

5

ALGORITHMIC MEMORY AND THE RIGHT TO BE FORGOTTEN

REMEMBERING TO FORGET

On March 13, 2014,¹ the European Court of Justice issued a judgment in favor of the plaintiff on case C-131/12 about 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.”² The ruling directly addresses the role of algorithms in the processing of social information, and raises a lively debate around the consequences of digitalization for memory.

The judgment reacted to a complaint lodged by a Spanish citizen against Google. The company was accused of infringing upon 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 matter 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 Google should be held responsible for the processing of personal data, and should be forced to suppress links to web pages

containing information on the person in question, even if that information remains available—lawfully published—on web pages where it is hosted.

The problem to which the European Court responded with its ruling 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 amounts of data, with no apparent limit to 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* (right to be forgotten) which has a long legal tradition emerging out of French law. This right protects the will of a citizen who has been convicted of a criminal act and has paid the debt to society to no longer be 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 one's future open—a right to reinvention that protects the future of the person from a colonization by the past.³ The nineteenth-century philosopher Friedrich Nietzsche knew it very well when he spoke of the “need of oblivion for life” as even more important than the ability to remember⁴—because without forgetting, one would remain bound to an eternal presence of the past that does not allow for building 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 links to the personal data of those who request it—unless that information has public relevance. However, with search engines giving access to the voluminous data available online, the right to be forgotten protected by the European Court becomes much more extensive than the classic right to be forgotten, both materially and socially: it concerns any act (especially those inconsequential on the penal level yet relevant for image and reputation) and includes any person (not only criminals but each of us, particularly teenagers).

The forgetting of anyone, though, also affects the forgetting of others—such as those who are involved in the same event and may not want it to be forgotten, or those who may become affected in the future or have an interest in similar events and want to preserve access to the relevant information. The protection of individual forgetting collides with the

right to information and with the creation of a reliable shared public sphere.⁵

The ruling of the European Court states that the right to privacy overrules the 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 web society.⁶

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,⁷ on the basis of a principle that holds that the responsible entity is “the natural or legal person” who “determines the purposes and means of the processing of personal data . . . whether or not by automatic means” (§4). Google, 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 this, although it distinguishes the processing of data by Google from the processing by publishers and journalists. Even if Google does not direct data processing, search engine activity makes data accessible to internet users, including those who would not have otherwise found some particular 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 merely publishing the information. The processing of data by Google is more subtle but more dangerous than that carried out by publishers and journalists; therefore, the company is charged with suppressing links to people who require those pages to be forgotten, even if the publication is lawful and the information remains available.⁸

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 computer memory

storage or remembering?⁹ Ascribing to Google the management of the right to oblivion implies a clear choice: data are considered forgotten if they are made difficult to find, 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,¹⁰ without any of the protections that legislation provides to media, archives, and other communication tools.¹¹ 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 data without knowing and without controlling it; thus, 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 in the same way as newspapers, publishers, and libraries, which select and organize the information to be disclosed. Search engines are purely passive intermediaries that merely provide access to materials they did not choose and do not know. The information that 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 anything other than the search engine. Search engines give access to information they produced themselves.¹² But how do algorithms produce and manage it?

DATA-DRIVEN AGENCY

Contemporary legislation collides with the new forms of agency in the digital world.¹³ The actor that selected and produced the additional information—the ranking—in Google is an algorithm such as PageRank that uses the available signals to produce information that was foreseen neither by its programmers, nor by content authors or search users. The information produced, if it was known to anyone, was known only to the algorithm itself—yet does it make sense to say that the algorithm *knows* it? And does it make sense to hold an algorithm accountable?

As discussed in chapter 1, algorithms deal with data in a different way than humans. Whereas human information processing refers to meaning, machine-learning practices allow algorithms to produce information

that does not start from meaningful elements. Algorithms do not process information, they only process data. *Data* by themselves are not meaningful. They are just numbers and figures, digital digits that only become significant when processed and presented in a context, producing information. Information requires data, but data is not enough to have information. The same data (e.g., about stock market movements) can be informative or not for different people in different contexts. Referring to Bateson's definition of "information" as a "difference that makes a difference,"¹⁴ we can say that data are differences (stock prices going up/stock prices going down) that become informative when they matter to someone in a given moment (who, e.g., decides to sell assets, or chooses not to invest).

Algorithms only process differences, from whatever source and with whatever meaning. They need only the data that they get from the web, deriving them from what we think and also from what we do without thinking and without being aware of it. Digital machines are able to identify in the materials circulating on the web patterns and correlations that no human being has identified, processing them in such a way as to be informative for their users. Human beings, however, need information. When communicated to users, the results of algorithmic processing generate information and have consequences,¹⁵ but outgoing information does not need incoming information: the revolutionary communicative meaning of big data is its ability to produce information from data that is not itself information. In Mireille Hildebrandt's words, "We have moved from an information society to a data-driven society."¹⁶

THE MEMORY OF A WEB-BASED SOCIETY

Whereas in the past the problem of memory was the inability to remember, now the problem of social memory is increasingly connected to the inability to forget.¹⁷ Especially since the spread of Web 2.0, with its virtually unlimited capacity to store and process data, the web seems to allow for a form of perfect remembering. Indeed, our society seems to be able to remember everything.¹⁸ The default value that holds automatically unless you opt out, which demands neither energy nor attention, is now remembering—not forgetting.¹⁹ It's become much easier and cheaper to

remember; remembering has become the norm. We decide to forget only as an exception, if it becomes necessary.

Think of our everyday practices on the web while dealing with texts, pictures, and emails. We lack the time to choose and to forget. By not making the decision to preserve anything, we habitually preserve *everything*, as the machine invites us to do. To choose and to decide to forget requires more attention and time. Usually there is no need to eliminate content, thanks to the availability of powerful techniques for searching out interesting information in the mass of data as and when the need arises—for example, in locating a particular message among a cache of saved emails. We therefore remember everything, recording it in the spaces (in the cloud) of a web which by itself does not have any procedure to forget.²⁰ The judgment of the European Court reflects this approach: the problem is the accessibility of citizens' data in the indelible archives of the web, and the law wants to create the ability of the web to forget (and the possibility that citizens be forgotten).²¹

But does it make sense to say that the web has a limitless memory, or even that it has a memory? The difficulties in implementing an effective regulation of forgetting are related to the fact that memory is not just storage, and efficiency in memory is not equivalent to unlimited data. Memory implies focusing on and selecting data to produce information that refers to a meaningful context. Memory thus requires both the ability to remember *and* the ability to forget.

This double nature of memory—remembering requires forgetting—is not always adequately taken into account. In common parlance and even in a large part of the scientific literature on the topic, memory ostensibly refers to the management of *remembering*. Increasing memory is understood as an increase in the number of memories or as strengthening the ability to remember. In this view, forgetting appears only as the passive negation of memory;²² if remembering increases, forgetting decreases, and vice versa. The opposite idea, that *forgetting* is a key component of memory, required for abstraction and reflection, is not new, although it has always remained in the shadows. From Themistocles in the sixth century BCE onward, there have always been voices claiming that the ability to forget is even more important than the ability to remember.²³

Remembering and forgetting, they argue, are the two sides of memory, each essential for its functioning.²⁴

This changes our understanding of forgetting. From this perspective, it is not simply erasure of data but an active mechanism that inhibits the memorization of all but a few stimuli, enabling one to focus one's attention and to autonomously organize information in accordance with one's own processes.²⁵ Forgetting is needed to focus on something and use past experience (that is, remembering) to act in a flexible, context-appropriate manner, rather than either starting from scratch each time or, indeed, always doing the same thing whenever a similar situation occurs.²⁶

The web, which stores all data in a kind of eternal present, is not able to forget, yet is also not even able to properly remember.²⁷ In dealing with data, algorithms behave like the mnemonist studied by Luria,²⁸ or like people living with hypermnesia, who cannot forget.²⁹ Like these individuals, algorithms are not able to activate the mechanism that distinguishes what they are interested in remembering from what they are not. However, memory is actually remembering *and* forgetting. Algorithms do not properly remember and do not properly forget; they merely calculate.

When algorithms allow us to forget (as they indeed do—we get from Google, for example, selective lists of links to sites that may interest us), they do it not because they learn to forget, but because their procedures “import” selections made by users to guide their own behavior.³⁰ The criteria for deciding which sites are relevant and should appear first in a list of search results are not produced by the algorithm and are not even decided from the beginning by programmers; instead, they are derived from the choices of previous users. A website is considered relevant to the algorithm if many web users connected to it many times.³¹ The algorithm forgets what had been forgotten by users.³²

FORGETTING WITHOUT REMEMBERING

How can we deal with a social memory driven by algorithms? How can we ensure both 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 influence of digital media has been a shift away from problems of analog memory. Traditional societies were always concerned with protecting the ability to remember (storing and retrieving data), while today we are primarily concerned with protecting the ability to forget.³³ But the two sides of memory have an interesting asymmetry, known since ancient times. You can decide to enhance remembering, and with *ars memoriae* we have for thousands of years developed elaborate techniques to do so.³⁴ But we do not have an *ars oblivionalis*—an art of forgetting—that would be an effective technique to enhance the ability to get rid of memories.³⁵ If you want to forget and decide to enhance that process, the most immediate effect is the opposite of the one intended because this draws attention to the content at stake, further cementing the initial memory.³⁶ For the web it is called the “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. Remembering to forget is paradoxical, and deciding to make something be forgotten, almost impossible.

On the web, this kind of boomerang effect has been observed. Reputation management sites on the web (e.g., reputation.com) warn that attempts to remove content are often counterproductive.³⁷ Once a request to “forget” has been accepted by Google, and a search on that particular person is performed, among the results appears a warning that certain contents have been removed in the name of the right to be forgotten. The obvious consequence is an increase in curiosity and interest in that content. Sites quickly emerged (like hiddenfromgoogle.com) that collect the links removed by virtue of this right to oblivion. Wikipedia has also released a list of links to articles that Google has removed from its search engine in accordance with the “right to be forgotten.”³⁸ Ironically, these “reminders” of the contents that the law requires be forgotten are perfectly legal because the ruling prohibits only the retaining of links to particular pages, and not to the contents of the pages themselves. Those pages continue to be available on newspaper websites or other sources that had diffused them.

Hindering remembering is not enough to induce forgetting. The paradox of remembering to forget must be circumvented in an indirect, more

complex way. The practice of using memory techniques (mnemotechnics) itself recognized that in order to reinforce forgetting, one should rather multiply the range of available memories.³⁹ If one increases memories by number, each piece of information is lost in the mass and becomes difficult to find, to the point where it becomes lost as if it were forgotten. This practice had never been able to yield a genuine technique (an *ars oblivionalis*) because of the limits of the human capacity to store and process data (to remember), which would be overloaded by such 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 capacity for managing data, which, while being the basis of their excessive remembering, 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 content or making them unavailable. This is the direction some recent techniques for protecting privacy is going, which is often understood as protecting forgetting. Strategies of obfuscation have been designed to produce misleading, false, or ambiguous data parallel to each transaction on the web⁴⁰—in practice, multiplying the production of information to hinder a meaningful contextualization. If, together with every search for information on the web, or together with any input of information on social media like Facebook, a dedicated software program produces a mass of other entirely irrelevant operations, it will be difficult to select and focus on relevant information—that is, to remember.⁴¹

These techniques, however, require a prior selection of the memories you want to forget, for which the obfuscation process is activated. Yet in many cases, one may want to forget memories that one had never thought needed to be forgotten, and these are the cases targeted by the legislation about the right to be forgotten.⁴² There are services that adopt the same approach to produce an equivalent of forgetting after the fact. They act directly on Google's search results through the multiplication of information. When a person has been publicly shamed on the web, the service produces sites laden with fictitious or irrelevant information, with the explicit purpose of pushing the sensitive information in question so far down the search results that it effectively vanishes.⁴³ For example, the

service ReputationDefender starts from the assumption that “deleting is impossible.”⁴⁴ To combat negative or undesired items about a person, it generates a wide range of unique, positive, high-quality 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 both remembering and forgetting, requires constant maintenance. Mnemotechnics work only by taking due care of and maintaining the palaces and caves of memory.⁴⁵ Memory athletes should not stop training.⁴⁶ Similarly, an effective artificial forgetting must always be renewed because Google constantly changes its algorithms and its targets.⁴⁷ 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, yet address the issue of forgetting from the perspective of information management—of how it is possible to forget 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 a search engine. When the request of a citizen is accepted, this 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 a library catalog, while at the same time preserving the books and other materials. This solution corresponds to the legislative 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."⁴⁸ To preserve the openness of the future, one would not want to lose the past. All data are still stored at the respective sites, although 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 on any of its sites outside Europe.

But there are deeper, more fundamental problems. Google's indexing, as with the catalog of a library, delivers information. The algorithm itself, however, "feeds" on data, which are much more diffuse and much more extensive than the information understood and thought by someone at some time.⁴⁹ Algorithms derive data from the information available in materials online (texts, documents, videos, blogs, files of all types), and from the information provided by users: their requests, recommendations, comments, chats. Algorithms are also able to extract data from information about information: the metadata that describe content and properties of each document, such as their title, creator, subject, description, publisher, contributors, type, format, identifier, source, language, and much more. Each of these bits of data refer to a different context than the original information, a context of which the author is usually unaware and had not explicitly intended to communicate. The Internet of Things and other forms of ambient intelligence also produce a multitude of data that individuals are not aware of, 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 which are largely independent of the intent or the original context for which they were produced, processing them to find correlations and patterns by performing calculations that the human mind could not realize nor understand, but which become

informative. Such secondary uses of data also make it possible to gain information relevant for the profiling and surveillance of citizens.

In these processes, algorithms use the “data exhaust” or the “data shadows” generated as a by-product of people’s activities on the web and, increasingly, in the world at large.⁵⁰ It is a sort of data afterlife that goes far beyond the representational quality of numbers and of information and depends on the autonomous activity of algorithms.⁵¹ Each difference makes a difference in many different ways, becoming increasingly independent from the original information. 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; however, it is also a grave threat to the freedom of self-determination of individuals and to the possibility of an open future. Information may be rendered inaccessible to indexing in accordance with the right to be forgotten, while data continue to be remembered and used by the algorithms to produce different information.⁵² Moreover, the implementation of the right to be forgotten itself involves collecting lots of metadata about which personal data is being used for what purpose. This process reveals personal preferences that, albeit anonymized, can be exploited by others for profiling.⁵³

CONCLUSION

Can one remember without forgetting? In order to remember better, is it necessary to forget less, or does the efficiency of memory depend on the ability to coordinate two different and correlated abilities, the ability to remember and the ability to forget? These questions cannot be answered without taking into account the information and communication technologies available at any given time, starting from the powerful and revolutionary tool of writing. For many centuries, increasingly refined technologies such as printing and systems for information storage had to deal first and foremost with the problem of reinforcing the ability to remember, removing from sight the related problem of the ability to forget, a problem with information use that has accompanied Western civilization since it began in ancient Greece.

Today, digital techniques bring forgetting to the forefront. The memory of our society is entrusted not only to texts and archived materials, but also to the tools that make it possible to access and distribute individual's content on the web—that is, to the algorithms that participate in communication. With their contribution, we can find, store, and access a quantity and variety of content that previously would have been unthinkable, creating a form of memory that remembers very much. This memory, however, does not seem to forget enough, unless a regulation—like the one pursued by the European Court of Justice—forces it to do so.

Finding the right balance is not easy. The attempt to create a digital form of forgetting brings out all the puzzles and paradoxes that had been latent for so many centuries: in the human form of memory, in order to reinforce forgetting, one must first remember—remember to forget. But algorithms that create the problem can help solve it. Digital tools remember so well because they work differently from human intelligence. And for the same reason they can forget differently: they can forget without remembering. Algorithms participating in communication 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. This requires a radical change in perspective. It does not solve all the problems of digital memory and of the difficulty in controlling the continuous production of an excess of data, but moves these problems to a different and much more effective level: from the reference frame of individuals to that of communication.

This is a section of [doi:10.7551/mitpress/14189.001.0001](https://doi.org/10.7551/mitpress/14189.001.0001)

Artificial Communication

How Algorithms Produce Social Intelligence

By: Elena Esposito

Citation:

Artificial Communication: How Algorithms Produce Social Intelligence

By: Elena Esposito

DOI: [10.7551/mitpress/14189.001.0001](https://doi.org/10.7551/mitpress/14189.001.0001)

ISBN (electronic): 9780262368865

Publisher: The MIT Press

Published: 2022

The open access edition of this book was made possible by generous funding and support from the MIT Libraries.



The MIT Press

© 2022 Massachusetts Institute of Technology

This work is subject to a Creative Commons CC-BY-NC-ND license.



Subject to such license, all rights are reserved.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Avenir by Jen Jackowitz. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Esposito, Elena, author.

Title: Artificial communication : how algorithms produce social intelligence / Elena Esposito.

Description: Cambridge, Massachusetts : The MIT Press, [2022] | Series: Strong ideas series | Includes bibliographical references and index.

Identifiers: LCCN 2021013271 | ISBN 9780262046664 (hardcover)

Subjects: LCSH: Telecommunication—Social aspects. | Artificial intelligence—Social aspects. | Online identities. | Social intelligence.

Classification: LCC HM851 .E765 2022 | DDC 303.48/33—dc23

LC record available at <https://lcn.loc.gov/2021013271>

10 9 8 7 6 5 4 3 2 1