ORIGINAL ARTICLE

Intelligent Technologies and Methodological Transformations in the Social Sciences

Xuan Qin[1]

Abstract

Through a systematic review of the paradigm of contemporary social science research, this paper explores the changes that technologies and algorithms bring to traditional social science methodology. There is a large amount of literature discussing how the application of newly developed technologies provides new solutions to old problems, but little research has been done on how they reshape old problems into new ones. Specifically, how does the development of research tools restructure the criteria for a good study and determine the right questions to ask, and how do computer simulations and artificial intelligence reform social science methodology by changing our presuppositions about humans and society? This article intends to fill the gap by analyzing the evolution of social science methodology and its epistemological impacts.

Keywords Social science · Methodology · Big data · Artificial intelligence · Computer simulation

With the comprehensive penetration of the Internet into human lives and the extensive use of digital devices, a large amount of data related to human behavior and value preferences have been recorded, providing important materials for research in the social sciences. The emergence of big data and the updating of analytical techniques have stimulated the transformation of social science research. A large amount of literature illustrates the application of newly developed data processing tools, discusses how they contribute to solving problems relevant to human society, and reflects on the ethical dilemmas arising therefrom, such as privacy protection and algorithm discrimination. The development of technology not only provides new instruments to solve old questions but also reshapes the questions by changing the way we think and inquire. However, there are few studies about the reform of social science methodology from an epistemological perspective. How does the

✉ Xuan Qin
  qinxuan@fudan.edu.cn

1 Fudan Institute for Advanced Study in Social Sciences, Shanghai, China

development of technology change the criteria for good research, decide which are the right questions to ask, and even change our understanding of humans and society? This article intends to fill this gap by analyzing the evolution of social science methodology and its epistemological impacts. Specifically, the positivist research paradigm pursues the objective neutrality of description, while qualitative research is committed to investigating idiographic cases embedded in specific cultural environments and interactive contexts. Through constructing a parallel artificial society to real human society and observing the free interaction of agents within it under preset parameters in a bottom-up dynamic, computer simulation technology breaks through the boundaries between traditional objective and subjective constructs, between the outside world and the inner self. The emergence of artificial intelligence and intensive data-driven research, on the basis of computer simulation, not only breaks the boundary between the real and artificial worlds but also ends the human monopoly on reason.

According to Jim Gray, human science has experienced four paradigms of the experiment, theory, computational simulation and big data (Tansley and Tolle 2009). Mi applied this structure to social science and derived its four counterparts: quantitative research, qualitative research, social computational simulation and big data-driven research (Mi et al. 2018). In this article, I follow the four paradigms structure and discuss the relationship between them. Each paradigm demonstrates a development with respect to the preceding paradigm(s), but at the same time each to some extent rejects some of the core commitments of the precedent(s).

1 Positivism and three models of scientific explanation

Since the Enlightenment, the development of the natural sciences and their core principles about knowledge have triggered new methods to solve problems in the fields of the humanities and social sciences. Following the example of the natural sciences, theories in the social sciences reformed their research methods and created new research paradigms distinguished from their prescientific or metaphysical traditions. Their aim is to achieve the same precision, coverage, and certainty as the natural sciences in questions relating to human society. In general, this shift advocates using empirical data as the only reliable source of knowledge and constructing models based on the inductive method of "observation-hypothesis-test" to describe, predict, or infer facts about human society.

Naturalism and positivism constitute the deepest characteristics of social science in its initial period. Naturalism believes that "social sciences should be like the natural sciences in some important way" (Risjord 2014, p. 8). To understand human behaviors and social norms, there is no need for special theories that are different from natural sciences; we can obtain answers to problems by strictly following the logic of the natural sciences. Positivism takes a similar view. In the 1830s and 1840s, positivist scholars such as Comte believed that all knowledge about facts should be based on empirical observation and empirical data, experience is the only way to obtain certainty, and acquiring knowledge through empirical evidence is the highest stage of human intellectual development (Comte 1876). Since then, statistics

and probability theory have been introduced into social science research, forming the methodological basis of the quantitative research paradigm. Quantitative research, through data collection and measurement, "mathematizing" the observation of human behaviors and social norms, and through modeling using hypothesis tests and causality analysis, attempt to make objective and neutral descriptions and predictions of social problems. The blooming of quantitative research is the concentrated embodiment of the belief of using scientific methods to study society.

When we further ask whether the "appropriation" of natural scientific methods by quantitative research in the field of social science is legitimate, the question can be expressed in the following form in the discussion of methodology: is the analysis of facts, probability and causal mechanisms by natural scientific explanation models applicable to social scientific explanation? Widely applied scientific explanation models include the Deductive-Nomological Model, the Statistical-Relevance Model and the Causal Mechanism Model.

1.1 Deductive-Nomological Model

The Deductive-Nomological (D-N) Model proposed by Hempel posits that an effective scientific explanation is a deductive process from explanans—the premises to explain—to explanandum—an event or theory to be explained. First, the explanans must contain at least one universal law. Second, the deductive process must be verified by experience. For Hempel, when both conditions are met, it constitutes a valid statement (Hempel 1980).

When we apply the D-N Model in the social sciences, however, the deductive processes of most social science fields, including biology and economics, cannot strictly meet the conditions proposed by the D-N Model. Even in these disciplines, there is no universal consensus on "law" as there is in the natural sciences. Hempel then proposed the Deductive-Statistical (D-S) Model and the Inductive-Statistical (I-S) Model to supplement the D-N Model. Of them, the D-S Model can be considered a weakened D-N Model because it retains the requirement of universal law and shares the same logic. However, the I-S Model replaces the absoluteness of law with possibility and introduces the idea of probability. In the I-S Model, as long as there is a certain probability correlation between the explanans and explanandum, the explanation is valid. The I-S Model and its core concepts exhibit a shift from incontestable certainty to expectation in scientific explanation (Salmon 1989). In addition, "prediction" has become a research issue as valuable as "description" in the sciences and social sciences.

1.2 Statistical-Relevance Model

Based on the I-S Model, Salmon developed the Statistical-Relevance (S-R) Model, which describes the nature of statistics in a more accurate way. In his view, it is inappropriate for the I-S Model to measure the effectiveness of an explanation in terms of probability. He argues that statistical relevance, rather than probability, is what makes up an explanation. Taking the recurring example of John Jones and his

recovery, according to the I-S Model, if it is a statistical law that the probability of recovery under treatment is high, then the information about Jones's treatment and recovery can be used to provide an I-S explanation of Jones's recovery. To Salmon, this is not a good explanation because even without treatment, some people can spontaneously recover, and there is no absolute correlation between treatment and recovery. Therefore, the key is whether the probability of a cure after treatment is greater than that before treatment, that is, whether the treatment is "relevant". In mathematical language, that is, under condition A, C is related to B if and only if $P(B|A.C) \neq P(B|A)$ (Salmon 1971).

1.3 Causal Mechanical Model

From the I-S Model to the S-R Model, the discussion around probability is further developed on the basis of the D-N Model, focusing on the construction of interpretation with the concept of relevance. One of the main criticisms of this approach is that relevance (or even certainty) does not equal causation. Providing relevant explanations for the phenomenon does not necessarily explain the real cause of the phenomenon. Take Salmon's example of males and birth control. According to the D-N Model, males who take birth control pills regularly fail to become pregnant; John Jones is a male who has been taking birth control pills regularly; John Jones fails to become pregnant. This deduction with 100 percent certainty does not tell us that the fundamental reason why Jones fails to become pregnant is that men cannot become pregnant in the first place (Salmon 1971). In this case, we do not need to rely on the D-N Model, we just need to think directly in terms of causality to provide an explanation for the phenomenon. Therefore, Salmon proposed the Causal Mechanical (C-M) Model with the causal mechanism as the core. He believed that the essence of causality was a physical process that provided explanations for the relationship between different events in space and time (Salmon 1984). Previous models confused deduction, induction, and causality and erroneously understood natural scientific and social scientific explanations as a process seeking certainty. In Salmon's opinion, the task of scientific explanation is to explain the causal mechanism of the occurrence of events. The C-M Model avoids the constraints of certainty and can explain the triggering conditions and uncertain consequences of phenomena more flexibly. However, the C-M Model is not without criticism. On the one hand, the causal process is not an isolated objective reality. It contains a series of paradigms, knowledge backgrounds and presuppositions. Ignoring the existence of these presuppositions will affect the objectivity of interpretation. On the other hand, as Pearson criticized, the explanatory power of causal mechanism models decreases as the number of factors introduced increases, thus reducing the model's ability to deal with complex realities (Persson 2012).

In general, from the pursuit of universal law to the introduction of probability and correlation and to the emphasis on causal mechanisms, the changes and improvements in the models of scientific explanations create more space for social science. By emulating natural science methodologies, quantifying facts about human behaviors and human society, and seeking rules and commonalities from them, social

science follows the principle of "hypothesis-test" and tries to obtain conclusions with universality and objectivity. However, this appropriation of methodologies has attracted much criticism over the past century because of the fundamental differences between the natural and social sciences. First, the objects of study in the social sciences are more complex in composition than those in the natural sciences and cannot be simplified as numbers and formulas. For example, rational choice theory and new institutionalism quantify individuals with free will and moral capacity as preference recipients in pursuit of profit and set the plural value systems that human beings may possess as a single sequence, which is an oversimplification of humans (Hall and Taylor 1996). Second, one of the most important criteria in natural science is the replicability of experiments. In social science research, most events are one-time occurrences. Even if control variables are introduced into the models to weaken the interference caused by errors, reproducibility in a strict sense is difficult to reach in social science research. Third, data collection in the social sciences and its dependence on historical statistical data make it hard for statistical units to meet the standard of homogenization, thus affecting their computational efficiency (Mi et al. 2018). Based on reflections of these criticisms, qualitative research on the basis of interpretivism became active again in the middle of the twentieth century, advocating a solution to the problems of the social sciences with special methods distinct from the research paradigm of the natural sciences.

2 Interpretivism and Unique Methodology for Social Sciences

Different from quantitative methodologies, qualitative studies take an epistemological anti-naturalism stance in some fundamental aspects (Risjord 2014). In their opinion, social science should not completely follow the methods and models of natural sciences but rather should establish its own research paradigm according to the specific characteristics of its research objects.

Quantitative studies based on the epistemology of naturalism and positivism seek objectivity similar to that of natural sciences. This objectivity requires that the description of social sciences be tested in the empirical world and remain neutral in value. The epistemological claims of qualitative research differ from those of quantitative research in at least two respects.

2.1 Description and Interpretation

First, qualitative researchers argue that the goal pursued by quantitative research, namely, an objective and neutral representation of empirical phenomena, is itself unattainable. In the scope of the natural sciences, meanings are often clear and explicit, leading to consensus on their measurement. For example, there are no great disputes between different researchers on the question of what volume is and how to calculate it. However, most of the widely used conceptions in the social sciences (e.g., fairness, activity, happiness, etc.) have no observable physical reality in the empirical world. There is a difference between the natural kind and the social kind. Unlike natural

classes (e.g., color, quality, etc.), social classes are nominal and rely on researchers' understandings and interpretations. How to translate these concepts depends on the different academic backgrounds, theoretical presuppositions and sociocultural environments of researchers. Even if theoretical consensus is reached, actual results differ due to measurements. For example, as one of the most widely used data collection methods in quantitative research, the validity of questionnaires depends on the assumption that different respondents have the same understanding of the questions, language and the concepts it contains so that their answers can be commensurable and compared. However, this is often not the case. How respondents understand a questionnaire depends on their own perspective and the translation of the questionnaire's language. The introduction of the questionnaire, wording of the conversation, and the time and place when the survey takes place all impact the measurement results. Thus, qualitative research distinguishes between two languages, the language of the interviewers and the language of the interviewees, and equivalence between the two languages cannot be taken for granted. The process of connecting the two languages includes the processes of interpretation by interviewers and interviewees. Since the language of the social sciences does not have the precision of the language of mathematics, the interpretation can never be neutral.

Thus, in contrast to the positivist pursuit of objectivity in quantitative research, qualitative researchers take a more interpretive stance. Interpretivism holds that there is no neutral language that can make an objective description of phenomena in the social sciences and that the translation from phenomenon to theory depends on the construction of the researchers (Risjord 2014). Moreover, social life is made up of values, norms, and principles, so one of the tasks of social science is to identify and clarify these values and systematically show the connections between them. The looping effect proposed by Ian Harkin can also illustrate this point (Hacking 1995). He believed that, in contrast to the constant laws of nature, the laws of society were in a dynamic process of constant change. Theories in the social sciences describe and provide explanations for their subjects, which in turn are adopted or rejected by their subjects, thus influencing their understanding of themselves. The self-interpreting and self-modification capacity of human subjects lead to an interaction between the social sciences and their subjects. This political nature of the social sciences inevitably makes them interact and participate with its subject—society—itself (Hacking 1995). There is no such thing as a completely neutral social science with unbiased values. The reflexivity of the social sciences determines the need to establish research paradigms different from those of the natural sciences. Therefore, common research tools in qualitative research, such as interviews, surveys, fieldwork, etc., attempt to reveal the life background and value system embedded in an answer through more in-depth communication and interaction with the respondents. That is, rather than a simple answer, qualitative research aims to reveal the whole process of "translation".

2.2 Nomothetic and Idiographic

In addition to objectivity, quantitative research aims to achieve universal validity similar to that of the natural sciences; that is, by processing data from small samples,

reliable knowledge and inferences about the whole can be obtained. This quest for universal validity is called the nomothetic methodological paradigm. Qualitative research, by contrast, is idiographic and concerned with the unique characteristics of a case in a specific situation. From a statistical point of view, the universality pursued by quantitative research depends on the premise of random sampling, which, by its very nature, requires that the sample be independent of the specific context and circumstances in which it occurs. However, what qualitative research tries to explore through in-depth field study is precisely the knowledge and the historical environment in which it is embedded (Tracy 2010).

In the opinion of qualitative researchers, the pursuit of objectivity and universality falls into the trap of positivism. Qualitative researchers hold that the social sciences should establish research objectives that are in line with their own disciplinary characteristics. Qualitative researchers regard human beings as active subjects with the capability of self-correction and creating meaning in the process of communication and interaction. Human beings and human society are in a constant dynamic process so that the knowledge of the social sciences is not static; there is no phenomenon and truth to be revealed, but instead the flow of developing and changing processes. Therefore, the goals of the social sciences should be to describe how different groups, organizations and individuals in society influence and shape each other and to focus on the "process" rather than just static conclusions. Moreover, in light of the previous discussion that it is an unattainable ideal to pursue an objective description that is completely consistent with empirical facts, proponents of qualitative methods need to come up with a new standard to answer the question "what is good social science research". Instead of evaluating by "a more accurate description of the empirical world" as traditional positivists do, qualitative researchers have again appealed to "reflexivity," defining good research as research that is more self-reflective and constantly self-revising. The method of bracketing, for example, requires researchers to recognize and reflect on their own perspective and to be alert to the presuppositions derived from the perspectives (Ahern 1999). These perspectives include one's beliefs, social classes, social roles, groups to which one belongs, and related interests. Good qualitative research requires researchers to recognize these perspectives, reduce the errors from them as much as possible, and clarify the possible impacts due to these special perspectives. As a compensatory method, however, bracketing is incomplete in the implementation of "reflexivity" because its goal is still close to objectivity. The fundamental problem is that researchers often fail to recognize their own perspectives and limitations on their own. Therefore, a more thorough and radical approach is to simply accept and acknowledge the subjectivity of research. Rather than creating a false sense of objectivity, this approach acknowledges that social scientific research inherently concerns individuals with stances and perspectives, so what needs to be done is to exhibit those perspectives and regard them as a part of the research. This kind of methodology highlights idiographic value and emphasizes that research is a product jointly created by researchers and interviewees in continuous interaction.

However, it should be noticed that although most of the literature regards interpretivism as the basic rule of the qualitative approach, exceptions do exist that integrate qualitative research and positivism together (Xie 2019). According to them,

both quantitative and qualitative approaches can be used within positivist or interpretivist paradigms.

3 Computer Simulation and Artificial Society

The criticism of positivism and the emphasis on social dynamic processes in qualitative research provide the soil for the application of complexity theory in the social sciences. Herbert Simon defines social science as hard science, for which its dynamic and interactive nature makes reductionist methods of analysis unsuitable (Ahern 1999). The method of reductionism divides the system into small subsystems through the dismantling of the system, analyzing each subsystem and synthesizing the results of analysis, through which knowledge about the whole can be obtained. However, human society is in the process of dynamic changes, so the interaction between subsystems and the interacting characteristics of the system itself will be omitted in the dismantling and restoring, which will affect the analysis results. Therefore, a system-based holistic perspective is important in dealing with the problems of human society. At the same time, the widespread application of computer technology in the 1980s provided an opportunity to develop new methods of data processing. In this context, new paradigms such as computer simulation have gradually entered the field of social science research and have been used to solve practical problems.

In general, computer simulations solve problems and draw conclusions by modeling social structures and observing individual/group interaction processes under different conditions and parameters. Different from regarding the real world as an object of observation in traditional empirical research, computer simulation technology establishes a virtual "human society". By setting parameters for the agents, the environment and the interacting rules, computer simulation makes it possible for agents to behave and interact freely in a specific environment to actualize the observation of the interacting process and results. Such a method can theoretically overcome the difficulties of reductionism and better solve the defects of the traditional research paradigm, such as the linear simplification of the functional relationship between variables and the omission of the two-way feedback mechanism. An artificial society based on computer simulation, as a parallel world in contrast with the real world, provides a new paradigm and analytical tools to deal with complex systems so that researchers can better explore the cause and effect relationships in human society.

Nigel Gilbert explained different types of simulation models to solve different problems (Gilbert and Troitzsch 2005). Examples include microanalytical simulation models, system dynamic simulation models and agent-based modeling and simulation. The microanalytical simulation model starts from the real individual and family samples, updates the sample data constantly, and obtains the overall statistical value through calculation. System dynamic simulation moves forward on the basis of microanalytical simulation to process and model the interaction process to obtain information related to interaction. However, system dynamics simulation deals with variables rather than agents, so it is difficult to obtain analysis and conclusions about

agents. To overcome this difficulty, the third type, agent-based modeling and simulation, attempts to simulate agents directly. The method is developed from cellular automata. During agent-based simulation, subjects are understood as individuals with autonomy, heterogeneity, limited rationality, interactive learning ability and different belief systems who make different choices. It sets parameters for these individuals to target and act in different environments, with reference to different rules of interaction. By controlling for time and environment, agent-based simulations are able to observe the interactions between individuals at the micro level in specific contexts and then model individual behavior in the macro system (Luo 2020).

3.1 Reform in Methodology

From an epistemological point of view, there are some core differences between computer simulation and traditional social science methods. The quantitative paradigm takes the empirical world as the only reference and pursues the objective description as closely as possible to the empirical world. Qualitative research is skeptical of our ability to attain such objectivity, shifting its focus from the empirical world to its subjective interpretation and emphasizing the inevitability of subjectivity in language. By imitating an "artificial society", computer simulation technology breaks through the boundary between the objective and the subjective, the external world and the internal self in the traditional sense, becoming a third approach independent of the first two paradigms. Specifically, the social model created by computer simulation technology is not an exact imitation of the empirical world but an artificial society "generated" by the spontaneous evolution of individuals under specific parameters and action rules. It is another dimension parallel to the real world. Its subjectivity comes from the fact that it is a completely artificial creation that has never truly happened in the real world, and its objectivity is based on the same potential as that of the real world. The epistemological basis here is that the artificial world and real world are two different contingencies from the same group of premises (parameters): "the understanding of artificial society has been moving toward the conception of 'multiverse society', which regards artificial society as a kind of reality and a possible alternative of real world or even the possible realization of society outside the earth. This connotation of artificial society is consistent with 'multiverse phenomenon' in the theory of artificial life" (Wang and Lansing 2004). With sufficiently precise presets, computer simulation and the idea of an "artificial society" can be of the same validity as a real society, even if it is virtual and "generated" by interactions between individuals under specific rules of action or even if it is entirely hypothetical. Therefore, computer simulation is not only progress of technology but also a revolution of human cognition and its relationship with the outer world. By acknowledging legitimacy to simulation, the artificial agent replaces the human, and the interactive system composed of the artificial agent replaces real society. Although the results of the model generated by the computer simulation do not represent observations in real time in a real place, they provide observations in the potential sense, thus providing a reference for policy forecasts.

In addition, unlike the reductionist paradigm of dismantling and then restoring a real-world system to reveal the underlying causality, computer simulations do not specify precise paths and outcomes, providing an open development direction for the interaction of agents in the system. Computer simulations empower the actors in the system, allowing them to interact freely according to the written rules of action and observing how they gradually change the entire social structure. The retroactive method of reductionism research is top-down, while computer simulation is bottom-up for the generation and cultivation of artificial society (Epstein and Axtell 1996). It models subjects, sets environmental parameters, directly observes communication and interactions between subjects, and shows how individual behavior at the micro level leads to systematic evolution and emergence at the macro level to understand how human society gradually grows out of predetermined basic rules. Therefore, computer simulations do not yield a certain outcome or an optimal solution but a variety of alternative probabilities and possibilities. By showing these possible options and how they relate to preconditions, computer simulations provide policymakers with technical tools to predict the effects of policies.

3.2 Difficulties and Challenges

As a new tool to analyze social science issues, computer simulation has overcome some limitations of traditional social scientific methodologies, such as the lack of data, limited data collection methods, and the inability to accurately control the external environment and carry out repeated experiments. However, even after decades of development, there are still some problems and difficulties with computer simulation technology, resulting in its failure to fully realize its theoretical assumption. First, in terms of model judgment, computer simulation is unable to effectively verify the model, which is one of the most widely disputed difficulties that simulation experiments face as a technical means. The basic assumption of positivism is to take the empirical world as the fundamental basis to test all theories. By taking the empirical world as the ruler, theories are falsifiable and thus meet the criteria of science. The methodological principles that computer simulations follow do not grant the empirical world this privilege, and convergence with the empirical world was never the goal of computer simulations in the first place. For simulations, the results point to a parallel world with the same potential rather than the empirical world itself. However, without the empirical world as a measure, how can we distinguish between the results obtained from the simulation experiment as false or an unrealized parallel possibility? The lack of measure leads to the difficulty of model validation, which hinders computer simulation from becoming an effective tool for social science research (Helmreich 2000). Second, and related to the difficulty in model validation, is the difficulty in initial parameter setting. Computer simulation is an open process that allows subjects to fully interact with each other, but the model still depends on some initial parameters, such as the subject's orientation, rational ability, value system, the basic situation of natural and social environment, the interaction rules between the subjects and so on. The subsequent free interaction is based on these parameters. How to select and set parameters often depends

on the understanding and selection of the existing research and relevant literature, which inevitably carries a certain subjective bias. The difference in parameters will significantly affect the results of the experiment. Such parameter uncertainty brings difficulties for computer simulation to provide reliable and stable experimental conclusions. At present, the common practice is to constantly change the parameters to repeat the simulation. This method can relatively reduce the errors in parameters, but it cannot completely eradicate the problem.

4 Data-Intensive Social Science

According to Jim Gray, human science is now moving toward a fourth paradigm, which is data-intensive scientific discovery (Tansley and Tolle 2009). Social science researchers borrowed this idea and began to apply big data and its technical tools to the study of human society. In 2001, when Doug Lenny analyzed the opportunities and development of data growth, he pointed out three challenges, which were later known as the 3V characteristics: Volume (increasing data amount), Variety (increasing data types), and Velocity (increasing data growth rate) (Laney 2001). Massive and constant high-speed growth of data established the requirements for hardware storage and management capabilities at that time. Later, with the continuous change of distributed storage and computing systems, clustering technology, real-time computing and streaming data processing and other technical means, the storage problem of massive data has been solved and has been gradually applied to the process of data analysis. In 2011, the International Data Corporation (IDC) proposed the feature of Value (low density of data value) to further explain the definition of big data (Gantz and Reinsel 2012). With the addition of Veracity (questionable authenticity of data), what is now known as the 5V features were formed—the widely accepted five defining elements of big data.

4.1 Reform in Methodology

4.1.1 Induction (Data-Driven) Versus Deduction (Theory-Driven)

The methodological innovation brought by big data is an "inverse operation" that lies in the shift from being theory-driven to data-driven. Traditional quantitative research usually has a theoretical framework and theoretical hypothesis and then collects data to verify the hypothesis with empirical reality. This approach, opposed to inductive reasoning, is based on deductive logical reasoning to test and falsify theories (Mahmoodi et al. 2017). However, with the comprehensive penetration of the Internet and the Internet of Things (IoT) to human life, as well as the wide use of different types of digital devices, a large amount of information about the physical world and human behavior has been digitized and recorded, providing an unprecedented volume of data for scientific research. These massive data stimulated the innovation of the research paradigm that no longer relies on special theoretical assumptions and theoretical presuppositions but rather relies directly on the data,

allowing the computer to identify and extract useful variables from a large amount of disordered data and to discover rules between them through algorithms. Similar to computational simulation, it allows for a bottom-up fashion to learn from empirical resources directly and derive theories from data. In 2008, Chris Anderson used Google as an example to explain the "end of theory" and pronounced traditional research methods obsolete. In his view, the usual hypothesis-test approach to find causality is being replaced by an approach that is free from predetermined constraints and entirely dependent on the data itself. He argues that Google does not need theoretical assumptions or models, but only through statistical analysis of large-scale data can a good match between content and advertising be achieved. In this process, "why is this page better than that page" is a question that does not need to be explained clearly, and the whole process does not need the intervention of semantic analysis or causal analysis. If the numbers say it is good, and there is enough information to track and ensure that result, then it is good (Anderson 2008).

4.1.2 Bigness Versus Representativeness

Similar with the tension between deduction and induction, another methodological shift triggered by big data is the sampling principle from representativeness to bigness. According to Mahmoodi, "representativeness does not only regards, first, the sampling of participants, but also, second, the sampling of environments participants are observed in, third, the kinds of stimuli participants are exposed to in given environments, and fourth, the states of minds and behaviors participants are able to and typically express in given environments" (Mahmoodi et al. 2017). Traditional social science researches emphasize the importance of sampling method to ensure data representativeness, while one of the core advantages of big data-driven research comes from its promise of a "full sample": it is no longer a "sample" if it is big enough to cover the "all". With new data mining methods, big data researchers expect to bypass the process of sampling and directly obtain all the data about their research subjects. However, it should be noticed that due to privacy protections and user behavior differences, it is difficult to meet the full sample requirements. For example, the use of technology such as crawlers to collect online data leaves out users who cannot use the Internet or those who are more sensitive to data privacy. From the point of view of reducing bias, full-sample data are better than sampled data, but sampled data with high quality are better than large but uneven data (Brenner and Smith 2013). Therefore, if the systematic bias in the representativeness of big data cannot be corrected, the reliability of the research results will be affected. But in an idealized design, the emergence of big data provides a completely new way to solve the sampling problem.

4.1.3 Human Reason Versus Artificial Intelligence

Although in practice over the last decade, such a methodological initiative has been considered an overly optimistic ideal due to the limitations of data quality and technical capabilities, it broadened our understanding of how to acquire knowledge as a meaningful direction. The core belief of artificial intelligence is similar

to this methodology: replacing rational human analysis process with machines and algorithms. The development of artificial intelligence has gone through different stages focusing on reasoning, knowledge and learning, while machine learning is the most cutting-edge research field with enormous potential. Among the several types of machine learning, unsupervised learning enables machines to adjust model parameters and obtain results through the process of self-training and self-learning. This process is independent of human participation and even beyond human intelligence. Specifically, the machine learning of artificial intelligence generally goes through the three stages of labeling, training and application. In the labeling stage, the machine obtains a large amount of text, pictures, voice and video information and forms a large data set through manual labeling or automatic generation. In the training stage, machines identify commonalities in the initial data set and categorize it into several rule sets by self-adjusting the parameters and models. In this process, machines iteratively improve model accuracy from data feedback. That is, when the output result does not meet the preset target, the machine will adjust the boundary conditions and rules set by itself without human intervention to realize self-production. In addition, because of the large volume of data, there are numerous calculations and human programmers do not have the ability to intervene and influence the calculation process. What happens from data input to the final output is often difficult to explain step by step. Therefore, compared with the innovations in traditional social scientific methodology realized by computer simulation, artificial intelligence takes the process a step further. It not only cancels the boundary between the real world and algorithm-based artificial society but also cancels human possession and monopoly of reason. If computer simulation redefines natural existence by giving artificial society empirical legitimacy, machine learning redefines humans by becoming an extension of human reason.

4.2 Difficulties and Challenges

The exciting prospects of the paradigm are tempered by the many difficulties of putting it into practice. The emergence of new methods of data mining and data collection, as well as the arrival of the era of big data, have provided huge amounts of data for social science research, but their quality varies, and most of them cannot be used directly for academic research. Big data research theoretically pursues the "full sample" and "self-dependent" methods, but there is still a long way to go to reach this goal.

The first problem is that of data accessibility. The demand for personal data in social science research inevitably runs into the issue of personal privacy and its protection (Kitchin 2014). For example, at which point should the mining of personal data stop? Should there be limitations and boundaries? What type of data collection requires agreements from individuals in advance? How are these agreements defined? Can digital traces with preferences and attitudes left inadvertently be considered as a form of voluntary disclosure and used directly? New technologies permeate and change the lives of individuals in a variety of ways, blur the boundaries of personal intention and subjective faults, and create new

challenges for the improvement of laws and norms. At present, there a complete legal system for the unified management of data development, use and sharing has not been designed.

A second problem is data accuracy. Traditional data collection methods, such as questionnaires and interviews, can improve the authenticity of responses as much as possible through direct communication and interaction with interviewees. However, the data collected by web data mining technology may be mixed with false, redundant and meaningless information. Because of the large magnitude of data, it is impossible to communicate and check with individual data sources, and it is difficult to understand the social environment and psychological motivation of these responses. Therefore, the processing and identification of these data are, to some extent, subjective. In recent years, the development of artificial intelligence technology has enabled it to provide text analysis and emotion recognition services to screen and clean texts, speeches, images and other unstructured data in a short period of time. However, at present, the accuracy of this processing is still limited, and errors exist.

Another widely discussed issue is the self-reinforcing effect of social discrimination on the part of machine computing without a value-correcting mechanism (Custers 1866). Big data-based AI relies on machines' self-learning, and the initial material for machine learning comes from the data left by the public on the Internet or the Internet of Things. In other words, the current human social behavior patterns are absorbed, summarized and copied by the machine algorithm and influence decision-making for future behavior. For example, the data analysis of human behavior about Internet users allows machines to easily categorize people into different groups, which motivates businesses and companies to offer different products and services to different groups of people for profit (Turow 2012). Such divisions, created and reinforced by data algorithms, emerge in areas as diverse as housing, employment, and health care and affect every aspect of human life (J. S. Winter 2014). The unwitting exposure of data, such as where you live, your consumption preferences, your schedule, or who you talk to, can be collected, recorded, and analyzed to identify you into a specific category with differentiated treatment (J. S. Winter 2015). In the name of improving efficiency and quality of service, algorithms deprive individuals of choice and create path dependence, thereby reinforcing and consolidating social stratification. On the other hand, as discussed earlier, there are problems with data integrity. To date, data capture technology is not developed enough to cover all the behaviors of all the people, and what is omitted from the data collection model accidentally is often not specific individuals but groups of people with specific characteristics. For example, Internet information is captured, and elderly groups and low-income groups without access to the Internet are structurally omitted. In the absence of structural data, treating that data as a "totality" to aid decision-making will deepen the "data gap", further isolating these excluded groups. Therefore, introducing a bias correction system to avoid the failure of algorithms in value judgments is another important challenge.

5 Conclusion

From the positivist paradigm represented by quantitative research to the interpretivist paradigm represented by qualitative research, to the emergence of computer simulation technology and the concept of "artificial society" and to the paradigm shift driven by artificial intelligence and big data, this paper systematically reviews the development of social science methodologies. On the basis of this, the paper discussed the impact of algorithms and new technology on traditional social science research methods.

Positivism advocates that social science research adopt the methodology of natural science. It believes that knowledge should be based on empirical observation and empirical data, and through technical means such as hypothesis testing and causality analysis, seeks to provide objective and neutral descriptions of the problems in human society. Different from positivism, qualitative research holds that there is no objective, neutral and value-unbiased research method. Qualitative research adopts the position of interpretationism, views society as a constantly flowing interactive process, and commits to revealing individual behaviors and meanings embedded in specific cultural environments and political contexts, recognizing that subjectivity is inevitable in research. In this context, computer simulation technology, by constructing an "artificial society" parallel to the real world, breaks through the boundary between the objective and the subjective, the external world and the internal self in the traditional sense; it is the third approach independent of the first two paradigms. Specifically, by allowing individuals within the system to interact freely under specific predetermined parameters, computer simulations generate artificial societies in a bottom-up manner and observe how this process occurs. Last but not least, artificial intelligence and intensive data-driven research, building on computer simulations, remove not only the boundary between the real world and artificial society but also the human monopoly on reason. Free from the constraints of theoretical presuppositions and theoretical frameworks, artificial intelligence as an inverse operation compared with traditional methodology, starts directly from data, through self-learning and self-adjustment, to identify and extract useful variables and discover the rules between them. In contrast to the idealized paradigm shift in theory, there are various difficulties with artificial intelligence encountered in practice. Although big data provides researchers with an unprecedented volume of data, there are still problems with its accessibility, representativeness and accuracy. The self-reinforcing nature of machine learning and analysis to establish behavior patterns has also been criticized. Only when these problems are satisfactorily solved can these technologies and new paradigms provide a solid support for academic research in the social sciences.

Declarations

References

Ahern, K.J. 1999. Ten tips for reflexive bracketing. *Qualitative Health Research* 9 (3): 407–411.

Anderson C (2008) The end of theory: the data deluge makes the scientific method obsolete. Wired Magazine. https://www.wired.com/2008/06/pb-theory/

Brenner J, Smith A (2013) 72% of online adults are social networking site users

Comte, A. 1876. *Course of positive philosophy*. London: George Bell & Sons.

Custers, B. 1866. Data dilemmas in the information society: introduction and overview. In *Discrimination and privacy in the information society*, ed. B. Custers, T. Calders, B. Schermer, and T. Zarsky. Berlin: Springer.

Epstein, J.M., and R. Axtell. 1996. *Growing artificial societies: social science from the bottom up*. Washington: Brookings Institution Press.

Gantz J, Reinsel D (2012) The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the fareast. http://www.emc.com/leadership/digital-universe/2012iview/index.htm

Gilbert, N., and K. Troitzsch. 2005. *Simulation for the social scientist*. UK: McGraw-Hill Education.

Hacking, I. 1995. The looping effects of human kinds. In *Causal cognition: a multi-disciplinary debate*, ed. D. Sperber, D. Premack, and A. James Premack, 351–394. New York: Oxford University Press.

Hall, P., and R.C.R. Taylor. 1996. Political science and the three new institutionalisms. *Political Stud* 44 (5): 936–957.

Helmreich, S. 2000. *Silicon second nature: culturing artificial life in a digital world*. Berkeley: University of California Press.

Hempel, C.G. 1980. *Aspects of scientific explanation and other essays in the philosophy of science*. Illinois: The Free Press.

Jun, L. 2020. Computing, simulation and experiment: three research methods of computational social science (计算·模拟·实验: 计算社会科学的三大研究方法). *Academic Forum* 1: 35–49.

Kitchin, R. 2014. Big data, new epistemologies and paradigm shifts. *Big Data Society* 1 (1): 1–12.

Laney, D. 2001. 3D data management: controlling data volume, velocity, and variety. *META Group Research Note 6*.

Mahmoodi, J., et al. 2017. Big data approaches in social and behavioral science: four key trade-offs and a call for integration. *Current Opinion in Behavioral Sciences* 18: 57–62.

Mi, J., C. Zhang, D. Li, and T. Lin. 2018. Fourth research paradigm: transformation of social science research driven by big data (第四研究范式: 大数据驱动的社会科学研究转型). *Academia Bimestrie* 2: 11–27.

Persson, J. 2012. Mechanistic explanation in social contexts: Elster and the problem of local scientific growth. *Social Epistemology* 26 (1): 105–114.

Risjord, M. 2014. *Philosophy of social science: a contemporary introduction*. London: Routledge.

Salmon, W.C. 1971. Statistical explanation. In *Statistical explanation and statistical relevance*, ed. W.C. Salmon, 29–87. Pittsburgh: University of Pittsburgh Press.

Salmon, W.C. 1984. *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.

Salmon, W.C. 1989. *Four decades of scientific explanation*. Minneapolis: University of Minnesota Press.

Tansley, S., and K.M. Tolle. 2009. *The fourth paradigm: data-intensive scientific discovery*. WA: Microsoft Research.

Tracy, S.J. 2010. Qualitative quality: eight 'big-tent' criteria for excellent qualitative research. *Qualitative Inquiry* 16 (10): 837–851.

Turow, J. 2012. *The daily you: how the new advertising industry is defining your identity and your worth*. Berkeley: Yale University Press.

Wang, F., and J. Lansing. 2004. From artificial life to artificial societies——new methods for studies of complex social systems (从人工生命到人工社会——复杂社会系统研究的现状和展望). *Complex Syst Complex Sci* 1: 33–41.

Winter, J.S. 2014. Surveillance in ubiquitous network societies: normative conflicts related to the consumer in-store supermarket experience in the context of the internet of things. *Ethics Inf Technolo* 16 (1): 27–41.

Winter, J.S. 2015. Algorithmic discrimination: big data analytics and the future of the internet. In *The future internet*, ed. R. Jenifer Winter and Ono,. Cham: Springer.

Xie, L. 2019. Re-discussion on the relationship between quantitative approach and qualitative approach in the field of social research. *Hebei Acad J* 39 (2): 160–170.

Xian Qin Xuan Qin is a lecturer and research fellow in the Institute for Advanced Study in Social Sciences in Fudan University. Her book, Reinforcing Authoritarianism through Democracy, was published by Palgrave Pivot in 2021. Her publications are found in journals including International Political Science Review, The China Quarterly, Chinese Political Science Review, and Japanese Journal of Political Science.