



HAL
open science

Epistemic Injustice

Gloria Origgi, Serena Ciranna

► **To cite this version:**

Gloria Origgi, Serena Ciranna. Epistemic Injustice: The case of digital environments. Ian James Kidd; José Medina; Gaile Pohlhaus. The Routledge Handbook of Epistemic Injustice, Routledge, 2017, 9781315212043. hal-03937653

HAL Id: hal-03937653

<https://hal.science/hal-03937653>

Submitted on 19 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Epistemic injustice: the case of digital environments

Gloria Origgi (IJN) & Serena Ciranna (IJN)
correspondent author: Gloria Origgi gloria.origgi@gmail.com

Published in : *Routledge Handbook of Epistemic Injustice*

1. Internet as a source of epistemic injustice

In her insightful and worldly acclaimed work on epistemic injustice, Miranda Fricker argues that people can be distinctively wronged in their capacity as knowers. Much of the discussion around the notion of *epistemic injustice* has revolved around power relations between different groups of people. In this chapter we would like to take a different perspective on epistemic injustice, by applying it to the context of human/ICT interactions. New technologies may be a source of epistemic harm, by depriving people of their credibility about *themselves*. The massive gathering of big data about our own identity and behaviour creates a new asymmetry of power between algorithms and humans: algorithms are perceived today as *being better knowers of ourselves* than we are, thus weakening our entitlement to be credible about ourselves. We argue that these new cases of epistemic injustice are, under many aspects, *more centrally epistemic* than other cases described in the literature because they wrong us directly in our epistemic capacities and not only in our dignity as *knowledge givers*. The examples of epistemic harm we will discuss undermine our epistemic confidence about our self-knowledge, a kind of knowledge that has been considered for a long time as markedly different from all other kinds of knowledge because of its infallibility and self-presentness. We are dwindled as knowers, especially in the most intimate part of our epistemic competence. This is the case for both kinds of injustice that Fricker defines: *testimonial* and *hermeneutical*. But, before presenting specific case, we would like to explain why we think that the ICT examples are more centrally epistemic than other case analyses in the literature and how they may help to contribute to answer to some objections raised about the “epistemic” nature of the *injustice* committed towards knowledge givers and to illustrate in a clearer way the idea of *epistemic objectification*.

2. To what extent epistemic injustice is truly “epistemic”?

While it is very clear how people are epistemically wronged when they are denied access to knowledge on the basis of discrimination (for example, women, blacks and other groups who were

not admitted to higher education until the second half of the XX century), it is less clear—as many authors have pointed out¹—to what extent being wronged in your capacity as a *knowledge giver*, as in the cases Fricker names *testimonial injustice*, is a genuinely *epistemic* harm. As she writes: “The capacity to give knowledge to others is one side of that many-sided capacity so significant in human beings: namely the capacity for reason. So wronging someone as a giver of knowledge—by perpetrating testimonial injustice—amounts to wronging that person as a knower, as a reasoner, and thus as a human being”². But is this wrong an epistemic wrong? Who is epistemically hurt? It seems that the receiver of information, due to her biases and prejudices, will end up with less information than she would have had if she had considered the speaker at her face value instead of applying biased filters to her credibility assessment³. Thus, in a sense, the hearer inflicts to herself an epistemic offence by consciously or unconsciously deciding to underestimate the (potentially true) testimony of a certain person. By harming herself *epistemically*, she clearly harms also the speaker *morally* or *socially*, lowering her status as a less authoritative knowledge giver. However, she does not harm her *epistemically*, because if the informant knows the truth and she is not believed, she will go on knowing the truth, that is, her knowledge would not be diminished. Someone who knows who raped her in that library that night can be denied the right to be believed because of prejudices about her sex or her ethnicity, but this doesn’t change the facts: if she knows who was the raper, she will go on knowing this fact and those who are epistemically hurt are the skeptics who turned down her testimony.

The same goes for hermeneutical injustice. Fricker defines this second type of injustice in the following way: “Wherein someone has a significant area of their social experience obscured from understanding owing to prejudicial flaws in shared resources for social interpretation”. The situation here is quite different than that of testimonial injustice. It is much less “asymmetrical” : there is no inquirer and informant with asymmetrical information to share. There is a social group where everybody is “cognitively disadvantaged” for lack of an hermeneutical resource that would benefit especially one part of this social group, the less powerful, to make sense of a painful social experience. Fricker herself insists on this dimension of shared understanding: “Our shared understandings

¹ Cf. P. Engel (2016) *La fable de l'injustice épistémique*, unpublished manuscript; K. Hawley (2010) “Knowing How and Epistemic Injustice”.

² Cf. Fricker (2007), p. 44

³ This objection has been raised by G. Paulhaus (2014) when she says: “. The difficulty of detailing testimonial injustice as an intrinsic epistemic harm lies in the fact that if we take away all of the secondary harms that may happen to someone who is not believed when she ought to be and we look at the situation strictly from an epistemic point of view, we are left with at least one person who remains ignorant of some piece of information and another person who has the information that the first one lacks. So wherein lay the intrinsic epistemic harm to this second person, to the one who knows?”, p. 101.

as reflecting the perspectives of different social groups [...] Relations of unequal power can skew shared hermeneutical resources so that the powerful tend to have appropriate understandings of their experiences ready to draw on as they make sense of their social experiences, whereas the powerless are more likely to find themselves having some social experiences through a glass darkly, with at best ill-fitting meanings to draw on in the effort to render them intelligible.”⁴. The “shared understanding” of a society is shaped by power relations and may hide the domination of one group on another. Yet, the lack of “shared resources of social interpretation” affects the society as a whole, not only the victims of domination. And Fricker acknowledges this in chapter 7, when she describes sexual harassment as the central case of hermeneutical injustice: “The lack of proper understanding of women's experience of sexual harassment was a collective disadvantage more or less shared by all. Prior to the collective appreciation of sexual harassment as such, the absence of a proper understanding of what men were doing to women when they treated them like that was ex hypothesis quite general”⁵.

Women who were hurt by sexual harassment were clearly wronged, but not *epistemically* wronged as a social group, because the whole society lacked these hermeneutical resources. Fricker argues that, even if it is true that some cognitive disadvantages are shared by the whole society, only some of them are *injustices*: “ For something to be an injustice, it must be harmful but also wrongful, whether because discriminatory or because otherwise unfair. In the present example, harasser and harassee alike are cognitively handicapped by the hermeneutical lacuna—neither has a proper understanding of how he is treating her—but the harasser's cognitive disablement is not a significant disadvantage to him.”⁶

One may argue that the harasser’s cognitive disablement is a clear disadvantage to him. Take the example of the “collective harassment” which took place in Germany in the evening of the New Year 2016 in the city of Cologne⁷: the attacks on women at the city's central railway station shocked Germany and the rest of the world. About 1,000 drunk and aggressive young men were involved in various forms of harassment, rape in one case, and various robberies. The behavior of those men, mainly immigrants from North Africa, were then used as an argument against immigration in Europe by many reactionary parties in Europe. The lack of hermeneutical resources of the perpetrators of the acts was something that clearly wronged not only the victims but also themselves.

⁴ Cf. *ibidem*, p. 148

⁵ Cf. *ibidem*, p. 151

⁶ Cf. *ibidem*, p. 151

⁷ For a report on the events, cf: <http://www.bbc.com/news/world-europe-35231046>

Fricker's solution is to appeal to the work of Edward Craig (1999) and Bernard Williams (2002) and to count being treated as an informant - as a respectable epistemic agent - among the central features of the cognitive distribution of labour on which societies are based. Being an epistemic subject is also being considered as a reliable informant.

Yet, the importance of virtuous behaviour in order for knowledge to thrive in a society seems to be crucial for both parties: the society as a whole will be damaged by not considering informants at their value. The epistemic wrong seems to be distributed within the whole group, although the victims will be harmed by a special moral wrong, that is, a lack of respect and of consideration. Conceived in this way, cases of epistemic injustice need a socio-epistemological approach to knowledge to be identified.

In the case we are going to present in our contribution, victims are clearly wronged as epistemic subjects, without need to extend the essential properties of epistemic subjects to the social dimension. People are epistemically diminished as *individual* knowers: the knowledge of themselves *objectifies* them in a new way. Their identity becomes a virtual object, a "statistical double" that companies can buy and sell without people being able to have their voice heard in this process. What we are mainly concerned with here is the way in which we deal with our online identities and the possibilities that the production of these new identities create for *new kinds* of epistemic injustice, which deserve a closer analysis. Given that this may seem too big a departure from the central tenets of the ongoing debate—which focuses on the social dimension of our epistemic life and the power relations between groups—we hope to convince the reader that our "extended application" of the concept to these cases may open an interesting new line of debate and a number of relevant case studies.

2. *Online identity prejudices*

One of the major cognitive, social and epistemic revolutions of the last 20 years is the massive immersion of our cognitive life in digital environments. The transition to an Internet-mediated society of knowledge poses new challenges to epistemology. How information is searched, used, communicated and assessed on the web? What are the roles and effects of search engines, cognitive outsourcing, online monitoring and trust on our cognitive life and on our status as "competent informants"? These are very broad questions that we don't want to tackle within this short essay. Rather, we will concentrate on those aspects of our new "digital condition" that impinge upon the very idea of epistemic injustice.

Consider this case. I'm trying to find the address of a restaurant I had lunch at a few days ago. I activate my *Google Maps Timeline*, an application of Google Maps that helps you find the places you have been at and the routes you have traveled by recording all the chronology of your movements. I realise with surprise that Google has recorded a car trip I don't remember having made. I do not have a car and usually use my bicycle and public transports to move around the town. Also, Google indicates a trip out of town, and I don't have memories of a recent day trip to the countryside. I make an effort, maybe I am confused and I am just forgetting something: where was I on Sunday, whom was I with? Did we take Paul's car to go the Parc de Sceaux? Or was it two weeks ago? I feel uneasy: a sort of short-circuit is going on in my mind: I go back and forth from the information Google is giving me and the ones I have stored in my memory; I don't know which one I should trust more. In a sense, I trust the accuracy of Google reports more than I trust my own epistemic capacities of retrieving memories of the places I have visited recently. But is it infallible? And how fallible is my own introspective access to my mental states? Do I have to give up my first-person privileged access to my memory on the basis of an act of deference to the authority and infallibility of Google? I am hesitant. What would be the interest of Google to attribute to me a fake memory? I don't want to become paranoid. From an objective point of view, it is far more probable that I had forgotten about this event, that a sort of "blank" in my mind made me overlook this short trip, rather than that Google had recorded by mistake an event that occurred to someone else. Yet, this puzzles me. I know that Google is a neutral witness of my movements. It has recorded all these details about my locations along the years just because I have asked it to do so, or because I simply have not modified the *default options* of my Google account when I have started using it. In this sense, I have a reason to trust it. But the infallibility argument is a little too strong, especially when it conflicts with my own memories and the authority I have on my mental states. If I have myself hesitated in this case about whom to trust, what would this have been in the case of a third person who had to decide between Google's accuracy and the accuracy of my memory? What about a testimonial use of this situation? If that trip out of town happens to be the piece of evidence that condemns me for a crime, who should the judge trust? My word or Google's? In the context of testimony, let's say for example in a trial, this is clearly a case that opens the possibility of a huge testimonial injustice: a judge who should decide between the testimony of Google and ours would surely prefer the first one, and even use it as a piece of convicting evidence, even if we are sure we did not make that car trip. Our certainty as knowers of ourselves is undermined and our word is not taken as reliable compared to the reliability of an automatic system.

Let us consider another example. This is not a fictional illustration but a real case example we found on a blog⁸. A manager asks on a blog how to behave in the following awkward situation: while interviewing some candidates for a job, she “googled” one of the shortlisted candidates and found some naked pictures on her MySpace account. The manager asks herself and on the blog: is this girl aware of her “image” on the web? Did she make these images spread intentionally or wasn’t she aware of the fact that on MySpace they were not private? Should the manager talk to the girl about this? Is this legitimate or is it a violation of her privacy, of “her space”? It seems plausible to think that, if the candidate will be excluded from the race without further explanation, this could be perceived as unfair. The candidate would be deprived of the right to justify herself and explaining the presence of these pictures on the web. Maybe she did not put them there, someone else did. The candidate might just be unaware of her reputation on the web and also on the ways of dealing with it and withdrawing data she doesn’t want to show, or doesn’t want to show anymore. What to do?

As the examples above show, we lack knowledge of our own online identity and behavior. This lack of knowledge is due at least to two factors:

- 1) By simply being present on line, each web user generates a huge number of avatars and statistical profiles that constitute his/her online identity.
- 2) The data generated belong to various platforms with different normative frameworks in terms of the rights of the users to control these data. The users are very often unaware of the policies of the platforms and of their rights with respect to them.

This may cause actual epistemic harm because the representation and the selection of the raw data gathered in order to be intelligible to the users is done through algorithmic procedures that are determined by the owners of the platforms according to their specific interests. This results in a representation of the profile of the users that is inevitably biased and partial and does not reflect in an “objective” way what the users want to be known about themselves. More precisely, the use of big data may create a systematic *informational prejudice* against the subjects concerned. According to Hookway (2010), informational prejudice occurs when a person is prejudicially judged to lack the ability to provide information relevant in a given context and hence as being an unsuitable partici-

⁸ <http://www.askamanager.org/2012/03/yes-near-nude-shots-on-your-myspace-page-might-hurt-your-job-prospects.html>

pant in collective epistemic activity. In this case, the informational prejudice concerns the capacity of a person to provide relevant information about *herself*.

Furthermore, the ability to determine the demographic traits of individuals through algorithms and aggregation of online data has the potential downside of systematic discrimination of groups. As A 2014 research paper on Big Data states: “Approached without care, data mining can reproduce existing patterns of discrimination, inherit the prejudice of prior decision-makers, or simply reflect the widespread biases that persist in society. It can even have the perverse result of exacerbating existing inequalities by suggesting that historically disadvantaged groups actually deserve less favorable treatment.”⁹

If algorithms for data mining are seen as ways of eliminating human biases about themselves and their capacity of decision making, the way in which the data are collected may largely reflect existing prejudices within a society and have also the effect of “rigidifying” social categories by making more difficult to change them.

3. *Statistical Döppelgänger. Who owns them?*

In 2009, Google announced on its blog a revolution on the web: *personalisation*¹⁰. The content of the webpages and of the advertisements started to be adjustable to the users’ interests. In order to achieve this breakthrough result, *Google* adapted some viral marketing strategies of gathering data about the users through the “tracks” they leave spontaneously on the Web and create profiles. Google is able today to provide services that are more and more coherent with the users’ *center of interests*, because, as its CEO, Eric Schmidt, declared to the *Wall Street Journal* in 2010, “Google knows more or less who we are, what are our interests and who are our friends”¹¹. To achieve this result, Google exploits a number of *indicators*: localisation, browser, chronology of the pages visited by the user¹². The algorithmic processing of these indicators makes targeted advertisement possible. This is why gathering users’ data has become such a huge market opportunity in the digital

⁹ Barocas, S., Andrew D.S., “Big Data’s Disparate Impact”, *104 California Law Review* 671 p.674 online at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899

¹⁰ *Google Blog*, 4 December 2009, <https://googleblog.blogspot.fr/2009/12/personalized-search-for-everyone.html>

¹¹ Holman W.J. Jr, “Google and the Search for the Future. The Web icon's CEO on the mobile computing revolution, the future of newspapers, and privacy in the digital age, *Wall Street Journal*, Aug. 14, 2010

¹² Eli Parisier, *The Filter Bubble. What The Internet is hiding from you*, Penguin press 2011

economy. The *centers of interest* of the users are determined through a “profiling” process that uses as raw material the tracks that users spontaneously leave on the web. The accuracy of profiling is due to the automatic storing and archiving of a mass of data such as: browsing history, IP addresses, social network activity, email content, keywords used in searching the web. This “immaterial capital” constitutes what is called *Big Data*. Some of them are just *observed*, that is, automatically gathered by the platform which keeps them stored. Some are *communicated* to the users. Although the users’ active behaviour is the very “substance” of these data, it should be noted that the user is “passive” in the data gathering process: most information stored about himself is deduced from his actions and automatically extracted. Once the users’ tracks are gathered, the *data-mining process*, that is the interpretation and processing of the data, makes emerge a number of correlations through which the interests of the users are individuated in order to anticipate their future actions. In some cases, this process ends up with the creation of a sort of “statistical double” of the user, a sort of *alter-ego* that is modified and refined through each new action. These “predictive profiles” are the essential ingredient of online marketing strategies: from the most visited webpages, an algorithm can deduce localisation, sex, age and nationality of a user. The algorithmic profiling, as it has been shown¹³, is not infallible: false predictions are possible and an incomplete or too biased treatment of the data may end up with an ill-targeted advertisement. Although predictive profiling has been mainly used in marketing, there are other areas in which these techniques become the more and more relevant: education, risk management, insurance, health and finance. The possible uses and misuses of our statistical doubles, their correspondence to who we are and what we know about ourselves, the rights and duties we have towards them, are a new area of investigation in which considerations of epistemic injustice are central.

An example of possible epistemic injustices generated by the lack of ownership of our personal doppelgänger is the one raised by the case of a young girl who started receiving from the big retailer Target a series of targeted coupons that are usually sent to pregnant women in their second trimester of pregnancy. Her father, who discovered the advertisements in his daughter’s mail, got upset, went to Target and asked to talk to the manager, in order to clarify why they were sending that coupons to his daughter. The young girl, the father explained, was still in high school. He asked if Target was trying to encourage her to get pregnant. The manager apologized: he didn’t know why Target had sent such advertisements to the man’s daughter. When, a few days later, the manager called

¹³ « La publicité ciblée en ligne », Cnil, Communication présentée en séance plénière le 5 février 2009 Rapporteur M. Peyrat
https://www.cnil.fr/sites/default/files/typo/document/Publicite_Ciblee_rapport_VD.pdf

him to apologize again, the man said that he had finally talked to his daughter and that she had confessed to him that she was pregnant.¹⁴

Using data to predict a woman's pregnancy can provoke an epistemic harm to the targeted person by anticipating information she doesn't want to know at that time or she doesn't want to share. The level of informational disclosure about ourselves should be something we keep control of.

4. The hermeneutical injustice of algorithmic profiling

Algorithmic profiling is a source of statistical knowledge whose main objective is to obtain information about the users in order to anticipate their behaviour. The algorithmic and statistical nature of this form of knowledge makes of it a new standard of "objectivity" and "impartiality": the statistical doubles generated by this process are in a sense a more objective representation of ourselves than that we would be able to produce intentionally. It is our opinion that this claimed "objectivity" can be a source of epistemic injustice.

Let us first elaborate more about the problematic role of the users in the algorithmic profiling, a process that is made possible by one's own intentional interactions with the web. First of all, in the production of his digital double, the user is deprived of his intentionality: he is no more an intentional subject: his intentional actions are interpreted by the algorithms as traces of something else, and treated as a "symptom" of some hidden pattern of behaviour that is statistically relevant. This lack of intentionality is accompanied by a poor understanding of the terms and conditions of data sharing on a certain platform. These conditions are written in standard forms that have to be accepted in order to register to a platform, but data show that most people accept them by default without reading the details thus without a proper awareness of the various treatments of data that the platform allows for itself. Second, the user is "alienated" from the data he has furnished: he doesn't have access to them anymore, he does not know how they will be used and, in those cases in which he can access the data, it is through the applications the platform provides to visualise these data he receives, a "biased" statistical information that has been algorithmically treated in a way that is inaccessible to him.

A recent survey¹⁵ showed that 59% of the users declare that they are not able to deal with their personal data online due to lack of knowledge about the methods of data gathering and the

¹⁴ Cf. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=0

¹⁵ <http://www.serdalab.com>

rights they have to access their own data. Some of the most popular uses of algorithms, like for example the automatic selection of job applications through algorithms that use a series of parameters to “evaluate” the eligibility of the application (length, consistency, keywords, etc) is often unknown to the applicants. The use of big data by companies in order to inspect their employees is also massively unknown. Roughly, there is a tension between the *explicit profile* the users produce and think they have control over and the *predictive profile* their statistical avatar produces and companies and platforms have control over. This creates a possible ground for epistemic injustices, as the example below shows.

Take the case of Google we mentioned earlier. Google offers a variety of services that allow it to deal with both our explicit profile and our predictive one. Recently, it has created a new application, *Google Take Out*, which allows the users to “export” their data in a file. *Take Out* is a simple interface from which the users can select some of the services (*Maps, Gmail, Photos*) that have gathered data on them and obtain those data. However, the procedure is opaque: what is accessible and what is not is very unclear and the way the data are presented to the users are in a form of a new *explicit* “statistical double”, while the *predictive* statistical double is hidden by the company.

Many users feel uneasy and sometimes paranoid, to the point of withdrawing their profiles from some platforms, about some sort of possible “harm” that the use of the big data generated by their own behaviour may cause to themselves without being exactly aware of what kind of dangers they are facing and which are their options (legal, technological, behavioural) to avoid them. As we have said, this is due mainly to three factors:

- 1) They ignore the systematic data gathering by the companies and are surprised—even frightened—by some “mysterious” correlations that emerge through their interaction with the platform (like showing advertisements that are the more and more related to their interests and present concerns).
- 2) They don’t know clearly who owns their predictive algorithmic double: the platforms? the advertisement companies who buy these data from the platforms? Internet? The State? Other entities?
- 3) They are not familiar with the techniques of data mining.

The impossibility for the subject to manage his own online identity, the fears and fragilities that this lack of control generate is a form of *hermeneutical injustice*. The very fact that a category such as “*predictive online profile*” or “*predictive statistical double*” doesn’t exist and thus it cannot be the object of a debate and of an attribution to rights and duties is a hermeneutical gap in our contemporary digital societies. We feel uneasy about the fact that our data can be exploited in a way we do

not control and, on the other hand, there are no hermeneutical resources to talk about this and to develop an awareness of what are our rights not only on our own online identity, but also on our online statistical double which is, at the moment, wholly in the hands of the platforms that control it. It has also been argued that the way in which we are represented by our statistical doubles may create a “representational crisis” in the subjects of whom they are doubles.

This creates a double hermeneutical injustice:

1. The inability of most people to manage their online identities can make it difficult for them to retain hermeneutical control over their online (and, by extension, ‘real-world’) identities.
2. Many people lack the appropriate hermeneutical resources to discuss the epistemic implications of new digital realities.

In a recent article, the French sociologists Antoinette Rouvroy and Thomas Berns argue that in the digital age our form of representation is going from a “modern rationality in which we try to understand phenomena by reconstructing their causes” to a “post-modern rationality in which we don’t try to understand phenomena, just to anticipate and predict their eventuality”¹⁶. This may have a backlash on the way in which people represent themselves: the predictive statistical double is a sort of “anticipation” of our future behaviours, instead of an understanding of our past trajectory that may deeply change our autobiographical memory and our narratives of autobiographical narration.

Furthermore, the many uses of a *quantified self* in a series of successful applications (hours of sport, numbers of kilometres per day, variations of weight) can represent a new form of insight on ourselves, but also a loss of control of the subject on his self-knowledge. Here we go back to the potential epistemic harm of being considered less reliable than algorithms about ourselves that we have discussed in section 2. In terms of “hermeneutical injustice” we face the problem of dealing with a statistical double of ourselves that can harm our self-directed hermeneutical practices, either by “trivialising” them and making us feeling as part of a general category—such as “people who are underweight” or “people who have a cardiac fragility”—or by “alienating” us from the hermeneutical practices regarding ourselves that help us to make sense of who we are.

Again, the lack of the necessary hermeneutical resources in our societies to be able to talk about these new phenomena creates new forms of epistemic injustice which, in our analysis, are truly epistemic because they harm subjects in their very capacities as knowers of themselves. How can these resources be generated within our societies? Why an “online generation” as the generation

¹⁶ A. Rouvroy, T. Berns « Gouvernamentalité algorithmique et perspectives d'émancipation », *Réseaux* Vol. 31/177 (2013), p. 193

of people born after 1990 can still live with such an opacity with respect to the management of our statistical doubles generated by big data? Instead of encouraging forms of “resistance” that imply a withdrawal from the social media and other useful applications, we prefer to suggest that a policy of “transparency” and “disclosure” about the use of our statistical doubles should be reinforced by new legislations and strongly requested by citizens. This requires also the work of intellectuals and academics to elaborate new hermeneutical resources around which new forms of awareness can be raised, as we are trying to do even in this modest contribution, by putting forward and inspecting the ideas of “predictive self” and “statistical doppelgänger”.

5. Testimonial injustice online: Whom to believe?

Many online interactions are testimonial in nature (we give information to others), but often in ways that are unstructured by more usual ‘real-world’ norms and practices; online epistemic exchanges are often unregulated and unstructured; and highly prone to stereotype, prejudice, and bias; and so on. And, as we have seen in the reports on Big Data, even in the case of more regulated interactions, such as the ones with online retailers or websites of job offers, the risks of reinforcing existing societal biases instead of “rationalizing” society are many. Whereas data mining is often considered a way of eliminating human biases from the decision-making process, it may reinforce many societal prejudices and “naturalise” them in their pretended objectivity.

One of the primary objective of data-mining techniques¹⁷ is the anticipation of our behaviour. The anticipation of our behaviour is a rich market for different kinds of companies, such as companies that want to target their sales to insurance companies and other service companies. For example, many insurance companies are trying to establish partnerships with the companies that produce quantified self data in order to extract health profiles of their clients in a much more efficient way¹⁸. The potential discriminations and injustices that the use of predictive data may create in the future is something that should be socially and politically addressed. And also epistemically. The use of our predictive statistical doubles in order to infer our state of health in ten years may harm our credibility as informants about ourselves and our expectations concerning our future. If a company has to choose between trusting the results of an algorithmic profiling and trusting the words of their clients (even if based on a traditional collection of evidence) it is easy to guess what it will prefer. The testimonial injustice we are facing in the digital age is that our testimony about ourselves could be discredited in favour of a more “objective” and quantified representation of our biography.

¹⁷ Cf. C. Duhigg <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>

¹⁸ Cf. Eli Parisier, *The Filter Bubble. What The Internet is hiding from you*, Penguin press 2011

Conclusion

Much of our testimonial and hermeneutical practice is now done online. This implies that online environments constitute not only new spaces in which epistemic injustices can be perpetrated, but also they can generate and sustain new and distinctive sorts of epistemic injustice. Given the lack of control over the identities we produce by interacting on the web, the fast development of data-mining techniques to track our actions and predict our behaviours harms our capacity of making sense of ourselves and our control on our self knowledge. This results in new forms of hermeneutical and testimonial injustice. The forms of injustice we analyse here are *essentially epistemic*: they harm us in our capacity for knowledge and self-knowledge and make us less credible as witnesses of our own lives.

References

- Barocas, S., Andrew D.S. (2016) “Big Data’s Disparate Impact”, *104 California Law Review* 671.
- CNIL, (2009) [“Marketing ciblé sur Internet : vos données ont de la valeur”](http://www.cnil.fr), *www.cnil.fr* (Accessed: August 17, 2016).
- CNIL, (2009) [“La publicité ciblée en ligne”](http://www.cnil.fr) *www.cnil.fr* (Accessed: August 17, 2016).
- Craig, E., (1990) *Knowledge and the State of Nature*, Oxford: Clarendon Press.
- Duhigg, C. (2012) [“How Companies Learn Your Secrets”](http://www.nytimes.com), *The New York Times Magazine*, February 16, (August 22, 2016).
- Engel, P. (2016) *La fable de l’injustice épistémique*, unpublished manuscript.
- Fricker, M. (2007), *Epistemic injustice, power and the ethics of knowing*, New York: Oxford University press.
- Green, A., (2012) [“Should I tell a job candidate that near-nude shots on her MySpace page are hurting her prospects?”](http://www.askamanager.com), *Ask a manager* [Blog], March 6, (Accessed: August 17, 2016).
- Hawley, K. (2010) “Knowing How and Epistemic Injustice », in Bengson, J., Moffett, A., (eds) *Knowing How: Essays on Knowledge, Mind, and Action*, New York: Oxford University press.
- Holman W. Jenkins, Jr., (2010) [“Google and the Search for the Future. The Web icon's CEO on the mobile computing revolution, the future of newspapers, and privacy in the digital age”](http://www.wsj.com), *Wall Street Journal*, August 14 (Accessed: 03 August 2016).
- Hookway, C., (2010) ‘Some varieties of epistemic injustice’, *Episteme* 7: 151–163.

Horling B., Kulick M., (2009) “Personalized search for everyone”, *Google Official Blog*, December 4, (Accessed: July 29, 2016)

Origi, G. (2012) “Epistemic Injustice and Epistemic Trust” *Social Epistemology*, vol. 6, issue 2.

Pariser, E. (2011) *The Filter Bubble. What The Internet is hiding from you*, New York: Penguin press.

Pohlhaus, G.Jr., (2014) “Discerning the Primary Epistemic Harm in Cases of Testimonial Injustice” *Social Epistemology* Vol. 28, Iss. 2.

Rouvroy, A., Berns, T., (2013)« Gouvernamentalité algorithmique et perspectives d'émancipation », *Réseaux* Vol.31 177.

SERDA LAB, (2014) “Données personnelles, comprendre la nouvelle donne”, *www.serdalab.com* (Accessed: August 10, 2016)

Williams, B., (2002) *Truth and Truthfulness: An Essay in Genealogy*, Princeton University Press