Volume 22 Issue 35
Originally issued as C4



Global Journal of human-social science: C
Sociology & Culture
Volume 22 Issue 4 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-460X & Print ISSN: 0975-587X

The Problem of Scientific Knowledge in Sociology: Big Data, Representativity and Abduction

By Cleto Corposanto
Magna Graecia University

Abstract- There are several aspects that link the themes of science and knowledge from a sociological perspective, and which relate our - each one of us - being a singular, unique, and unrepeatable individual to the fact that we are nonetheless necessarily part of a larger whole, of a collectivity on which we certainly depend. The first is certainly that of reasoning around the central theme of the construction of social relations: I have tackled various types and of various natures, from different angles, but it is always a question of relations (Corposanto 2016, 2018, 2021). Because after all, since we are part of a complexity in perpetual movement, without relationships we would be a useless part of a whole. The second, only apparently separate from the first (but part of the same, in a symbiosis of scientific love) is that of the tools of knowledge and the technical needs of the researcher: in a single term, methodology. The two threads chase each other, intertwine and disjoin, and then necessarily intertwine again, to compose a framework of inescapable eclecticism for a sociology that must constantly come to terms with continuous reductions and re-compositions.

GJHSS-C Classification: *DDC Code: 301 LCC Code: HM24*



Strictly as per the compliance and regulations of:



© 2022. Cleto Corposanto. This research/review article is distributed under the terms of the Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0). You must give appropriate credit to authors and reference this article if parts of the article are reproduced in any manner. Applicable licensing terms are at https://creativecommons.org/licenses/by-nc-nd/4.0/.

The Problem of Scientific Knowledge in Sociology: Big Data, Representativity and Abduction

Cleto Corposanto

Abstract- There are several aspects that link the themes of science and knowledge from a sociological perspective, and which relate our - each one of us - being a singular, unique, and unrepeatable individual to the fact that we are nonetheless necessarily part of a larger whole, of a collectivity on which we certainly depend. The first is certainly that of reasoning around the central theme of the construction of social relations: I have tackled various types and of various natures, from different angles, but it is always a question of relations (Corposanto 2016, 2018, 2021). Because after all, since we are part of a complexity in perpetual movement, without relationships we would be a useless part of a whole. The second, only apparently separate from the first (but part of the same, in a symbiosis of scientific love) is that of the tools of knowledge and the technical needs of the researcher: in a single term, methodology. The two threads chase each other, intertwine and disjoin, and then necessarily intertwine again, to compose a framework of inescapable eclecticism for a sociology that must constantly come to terms with continuous reductions and re-compositions. The proposal then becomes that of an inclusive sociology, epistemologically tolerant, without any claim to be exhaustive in its space-time arguments (which, moreover, as is clear from particle physics, are themselves social constructions lacking the requirements of objectivity and truth in themselves). A sociology, however, that is open to the versatility of knowledge and the certainty of the absence of linearity in conclusions, to the awareness that there is no true paradigm that does not at the same time presuppose a possible error, and finally that the gaze, albeit fleeting, on the social world must nevertheless try to make the maximum effort to be credible, even before being plausible. While starting from an ineliminable and - perhaps - the only certainty in the necessary premises: that of the complexity, of things, of the scenarios, of the approaches required and of the analysis of the relationships between things and events.

I. Introduction

Often, to achieve tangible results, as scientists we pretend to ignore the fact that reductionism is only a partial approach to explanation: the hyper-focus on the parts does not allow us to grasp the plus that comes from interaction, that sort of 'emergent reality' from the whole that is probably also at the basis of the very birth of biological life on our planet. And so, all our efforts are aimed at unveiling (somewhat technicistically) single aspects, which are often misleading with respect to the vision of the whole (and with respect to which, in fact, we are often unable to find an answer except, perhaps, in the form of an apparent serendipity). The problem, perhaps, lies in the fact that we do not have the right tools, we have a toolbox that is vastly undersized in relation to our cognitive needs; a bit like what happened with the knowledge of the cosmos, before Hubble (and even before that to those who dealt with space until the advent of the telescope). Despite this, we arm the idea that there can be nothing but uniqueness in the scientific method, regardless of whether we study planets, black holes, trees, the atmosphere, horses, men, soldiers, and wars. This is the approach that tends to separate man from nature (but also mind from brain), with very often disappointing results: we consider our species, *homo sapiens*, as the ruler of the world (with all the comforts that would come with it) but then all it takes is a bacterial infection or the strengthening of a virus to bring everyone to their knees (Quammen, 2012). And it pushes neuroscience to consider the brain exclusively in terms of neurons and synapses that can only be described in electrical and chemical terms, practically eliminating the concept of mind altogether; and, on the other hand, the great deployment of forces that, by separating the mind from the brain, makes it an object that can be analyzed, as if it were an objectively existing reality regardless. In this way we fail to grasp what emerges precisely from the complexity of things, from their interrelation, which as an 'emergent reality' (Morin 1973) produces what we call soul and conscience (which are also part of us and help us in the regulation of our being social animals).

The problem lies, as we said, probably in the choice of tools in the toolbox. Which is not unique, on the contrary, but certainly as it stands it appears inadequate: it must in fact adapt to a multiplicity of approaches depending on whether scientists, first and foremost, deal with inanimate objects or, precisely, those endowed with soul, consciousness and consequently emotions.

Scientific reasoning - supposedly unique, perfect, objective - still solidly bases its cornerstones on the consequences of the evident success (theoretical and practical) achieved over the centuries since Galilean intuition. In the course of time, the granitic conviction that scientific success can depend exclusively on a single, simple principle of method has in fact been slightly shaken; a solid scientistic basis remains regardless, but the calls for rethinking and eclecticism even in the methodological approach are beginning to be 'important'. It therefore seems entirely cogent to emphasize the difference between the 'inanimate'

Author: Professor of Sociology – UMG – Magna Graecia University – Catanzaro, Italy. e-mail: cleto.corposanto@unicz.it

© 2022 Global Journals

sciences - which have gradually seen the level of complication increase, moreover, due to the fact that the more the system of knowledge grows, the more dark areas to be revealed - and the 'animate' sciences, which have as their object of interest people, sensitivities, ethics, behavior and social actions.

From this perspective, the pandemic could also play a regenerative aspect on the social sciences, in particular on methods and consequently on relations with other scientific disciplines, just as the virus brings, in its devastating course, opportunities for rebirth for societies and their vital organizations. A sort of stress virus also for what concerns the social sciences, therefore, which arrives at a historical moment of evident difficulty of the same in general and of Sociology in particular. A crisis that stems from afar, from a scientistic drift on which much has been focused, and which has had the opposite effect to the one desired, causing the capacity for scientific recognition of the natural vocation for knowledge of the mechanisms that regulate social action to implode rather than to enhance it. In the meantime, the subjective nature of the pandemic seems clear: not only - or not so much - because some of the measures to contain it concern the individual and collective social sphere (and thus significantly affect our own room for manoeuvre within social relations), but above all because its origin can only be interpreted by combining bio-virological studies with those on our collective behavior and on many of the choices that have characterized our recent development models. Viruses have existed in nature for millions of years, and it is only the behavior of the most important animal species that causes them, through wrong choices, to move from one place to another. With consequences that we have seen can be disastrous. The social aspects are therefore not simply a possible 'cure' but can be analyzed ex ante, and from this point of view constitute a formidable aspect of 'preventive medicine (not in the strictly pharmacological sense of the term)'. The discourse thus remains centered on the method, which has always been the main flaw in the accreditation of sociological research in the scientific sphere. From this point of view, the choice of multi-paradigmicity flaunted by the scientific community appears more and more like an attempt to request the legitimization of positions that no one wantsto discuss so as not to run the risk of losing important room for manoeuvre - and academic power. And while we sit here discussing the prevalence of quantitative over qualitative (or vice versa, it is the same thing), of standard and non-standard, of intrusive and periscopic, of objective and constructive, we are gradually slipping away from the main stage, that of recognized scientific knowledge. The fault of the sociologists or Sociology? Of the sociologists, I have no doubt.

Yet, the signs of a way out have been there for some time. A possible path, an overcoming of the useless dualism that has torn apart scientific credibility. If the problems of credibility, scientific credibility are not solved first, we will go nowhere.

A new paradigm, which anticipated 'in theory' what could have happened - and which punctually did happen - is that relating to the use of large masses of data. We started talking about it at the dawn of the new millennium, when Big Data did not yet exist. When the discussion was still centredon the concepts of statistical representativeness and its real capacities to respond to a sociological representativeness that was only other because - fortunately - we are endowed as people with brains unlike the black and white balls of probabilistic experiments, some began to show interest in the possibility of going further. As is often the case when one finds oneself mired in a seemingly irresolvable dualism, very often the way out lies elsewhere. I was helped by the first ANNs, the artificial neural networks, mathematical models that simulated the behavior of their sisters ANNs, the natural ones, made of neurons and synapses. I dealt with this many years ago (Corposanto 2001), proposing a new paradigm of interpretative data analysis aimed more at a sort of incorporation of the classical, qualitative and quantitative approaches (and also periscopic and intrusive, which I also tried my hand at a few years later) than at overcoming them. The reasoning was simple: do I trust the results obtained from a good number of cases (statistically speaking) processed using strictly quantitative methods in deference to the principle of the uniqueness of the scientific method, or do I rather find the results of a few in-depth qualitative interactions more suitable based on a grounded theory that reverses the hypothetical-deductive perspective? Based on which principle do I choose? I suggested, then as now, to rely on the only model that, instead of arguing about the method, reasoned about the result. The ANNs were the basis for observing exactly what the trend of a phenomenon was based on different variables - whether qualitative or quantitative, even considered together, thus overcoming the limit of their operational 'contamination': the model 'learned' from reality data and was thus able to identify extremely precise predictive paths. It was the keystone, albeit only theoretical. I have never been convinced by strictly mathematical approaches to human behavior because data, despite what some people continue to think, do not speak for themselves: but it was still a breakthrough. I remain convinced that the great capacity for sociological imagination plays a central role in the capacity for sociological analysis but can be usefully employed in the choice of aspects, variables, and models of interest from time to time. That breakthrough was the basis on which the so-called multi-agent models, simulative models, were born, and it is the one on which today's network has developed allowing great capacities of analysis (Manzo 2022), also thanks to the

© 2022 Global Journals

aid of mixed methods, on equally great quantities of variables/data/ information that can be found.

This is how a 'neutral' methodological approach - from the point of view of the origin of the dataset and therefore also of the scientific disciplines that can draw information from it - brings different scientific approaches back on the same level, no longer hard or soft as a sort of scientific-academic allotment has always maintained (Corposanto & Molinari, 2022).

In this perspective, sociologists can once again occupy a leading position in the scientific debate, making use of their ability to read in advance the situation to be analyzed (the hypothesis formulation phase), carrying out an adequate intervention plan (by means of imagination) and being able to count on an apparatus of techniques that today appear more adequate to grasp the meaning of things (Wright Mills 1953). If I want to understand the state of mind of the people who are experiencing a particular situation, I can then work with a standard method (questionnaire and data analysis), through the reconstruction of interviews and/or life stories (to investigate how social reality settles in individual consciences) or I can resort to millions of information from different sources (blogs, videos, messages, photos, comments, tweets, etc.) to grasp the essence of things.

II. Datism & Big Data

Algorithms were born as a general orientation tool within what is generically identified as Big Data.They have long since officially become part of the interest of social researchers (as well as, of course, of those who 'monetize' with information) even if they have not yet been fully exploited. In many cases, it is the very role of the researcher that is at the center of the debate, rather than Big Data itself: for some, in fact, it could ultimately be a true paradigm shift - in Thomas Khun's own sense - even though this is certainly not inevitable, nor is it easily framed from a theoretical point of view (Kuhn 1962). For many, this would be a new empiricism that would go in the direction of a full positivist and post-positivist fulfilment: the realization of a project of social control and prediction made possible by the otherwise incalculable amount of available data. But it seems that no one can do without Big Data anymore, for many reasons. And so, starting from the same economic sustainability of the research, working on Big Data could rather help to solve some methodological problems peculiar to quantitative social research: the interviewer effect (already recalled about the concept of intrusiveness and linked to the Hawthorne effect) but also the so-called social desirability. It is therefore necessary to place oneself in the perspective of working on Big Data and not with Big Data: in this way, the researcher could exploit the information potential of Big Data without negotiating his key role in the process. This would make it possible to utilize this enormous data resource by overcoming some of the problems that have been raised in this regard over the past few years. One of the many interesting speeches on this subject was certainly that of Chris Anderson, editor of the computer science magazine 'Wired', who (provocatively) warned researchers in 2008 that correlation would soon supplant causation due to the gigantic amount of data available: “Petabytes allow us to say: Correlation is enough. Correlation supersedes causation, and science can advance even without coherent models and unified theories”. Evidently, in other scientific circles (those who like to remember how 'the data speaks for itself'), the opportunity of Big Data was seen as a welcome one: “Big data is about what, not about why. We don't always need to know the cause of phenomenon: rather, we can let data speak for itself”, Viktor Mayer- Schönberger, Professor of Internet Governance and Regulation at Oxford University and Kenneth Cukier, Data Editor for 'The Economist' wrote in 2013. Of course, one must also reflect on the relationship between Big Data and representativeness in comparison to traditional statistical sampling methods. We have dealt with this in detail, showing how statistical representativeness is actually a 'myth' when it comes to understanding social actions: “It is one thing if we consider sampling as a simple extraction from an urn but it is a different matter if we follow the same procedure by interviewing people that, unlike the dice boxes of the lottery, may refuse to contribute to their task. Besides, it should be pointed out that a ‘random sampling’ with individuals is statistically representative only if the population is well known in its entirety and a list has been provided. Under these circumstances, it is quite evident that carrying out the inference procedure on the outcomes obtained may result rather difficult” (Corposanto & Molinari, 2022). A great bath of what has been appropriately called 'dataism', in the empirically founded belief that data can speak for itself, in an entirely new mode of scientific knowledge comparable to a kind of 'exploratory science'. Now, while there is no doubt that Big Data analysis can be of great interest for all fields of research, it is equally certain that, on its own, it is absolutely insufficient. Relying on interpretations of the recurrences and concordances in large amounts of data, if it can be inviting from a descriptive point of view, certainly contravenes the scientific process of hypotheses and theories to be tested (if taken in this way, only at an exploratory level). Dataism, however, is certainly a fascinating mode of scientific approach, especially since it would envisage that 'uniqueness of method' that has always been one of the main aims of the major philosophers of science: the problem is that, although we are all the offspring of a biological evolutionary process, the fact that as *Homo sapiens* we have evolved over other living organisms, far from giving us the

© 2022 Global Journals

prerogatives of useless and even harmful primacy, certainly assigns us the ability to relate to each other in culturally determined ways, which end up characterizing our actions (both individual and collective) in the sign of a complexity that, precisely, a uniqueness of method would not be able to fully grasp. A certain, unique, and incontrovertible scientific paradigm would be nice. It would be the driving force from which would spring certain and unassailable dogmas; the problem is that quantifying emotions, feelings, states of mind and suppositions is an a priori reductionist operation that contrasts, precisely, with the richness that the nature-culture symbiosis (mediated by time and place) characterizes our being social actors. Dataism, in short, would go exactly in the direction of an algocracy that is useful in many situations related to our daily lives, but which will certainly never be able to completely take the place of human argumentation. This despite assertions of the goodness of a model that produces data in large quantities in which writing is the condition of possibility of all reality. For what appears certain, is that what has been called 'surveillance capitalism' does not have for certain only a commercial purpose, let us say, of the traditional kind: the use that can be made of documents and data, in addition obviously to commodifying humanity using services that are only apparently free, appears well described by the Foucaltian analysis of the Panopticon (albeit in completely different times and situations) (Foucault 1975). The latter, far from needing to be treated with a logic that unveils their complication, rather see the growing awareness of considering their complexity, which can almost never be addressed, as happens in the other case, with systems of cause-effect explanation, which are also usually linear.

III. Conclusion

A complex toolbox, therefore, the one that seems to be the best one for dealing with the storms of explanation in such diverse sciences and fields, cannot but extend, in the meantime, also to abduction (safeguarding, however, approaches of deduction and induction), which somehow seems to be congruous in many situations in which there is a need to reconstruct premises starting from rules and results that are in some way known. In fact, therefore, ab- deduction does not seek to make predictions, it does not seek probability but possibility, it does not calculate but asks questions and seeks answers. This is what we all do naturally daily: abduction is a form of reasoning that deals with probabilities and likelihoods. The logical conclusion of sound abductive reasoning is therefore a hypothesis that provides the best explanation of a whole series of known facts. This is because if thought is naturally inferential, abduction is somehow the only inference that can move it forward,to think about being in the future. It is in essence a situation of perpetual tension towards explanation, which certainly runs in the direction of providing answers to questions that would otherwise risk remaining unanswered. In a way, it is a question of setting the state of initiation of adesign process; and, as Peirce correctly points out (Peirce 1940), abduction can certainly represent the initiation, whereas induction can be considered the closure (obviously where the process can be completed in this direction). Abduction occurs, then, when thought makes a lateral movement (or even when it proceeds backwards, in which case it is also called retroduction). What remains is that the point of arrival of these three types is different: if for an induction it is a synthesis and for a deduction a thesis, that of an abduction takes the form of a hypothesis. It is a question of broadening perspectives, in short. The paradigm of simplification - and of the uniqueness of the method - resembles a paradigm of the search for the maximum utility of profit, in a common government of science, technique and - fatally - economics and markets. It is then a matter of embracing multidimensional explanatory possibilities, in the face of the objective complexity of the frameworks on which one works. A situation that is as well known today as in the past: it is perhaps worth re-evaluating Heidegger's path. The concept of *Vorwissenschaft* (preliminary science), which first transited into *Hermeneutik der Faktizität* (hermeneutics of facticity) and later landed in *Existenziale Analytik* (existential analytics), can contribute to restoring to sociological thought that interpretative brilliance that scientism seems to have decisively contributed to drying up.

Bibliography

1. Corposanto C. (2001), La classificazione in Sociologia. Reti neurali, Discriminant e Cluster Analysis, Franco Angeli ed.
2. Corposanto C. (2016), From the originally Twaddle's Triad to a new ESA Model: the Sonetness, *Journal of Health Education Research & Development*, ISSN: 2380-5439, 2016
3. Corposanto C (2018), Awareness of the disease, *J Community Med Health Educ*, n.8, DOI: 10.4172/ 2161-0711-C2-034
4. Corposanto C. (2021), Le relazioni pandemiche. Istruzioni per l'uso, Rubbettino ed
5. Corposanto C., Echeverria J, Fotino M. (eds)(2021), Covid19. Sociological scenarios, TheDiagonales ed, ISBN: 978-88-945357-2-3
6. Corposanto C, Molinari B (2022), How does the error from sampling to Big Data change?, ISR – Italian Sociological Review, vol.12, n 7/S
7. Foucault M. (1975), Surveiller et punir. Naissance de la prison, Editions Gallimard, Paris.
8. Heidegger M. (1923) Ontologie. Hermeneutik der Faktizität, GA 63, 1988; trad. it.

© 2022 Global Journals

9. di G. Auletta, Ontologia. Ermeneutica della fatticità, Guida, Napoli 1992.
10. Kuhn T.S. (1962), The Structure of Scientific Revolutions, The University of Chicago
11. Manzo G. (2022), Agent based Models and Causal Inference, Wiley
12. Morin E. (1973), Le paradigm perdu: la nature humaine, Les Éditions du Seuil
13. Peirce C.S., (1940), The Philosophy of Peirce: Selected Writings, a cura di J. Buchler, London: Routledge & Kegan Paul; 1955, Philosophical Writings of Peirce, New York: Dover Publications.
14. Quammen D. (2012), Spillover, WW Norton & Company Inc.
15. Wright Mills C. (1953), The Sociological Imagination, Oxford University Press.

© 2022 Global Journals