Model-Based Demography: Towards a Research Agenda

Daniel Courgeau, Jakub Bijak, Robert Franck, Eric Silverman

Abstract   This chapter aims to contribute to the debate on the role of model-based approaches, such as agent-based modelling, in the future of demography. First we call attention to the developments of the discipline since the 17th century, and we describe its four successive paradigms related to the period, cohort, event-history and multilevel perspectives. We argue that these paradigms are complementary and that demography, since its beginnings, has subscribed to the classical scientific research programme launched by the promoters of modern science. Next, we examine how simulation modelling developing in population sciences recently, may help to respond to three main challenges: how to overcome complexity in social research; how to reduce its uncertainty; and how to reinforce its theoretical foundations. We sketch a model-based research programme for demography, looking specifically at interactions between various population systems. We then show how this approach might conform to the classical scientific research programme, in order to take advantage of its benefits.

1. Introduction

Over its 350-year history, demography has progressed through successive paradigmatic changes, from period analysis (Graunt 1662) to multilevel analysis in the more recent period (Courgeau 2007). Currently, the prominence of agent-based models (ABMs) has indicated an increased focus on individual behaviours and in-

D. Courgeau
Research Director Emeritus, Institut national d'études démographiques, Paris, France

J. Bijak (✉)
Social Sciences, University of Southampton, Southampton, United Kingdom
email: j.bijak@soton.ac.uk

R. Franck
Professor Emeritus, Université catholique de Louvain, Louvain-la-Neuve, Belgium

E. Silverman
School of Computing, Teesside University, Middlesbrough, United Kingdom

teractions in the study of populations, and also a desire to bolster the theoretical foundations of demography (Burch 2003ab; Silverman, Bijak and Noble 2011). Here we posit that ABMs have a potential to become a manifestation of a broader, model-based research programme, which would be much more heavily reliant on computer simulations as a tool of analysis. The key advantage of such methods is that they allow examining interactions between various elements of complex population systems. In our view, such a model-based approach, while firmly rooted in the multilevel paradigm, can form the foundation of the next step in the cumulative progression of demographic knowledge.

This chapter proceeds first by detailing the successive paradigmatic changes evident in the history of demography in Section 2, and then by describing the challenges of studying uncertainty, complexity and interactions in population systems in Section 3. In Section 4 we discuss the ways of conforming computational methods to the classical scientific programme outline, and make the case for model-based demography as a new research programme for the discipline. Finally, in Section 5 we propose a research agenda to address the challenges ahead.

2. Cumulativity in Demography

Since the origin of demography in the 17th century, the field has progressed through a series of paradigmatic changes. Here we use the term paradigm in a somewhat different sense from Kuhn (1962), and from its current usage. We want to point out the methods by which the phenomena observed within a population have been related to the set of key parameters (fertility, mortality and migration) used in demography to explain population growth, decline or stabilization. There are four main methods, each implementing a limited scope of notions which we call the paradigm of the selected method (Courgeau and Franck, 2007). In this section we point to the differences between these four paradigms, and to their possible complementarity.

It is important to recall the path taken by Bacon in 1620 in his elaboration of an inductive method for scientific thought. He presented it in contrast with the dominant way of thinking in fashion at this time (Bacon, 1620, aphorism 19):

> There are and can be only two ways of searching into and discovering truth. The one flies from the senses and particulars to the most general axioms, and from these principles, the truth of which it takes for settled and immovable, proceeds to judgement and to the discovery of middle axioms. And this way is now in fashion. The other derives from the senses and particulars, rising by a gradual and unbroken ascent, so that it arrives at the most general axioms at last of all. This is the true way, but as yet untried.

The first way generates what Bacon called the four Idols, where axioms are not grounded on a meticulous observation of the properties of nature to be studied, but rather on prejudices – unverified notions of human understanding. As

Bacon said[1], such axioms should not "avail for the discovery of new works, since the subtlety of nature is greater many times over than the subtlety of argument". As already argued elsewhere (Courgeau et al. 2014) 'Idols' may exist in various areas of contemporary population sciences, for example in the form of behavioural genetics, postmodern theory, hereditarism, or modern hermeneutics.

On the other hand, the Baconian "second way" became the modern scientific way of thinking, rising from a meticulous observation of studied facts to the "formation of ideas and axioms by true induction". This method of induction[2] consists of discovering the principles – the 'first' axioms, the 'lesser', and the 'middle' in Bacon's terms – of natural or social properties by way of experimentation and observation. The Baconian induction rests on the requirement that without these principles the properties observed would be different (Franck, 2002).

Graunt (1662) was the first to apply this method for the study of human populations. He no longer considered that phenomena such as births, illnesses and deaths were to be seen as God's secret and therefore out of bounds to scientific scrutiny. He studied each event not as a unique one but as one occurring to a statistical individual, with only a few characteristics. These abstract events became fertility, morbidity and mortality, and lost any direct attachment to a given individual. This was the only way to begin a scientific study of population, called by Petty (1690) *political arithmetics*, which prevailed for around 200 years. Graunt's research paved the way for demography, epidemiology, political economics, and for population sciences more generally.

Graunt's demonstration of the links between probability and population science was also vital. Probability was first addressed in 1654 by Pascal and Fermat, but their results were published later (Pascal, 1665), and it was in fact Huyghens (1657) who first published a treatise on games, with a foundation for objective probability. Graunt uses this concept in order to estimate the population of London from the number of deaths, using an estimation of the probability of dying (Courgeau, 2012). The probability of an abstract event in a human life was used for the first time, facilitated by the notion of a *statistical individual*.

We can conclude that the population sciences were without a doubt born in England, and subsequently led to a more general school of scientific thought on population problems. From the end of the 17th and throughout the 18th century, this way of thinking developed through the work of many leading European researchers such as Halley, Süssmilch, Euler, Moheau, and so on (Courgeau, 2007).

During the 18th century a new concept of epistemic probability was introduced, first by Bayes (1763), and then refined by Laplace (1774, 1812). In fact, the objective probabilistic approach was already showing how through successive trials, the estimated frequency tends towards such a probability, as is the case for fair games for which we can determine an *a priori* probability. However, as had been already recognised in the preceding century, such a hypothesis was difficult

---

[1] Citations in this and in the next paragraph come from Bacon (1620), aphorisms 24, 39 and 40.

[2] *Induction* is not taken in the sense of Mill (1843) and his followers, i.e. generalisation from particular facts. In Bacon's sense, induction designates the complete research process (Section X.4).

to justify for events in human life. A new approach was necessary for such events, where all we know is the sample observed. Not only is the population from which it is drawn unknown, but its very existence is a hypothesis. By using a *prior* probability in order to estimate a unique posterior one, the epistemic approach permitted answering these questions clearly. Laplace applied it to many phenomena, including a number of population science ones.

In 1809, in Germany, Gauss proposed the method of least squares, which was mainly used in astronomy at this time. Following the work of mainly British but also German and French biological and social scientists[3], it became widely used: by Galton and Pearson in population genetics; Lexis and Quetelet in demography; Edgeworth in statistics; Durkheim in social demography; and Yule in economic demography. Finally, at the beginning of the 20th century Fisher, a statistician and population geneticist, developed the maximum likelihood theory and theory of statistical inference (see Courgeau, 2012). Population science was coming back to an objective approach, as the development of censuses through the 19th century permitted the use of exhaustive samples.

The problem with many statistical tools from the 19th century, such as the least squares method, is often that they assume a particular mathematical structure among a limited set of macro-variables, irrespective of whether that structure exists in the real world. This was already the case for Durkheim's study of suicide in Prussia (1897). This may lead to what is called an *ecological fallacy*, meaning that aggregate data, as a rule, cannot be used to study individual behaviour. The only instance where this is possible is when the probability of experiencing the event is independent of the area studied and when the population is large enough to cancel out any random difference that may appear (Courgeau, 2007).

Another issue here is related to the type of observations, which are exclusively *cross-sectional or period-based*. After Courgeau (2007), we can conclude that the paradigm of the cross-sectional approach may be defined as follows: the social facts of a period exist independently of the individuals who experience them, and can be explained by various characteristics of the surrounding society, such as economic, political, religious, or social aspects. This cross-sectional paradigm prevailed in demography till the end of World War II.

The next change came from the US, where population scientists set up a new perspective of *cohort analysis*, following the pioneering sociological work by Mannheim (1928), which introduced the individual's lived time; Whelpton (1949) and Ryder (1951) were the first to promote this approach, and Henry (1959) formalised its theoretical underpinnings. The resultant paradigm is defined by the following postulate: "the demographer can study the occurrence of only a single event, during the life of a generation or a cohort, in a population that preserves all its characteristics and the same characteristics for as long as the phenomenon manifests itself" (Courgeau, 2007, p. 36). We will not go further into this approach, as "for the analysis to be feasible, the population must be regarded as homogeneous

---

[3] These fields were not so clearly defined at this time: scientists were working in different social or biological sciences and in statistics simultaneously.

and the interfering phenomena must be independent of the phenomenon studied" (Courgeau, *idem*, discussing Henry, 1959, and Blayo, 1995). These conditions are restrictive, and led to a new approach permitting us to set such hypotheses aside.

To be able to consider heterogeneous cohorts and to introduce dependencies between phenomena, it became necessary to introduce statistical methods able to analyse different processes simultaneously and look at numerous characteristics of the studied individuals. The general theory of stochastic processes was first developed by the US statistician Doob (1953) and was applied to demographic processes by Aalen (1975) in Sweden. In demography, it was incorporated through the introduction of *event-history analysis* (Courgeau and Lelièvre, 1992).

In the event-history paradigm, "individuals follow complex, life-long trajectories that depend, at a given instant, on their earlier trajectories and on the information they had acquired in the past" (Courgeau, 2007, p. 58). We can identify the factors at work – both demographic and non-demographic – and analyse their effect on individual behaviour in more detail. In order to do that, it is necessary to use surveys that follow individuals along a large part of their life and to collect data on events, and on the characteristics, fixed as well as time-dependent, which may affect these events. However, here we cannot view an individual trajectory as the outcome of a process specific to each person. As we observe only a single outcome (the individual trajectory), the process is not identifiable.

In this case, we must adopt a collective point of view: all individuals are assumed to follow the same random process, the parameters of which we can estimate from the observation of a sample of individuals with their own characteristics. At first glance, this assumption seems quite bold. However, it is important to realize that this is not a hypothesis about observed people, but about the construction of a process underlying a set of trajectories. In this case, two observed individuals do not necessarily follow the same process, whereas two statistical individuals with the same characteristics do so automatically, as random sampling units with identical characteristics (subject to identical selection conditions). Such an approach again may require adopting a Bayesian point of view (Ibrahim et al., 2001), as it looks at many characteristics measured on a sample of restricted size.

However, the use of individual characteristics may lead to an *atomistic fallacy*, as opposed to the *ecological fallacy* of cross-sectional studies. By concentrating on individual characteristics, we disregard the context in which human behaviours occur. As noted by Courgeau (2007), context clearly may influence individual behaviour, and therefore isolating individuals from the constraints imposed by the social networks of the living environment seems misleading.

We must then introduce the different types of groupings of individuals found in all human societies: social groupings, such as the family, networks of contacts (or, more generally, social networks), etc.; economic groupings, such as the firm or the organisation where a person works; educational groupings, healthcare groupings; political groupings; etc. In order to consider not only the individual but different groupings we must develop new methods of *contextual and multi-level analysis*. These models have been elaborated by American (Mason et al., 1983) and English (Goldstein, 1987) researchers.

Multilevel approaches have permitted us to solve the apparent contradiction between aggregate models and the individual, event-history perspective. Thanks to their properties, we can combine the results of the analyses at the aggregate and individual level by clarifying the apparent paradox between them. As observed by Courgeau (2007, pp. 79–80):

> The new paradigm will therefore continue to regard a person's behaviour as dependent on his or her past history, viewed in its full complexity, but … this behaviour can also depend on external constraints on the individual, whether he or she is aware of them or not.

This paradigm allows for removing the two fallacies mentioned before (*idem*):

> The ecological fallacy is eliminated, since aggregate characteristics are no longer regarded as substitutes for individual characteristics, but as characteristics of the sub-population in which individuals live and as external factors that will affect their behaviour. At the same time, we eliminate the atomistic fallacy provided that we incorporate correctly into the model the context in which individuals live.

As we have demonstrated previously demography has advanced effectively thanks to the introduction and refinement of successive paradigms. Each paradigm takes the shortcomings of its predecessors as a starting point and offers a method for surmounting them – without, however, erasing all the knowledge attained through earlier paradigms. Indeed, for some questions that a population scientist may wish to ask, cross-sectional analysis can suffice just as any other form of analysis may be sufficient for other issues. The same is true for some questions asked by the physicist that may be answered perfectly by Newtonian physics, without taking into account Einstein's physics.

However, in demography these developments have not led to a patchwork landscape of competing approaches, but instead to a *cumulativity* of knowledge, despite being far from linear. This is because different paradigms take a different point of view on the studied phenomena, partly preserving some of the results of the previous ones, as the multilevel analysis compared with cross-sectional and event history analysis. As Courgeau (2012, p. 239) has put it:

> Cumulativeness of knowledge seems self-evident throughout the history of population sciences … the shift from regularity of rates to their variation; the shift from independent phenomena and homogeneous populations to interdependent phenomena and heterogeneous populations; the shift from dependence on society to dependence on the individual, ending in a fully multilevel approach. Each new stage incorporates some elements of the previous one and rejects others. The discipline has thus effectively advanced thanks to the introduction of successive paradigms.

Each of the four paradigms frames the relationship between observations and scientific object differently, and in so doing allows for new methodologies that can alleviate difficulties associated with other methods, as summarised in Table 1. The scientific objects of enquiry in population sciences, such as mortality, fertility, nuptiality, migration and so on, are independent of the theory used to treat them. By contrast, the relationships assumed to exist between these objects are strongly dependent on the key theory underpinning each paradigm: independence between them in cohort analysis, heavy dependence between them in event history

analysis. Yet, as argued before, each paradigm also occupies a different context, and therefore previous paradigms remain relevant despite the proliferation of new ones.

Table 1 The four paradigms of demography – a summary

No.	Paradigm	Period	Key focus
1	Period (cross-sectional)	1662–	Population-level (macro) phenomena, observed and measured according to the historical time
2	Cohort (longitudinal)	1950s–	Population-level phenomena, observed and measured along the lifetime of individual cohorts
3	Event history	1980s–	Individual-level (micro) phenomena, observed and measured according to the individual time
4	Multilevel	1980s–	Individual, population, and interim-level phenomena, observed and measured from multiple perspectives

The evolution of successive paradigms is an ongoing process, and the paradigms themselves are in a constant need of improvement and refinement, in order to be able to answer emerging research questions. Even the multilevel approaches do not address questions related to interactions between various elements of increasingly complex population systems. In particular, micro-level rules may be hardly linked with aggregate-level rules, while macro-level rules cannot be modelled exclusively with an individual approach, since they transcend the behaviour of the component agents (Holland, 1995). As Conte et al. (2012, p. 336) said, in their *Manifesto of computational social science*, such a micro-macro link:

> … is the loop process by which behaviour at the individual level generates higher-level structures (bottom-up process), which feedback to the lower level (top-down), sometimes reinforcing the producing behaviour either directly or indirectly.

We will add that in some cases it can go in the opposite direction of the producing behaviour, leading to "perverse effects" as shown by Boudon (1977).

We must go further, however, as the effects of aggregation levels are always defined with respect to the individual. For example, a series of individual actions in a community may foster awareness of a problem that concerns the entire community. This may lead to political measures, taken at more aggregated levels. These measures will naturally affect individual behaviours, generating new actions to offset their perverse effect, and so on. The multilevel approach as described above does not allow for inclusion of this two-way flow. More generally it is necessary to identify the different levels as truly different systems of agency, i.e. of collective action with different goals, specific resource interdependencies between members and specific social processes that help members to manage dilemmas at each level. We will see in the following sections, how a model-based research programme may answer these challenges.

3. From Empirical to Model-Based Demography, and Back: Uncertainty, Complexity and Interactions in Population Systems

The recent evolution of demography and population studies has coincided with shifting perspectives on the epistemological challenges facing the studies of human populations. In particular, demographers are now paying ever more attention not only to different levels of analysis, but also to the uncertainty and complexity of population phenomena, which are discussed in this section.

Demographic phenomena – as all other aspects of social reality – are inherently *uncertain*, but to a slightly lesser degree than is the case in other areas of social sciences, such as sociology or economics. This comparative advantage of population science is largely due to the strength of the underlying relationships, such as population accounts and persistence of demographic patterns in time, and is helped by the strong empirical slant of population science (Xie, 2000; Morgan and Lynch, 2001). Still, particular areas of demographic interest differ with respect to their uncertainty: out of the three main components of population change, mortality is usually thought to be the least uncertain, while migration is the most (e.g. NRC, 2000). The explicit acknowledgement of the uncertainty challenge has led to a renaissance of statistical demography since the 1980s, and to the "return of the variance" to demography – an important methodological perspective for all four paradigms mentioned above (Alho and Spencer, 2005; Courgeau, 2012)[4].

Uncertainty is vastly augmented by social reality becoming increasingly *complex*. Hence, appropriate tools are required to analyse the associated complexities in more depth. In demography, the debate on the complexity versus the parsimony of demographic models has been present especially in the context of predictions (Ahlburg, 1995; Smith, 1997; Lutz 2012). However, the evidence regarding the relative performance of models of varying complexity is inconclusive. For predictive applications it may be tempting to apply Occam's razor and opt for simple models that describe the uncertainty relatively well (Bijak, 2010). On the other hand, despite its importance, prediction is not the only goal of enquiry in population science (Xie, 2000). If the perspective shifts towards explanation, exploration, or other non-predictive applications, a different approach is required[5].

From a statistical point of view, *model uncertainty* needs to be acknowledged as well (Raftery, 1995). If the models themselves are to be formally recognised as yet another source of uncertainty in population studies, next to the underlying processes, parameters, and inherent randomness, the most natural and coherent way of describing all these sources is via Bayesian statistical inference and epistemic probability. Within the Bayesian paradigm there exist several ap-

---

[4] Similarly, acknowledgement of the role of space in demography has led to the multi-regional perspective within the cohort paradigm (Rogers 1975), later extended to the multi-state case.

[5] See Epstein (2008) for "sixteen reasons other than prediction to build a model". Conte et al. (2012) highlight the capability of "generative" models to reproduce qualitative regularities observed in the real world (the stylised facts).

proaches to model error: from a formal *model selection* out of several competing possibilities, and the related *model averaging* (Raftery, 1995); to including an additional *model discrepancy* (*inadequacy*) term in the modelling process (Kennedy and O'Hagan, 2001). In addition to the appealing prospect of reconciling quantitative and qualitative information in a formal way, Bayesian statistics allows for the inclusion of subjective opinion in the process of statistical inference.

On a larger scale, Bayesian statistics also provides a possible way of reconciling the empirical and computational approaches by returning to empiricism, yet at a different level of analysis. All computational models, no matter how complex, have inputs (parameters) – and outputs (quantities of interest). Their mutual mapping enables statistical analysis. There are techniques available for this purpose, chiefly *Bayesian melding* (Poole and Raftery, 2000), and approaches based on *Gaussian process emulators*, also Bayesian (Kennedy and O'Hagan, 2001; Oakley and O'Hagan, 2002). Both have already been prototyped in demographic applications – the former by Alkema et al. (2007) and Clark et al. (2012), and the latter by Bijak et al. (2013), Silverman et al. (2013), and Hilton and Bijak (Chap. X, this volume). The application of such methods allows for analysing the properties of complex computational models within a formal statistical framework, which would not be possible with more traditional approaches.

As demography has started incorporating insights regarding its own epistemological limits, new approaches to modelling have begun to flourish. The perspective of population science becoming a *model-based science* (Burch, 2003b) has become appealing[6], mirroring similar movements within the study of biological systems and evolution (Levins, 1966; Godfrey-Smith, 2006). As argued by Xie (2000), there are certainly insights to be gained from examining the successes and failures of modelling efforts in population biology (see also Silverman, 2016).

Previous efforts have outlined various approaches toward modelling the *complexity* of population processes, amongst which we can identify two broad trends: *social simulation* and *systems sociology* (Silverman and Bryden, 2007; Silverman, 2016). The former is concerned with the application of novel modelling techniques, primarily agent-based models, to specific populations and situations. The latter is a primarily theory-driven enterprise, investigating the consequences of various foundational social theoretic positions – along the lines of the 'opaque thought experiment' role for simulations proposed by Di Paolo et al. (2000). Within demography and population sciences, the desire to remain empirically relevant – and to strengthen that relevance through more reliable and nuanced predictions – has led to a focus on social simulation more than systems sociology approaches. Micro-simulations, based on empirical transition rates or probabilities for *simulated (virtual) individuals* (Willekens, 2005), clearly belong to this class.

Within simulation approaches, we also need to distinguish between *weak simulations* and *strong simulations* (Huneman 2014)[7]. Weak simulations serve to

---

[6] Burch (2003b) points to Nathan Keyfitz (1971) as the pioneer of the model-based demography.

[7] Following Huneman (2014), we give these terms slightly different meanings than for example Thagard (1993, p. 6), for whom the weak simulation is "a calculating device drawing out the

test some theory or hypothesis, when the system studied cannot be easily modelled by mathematics or when data are limited or unavailable. They are top-down models, which start from setting the hypotheses and assumptions. Strong simulations, on the other hand, aim to "explore the possible outcomes of a simple model" without any reference to a pre-existing theory or hypothesis (*idem*, p. 72). Many existing agent-based models often proceed in this way, where simulations are used with no pre-existing theory to explain the modelled phenomena, but only some intuitive rules. These models are built from the bottom-up: low-level interactions are supposed to produce high-level complex behaviour. As argued by Conte et al. (2012), such 'generative explanations' are often arbitrary – they also suggest that simulation models need to become much more empirical, in order to provide solid micro-foundations for the social mechanisms they attempt to model.

The presence of emerging properties and of 'downward feedback' or causation (from macro to micro) in complex models means that we cannot obtain the macro-level patterns by simply aggregating the micro-level outcomes. Instead, we need to model both levels jointly. Therefore, from the point of view of the demographic paradigms, we remain firmly within the realm of the multilevel analysis, only using different tools (simulations) to explore multiple layers of population processes at the same time. Conte et al. (2012, p. 342) suggested that:

> …simulations must be accompanied by micro-macro-loop theories, i.e., theories of mechanisms at the individual level that affect the global behavior, and theories of loop-closing downward effects or second-order emergence.

A part of the strength of simulations lies in a potentially wide variety of ways to represent the same problems using a relatively simple set of techniques. However, there is a real danger that the models can be constructed in an arbitrary way, not linked to the observations of the properties of the population systems of interest, and thus become manifestations of Baconian 'idols'. This problem can be exacerbated if the models lack an explicit documentation of their construction and core assumptions when simulation results are presented[8]. In such situations, even models with well-grounded and well-justified assumptions, may seem arbitrary.

Agent-based models are capable of analysing systems of interacting elements through computational modelling. A part of the appeal of such models is their capacity for explanatory power (see Burch 2003a,b; Silverman et al., 2011). As such, agent-based models by their very nature are intended to represent the import and impact of individual actions on the macro-level patterns observed in a complex system, and vice versa, showing a potential promise to transcend different levels of analysis. Such methods can further theoretical understanding of population processes (Burch, 2003a; Chattoe, 2003), and using these methods to break from the over-reliance of some micro-simulation models on empirical data at the

---

consequences of mathematical equations that describe the process simulated," while a strong simulation "itself resembles the process simulated" (see also Brenner and Werker 2007).

[8] For a discussion of the ABM documentation standards, and the ODD framework ("Overview, Design concepts and Details"), see Grimm et al. (2006), as well as Chap. X in this volume.

expense of reasonable theoretical explanations and mechanisms[9] (Silverman et al., 2011). As mentioned earlier, however, to take a full advantage of this potential, we need to look at how these different levels of aggregation interact, in order to better explain social facts. Simple aggregation of individual-level rules to generate and validate macroscopic patterns – as often implicitly done in existing agent-based models – is not sufficient (Conte et al. 2012).

In population sciences, there are many systems comprised of *interacting individuals, groups, or institutions* which are worthy of enquiry. Population sciences can become model-based by making those interactions between different levels in population systems an explicit object of interest. In so doing, our models would become capable of representing complex, interacting behaviours at various levels, and investigating the roles of different elements of population systems in shaping the observed demographic outcomes. Such models of multilevel interacting systems would have clear potential for contributing to theory-building within population sciences, and perhaps even social science more broadly.

Recent years saw an ever-increasing interest amongst population scientists in new modelling methodologies for complex social realities, many of these inspired by agent-based computational approaches (see Billari and Prskawetz, 2003; Aparicio-Diaz et al., 2011; Kniveton et al., 2011; Willekens, 2012; Bijak et al., 2013; Silverman et al., 2013). The movements toward computational complexity have been matched by a shift coming from the other direction, as agent-based modellers have branched out into areas traditionally covered by statistical approaches in population science (see e.g. Axtell et al., 2002; Geard et al., 2013).

Of course, model-based approaches come with their own shortcomings – in particular, models attempting to represent the complexities of particular population systems are naturally dependent on sensible theories regarding these systems, and on their representation. However, such theories are not only many and varied, but can be notoriously difficult to formalise (Klüver et al., 2003), and validate[10], especially in social science realms (see Moss and Edmonds, 2005). Without such theories, it may be difficult to build an adequate model of the systems under study. A possible way forward from this conundrum is to reconnect to the classical research programme which promotes some sort of *functional-mechanistic analysis* (Franck 2002a); this will be discussed in the next sections of this chapter.

A clear strength of population science, and one of the keys to its success, is its applied character, responding to the direct needs of policy makers (Xie, 2000; Morgan and Lynch, 2001; Hirschman, 2008). The methodological developments outlined above can only further this practical, utilitarian aspect of demo-

---

[9] The problem here is not the empirical basis of such models – quite the contrary – but unrealistic mechanisms. Particularly problematic are Markovian assumptions of the lack of memory, where simulations are based on homogenous matrices of transition probabilities. Examples of microsimulation models that allow for heterogeneous transition patterns or mechanisms, e.g. of partnership formation, include SOCSIM (http://lab.demog.berkeley.edu/socsim/).

[10] After Franck (2002a), we interpret validation as a continuous process, rather than an achievable state.

graphic enquiries. The Bayesian approach naturally allows formal statistical decision analysis, which can offer practical support to various decisions which require numerical input, for example for planning purposes (Alho and Spencer, 2005; Bijak, 2010). On the other hand, the model-based approach, especially coupled with statistical analysis, allows the decision makers to trial a range of policy "levers" in a simulated environment. Such experimentation *in silico* would consist of generating coherent scenarios, where mechanistic rules governing the behaviour of *simulated individuals* would be coherent with the empirical patterns for *statistical individuals* observed through a scientific lens (Courgeau, 2012).

Demography needs more simulations to be able to answer new research questions, but in order to suit the goals of the discipline, such simulations would need to be grounded in the observables, and the models would need to be built inductively (bottom-up), rather than starting from hypotheses and assumptions. To address this challenge for the future of demography and population sciences we propose a *model-based* research programme, firmly rooted in the wider *functional-mechanistic* approach. If agent-based models, as introduced above, are to belong to this programme, they need to be empirically based and scientifically rigorous.

As a part of this research programme, we posit that demography should investigate the *interactions* between various population systems and the functional *mechanisms* behind them. The interactions and mechanisms are best described by formal models based on data and theory-based rules, derived from observations of system properties by following the Baconian inductive method. This approach can augment the capabilities of the multilevel paradigm, whilst broadening the scope of scientific exploration in demography. In particular, it can enable population sciences to enhance the theoretical base of the discipline, whereby theories represent formal conceptual systems rather than necessarily empirical ones (Franck, 2002a; Burch, 2003b).

4. Conforming Model-Based Approaches to the Classical Scientific Programme Following the Baconian Inductive Method

How may the *model-based* approaches we propose for demography conform to the classical scientific research programme? This programme is at present generally ignored by social scientists – as well as overlooked by philosophers of science – because it has been distorted by the empiricist tradition in philosophy, where the empiricism promoted by David Hume and John Stuart Mill has substituted for the classical empiricism. Francis Bacon's programme was shared by the other pioneers of modern science, Galileo, Descartes, Newton, Huyghens, Graunt, and others. This programme tacitly continues to guide research in the natural sciences today just as in the past, yet in the social sciences it is often abandoned.

To start with an illustrative example, consider the famous model of a neuron (McCulloch and Pitts, 1943), the ancestor of the now-ubiquitous artificial neural networks, which is shown in Figure 1.
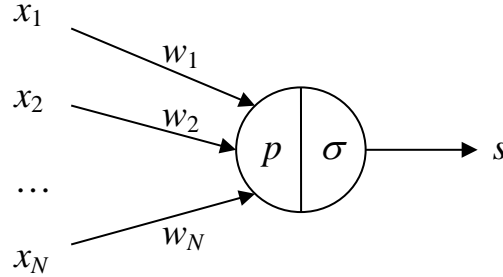

Fig. 1: The McCulloch and Pitts (1943) model of a single neuron

Source: Adapted from Franck (2002a: 143), with permission of Springer

The model in Figure 1 represents one neuron. Yet what is represented? Not the soma, nor the axon, nor the dendrites, nor the gene nuclei, nor the membrane, nor the shape of the neuron, nor the way that the various parts of the neuron work together. Starting from the observation of some main properties of the neuron, McCulloch and Pitts tried to represent its *functional* architecture, *without which these properties could not come about as they do*.

Five functions were identified: receiving the stimuli x_1, x_2, … x_N; weighting them by synaptic coefficients w_1, w_2, … w_N; calculating the sum of weighted stimuli received (p); fixing a threshold of stimulation (σ) below which transmission does not occur; and finally computing the exit signal s. These functions are arranged in a specific order: the weighting of stimuli must precede the calculation of the weighted sum, and so on. Thus, more precisely, the McCulloch and Pitts model represents *the functional structure of the process generating the observed outcomes* of the neuron. Note that such a model ignores the combination of factors or *causes* that fulfil these functions – it is wholly conceptual.

Let us now have a look at *reverse engineering*[11]. Reverse engineering follows a similar path: inducing the design of a device from its end products. Its original aim is to make a new device that does the same thing as the device studied. At first glance, modelling the *functional* structure of a device from its products in order to make similar products through new procedures is very far from what McCulloch and Pitts achieved, since they were not driven to manufacture some artificial neuron. Yet they have followed the same *method*: they inferred from the neuron's properties the structure of functions without which these properties could not be as they are (Franck, 2002b).

---

[11] *Reverse engineering* denotes today diverse research practices varying with the areas of application. We refer to its initial sense.

This method is the one which has been conceived by the classical programme of scientific research: from the sustained observation of some property of nature (light, heat, motion…) we try to infer – to *induce*[12] – the *functional* structure – in classical terms the *axiom*, *form, principle,* or *law* – which rules the process generating this property. We may, at present, qualify this method as *functional-mechanistic* to underline that it aims to model *the structure of functions* that rules the *mechanism* – the process – generating some property of nature.

For social properties, the method involves modelling the structure of the social functions (the 'first' one, the 'lesser' and the 'middle' in Bacon's terms) that rule the social process generating these properties, and without which these social properties could not become as they are. For example, regarding variations of population size and structure, for example, demographers uncovered the 'first' principle of the generating processes, namely some combination – which remains to be discovered – of three functions: *fertility, mortality, and migration*.

The 'law' of *supply* and *demand*, as another example, is the 'first' structure of functions which was inferred (*induced*) by Adam Smith from the observation of markets: it rules the process of social exchanges generating the market. Karl Marx inferred the general structure of *functions* ruling the process that generates industrial production from a thorough historical study of the technical and social organisation: this 'first' principle consists of *separating labour and capital.* Finally, Durkheim inferred the *integration theory* from a sustained statistical analysis of the differences in suicide rates between several social milieus: the social process which generates suicides, whichever their causes, is ruled by the *integration* of the individual agents. The application of the classical programme led to these prominent *theoretical* results at the height of *social* sciences.

Next, the functional structure governing the process generating some social property, once established as well as possible, may guide us in identifying and modelling the social factors which – in some singular, historical situation – have contributed to that process. We may restrict our causal investigation to those variables which plausibly contributed to the combination of functions required for generating the property under study. For example, what social factors (events, agents' behaviour, etc.) led to a weakening of the *integration* of people in some social milieu, and contributed to the increase of suicide? Another example: when we investigate the ups and downs of the market, we no longer ought to interrogate every plausible factor influencing these variations; instead, it may suffice to investigate and model the factors implied in supply and demand. In demography it is the functions of fertility, mortality and migration which actually delimit its parameter space and channel the *empirical* investigation of demographic properties.

---

[12] Bacon's *induction* is regularly confounded with induction by philosophers in its usual sense of generalisation. Bacon wrote: "In establishing axioms, another form of induction must be devised than has hitherto been employed, and it must be used for proving and discovering not first principles (as they are called) only, but also the lesser axioms, and the middle, and indeed all. For the induction which proceeds by simple enumeration is childish; its conclusions are precarious and exposed to peril from a contradictory instance; and it generally decides on too small a number of facts, and on those only which are at hand." (Bacon 1620; aphorism 105).

Against this background, we propose that the model-based research programme should proceed in accordance with the classical inductive programme, which we qualify as functional-mechanistic. Model building should start with a collection of all relevant empirical information about the social property under study. This would serve as the basis with which to infer the formal functional structure of the social property in question. Once the structure is modelled it can serve to guide the modelling – also simulation modelling – of the interactions between the systems of individuals, groups and institutions, combining the bottom-up and top-down relationships, and feedbacks between them (Franck 2002a).

The key stages of the inductive functional-mechanistic approach are shown in Figure 2. The solid arrows denote the four main stages of the process. Their implementation leads to the execution, and analysis of a computational model designed on the basis of a functional structure of the population processes under study, and derived from empirical observations. The dotted lines depict a possible feedback: the model results can guide the process of further data collection on particular aspects of the process that have not been included in the model. Obtaining more information would enable revisiting the inferred functional structure, as well as the resultant conceptual, mathematical, and computational models. Thus, the proposed approach could be seen as iterative, with successive modelling iterations enabling the researchers to identify gaps in existing knowledge and to fill these gaps by conducting further observations of the processes of interest.
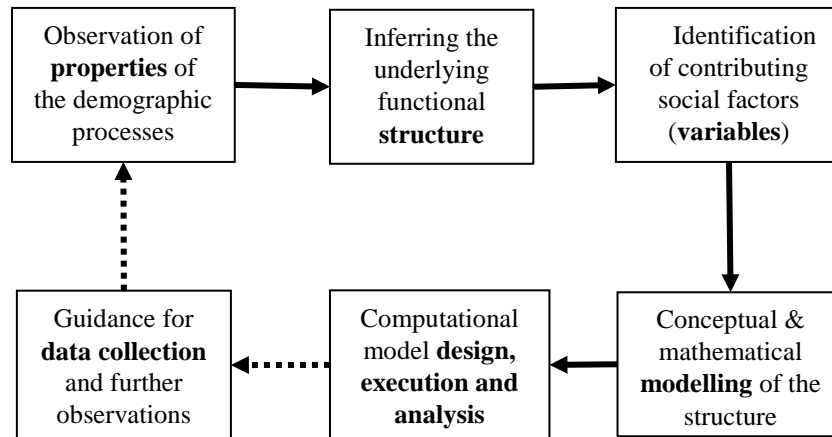

Fig. 2 Key stages of the inductive functional-mechanistic approach to model-based demography

The proposed approach is in line with the suggestion of Conte et al. (2012, p. 342) that "… data can be used to check and validate the results of simulation models and socio-economic theories, but a further step in using them is to take them into account already at the modelling stage." A careful experimental design becomes therefore a crucial part of the modelling process, and statistical methods, such as those discussed earlier in this chapter, need to become embedded in the model construction process, rather than being used only as a validation tool.

5. Towards a Research Agenda for Demography

The resulting research agenda we would like to propose for demography is based on three key pillars: (1) adherence to the classical programme of scientific enquiry; (2) enhancement of the ways in which demographic phenomena are measured and interpreted; and (3) the use of formal models, based on the functional-mechanistic principles, as fully-fledged tools of population enquiries.

According to several authors demography should become *interdisciplinary* in order to compensate its perceived shortcomings (e.g. Petit and Charbit, 2012). Demography should borrow *information, methods,* and *theories* from other social sciences. Although we approve a cautious recourse, when needed, to other disciplines, we have to underline that the solution to the weaknesses of current demographic practices cannot be found in theories and methods borrowed from other social sciences, since the last share similar weaknesses (e.g. Moss and Edmonds, 2005). All of these are confronted, just as in demography, with the *complexity* of society; they suffer from *uncertainty* in collecting and treating information often more than in population sciences; and their *theories* are hard to settle. The history of social sciences since the mid of 20th century teaches us that many innovative 'theories' had a generally short life, or at best remain heavily controversial. Moreover, the flowering of such 'theories' nourishes the proliferation of heterogeneous explanations and seem to increase the *complexity* and the *uncertainty* which undermines social sciences.

On the contrary, the first of the pillars of the proposed research agenda – the classical programme of scientific research – helps overcome the *complexity* of society, it reduces the *uncertainty* of the models we are building and of the explanations we are advancing, it establishes the *theoretical* component of research, and it discloses the way to *generalize* social models, something which is reputed to be an inaccessible goal in the social area. This is one of the reasons why we recommend applying this method in particular in demography and population sciences.

The belief that knowledge is something like a copy or an image is widespread. The classical programme conveys a different concept of science: scientific research is not intended to improve or to extend our image of reality. Instead, scientific research consists of discovering the principles governing the processes that generate some properties of nature or of society. We need to collection the best information on some *property* of nature – not about nature as a whole – in order to discover the principles governing the process that generates this property. The same applies for the social sciences: we need to collect the best information about some social *property* of human populations, not about human populations as a whole. Moreover, it is not merely information about this social property that will reinforce our scientific knowledge. Collecting information must be augmented by research on the principles – i.e. the combination (or *structure*) of *functions* – commanding the process generating this property. When selected in this way, the required informative intake varies with the property under study and it is drastically reduced, restricting at the same time the complexity of the task. The classical

research programme also restricts the *theoretical* approach to some social property to the modelling of *the structure of functions necessary* to generate this property.

This approach provides a major criterion for selecting information, by restricting our casual investigation to those variables which plausibly contributed to performing the combination of functions required for generating the property under study. This is yet another way to overcome complexity that is delivered by the classical programme of science.

Assumptions are an important source of uncertainty and nourish the proliferation of explanations in the social sciences[13]. The classical research programme recommends setting aside any assumptions in the guiding of scientific investigation (Baconian 'idols'). Thus it eliminates the root of any uncertainty arising from assumptions. Managing research without assumptions seems difficult – but can be done if we resist the urge to make hypotheses. Newton argued that "Hypotheses are not to be regarded in experimental Philosophy" (after Ducheyne, 2005, p. 124). This way of thinking is not in fashion today in the social sciences, yet it is commonplace in the natural sciences.

The classical research programme substitutes *induction* for the *hypothetical-deductive* approach. As we have seen, *inductive research* in its classical sense consists in inferring, from the sustained observation of and experimentation on some property of nature or population, the *functional* structure – in classical terms the *axiom*, *form, principle,* or *law* – which rules the process generating this property, and *without which* this property *could not* come about as it does. This criterion of *necessity* which guides the *inductive* investigation of the *functional structure* ruling the social processes which generate some social property is the best guarantor of the relevance of some *theory*, be it in the natural or in the social sciences.

Besides reducing by far the nagging worries of *complexity* and *uncertainty* characterising the social *sciences* today, the classical programme of science provides another huge advantage. By focusing on the process – or 'mechanism' – generating some natural or social property, the functional structure is treated independently of the causal structure. Modelling each of these structures separately allows us to disclose the way to *generalise* social models. Causal structures may never be generalized since populations are diverse and changing; causal structures are at best relatively constant. But a functional structure may be generalized in the sense that, *whenever the same property occurs*, the functional structure of its generative process ought to operate, insofar it has been established that this structure is really required for the property to appear[14]. This is the core sense of the *universality* and of the *necessity* of natural *laws*, but it has been regrettably distorted by the Humean empiricist tradition. For the founders of modern science the term *law*

---

[13] Formulating and testing hypotheses is not wrong, in our opinion, as long as it is based on empirical observations. However, throughout the present chapter we plead for abandoning the hypothetical-deductive approach and for substituting it with the classical induction.

[14] The property itself may not be generalized, of course.

was simply a metaphorical synonym of *principle*[15]. Thus laws are fully attainable by the social sciences, just as by the natural sciences, insofar we are willing to return to the classical concept of scientific laws.

The second pillar of the tentative research agenda we wish to advance comprises a question: how can we make better use of the measures achieved in demography? One of the main tasks of demography is to measure human populations: their size, density, rate of change, composition, various distributions, as well as the possible causes and consequences of changes in these factors[16]. In order to achieve this, demography rests largely on statistical analysis. Yet, *measuring* provokes an increasing dissatisfaction today. This is due, in our opinion, to a distorted view of measurement and quantification. *Measuring* population properties is *judging*, by way of comparisons, the amplitude of these properties regarding their *potentialities*. Measuring also guides the *induction* – in the sense of the classical research programme – of the principles governing these properties.

The term *"potentialities"* refers to what possible effects something might generate in certain circumstances. These *potentialities* are what we have in mind, explicitly or not, when we are measuring – or *judging* – some social property[17]. Measuring may not be confounded with – nor reduced to – the mathematical, statistical or other means by which measuring is carried out (for example censuses, surveys or vital registration systems; see Courgeau 2013).

In essence, we ought to multiply the measures of the social properties under study – all sorts of measures which are adequate – and to improve the *quality* of our measures in order to reinforce the quality of our judgments about their potentialities. This recommendation is exactly the opposite of what was proposed for the future of demography in recent years by some demographers (e.g. Tabutin 2007; Charbit and Petit 2011; Petit and Charbit 2012, and others): they wish to reduce the importance of *measuring* in demography and to increase our confidence in judgments – assumptions – conceived without measures in other disciplines, and somewhat abusively called *theories*.

Now we reach the third pillar of our tentative research agenda: to promote the *model-based* work programme, based on the functional-mechanistic approach outlined in this chapter. This approach carries with it substantial promise: it complements the four extant paradigms while incorporating insights gained from model-based science. Besides, as we see in other areas of model-based science, the deployment of this kind of approach likely will influence future data col-

---

[15] The principles are traditionally named *theories*; this tradition goes back to Plato's *theoria*, and reserves to the term *theory* the restricted sense of a corpus of principles. This is far from its present use describing as a 'theory' every sort of conceptual hypothesis, or model, or explanatory 'mechanism'.

[16] See for example the following definition of demography (IUSSP, 1982): "the scientific study of human populations primarily with respect to their size, their structure and their development; it takes into account the quantitative aspects of their general characteristics".

[17] Plato, who was familiar with the concept of *number* developed by the Pythagoreans, developed at length the idea that *measuring is judging*, and that we ought to recourse to measures in order to act wisely in politics as well as in private life (see Bassu 2009, 2011).

lection in demography and other population sciences, not only from surveys and other traditional sources, but also controlled experiments (Conte et al. 2012).

The model-based approach provides us with the means to expand the range of benefits already provided by multilevel modelling. We gain deeper insights into the interactions between various population systems, and we also gain the capacity to explore the parameter space of the simulations by generating "what-if" scenarios. Simulation parameters – once they result from the functional-mechanistic approach – govern the way in which the complex, interacting social processes in the model work, and therefore exploring the parameter space enables us to investigate numerous such scenarios, which could represent policy changes, individual behavioural changes, societal-level changes, and similar (Silverman et al., 2013). Given the construction of these simulations, running them under varied scenarios can illustrate the unforeseen, non-linear impact of changes to these complex processes. This *scenario generation* capability, when coupled with uncertainty quantification, allows us to extend the utility and policy relevance of empirically-grounded population models beyond what is accommodated by the traditional approaches. In addition to addressing the ecological and atomistic fallacies, which is already the case in the current multilevel paradigm, we could now analyse different layers of interactions between population systems.

Such approaches, relying as they do upon inference about systems and interactions between them, are also well-suited to integrating both quantitative and qualitative data into the same simulations, as mentioned before. For example, qualitative information can be gathered from individuals within the population under study, as a means of gaining understanding regarding individual behaviours, intentions, and goals, and these can inform the behavioural rules in the simulated population. Further, qualitative data can even be used to guide the construction or modification of the model itself (e.g., Polhill et al. 2010).

In this chapter, we have discussed what we believe are the key elements of model-based approaches – such as their inductive character – that would be necessary for them to become a real addition to the toolbox of population sciences. If the future demography is to examine complex, multilevel interactions of different elements of population systems seriously, computational approaches are the methodology of choice. However, the models constructed would need to conform to the rigours of scientific enquiry, rather than being based on arbitrary assumptions which often lack empirical basis. The model-based work programme, rooted in the functional-mechanistic approach, offers a general analytical framework to guide this process. Besides, more attention needs to be paid to the role of different levels of analysis, and interactions between them. If this is done correctly, the multilevel paradigm will gain very powerful analytical tools to study new research questions, related to the behaviour of complex population systems.

The next step in developing model-based demographic approaches must consist of proposing some concrete solutions, analytical formalisms and practical guidelines for the modellers. Although this topic remains beyond the scope of the current chapter, in the literature there are already some promising suggestions in that regard. For example, Casini et al. (2011) have proposed using recursive

Bayesian networks as an analytical formalism for building "models for prediction, explanation, and control", which are capable of describing functional mechanisms and causal relations, and of analysing uncertainty in coherent, probabilistic terms. In practice, the process of model-building can be iterative, as shown in Figure 2: we could start with a first approximation of a model that reproduces some well-established qualitative features of the modelled phenomenon ('stylised facts'), but should not stop there: the model could then be refined by including increasingly more data as they become available. These propositions are clearly worthy of investigating in the demographic context.

Of course, it is unrealistic to expect that every piece of model-based demographic research should contain all the elements discussed above. However, as future studies progress – and as populations under study continue to shift following ever-changing and interacting social processes – model-based approaches to demography will bring about further opportunities for constructing and verifying the models. In this respect the linkage between empirical data on population structures and modelling the social mechanisms and interactions at the root of these structures becomes ever more important – and perhaps more powerful.

References

Aalen, O.O. (1975). *Statistical inference for a family of counting processes*. PhD thesis, Berkeley: University of California.

Ahlburg, D.A. (1995). Simple versus complex models: Evaluation, accuracy and combining. *Mathematical Population Studies*, 5(3), pp. 281–290.

Alho, J. M. and Spencer, B. D. (2005). *Statistical Demography and Forecasting*. Berlin–Heidelberg: Springer.

Alkema, L., Raftery, A.E., and Clark, S.J. (2007) Probabilistic projections of HIV prevalence using Bayesian melding. *Annals of Applied Statistics*, 1(1), pp. 229–248.

Aparicio-Diaz, B., Fent, T., Prskawetz, A. and Bernardi, L. (2011). Transition to Parenthood: The Role of Social Interaction and Endogenous Networks. *Demography*, 48(2), pp. 559–579.

Axtell, R., Epstein, J., Dean, J., Gumerman, G., Swedlund, A., Harburger, J., Chakravarty, S., Hammond, R., Parker, J., and Parker, M. (2002). Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley *Proceedings of the National Academy of Sciences of the United States of America*, 99 (suppl. 3), pp. 7275–7279.

Bacon, F. (1620). *Novum Organum*, London: J. Bill. English translation by Spedding, J., Ellis, R.L., Heath, D.D. (1863) in *The works* (Vol. VIII), Boston: Taggard and Thompson.

Bassu, S. (2009). Metretique, éthique et politique: le Protagoras et le Politique de Platon. *Dissertatio,* 29: 85–114.

Bassu S. (2011). Ordre et mesure, *kosmos* et *metron* de la pensée archaïque à la philosophie platonicienne. In : S. Alexandre and E. Rogan (eds.), *Ordres et désordres.* Zetesis – Actes des colloques de l'association, pp. 1-23. http://www.zetesis.fr/actes/spip.php?article25

Bayes, T. R. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, *53*, pp. 370–418.

Bijak, J. (2010). *Forecasting International Migration in Europe: A Bayesian View.* Springer Series on Demographic Methods and Population Analysis, vol. 24. Dordrecht: Springer.

Bijak, J., Hilton, J., Silverman, E., and Cao, V. (2013). Reforging the Wedding Ring: Exploring a Semi-Artificial Model of Population for the United Kingdom with Gaussian process emulators. *Demographic Research*, 29(27), pp.729–766.

Billari, F., and Prskawetz, A., eds. (2003). *Agent-based computational demography. Using simulation to improve our understanding of demographic behaviour.* New York: Physica-Verlag.

Blayo, Ch. (1995). La condition d'homogeneite en analyse demographique et en analyse statistique des biographies, *Population*, 50(6), pp. 1501–1518.

Boudon, R. (1977). *Effet pervers et ordre social*. Paris: Presses Universitaires de France.

Brenner, Th, and Werker, C. (2007). A Taxonomy of Inference in Simulation Models. *Computational Economics*, 30(3), 227–244.

Burch, T. (2003a). Data, models, theory and reality: the structure of demographic knowledge. In *Agent-based computational demography. Using simulation to improve our understanding of demographic behaviour*, in Billari, F., Prskawetz, A., eds., Heidelberg, New York: Physica-Verlag, pp. 19–40.

Burch, T. (2003b). Demography in a new key: A theory of population theory. *Demographic Research*, 9(11), pp. 263–284.

Casini, L., Illari, P.M., Russo, F., and Williamson, J. (2011). Models for Prediction, Explanation and Control: Recursive Bayesian Networks. *Theoria*, 26(1)/70/, pp. 5–33.

Charbit, Y, Petit, V. (2011). Towards a comprehensive demography: rethinking the research agenda on change and response. *Population and Development Review*, 37 (2), pp. 219–239.

Chattoe, E. (2003). The Role of Agent-Based Models in Demographic Explanation. In: Billari, F., and Prskawetz, A., eds., *Agent-based computational demography. Using simulation to improve our understanding of demographic behaviour*, Heidelberg, New York: Physica-Verlag, pp. 41–54.

Clark, S.J., Thomas, J.R. and Bao, L. (2012). Estimates of age-specific reductions in HIV prevalence in Uganda: Bayesian melding estimation and probabilistic population forecast with an HIV-enabled cohort component projection model. *Demographic Research*, 27(26), pp. 743–774.

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, Loreto, V., Moat, S., Nadal, J.-P., Sanchez, A., Nowak, A., Flache, A., San Miguel, M., and Helbing, D. (2012). Manifesto of computational social science. *European Physical Journal Special Topics*, 214, pp. 325–346.

Courgeau, D. (2007). *Multilevel synthesis. From the group to the individual*. Dordrecht: Springer.

Courgeau, D., (2012). *Probability and social science. Methodological relationships between the two approaches*. Methodos Series 10, Dordrecht : Springer.

Courgeau, D. (2013). La mesure dans les sciences de la population. *Cahiers Philosophiques*, 135(4), pp. 51–74.

Courgeau, D., Bijak, J., Franck, R., and Silverman, E. (2014). Are the four Baconian idols still alive in demography? *Revue Quetelet / Quetelet Journal*, 2 (2), pp. 31–59.

Courgeau, D., Franck, R. (2007). Demography, a fully formed science or a science in the making, *Population-E*, 62 (1), pp. 39–45. (La démographie, science constituée ou en voie de constitution? Esquisse d'un programme. *Population*, 62(1), pp. 39–46.)

Courgeau, D., Lelièvre, E. (1992) *Event History Analysis in Demography*. Oxford: Clarendon Press.

Di Paolo, E. A., Noble, J., and Bullock, S. (2000). Simulation models as opaque thought experiments. In: Bedau, M., McCaskill, J., Packard, N., & Rasmussen, S. (eds.), *Proceedings of the 7th International Conference on Artificial Life*. Cambridge MA. MIT Press, pp. 497–506.

Doob, J.L. (1953). *Stochastic processes*. John Wiley & Sons: New York / Chichester.

Ducheyne, S. (2005). Bacon's idea and Newton's practice of induction. *Philosophica*, 76, pp. 115–128.

Durkheim, E. (1897). *Le suicide*. Paris: Alcan.

Epstein, J.M. (2008). Why Model? *Journal of Artificial Societies and Social Simulation*, 11 (4), article 12. http://jasss.soc.surrey.ac.uk/11/4/12.html

Franck, R. (ed.) (1994). *Faut-il chercher aux causes une raison? L'explication causale dans les sciences humaines*. Paris : Librairie Philosophique Vrin.

Franck, R. (ed.) (2002a). *The explanatory power of models. Bridging the gap between empirical and theoretical research in the social sciences.* Methodos Series vol. 1. Boston/Dordrecht/London: Kluwer Academic Publishers.

Franck R. (2002b). Computer simulation and the Reverse Engineering method. Conclusions of Part II. In: R. Franck (ed.), *The Explanatory Power of Models*. Methodos Series vol. 1. Dordrecht/Boston/London: Kluwer Academic Publishers, pp. 141–146.

Franck R. (2004). Introduction. Modelling is the Key to Science. In: M. Negrotti (ed.) *Yearbook of the Artificial*, vol. 2: 'Models in Contemporary Science'. Bern/Berlin/Bruxelles/Frankfurt/ New York/Oxford/Wien: Peter Lang, pp. 9–21.

Geard, N., McCaw, J.M., Dorin, A., Korb, K.B. and McVernon, J. (2013). Synthetic Population Dynamics: A Model of Household Demography. *Journal of Artificial Societies and Social Simulation*, 16 (1), article 8. http://jasss.soc.surrey.ac.uk/16/1/8.html

Godfrey-Smith P. (2006). The strategy of model-based science. *Biology and Philosophy*, 21(5), pp. 725–740.

Goldstein, H. (1987). *Multilevel models in educational and social research.* London: Arnold.

Graunt, J. (1662). *Natural and political observations mentioned in a following index, and made upon the bills of mortality.* Tho. Roycroft: London.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S., Huse, G., Huth, A., Jepsen, J.U., Jørgensen, C., Mooij, W.M., Müller, B., Pe'er, G., Piou, C., Railsback, S.F., Robbins, A.M., Robbins, M.M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R.A., Vabø, R., Visser, U., and DeAngelis, D.L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198, pp. 115–126.

Henry, L. (1959). D'un problème fondamental de l'analyse démographique. *Population*, 14(1), pp. 9–32.

Hirschman, Ch. (2008). The future of demography. *Asian Population Studies*, 4(3), pp. 233–234.

Holland, J.H. (1995). *Hidden order.* Reading, Mass.: Addison-Wesley

Huneman, P. (2014). Mapping an expanding territory: computer simulations in evolutionary biology. *History and Philosophy of the Life Sciences*, 36(1), pp. 60–89.

Huyghens, C. (1657). *De ratiociniis in ludo aleae*. Leyde: Elzevier.

Ibrahim, J.G., Chen, M.-H., Sinha, D. (2001). *Bayesian survival analysis*. New York: Springer.

IUSSP [International Union for the Scientific Study of the Populations] (1982). *Multilingual Demographic Dictionary*, second edition, Liège: Ordina.

Kennedy, M., and O'Hagan, T. (2001). Bayesian Calibration of Computer Models. *Journal of the Royal Statistical Society, Series B*, 63(3), pp. 425–464.

Keyfitz, N. (1971). Models. *Demography*, 8(4), pp. 571–580.

Klüver, J., Stoica, Ch., and Schmidt, J. (2003). Formal Models, Social Theory and Computer Simulations: Some Methodical Reflections. *Journal of Artificial Societies and Social Simulation*, 6 (2), article 8, http://jasss.soc.surrey.ac.uk/6/2/8.html

Kniveton, D., Smith, Ch. and Wood, S. (2011). Agent-based model simulations of future changes in migration flows for Burkina Faso. *Global Environmental Change*, 21 (Suppl. 1), pp. S34–S40.

Kuhn, T. (1962). *The Structure of scientific revolutions*. Chicago-London: The University of Chicago Press.

Laplace, P.S. (1774). Mémoire sur la probabilité des causes par les événements. *Mémoires de l'Académie Royale des Sciences de Paris*, Tome VI, pp. 621–656.

Laplace, P.S. (1812). *Théorie analytique des Probabilités, 2 vols*. Paris: Courcier Imprimeur.

Levins, R. (1966) The Strategy of Model Building in Population Biology. *American Scientist*, 54(4), pp. 421–431.

Lutz, W. (2012). Demographic Metabolism: A predictive theory of socio-economic change. *Population and Development Review*, 38 (Supplement), pp. 283-301.

Mannheim, K. (1928). Das Problem der Generationen. *Kölner Vierteljahrshefte für Soziologie*, 7, pp. 309–330.

Mason, W.M., Wong, G.W., and Entwistle, B. (1983), Contextual analysis through the multilevel linear model, in S. Leinhart (ed.), *Sociological Methodology 1983-1984*, Jossey-Bass, San Francisco, pp. 72–103.

McCulloch, W.S. and Pitts, W.H. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5(4), pp. 115–133.

Michel, J.-B., Shen, Y.K., Aiden, A.P., Veres, A., Gray, M.K., Brockman, W., The Google Books Team, Pickett, J.P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M.A., and Aiden, E.L. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331(6014), pp. 176–182.

Mill, J.S. (1843). *A system of logic, ratiocinate and inductive, being a connected view of the principles of evidence, and the methods of scientific investigation*, vol. I, London: Harrison.

Morgan, S.P., and Lynch, S.M. (2001). Success and Future of Demography. The Role of Data and Methods. *Annals of the New York Academy of Sciences*, 954, pp. 35–51

Moss, S., and Edmonds, B. (2005). Towards Good Social Science. *Journal of Artificial Societies and Social Simulation*, 8 (4), article 13. http://jasss.soc.surrey.ac.uk/8/4/13.html.

NRC [National Research Council], (2000). *Beyond six billion: Forecasting the world's population*. Washington, DC: National Academies Press.

Oakley, J. and O'Hagan, A. (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, 89(4), pp. 769–784.

Pascal, B. (1665). *Traité du triangle arithmétique, avec quelques autres traités sur le même sujet*. Paris : Guillaume Desprez.

Petit, V., Charbit, Y. (2012). The French school of demography: contextualising demographic analysis. *Population and Development Review*, 38 (supplement), pp. 322–333.

Petty, W. (1690). *Political arithmetick*. London: Robert Clavel & Hen. Mortlock.

Polhill, J.G., Sutherland, L.-A., and Gotts, N.M. (2010) Using Qualitative Evidence to Enhance an Agent-Based Modelling System for Studying Land Use Change. *Journal of Artificial Societies and Social Simulation* 13(2), art. 10. http://jasss.soc.surrey.ac.uk/13/2/10.html.

Poole, D., and Raftery, A. E. (2000). Inference for deterministic simulation models: The Bayesian melding approach. *Journal of the American Statistical Association*, 95(452), pp. 1244–1255.

Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, pp. 111–163.

Raftery, A.E., Li, N., Ševčíková, H., Gerland, P., and Heilig, G,K. (2012). Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences*, 109: 13915–13921.

Rogers, A. (1975) *Introduction to Multiregional Mathematical Demography*. New York: John Wiley & Sons.

Ryder, N.B. (1951). *The cohort approach. Essays in the measurement of temporal variations in demographic behaviour*. Ph. D. Dissertation, New York: Princeton University.

Silverman, E. (2016). *Methodological investigations in agent-based modelling, with applications to social sciences*. Forthcoming.

Silverman, E., Bijak, J., Hilton, J., Cao, V., and Noble, J. (2013). When Demography Met Social Simulation: A Tale of Two Modelling Approaches. *Journal of Artificial Societies and Social Simulation*, 16(4), art. 9. http://jasss.soc.surrey.ac.uk/16/4/9.html

Silverman, E., Bijak, J. and Noble, J. (2011). Feeding the Beast: Can Computational Demographic Models Free Us from the Tyranny of Data? In Lenaerts, T., Giacobini, M., Bersini, H., Bourgine, P., Dorigo, M., and Doursat, R. (eds.), *Advances in Artificial Life, ECAL 2011: Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems*. Cambridge MA: MIT Press, 747–754.

Silverman, E. and Bryden, J. (2007). From artificial societies to new social science theory. In: Almeida e Costa, F., Rocha, L.M., Costa, E., Harvey, I. and Coutinho, A. (eds.) *Advances in Artificial Life, 9th European Conference, ECAL 2007 Proceedings*. Berlin–Heidelberg: Springer, pp. 645–654.

Smith, S.K. (1997). Further thoughts on simplicity and complexity in population projection models. *International Journal of Forecasting*, 13(4), pp. 557–565.

Tabutin, D. (2007). Whither demography? Strengths and weaknesses of the discipline over fifty years of change. Followed by a debate on the future of the discipline, by G. Caselli and V. Egidi, D. Courgeau & R. Franck, J. Hobcraft, and J. Hoem, *Population-E*, 62 (1), pp. 13–56.

Thagard, P. (1993). *Computational Philosophy of Science*. Cambridge MA: MIT Press.

Willekens, F. (2005). Biographic forecasting: bridging the micro-macro gap in population forecasting. *New Zealand Population Review*, 31(1), pp. 77–124.

Willekens, F. (2012). Migration: A perspective from complexity science. Paper for the Complexity Science for the Real World workshop on migration, Chilworth, UK, 16 February 2012.

Xie, Y. (2000). Demography: Past, Present and Future. *Journal of the American Statistical Association*, 95 (450), pp. 670–673.