Surveillance & Society

Article

Why data is not enough:
Digital traces as control of self and self-control

Tyler Reigeluth

Université Libre de Bruxelles, Belgium.
tyler.reigeluth@ulb.ac.be

Abstract

As an alternative to the seemingly natural objectivity and self-evidence of "data," this paper builds on recent francophone literature by developing a critical conceptualization of "digital traces." Underlining the materiality and discursiveness of traces allows us to understand and articulate both the technical and sociopolitical implications of digital technology. The philosophies of Gilbert Simondon and Michel Foucault give strong ontological and epistemological groundings for interpreting the relationships between technology and processes of subjectification. In this light, digital traces are framed as objects and products of heteronomous interventions, the logics of which can be traced through the programs and algorithms deployed. Through the empirical examples of "Predictive Policing" and "Quantified Self" digital traces are contrasted with the premises and dreams of Big Data. While the later claims to algorithmically correlative, predict and preempt the future by reducing it to a "what-is-to-come," the digital trace paradigm offers a new perspective on how forms of self-control and control of the self are interdependent facets of "algorithmic governmentality."

Where is the Life we have lost in living?
Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?
T.S. Eliot, in *Choruses from 'The Rock'* (1934)

Introduction

Elemental and given, data in the digital age appears to be knowledge's natural resource, its fundamental building block and its universal unit of measure. Digital technologies have opened unprecedented horizons for the production and flow of "information," increasing both their speed and volume. More and more data is left behind in our digital wake and, as its management claims dominion over information and knowledge as forms of abstraction, little of what we do in our everyday lives seems to escape the territory of digital data. With the proliferation of digital technologies and the rise of Big Data, this tendency towards a naturalization of data, endowed with inherently "objective" qualities and capable of "speaking the truth," has compressed qualitative and epistemological differences between data, information and knowledge to such a degree that we might prolong T.S. Eliot's cascade of questions by asking: "Where is the information we have lost in data?"[1] Big Data's bigger promise is that the mass aggregation of data, its

[1] I do not lay claim to this added question, several commentators have extended Eliot's interrogation, namely http://chrisbatt.wordpress.com/2010/07/27/where-is-the-wisdom-we-have-lost-in-knowledge-where-is-the-knowledge-we-have-lost-in-information-and-where-is-the-information-we-have-lost-in-data/, accessed 12/06/2013;
and http://affordance.typepad.com//mon_weblog/, accessed 15/03/2013.

Reigeluth, T. 2014. Why data is not enough: Digital traces as control of self and self-control. *Surveillance & Society* 12(2): 243-354. http://www.surveillance-and-society.org | ISSN: 1477-7487

infinitesimal dimensions and its inductive correlations, bring us closer than ever to the "real." This new self-evidence of "objectivity" should not disguise the discursive and technological processes that scaffold its construction. As the tenants of Big Data seek ever further to subsume life into data, the space and time for understanding how the latter is programed, algorithmically computed, visually represented and discursively modeled is shrinking.

Building on francophone research on the subject of "digital traces," this paper will outline the technological and discursive dimensions of this construct that make it particularly relevant and heuristic for a critical approach of governing (through) digital technologies. Given the expansive scope of both terms—"digital" and "trace" —the following reflections should be read only as early explorations of a much larger undertaking. As assemblages possessing certain intents and references, digital traces are both objects and products of interventions, the normativity of which must be objectified and approached critically. The technological determinism that has convinced us, for better or for worse, that our digital lives automatically produce data which is incapable of lying, is the blinding effect of black-boxing technology. Opening this black box implies not only a better understanding of the materiality of digital technologies, but also an unveiling of the political and economic functions which are embedded in their programing. This article[2] will show how digital traces can be conceptualized as forms of control of the self and self-control.

In search of data

In 1995 digital prophet and founder of MIT's media laboratory, Nicholas Negroponte, wrote in his *Being Digital* that, thanks to binary code, information transmission had effectively become a question of "moving bits and not atoms" (Negroponte 1995). This analogy by opposition between bits and atoms is noteworthy for two reasons. First, it implies that digital information is immaterial or that it has been dematerialized into a virtual weightlessness. Second, bits are distinguished but equated to the elemental building blocks of matter and become the fundamental informational code upon which all else emerges. This view is misleading in that it simultaneously partakes in the paradoxical—and now commonplace—abstraction and naturalization of digital information which is, on the one hand, seen as ethereal or intangible and, on the other, accepted as the "essence" of digital technology. A cursory technical exploration will give us the material ground to refute such misrepresentative metaphors and suggest alternate ways of conceptualizing digital information.

Data is one of those slippery terms, the polysemy of which requires careful manipulation. According to Christian Fauré, data has at least "four faces" —all sharing the common etymological heritage of a "given" which the subject possesses or acquires from the outside world: *i)* in a cognitive approach, data is the raw material which can be handled and calculated by the mind; *ii)* with the advent of computing and software it became the storable information that was written *for* machines; *iii)* more recently it has become the process of information transfer, of *giving* information from one component to another; and *iv)* metadata, or data on data, allows for the categorization and indexing of masses of data (Fauré 2013). Needless to say, these four dimensions are not exclusive from one another and reflect certain technical and material possibilities. The problem as to what can and cannot be considered digital data is best resolved by starting at the lowest denominator: a *bit* or binary digit. In material terms a bit corresponds to the state of a silicon[3] transistor. Schematically speaking, the transistor transforms the continuous analogue signal into a discrete digital signal which can be translated in binary values: yes/no; true/false; 0/1. Each value corresponds to a very concrete state on an electrical circuit—open or closed. With computers, logic

[2] It should be noted that this article does not aim to be a rigorous empirical analysis. Many of the examples or theoretical references used are embedded in a larger research project and are mentioned here to draw attention to French-speaking literature on the subject and gesture towards challenging perspectives.

[3] Silicon is the element which gave its name to the eponymous valley.

becomes the material order of operations, the function of which is to switch from one state to another. Today there are literally tens of millions of electrical circuits etched onto tiny silicon chips and, according to Moore's Law, every 18 months the circuit density on a chip doubles. The history of computing technologies is inextricable to that of the miniaturization and densification of transistors, processes which are eminently material, especially if one considers that the limit to Moore's law is expected to be reached in the near future. As it currently stands, this limit, beyond which the transistor could no longer fulfill its function, is unsurpassable. Furthermore, the constant densification of integrated chips has been plagued by the recurring problem of overheating. Such technical concerns have pushed IT engineers and computer scientists to search for other raw materials, different computing architectures and more efficient machines. With the ever-growing proliferation and interconnection of digital devices and the construction of massive data centers or warehouses, it becomes more and more evident that there is little that is immaterial about digital data. On the contrary, and insofar as virtually anything can be discretized on and by electronic devices, digital technologies seem to involve an intensive and extensive materialization of our social interactions and experiences through a dense technical networking of services, objects, institutions and people.

In this regard, the material determinations of digital technologies also make them profoundly malleable and essentially *programmable* (Bachimont 1999). A bit all by itself is insignificant. Only when it is syntaxically assembled and organized can it become semantically representable and interpretable. A single piece of data may be composed of thousands of bits, and while this may correspond to a certain syntaxic logic it does not mean that the piece of data has any semantic value. A single pixel can comprise up to 24 bits, but this does not make it meaningful if it is not matched with other pixels. Programing involves a process of writing *for* machines, of inscribing in their functioning certain patterns, interfaces and logics which in turn condition its user's possible interactions. If we consider that an algorithm is, basically, a cooking recipe or an itinerary involving a series of steps and directions which must be taken in order to achieve a certain output, then digital technologies require and produce particular discursive forms. As Bachimont underlines, algorithms turn "spaces of possibilities" into a necessary and verifiable temporal sequence. In other words, "The algorithm specifies that, the initial conditions [input] being met, the result [output] cannot fail to be achieved following a given complexity. The program, therefore, is a means of certifying the future by eliminating the uncertainty and improbability of the latter so as to make it controllable" (Bachimont 2008) [my own translation]. Algorithms involve a very specific temporal relationship aimed at reducing the relative indeterminacy of the future to a predictable and computational sequence of *that which is to come*. The paradox is that by black-boxing its computational complexity and reducing it to a series of inputs and outputs, algorithmic processes actually sustain a margin of unpredictability and uncertainty insofar as their increased speed, complexity, automation and adaptability generate *surprising* functions and patterns.[4] To better understand the conditions in which a large scale and generalized use of algorithmic handling of information can be deployed a few technical considerations should be emphasized.

Indeed, another compelling example of the plasticity and limitations of digital technologies is the internet, which was developed under the auspices of the Department of Defense, the MIT and AT&T in order to establish a decentralized and resilient communications network. Differentiating itself from more traditional centralized or broadcast-type networks, the internet relies on *packet switching* rather than *circuit switching*; instead of reserving a given circuit for the duration of a communication as with the telephone, the internet breaks information into packets of bits which are routed across different circuits and are reassembled upon reaching the client. In fact, the basic architecture and topology of the internet assumes that every computer is both a client and a server, in that it is capable of receiving as well as sending data. The original principle of the internet was that everyone would structurally have the

[4] This was one of the recurring debates that resurfaced between logicians and computers scientists at the recent conference in Paris on the History and Philosophy of Computing, http://hapoc2013.sciencesconf.org/resource/page/id/4.

possibility to host their own webpage and their own e-mail service, but with the advent of ADSL (*asymmetrical digital subscriber line)* and mobile "internet," upload bandwidth has been limited while unique and fixed IP addresses have been thrown to the wind, turning the everyday internet user progressively into a broadcast receiver and an information consumer.

Theoretically, the web and e-mail are only services which use the internet's structure, but they are *not* the internet. However, technical and material standards—such as ADSL—promoted by certain economic and political strategies have precipitated the rise of the Web 2.0, where "user generated contents" are actually hosted by third party servers belonging to companies that have no other business plan than monetizing their users' data. Similarly, it is worth recalling that the web's protocol—HTTP (*HyperText Transfer Protocol*)—was originally a stateless protocol, meaning that websites or web sessions had no recollections of your previous visits. The use of cookies—small text files that are stored on your navigator—makes it possible for websites and search engines to remember previous requests, entries and queries for ergonomic and practical purposes (completing a form, repetitive tasks, etc.), but also as embedded tracking techniques which aim to "personalize" and "adjust" web experiences for myriad interests and objectives. As Windley states, "Chances are, you've got hundreds of cookie-base identities that you're largely unaware of" (2005: 52). "Lightbeam," an application recently developed by Mozilla, allows its user to visualize the constellations of third party websites that potentially receive data when you visit a given page. For every ten websites actually visited there are easily another twenty that have access to some form of tracking data pertaining to your navigation habits. Unsurprisingly, the most prominent of these services are: Google-analytics, Facebook and Twitter, specialized in the statistical and algorithmic handling of users' data.

In this respect, the centralization of data, materialized by the construction of massive data centers, corresponds to concrete technical decisions, public policies and commercial norms.[5] Not only do most internet users leave their opened e-mails, their pictures, videos and blogs with perfect strangers, but the very objects that assist them in their daily tasks emit geo-localization points and other parcels of data which can be algorithmically correlated to produce individualized patterns and profiles. The point here is to emphasize that more and more of our interactions in digital "environments" are programed to produce data at unprecedented scales and speeds. This refers to the three central handling stages—distinguished here for analytical purposes, but empirically much more blurry—of "algorithmic governmentality," put forth by Antoinette Rouvroy and Thomas Berns. The first step is *data warehousing*, whereby massive quantities of data are systematically collected from a plethora of practices and interactions and stored, regardless of their utility or significance; after which comes *data mining,* effectively extracting correlations and patterns through algorithmic calculations, the inductive self-evidence of which lies in the sheer quantity of data that is handled; and finally there is *profiling,* which aims not so much at identifying specific persons, but at correlating scattered occurrences together so as to reduce behaviors themselves to necessary and temporally successive algorithms (Rouvroy and Berns 2010, 2013). The implications of algorithmic governmentality in terms of processes of subjectification and control will be detailed at greater length in the third part of this paper, but for the time being it is enough to underline that the emergence of Big Data is socio-technical in nature, and that the instruments that are used to govern shape and mold the way we govern and are governed. Understanding this relationship requires a renewed conceptual framework.

Technology and subjectification

Technology does not produce a societal change in and of itself, nor does it envelop a society in some functional essence as the term "information society" suggests. Reducing a technology to a machine's

[5] An in-depth comparative international analysis would undoubtedly help determine the relationships between national policy settings and supranational "governance" strategies.

apparent function—i.e. "digital technology is essentially concerned with information and communication, its function is stable and inherent"—blinds us to its function*ing* and leads us to neglect the social processes through which it is structured and that it in turn structures. Anchoring our understanding of technology and subjectification in a relational ontology, or an "ontogenesis" as French philosopher Gilbert Simondon (1969, 1989) powerfully suggested, gives us strong epistemological groundings for avoiding both technocentric and anthropocentric conceptualizations. Without elucidating the incredible empirical and theoretical richness of Simondon's philosophy,[6] one of his most inspiring considerations was that a technology is a reticular mode of existence in which a technical object extends itself through the gestures, dispositions, cognitive schemas and analogue experiences of its user. Technology is considered as a structuration engaged in a process of becoming—instead of an "availableness" or "ready-to-handedness" of an instrument (Heidegger) or as something having an inherently transformative logic. This turns us toward the more pressing and challenging problem of determining not only how technology governs our behaviors and interactions, but also, and more importantly, how our norms and social structures regulate the uses and functions invested in a certain technology which in turn structures our social and cognitive schemes.

Among Simondon's major philosophical contributions, undoubtedly one of the most significant is his conception of information not in substantive or content-related terms, but as a process of mediation between disparate orders of magnitude. *In*formation is the structuring relationship between matter and form. Even more fundamentally, the individual, be it technical or human, is not a term involved in a relationship but rather *is* a singular relationship and *the task of an ontogenetic theory is to follow the relationship throughout its becoming*. As an enthusiastic reader and acute appraiser of cybernetic and information theories, Simondon refuted their stabilized view of information as "message" and namely criticized the fact that cybernetics overgeneralized the feedback mechanism as the pervasive and exclusive form of regulation and control. The degree to which cybernetic representations still lend very strong, albeit latent, credence to the idea that data can speak the truth and that from its undifferentiated mass, norms can emerge magically, seamlessly, "inoffensively" (Berns 2011) will be elucidated in the third part of this paper.

Simondon's multi-dimensional and relational perspective parallels, in certain aspects, Michel Foucault's study of processes of subjectification[7] as technically mediated and historically situated practices. Indeed, if Foucault was able to consider "techniques and practices" of the self (2001, 2009) in their different historical and normative contexts, it is because he tied the ethical and the technical dimensions of the self to one another, insofar as "*Self* is evidently not, in this context, the name of an identity or a position, it is the very matter of the experimentation of the *tekhnê*—and it is also the result, the product which is constantly reworked, modified, folded to the creative logic of an endless becoming" (Revel 2009) [my own translation]. In this regard, and similarly to Simondon, the "self" is not the term used to describe a stabilized unit of existence or a demarcated identity possessing origin or teleology, but a field of relations laboring its own transformations. The self is a process of selecting, incorporating and unifying disparate and "exterior" elements; it is not an "interior" substance or some nuclear essence. More categorically, the relation to one's self is always simultaneously a relation to others, whereby the subject can become object for another and inversely. For Foucault, subjectification through self-knowledge and self-care is indivisible from forms of subjugation and dependence on others (Combes 2011). It is precisely this relationship between government of self and government of others (Foucault 2001) that Foucault qualifies as "governmentality" or as he explains, "I call 'governmentality' the encounter between techniques of domination exercised over others and techniques of the self" (quoted in Combes 2011: 57) [my own

[6] For an introduction see, for example: Hottois (1993) and Combes (1999). [English translations of Simondon remain relatively limited as of now, but translations of his major works should be forthcoming.]

[7] This cross-hatched interpretation of Simondon and Foucault is explored in Combes' (2011) *La vie inséparée. Vie et Sujet au temps de la biopolitique*.

translation]. These reflections could be framed in a single question: to what degree are techniques of the self also modes of control and government by, for and on others? Following these considerations, it is quite clear that the self is not a "given" pre-existing its objectification and that, as such, it cannot be represented and computed through bits of data.

Digital technologies hold neither the promise of a free and horizontal society, nor do they spread through society like a plague. As Simondon would point out, these two visions still posit an ontological juxtaposition between "technology" on the one hand and "society" or "culture" on the other. The determining challenge that digital technologies pose is not whether they liberate or subjugate the individual more or less than other technologies, but to develop a critical understanding of the relations and interactions programed in their functioning. Nonetheless, as digital networks expand and intensify, as data production and handling reach new levels, and as statistical models defy old models of control and modulation by moving away from populations and towards what Gilles Deleuze called *dividuals* (1990), a pervasive shadow of information is growing like a digital unconscious (Beer 2009) withholding unprecedented secrets and insights into subjects and objects. Indeed, Big Data's vaunted prospect is to unearth and discover what has never been observed by abandoning the rigid hypothetico-deductive methods of reasoning and embracing new inductive tools that rely on vast expanses of arable data. In fact, the tenants of Big Data seem to believe that their greatest feat is to have "discovered" a new ecosystem where data is naturally produced and observable. Alex Pentland—Big Data and mobile computing pioneer, World Economic Forum advisor and MIT's Human Dynamics Laboratory founder—claims that,

> While it may be useful to reason about the averages, social phenomena are really made up of millions of small transactions between individuals. There are patterns in those individual transactions that are not just averages, they're the things that are responsible for the flash crash and the Arab spring. You need to get down into these new patterns, these micro-patterns, because they don't just average out to the classical way of understanding society. We're entering a new era of social physics, where it's the details of all the particles—the you and me—that actually determine the outcome.
>
> (Pentland 2012)

Pentland's prophecy is an expressive illustration of the faith that Big Data proponents place in bits, seemingly forgetful of the fact that the data collected or "discovered" will always depend on the tools, programs and interfaces deployed, and that the information gleaned is not a given, but the trace of a relationship. When the founder of the world-wide-web, Tim Berners-Lee, proclaimed "We want raw data, now!" at a TED conference in 2009, he was asking for data to be liberated as if it naturally existed in the digital ecosystem, as if "raw data" were not a confounding oxymoron.[8] The new data economy seems to magically complete a virtuous circle, whereby more data gives more information and more information equals more knowledge. This self-fulfilling logic is not only a theoretical stance, but is equally prevalent in emerging practices of self-control, surveillance and preemption.

Digital traces

A recent strand of francophone research has emerged around the question of "digital traces" (*traces numériques*), underscoring social science's growing interest for the implications between digital technology and social transformations. This research also highlights the semantic variations and differences in approaches that divide linguistic communities. In the English language, very little academic research or media discourse refers to digital "traces." Instead the preferred lexical field seemingly revolves around "data." One term or lexical paradigm is not necessarily inherently better than another, especially considering that every language has its relevant and specific meanings for homologous words, but a

[8] See Gitelman (2013).

challenging and critical perspective of our digital interactions can be developed by analyzing information as the traces that are produced, abandoned or captured in digital environments. Before reviewing some of the literature on digital traces and framing them in light of algorithmic governmentality, a brief discussion of the term "trace" is needed.

Just like the word "data," "trace" collects numerous meanings and appendices (*to trace, track, traceable, traceability, tracing, etc…*) and seems to connote an isolated object as well as an action or a process. Serres sees at least four different denotations: *i)* a *print* indicating a physical or figurative passage; *ii)* a *mark,* a clue or a remainder left by a passed event; *iii)* an *infinitesimal quantity* as in a chemical trace substance; and *iv)* a *geometrical intersection point* (Serres 2002). According to Serres, each of these dimensions can be extrapolated into its own paradigm associated to its own field of study and theorist. Without detailing or debating this typology, its unifying trait is that a "trace" corresponds to some minute detail or seemingly insignificant fragment such as the chemist's residue, the detective's clue, the historian's indices, or the psychoanalyst's symbol. Sign, mark, document, inscription…what actually distinguishes the "trace" from this vast constellation of meanings?

The difficulty in defining the word "trace" is reflected in most academic or media discourses. As though involved in an infinite regress of meaning, a trace is used and defined interchangeably with some of its most common synonyms, effectively voiding any conceptual density or specificity it may retain. Despite the minor variations between authors, there appears to be a relative agreement that a "digital identity" is the collection or the sum of digital traces—be they written, audio or video documents, logins, online purchases, or browsing sessions—that are left behind, deliberately or unconsciously, throughout the network of a user's online relationships and exchanges (Ertzscheid 2009). Similarly, there seems to be little disagreement about the sweeping claim that "we cannot *not* leave traces" (Merzeau 2009a, 2009b), and that producing traces has become an inevitable byproduct—if not an integral part—of communicating in the "information society." As digital devices and databases hoard more and more data through the increased interconnection and interoperability of services, machines, software and institutions, the normal state of existence according to Merzeau is no longer forgetting but remembering (Merzeau 2009a). Furthermore, because our identities are collections of digital traces and because traces are involuntarily and ubiquitously produced, who we are can longer be considered in representational or subjective terms, but as objects or goods which are directly accessible and effective forms of reality (Mondoux 2011: 52). The paradox that these studies run into and have difficulty in problematizing is that while digital technologies have reified our identities they have also turned them into post-modern assemblages and constructions of traces. Following these authors it appears as though our modes of existence have shifted radically, presumably unhinging our historically structured conceptions of identity, memory and normalization. Digital transformations have indeed fashioned and configured new spatiotemporal modes of sociability and control, but the weight of such sweeping anthropological claims is unbearable for such weakly structured conceptual scaffolding. Instead, adopting the "trace" as a paradigmatic lens should actually imply seeing digital technology in continuity with "previous" or existing social, political and economic structures, and not only in terms of change, revolution or novelty.

Recent research in computer science, knowledge engineering and ergonomics has focused on the modeling and visualization of digital traces as instruments to enhance and maximize the user's computer contact. From this perspective, a digital trace has diverse sources (logs, cookies, IP addresses, browsing histories, length and time of access, etc.) and multiple objectives (adapting the user's environment, enriching the user's interactions, or assisting the user's activity), which can be articulated through a *tracing system* capable of determining the user's problematic, effective, "non-pertinent" or "redundant" machine interactions (Laflaquière 2009). According to Laflaquière, implementing a tracing system requires three steps: *i)* collecting digital interaction traces which requires sorting, cleaning and organizing the different data sources; *ii)* transforming digital interaction traces into interpretable forms by associating and synchronizing the different sources through visualization techniques, statistical tools and the user's

implication; *iii)* presenting digital interaction traces that are interactive and shareable representations of the user's activities (Laflaquière 2008: 63-68). Whether the digital interaction trace is used as a tool to increase the user's reflexivity, to personalize content or to develop automatic recommendation systems, it is obvious that the trace becomes a modeled correlation of disparate elements. As Laflaquière explains when discussing online tracing systems, "The ultimate goal is to be able to anticipate the user's browsing" and "to automatically suggest possible browsing paths" (2008: 47) [my own translation]. The correlative logic underpinning the implementation of tracing systems is to adjust the digital environment to the user—or inversely—through countless and real-time micro corrections and tunings, the ideal outcome being a perfect equation between the environment and the user. Laflaquière's research has the advantage of showing the fragility and programmability of digital traces, since they do not exist on their own and they are always the result of processes of compilation, selection, and interpretation. However, the correlative logic that links the user to the digital environment has an uncanny cybernetic ring to it, whereby homeostasis and stability of a system is achieved through the constant correlation and negative feedback of its parts. There are certain similarities between some of Laflaquière's claims and Google's CEO, Eric Schmidt's statement: **"**I actually think most people don't want Google to answer their questions, they want Google to tell them what they should be doing next." Following Bachimont's aforementioned definition of an algorithm as a space of possibilities transformed into a predictable and calculable temporal sequence, it is patent that the correlative logic underlying this particular conception of digital traces involves a relationship to time aimed at evacuating unknown elements and contingencies which could disturb the algorithm's "normal" course.[9]

It would appear that digital traces are fragments of past interactions or activities which, when correlated together, allow a preemption and prediction of future behaviors. If this logic is extended to its overarching implications, then intervening on traces displaces the focal point of normalization and regulation processes from the individual or "subject" to their environment. In this respect, compliance and conformity are achieved not through the direct imposition of coercive norms, but through the indirect modulation of the individual's stimuli and relations. The fable of "self-determination" has donned new clothes. The latent irony in this formulation of digital traces is that they are distinguished from naturally available givens in that they are programed, modeled and visualized from correlation of disparate data sources, but are concurrently supposed to produce an equivalence or homeostasis excluding any form of difference or gap between the individual and his environment. In this imagined state of self-regulation where the ways an individual *should* behave constantly adapt to how she *does* behave, the question remains, as we are about to see, remains: what potentials for subjectification are left when there is nothing left to subjectify? If the ideal individual is perfectly correlated and immanent to his environment, if her singularity can be reduced to the degree to which she fulfills these correlations, then is it actually possible for a subject to exist in its ethical and political dimensions?

Tracing control of self and self-control

It is precisely this correlative and preemptive rationality that the champions of Big Data believe they can extrapolate to all facets of "life," effectively turning individuals into algorithmic relationships. By mining beneath the subject's awareness and by quarrying with techniques that the subject has little or no grasp over, Big Data intends to bypass the interpretative and problematic moment of knowledge—"subjectivity"—by directly accessing reality on its most self-evident, undeniable and "objective" plane. To this effect, Big Data actually reifies and essentializes relationships into computable units (Rouvroy and Berns 2013), basically reducing their unpredictability, potentiality and becoming to successions of "what is to come"; thereby considering possibility as already contained in actuality (Rouvroy and Berns 2013:

[9] This depiction of algorithms needs to be nuanced by developments in heuristics and machine-learning in which algorithms produce less than optimal yet "good enough" solutions for extremely complex or large datasets. In this case the use of experience, intuitive shortcuts and "fuzzy logic" actually creates surprising or unpredictable outputs.

182) and simply in need of liberating. In this sense, "algorithmic governmentality"[10] relies on the dream that reality, if correctly probed and recorded, will reveal its own passive, inoffensive and non-coercive normativity (Rouvroy and Berns 2013), to which the individuals need only adapt as painlessly and seamlessly as possible. This implicit necessity for individuals to correspond, without gap or contrast, to their digital traces, exposes the prospect for new social stratifications and cleavages to emerge around the ability one has to be what their traces say they are. The new social outcast could very well become the one who is unable or unwilling to be "oneself," and who by denying the "objectivity" or "undeniability" of one's traces, is denying one's "self."

The political, economic and social implications of this seemingly hypothetical or conjectured algorithmic governmentality are already flourishing. Recent pilot crime-prevention programs in US and British cities have apparently turned Philip K. Dick's science-fiction story, *Minority Report*—in which law enforcement is a prescient organization that stops crimes before they can be committed—into reality.[11] Building on models used for predicting earthquake aftershocks, the real-world crime-prevention program called PredPol (or Predictive Policing), allows police to follow real-time crime patterns and establish extremely precise—although perhaps not relevant—"hot-spots" throughout the city. PredPol claims to differ from existing statistical policing software in that it does not chart or map out crime occurrence across a city and over time, but actively interacts with "databases of intentions" (Battelle 2010) and correlates disparate sources in order to produce a "sensitive prediction" aimed at legitimizing preemptive policing strategies. Eric Sadin notes that generalizing algorithmic anticipation of behavior blurs the public and private spheres of intervention more than ever, as both marketing and surveillance apply the same tools and strategies, and often cross their databases and share "best practices" (Sadin 2009). This trend is compounded by the fact that algorithms have become a market unto themselves, in which the competition and the proprietary logic appear to seriously hinder attempts to understand, criticize or re-appropriate them. Algorithmic governmentality partakes, to a certain extent, in what Andrejevic, among others, has called "digital enclosures" whereby the interoperability of digital platforms and devices relies on capturing communication in proprietary spaces (Andrejevic 2009). The more one communicates, "likes," "checks-in," "tweets," "tags," shops or listens to music, the more value is produced for the company storing "your" data and the more targeted its recommendations will be, the more precise its anticipation will be, and the more perfectly one will fulfill his or her "personalized" recommendations. One example, among myriad others: mobile telephone providers can record up to 200 geo-localization points per day from a smartphone, allowing them to effectively infer your daily habits, interests and preferences.[12] Eric Schmidt illustrates the pervasiveness of the phenomenon with his usual candor, "We know where you are. We know where you've been. We can more or less know what you're thinking about."

Governing through digital traces and tracking technologies is not only a question of intervening on the individual's environment through coercive and vertical means of control such as policing strategies or biometrics, it is also a question of learning and feeding back into the loops of the individual's self-control. Indeed, another emerging form of digital tracking is epitomized in the international platform Quantified Self, the motto of which is "Self-knowledge through numbers." QS organizes group meetings in many countries worldwide and hosts a prolific website,[13] giving its *"self-trackers"* the opportunity to share and discuss the personal problems or questions they have tried solving by recording, correlating and visualizing quantities of data over a significant period of time. Most "self-trackers" begin recording an activity or habit because there is some behavior, health issue or uncertainty they wish to elucidate or

[10] It should be noted that while the conceptual framework is fundamentally the same, my use of the term does not necessarily and entirely reflect that of Rouvroy and Berns.

[11] See: http://www.theguardian.com/science/2013/jul/01/how-algorithms-rule-world-nsa (accessed 10/11/2013); and PredPol's website: http://www.predpol.com/ (accessed 06/02/2013).

[12] See: http://www.youtube.com/watch?v=Gv7Y0W0xmYQ (accessed 02/03/2013).

[13] See: http://quantifiedself.com/ (last accessed 24/02/2013).

remedy, but continue tracking their activities because they believe that their data will express something the individual cannot. Self-tracking experiments are invariably accompanied by a device—the fork that measures the number of bites per minute you take, the gadget that records your sleep rhythms and blood pressure, the program that holds you to your promises—which is networked to a social media interface, a blog or a dedicated website, through which the self-tracker can share and visualize the results of his or her experiment. The recording, measurement, visualization and publication of one's data are inseparable elements of a self-reporting relationship to the self,[14] similar to practices of confession or avowal (Foucault 2009). There is a shared conviction among self-trackers that the devices recording their behaviors are more honest than any friend, mentor or counselor will ever be, to such a degree that the objectivity of their data can be "unbearable." The act of reporting oneself involves a discursive (re)production of traces, the veracity of which is achieved precisely because the logical and semantic compositions are masked by the speed and opacity of algorithmic computations. Again, the irony of algorithmic rationality is that it offers unprecedented possibilities for constructing formalized models of "reality," while simultaneously blurring the very processes that produce a given output or result (Bachimont 2008).

While Quantified Self's aspirations should not be overgeneralized, the rapidly growing popularity of the self-help or self-knowledge devices and applications it promotes does shed light on what I call *digital discursiveness,* the three characteristics being: *i) measurement* as the objectification of the self and criterion of "reflexivity"; *ii) reticulation* of the tracking results as a relational technique, or a relationship to the self as self-reporting; and *iii) visualization* of digital traces as the exploration and discovery of one's "interiority." Of course, the question remains as to whether these forms of "reflexivity" and "interiority" are actually articulated with the ethical and political aspects of subjectification. The age-old themes of "self-knowledge," of "looking inside" and staying "true to oneself" have not entirely vanished under the shroud of digital technology. Instead their techniques and practices are reconfiguring into other forms of (self-) control, and appear to increasingly accentuate that the subject's relative freedom is also an object of control, that the subject is a singular ethical problem at the confluence of disparate determinations. Reframing the subject in relational terms means considering her in a co-constitutive process of becoming with her environment. It is this problem, this margin of indeterminacy and becoming, which is materialized in our traces. Cardon has underlined that data is not so much studied in light of its materiality or what it does, as it is staged for the hopes and fears of larger transformations (Cardon 2012: 138). Considering data in its material and discursive dimensions is precisely what I have called a trace. The trace can, and must, be objectified, denaturalized, criticized and tracked. Indeed, traces can be analyzed in both their discursive and material dimensions, ultimately showing us that the two are inextricable. To use Simondon's language, our traces are *in*-formation; they are mediations between what we know and what an object lets us know, between what we can say and what an environment tells us. If a trace manifests an absence by providing putative evidence that someone or something was, but no longer is present and opens spaces for different interpretations and intents, then there is potentially *some trace* in anything at any moment. A trace is not an isolated object, mark or thing; it is the potentiality inherent to all becoming, actualized only through the social processes of interpretation, conflict and subjectification. The aforementioned examples illustrate how essentially the same data can be modeled, visualized and interpreted into different traces according to their intents and references.

Conclusion

This paper has developed, as synthetically as possible, a critical and multidimensional understanding of digital traces, contrasted to the technical and theoretical shortcomings of "data." Some often neglected technical aspects of digital technology and the internet enabled me to explicate some of the material

[14] See practices of Lifelogging and Lifecasting exemplified by Nicholas Feltron who publishes yearly reports of his life: http://www.feltron.com/ (last accessed 13/02/2013).

conditions of possibility in which Big Data and algorithmic governmentality have emerged. By exploring their origins, transformations and limitations, it became clear that digital networks actually intensify the materialization of communication, and that the proliferation of devices, services and datacenters depends on socio-political strategies and information structures. This fits with the larger thesis that centralization and privatization of the internet's architecture are two faces of the same phenomenon.

Furthermore, a few directions for a theoretical approach to human-technology relationships were suggested by mobilizing Gilbert Simondon and Michel Foucault. While Simondon considers technology in non-functionalist and non-essentialist terms as extended through gestures, schemas and social structures of the *milieu* within which it operates, Foucault considers technology in light of practices and techniques of government of the self and others. Yet both see a given technology extending beyond machines and artifacts and as embodied in behaviors and institutions. This philosophical grounding allowed a proper explanation of the term "algorithmic governmentality."

By considering the predictive and correlative—what I condensed as "cybernetic"—rationality of the algorithm, I reviewed two strands of francophone literature devoted to the question of digital traces. The first, mainly representative of a social science approach, frames digital identity as an assemblage of traces, which are automatically and ubiquitously produced; the second, a perspective developed in computer science and knowledge engineering, sees digital traces as ergonomic forms of reflexivity, assistance and recommendation which are intended to adjust the user to his or her digital environment. Revealing some of the inadequacies of these two outlooks, I extrapolated that a trace does not exist by itself, that it is fundamentally programmable and that it constitutes a material mediation determined by the intent and reference of its interpretation. A trace is both object and product of interventions and representations. However, it would be misguided to assume that every trace is intentional or referential—thus positing a preeminent subject who reads texts only in their contexts. Instead we must see this discursive layer as shaped by certain technological and material structures. In other words, digital technologies are discursive only if we consider their conditions of existence, their functioning and their processes of becoming. Interpretations are not conscious and independent acts of a posited subject.

Finally, I examined two contemporary examples of Big Data applications, illustrating particular normative and discursive dimensions of algorithmic governmentality. Predictive policing and "self-trackers" are two distinct but related phenomena, in that they both consider the self-evidence and objectivity of data to be the legitimizing factor for its compilation and correlation, as if data can spontaneously and irresistibly speak the truth and provide knowledge. Algorithmic governmentality moves us away from classical statistical populations towards the populations of relationships that inhabit an individual, a behavior, an imminent deed. These relationships are reduced to computable units capable of being plugged into the algorithm's syntax. The individual's singularity is thus reduced to a particular syntaxical arrangement of traces that can be represented and modulated accordingly. This is not to say that the individual can be equated to an algorithm, but that thinking, representing, intervening and governing algorithmically produces certain effects and transformations on what it means to be a "subject."

If critically conceptualized, "digital traces" offer perspectives for objectifying and distancing their supposed obviousness. Before being the material used to program the future and preempt behaviors, a trace is the uncertain, conflictive and problematic manifestations of our becoming. Understanding the trace as *in-formation*—instead of information that is already there and needs only to be adequately captured and represented—should give us insight on how and why something becomes a trace, as it focuses our attention on the mediations and operations between disparate scales of reality. This tentative exploration should be read more as a provocation to rethink our interactions with digital technology than as the presentation of a comprehensive study. Tracking the processes through which digital traces are programed, modeled and visualized calls for a critical sociology of the structural relations that determine the conceptions and functionings of digital technology.

References

Andrejevic, M. 2009. Surveillance in the Digital Enclosure. In: *The New Media of Surveillance*, eds S. Magnet and K. Gates, 18-40. New York: Routledge.

Bachimont, B. 1999. De l'hypertexte à l'hypotexte : les parcours de la mémoire documentaire. In: Mémoires de la technique et techniques de la mémoire. *Technologies, Idéologies, Pratiques*, eds C. Lenay and V. Havelange, 195-225. Erès.

Bachimont, B. 2008. Signes formels et computation numérique: entre intuition et formalisme. Available at: http://www.utc.fr/~bachimon/Publications_attachments/Bachimont.pdf (last accessed 15/01/2013).

Battelle, J. 2010. The database of intentions is far larger than I thought. Available at: http://battellemedia.com/archives/2010/03/the_database_of_intentions_is_far_larger_than_i_thought.php (last accessed 01/04/2013).

Beer, D. 2009. Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media and Society* 11: 985–1002.

Berns, T. 2011. Transparence et inoffensivité du gouvernement statistique. Available at: http://www.raison-publique.fr/article447.html (last accessed 03/10/2011).

Cardon, D. 2012. Regarder les données. *Multitudes* 49: 138-142.

Combes, M. 1999. *Simondon. Individu et collectivité*. Paris: Presses Universitaires de France, coll. Philosophies.

Combes, M. 2011. *La vie inséparée. Vie et sujet au temps de la biopolitique*. Paris: Editions Dittmar.

Deleuze, G. 1990. Post-scriptum sur les sociétés de contrôle. Available at: http://infokiosques.net/imprimersans2.php3?id_article=214 (last accessed 01/02/2013).

Ertzscheid, O. 2009. L'homme est un document comme les autres: du World Wide Web au World Life Web. *Hermès* 53: 33-40.

Fauré, C. 2013. Quatre visages des data. Available at: http://www.christian-faure.net/2013/04/18/quatre-visages-des-data/ (last accessed 18/04/2013).

Foucault, M. 2001. *L'herméneutique du sujet*. Paris: Seuil/Gallimard, Paris.

Foucault, M. 2009. *Le courage de la vérité.* Paris: Seuil/Gallimard.

Gitelman, L., ed. 2013. '*Raw data' is an oxymoron.* Cambridge, MA: MIT Press.

Hottois, G. 1993. *Simondon et la philosophie de la « culture technique »*. Bruxelles: De Boeck.

Laflaquière, J. 2009. Conception de système à base de traces numériques dans les environnements informatiques documentaires, Thèse de l'Université de Technologie de Troyes.

Merzeau, L. 2009a. Présence numérique: les médiations de l'identité. *Les Enjeux de l'information et de la communication* 2009: 79-91.

Merzeau, L. 2009b. De la surveillance à la veille. *Cités* 39: 67-80.

Mondoux, A. 2011. Identité numérique et surveillance. *Les Cahiers du numériques* 7: 49-59.

Negroponte, N. 1995. *Being Digital.* New York: Alfred A. Knopf.

Pentland, A. 2012. Reinventing society in the wake of Big Data. Available at: http://edge.org/conversation/reinventing-society-in-the-wake-of-big-data (last accessed 20/01/2013).

Revel, J. 2009. Michel Foucault: repenser la technique. *Tracés* 16: 139-149.

Rouvroy, A. and T. Berns. 2013. Gouvernementality algorithmique et perspectives d'émancipation. *Réseaux* 177: 163-196.

Rouvroy, A. and T. Berns. 2010. Le nouveau pouvoir statistique. *Multitudes* 40: 88-103.

Sadin, E. 2009. *Surveillance globale*. Paris: Flammarion-Climats.

Serres, A. 2002. Quelle(s) problématique(s) de la trace? Available at: http://archivesic.ccsd.cnrs.fr/docs/00/06/26/00/PDF/sic_00001397.pdf (last visited 13/02/2013).

Simondon, G. 1969. *Du mode d'existence des objets techniques.* Paris: Aubier-Montaigne.

Simondon, G. 1989. *L'individuation psychique et collective à la lumière des notions de Forme, Information, Potentiel et Métastabilité*. Paris: Aubier.

Windley, P. 2005. *Digital Identity.* California: O'Reilly.