# The Empirical Investigation of Non-Linear Dynamics in the Social World. Ontology, Methodology and Data

Graham Room[*]

**Abstract**

Across much of social science, linear models hold sway, but they have significant limitations. This article makes the case for studying social processes as co-evolving systems, involving non-linear dynamics. Co-evolution of species in the natural world is a blind process. In the social world in contrast, purposeful interventions by social actors are omnipresent, in their struggles for positional advantage. The article brings together co-evolving networks and purposeful social action in the "Contingent Historical Model." We seek to apply this model in ways that engage with both scholarly and policy concerns. If such investigations are to be fruitful, they must not only be elaborated theoretically, they must also be applied to empirical datasets. This article considers how this can be done, with what sorts of data sets and what forms of data analysis. It takes as its specific example the international datasets on patents, as revealing processes and patterns of technological innovation. It shows how such an approach can illuminate scholarly debates and develop indicators for policy makers. Finally, it offers an agenda for research into dynamic co-evolving systems across other empirical areas.

**Keywords**: non-linear dynamics; co-evolving systems; autocatalytic sets; patents and technological innovation; contingent historical change.

[*] Department of Social & Policy Sciences, University of Bath (United Kingdom); ✉ G.J.Room@bath.ac.uk; 
ⓘD https://orcid.org/0000-0002-7072-0180

## 1   Introduction

Across much of social science, the general linear model (GLM) holds sway. However, it has significant limitations. This article makes the case for studying social processes as co-evolving systems, involving *non-linear* dynamics. Co-evolution of species in the natural world is a blind process. In the social world in contrast, purposeful interventions by social actors are omnipresent, in their struggles for positional advantage. The article brings together co-evolving networks and purposeful social action in the "Contingent Historical Model."

We seek to apply this model in ways that engage with both scholarly and policy concerns. If such investigations are to be fruitful, they must not only be elaborated theoretically, they must also be applied to empirical datasets. This article considers how this can be done, with what sorts of data sets and what forms of data analysis. It takes as its specific example the international datasets on patents, as revealing processes and patterns of technological innovation.

Section 2 examines the assumptions that underlie the GLM, but also the weaknesses that are exposed, when those assumptions are not met. These weaknesses reveal the desiderata that any alternative must fulfill. In light of this, Section 3 examines models of co-evolving systems, as a form of *non-linear* dynamics. Section 4 elaborates the corresponding mathematics of autocatalytic sets, arguing that these must become a normal part of the social scientist's toolkit.

Co-evolution was first noticed by Darwinian biologists, as a blind process of the natural world. The complex dynamics of the social world are likewise in varying degree blind and emergent, without collective purpose. Nevertheless, purposeful interventions by social actors are omnipresent; this entails struggles for positional advantage and the exercise of power. Section 5 brings this together in the "Contingent Historical Model" (CHM) which we juxtapose to the GLM. This we seek to apply empirically and in ways that engage with both scholarly and policy concerns.

To do this involves three steps. Section 6 considers how the mathematics of co-evolving systems can be applied to empirical datasets and the requirements those datasets must fulfill. Section 7 applies these to databases of patents, as a vehicle for studying technological innovation. Section 8 examines how far an empirical network of co-evolving technologies can thereby be constructed, as a contingent historical process.

Section 9 brings together the results of this enquiry and the implications for future work.

## 2   The General Linear Model[1]

The task of social science is to explain social phenomena. This, it is commonly asserted, should involve measuring the effects of different "independent" variables on some "dependent" variable of interest. This can be presented diagrammatically. In **Figure 1**, the independent variables $x_1$, $x_2$ and $x_3$ shape the dependent variable $y$ (but with some effects exerted via the intermediate variable $z$).

This vision or ontology of the social world can be presented as an equation:

$$y = f_1(x_1) + f_2(x_2) + ... f_n(x_n) \qquad (1)$$

Quantitative social science applies this vision using regression analysis. It commonly casts the problem of explanation in terms of a set of linear equations; this is why it is often described as the "General Linear Model" (Abbott, 2001, Ch. 1). The GLM looks for the straight line that

---

1.    This section draws on Chapter 3 of Room (2016), written in collaboration with Graham K. Brown.
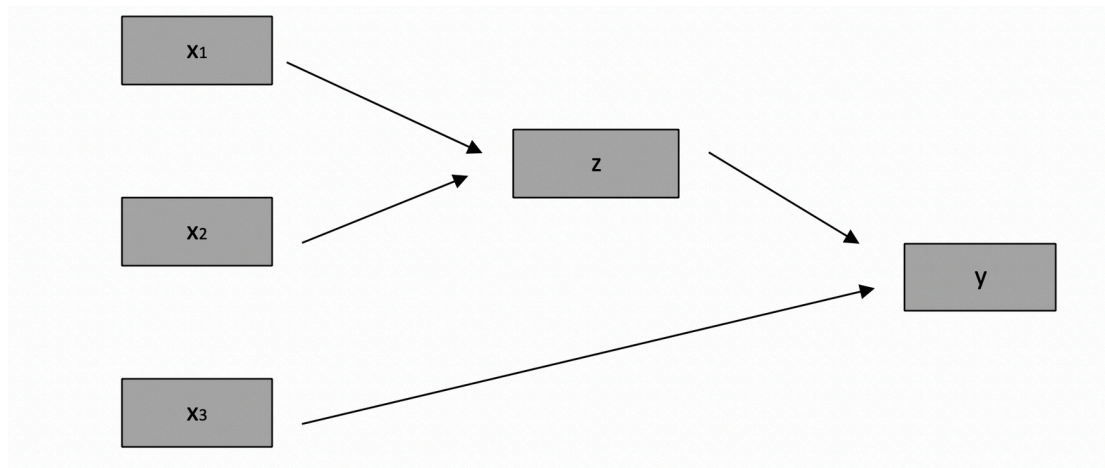
Figure 1: The General Linear Model

best estimates — and therefore "explains" — the dependent variable $y$ as the additive outcome of a number of independent variables $x_1 \dots x_n$ plus a random error term $u$:

$$y = b_1 \cdot x_1 + b_2 \cdot x_2 + \dots b_n \cdot x_n + u \qquad (2)$$

The $b$ coefficients measure the rate at which changes in the independent variables $x$ produce changes in $y$. The error term $u$ is a measure of how closely our equation captures the empirical data — and how much "noise" remains around its predicted values of $y$.

There are a number of assumptions involved here however, which should be made explicit.

1. The GLM assumes that the separate causal effects of the independent variables can be isolated — there are no significant interactions among them. When interactions are substantial, equation (1) cannot in general be solved mathematically; and statistical models such as equation (2) may produce biased estimators and reduced significance for each of the variables in question.[2]

2. The GLM assumes that the influence of the dependent variables $x$ on the dependent $y$ is downstream and one-way. There are no significant feedback processes allowing $y$ to influence $x$. And yet of course, in the real world such feedback processes are common. This may mean that instead of a downstream and uni-directional determination of $y$ by $x$, the directions of influence run upstream as well (**Figure 2**). Such self-reinforcing processes are well-recognized across the social sciences, including for example the economics of "cumulative causation" in Myrdal and Kaldor (Toner, 1999), system dynamics (Checkland & Scholes, 1990) and the social policy literature on the dynamics of social exclusion (Room, 1995). Nevertheless, across much of social science the mathematical modelling of such cumulative feedback is under-developed.[3]

---

2. It may be possible to deal with these problems through smart technical "fixes." If for example there is thought to be an interaction in the influence of two variable $x_1$ and $x_2$, we may create a third variable $x_3$, usually defined as the product of $x_1$ and $x_2$, which is then included in the regression analysis. Once this interaction term has been created, it allows us to maintain linearity in the estimation, if we assume that its influence is proportional to $x_3$. However, a limitation of this approach is that we need to stipulate *ex ante* what form the interaction
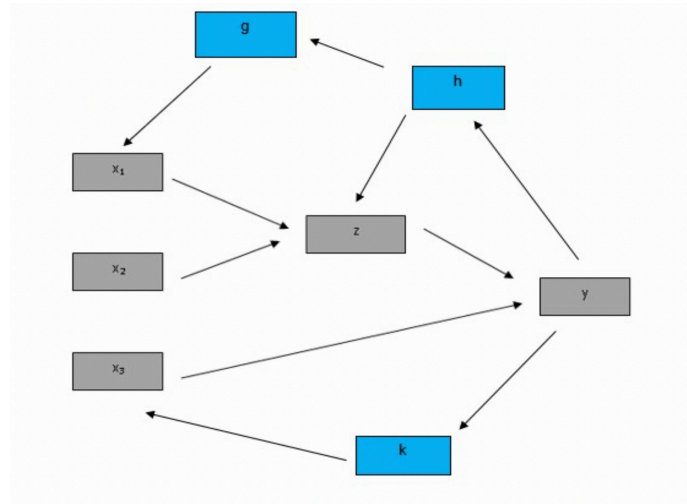
Figure 2: Feedback Processes

3. The GLM assumes that changes in the independent variables will produce broadly proportional changes in the dependent variable, across the entire range of observable values, within a rather simple "timescape."  In the presence of the afore-mentioned feedback processes however, timescapes are often more complex. Some effects of the independent variables are short-term and immediate, while others are long-term and delayed (Abbott, 2001, Ch. 1). Many social processes are replete with time lags, ratchets and path dependencies (Lieberson, 1987, Ch. 4).  Pierson (2004) points to the consequences of long-term and often slow changes in background social and economic conditions (pp. 74–77). There may be long periods of stasis, and then thresholds at which sudden avalanches of reconfiguration occur (as sometimes discussed in terms of "punctuated equilibria" (Baumgartner & Jones, 1993).

   It is possible for the GLM to handle some of these temporal complexities, using more sophisticated statistical techniques. Abbott's critics claim that he overlooks the advanced econometric techniques available to quantitative sociologists such as structural equation modelling (SEM), time series analysis and hierarchical linear models (Stolzenberg, 2003, p. 422).  Nevertheless, taken together, the GLM struggles, when applied to social sub-systems and processes with complex inter-connections.[4]

4. The GLM assumes, finally, that the independent and dependent variables are given and that the relationships among them are fixed.  This may well be appropriate, at least

---

   takes.

3. It is common to use lagged variables to build in some dynamics, with $y_{t-1}$ being used as a predictor of $y_t$. This does not however allow for the feedback processes from $y$ to $x$ that Figure 2 displays.

4. Further statistical advances in the decades since Abbott published his critique have extended even further the sophistication of econometric methods, including multi-level models that allow for hierarchical nesting of observations with higher level fixed effects (Rabe-Hesketh & Skrondal, 2008) and advances to selection bias two-stage models to include interaction effects (Brown & Mergoupis, 2011). While these advances extend the range of contexts in which linear methods might be appropriately applied, however, they do not overcome the underlying problems identified by Abbott (Brown, 2014).

in the short term.[5]   Sooner or later however, the relationships among the "variables" may change and the variables themselves amalgamate, divide or disappear and new ones emerge. For understanding these dynamics of wholesale reconfiguration, the techniques of the GLM are of little use.[6]

It is hardly surprising that the GLM exerts such a powerful sway across many disciplines. It is readily visualized; it can be formalized in terms of simple equations; it is convenient and tractable.

While the social world is not in general linear, linear methods can often be used to good effect, if they are understood as "local" approximations. Nevertheless, where the foregoing assumptions fail to hold, the general applicability of the GLM is put in question. Among natural scientists, there are clear and well-recognized limits on applying linear models to particular phenomena. Social scientists need similarly to have alternative strategies available, for when linear models are not appropriate (Jervis, 1997, pp. 34ff).

The weaknesses we have identified provide clues that point directly towards the alternative we require. It is after all precisely by confronting inconsistencies in this way that scientific understanding and practice are able to advance (Kuhn, 1970; Popper, 1994, Ch. 5; Tavory & Timmermans, 2014). The most appropriate alternative, we now argue, is to adopt a model of co-evolving systems, as a form of *non-linear* dynamics.

## 3   Co-Evolutionary Dynamics

The GLM finds its inspiration in classical physics. Evolutionary biology provides a quite different source of conceptual and methodological inspiration — albeit one which social scientists have interpreted and applied in a diversity of ways.[7]

The strengths of an evolutionary model align closely with the limitations of the GLM discussed in the previous section. With evolutionary science as our starting point, we now therefore consider co-evolutionary dynamics, as an alternative to the GLM.[8]

In his account of the diversification of species, Darwin was centrally concerned with processes of adaptation to different habitats. He depicted this visually as a "Tree of Life" (Darwin, 1859/1998). This offered successively sprouting branches and sub-branches, as particular "variations" adapted to and exploited different habitats, over many millions of years.

---

5.   It was for example the basis for Alfred Marshall's treatment of the short-term: the stock of capital in an economy was fixed, as were the relationships among different factors of production (Marshall, 1920, Book V, Ch. V). In the long-term however, these were all malleable.

6.   Advances in multi-level modelling allow us to disentangle the effects of differences in the characteristics of individuals living in different areas, from the higher-level effects of differences between the areas. What such methods cannot do however is unpick the interconnections of those neighbourhoods and the path dependencies involved in their interrelated histories.

7.   For an eloquent statement of the relevance of evolutionary models in social science, closely consistent with the argument of the present article, see Lieberson and Lynn (2002). See also Hodgson (2001, Ch. 22; 2002); Blaug et al. (2011).

8.   It could be argued that there are many other varieties of non-equilibrium social theory that could equally provide our point of departure and which enjoy some family resemblance with the evolutionary model adopted here. For overviews of that larger literature, for the social scientist approaching these matters for the first time, see Waldrop (1992), Johnson (2001), Ball (2004), and more recently Room (2011) and Johnson et al. (2017). A more comprehensive review might also include some "softer" strands of complexity writing in cultural studies: see for example Cilliers (1998).

Imagine looking down on the tree of life from above and viewing the top-most branches. Displayed there are the various species that are alive today (or we could consider any other horizontal "cut," representing the species that lived at another chosen period in the Earth's history). Across each of these cross-sections, the various branches (species) are connected in ecosystems of interdependence, involving dynamic synergies and arms races of co-evolution (Kauffman, 1993; Maynard Smith & Szathmary, 2000). These ecosystems typically involve populations far removed from each other across the evolutionary tree — for example, flowers and insects. They powerfully influence which species thrive and which are extinguished.

Figure 1 provided a visual representation of the GLM in its most basic terms. Such diagrams can provide powerful images that organize and direct our thinking about a given phenomenon. We now therefore consider a correspondingly parsimonious representation of evolutionary dynamics (**Figure 3**). This brings centre-stage variables that emerge, divide and disappear, as their interrelationships unfold. In subsequent sections we apply this to the social world.

**A** and **B** (at the bottom of the diagram) are two mutually adapted entities in the world of today. They might for example be the populations of two species such as bees and flowering plants, each benefitting as the other thrives.

We then pose the question: how did this mutual adaptation arise? We decline to treat it as a *causal* correlation, with the population of flowers "causing" the population of bees within a timeless environment. Instead, we seek to unpick the intricate and messy history of successive contingencies that has led to the mutual adaptations of today.

The upper part of the diagram reveals those historical contingencies. **A1** and **B1** were the ancestors of today's bees and flowers. As we know from Darwin, in each generation, variations are produced. In general however, as long as the environment remains stable (**Period -2**) they are unlikely to displace **A1** and **B1**.
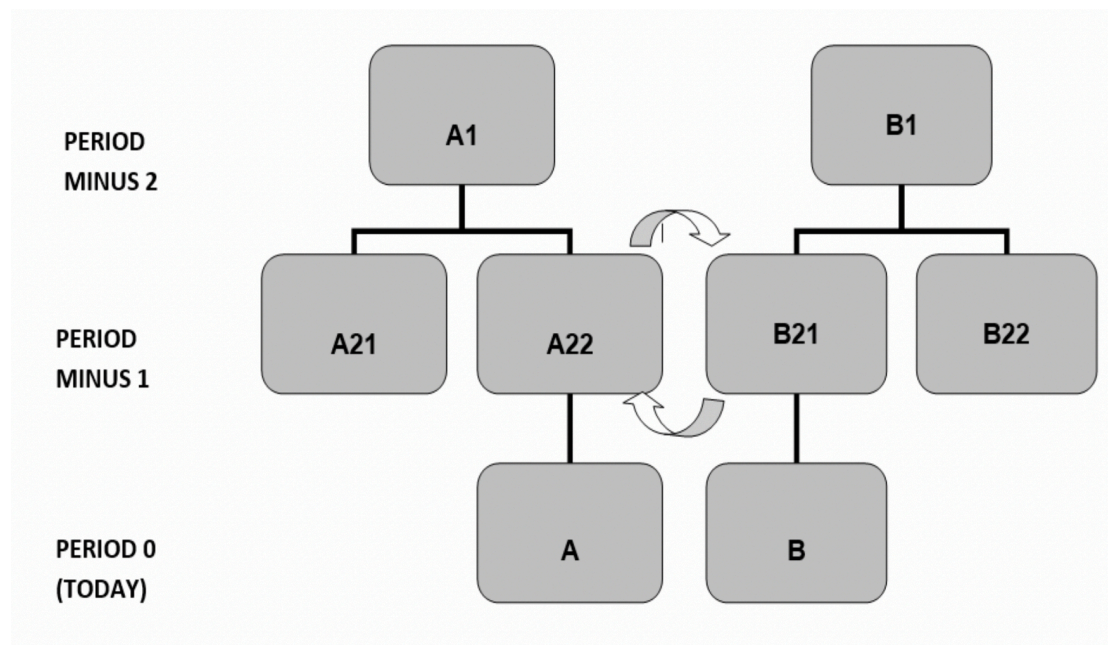


Figure 3. Co-Evolutionary Dynamics

It is when some environmental change occurs, at the start of **Period -1**, that we may expect

some of the variations (**A21**, **A22**, **B21**, **B22**) to be adopted as superior to **A1** and **B1**, which now become extinct. However, which of the variations **A21** and **A22** becomes preponderant depends in part on the new biotic environment constituted by the arrival of **B21** and **B22**; and *vice versa*. In short, what is crucial is which of the four sets of interactions between **A21** and **A22** on the one hand, **B21** and **B22** on the other, is of greatest mutual benefit.

In the diagram, we show the relationship of **A22** with **B21** as being this favoured pairing, this synergy or "elective affinity."[9] Each will now accelerate the flourishing of the other. Their flourishing will in turn deny resources to **A21** and **B22**, which in the "struggle for existence" become extinct. Hence we arrive at the bottom row of the diagram, **Period 0**, where **A** and **B** dominate. Here, by virtue of their domination, the environment is quite different from that in **Period -2** or even in **Period -1**. And indeed, **A** and **B** are themselves may be quite different in their capacities from their respective forebears **A1** and **B1**: perhaps barely recognizable (Shubin, 2008). Nevertheless, the domination of **A** and **B** is unlikely to last for ever; further rounds of interaction with the wider eco-system will eventually destabilize it, as new rounds of variation and selection are set in motion.

The dynamic synergies among particular elements have, as their obverse and corollary, the progressive disruption of other connections and elements of the eco-system — and the incorporation of those elements into the "empire" of the favoured elements (see **Figure 4**). The **A22–B21** axis becomes a vector of cumulative change, around which the wider eco-system is progressively reordered and re-configured. This also makes it a non-linear system with strong path dependency, where instead of the additive effects that are central to the GLM, change is multiplicative and self-reinforcing.

The dynamic synergy cannot however continue without limit; nor can the concomitant disruption and recycling of other elements. Some parts of the wider eco-system are too resilient and robust to be unpicked and re-worked; they constitute an "evolutionarily stable state" (Maynard Smith, 1982). On the other hand, the disruptive change that is driven by the elective affinity of **A22–B21** is forever opening up new possibilities for other elements of the ecosystem, other dynamic synergies that may eventually match, surpass or disrupt **A22–B21**.

In short therefore: the GLM finds its inspiration in Newtonian mechanics, with stable entities having causal effects on other stable entities. The process of co-evolutionary change elucidated by Darwin and his successors is quite different. Here the populations of different species share a common space, and interact with each other in a variety of mutual synergies and antagonistic threats. Through successive generations, the populations of these different species successively re-constitute each other, for the next round of the competitive struggle. These selection dynamics can shift; they will not necessarily be incremental and gradual; they can instead involve punctuated equilibria and tipping points and core shifts (Kauffman, 1993).

---

9. The term "elective affinity" was originally used in German chemistry of the Eighteenth Century, to refer to the way in which compounds interact and combine selectively with each other (Howe, 1978). The search for such affinities in chemistry was conducted in the shadow of Newton and in envy of physics and its claim to universal natural laws — just as our own account has been located within the larger debate about social science and the Newtonian antecedents of the GLM. Goethe took this idea of elective affinities into his novel *Die Wahlverwandtschaften*, applying it to sexual attraction. Kant in turn applied the idea to relationships among concepts; Weber to relationships between ideas and the interests of social actors. Perhaps surprisingly however, it does not seem to have been used in relation to biological co-evolution. In these various cases, "elective affinity" is not just a matter of complementarity or similarity; it is a dynamic synergy, in which elements that are especially favourable to each other enable the ensemble as a whole to flourish. It thus offers a dynamic of mutual selection, reinforcement and change. Crouch (2005, Ch. 3) has been a trailblazer in applying such a perspective within institutional sociology and anticipates much of what is said here.
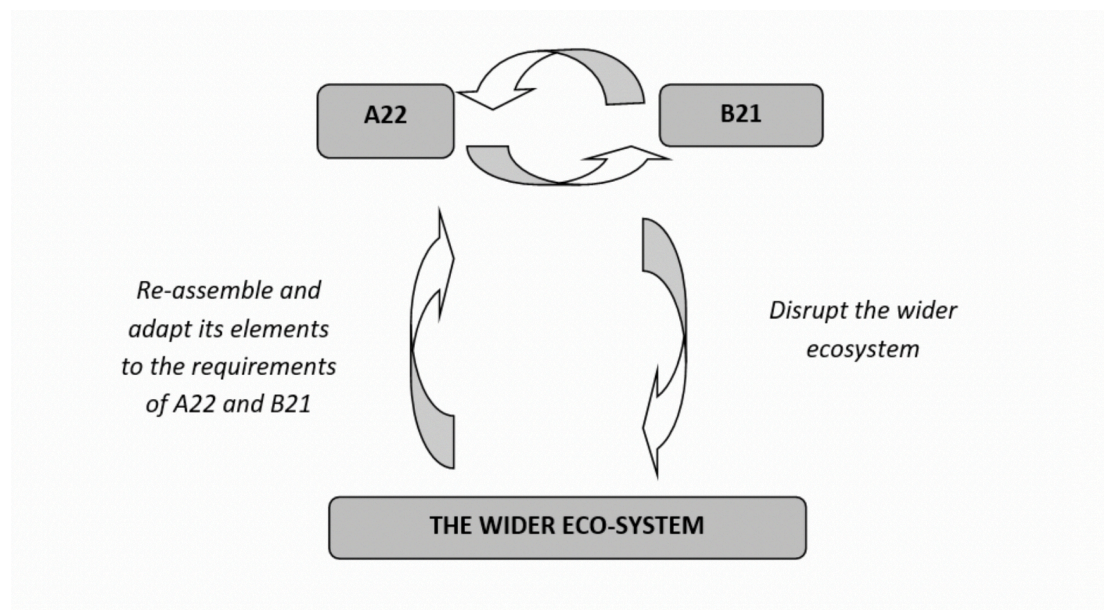
Figure 4. Creative Disruption.
The dynamic synergies between A22 and B21 disrupt and re-cycle the wider eco-system

There is a large scientific literature concerned with these co-evolutionary dynamics — conceptually, methodologically and empirically — and with their application beyond the biological realm. This literature has been central to the larger discussion of complex systems that has developed over recent decades.[10]

The sections that follow apply these insights to the modelling and empirical analysis of social and technological change.

## 4   The Maths of Co-Evolving Systems

We now require an empirical methodology, appropriate to the dynamics of co-evolving systems and applicable to the social world. In developing this methodology, we continue to take inspiration where appropriate from evolutionary biology, as well from social scientists who have sought to capture these dynamic processes in their empirical research.

In the GLM, the variables and their boundaries are fixed. So are their interrelationships. The practitioner of GLM seeks to distinguish and measure the effects of the independent variables on the dependent variable. Much energy has been devoted to clever ways of doing this, even under apparently unpropitious circumstances.

Co-evolutionary dynamics involve a quite different notion of causation. Here are phenomena whose configuration and frequency are driven by a dynamic that is synergistic not additive,

---

10.   Key contributions include Kauffman (1993) and Gavrilets (2004), elaborating the notion of fitness landscapes that co-evolve with each other; and Holland (1995) and Mitchell (1996), concerned with the genetic algorithms out of which such fitness landscapes are built. Solé and Boscompte (2006) explore complexity in ecological systems, using mathematics but accessible to non-mathematicians, and moving smoothly from species to networks to macroevolution. Gould and Eldridge (1977), Shubin (2008) and Odling-Smee, Laland, & Feldmann, (2003) deal with the contingencies of these evolutionary dynamics, the path dependency and the alternative "tapes of history" that might have played out (see also Section 5 below).

as depicted parsimoniously in Figure 3. It is the "elective affinities" among particular elements that lead to their progressive domination of the system in question and the corresponding re-configuration of the variables.

## 4.1 Qualitative System Dynamics

Powell has developed qualitative system dynamics (QSD) for the analysis of organizational change (Powell, 1992; Powell & Bradford, 1998); he builds on the work on system dynamics and causal loops by such writers as Checkland and Coyle (Checkland & Scholes, 1990; Coyle, 1996). He first maps the organizations of interest and the connections of interdependence among them. He then labels each line of interdependence, to indicate its direction, but also whether the relationship is direct or inverse — whether, in other words, an increase in some property or activity of the "upstream" node causes a change in the "downstream" node that is positive or negative.
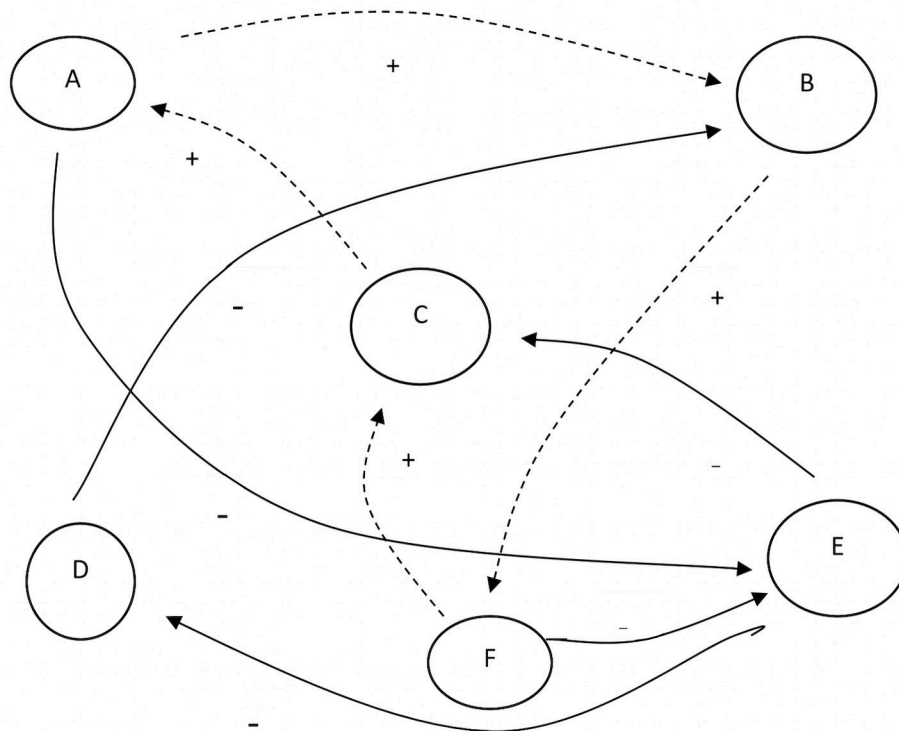


Figure 5. Runaway Loops

Within this map, Powell proceeds to identify those cycles whose links are all positive. These are cycles which loop back on themselves in self-reinforcing circles. When any one element starts increasing, the whole sub-system experiences explosive growth; when any starts decreasing, the sub-system experiences implosive collapse. Powell refers to these as "runaway loops." In **Figure 5** the runaway loop is marked as a dotted line. Each element in this loop accelerates the flourishing of its downstream neighbour, just as happened in Figure 3 between the elements

**A22** and **B21**. Meanwhile other cycles of interdependence loop back on themselves, in ways that dampen down change and stabilize the system as it presently exists.

Having identified the runaway loops, the next step is to assess how strong is the self-reinforcing dynamic. This will determine the speed at which the cycle "runs away." Second, we need to know how well-connected the sub-system or loop is to the system as a whole, so that its runaway loops have wider influence. There may be particular threshold effects: beyond a certain point, the runaway subsystem triggers other sub-systems, which in their turn begin also to "run away." The particular dynamics that emerge will be heavily dependent on the way that elements are connected to each other.

Nevertheless, Powell's QSD assumes that the configuration of elements — in Figure 5 for example — remains fixed. There are no novelties and no extinctions. He does not allow for the sub-system that runs away to change its configuration or to re-shape the larger system in which it is embedded. Powell advances our quest: but he does not provide us with a co-evolutionary dynamic.

## 4.2 Autocatalytic Sets

Jain and Krishna (2003) are interested in autocatalytic sets (ACS) and the key role these appear to have played in the origins of life. An ACS comprises a set of simple molecular organisms, none able individually to self-replicate, but each providing a catalyst for each of its fellows — a process of symbiotic and co-evolutionary "boot-strapping" for collective self-replication (see also Kauffman, 1993, Ch. 7). This is of central relevance to any Darwinian account of the origin of species.

Like Powell, Jain and Krishna employ a methodology of directed graphs (networks where the *direction* of the connection matters).[11] They model a population located at each of the different nodes, dependent on the growth of population at a number of other nodes. Their first step is to watch which nodes thrive and which do not — and the significance of the ACSs for this variation in fortune. They notice (Section 3) that the population of a node enjoys particularly rapid growth if it is part of an ACS. An ACS thus plays a role in Jain and Krishna similar to that of "runaway loops" in Powell.

So far, this analysis by Jain and Krishna — like that of Powell — does not allow for any reconfiguration of the system or any transformation of its elements. This we will refer to as the *"fast" dynamic*, as the nodes within a given configuration of the network thrive to different extents, their populations eventually reaching an equilibrium. Jain and Krishna then however apply a Darwinian rule, to effect a re-configuration of the network (Sections 4–5). This we will refer to as the *"slow" dynamic*, occurring only once the fast dynamic has reached an equilibrium. From among the nodes that have flourished the least, a randomly chosen few are extinguished; random new nodes and connections are then added, mimicking the mutations in Darwin. Such reconfigurations are of course central to any evolutionary framework.

This formal analytic of the fast and the slow dynamic thus captures co-evolutionary dynamics in elegant and parsimonious fashion. The fast dynamic sees the various nodes or species thriving differentially, by virtue of their interactions with their fellows. It is those nodes associated with ACSs that thrive the most; just as in Figure 3, it was elements **A22** and **B21** that thrived by virtue of their mutual affinity. As with Darwin, the slow dynamic sees those nodes or species that fail to thrive being extinguished, allowing new nodes or mutations of species to discover a niche for themselves.

---

11.   Also available at http://arxiv.org/PS_cache/nlin/pdf/0210/0210070v1.pdf
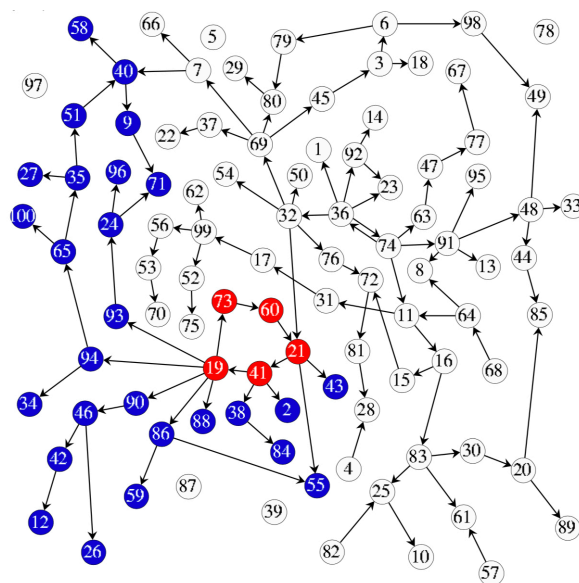
Figure 6. Jain and Krishna's Autocatalytic Sets

*Red nodes belong to the core of the dominant autocatalytic set of the graph, blue nodes to its periphery, and white nodes are outside the dominant autocatalytic set. This diagram shows run 6062 in the computational modelling of the graph (Jain & Krishna, 2003). Reproduced by kind permission of Wiley publishers and the authors.*

Jain and Krishna use computational models to simulate what can happen in such a dynamic system, depending on the parameters and algorithms adopted. For one illustrative moment in these successive transformations, see **Figure 6**. One ACS has developed, centred on the five nodes coloured red, but also benefitting a range of blue nodes, which thrive on their accelerated development, while not themselves contributing to the self-reinforcing feedback loops. The white nodes however are left without benefit from these elective affinities; they will therefore provide the low fitness candidates for extinction. The subsequent reconfigurations associated with the slow dynamic may occasionally enable new ACSs to form, undermining those which currently dominate, and producing dramatic shifts or collapses within the connective geometry of the system.

Computational models of this sort can serve as ideal types, illuminating the range of empirical dynamics to be found in the real world (Gilbert & Troitzsch, 2005). To apply such models empirically is likely to be rather demanding in terms of data, depending as it does on details of connections across a complex system, collected repeatedly and in timely fashion. This is an obvious example of where "big data" analysis may play a key role — on the one hand drawing on the large-scale administrative data that are routinely amassed and regularly updated; on the other hand, deploying modern computational capacity to scan such data sets, for the "runaway loops" and the ACSs they reveal (Mayer-Schonberger & Cukier, 2013).

### 4.3   Eigenvalues and Eigenvectors

A linear system tells a story of one-way influence or determination. Rising unemployment causes growing poverty; rising obesity causes higher rates of diabetes. The independent variables exert a force on the dependent variable — Newton's mechanics applied to the social world.

In a linear system, as captured in Figure 1, the $x$ variables are independent both of each other and of the dependent variable. What however if there are feedback processes, of the sort depicted in Figure 2, including a variety of "cycles" (loops), such as **z-y-h** and **x₃-y-k**? These cycles allow self-reinforcing forces to develop, similar to those highlighted by Powell and by Jain and Krishna. The influence of the independent variables on the dependent variable is no longer one-way; and the independent variables are no longer isolated from each other. Here instead is a network of interacting nodes, with the rate of activity on each node determined by the activity levels on the other nodes to which it is connected.

Mathematically, we can present the connections of such a network as a matrix (the "adjacency matrix"), with as many cells, both vertically and horizontally, as there are nodes. The cells of the matrix show whether any two nodes are connected and, if so, in which direction. This could, for example, be done for the networks in any of the Figures 2, 5 and 6.

In equation (1), the function $f_i(x_i)$ showed how $y$ derives from $x_i$. In equation (3) the adjacency matrix **C** plays a somewhat analogous role. The left-hand side of the equation is a vector whose component $da_i/dt$ measures the change in activity level $a_i$ on node $i$ that derives from the activity levels on the other nodes. It thus allows us to see how the activity levels on other nodes drive the change in activity level on whichever node is of interest.

$$d\mathbf{a}/dt = \mathbf{C} \cdot \mathbf{a} \qquad\qquad (3)$$

This matrix **C** of interactions can be expressed more parsimoniously, in terms of its **eigenvalues** and the corresponding **eigenvectors**. Together they summarize the dynamic evolution of the network with elegant simplicity. The eigenvalues display the strength of the self-reinforcing forces unleashed by the interactions among the nodes; the eigenvectors show where in the system these forces act and how they partition the network into corresponding zones.

A matrix such as this may have a number of different eigenvalues and corresponding eigenvectors. It is the eigenvalue of largest (absolute) value that will dominate the dynamics of the system as a whole. By identifying that eigenvalue and its corresponding eigenvector, we can calculate the final configuration of activity levels it will produce — the "attractor" or equilibrium towards which the fast dynamic is taking the system, under its present configuration.[12]

This is the fast dynamic. Now however the Darwinian cull of one or more low-fitness nodes is effected; new nodes arrive and occupy vacant positions. The slow dynamic has thus produced a network with a new configuration; the fast dynamic can be run afresh.

Again, when they study the consequences of the slow dynamic, with shifts or even collapses within the connective geometry of the system and the new ACSs that may form, Jain and Krishna attach major significance to the eigenvalues and eigenvectors of their network (Section 6 of their paper). In applying their approach empirically, the eigenvalues and eigenvectors will likewise be our key tool for making sense of the ACSs that develop and their likely dynamics.

## 5    Contingency and the Arts of Civilisation

Co-evolution was recognized by Darwinian biologists as a blind process of the natural world. It was also a *contingent* process, as Gould (1991) has conclusively argued; if the biological "tape of

---

12.    There is a parallel to be drawn with the GLM and equations (1) and (2). There we might estimate the importance of the different **x** variables in predicting the value of **y** and rank them by reference to the proportion of variance they explain. We might then, in a spirit of parsimony, include lower-ranked x variables only up to the point where we have explained our desired proportion of variance.

history" were to be "replayed" many times, altering some apparently minor detail in that chain of contingencies, each would have produced a quite different result. Similarly, when Jain and Krishna run successive simulations of their model, small shifts in the algorithms of extinction and mutation can produce quite different trajectories. The fruit of any such enquiry is not therefore a set of "universal truths," of the sort to which the GLM aspires, but instead some "timebounded truths" about the contingent dynamics of change (Brown & Langer, 2011).

Few of the "replays" of Gould's biological "tape of history" would have produced anything like *homo sapiens*. That however made all the difference. *Homo sapiens* now purposefully intervenes in the contingent dynamics of biological co-evolution, shifting the algorithms of extinction and mutation and the terms on which the tape of history is played. Thus in some degree, our species makes its own history, not least in regards to the biological world of which Darwin wrote and where human interventions are producing new mass extinctions.

This article however is concerned with the coevolutionary dynamics of *social* rather than *biological* life. The complex dynamics of the social world are in some degree as blind and emergent as the dynamics of natural world, without collective purpose. Nevertheless, purposeful interventions by social actors are omnipresent; and this entails struggles for positional advantage and the exercise of power. It is necessary therefore to consider how to conceptualize this purposeful agency in relation to the co-evolutionary dynamics of the social world.

Those who champion evolutionary ideas and want to apply them to the social world disagree as to *what it is that evolves*. For Dawkins, it is a matter of understanding social dynamics by reference to the demands of biological evolution. It is for example by reference to the "selfish gene," that we should understand the evolution of cooperation and altruism (Dawkins, 1976). Sloan Wilson (2008) likewise retains a strong focus on the biological substrate of human behaviour.

Evolutionary economists in contrast leave no place for biological selection.[13] In Darwinism, it is the genetic legacy of a species that is re-worked; in evolutionary economics, it is the technological and institutional legacy of a society (Potts, 2000; Hodgson, 2002; Crouch, 2005; Beinhocker, 2007). New "variations" in technology emerge from the "animal spirits" and inventiveness of entrepreneurs, in what Schumpeter described as "swarms of innovation."

Just as the pigeon breeders and horticulturalists described by Darwin (1859/1998, Ch. 1) looked out for novel characteristics in the offspring of each new generation, entrepreneurs are forever on the lookout for new technologies whose coevolutionary dynamics can open new markets and yield disproportionate returns. Such dynamics may entail co-evolution between different technologies; between new technologies and new markets; between new forms of industrial organization and new systems of public regulation, etc. To discover and nurture such dynamics is central to what we may call the arts of civilization (Bronowski, 1981: Chs. 2–4).[14]

How a given technological innovation will then fare — and how it may interact with other technologies and institutions — can never be entirely foreseen. Entrepreneurial ingenuity may *propose* new variants: but it is processes of differential selection through the market that *dispose*;

---

13. This does not mean overlooking that human beings are biological organisms. They feed on other organisms; they are vulnerable to the ravages of new viruses; much of their economic and social activity is geared to the collective management of these challenges (Flannery, 1994, Part 2). Nevertheless, the variations thrown up in their social and economic technologies — and then variously selected and retained — are not biological. It is in this narrow but crucial sense that the analysis of societal evolution can and should ignore the biology.

14. This is very much in the tradition of economics writing on "cumulative causation" including in particular Kaldor (Toner, 1999). It contrasts markedly with orthodox economics and its preoccupation with market "equilibrium" (Kaldor, 1972). Also relevant here is the literature on national innovation systems (Lundvall, Intarakumnerd, & Vang, 2006).

and these can seem just as collectively "blind" and devoid of overall intent as the processes of natural selection that drive speciation in the wild. To understand the economy in such evolutionary terms — and more generally as a complex system with emergent features — has attracted growing numbers of heterodox economists.[15]

Even as those collectively blind processes unfold however, strategically purposeful social actors will attempt to modify them to their own ends — depending on the resources and power at their disposal. This is a struggle for positional advantage. Attempts to apply evolutionary models to the social world have in general neglected such exercise of power. We however bring power centre-stage — all set within the hierarchical relationships of domination and dependence which characterize our human societies.[16]

If Figure 3 — our generic model of coevolutionary dynamics — is to inspire our social enquiry, it must be seen as playing out within those relationships of power and domination — and the institutional rules and regimes in which they are embodied. That is for example why, when we come to examine the co-evolution of technologies, we will also engage with the multi-level institutional processes within which new patents establish their intellectual property rights. In the biological world, the simple selective dynamics summarized in Figure 3 play out across complex multi-level food webs; in the social world the selective dynamics, through which technologies and capabilities develop, play out across contested multi-level institutional webs of positional advantage.[17]

We will henceforth speak of the **Contingent Historical Model** (CHM) as our alternative to the General Linear Model: integrating blind co-evolutionary dynamics with purposeful agency and the struggle for positional advantage (Room, 2011; 2016). It is this model that we seek to apply empirically, in ways that engage with scholarly concerns. We will also assess the practical value of Jain and Krishna for policy actors, in detecting early or "weak" signals of impending change, and applying the arts ofcivilizationn to the policy world as it unfolds.[18]

## 6   The Investigation of Empirical Dynamics

We now consider how the mathematics of co-evolving systems — in particular the model of Jain and Krishna — can be applied to empirical datasets — and the desiderata those datasets must fulfill.

---

15. These include Arthur (1994) Dopfer and Potts (2008), Hodgson (2004), Loasby (1999), Metcalfe and Foster (2004), Nelson and Winter (1982), Witt (2003), and beyond them Penrose (1959). For a more popular treatment see Beinhocker (2007). For my own critical reading, see Room (2011, Ch. 4).

16. This raises the question of how we are to characterize social action in relation to emergent phenomena within co-evolving systems (and indeed complex systems more generally). On this see my discussion of agile action and such transformational processes (Room, 2016, pp. 22–24) and compare Manzo (2007).

17. It is from this standpoint that I critically read the literature on evolutionary models in social science. That literature includes for example Thelen (2004) and Steinmo (2010), but also North (2005), Pierson (2004), Fligstein (2001) and Crouch (2005). For my own critical reading and attempt to re-work these contributions into an ontologically and methodologically appropriate form for social enquiry, see Room (2011, Chs. 6–8; 2012).

18. Here again, it is to the eigenvalues and eigenvectors of such co-evolving systems that the work of Jain and Krishna suggests we should look, in detecting such "weak signals" of change. This is analogous to the monitoring and control of complex engineering systems such as nuclear reactors, where engineers have a dashboard displaying (in effect) the eigenvalues and eigenvectors of the system, as they shift over time (Stewart, 1997, pp. 96–97; pp. 317ff). This involves discontinuous interventions — the seizing of critical moments, the throwing of particular switches — rather than a continuous and smooth process.

In the present article, we focus on databases of patents, and their adequacy for studying the empirical processes by which new technologies co-evolve. This empirical project is the initial focus of the DCICSS research group.[19] Case studies in other empirical areas will follow. Depending on their distinctive features, we may need to modify the model; and there are always likely to be trade-offs between the realism of the model and its mathematical tractability.

We construe the technologies of the modern world as a connected network of loci or nodes. The level of activity on each node is affected by that at the other nodes to which it is connected. This is the "fast dynamic" of Jain and Krishna. The pattern of outcomes in such a vast connected system cannot however be predicted in advance, as the simple aggregation of activity at the micro-level — it is blind and emergent and the outcome may be counter-intuitive. This is a general feature of complex systems (see for example Schelling, 1978; Squazzoni, 2012).

The "slow dynamic" is different. It involves the extinction of nodes with low levels of activity and the introduction of new nodes and connections. In Jain and Krishna, these are random novelties. It can also however be effected by purposeful selection by social actors. This may happen at the local level, for example when inventors produce new technological devices (Koenig, Battiston, & Schweitzer, 2008), a few of which find elective affinities that enable them to thrive disproportionately. It can happen on a larger scale, when the big actors of government and the corporate world consider the emergent macro-outcomes of the fast dynamic and purposefully intervene, intending to steer them in new directions.[20]

Against this conceptual background, **Table 1** summarizes the *desiderata* for the datasets we will require, when we study non-linear dynamics within co-evolving systems, using the Jain and Krishna model, and the tasks that will be involved. We will be able to judge, by reference to this Table, whether the datasets available to us, for any empirical case study, sufficiently meet these requirements.

As we shall see, the process of mapping data from empirical datasets into our model of co-evolving systems is itself highly demanding. We confront, first, the challenges facing any researcher who uses large databases, in checking the reliability and coverage of the data and their consistency between countries. More than this however, we are mapping these data not into the rather simple architecture of the GLM, as captured in equation (1), but into the tangled societal counterpart of the "tree of life," as captured in equation (3), re-shaped over multiple periods of co-evolutionary dynamics and purposeful agency, within a changing institutional architecture and political economy.

Section 7 will apply these desiderata to databases of patents, as a vehicle for studying technological innovation. Section 8 examines how far an empirical network of co-evolving technologies can thereby be constructed, as a contingent historical process. Section 9 will bring together our findings.

---

19. For details of the DCICSS project, centred at the University of Bath, see *Dynamics of Cumulative Innovation in Complex Social Systems*: https://www.bath.ac.uk/projects/driving-socio-economic-development/. It benefits from the contributions of Evangelou Evangelis, Orietta Morsili, Lorenzo Napolitano, Emanuele Pugliese, Alastair Spence, Paolo Zeppini.

20. It can also involve meso-level actors mobilising from below, to capture those global dynamics and impose agendas of their own (Kristensen & Zeitlin, 2005).

## 7   Patents as Technological Innovations

### 7.1   Introduction

Technological innovation is a key feature of the modern world, driving all else, as celebrated by such diverse luminaries as Adam Smith, Marx and Schumpeter. How might we study the dynamics of technological innovation, as innovators combine existing and new technologies in novel ways? What sort of datasets might allow us to track these changes over time and better anticipate their direction?

Databases of patents have developed over the last century and are harmonized internationally. Many scholars have used them for studying different aspects of technological innovation; there is a rich and self-critical literature on which we can build (Napolitano et al., 2018).

The World Intellectual Property Organisation (WIPO) is the forum charged with overall governance of the international patent system; it includes the major national patent offices and the European Patent Office (EPO). WIPO is responsible for the global patent classification system and its annual updates; EPO is responsible for publishing the global PATSTAT database, set within the latest version of the classification system.

A patent can be viewed from two standpoints (Strumsky, Lobo, & Leeuw, 2012). First, it constitutes a new capability, a *force of production*. It combines and applies knowledge and technologies in new ways, whether incremental or more radical. This combinatorial ontology is central to much of the innovation literature, including Schumpeter and Hayek, and more recently Teece (2009) and Potts (2000).[21]

Second, a patent is a claim to novelty — and one whose commercial potential the inventor wishes to protect. A patent application acknowledges the "prior art" on which it builds, its intellectual debts, but it also makes clear what is new. A patent thus constitutes a claim to intellectual property from which others will be excluded — part therefore of the *relations of production*. Once a patent is granted, the inventor may develop it commercially. It may alternatively be sold, with the new owners either exploiting or shelving it, for a period at least, to avoid it threatening their existing product lines and markets.

As both forces and relations of production, patents remind us that technological change takes place within particular legal and institutional settings. These vary between countries and over time; they are socially and politically constructed and contested (Polanyi, 1944). Relevant here is the literature on "varieties of capitalism" (Hall & Soskice, 2001) and "national innovation systems" (Lundvall, Intarakumnerd, & Vang, 2006). A study of patents may serve to illuminate the co-evolution of technologies and institutions in different political economies. This does not mean that these are mutually insulated national domains. On the contrary, the patent regimes of individual countries can have consequences for innovation elsewhere, not least through the interconnections woven by international companies and their production chains.

It follows that a study of patents should serve to illuminate major theoretical and empirical questions that are central to sociological enquiry. It should allow us to lay bare alternative trajectories of development and the scope for intervention by policy makers and corporate actors, in pursuit of those alternatives. In thus demonstrating the value of models of co-evolving

---

21.   This contrasts with much of orthodox economics, which assumes a production function at the frontier of technology, shifting in response to technical progress, by reference to which businesses assess the profitability of different production mixes, as technology takers rather than technology makers.

**TABLE 1**

***Steps in the extraction of data and the construction and analysis of the empirical network***

*BASIC COMPONENTS OF THE NETWORK AND ITS EVOLUTION*

- A directed network of nodes and edges, supporting populations with interdependent levels of fitness
- A "fast dynamic" that plays out the interactions among nodes and reveals the levels of fitness they each enjoy
- A "slow dynamic" which eliminates the least fit nodes and their edges and introduces novel replacements

*MINIMUM REQUIREMENTS FOR THE DATASET*

- Operationalise nodes, edges and "fitness" by reference to the database and over a succession of time periods
- Specify how the fast and slow dynamics will be observed within the data
- Recognise the database as a social artefact, the "arts of civilisation" applied to the social world, and specify how the database will be exploited in this regard.

*THE EMPIRICAL ANALYSIS*

*For each time period:*

- Construct the adjacency matrices of the network (depending on the database, this could be at various levels of granularity).
- Examine the network for the presence of autocatalytic sets (ACS).
- Measure the levels of fitness of the different nodes and examine whether these are higher for nodes involved in an ACS.

*For successive time periods:*

- Map the eigenvalues and eigenvectors of the network and how they change over time
- Evaluate the eigenvalues and eigenvectors as revealing different types of innovation

*THE SCHOLARLY AND POLICY DEBATES*

- Consider the significance of these findings for research debates on innovation and change, within the fields in question
- Assess the utility of such analytical methods for social scientists, and their practical value for policy actors, as "weak signals" of impending change and as tools for monitoring interventions
- Assess the implications for research into co-evolving systems across other empirical areas.

dynamics for the study of patents, we hope to show that such models deserve greater attention from the social science community at large.

## 7.2    Components of the Network

We have obtained privileged access to the PATSTAT datasets, which incorporate details of the patents registered each year, across all major patent regimes globally. We make selective use of these datasets to establish the components of our network.[22]

*Nodes*: The PATSTAT database allocates each patent, newly registered in a given year, to a particular class and sub-class of its overall classification system — a dendrogram that is readily searchable. These classes and sub-classes we take as the nodes of our network. We can then study our network at different levels of granularity, ranging from the overall classes down to finer distinctions among sub-sub-classes etc.

*Edges*: Each patent that is registered records other patents on which it draws and to which it is thus indebted: the "prior art" by reference to which its own distinctive novelty can be viewed. The PATSTAT database thereby links each patent (and the class and sub-classes within which it sits) back to the classes of those other patents, from whose prior art the patent in question benefits. This is a flow of know-how from those earlier patents and their patent classes to the latest innovations. We take these citations as the basis for building the edges or links of our network, representing flows of knowledge among patent classes.

## 7.3    Observing the Fast Dynamic

In the Jain and Krishna model, the interactions among nodes during the fast dynamic drive the populations that thrive at each node and thus their fitness. In our case study we take this as given by the number of new patents registered in a given class (node) during a given time period.[23]

Once we can with some confidence identify the dominant ACS that is emerging, we can also identify the equilibrium towards which this particular episode of the fast dynamic is taking the system. This will also reveal the least thriving technologies and patent classes and thus the nodes that are candidates for extinction. More generally, identifying the dominant ACS will reveal what we might call the *zone of intensive innovation* and the *zone of stagnation*.

## 7.4    Observing the Slow Dynamic

In the Jain and Krishna model, the algorithms of the slow dynamic extinguish one of the least fit nodes and add a new node, a random mutation. Only then do they run a new episode of the fast dynamic, on this modified network.

In the real world of patents, new nodes (in the form of new patent classes and sub-classes) appear more frequently. Every year WIPO publishes an updated version of the patent classification system, with EPO publishing the latest PATSTAT database, using that updated classification system. Both will typically include more than a few new classes or sub-classes.

---

22.    This particular application of the Jain and Krishna model to the PATSTAT data was originally proposed by Zeppini (2017).

23.    An alternative would have been to take the *total* number of patents in a given class (not just the newly registered) as our definition of fitness; but this would distract attention from the zones of most intense innovation with which we are most concerned: see section 8 below.

If the introduction of new nodes is here much more plentiful than the frugal mutations allowed by Jain and Krishna, the elimination of patent classes is in comparison much more grudging. Few patent classes or major sub-classes are ever wholly eliminated from the classification system, they are just left to stagnate. It seems reasonable, however, to focus primarily on the new zones that open up and the zones from which they primarily draw knowledge. Patent classes long sterile can reasonably be ignored, whether or not the WIPO classifications and the PATSTAT database retain them.[24]

For Jain and Krishna, new edges are associated with the newest nodes. It may be objected that in the real world of technological innovation, new edges are often associated with existing nodes. We might explore some corresponding mathematical variant of the Jain and Krishna model. Nevertheless, the sense of the model is surely this: that we should focus on the new nodes and edges that drive change more generally, and without which the existing system will tend to stasis.

Again therefore, our main focus is on zones of intensive innovation — zones where the annual addition of new nodes (classes and subclasses) and edges is most lively. We investigate how far these are home to the dominant ACSs — and whether, by watching the eigenvalues and eigenvectors of the system, we gain insight into the emerging zones of change.

## 7.5 Grasping a Contested Social Construct

Now however we face a rather different a sort of challenge. We see and discuss objects using the concepts and language available to us, not least in regards to new technologies. Such concepts are socially constructed and contested. This does not imply that those technologies have no reality of their own — a reality which (although mediated by those concepts) we may attempt to explore and analyze. Nevertheless, we must take account of the social processes by which these concepts are constructed and modified. This is all part of what we earlier referred to as the "arts of civilisation" — involving human reflection, experimentation and the growth of knowledge, all set within a struggle for positional advantage.

The international patent system is one of those processes. An inventor registers a patent, to establish a claim of intellectual property, but it may then take several years, before the patent is granted. During this period the inventor is in negotiation with the patent office, seeking an agreement as to what is new, about this particular patent — how the novelty should be conceptualized and classified, within the overall patent system. That system is itself however in flux.

WIPO publishes its annual revision to the classification system, adding sub-classes here, or merging sub-classes there, or even adding whole classes elsewhere. This ensures that the classification system keeps up to date with major reconfigurations of technologies; and it allows inventors to search more easily for the prior art, when they are making new applications for fresh patents.

---

24. A technology (class) such as the steam engine may at some point cease to be a zone of innovation and no longer attract any new patent applications. This is not because the steam engine has necessarily exhausted its potential for innovation, but rather because other technologies and the dynamics of economic development have diverted innovation and investment to other areas — this in response to new openings that have appeared, in part out of the successes of the steam age. It should never however be assumed that those old technologies will never have new applications — see for example windmills and ceramics, re-invented for the modern age (Tomlinson & Branston, 2014). Thus technological development can open up new technological vistas which cannot be wholly predicted — and which may indeed appear rather like a random new node arriving.

The EPO also publishes annually its latest PATSTAT database of patents worldwide. This publication re-allocates patents across the revised classification system, so that they can all be viewed by reference to this most up-to-date picture of the technology system, as it exists today. This includes not only the new patents that inventors applied for over the past year, but also the patents from previous years that were already on the database. An older patent may now be re-assigned to a class or sub-class different from that which it originally occupied; and to one, indeed, that may have been only recently created.

This presents us with a significant conceptual and methodological challenge, as we seek to map our network of nodes and edges and the dynamics in which they have been involved over successive years. This is especially the case for areas of technology that are zones of innovation — the most interesting for us — because it is here that the patent officers will in general have been most active in modifying and updating the system of classification. The classification system is itself evolving — and it is therefore a shifting vantage point from which to view the co-evolution of technologies.

## 8    Patents as a Contingent Historical Process

This conceptual and methodological challenge nevertheless opens up an important opportunity for our analysis of innovation. Far from the selection of patents as our case study being ill-advised, it will prove to have been particularly apt.

We have thus far been interested in how technologies co-evolve with each other, in autocatalytic processes as modelled by Jain and Krishna. Nevertheless, we grasp technologies through the institutions by which we organize them. The pace and direction of innovation depends, indeed, as much on the new institutions that are being invented — laws on e-commerce for example and on new forms of intellectual property — as on the new technologies themselves. The designers and inventors of these institutions (in government, in business, etc) are re-shaping the world, no less than the designers and inventors of technologies.

In this way, **institutions** and **technologies** interact and co-evolve with each other. Many new technologies emerge, only to find themselves on an institutional terrain that stifles them. In some other country, the institutional terrain may be more supportive. Some technologies and institutional forms may thus discover an "elective affinity" that enables both of them to thrive and together to dominate their socio-economic ecosystem.[25] These *elective affinities between technologies and institutions* are therefore as potentially significant, as those between different technologies, in generating autocatalytic effects and in driving processes of innovation.

The international patent system is one such institutional domain; its officers are one group of institutional inventors. How then shall we think of the autocatalytic dynamics involved in the development of new technologies and their registration as patents?

Consider **Figure 7**. We start from the **north-east quadrant** and proceed clockwise.

A new year has started. WIPO has updated the classification system it published 12 months earlier, involving a dendrogram of classes and sub-classes. EPO has published the latest version of the PATSTAT database, locating all patents, both recent and not so recent, within this most recent version of the classification system. It thus maps the technological capabilities of

---

25.   This co-evolution is not entirely blind. The inventors of new technologies pay attention to changes under way in laws on IPR, e-commerce etc. This will influence their decisions as to where they invest their time and creativity; and as new institutional spaces are created, they may shift their focus. Meanwhile, institutional inventors watch what new technologies are emerging, when they consider how to modify the legal and administrative environment.

the society. It also records the knowledge claims that are officially recognized, and hence the flows of know-how on which successive generations of patents relied — all organized in as parsimonious, searchable and up-to-date a form as the international patent officers consider possible. This mapping will now provide the stable framework that inventors can use, to search the "prior art" and to register their new patents over the coming months. It will also allow patent officers to evaluate those new applications.

> **We will ask: Q1:** *What technological capabilities do these patent classes capture — and what flows of know-how, from established patent classes (and sub-classes) to new patents, did they involve? We construct an empirical network, with* **patent classes as nodes** *and* **knowledge flows as edges.**

We return to this NE quadrant, after working our way through the other three.

This moment of publication is the starting gun for the registration of new patents. Within **the south-east quadrant**, inventors start registering new patents and populating the classes of the newly-updated classification system. They cite a variety of existing patents (and thus patent classes) as the antecedents on which they have drawn in novel ways.

This is the *continuous present of the technological fast dynamic*. Within this quadrant, the classes are given and fixed, just as with the nodes in Jain and Krishna. Inventors anchor their claims to novelty within this stable substrate, the prior art; it is only thanks to the parsimonious clarity and searchability of the classification system, that the novelty of their own invention can be rigorously demonstrated.

> **We will ask: Q2:** *How do the various nodes of the network (in this case, the classes and sub-classes) thrive differentially? What is their resulting* **fitness** *— the number of patents registered in each class or sub-class during the year in question?*

Even as the inventors register their grounding in the prior art, they also document the novelty of what they offer. Some of these novelties will however in due course, and taken together, abrade against and undermine the classification system, even though it has only recently been updated. Here therefore is a stable classification system, but one across which there progressively spreads a mass of prospective instabilities. In this way, the continuous present unfolds on the stable ground of the past, but also defies and challenges it.

In places, the sheer numbers of new patent applications threaten to overwhelm the categories' capacity to describe and contain the promised novelties, without blurring or losing their distinctiveness. In other places, the novelties defy the categorical boundaries and the distinctions they draw. This makes it increasingly difficult for the national patent officers to apply the classification system as it stands. We might indeed describe the pressures for such revisions, and their distribution across the array of patent classes and sub-classes, as the rate of activity — the fertility — of different parts of the *classification system*. This is the *institutional fast dynamic*, located in the **south-west quadrant**.

> **We might ask: Q3:** *What are the mounting pressures for revision to the classification system and how are they distributed across the different classes and sub-classes? What limitations and inconsistencies in the classification system do these pressures expose?*

| | INSTITUTIONS | | TECHNOLOGIES |
|---|---|---|---|
| **The slow dynamic – a single moment every 12 months** | **Institutional slow dynamic**<br><br>**Patent office simplifies the system of classification - so that it is parsimonious in terms of classes and cross-citations across old as well as more recent patents.  It reassigns all patents to their appropriate place within this amended classification system** | → | **Technology slow dynamic**<br><br>**The amended system of classes is published, for inventors to populate with new patents** |
| **The fast dynamic – a continuous process over 12 months** | ↑ **Institutional fast dynamic**<br><br>**Pressure mounts on patent offices to make ad hoc adjustments to the classification system** | ← | ↓ **Technology fast dynamic**<br><br>**(New) patents populate (new) nodes** |

Figure 7. Patents as CHM

Nevertheless, these micro-processes are largely invisible; we cannot view them directly within the patent publications.[26]

This is a fertile, creative and expert activity, a practical craft on the part of the national patent officers, as they consider what adjustments would be needed to provide all these new patent applications with an appropriate niche, capturing and displaying its distinctive novelty. As in all scientific work, such classification is an essential tool in building conceptual links for understanding and progressing the architecture of knowledge (Bowker & Star, 2002: Ch. 9). It goes beyond mere categorization and search; for it helps ensure that each patent comes to the notice of other relevant strands of invention, enabling cross-pollination and the discovery of further elective affinities.

That task is taken further at year end by their WIPO colleagues at international level, with their annual updates of the international classification system. Now, in the **north-west quadrant**, we move *from the fast to the slow dynamic*; from the video of the continuous present, to looking back from the end of each year, a careful snapshot, a *pause for constructive reflection*. WIPO reads the reports and recommendations from the national officers. It takes stock of the pressures they have encountered for adjustment of the classification system and the conceptual tangles these have produced. It considers how to simplify and untangle the web of citations as economically as possible. In information theory, this parsimony appears as a counterpart to the export of entropy in physics. Elements of the classification system serving no useful purpose are eliminated, as surely as the least fit species in an ecosystem.

> ***We might ask: Q4:*** *What rules and processes of sorting and re-working of the classification system do the national and international patent officers follow, in their efforts to maintain a parsimonious, searchable and consistent system?*

It is useful to distinguish between new patents registered in a given year, using the classification system published at the start of that year, and patents registered in earlier periods, but now re-classified by reference to the current system. The *ad hoc* modifications made during each year, by the national patent officers within the south-west quadrant, are driven primarily by the former. The latter also matter however; and when the international office considers what annual revisions to make, it will need to weigh up how easily those earlier patents can be mapped into the revised classificatory matrix that the newest patents suggest, while still maintaining a parsimonious and searchable scheme.[27]

This is the institutional slow dynamic — with WIPO looking back and checking the consistency and elegance of the various ad hoc adjustments that their national counterparts have

---

26. Lafond has been one of the scholars who has made some empirical study of these processes (Lafond, 2014; Lafond & Kim, 2019). He notices that they follow a strict process of checking and search, when they evaluate any new patent application, and search for the relevant prior art. He notices also that faced by a flood of new novelties, they tend to adopt a pragmatic two-fold approach: on the one hand adding extra sub-sub-classes, to permit finer distinctions; on the other merging (sub-)classes into a completely new class or sub-class, when sufficient novelties put the distinction between them in question. The first of these involves an "incremental" change to the classification system, while the second involves changes that are to some extent more "radical."

27. Revisions to the classification scheme may be postponed, if they risk rendering it more complicated and less searchable. In the same way, the addition of extra epicycles to the Ptolemaic picture of the universe seemed necessary at the time, to deal with new astronomical observations, but only produced a more complicated picture of the heavens (Kuhn, 1970). At the very least, WIPO will want to take its time. They seem to have found that their 12-month cycle is about the right periodicity for the technology innovation system: allowing the classification to retain its freshness and relevance, while also not rushing to adopt every modification that national patent officers might suggest.

proposed, on the basis of the previous twelve months. The new classification of technologies that WIPO now publishes — and the new version of the PATSTAT database that the EPO then publishes — is the result of this reflection. The slow dynamic is however not just a look back at the preceding year, simplifying and pruning the tangles of the classification system, after the myriad local adjustments it has suffered over the past months. It also looks forward and establishes the new classificatory terrain for the year ahead. This is the **north-east quadrant**, the slow dynamic in relation to technological forms. The new classification system will provide the stable framework that inventors will use, to register their new patents and their debts to the prior art, over the coming months.

Our progress around the four quadrants has thus left us with four empirical research questions. Q1 and Q2, concerned with the technology slow and fast dynamics, can be addressed directly using the PATSTAT database. At the end of each year, we have the newly published WIPO classification scheme and the latest PATSTAT database, as well as their predecessors of 12 months earlier. Some of these patents will have been registered only during the last 12 months; others are of longer-standing. This enables us annually to construct an empirical technology network, showing patent classes and sub-classes as the nodes, the citations and knowledge flows as the edges, the numbers of patents as indicating the levels of activity, as viewed from the standpoint of the new and updated classification system. We can identify ex post the growth, consolidation or decline of autocatalytic sets through the year. This is the core of the empirical analysis in which we engage.

Meanwhile however Questions Q3 and Q4 concern the evolution of the classification system but cannot be directly addressed through the PATSTAT database. The latest PATSTAT database leaves in the background the classification system published 12 months ago, which nevertheless during the year served as the immediate point of reference for inventors. Those inventions are instead now displayed by reference to the revised classification system that WIPO produced during the final weeks of the year. That revision was driven by the stresses and strains which those inventors placed on the classification scheme, obliging the national and international patent officers to bring their collective wisdom, experience and ingenuity to bear — the arts of civilization. Thus inventors benefitted from the legacy of classifications transmitted from the past; but they also then exposed the limitations in that legacy, laid bare by the very novelties they sought to register.

Of that journey through the south-west and the north-west quadrants we have no direct knowledge, unless we mount a fieldwork enquiry of our own, surveying the national and international patent officers. We can however, albeit not without some effort, track how patents registered in previous years have over subsequent years been reallocated to new positions, as the classification system has been revised. Within the DCICSS research group, Napolitano and Pugliese (2017) have created just such a unified database, mapping each patent against all the codes it had in different classifications in different years, so as to track not only changes in the classification system but also the detailed flow of patents through that changing system.[28] This could be used to study indirectly the pressures exerted on the classification system through each 12 months by the latest cohort of inventions.

These are the selective dynamics through which technologies and capabilities develop, played out across contested multi-level institutional terrains, in a process which we earlier referred to as contingent historical development.

---

28. This dataset (PATSTAT 1400) uses raw data provided by the EPO's PATSTAT office. It involves relevant data extracted from versions of PATSTAT published between 2007 and 2017.

## 9   Conclusion

We now bring together the results of this study and its preliminary findings on technology networks.

The previous section displayed the co-evolution of technologies and institutions, patents and classifications, in a contingent historical process. Centre-stage are zones where a proliferation of new technologies is manifest in patents that challenge the established classification system. It is the dynamics of this zone of intense innovation, in both technologies and institutions, that we seek to capture, using our model of co-evolving networks taken from Jain and Krishna.

Notice the relationship between Figure 7 and the earlier Figure 3. Figure 3 offered a parsimonious representation of co-evolutionary dynamics. It was the starting point for our presentation of the Contingent Historical Model — a radical alternative to the General Linear Model, which has dominated much of social science. It looked back from the ecosystems of today to the elective affinities that had developed, among the populations of different species, during earlier periods. Those affinities enabled some species to thrive, while others failed and became extinct. Only by revisiting those earlier periods, was it possible to make sense of the complex food webs that we see today.

Figure 7 has taken us on an analogous journey through successive time periods, in order to make sense of the patent system we see today, with its wide range of interrelated technologies, classified to clear and parsimonious effect. The journey alternated between the slow and the fast dynamics, and between innovations in technologies and in classifications, as we took transverse cuts through successive annual data sets. Those cuts involve alternating snapshots and videos. The fast dynamic tells of a continuing present; the slow dynamic looks back from the end of each year, a pause for both reflection and foresight. We thus disassemble the co-evolutionary dynamic into four moments, each of which we study by appropriate interrogation of the PATSTAT databases for successive years. With the data laid bare, from this succession of vantage points, it becomes possible to construct the empirical network of co-evolving technologies and to investigate the autocatalytics of change.

The process of mapping data from empirical datasets into our model of co-evolving systems is thus non-trivial. Instead of the rather simple architecture of the GLM, we face the complex process of technological and classificatory co-evolution, by which the patent database is regularly re-woven. There was no certainty that the PATSTAT database could be bent to our research purposes. In the event however, the selection of this database for our empirical study has proved particularly apt.

This therefore is the principal empirical finding of the present article: not yet a systematic analysis of the autocatalytic empirics of technological development, but a necessary first step, in establishing the fitness-for purpose of this particular empirical database, and its exploitability in the service of the CHM. To this extent, the present article reports a conceptual and analytical approach and assesses the feasibility of a specific database for an empirical application of that approach.

Beyond this however, the DCICSS project has meanwhile published preliminary findings on the autocatalytic empirics of technological development, using the PATSTAT data (Napolitano et al., 2018).[29] This study incorporates many elements of the conceptual framework described here. It demonstrates that the evolution of the technology network does indeed involve a growing autocatalytic structure. It also confirms that the technology fields in the core of the

---

29.   https://royalsocietypublishing.org/doi/full/10.1098/rsos.172445

autocatalytic set display greater fitness, in terms of a greater number of patents. Finally, it reveals core shifts, whereby different groups of technology fields come to dominate the autocatalytic structure, only then to be overthrown. This points to radical innovation, with new combinations developing among distant as well as closely related technology fields. These are just the sorts of dynamic change that Jain and Krishna explore through their analysis of eigenvectors and eigenvalues. This empirical analysis will be extended in the next phase of the DCICSS work.

This demonstrates the promising possibilities for this form of analysis. The research also has potential benefit for policy makers, providing indicators — "weak signals" — of impending change, so that they can better steer the dynamics of autocatalytic innovation. It will need to be set within a wider institutional analysis, including the literature on national innovation systems (Lundvall et al., 2006) and varieties of capitalism (Hall & Soskice, 2001). It will also need to integrate co-evolutionary dynamics with a more thorough analysis of power and political economy (Room, 2016).

Beyond the present programme of work, the DCICSS project expects to apply the Jain and Krishna model, *mutatis mutandis*, to other fields of social, economic and technological change. It may of course be that in many potential areas of study, there are no appropriate databases immediately available. Nevertheless, this article has at least established the conceptual and methodological coherence and viability of such an empirical enterprise.

# References

Abbott, A. (2001). *Time Matters: On Theory and Method*. Chicago: University of Chicago Press.

Arthur, W.B. (1994). On the Evolution of Complexity. In G.A. Cowan, D. Pines & D. Meltzer (Eds.), *Complexity: Metaphors, Models and Reality* (pp. 65–81). Reading, Mass: Addison-Wesley.

Ball, P. (2004). *Critical Mass: How One Thing Leads to Another*. London: Heinemann.

Baumgartner, F.R., & Jones, B.D. (1993). *Agendas and Instabilities in American Politics*. Chicago: University of Chicago Press.

Beinhocker, E.D. (2007). *The Origin of Wealth*. London: Random House.

Blaug, M., Hodgson, G.M., Lewis, O., & Steinmo, S. (2011). Introduction to the Special Issue on the Evolution of Institutions. *Journal of Institutional Economics*, *7*(3), 299–315. https://doi.org/10.1017/S1744137411000270

Bowker, G.C., & Star, S.L. (2002). *Sorting Things Out: Classification and Its Consequences*. Cambridge, Mass: MIT Press.

Bronowski, J. (1981). *The Ascent of Man*. London: Futura.

Brown, G., & Langer, A. (2011). Riding the Ever-Rolling Stream: Time and the Ontology of Violent Conflict. *World Development*, *39*(2), 188–198. https://doi.org/10.1016/j.worlddev.2009.11.033

Brown, G.K. (2014). *The Limits of Linearity: A Modest Defence of the General Linear Model and of Its Critics*. Centre for Development Studies: University of Bath.

Brown, G.K., & Mergoupis, T. (2011). Treatment Interactions with Non-Experimental Data in Stata. *Stata Journal*, *11*(4), 545–555. https://doi.org/10.1177/1536867X1201100403

Checkland, P., & Scholes, J. (1990). *Soft Systems Methodology in Action*. Chichester: John Wiley.

Cilliers, P. (1998). *Complexity and Postmodernism*. London: Routledge.

Coyle, R. (1996). *Systems Dynamics Modelling: A Practical Approach*. London: Chapman & Hall.

Crouch, C. (2005). *Capitalist Diversity and Change: Recombinant Governance and Institutional Entrepreneurs*. Oxford: Oxford University Press.

Darwin, C. (1859). *The Origin of Species*. London: Wordsworth (reprinted 1998).

Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

Dopfer, K., & Potts, J. (2008). *The General Theory of Economic Evolution*. London: Routledge.

Flannery, T. (1994). *The Future Eaters*. New York: Grove Press.

Fligstein, N. (2001). *The Architecture of Markets: An Economic Sociology of Twenty-First Century Capitalist Societies*. Princeton: Princeton University Press.

Gavrilets, S. (2004). *Fitness Landscapes and the Origin of Species*. Princeton: Princeton University Press.

Gilbert, N., & Troitzsch, K.G. (2005). *Simulation for the Social Scientist* (2nd Edition). Maidenhead: Open University Press.

Gould, S.J. (1991). *Wonderful Life*. Harmondsworth: Penguin.

Gould, S.J., & Eldridge, N. (1977). Punctuated Equilibrium: The Tempo and Mode of Evolution Reconsidered. *Paleobiology*, *3*(2), 115–151. https://www.jstor.org/stable/2400177

Hodgson, G.M. (2001). *How Economics Forgot History*. London: Routledge.

Hodgson, G.M. (2002). Darwinism in Economics: From Analogy to Ontology. *Journal of Evolutionary Economics*, *12*, 259–281. https://doi.org/10.1007/s00191-002-0118-8

Hodgson, G.M. (2004). *The Evolution of Institutional Economics*. London: Routledge.

Holland, J. (1995). *Hidden Order: How Adaptation Builds Complexity*. New York: Basic Books.

Howe, R.H. (1978). Max Weber's Elective Affinities: Sociology Within the Bounds of Pure Reason. *American Journal of Sociology*, *84*(2), 366–385. https://www.jstor.org/stable/2777853

Jain, S., & Krishna, S. (2003). Graph Theory and the Evolution of Autocatalytic Networks. In S. Bornholdt & H.G. Schuster (Eds.), *Handbook of Graphs and Networks* (pp. 355–395). Weinheim: Wiley-VCH.

Jervis, R. (1997). *System Effects: Complexity in Political and Social Life*. Princeton: Princeton University Press.

Johnson, J., Nowak, A., Ormerod, P., Rosewell, B., & Zhang,Y.-C. (Eds.). (2017). *Non-Equilibrium Social Science and Policy: Introduction and Essays on New and Changing Paradigms in Socio-Economic Thinking*. Heidelberg: Springer.

Johnson, S. (2001). *Emergence: The Connected Lives of Ants, Brains, Cities and Software*. Harmondsworth: Penguin.

Kaldor, N. (1972). The Irrelevance of Equilibrium Economics. *The Economic Journal*, *82*(328), 1237–1255. https://www.jstor.org/stable/2231304

Kauffman, S. A. (1993). *The Origins of Order: Self-Organisation and Selection in Evolution*. Oxford: Oxford University Press.

Koenig, M.D., Battiston, S., & Schweitzer, F. (2009). Modeling Evolving Innovation Networks. In A. Pyka & A. Scharnhost (Eds.), *Innovation Networks. New Approaches in Modeling and Analyzing* (pp. 187–267). Heidelberg: Springer.

Kristensen, P.H., & Zeitlin, J. (2005). *Local Players in Global Games: The Strategic Constitution of a Multinational Corporation*. Oxford: Oxford University Press.

Kuhn, T.S. (1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lafond, F. (2014). *The Size of Patent Categories: USPTO 1976-2006*. Maastricht: UNU-Merit.

Lafond, F., & Kim, D. (2019). Long-Run Dynamics of the U.S. Patent Classification System. *Journal of Evolutionary Economics*, *29*(2), 631–664. https://doi.org/10.1007/s00191-018-0603-3

Lieberson, S. (1987). *Making it Count: The Improvement of Social Research and Theory*. Berkeley: University of California Press.

Lieberson, S., & Lynn, F.B. (2002). Barking up the Wrong Branch: Scientific Alternatives to the Current Model of Sociological Science. *Annual Review of Sociology*, *28*, 1–19. https://doi.org/10.1146/annurev.soc.28.110601.141122

Loasby, B. (1999). *Knowledge, Institutions and Evolution in Economics*. London: Routledge.

Lundvall, B.-A., Intarakumnerd, P., & Vang, J. (Eds.). (2006). *Asia's Innovation Systems in Transition*. Cheltenham: Edward Elgar.

Manzo, G. (2007). Comment on Andrew Abbott/2. *Sociologica*, *1*(2). https://doi.org/10.2383/24752

Marshall, A. (1920). *Principles of Economics*. London: Macmillan.

Mayer-Schonberger, V., & Cukier, K. (2013). *Big Data*. London: John Murray.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Maynard Smith, J., & Szathmary, E. (2000). *The Origins of Life: From the Birth of Life to the Origins of Language*. Oxford: Oxford University Press.

Metcalfe, J.S., & Foster, J. (Eds.). (2004). *Evolution and Economic Complexity*. Cheltenham: Edward Elgar.

Mitchell, M. (1996). *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press.

Napolitano, L., Evangelou, E., Pugliese, E., Zeppini, P., & Room, G. (2018). Technology Networks: The Autocatalytic Origins of Innovation. *Royal Society Open Science*, *5*(6), RSOS–172445. https://doi.org/10.1098/rsos.172445

Napolitano, L., & Pugliese, E. (2017). *PATSTAT1400 Multi-version Combined Database*. DCICSS.

Nelson, R., & Winter, S. (1982). *An Evolutionary Theory of Economic Change*. Cambridge, Mass: Harvard University Press.

North, G.C. (2005). *Understanding the Process of Economic Change*. Princeton: Princeton University Press.

Odling-Smee, F.J., Laland K.N., & Feldmann, M.W. (2003). *Niche Construction*. Princeton: Princeton University Press.

Penrose, E. (1959). *The Theory of the Growth of the Firm*. Oxford: Blackwell & Mott.

Pierson, P. (2004). *Politics in Time*. Princeton: Princeton University Press.

Polanyi, K. (1944). *The Great Transformation*. New York: Rinehart.

Popper, K.R. (1994). Models, Instruments and Truth: The Status of the Rationality Principle in the Social Sciences. In K.R. Popper (Ed.), *The Myth of the Framework: In Defence of Science and Rationality* (pp. 154–184). London: Routledge.

Potts, J. (2000). *The New Evolutionary Microeconomics: Complexity, Competence and Adaptive Behaviour*. Cheltenham: Edward Elgar.

Powell, J.H., & Bradford, J.P. (1998). The Security-Strategy Interface: Using Qualitative Process Models to Relate the Security Function to Business Dynamics. *Security Journal*, *10*, 151–160. https://doi.org/10.1016/S0955-1662(98)00020-4

Powell, P.L. (1992). Information Technology Evaluation: Is it Different?. *Journal of the Operational Research Society*, *43*(1), 29–42. https://doi.org/10.2307/2583696

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata*. College Station, TX: Stata Press.

Room, G., (Ed.). (1995). *Beyond the Threshold: The Measurement and Analysis of Social Exclusion*. Bristol: The Policy Press.

Room, G. (2011). *Complexity, Institutions and Public Policy: Agile Decision-Making in a Turbulent World*. Cheltenham: Edward Elgar.

Room, G. (2012). Evolution and the Arts of Civilisation. *Policy and Politics*, *40*(4), 453–471. https://doi.org/10.1332/030557312X133233392627832

Room, G. (2016). *Agile Actors on Complex Terrains: Transformative Realism and Public Policy*. London: Routledge.

Schelling, T.C. (1978). *Micromotives and Macrobehaviour*. London: W.W. Norton.

Shubin, N. (2008). *Your Inner Fish*. Harmondsworth: Allen Lane.

Sloan Wilson, D. (2008). Multilevel Selection Theory and Major Evolutionary Transitions. *Current Directions in Psychological Science*, *17*(1), 6–9. https://doi.org/10.1111/j.1467-8721.2008.00538.x

Solé, R., & Bascompte, J. (2006). *Self-Organization in Complex Ecosystems*. Princeton: Princeton University Press.

Squazzoni, F. (2012). *Agent-Based Computational Sociology*. Chichester: Wiley.

Steinmo, S. (2010). *The Evolution of Modern States: Sweden, Japan, and the United States*. Cambridge: Cambridge University Press.

Stewart, I. (1997). *Does God Play Dice? (2nd Edition)*. Harmondsworth: Penguin.

Stolzenberg, R. (2003). Book Review: Time Matters: On Theory and Method. By Andrew Abbott *Sociological Methods and Research*, *31*(3), 420–427. https://doi.org/10.1177/0049124102239082

Strumsky, D., Lobo, J., & Leeuw, S.v.d. (2012). Using Patent Technology Codes to Study Technological Change. *Economics of Innovation and New Technology*, *21*(3), 267–286. https://doi.org/10.1080/10438599.2011.578709

Tavory, I., & Timmermans, S. (2014). *Abductive Analysis: Theorizing Qualitative Research*. Chicago: University of Chicago Press.

Teece, D.J. (2009). *Dynamic Capabilities and Strategic Management*. Oxford: Oxford University Press.

Thelen, K. (2004). *How Institutions Evolve*. Cambridge: Cambridge University Press.

Tomlinson, P.R., & Branston, J.R. (2014). Turning the Tide: Prospects for an Industrial Renaissance in the North Staffordshire Ceramics Industrial District. *Cambridge Journal of Regions, Economy and Society*, *7*(3), 489–507. https://doi.org/10.1093/cjres/rsu016

Toner, P. (1999). *Main Currents in Cumulative Causation: The Dynamics of Growth and Development*. London: St Martin's Press.

Waldrop, M.M. (1992). *Complexity: The Emerging Science at the Edge of Order and Chaos*. London: Viking.

Witt, U. (2003). *The Evolving Economy*. Cheltenham: Edward Elgar.

Zeppini, P. (2017). *Autocatalytic Networks of Technologies*. Department of Economics: University of Bath.

**Graham Room:** Department of Social & Policy Sciences, University of Bath (United Kingdom)
🆔 https://orcid.org/0000-0002-7072-0180
✉ G.J.Room@bath.ac.uk; ⧉ https://researchportal.bath.ac.uk/en/persons/graham-room
Graham Room is Professor of European Social Policy at the University of Bath. He is author, co-author or editor of thirteen books, the most recent being *Agile Actors on Complex Terrains: Transformative Realism and Public Policy* (Routledge, 2016). He was Director of the Institute for Policy Research until December 2013. He was Founding Editor of the *Journal of European Social Policy* and is an elected member of the UK Academy of Social Sciences.