

Sociology Between Research and the Society of Data

Stefania Capogna

Link Campus University, Rome, Italy

This paper focuses on how big data guide the construction of the interpretative schemata we use to understand the world and act in it. To this end, the essay describes the most significant new research elements and frontiers that sociology is obliged to address at present. While the prevailing literature advocates the need to promote data literacy, the idea we wish to advance is that it is necessary to foster the comprehension of data along with an understanding of the role and the responsibility which sociology has intrepidly assumed since its foundation, that is, the study and explication of the complexity of the relationships characterising social life always and everywhere. Our intent is to make a proactive contribution to the study of the digitality to bring to light perverse, unexpected and/or unwanted effects associated with naive use of big data for research purposes.

Keywords: sociology, big data, framing, social research, mix methods

Introduction

The paper focuses on how big data guide the construction of the interpretative schemata we use to understand the world and act in it. Due to the so-called “datafication process” (Van Dijck, 2014), made possible by the digital revolution, the power of platforms (public or private) has increased. This power descends to their ability to capture, gather and deal with unbelievable amounts of data, which permits them more and more to sway and orient the culture of measurement, ranking, forecasting and evaluation. This trend has significant repercussions on:

- (1) The policy systems, increasingly oriented towards logics of auditing and accountability (Strathern, 2000);
- (2) The economic sphere which witnesses the contest aimed at controlling data to favour their commercial applications (e.g. the so-called G.A.F.A.M set: Google, Amazon, Facebook, Apple, Microsoft);
- (3) The political sphere which does not frown upon the exploitation of social platforms to make predictions and develop targeted propaganda;
- (4) The military control aimed at maintaining supremacy and security or field applications;
- (5) The ambition to build a transparent state based on world indicators (Ruppert, 2015);
- (6) The predictive logic fuelled by illusions of control of the world, made possible by the advancement of the science of forecasting;
- (7) The generalized need to develop digital competences for lifelong learning (Castillo de Mesa & Gómez Jacinto, 2021) to move consciously in the network, enhancing the positive aspects and contrasting the negative ones (Qi, 2021).

Starting from the premises mentioned above, we shall reflect upon the danger of the framing effect of the unconscious use of these resources in social research and on the crucial role that sociology might (and should play) to counteract these risks and bring to light the perverse, unexpected and unwanted effects of naive use of big data for research purposes.

While the prevailing literature advocates the need to promote data literacy, the idea we wish to advance is that it is necessary to foster the comprehension of the trends that the data aim to outline.

This essay does not present findings from an empirical study but uses the lens of Goffman's "framing effect" (1974) to analyse current developments related to the data society.

The two crucial issues upon which this article dwells are:

- (1) The implications for research that may enable it to better understand the "framing effect";
- (2) The need to overcome opposition regarding the method with a view to accommodating the investigational prospects.

The issue that guides the reasoning underlying this article may be summed up in the following question: what specific contribution can sociology make to the study of, with, through big data?

To this end, we shall endeavour to outline the most significant elements with which the new frontiers of research and sociology are obliged to measure themselves (§ 1). We shall strive to examine the concept of "interpretative schemata" (§ 2), the implications of method (§ 3), and the concept of big data confronted with the need to create new areas of competence (§ 4).

The Sociological Perspective

Faced with the need to take part in the race towards the new digital frontier, many social scientists and the discipline itself, seem to be timidly aware of the issue and the need to preside over this field of study, which, for far too long, was seen as an aspect of other fields of knowledge. The first scientific article to coin the term "digital sociology" was that written by Wynn in 2009.

Big data cannot be conceived trivially as a numerical representation/coding of individual behaviour styles (e.g. the behaviour of consumers/users of the network) seeing that the social reality itself has been and continues to be radically transformed by digital processes, which are never neutral as they manage to radically reconfigure the "fields of action" (Lewin, 1951) and the sphere of "social action" (Weber, 1922) which orient individual and collective action.

By reifying the data which transform personal information into commodities, the individual becomes the mere object of economic-financial, political and military speculation. Following Lukács (1967), who adopted and expanded further the concept of "reification" (Marx, 1867), people are contracted with what they themselves have voluntarily and/or involuntarily produced within the network. All this gradually grows independent of people until they find themselves dominated by it thanks to autonomous laws alien to them. Unwittingly, individuals base their online activities on a mere set/exchange of scattered and disconnected data (likes, followers, stories), without the obligation to build up a relationship which implies meeting the "other than the self" (Mead, 1934), acknowledging the person inside and behind the bit by applying that necessary process of empathic recognition (Stein, 1958) which is the basis of every communicative act and prosocial behaviour (Cipriani, 2021).

When a datum is treated as an object, the subjective, contextual, relational and situational nature of its construction, management, action and interpretation is lost sight of. We also fail to see the relative "fields of

force” (Lewin, 1951) where relational dynamics are always asymmetrical and subject, therefore, to the influence of power which, as we know, is the architrave of the position that from Marx (1867), passes through the Frankfurt School (Adorno & Horkheimer, 1949; Adorno, 1966; Horkheimer, 1947), functionalism (Parsons, 1969) and Weberian sociology (Weber, 1964), to arrive at the “power of uncertainty” (Morin, 1989).

This means that we are witnessing the re-emergence of that inveterate clash between objectivism and subjectivism treated widely by the prodromes of the discipline, from its foundations with Durkheim (1895) bearer of a positivist vision of reality given his attention to facts, and Weber (1922) inspirer of a sociology which comprises the concept of “social action endowed with meaning”.

Due to the proliferation of data, social investigation is faced with a two-fold paradox.

The first relates to research itself and paves the way to two opposite scenarios. On the one hand, we find a vast quantity of data that are more or less freely accessible, or pre-packaged to facilitate reading using data visualisation tools and applicable in various ways to the methodological rigour needed to guide analysis; on the other hand, we come across a significant loss of legitimisation¹, competence and capacity for action within the overall framework of the domains of scientific knowledge that avail themselves of this type of groundwork, so that, incredulously, even unwittingly, we witness the “plunder” of the conceptual tools of sociology used and consumed by neighbouring domains of knowledge.

The second involves the field of education and the acquisition of the skills indispensable to the conduct of social research in different contexts and for diverse purposes. In a world characterised by an extremely high level of complexity, systemic and inter-systemic turbulence, requiring an increasingly greater degree of critical analysis, transversal to various sectors, roles and professions, we find ourselves depleted when it comes to the need for “sociological imagination” (Mills, 1959). In other words, the greater the demand for widespread critical thinking in all sectors (education, work, health, politics, globalisation, technology), where sociology might play a leading role, the greater the difficulty the discipline seems to encounter when seeking to acknowledge and maintain its specific scope of action and educational purpose compared to other fields of scientific knowledge. While the challenge is conducted on the terrain of the dominion of the most advanced forms of technology, the sociological perspective retains the task of bringing to light the dark side lurking behind “infatuation” with the use of data.

The Framing Effects

The concept of interpretative scheme refers commonly to “frames of meaning”, that is, to pre-fabricated meanings, standardised interpretations of situations one needs to address. These schemata are stored by the memorising individual and stored in the social repertoire which one draws up during the course of one’s entire life. The greatest contribution to the definition of the interpretative scheme was made by Goffman (1974) who reinterpreted, by combining them, some of the proposals of phenomenological sociology and field theory (Lewin, 1951). Goffman pointed out that, in everyday life, no action is possible without referring to an interpretative scheme and that the immediate action a subject performs when an event enters into his/her “field of perception” is strictly due to the interpretative scheme adopted. These perceptual schemata have a number of fundamental characteristics which were summed up in a clear manner by Gallino (2002) as: the nature of inference, that is, of hypotheses, which empirical investigation may confirm further or

¹ For the sociology of legitimisation, see Cipriani (1987).

disprove, without this becoming important at the moment when the action directed by a specific interpretative scheme is carried out; the ability to direct action when an interpretative scheme is adopted even if the hypothesis assumed were to be proven false later. In this case, recourse to a particular scheme is not confined to interpreting the reality as “other” but as the reality itself as the basis needed to perform real acts that produce tangible consequences that have an effect on concrete objects and subjects belonging to the “social field” to which the subject belongs. One such case might be “virtual violence”, the cause of considerable social alarm today; the coherence, stability and resilience deriving from the interpretative scheme when it comes into positive contact with the subculture, the identity or the social status of the subject and/or with his experience. This way, by asserting itself as a “horizon of common sense” it represents an effective factor of communication for those who share the same scheme, while it represents an obstacle for those who oppose it. This is the case of the so-called “echo-chambers”; the motivating function that an interpretative scheme performs *a priori* representing *a posteriori* a form of rationalisation of the act carried out bestowing a characteristic of guilt-ridden legitimisation or justification on it. It often happens, however, that when the act carried out cannot be legitimised by the interpretative scheme adopted, because it is not acceptable according to the canons of the dominant local morality, one recurs to completely opposite frames of reference.

Referring again to Goffman (1974), it is possible to identify three types of priority interpretative schemata: the primary scheme which appears original and independent of any other previous or primary interpretation; the codified scheme based on material previously endowed with meaning based on a primary scheme, transformed consciously and consensually by the actors involved into a new situation. One example of this type of scheme might be the liturgical ceremony or situations consciously considered a joke, a game, which symbolically simulates another situation. An example of this type might be that of one of the latest online challenges associated with the *Squid Game* series; finally, those created to be purposely deceptive. In similar circumstances, an *ad hoc* interpretative scheme is adopted and transmitted to deceive someone, as is the case of “fake news”.

The interpretative scheme is not comparable conceptually to the definition of the situation introduced by Thomas and Znaniecki to explain the complete organisation of a person's life through processes used by the subject to gradually construct a coherent image of the world based on meanings mediated by the reference group/groups and progressively structured according to personal needs and means of control. With the extension of online life, the process of defining the situation is increasingly oriented by the type and variety of groups and channels frequented, contributing to feeding the aforementioned “echo chamber”. The interpretative scheme is a micro-sociological process that, at individual level, emerges and expires very rapidly, often in the handful of seconds it takes to interpret and respond to a given situation; the two authors define a situation as a macro-sociological process that develops over a long time span (years, decades) following a circular route of encounters-clashes between the introduction of new data sets into the sphere of individual experience and the definition of new situations within these sets (Thomas & Znaniecki, 1968, p. 542).

Based on this thesis, aimed at understanding the “framing action” carried out by the proliferation of big data and digital communication, in the following sections, we shall focus on the methodological implications for social research and on the usefulness of reinterpreting the digital space-world in the light of conceptual sociological tools.

The Implication of the Method

In the past, social surveys had the need to collect data. Today they are confronted with an excess of pre-packaged data, which nourish new development bases.

At the same time, sociological research has to deal with the hegemony of Western scientific knowledge, conveyed mainly through the English language. This dominance is due mainly to the emergence of English as the *lingua franca* of the world of science and to the current research-evaluation model powered by open science technological platforms, the most important of which, at a global level, are Anglophone, something which permits them to colonise the scientific debate (Zincke, 2014).

Among the most relevant difficulties introduced by big data and by the ease of with which it is possible to gain access to/use these information resources, we shall list some which affect social research in particular, especially the sector of education.

(1) The concept of “representativeness” is the basis of any serious research to explain social phenomena. The lack of representativeness, statistically speaking, does not hamper the conduct of meaningful analyses of phenomena that may assume a thematic value, investigate original and/or emerging research pathways, bring hidden needs to light, penetrate covert dynamics requiring alternative approaches. The kind of social research that favours the use of digital infrastructures clashes, first of all, with the issue of the digital divide that we are well aware of, has an impact at different levels and creates a plethora of people who, by being excluded/remaining invisible cannot be reached using online research.

(2) The concept of “reliability”, strictly linked to the issue of “veridicity”, seeing that big data exploit the following two main channels.

(a) Self-production is linked to movements carried out on the web, given the traceability guaranteed by the digital medium and using which our every action, consumption, question, purchase, consultation leaves unambiguous “trace” (Gray & Gómez-Barris, 2010). Because our digital presence is increasingly subtle and invisible, we are not generally aware of how pervasive the digital in our daily lives.

(b) The tendency people have to knowingly/unknowingly or necessarily transfer personal data when accessing both public and private sites and/or resources (e.g. cookies, management of privacy, likes).

(3) Another element associated with the truthfulness of the data is the “arbitrariness” that characterises the provision of personal information, which can be easily falsified. However, even when the issue of voluntary falsification is excluded, when information is released voluntarily it is evident that the data are filtered according to what the subject wishes to reveal and share at that time, that is when providing an image of him/herself. This contributes to orienting the relational dynamics in the virtual/real counterpoising. It is no coincidence that we speak of the society of images.

(4) A further essential element is the “reproducibility of results”, a fundamental prerequisite of the scientific method (Popper, 2002). All digital platforms/resources/technology needs to be seen as social artefacts, expressions, therefore, of a specific worldview. As such, they are never neutral, but a kind of “black box” always based on an asymmetrical relationship because the user, even when she/he is a researcher, is not acquainted with, has no access to the logic, objectives, constraints, and the different rationalities determine certain choices. One is constantly faced with an unavoidable degree of opacity that clashes with the cardinal principle of reproducibility and control, along the entire chain of theoretical-methodological choices that are the basis of data collection. In other words, the new “worlds of data” pose an irremediable problem regarding the

penetration and understanding of the deep-seated reasons that inform and contribute to their acquisition and diffusion.

(5) Within this centrifugal vortex, social research, even concerning education, is losing its control over the format of data construction processes, over possible distortions and/or other sources of error regarding their production. Social researchers have no control over the data chain; they do not act as designers of the tools used to collect data (Salganik, 2018). The operational transformation of constructs into variables is no longer in their hands as used to be during the conduct of classical surveys. Engineers now manage it, computer scientists, statisticians and other professionals involved in daily practices of hardware implementation, even determined by sensors and programmes (apps, algorithms) designed to feed detection platforms automatically.

Finally, the issue of validity introduces the relationship between theory and empiricism, expressed through coherence between a concept and the ability of an instrument to measure what it is meant to detect. Criticism claims, generally, that the development of research methods based on big data orients data-driven rather than theory-driven research and runs the risk of working on superficial information, without sufficient awareness of the influence determined by the context, the environment, by the data-generation tool itself, by the specific fields of action within which it is constructed.

Desrosières (2016) explored the topic holding that it was necessary to distinguish between measurement and quantification. In his opinion, quantification comes before all else and provides the defining process capable of leading to the development of univocal, standard concepts, the development of procedures of classification and measurement. The defining process develops along with a continuous negotiation, coordination and critical review involving different actors from diverse disciplinary and professional fields and leading to the construction of a system of shared theoretical-methodological conventions. This means that quantification is anchored in a specific logic of justification that legitimises the detection and measurement system used to evaluate a given phenomenon. This conventional logic is guided by principles of value capable of orienting the different perspectives through which it is possible to problematise the social world. From this, it is possible to deduce that the measurement system incorporated in a digital platform/device reproduces this perspective, and in a novel fashion that false consciousness Adorno (1956) wrote about. Considering measurement and quantification, this way permits us to penetrate the “black-box” of big data in order to reflect on:

- (1) What lies behind systems of measurement (Salais, 2016);
- (2) The processes and rational elements that underscore the construction of automatic survey systems;
- (3) The logic of open data;
- (4) The internal coherence of theoretical-methodological structures;
- (5) The methods for the communication and consultation of data;
- (6) The limits and approximations connected to the numerical language just as to the alphabetic language (Russell, 1927).

Moving on to a different order of considerations, we should remember that, in general, driven by the pressure of rapidity, economy, and the possibilities provided by the power of ITC, research with and through digital platforms tends to focus on its structural characteristics, assuming, for example “that the structure of the network as such practically determines the action” (Uzzi, 1997, p. 63). This proposition often underpins research regarding career advancement and business opportunity networks, starting by attributing an advantage to those “closely connected in social networks” (Collins, 1998, pp. 44-45). However, this type of perspective

expresses an overdetermined vision that does not take into account the transformative and emancipatory powers inherent in subjective action, by which we mean the ability of a person to operate consciously in the world, to achieve goals oriented to personal standards and values, to operate actively and transformatively in the reference context, following a design anchored to that person's own life project (Bandura, 2000). By overcoming a naive concept of using these resources, one would be helped to understand the interpretative, negotiation and implementation perspective of the actors involved. These different rationalities inform the processes and their implications when it comes to collective action and the common good.

Although the results of the processes which generate the conventions of use to research maybe more or less explicit and/or accessible employing codes or methodological guidelines, Salais (2016, p. 119) recalled that most of the processes of negotiation and definition remain invisible during all the subsequent phases and tend to be forgotten as the platform/device becomes independent of its creators/programmers. This way, the instruments gradually assume power of their own, are considered non-contestable, non-modifiable (because the fundamentals of and/or the reasons for the initial choices underscoring the process are lost, causing inevitable drifts during the employment of subsequent applications). The domestication process (Jedlowski, 2005) does the rest, transforming them into a habit, making users forget their invisible prescriptive strength, normalising their presence and use in daily practice.

Some authors (Thévenot, 1983; Diaz-Bone & Didier, 2016) had introduced the notion of the “statistical chain” to describe how different actors and situations are involved in the process leading to the proliferation of these IT resources, to show how datafication and quantification can lead to unsound data when even one of the actors involved changes a convention along the chain. Conventions are relevant to guarantee methodological consistency and, if properly devised, communicated and observed, can represent act as a guarantee to the principle of transparency which should inform democratic society and all kinds of public responsibility.

An ulterior interpretative key (Van Dijck, 2014; Gray, Gerlitz, & Bounegru, 2018) shifts the emphasis to the role played by data-collection infrastructures to throw light on their function as an information/communication/strategic coordinating interface between different spheres (political, bureaucratic, state; economic, managerial) and daily epistemic practices. These infrastructures are conventional, institutionalised resources that guide and organise the datafication of social phenomena. One example is third-mission evaluation systems applied to schools, universities and research. Concerning the concept of framing illustrated in the previous paragraph, an essential feature of this kind of infrastructure is that it provides pre-configured frameworks, the so-called dashboards and/or graphic representations, useful because they provide a portrait of the reality capable of informing policy and/or strategic choices. These frameworks evolve through negotiation, engagement, and mediation processes between the interests of those who help design digital infrastructures. Through these processes of construction of knowledge (Berger & Luckmann, 1966), preordained schemes and criteria are elaborated for the interpretation and evaluation of the information made available, slowly becoming tacit beliefs and incorporated into the same digital processes and artefacts. This way, the more or less public and/or accessible systems of infrastructures are “objectified”, constituted as “data worlds”, that is, specific and complex sets of interactions between actors, institutions and practices that impose themselves on the interpretation of the reality and of organisational and social processes². They impact

² On this issue, see also Diaz-Bone, Horvath, & Cappel (2020).

prescriptively as a force that binds and slowly becomes autonomous. Seeking to deal with this state of affairs, some authors speak of the “crisis” of sociology (Frade, 2016; Savage & Burrows, 2007).

Within this ambit, myriad ideas have emerged to denounce the fact that:

(1) The main institutional context of innovation and methodological competence has moved from the context of the academic social sciences to the world of private companies (Van Dijck, 2014; Zuboff, 2015), with evident distortions regarding reasons for sociological investigation;

(2) The role of quantitative data and information has ceased to be that of informing (national) policies and fostering an understanding of social phenomena and transitions to use numbers as benchmarks; to individualise “surveillance” and produce models of prognostication (Zuboff, 2015) and the assessment of evaluation of productivity;

(3) Ownership of data and responsibility for the definition and development of research have moved from the public to the private sector, which is tantamount to the subjection of the political-institutional system to that dictated by an economic-financial and/or technocratic speculative order (Srnicek, 2017);

(4) It is necessary to enter into this “black box” which requires a limited sometimes incorrect and ethically disturbing type of knowledge (Kitchin, 2014; Kitchin & McArdle, 2016);

(5) Big data pose unavoidable, new methodological challenges to sociology (DiMaggio, 2015; Lee & Martin, 2015; Marres & Gerlitz, 2016; Williams, Burnap, & Sloan, 2017).

Although many of the latest techniques for dealing with digital data are evolving, no established protocol exists as yet to inform us how avail ourselves of this new wealth of sources to promote sociological research.

Defining Big Data

Although it is not possible to identify who coined the word “big data”, its birth dates back to around 1990 and John R. Mashey, an employee of Silicon Graphics. However, the diffusion of the term stems from an analysis carried out by the McKinsey Global Institute (2011), which was the first to illustrate the potential of big data when applied to business.

Big data introduced many significant new elements that immediately posed essential challenges to research in all fields compared to previous period.

It is possible to pinpoint three essential developmental phases.

The first phase (1963-2000) was influenced by rapid development of the statistical sciences aimed at orienting the development of Relational Database Management Systems (RDBMS) to store, extract and optimise data.

The second phase began in 2000s, with the advent of the Internet, the spread of the web on a large scale and the consequent possibility of collecting information in a completely new way.

The current third phase, defined as “platform society” (Van Dijck, Poell, & de Waal, 2018), contemplates, alongside the enormous developments of all digital facets, the capitalization of web-content based, to recover/monetize the information. Over the past few years, we have witnessed a flourishing of unprecedented connections and collaborations which strive to foster cross-cutting collaboration within these fields. The most significant impact has been that based on the triangulation between avant-garde statistical methods, the development of the computational sciences and innovative theories within the practical field of applications (King, 2016). However, what appears extremely timid is sociology’s contribution, which is asked to comprehend the distinctive theoretical-empirical contribution within this renewed, complex framework.

Having decided to dwell on big data, we need to define them briefly. One definition that is generally shared is “disorderly data”. This alone would suffice to undermine the social sciences and sociological research, shake them at their very foundations, seeing that they explore “orderly”, clearly defined, punctually controlled and reproducible data formats as required by the “scientific method” (Popper, 2002). Big data are commonly defined in terms of the so-called 3Vs (Laney, 2001):

(1) Volume as they are capable of collecting huge amounts of data in real time and without any human intervention;

(2) Variety because they come from diversified sources (platforms) that do not usually intercommunicate;

(3) Velocity seeing that they self-produce in different ways and modalities, at a whirlwind pace.

However, the use of big data poses a number of problems in terms of (Marr, 2014):

(1) Veracity that questions their use by us when seeking to inform a survey correctly, considering that they are subject to different types of alteration;

(2) Value since, as was the case when mines were discovered during the race for colonial and political-economic-military exploitation, data permit the possibility to capitalize their value through processes of business analysis;

(3) Variability, meaning that these data fluctuate as trends differ like those determined by a historical period, by seasonality, by the fact that they come from different sources that need to be linked, “cleansed”, transformed, contextualised (McNulty, 2014), “historicised” as Weber (1964) explained when seeking to clarify the thin line that unites and distinguishes between historical sociology and sociological history.

Finally, big data are characterized by the fact that they are:

(1) Unique because each micro-datum is uniquely identifiable (Dodge & Kitchin, 2005), thanks to the fact that digitality guarantees traceability. This element is enlarging the frontier of exploitation, further giving rise to new phenomena (markets), for example, that of the cryptocurrencies, bitcoins;

(2) Relational (Boyd & Crawford, 2012) as they can be linked to data from other sources;

(3) Extensible in that information can be easily added and/or changed;

(4) Scalable since, thanks to their structure, they can be expanded rapidly and without any additional cost (Marz & Warren, 2012).

This state of affairs invites sociology to intervene more decisively on these issues and to make that indispensable and distinctive contribution to the understanding of data and the formation of adequate skills. Skills that go beyond the logic of mere data literacy with which one intends the ability to read, understand, create and communicate data, to prioritise thinking about their understanding.

The prevailing debate tends to explain digital literacy as the outcome of a combination of the following three areas of competence as suggested by the Advisory Group on Data Revolution for Sustainable Development:

(1) Information literacy means being able to recognise a need for information, the ability to seek, evaluate, use information in a conscious way in order to create new knowledge;

(2) Technical skills meaning set of skills or technical knowledge used to perform practical tasks in the fields of science, the arts, technology, engineering, mathematics, the so-called hard sciences;

(3) And statistical skills meaning the ability to know how to manipulate, organise, explore, work with data and communicate them.

In this perspective, the social view is missing. If the analysis and management of similar data were merely technical and technological, this short-sighted interpretation would mean losing sight of the human and humanistic dimension underlying the language of numbers and failing to see the symbolic violence (Bourdieu & Passeron, 1977; Bourdieu, 2003) which often lurks within simplistic interpretations. The new “worlds of data”, outlined here, question the dominant canon along with reductionist perspectives, and sociology is duty-bound to bring to light the social implications concealed behind each number.

As Simon explained in 1969 “in a society rich in information [...] something must be missing: this something is attention”, and attention requires time, in-depth dialectical action within the framework of an inter-and-multi-disciplinary perspective the only one capable of leading to conscious work with and of data.

This lack of this kind of indispensable interpretation of relational dynamics on the part of the prevalent debate that focuses on the centrality of data literacy skills, proper to the epistemological frameworks that inform the discipline, confirms the fact that sociology has been timid and tardy to participate in the debate and intervene in the field of application. While, more and more, big data assume the value of “knowing capitalism” for the social sciences, they remain an open question (Goldthorpe, 2016, pp. 80-81).

Conclusions

To sum up the reasoning carried out so far, we will try to rethink the contribution of sociology in the complexity outlined to reduce the framing effect and/or simplistic interpretations.

Applying a sociological perspective to the analysis of big data means maintaining a critical view of their informative capacity and the platforms they use to collect and diffuse data. Moreover, any analysis, of whatever nature it may be, which seeks to base itself on these sources, needs to avail itself of pragmatic, non-naïve, lateral thinking of it to understand the possibilities and limits of big data.

Some scholars (Gehl, 2015) propose avoiding any a-critical attitude towards the results of similar surveys and examining the sociotechnical processes involved along the “data construction chain”. Dutiful attention needs to be paid, especially to systems used to historicise data through real-time, self-powered updates, that many use to inform their understanding of the world and the lines of action they adopt. As mentioned previously, historicization and comparison are among the cornerstones of Weberian sociology (Weber, 1964) to understand the impact of events upon the tensions existing between their subjective thrust and historicization.

In other words, this means identifying biases, blind spots, lack of information; it also involves adopting a situational approach aimed at contextualizing the process of affirmation and development of the data and being willing to explore meaning, to think creatively about how to deal with this epochal and paradigmatic change in the methodology of social research.

This requires that the researcher be able to intervene to cleanse the data, remove any semantic ambiguity, correct possible distortions, clarify its interpretation and endeavour to problematise what remains in the background, that is, what lies behind, before and within such easily accessible data.

Suppose we assume that sociology involves analysing and understanding social phenomena and that as social research, it is a combination of theory and empiricism that sees no fractures and clashes. We can acknowledge the principle of circularity between these two dimensions of sociological action as it seeks to link the particular to the general and explain the individual using universal principles. This effort is represented emblematically by the famous book on *Suicide* (Durkheim, 1897), which many acknowledge as the cornerstone of sociology.

Social research should act by applying a symphonic approach (Halford & Savage, 2017) to the conscious and iterative assemblage of data, method and theory, description and interpretation, in a circular mode where theory and empiricism, description and interpretation interpenetrate and reinforce each other. This means favouring a cross-fertilisation between different epistemic apparatuses where sociology can play a role of driving force in the diffusion of a new type of “data culture”³ and in the diffusion of “sociology of data”⁴ capable of informing processes, policies and choices of collective interest at all levels to have an complete comprehension of what lies behind, within and beyond the data. This shifts the focus away from the issue of access to information/content.

The recent emergence of digital data as a new source for understanding the world we find online is promising but analysing it only is insufficient to satisfy the intent of sociological research to explore broader aspects of society and how behaviour styles have changed as a result of the affirmation of the digital society (Lupton, 2015). To analyse this complexity, it has become necessary to deal with inhomogeneous types of data and engage in projects of research based on an integrated perspective. So, mixed methods have emerged as a valid alternative and provided an opportunity for sociologists. In traditional sociology, quantitative and qualitative methodologies differed in the value attributed to research orientation. As Hamberg explains (Hamberg, Johansson, Lindgren, & Westman, 1994, p. 178), quantitative research focused on generalising hypotheses by examining their internal validity, while qualitative research focused on reliability and credibility to produce transferability. The two perspectives were not necessarily conflictual, nor did they exclude each other though their relationship was not one of reciprocal substitution. However, communication between these two methodologies has been lacking because they have been divided into “two sub-cultures”. The challenge of big data makes it necessary to bridge this gap in favour of research and training pathways capable of promoting the adoption of integrated research methodologies. We can apply the symphony metaphor where a variety of tempos, methods and instruments become its distinctive, invaluable characteristic.

According to Creswell and Plano Clark (2017, p. 5), the principles of the mixed methods approach should follow four steps: rigorous collection and analysis of qualitative and quantitative data in response; integration, mixing or combination of the two data forms and their results; organisation of these procedures into specific research designs; and placing these processes within a precise epistemological and methodological framework.

In conclusion, some critical issues of particular importance to the discipline can be recalled.

The question arises of the data’s uses and usefulness, relative access to the sources they generate, and issues of cost and ownership. Unlike the engineering and natural sciences and the many opportunities they have of collaborating with large private companies, academic and social research has very few, particularly when it comes to education and the acquisition of funds to support dedicated research infrastructure. Similarly, few public initiatives and state organisations support the new exigencies of social research to build data structures comparable to those provided by digital platforms the way they do other highly profitable fields.

The issue arises, therefore, of the ability sociological research may possess to inform policies aimed at avoiding/reducing the polarisation of phenomena that broaden previous inequality gaps and enlarge the area of social injustice while, at the same time, guiding the development of new social models bent on the institution of a new type of humanism and conscious, sustainable use of technology.

³ For an in-depth analysis of data culture, information and infrastructures, see, among others, Edwards et al. (2011); Bowker & Star (2000).

⁴ On this issue, see, among others, Iliadis & Russo (2016); Zhu et al. (2021).

This state of affairs generates at least two perverse effects (Boudon, 1977) upon different levels, first of all:

(1) The propensity to modify the institutional mandate of those charged with managerial tasks, paying greater attention to society on the whole and to the institutions charged with the presentation of data, making them framing tools of public legitimisation and functional to self-approval and the acquisition of more resources;

(2) The risk to sway the times, ways and opportunities available to those wishing to carry out social research in this area to the point of determining the themes deemed accessible and prevalent, outlining a sort of research environment designed to guide those who investigate these fields by pre-constructing the very possibilities of doing research and querying the data. In this way, the platform society (Srniczek, 2017) comes to address the narrative and the very way in which this narrative is to be achieved.

Today more than ever, in the face of the complexity-generated risks, sociology is called upon to provide valuable tools facilitating analysis and understanding of the reality, starting from the need to overcome the clash between sociology and applied research, in order to explore new frontiers of investigation capable of surpassing present-day fragmentation. This is possible only by triangulation between theoretical perspectives, methodologies, and tools such as enhancing the distinctive contribution sociology can make to understanding and managing today's changes.

Decades of ideological clashes between methods, quantitative and qualitative data, systems and actors, objects and subjects, the macro and the micro, theory and practice, deduction and induction, academia and practice, have betrayed and impoverished the specific mission of sociology, that of interpreting the social reality and, possibly, of helping to understand it. Today, sociology should assume the task of inducing those who embrace the dominant view of the use/application of big data to the real world (with its predictive, "a-theoretical" "infatuation" and tendency to use algorithms in an affirmative, a-critical way) to reflect upon ethical issues, promote exploratory research, reflective, critical methodologies which seek to bring the hidden side of data to light. Agree with Kim (2019) and Lyle (2017), the challenge for sociology is to rethink its identity as a holistic science which means making itself available to the contamination of the increasingly well-established science of data to favour contamination of perspectives and methods such as to permit it to play a central role in the debate regarding these issues.

Funding

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

References

- Adorno, T. W. (1956). *Lezioni di sociologia*. Turin: Einaudi.
- Adorno, T. W. (1966). *Dialettica negativa* (trad. id.). Turin: Einaudi.
- Adorno, T. W., & Horkheimer, M. (1949). *Dialettica dell'illuminismo*. Turin: Einaudi.
- Bandura, A. (2000). Self-efficacy: The foundation of agency. In W. J. Perrig and A. Grob (Eds.), *Control of human behavior, mental processes, and consciousness: Essays in honor of the 60th birthday of August Flammer* (pp. 17-34). Mahwah, NJ, USA: Lawrence Erlbaum Associates.
- Berger, P. L., & Luckmann, T. (1966). *La realtà come costruzione sociale*. Milano: Il Mulino.
- Boudon, R. (1977). *Effets pervers et ordre social*. Paris: Presses Universitaires de France.
- Bourdieu, P. (2003). Sul potere simbolico. In Boschetti, A. (2003). *La rivoluzione simbolica di Pierre Bourdieu* (pp. 119-129). Venice: Marsilio [*Sur le pouvoir symbolique*, Annales, n. 3, maggio-giugno 1977, pp. 405-411].
- Bourdieu, P., & Passeron, J.-C. (1977). *Reproduction in education, society and culture*. London: Sage.

- Bowker, G. C., & Star, S. L. (2000). *Sorting things out. Classification and its consequences*. Cambridge: MIT Press.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication and Society*, 15(5), 662-679.
- Castillo de Mesa, J., & Gómez Jacinto, L. (2021). Digital competences and skills as key factors between connectedness and tolerance to diversity on social networking sites: Case study of social work graduates on Facebook. *Current Sociology*, pp. 1-17.
- Cipriani, R. (1987). The sociology of legitimation: An introduction. *Current Sociology*, 35(2), 1-20.
- Cipriani, R. (2021). The other, altruism and empathy. Variety of prosocial behaviour. *The American Sociologist*, 09 June 2021.
- Collins, R. (1998). *The sociology of philosophies*. Cambridge, MA: Harvard University Press.
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research*. Los Angeles: Sage Publications.
- Desrosières, A. (2016). The quantification of the social sciences. An historical comparison. In I. Bruno, F. Jany-Catrice, and B. Touchelay (Eds.), *The social sciences of quantification* (pp. 183-204). Cham: Springer.
- Diaz-Bone, R., & Didier, E. (2016). The sociology of quantification—Perspectives on an emerging field in the social sciences. *Historical Social Research*, 41(2), 7-26.
- Diaz-Bone, R., Horvath, K., & Cappel, V. (2020). Social research in times of big data. The challenges of new data worlds and the need for a sociology of social research. *Historical Social Research*, 45(3), 314-341.
- DiMaggio, P. (2015). Adapting computational text analysis to social science (and vice versa). *Big Data & Society*, 2(2), 1-5.
- Dodge, M., & Kitchin, R. (2005). Codes of life: Identification codes and the machine-readable world. *Environment and Planning D: Society and Space*, 23(6), 851-881.
- Durkheim, E. (1895). *Le regole del metodo sociologico*. Turin: Einaudi (2008).
- Durkheim, E. (1897). *Il suicidio*. Studio di sociologia. Milano: Rizzoli (2009).
- Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science friction: Data, metadata and collaboration. *Social Studies of Science*, 41(5), 667-690.
- Frade, C. (2016). Social theory and the politics of big data and method. *Sociology*, 50(5), 863-877.
- Gallino, L. (2002). *Dizionario di sociologia* (XI edizione). Turin: UTET.
- Gehl, R. (2015). Critical reverse engineering: The case of Twitter and TalkOpen. In G. Langlois, J. Redden and G. Elmer (Eds.), *Compromised data: From social media to big data*. London: Bloomsbury.
- Goffman, E. (1974). *Frame analysis. An essay on the organization of experience*. New York: Harper and Row.
- Goldthorpe, J. (2016). *Sociology as a population science*. Cambridge: Cambridge University Press.
- Gray, H., & Gómez-Barris, M. (2010). *Toward a sociology of the trace*. Minneapolis: University of Minnesota Press.
- Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, 5(2), 1-13.
- Halford, S., & Savage, M. (2017). Speaking sociologically with big data: Symphonic social science and the future for big data research. *Sociology*, 51(6), 1132-1148.
- Hamberg, K., Johansson, E., Lindgren, G., & Westman, G. (1994). Scientific rigour in qualitative research: Examples from a study of women's health in family practice. *Family Practice*, 11(2), 176-181.
- Horkheimer, M. (1947). *Eclisse della ragione*. Turin: Einaudi.
- Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, 3(2), 1-7.
- Jedlowski, P. (2005). *Un giorno dopo l'altro. La vita quotidiana tra esperienza e routine*. Milan: Il Mulino.
- Kim, Y. (2019). New opportunities for sociological research: A discussion of the usefulness of mixed methods with data science. *Journal of Asian Sociology*, 48(3), 343-358.
- King, G. (2016). Preface: Big data is not about the data! In R. M. Alvarez (Ed.), *Computational social science: Discovery and prediction* (pp. 1-4). Cambridge: Cambridge University Press.
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. London: SAGE.
- Kitchin, R., & McArdle, G. (2016). What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1) 1-10.
- Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. META Group Research Note, 6.
- Lee, M., & Martin, J. L. (2015). Surfeit and surface. *Big Data & Society*, 2(2), 1-3.
- Lewin, K. (1951). *Field theory in social science*. New York: Harper & Brothers.
- Lukács, G. (1967). *Storia e coscienza di classe*. Milan: Sugar.
- Lupton, D. (2015). *Digital sociology*. London: Routledge.
- Lyle, K. (2017). Shaping the future of sociology: The challenge of interdisciplinarity beyond the social sciences. *Sociology*, 51(6), 1169-1185.

- Marr, B. (2014). *Big data: The 5 Vs everyone must know*. Retrieved from <https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know>
- Marres, N., & Gerlitz, C. (2016). Interface methods: Renegotiating relations between digital social research, STS and the sociology of innovation. *Sociological Review*, 64(1), 21-46.
- Marx, K. (1867). *Das Kapital*. Moscow, USSR: Progress Publishers.
- Marz, N., & Warren, J. (2012). *Big data: Principles and best practices of scalable realtime data systems* (MEAP edition). Westhampton, NJ: Manning.
- McKinsey Global Institute. (2011). *Technologies, jobs and the future of work*. McKinsey & Company. Retrieved from <http://www.eyesreg.it/2016/big-data-e-scienze-regionali-una-relazione-inesplorata/>
- McNulty, E. (2014). *Understanding big data: The seven V's*. Retrieved from <http://dataconomy.com/seven-vs-big-data/>
- Mead, G. H. (1934). *Mind, self and society*. Chicago: University of Chicago Press.
- Mills, C. W. (1959). *Immaginazione sociologica*. Milano: Il Saggiatore.
- Morin, E. (2001). *Per uscire dal ventesimo secolo*. Bergamo: Lubrina, 1989 (ed. orig. *Pour sortir du XX si ècle*. Paris: Seuil, 1981).
- Parsons, T. (1969). *Structure and process in modern societies*. Glencoe, Illinois: The Free Press.
- Popper, K. R. (2002). *The logic of scientific discovery*. London: Routledge.
- Qi, X. (2021). Social networks as contexts for engagement and initiative: An empirical investigation. *Current Sociology*, pp. 1-8.
- Ruppert, E. (2015). Doing the transparent state: Open government data as performance indicators. In R. Rottenburg, S. E. Merry, S.-J. Park, and J. Mugler (Eds.), *A world of indicators: The making of governmental knowledge through quantification* (pp. 1-18). Cambridge: Cambridge University Press.
- Russell, B. (1927). *The analysis of matter*. London: Routledge.
- Salais, R. (2016). Quantification and objectivity. From statistical conventions to social conventions. *Historical Social Research*, 41(2), 118-134.
- Salganik, M. J. (2018). *Bit by bit. Social research in the digital age*. Princeton: Princeton University Press.
- Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology*, 41(5), 885-899.
- Simon, H. (1969). *Designing organizations for an information-rich world*. Pittsburg, Pennsylvania: Carnegie Mellon University.
- Srnicek, N. (2017). *Platform capitalism*. Cambridge: Polity Press.
- Stein, E. (1958). *Il problema dell'empatia* (tr. it. di E. Costantini, E. S. Costantini). Rome: Studium.
- Strathern, M. (2000). The tyranny of transparency. *British Educational Research Journal*, 26(3), 309-321.
- Th évenot, L. (1983). L'économie du codage social. *Critique de l'économie politique*, 23/24, 188-222.
- Thomas, W. I., & Znaniecki, F. (1927). *Il contadino polacco in Europa e in America* (Vol. 5, Chicago e Boston 1918-1920). Milan: Edizioni di Comunità 1968, vol. II. Spec. Introduzione.
- Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, 42(1), 35-67.
- Van Dijck, J. (2014). Datafication, dataism and dataveillance. Big data between scientific paradigm an ideology. *Surveillance & Society*, 12(2), 197-208.
- Van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. New York, NY: Oxford University Press.
- Weber, M. (1922). (postuma; trad. it. 1961). *Economia e società* (Vol. 2). Milan: Edizioni di Comunità
- Weber, M. (1964). Engisch, K., Pfister, B., & Winkelmann, J. (Eds.). *Max Weber. Gedächtnisschrift der Ludwig-Maximilians-Universität München zur 100. Wiederkehr seines Geburtstages 1964* [in honour of the centenary of his birth]. Berlin: Duncker und Humblot.
- Williams, M. L., Burnap, P., & Sloan, L. (2017). Crime sensing with big data. The affordances and limitations of using open-source communications to estimate crime patterns. *British Journal of Criminology*, 57(2), 320-340.
- Wynn, J. R. (2009). Digital sociology: Emergent technologies in the field and the classroom. *Sociological Forum*, 24(2), 448-456.
- Zhu, J., Ni, P., Tong, G., Wang, G., & Huang, J. (2021). Influence maximization problem with echo chamber effect in social network. *IEEE Transactions on Computational Social Systems*, 8(5), 1163-1171.
- Zincke, C. R. (2014). Local and global communications in Chilean social science: Inequality and relative autonomy. *Current Sociology*, 62(5), 704-722.
- Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75-89.