Chapter 1
Introduction: Scientific Discovery in the Social Sciences

Fernand Gobet, Mark Addis, Peter C. R. Lane, and Peter D. Sozou

Abstract Is it reasonable to talk about scientific discoveries in the social sciences? This chapter briefly reviews the status of scientific research in the social sciences and some of the arguments for and against the notion of scientific discovery in those sciences. After providing definitions of "scientific discovery" and "social sciences", the chapter notes the large variety of epistemological views and methodologies drawn on by the social sciences. It discusses the extent to which the social sciences use precise formalisms for expressing theories. Critiques of the use and reliability of the scientific method in the social sciences are discussed. In spite of these critiques, it is argued that it is possible to speak of scientific discovery in the social sciences. The chapter ends with a preview of the book.

Keywords Deconstruction · Falsification · Formal theory · Grounded theory · Information theory · Postmodernism · Pseudoscience · Psychoanalysis

1.1 Introduction

When asked about scientific discoveries, people typically think of Pasteur's discovery of vaccination, Le Verrier's discovery of Neptune and Einstein's discovery of the theory of relativity. Discoveries from the social sciences are rarely mentioned, if at all. The social sciences are also conspicuously absent from lists of the most

F. Gobet (✉)
Department of Psychological Sciences, University of Liverpool, Liverpool, UK
e-mail: fernand.gobet@liverpool.ac.uk

M. Addis · P. D. Sozou
Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science, London, UK
e-mail: m.addis@lse.ac.uk; p.sozou@lse.ac.uk

P. C. R. Lane
Department of Computer Science, University of Hertfordshire, Hatfield, UK
e-mail: p.c.lane@herts.ac.uk

M. Addis et al. (eds.), *Scientific Discovery in the Social Sciences*,
Synthese Library 413, https://doi.org/10.1007/978-3-030-23769-1_1

influential scientific discoveries in books and on the internet. However, there are many discoveries in the social sciences. Just to take the fields of linguistics and cultural studies, advances have ranged from the nineteenth century theory of laryngeals in Indo-European linguistics to the current controversy about the Luwian civilization, which offers a radical different view of the Mediterranean in the Bronze Age. Why are such discoveries missing?

Scientific discovery can be defined as the discovery of new objects, phenomena, mechanisms, cures, technologies and theories (including the unification of theories). It involves a range of activities and methods, including observations, formation of taxonomies, finding empirical rules and devising theoretical explanations (Sozou, Lane, Addis, & Gobet, 2017). Whilst philosophers of science have focused on how to best test and falsify theories, following Popper's (1959) influence, psychologists have studied the mechanisms leading to scientific discoveries. For example, Langley, Simon, Bradshaw, and Zytkow (1987) emphasise heuristic search whilst Simonton (1999) argues that search is essentially random but that successful scientists use efficient selection processes.

The social sciences can be defined as the application of scientific methods to the study of societies and the individuals within societies. They cover a wide range of fields of which a good summary is offered by the sciences represented in this book: anthropology, business, economics, law, liberal and performing arts, management, psychology and sociology. Other fields have important things to say about the social sciences – including computer science, mathematics and philosophy – as they do in this book. For example, Maymin (2011) argues that polynomial time computational complexity theory provides strong evidence for the efficient market hypothesis in financial economics being false. The social sciences use a dizzying variety of epistemological views and methodologies. In some cases, the boundaries between the natural sciences and the social sciences can be fuzzy. For instance, subfields of psychology such as the study of cognitive processes using brain imaging technologies clearly belong to the natural sciences, whereas other subfields of psychology such as psychoanalysis find a more suitable home in the social sciences.

1.2 Popper – An Incorrect Critique?

As is well known, philosopher of science Popper (1959) directed his attack against pseudo-science with a condemnation of two subfields belonging to the social sciences: Marxism and psychoanalysis. His key argument was that theories cannot be proven: they only can be refuted. However, for testing and possibly refuting scientific theories, it is necessary that these theories are formulated precisely. Popper's argument was that this was not the case with Marxism and psychoanalysis: these theories were formulated in such a way that it was always possible to generate post-hoc explanations to account for any recalcitrant empirical data. Some authors argued that Popper's criticism was incorrect and that, in fact, theories in psychoanalysis can be tested and refuted. Grünbaum (1984) showed that psycho-

analysis did make testable predictions. For example, Freudian theory postulates that repressed homosexuality leads to paranoia. Thus, a falsifiable – and, as it turns out, incorrect – prediction is that a reduction in repression of homosexuality should lead to a reduction of paranoia. However, even Grünbaum agrees that some schools of psychoanalysis develop theories that are not refutable. In a similar way, some scholars have claimed that Popper's objections to Marxism are unjustified and that theories in Marxian economics are capable of testing and refutation. Roemer (1981) formulated Marxian economics using neoclassical economics as a theory of dynamic macroeconomics processes to explain class and exploitation. Despite this, there are a number of versions of Marxism which are expressed in fundamentally ideological terms where the theories produced cannot be refuted.

It is important to note that, within individual social sciences, there is a large variability in the extent to which precise formalisms are used for expressing theories. For example, in economics, theories of microeconomics tend to be expressed mathematically, while behavioural economics tends to have more descriptive formulations. Similarly, in psychology, whilst some theories are implemented as mathematical or computer models, most are expressed informally and lack precision. By the same token, there is considerable variety in the types of methodologies used in the social sciences. In particular, in many social sciences there is a tension between quantitative data and qualitative data. For example, psychological research based on brain imaging uses vast amounts of quantitative data whilst other subfields of psychology, such as existential psychotherapy, use exclusively qualitative data such as introspection and case descriptions.

1.3 Are the (Social) Sciences Reliable?

Our brief discussion has so far assumed that the natural sciences are a good model for the social sciences to emulate – emphasis on precise theories, tests with empirical data and sound methodologies. But there has also been strong scepticism in the social sciences about the reliability of the natural sciences themselves, and indeed science in general. In the field of science and technology studies, Latour and Woolgar (1979) argued that science is not objective but socially constructed; rather than consisting of principles and procedures, science is a culture. Thus, science is not about refuting or verifying theories but about making alliances. A similarly negative analysis is that offered by Derrida's (1967) deconstruction, where the natural, "obvious" meaning of a text is critiqued and deconstructed to show that language cannot carry truth unambiguously. While influential, these postmodernist critiques cannot at face value be correct. For example, progress in technology, such as in computer science and space exploration, has produced scientific theories that are not devoid of truth nor solely socially constructed: computers can beat the best human players at chess and Go, and space probes have been sent to Mars and beyond.

It remains the case that the terms "mechanism" and "law" are frowned upon in several social science circles, primarily in anthropology and sociology, where such terms are often uttered inaudibly and with trusted colleagues. Rather than focusing on solving problems, emphasis in these fields is put on identifying, understanding and possibly deconstructing them. Some popular methodologies explicitly negate the cumulative construction of scientific knowledge. The textbook example is "grounded theory", an inductive methodology according to which pre-existing theoretical views should be ignored and qualitative data should be analysed in such a way that new theories emerge. The theories are "grounded" in the data in the sense that they are specific to these data (Strauss & Juliet, 1994). Generalisation and incremental development of theories are thus excluded by definition.

A final critique of the possibility of objective science is the idea that scientific theories reflect biases and prejudices of the individuals enjoying power, a line for example adopted by feminism with respect to "male science". Whilst it is possible that the prevailing intellectual and political environment can influence the probability of different sorts of ideas being put forward in science, the effect is likely to be stronger with respect to the kind of scientific objects studied in the social sciences.

Overall, there is no doubt that these criticisms have had considerable influence on the social sciences. At the same time it is obvious that if scientific theories cannot be built or if there is no objective reality, then it is not possible to speak of scientific discoveries. There are other consequences as well, of which two will be briefly mentioned here. The first is that the self-corrective nature of science would disappear if the criticisms mentioned above are correct. The second relates to research funding. Funders in many countries emphasise the impact of research: curiosity and creation of new knowledge is not enough, and research must have measurable implications for society, such as better health, less crime, a booming economy and improved instructional methods. One might disagree with this notion of impact, and rather argue that science should be about exploration and understanding and should not be systematically pressured to produce measurable practical applications, but the point is that, without objective experiments or observations, and indeed testable theories, it is hard to see how any impact could happen. Of course, organising a conference on discovery in the social sciences and editing a book on this topic implies that we believe that there are indeed discoveries in the social sciences. It also implies that we believe that there exists an objective reality, which might be far from our current scientific understanding, but which is there nonetheless.

1.4 Preview of Chapters

The following chapters are based on papers presented at the *Scientific Discovery in the Social Sciences* international conference, an interdisciplinary event which was organised by the editors of this book and held at the London School of Economics and Political Science on 30–31 January 2015. The conference brought together some

of the leading authorities in the field of scientific discovery in the social sciences from computer science, economics, management, philosophy and psychology.

The first part of the book deals with the type of methods – broadly construed – used in scientific discovery. In Chap. 2, Peter Abell and Maria Koumenta discuss the opposition between quantitative and qualitative methods. While causal inference has traditionally been the province of quantitative methods, Abell and Koumenta argue that qualitative causal inference has an important role to play as well. It is possible to develop causal models using qualitative data. As a specific method, they describe Bayesian Narratives based on ethnographic data. The chapter also discusses the limits of quantitative and qualitative methods.

Just like the natural sciences, the social sciences use latent constructs (or theoretical terms) to develop theories. Not directly observable, these variables are essential in making predictions. In Chap. 3, Clayton Peterson provides a philosophical analysis of one of the main methods for identifying latent variables: factor analysis. He provides a discussion of both exploratory and confirmatory factor analysis, and identifies some of the statistical and philosophical issues associated with these techniques. In the end, pragmatic considerations are emphasized in using these techniques.

In Chap. 4, Michael Stuart broadens the scope of what counts as a method of discovery and investigates what is needed for letting machines make scientific discoveries in the social sciences. Compared to the natural sciences and mathematics, the social sciences impose new constraints for scientific discoveries. As noted above, in several subfields the data are not quantitative but rather qualitative, and the difficulty is in reaching a correct interpretation. Whilst standard quantitative algorithms are ill-suited for such data, there is nothing in principle to prevent the development of new algorithms for dealing with them. Considering the kind of abilities that computers would need to successfully interpret qualitative data, Stuart concludes that there is one common prerequisite: imagination.

Chapter 5 deals with an extreme "method" for scientific discovery: fraud. Ben Trubody uses Kuhn's (1970) concept of a scientific paradigm to argue that, depending on the strength of a given scientific paradigm, it is easy or difficult to "successfully" commit scientific fraud. Whilst examples of scientific misconduct exist in all sciences, including the "hard" sciences such as physics, it is easier to commit fraud (such as omitting anomalous data points or even faking data) in the softer sciences. One reason is that the difference between contributing expertise and interactional expertise (Collins & Evans, 2007) blurs in those sciences. Several examples are provided from social psychology. To remain undetected but still have some influence on a field, fraudsters need to reach a fine balance between making reasonable claims and claims that are significant.

The second part of the book presents examples of scientific discovery in specific fields. In Chap. 6, Catherine Greene sheds new light on the role of information in financial markets. The concept of information plays a central role in the Efficient Market Hypothesis – that share prices incorporate all available information. However, Greene notes two important ambiguities in the way the concept of information is used: there is disagreement not only about the nature of information

but also about what information means for different individuals. Using Skyrms's (2013) definition that information is whatever changes probabilities, she discusses examples showing that these two ambiguities raise serious doubts about previous analyses of the Efficient Market Hypothesis.

Chapter 7 addresses the issue of scientific discovery in macroeconomics. Tobias Henschen notes that, in this field, scientific discoveries aim to explain causal relations. However, the data available are not strong enough for testing the presence of causal relations. Henschen argues that, as a consequence, it is incorrect to describe scientific research in macroeconomics as following Lakatos's (1970) notion of a research program. Rather, research in macroeconomics fits Kuhn's (1970) description of scientific research. Specifically, it is motivated by ideology and not empirical data. As Kuhnian logic has motivated many policies that turn out to be failures, Henschen argues that macroeconomics should use a different type of logic – a variant of Popper's (1963) situational logic.

In Chap. 8, Jakub Motrenko first discusses why sociologists rarely talk about scientific discoveries in their field. He then argues that scientific discoveries do indeed happen in sociology, and typically take the form of "That-What discoveries" (Kuhn, 1962). With such discoveries, a new object is found first, and a theory is developed to explain it. Adapting Kuhn's notion to sociology, Motrenko applies it to the study of Solidarność (Solidarity), the Polish trade union which successfully challenged Poland's communist government in the 1980s. He shows how research carried out by Polish sociologists can be described as That-What discovery.

The third and final part of the book deals with formalising theories in the social sciences. In Chap. 9, Maria Dimarogkona, Mark Addis and Petros Stefaneas argue that the theory of institutions offers a powerful means for formalising theories in the social sciences. This theory, originally developed in computer science for addressing the problem of the vast number of logics developed in this field, has the advantage that it can be used independently of the specific nature of the underlying formal language. It thus presents an elegant solution to the philosophical debates about the relative merits of the syntactic and semantic views of scientific theories. The authors discuss several advantages of using this approach for formalising scientific theories in the social sciences, including formalising systems that use a different theoretical vocabulary and facilitating analysis at different levels of abstraction.

Building computational models is an important way of developing theories in psychology and cognitive science. Such models offer important advantages, such as precise predictions and the ability to simulate the behaviour under study, whatever its complexity. However, this endeavour is notably difficult and time consuming. In Chap. 10, Mark Addis, Fernand Gobet, Peter Lane and Peter Sozou describe a novel way to develop computational models, whereby computational models are semi-automatically generated by genetic programming algorithms, which use analogues of Darwinian selection processes. The chapter also addresses several philosophical issues raised by this approach, including the nature of the explanations proposed by this approach and how they relate to the notion of fictional modelling.

While Chap. 10 focuses on data obtained under experimental control, Chap. 11 considers the use of correlational data, which are omnipresent in the social sciences.

Specifically, Pat Langley and Adam Arvay focus on modelling research aimed as inducing numeric laws. Whilst early research focused on finding mathematical relationships between variables independently of domain knowledge, more recent efforts have aimed at using such knowledge for explaining data, in particular with the use of multivariate time series. Applications have been mostly carried out in biology and ecology, and Langley and Arvay discuss ways similar methods could be used in the social sciences. These methods offer the prospect of discovering new explanations for data that characterize dynamic systems evolving as a function of time.

Acknowledgments The *Scientific Discovery in the Social Sciences* international conference and the preparation of this book was supported by Economic and Social Research Council grant ES/L003090/1.

References

Collins, H. M., & Evans, R. (2007). *Rethinking expertise*. Chicago: University of Chicago Press.
Derrida, J. (1967). *De la grammatologie*. Paris: Les Éditions de Minuit.
Grünbaum, A. (1984). *The foundations of psychoanalysis. A philosophical critique*. Berkeley, CA: University of California Press.
Kuhn, T. S. (1962). The historical structure of scientific discovery. *Science, 136*, 760–764.
Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
Lakatos, I. (1970). Falsification and the methodology of scientific research programs. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge, UK: Cambridge University Press.
Langley, P., Simon, H. A., Bradshaw, G. L., & Zytkow, J. M. (1987). *Scientific discovery*. Cambridge, MA: MIT press.
Latour, B., & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. Los Angeles: Sage.
Maymin, P. (2011). Markets are efficient if and only if P=NP. *Algorithmic Finance, 1*, 1–11.
Popper, K. (1959). *The logic of scientific discovery*. New York: Basic Books.
Popper, K. (1963). *The poverty of historicism*. London: Routledge.
Roemer, J. (1981). *Analytical foundations of marxian economic theory*. Cambridge, UK: Cambridge University Press.
Simonton, D. K. (1999). *Origins of genius*. Oxford, UK: Oxford University Press.
Skyrms, B. (2013). *Signals: Evolution, learning and information*. Oxford, UK: Oxford University Press.
Sozou, P. D., Lane, P. C., Addis, M., & Gobet, F. (2017). Computational scientific discovery. In L. Magnani & T. Bertolotti (Eds.), *Springer handbook of model-based science* (pp. 719–734). New York: Springer.
Strauss, A., & Juliet, C. (1994). Grounded theory methodology: An overview. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (1st ed., pp. 273–284). Thousands Oaks, CA: Sage.