008), who distinguishes the interest to forget from the interest (or even the right) to be forgotten. The first aspect concerns the possibility of projecting an open future, the second the desire not to be bound by the past (or by certain portions of it) in one’s social identity. In the debate on the European ruling, however, the two issues overlap and in this context I will not consider the differences. Nobody, however, wants to forget or be forgotten altogether, and in this sense the expression “right to be forgotten” is somehow misleading.
 16. Unlike other tools such as AltaVista, Google is not based on a predetermined semantic data grid, but derives its classification from the data examined: meaning is the result of the processing, not its premise.
 17. Narratively reproduced as Funes in Borges’ (1962) famous short story.
 18. And if relevant users link to it—while users are relevant if they themselves are linked by others (Page et al., 1999).
 19. And somehow learns, because machine learning techniques use selections also to orient the subsequent behavior – but not because it understands the contents.
 20. Cf. Eco (1987). For the web it is called “Streisand effect,” similar to the one known and widely studied about censorship—the reason why one should usually refrain from suing defamatory articles, to avoid spreading the news even more: politicians, actors, and all public figures know it very well.
 21. An “index of the de-indexed”: cf. Binns (2016).
 22. With a similar attitude, FaceCloak generates fictitious information to oppose the excess of individual transparency on Facebook by creating, parallel to sensitive data, a series of completely irrelevant invented information. There are also procedures which generate whenever you do a query on Google a series of parallel ghost queries that make it difficult for companies to identify your pattern of preferences.
 23. The most complex aspect of reputation management is repair (Woodruff, 2014). Ausloos (2012) observes that the right to be forgotten only provides *ex post* solutions to privacy issues.

24. reputationdefender.com.
25. “Most of the Digital Universe are unstructured data”: IDC—IVIEW 2010.
26. The same problems arise in the debate about digital privacy, which in its most refined forms also refers to the problem of preserving the self-determination of individuals as possibility of reinvention (Solove, 2007a, 2007b, 2011). Here the advocates of privacy claim the protection of “contextual integrity” (Nissenbaum, 2004: 137): you should not allow the use of data in contexts that are inappropriate to the original one. The difficulty, however, is that in many cases the working of algorithms completely disregards the context.
27. According to Ausloos (2012: 11), the very idea that individuals can have control of their data might be only a remainder of the preinternet era where the default was forgetting.

References

- Adkins L and Lury C (eds) (2012) *Measure and Value*. Malden, MA and Oxford: Wiley.
- Agrawal R (2005) Interview with Marianne Winslett. Available at: <http://www.sigmod.org/publications/interview/pdf/D15.rakesh-final-final.pdf> (accessed 15 April 2016).
- Ambrose LM (2012) You are what Google says you are: The right to be forgotten and information stewardship. *International Review of Information Ethics* 17(7): 21–30.
- Ambrose LM (2015) Forgetting made (too) easy. *Communications of the ACM* 34(June): 34–35.
- Amoore L and Piotukh V (2015) Life beyond big data: Governing with little analytics. *Economy and Society* 44(3): 341–366.
- Anderson MC (2003) Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language* 49: 415–445.
- Assmann J (1992) *Das kulturelle Gedächtnis. Schrift, Erinnerung und politische Identität in frühen Hochkulturen*. München: Beck.
- Assmann J and Hölscher T (eds) (1988) *Kultur und Gedächtnis*. Frankfurt a.M.: Suhrkamp.
- Ausloos J (2012) The ‘right to be forgotten’ – Worth remembering? *Computer Law and Security Report* 28(2): 143–152.
- Barone F, Zeitlyn D and Mayer-Schönberger V (2015) Learning from failure: The case of the disappearing Web site. *First Monday* 20(5). <http://dx.doi.org/10.5210/fm.v20i5.5852>.
- Bateson G (1972) *Steps to an Ecology of Mind*. New York: Ballantine Books.
- Binns R (2016) How to be open about being closed. *Limn* 6. Available at: <http://limn.it/how-to-be-open-about-being-closed/> (accessed 5 November 2016).
- Blanchette JF and Johnson DG (2002) Data retention and the panoptic society: The social benefits of forgetfulness. *The Information Society* 18: 33–45.
- Blumenberg H (1957) Nachahmung der Natur. Zur Vorgeschichte der Idee des schöpferischen Menschen. *Studium Generale* 10: 266–283.
- Boellstorff T (2013) Making big data, in theory. *First Monday* 18(10). <http://dx.doi.org/10.5210/fm.v18i10.4869>.
- Bolzoni L (1995) *La Stanza Della Memoria. Modelli Letterari e Iconografici Nell’età Della Stampa*. Torino: Einaudi.
- Borges JL (1962) Funes the Memoriosus. In: *Labyrinths*. New York: New Directions, pp.148–154.
- boyd d and Crawford K (2012) Critical questions for big data. *Information, Communication and Society* 15(5): 662–679.
- Brunton F and Nissenbaum H (2015) *Obfuscation: A User’s Guide for Privacy and Protest*. Cambridge: The MIT Press.
- Burrell J (2016) How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society* 1: 1–12.
- Chun WHK (2011) *Programmed Visions: Software and Memory*. Cambridge: The MIT Press.
- Custers B (2016) Click here to consent forever: Expiry dates for informed consent. *Big Data & Society* 2: 1–6.
- Davis R (2010) Moving targets: Web preservation and reference management. In: *Innovations in reference management workshop*, January. Available at: <http://www.ariadne.ac.uk/issue62/davis/> (accessed 20 March 2016).
- Eco U (1987) An ars oblivionalis? Forget it! *Kos* 30: 40–53.
- Erdelyi MH (1996) *The Recovery of Unconscious Memories. Hypermnesia and Reminiscence*. Chicago, IL: University of Chicago Press.
- Esposito E (2002) *Soziales Vergessen*. Frankfurt a.M.: Suhrkamp.
- Esposito E (2014) Algorithmische Kontingenz. Der Umgang mit Unsicherheit im Web. In: Cevoloni A (ed.) *Die Ordnung des Kontingenten. Beiträge zur Zahlenmäßigen Selbstbeschreibung der Modernen Gesellschaft*. Wiesbaden: Springer VS, pp. 233–249.
- Etzioni O (2016) Deep learning isn’t a dangerous magic genie. It’s just math. *wired.com*, 15 June.
- Floridi L (2017) L’ultima legge della robotica. *La Repubblica Robinson* 12: 5–7.
- Floridi L and Sanders JW (2004) On the morality of artificial agents. *Minds and Machine* 14: 349–379.
- Foer J (2011) *Moonwalking with Einstein*. London: Penguin.
- Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski PJ and Foot KA (eds) *Media Technologies*. Cambridge: The MIT Press, pp. 167–194.
- Gitelman L (ed) (2013) “Raw Data” is an Oxymoron. Cambridge: MIT Press.
- Grossman L (2010) How computers know what we want – Before we do. *Time*, 27 May.
- Halbwachs M (1952) *Les Cadres Sociaux de la Mémoire*. Paris: P.U.F.
- Hammond K (2015) *Practical Artificial Intelligence for Dummies*. Hoboken, NJ: Wiley.
- Hildebrandt M (2015) *Smart Technologies and the End(s) of Law*. Cheltenham: Elgar.
- Hulbert JC and Anderson MC (2008) The role of inhibition in learning. In: Aaron S, Benjamin J, de Belle S, et al. (eds) *Human Learning*. New York: Elsevier, pp. 7–20.
- Kitchin R (2014) Big data, new epistemologies and paradigm shifts. *Big Data and Society* 2: 1–12.
- Koops BJ (2011) Forgetting footprints, shunning shadows. A critical analysis of the right to be forgotten in big data practice. *Scripted* 8(3): 229–256.

- Koops BJ, Hildebrandt M and Jaquet-Chiffelle DO (2010) Bridging the accountability gap: Rights for new entities in the information society? *Minnesota Journal of Law, Science and Technology* 11(2): 497–561.
- Lachmann R (1991) Die Unlösbarkeit der Zeichen: Das semiotische Unglück des Memoristen. In: Haverkamp H and Lachmann R (eds) *Gedächtniskunst: Raum-Bild-Schrift*. Frankfurt a.M.: Suhrkamp, pp. 111–141.
- Langville AN and Meyer CD (2006) *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton, NJ and Oxford: Princeton U.P.
- Lepore J (2015) The Cobweb. Can the Internet be archived? *The New Yorker*, January, 34–41.
- Luria AR (1987) *The Mind of a Mnemonist: A Little Book About a Vast Memory*. Cambridge, MA: Harvard University Press.
- Mayer-Schönberger V (2009) *Delete: The Virtue of Forgetting in the Digital Age*. Princeton, NJ: Princeton U.P.
- Mayer-Schönberger V and Cukier K (2013) *Big Data. A Revolution That Will Transform How We Live, Work, and Think*. London: Murray.
- Mittelstadt BD, Allo P, Taddeo M, et al. (2016) The ethics of algorithms: Mapping the debate. *Big Data & Society* July–December: 1–21.
- Nabi Z (2014) Resistance censorship is futile. *First Monday* 19(11)<http://dx.doi.org/10.5210/fm.v19i11.5525>.
- Nietzsche F (1874) *Unzeitgemässe Betrachtungen. Zweites Stück: Vom Nutzen und Nachteil der Historie für das Leben*. Berlin and New York: de Gruyter, 1972.
- Nilsson NJ (2010) *The Quest for Artificial Intelligence. A History of Ideas and Achievements*. New York: Cambridge University Press.
- Nissenbaum H (2004) Privacy as Contextual integrity. *Washington Law Review* 79: 79–119.
- Page L, Brin S, Motwani R, et al. (1999) *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report, Stanford Infolab. Available at: <http://ilpubs.stanford.edu:8090/422/> (accessed 5 February 2016).
- Parker ES, Cahill L and McGaugh JL (2006) A case of unusual autobiographical remembering. *Neurocase* 12: 35–49.
- Ricoeur P (2004) *Memory, History, Forgetting*. Chicago, IL: University of Chicago Press.
- Rogers R (2013) *Digital Methods*. Cambridge and London: The MIT Press.
- Ronson J (2015) *So You've Been Publicly Shamed*. New York: Riverhead.
- Rosen J (2012) The right to be forgotten. *Stanford Law Review Online* 88: 64.
- Rouvroy A (2008) Réinventer l'art d'oublier et de se faire oublier dans la société de l'information? In: Lacour S (ed.) *La sécurité de l'individu numérisé Réflexions prospectives et internationales*. Paris: L'Harmattan, pp. 249–278.
- Siegel E (2016) *Predictive Analytics. The Power to Predict Who will Click, Buy, Lie or Die*. Hoboken (NJ): Wiley.
- Silver D and Hassabis D (2016) Alpha go: Mastering the ancient game of go with machine learning. Available at: <https://research.googleblog.com/2016/01/alphago-mastering-ancient-game-of-go.html> (accessed 27 January 2016).
- Simon J (2012) Epistemic responsibility in entangled socio-technical systems. In: Dodig-Crnkovic G, Rotolo A, Sartor G, et al. (eds) *Social Computing, Social Cognition, Social Networks and Multiagent Systems. Social Turn. AISB/IACAP World Congress 2012, Birmingham (UK): SNAMAS*, pp. 56–60.
- Solon O (2012) Weavrs. The autonomous, tweeting blog-bots that feed on social content. *Wired.co.uk*, 28 March.
- Solove DJ (2007a) *The Future of Reputation. Gossip, Rumor and Privacy on the Internet*. New Haven, CT and London: Yale U.P.
- Solove DJ (2007b) 'I've got nothing to hide' and other misunderstandings of privacy. *San Diego Law Review* 44: 745–772.
- Solove DJ (2011) Speech, privacy, and reputation on the internet. In: Nussbaum M and Levmore S (eds) *The Offensive Internet: Speech, Privacy, and Reputation*. Cambridge, MA: Harvard U.P., pp. 15–30.
- Toobin J (2014) The solace of oblivion. In Europe, the right to be forgotten trumps the Internet. *The New Yorker*, 29 September.
- Vaidhyanathan S (2011) *The Googlization of Everything (And Why We Should Worry)*. Berkley and Los Angeles: University of California Press.
- van Hoboken J (2013) *The Proposed Right to be Forgotten Seen from the Perspective of Our Right to Remember, Freedom of Expression Safeguards in a Converging Information Environment*. Available at: http://www.law.nyu.edu/sites/default/files/upload_documents/VanHoboken_RightTo%20Be%20Forgotten_Manuscript_2013.pdf (accessed 10 November 2015).
- Weinrich H (1996) *Gibt es eine Kunst des Vergessens?* Basel: Schwabe & Co.
- Weinrich H (1997) *Lethe. Kunst und Kritik des Vergessens*. München: Beck.
- Wilding J and Valentine E (1997) *Superior Memories*. Hove: Psychology Press.
- Woodruff A (2014) Necessary, unpleasant, and disempowering: Reputation management in the internet age. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. New York, NY: ACM.
- Yates FA (1966) *The Art of Memory*. London: Routledge & Kegan Paul.