Perspectives of the Computational Approach as a Method for Criminological Research

FABRIZIO CACCAVALE[*]

SUMMARY: *1. Statistics as a Consolidated Tool of Criminological Research – 2. Data Processing and Their Interpolation: The Computer Tool – 3. The New Frontiers of the Computational Method in the Criminology Domain*

1. STATISTICS AS A CONSOLIDATED TOOL OF CRIMINOLOGICAL RESEARCH

In criminological research methodology, the use of quantitative methods plays a very important role contributing to the very spirit of the subject matter. Studies with a sociological matrix on crime, already existing from the 1800s, try to explain the essence of it through the analysis of the "ultimate cause" of its manifestations, and the use of systematised data for its description becomes a significant moment in the research. In fact, "the image of criminology as a synthetic science is justified if it is understood in terms of empirical science, characterised by the inductive method and founded on observation, that is, if criminology is considered the result of a systematisation of assertions that describe observations and if we believe that criminological knowledge must be built on the basis of principles of observatism"[1].

The studies carried out by Quetelet (from Belgium) and Guerry (from France) are recognized as the first important statistical studies[2], and these became of utmost importance in the criminological domain for qualitative and quantitative knowledge of delinquency, for both verifying the interpretations of this phenomenon offered on a theoretical level and for offering measures aimed at preventing and containing its spread.

---

[*] The Author is assistant professor of Criminology and criminal law at "Federico II" University, Naples (Italy).

[1] M. COCO, F. MICHELUZZI, G. PISAPIA, *Criminologia. Norme e regole*, Torino, UTET, 2003, pp. 11-12.

[2] Adolphe Quetelet and Andrè-Michel Guerry are recognised as the founders of moral statistics, embryo and founding element of criminal sociology, and promoters of the Cartographic or Geographic School who interpret the delinquency phenomenon with a statistic geographic representation of its distribution analysing the relationship existing between socio-demographic factors and fluctuations in crime indexes (*amplius*, A.J. QUETELET, *Physique sociale ou Essai sur le développement des facultés de l'homme*, Bruxelles, Hayez, 1835; A.-M. GUERRY, *Essai sur la statistique morale de la France*, Paris, Grochard, 1833).

The main source, and also the one more easily accessible, from which to obtain these data is represented by the official measurements furnished by formal and social control agencies and ordered by bodies in charge of the study of the spread of the behaviour and attitudes of the collectivity throughout the national territory[3]. However, the most important limit that weighs on research conducted by using official statistical data derives from the fact that that they are only representative of those kinds of behaviour that are known to the social control agencies. The reality, instead, is also made up of a whole series of crimes which, eluding the spotlights of the institutional system of the control of the crime, remain within the cone of shade of the unknowable and, therefore, hidden. Hidden crime, therefore, constitutes the deficit of official knowledge, and can be defined as a set of those crimes that have effectively been committed but not recorded. To overcome the limits to research deriving from this so-called "dark number" of criminality[4], and to better quantify the spread of the delinquency phenomenon in time or space – considering that the data officially coming from the selective

[3] In Italy, for example, the *Istituto Nazionale di Statistica* (National Institute for Statistics) which collects and processes data in the legal domain, by gathering up-to-date information on the fundamental data and flows of the demand for justice in Italy and on the capacity of the system to respond. As far as specifically criminal matters are concerned, they include: statistics of criminal activities, crime statistics, statistics of accused persons who have been convicted, statistics of criminal proceedings and statistics of criminal military proceedings.

[4] The index of dark crime, that is, the relationship between crimes effectively committed and officially recorded, varies over time and according to the type of crime and concern, above all, those which, due to their nature, can only be brought to attention through criminal charges, which if encouraged by individual factors or favourable situations (hope of recovering the stolen goods, compensation from insurance, aiding the criminal's capture, the victim's civic conscience, faith in the authorities, etc.), are often not presented for a series of reasons, for instance, fear of reprisals, of a public scandal, the code of silence, or even compassion for the criminal, justification of his/her action for ideological or religious reasons, or political solidarity towards the criminals. But not only. The size of the number of hidden crimes may also depend both on the manner in which a certain behaviour is defined as criminally significant and the ways in which this definition is applied in practice by the institutional apparatus responsible for social control. On the amount of the "dark number" and, in particular, on the variables that motivate it and condition its entity, cfr., among others, S. AMBROSET, G.V. PISAPIA, *Numero oscuro della devianza e questione criminale*, Verona, Bertani, 1980, pp. 38-39; L. RADZINOWICZ, J. KING, *La spirale del crimine. L'esperienza internazionale*, Milano, Giuffrè, 1981, pp. 46 ff.; G. FORTI, *Tra criminologia e diritto penale. Brevi note su "cifre nere" e funzione generalpreventiva della pena*, in Marinucci G., Dolcini E. (a cura di), "Diritto penale in trasformazione", Milano, Giuffrè, 1985, p. 67 ff.; N. WEINER, M.E. WOLFGANG, *Le fonti dei dati in criminologia*, in Ferracuti F. (a cura di), "Trattato di criminologia, medicina criminologica e psichiatria forense", Milano, Giuffrè, 1987, 1 (Le

action of the social control agencies are only representative of a modest part of the real numeric size of the criminality – noteworthy scientific efforts have been made: also through social science techniques, methods for detecting committed crimes, provided with greater reliability and penetration compared to traditional methods, are basically "self-report research" – that concern potential authors of deviant behaviours – and "victimisation surveys" – that look at possible victims of a crime. In *self-report studies*, whose origins go back to the end of the 1940s in the United States and spreading subsequently to many European and non-European countries, sample surveys are conducted that usually use structured self-administered and anonymous questionnaires, or else the interview system, through which certain social groups are invited to report any participation in delinquent activities and, if so, to provide information about its frequency and characteristics, as well as, possibly, the social and judicial reactions resulting from it. *Victimization surveys* became especially popular in English-language speaking countries from the second half of the 1960s and involve questionnaires given to representative samples of the population in order to identify whoever among those interviewed or their families had been a victim of crime within a specific period of time, to learn about any complaints made, to collect information about the dynamics of the act and the consequences it had.

   Going into more detail into the advantages deriving from these research techniques as well as the problematic nodes they present fall outside the scope of this contribution[5]; here, it is sufficient to keep in mind that the

---

radici, le fonti, gli obiettivi e lo sviluppo della criminologia), pp. 125-126; S. CORRADO, *Statistica giudiziaria*, Rimini, Maggioli, 1993, p. 154.

   [5] On controversial issues regarding the methodology of self report researches, see, among others, D.S. ELLIOTT, S.S. AGETON, *Reconciling Race and Class Differences in Self-reported and Official Estimates of Delinquency*, in "American Sociological Review", 1980, n. 45, p. 95 ff.; M. HINDELANG, T. HIRSCHI, J.G. WEIS, *Measuring Delinquency*, Sage, Beverly Hills, 1981; D.P. FARRINGTON, R. LOEBER, M. STOUTHAMER-LOEBER, W.B.V. KAMMEN, L. SCHMIDT, *Self-reported Delinquency and a Combined Delinquency Seriousness Scale Based on Boys, Mothers and Teachers: Concurrent and Predictive Validity for African-Americans and Caucasians*, in "Criminology", 1996, n. 34, p. 493 ff.; D. JOLLIFFE, D.P. FARRINGTON, J.D. HAWKINS, R.F. CATALANO, K.G. HILL, R. KOSTERMAN, *Predictive, Concurrent, Prospective and Retrospective Validity of Self-reported Delinquency*, in "Criminal Behaviour and Mental Health", 2003, n. 13, p. 179 ff. In relation to victimisation surveys, see, SKOGAN W.C. (ed.), *Sample Surveys of the Victims of Crime*, Cambridge, Ballinger, 1976; R. SPARKS, *Research on Victims of Crime: Accomplishment, Issues, and New Directions*, Washington, U.S. Government Printing Office, 1982; J.G. WEIS, *Crime Statistics. Reporting Systems and Methods*, in Kadish S.H. (ed.), "Encyclopedia of Crime and Justice", New York, The

"institutionalistic" approach is that of those who argue that the most valid data are those collected by official institutions since crime and the criminality are only significant formulations at the time in which they are formally recognised. This is in contrast with the "realist" position which leans towards the need for an integrative supplement of official statistics rather than their institutional validation. This latter approach, in fact, is concerned with the statistical description of delinquency that is as close to reality as possible compared to the former which focuses on the use of criminal statistics as the output of agencies dealing with criminal complaints.

2. DATA PROCESSING AND THEIR INTERPOLATION: THE COMPUTER TOOL

Initially, the computer was timidly adopted for processing, classifying and displaying numeric data; subsequently, with the development of proper software and greater computing power, it increasingly became more widespread, up to the point of profoundly and inescapably taking root in the operational modes of criminological research. However, apart from requiring advanced technical skills, software in use often necessitated exhaustive study and appropriate exercise to be effectively applied.

In particular, in the final decade of last century, I used a personal computer to store data relating to a study on authorisations to proceed in the Italian Parliament: a very large mass of knowledge and information (parties that the persons under investigation belonged to, place of their electoral college, type of crime they were accused of for which an authorisation to proceed was sought, etc.) that allowed me to describe and assess the actual qualitative and quantitative extent of the "parliamentary delinquency" phenomenon and its effective evolution, above all in the medium- and long-term. In practice, it was an investigation that acquired raw data and reprocessed to provide, apart from an initial and general classification, their organic and periodic quantisation[6].

Free Press, Vol. 1, 1983; U. GATTI, M.I. MARUGO, G. FOSSA, V. MATERAZZI, *Le inchieste di vittimizzazione: problemi metodologici e primi risultati di uno studio-pilota condotto nella città di Genova*, in "Rassegna italiana di criminologia", 1991, n. 4, p. 363 ff.; M. BARBAGLI, *Immigrazione e criminalità in Italia*, Bologna, il Mulino, 1998.

[6] M. D'ANIELLO, F. SCLAFANI, with the collaboration of F. CACCAVALE, *Autorizzazioni a procedere. Analisi del fenomeno nel Parlamento italiano 1948-1991*, Napoli, ESI, 1991; F. SCLAFANI, M. D'ANIELLO, F. CACCAVALE, *Le autorizzazioni a procedere nel Par-*

Subsequently, this type of research and analysis of criminality regarding its extent has been refined by a statistics-type analysis: dedicated software has, on the one hand, offered an ever increasing number of processing possibilities and, on the other, it has undergone a progressive process of simplification in use, that has enabled frequent and repeated use from the time the user has also been able operate without the assistance of specialised personnel. As a result, "multivariate data analysis" and "calculation of Chi-squared distribution on the significance of the correlation of factors", for example, represent terminological names, an expression of statistical techniques by now firmly acquired by scholars in criminology and which they find difficult to exclude in giving a more complete and consistent representation of the (criminal) reality they intend to investigate.

In this regard, forgive me if I once again refer to a recent study I carried out relating to Naples that was part of a wider ISRD 2 international research project (*International Comparative Self-Report Delinquency Study*) promoted and organised by the Research and Documentation Centre of the Dutch Ministry of Justice, in which numerous European countries participated. In particular, as far as Italy is concerned, given that in Italy there are marked difference in social phenomena from region to region, with high economic, social and cultural specificities, in the various geographical areas, the research was carried out in 15 cities selected on the basis of the number of inhabitants and their geographical position appertaining to the north-south axis. The statistical data was processed using the *SPSS 15.0 for Windows*, a powerful and flexible software that allowed extremely deep and wide analysis[7].

From what has already been discussed, it emerges that the use of computer tools - so far having occurred essentially for processing statistics – has taken an ancillary place compared to traditional criminological goals: the method consisted of collecting data from concrete reality through its direct observation; their interpolation and combination through the use of the computer, then, for obtaining a statistical-type logical representation of them for learning about criminality, its prevention and validating the relative interpretation theories.

---

lamento italiano. *La Camera dei Deputati*, in "Medicina legale - Quaderni camerti", 2002, p. 485 ff.

  [7] F. CACCAVALE, *I comportamenti devianti dei giovani nella realtà napoletana*, in Gatti U., Gualco B., Traverso S. (a cura di), "La delinquenza giovanile in Italia. I risultati di una ricerca multicentrica", Lecce, Pensa Multimedia Editore, 2011, p. 185 ff.

3. THE NEW FRONTIERS OF THE COMPUTATIONAL METHOD IN THE CRIMINOLOGY DOMAIN

In the early 1990s, the epistemological transformation due to the spread of computational approaches gave birth to a new methodology. And if the legal sciences as a whole showed themselves to be substantially indifferent to this metamorphosis, Criminology, on the contrary, particularly in the United States, has shown a certain interest in this approach. This is due to the opportunities offered by computational social science methods to disciplinary contexts where the scholar's attention is concentrated not only on abstract rules but also on the interaction of the latter with social dynamics.

It has been the origin, therefore, of what has been defined as *computational criminology*: a hybrid mixture of criminology, computational sciences and applied mathematics, that aims at finding answers to social uneasiness and alarm deriving from the commission of criminal offences, from the least serious (but not less worrying) like crimes included in micro-criminality, to forms of crime featuring a higher level of danger, such as organised crime and international terrorism. In particular, computational criminology involves the use of the power of computer processing to identify existent and emerging models; explore crimino-genetic and crimino-precipitating factors; identify terrorist networks, organised crime and social gangs; clarify the contours of computer crime. The algorithms are developed by using various exclusively mathematical and statistical techniques, like computational topology, the hypergraph, the discovery of knowledge in databases (KDD), the social networks analysis (SNA), agent-based simulations, the dynamic analysis of information systems, and so on[8].

Amongst these, which therefore represent the new way to use the computer, in particular, one of considerable interest stands out: *agent-based social simulation*. Originating from the encounter of suggestions coming from a heterogeneous set of research sectors, the objective of agent-based simulation is to study society through computational models directed towards discovering how single individuals or groups of persons, by interacting among themselves, bring about complex social phenomena. They can be described, explained or predicted – according to this approach – by reproducing the dynamics that generate them within artificial societies simulated by the computer. *Agent-based simulations are, in fact, true virtual reconstructions that*

---

[8] P.L. BRANTINGHAM, *Computational Criminology*, Paper presented at the European Intelligence and Security Information Conference - EISIC (Athens, 12-14 September 2011).

permit a real phenomenon to be modelled and all the possible evolutions of it to be studied based on changes in the variables in play. Through this new model of knowledge, the most varied phenomena can, in this way, be studied: economic processes, cultural and market trends, emergence and spread of social rules; and this with a methodology that, for its intrinsic characteristics (high level of formalisation, use of computational techniques, resort to experiments), offers social sciences the advantages traditionally linked to the practice of the scientific method: theory falsifiability, inter-subjective communicability, cumulativity of results. The study of delinquency and criminal policy, in particular, has been marked by how, in order to prove the need for complementariety between the experimental and the simulation approach, the use of simulation would appear to be especially useful in all those cases where empirical research is hampered by difficulties deriving from experiments being performed in the field and it would constitute a kind of comparatively inexpensive pre-test for evaluating *ex ante* the different crime prevention programs for the purpose of suggesting any changes before effectively carrying out the empirical test[9]. And, in this regard, it should be noted that the quantitative datum used for processing statistics in the criminal domain only permits an ex post representation of the crime: the mapping described in this way gives us a photograph of delinquency rebus sic stantibus, fixed in the immobile static nature of that particular moment in history coinciding with the act of its acquisition, which can indeed serve in terms of prevention or in the unveiling of the ways in which delinquent processes unfold, but does not offer a concrete, dynamic and articulated vision of social reality and the crimes that occur in it. Agent-based computational science, with the possibility of a simulation approach - which exploits data interpolating them dynamically, together with other socio-environmental factors, in order to obtain a sort of their temporal projection also extending into the future - manages instead to make social dynamics "live", enabling an interpretation of prevention that can take advantage of the knowledge in real-time of the action-reaction links, albeit only in simulated terms between human forms of behaviours and between them and their surroundings[10].

---

[9] V. PUNZO, *Scelta razionale e sociologia del crimine. Un approccio critico e un modello di simulazione ad agenti*, Milano, Franco Angeli, 2012, pp. 194-195, that represents one of the early attempts in Italy to unite the sociological explanation of crime with the modelling of the behaviour of social actors.

[10] On this matter: "Like other modeling approaches, simulation modeling involves the creation of a simplified representation of a social phenomenon (…). The most familiar type

Asking the question in these terms, the agent-based computational approach presents advances of great interest to criminological science. Thus, in the United States, pioneering work that has adopted agent-based simulation dates back to 2001. Since 2004, in particular, the use of this methodology in the field of environmental crime (which can briefly be defined as the study of crime, criminality and victimisation primarily with reference to certain places, and secondly, as the way in which individuals and organizations create space for their activities, being influenced and influencing in turn that same environment). There are two basic objectives to be achieved: on the one hand, to formalise and test specific theories of deviance, complete in their abstract definition, also to highlight any conceptual gaps or to further hone them before empirical validation, and on the other, to serve as a tool for the analysis and evaluation of policies on the fight against crime and the prevention of deviant behaviour.

Some scholars believe simulation is the solution for conducting research in the social sciences domain, in addition to the more traditional verbal and mathematical/statistical representation of theories; and if the simulation and statistical models have the same ratio to better understand a social phenomenon[11], the former presents a series of advantages compared to the latter: in the first place, the possibility of examining the entire dynamic his-

---

of model is a statistical one (e.g., a regression model) in which input data are 'run' via a statistical program and values are output that describe the relationships among the input data. In contrast, simulation models are themselves computer programs that incorporate the critical aspects of the social phenomenon being modeled. The program is run and the output data are analyzed using standard statistical techniques. Some advantages of simulation models include the ability to examine the 'entire dynamical history of the process under study' since information about the dynamics can be collected as the model runs (...). In addition, simulation allows heterogeneity among individuals that more closely approximates the variety found in everyday life and is able to accommodate the non-linear relationships present in dynamic and complex interactions" (E.R. GROFF, *Simulation for Theory Testing and Experimentation: An Example Using Routine Activity Theory and Street Robbery*, in "Journal of Quantitative Criminology" 2007, n. 23, p. 78).

[11] "The logic of developing models using computer simulation is not very different from the logic used for the more familiar statistical models. In either case, there is some phenomenon that we as researchers want to understand better. This is the 'target'. We build a model of the target through a theoretically motivated process of abstraction (this model may be a set of mathematical equations, a statistical equation, such as a regression equation, or a computer program). We then examine the behavior of the model and compare it with observations of the social world. If the output from the model and the data collected from the social world are sufficiently similar, we use this as evidence in favor of the validity of the model (or use a lack of similarity as evidence for disconfirmation)" (N. GILBERT, P. TERNA,

tory of the process under investigation as the information about it can be collected during the processing of the simulation itself[12]; secondly, the simulation allows for heterogeneity among simulated agents which more closely resembles the variety of real world and is able to represent non-linear relations present in dynamic and complex interactions[13]; thirdly, the simulation model can be used in situations where there is little or even no empirical data available whilst statistical models require empirical or simulated data[14]. Then, the additional advantage is that in comparison to empirical research, the simulations have a minimal cost[15].

However, the methodology in question has some limitations, the first of which is that which emerges from these kinds of research is that it is constrained by the assumptions and rules on which the model is based. Thus, agent-based models reflect the quality of the available theoretical and empirical research, data sources and the choices on the basis of which they are implemented in the model. In particular, the represented relations, data sources, the values of the parameters and decision-making rules within the model affect the observed relations within it and their results. In fact, agent-based models rest on random numbers and random numeric distributions for providing a stochastic element for the simulation. Similarly to the choice of the values of the parameters, the choice of the distribution (for example, uniform, normal, etc.) and the time of the distribution (for example, average, standard deviation, etc.) have repercussions on the results of the model. Another limitation concerns the meaning of that which emerges from an

How to Build and Use Agent-based Models in Social Science, Discussion Paper of 18 May 1999, http://web.econ.unito.it/terna/deposito/gil_ter.pdf, pp. 3-4).

[12] "Since the model is 'solved' merely by executing it, there results an entire dynamical history of the process under study". (R. AXTELL, *Why Agents? On the Varied Motivations for Agent Computing in the Social Sciences*, Working Paper, November 2000, http://www.brook.edu/es/dynamics/papers/agents/agents.pdf, pp. 2-3).

[13] J.M. EPSTEIN, R. AXTELL, *Growing Artificial Societies*, Washington, Brookings Institution Press, 1996.

[14] E.R. GROFF, *'Situating' Simulation to Model Human Spatio-temporal Interactions: An Example Using Crime Events*, in "Transactions in GIS", Vol. 11, 2007, n. 4, p. 507 ff. In particular, "simulation modeling can be used in situations where little or no empirical data are available. Statistical models require data, either empirical or simulated" (p. 508).

[15] E.R. GROFF, *Simulation for Theory Testing and Experimentation: An Example Using Routine Activity Theory and Street Robbery*, cit., p. 79.

artificial society: in fact, it is not an empirical test of the theory but rather the extent to which the theory is plausible[16].

Precisely it is this last statement that focuses on the crux of the whole matter, or on what the real contribution of computational simulation is to the path of criminological development in research into the mechanisms explaining delinquent behaviour. Already R. Penrose claims he is able to decree the demise of the theory that machines are able to imitate human thought and practices and in his essay *The Emperor's New Mind*[17] proposes an argument of an algorithmic nature whereby certain characteristics of human intelligence - amongst others, creativity, intuition and consciousness, or rather self-consciousness - cannot be trapped in patterns, or to be more precise, in patterns of an algorithmic nature. The answer seems to be that scholars declare that which they are effectively working at is to teach machines to solve specific problems of a cognitive nature, which if solved y humans, require some degree of intelligence. Today, in fact, it is actually possible to talk about intelligence in agent-based simulation, where everyone does not have a general picture of the environment and of the other agents, but is able to act and coordinate with neighbours, sometimes without an initial target, subsequently discovering it in collaboration with those neighbours.

If all this is true, then it is the creation of a new intelligence, that, most probably, has a different nature from human intelligence. The multidisciplinary context proposes a synthesis which is not merely the sum of disciplines, but a truly new field of study, namely, knowledge with its own ontology that, before proposing solutions to problems, gives the possibility to model them. And it is precisely at this point that agent-based modelling arose, that which can be defined as an artificial vision integrated with existing disciplines. Ultimately, today we have the ability to manipulate multidisciplinary knowledge that has its own *status quo* and is able to continue on its own two feet.

In this way, a series of research horizons are unlocked, whose borders are still now difficult to trace, that are exciting for their amplitude and for the wealth of discoveries it is possible to obtain. The risk, however, is that social simulation, with the creation of an "artificial" world that overlaps the "natural" world, complicate research, and a great deal: how far is it worth-

[16] *Ibidem*.

[17] R. PENROSE, *The Emperor's New Mind: Concerning Computers, Mind and the Laws of Phisics*, Oxford, Oxford University Press, 1989.

while devoting resources in terms of economics and intelligence, to the understanding and interpretation of a wide range of dynamics that are and remain internal to virtual agents, that are, nevertheless, "other" than human agents?