Papers on Economics & Evolution #0303, Max Planck Institute, Jena, 2003

The Use of Simulations in Developing Robust Knowledge about Causal Processes: Methodological Considerations and an Application to Industrial Evolution

Thomas Brenner[1] and Johann Peter Murmann[2]

Abstract: The paper develops a simulation method that can be used to develop knowledge inductively and test hypotheses about causal processes in social systems. Combining existing simulation methods in novel ways, it describes in detail the uses and requirements for carrying out history-friendly simulation experiments. The second part of the paper illustrates the value of the methodology by developing inductively information about the precise causal impact of (1) the initial number of organic chemists at the start of the industry (1857) and (2) the responsiveness of the German university system on the German global market share in the synthetic dye industry in 1913.

1. Introduction

Because of dramatic increases in the power and decreases in the cost of computing, mathematical simulations have become more popular in all branches of science. While scientists employ computer simulations in a variety of ways, two approaches to simulations have become dominant in recent years (Jacobsen and Bromson 1997; Troitzsch 1999). The first tries to recreate reality as closely as possible. The purpose of this kind of simulation is to make exact predictions about how the modeled system will behave under different conditions (Richards 2002). This approach is mainly used in natural sciences and engineering. Since much less is known about the causal mechanisms that drive social phenomena and since social phenomena are more complex, social scientists typically cannot construct simulation models that recreate reality and predict how a social system will behave (Meyer 2000 and Eliasson and Taymaz 2000 describe some exception). For this reason most simulation models in the social sciences have been used for a second purpose: They aim to reproduce stylized facts and examine the relationship between factors that are surmised to have been the driving forces in a social process in order to increase our understanding of such processes (Carley 1995, Levinthal 2001, Macy and Strang 2001).

The major challenge for simulations in the social sciences is that the laws according to which systems develop change over time (Tilly and Stinchcombe 1997). Simulation models that rebuild social reality to make predictions about the future, therefore, are not only rare but also highly controversial. When the aim is not to predict the future but to model history, this problem

[1] Max Planck Institute for Research into Economic Systems, Evolutionary Economics Group, Kahlaische Strasse 10, 07745 Jena, Germany, E-mail: brenner@mpiew-jena.mpg.de.
[2] Kellogg School of Management, Northwestern University, Evanston, IL 60208, United States of America, E-mail: jpm@northwestern.edu.

does not exist because, at least in principle, changes in the rules of development can be approximated through detailed historical research. In recent years, evolutionary economists have begun to test the validity of theoretical explanations offered for the development path of an economy or an individual industry by constructing so-called "history-friendly" simulation models (Malerba, Nelson, Orsenigo, and Winter 1999; Malerba and Orsenigo 2002; Nelson and Winter 1982).

There are three compelling reasons for examining causal explanations for observed historical developments with the aid of a simulation model. First, history represents only one realization of a stochastic process. If we were able to rerun history, the outcome could be completely different even if the initial conditions had been exactly the same. There are, however, some developments that are determined entirely by the initial conditions despite stochastic influences. These processes have received increasing attention over the last two decades and are now known under the general heading of path-dependency (Arthur 1994; David 1985). To develop valid theories of social processes, it is important to distinguish between developments that are caused by the initial conditions, developments that are triggered by point interventions, and developments that occur merely by chance. Simulations offer a powerful tool to develop such insights.

Second, simulation models make it possible to develop counterfactual analyses that cannot be carried out by the human brain. Most verbal explanations for social developments at least implicitly contain the notion that if a certain factor had been absent, the outcome of the social process would have been different. When multiple agents interact over longer periods of time, verbal articulations about what would had happened if certain factors in the initial conditions had been different can readily go wrong. Human brains are excellent pattern recognizers but they are very bad at making sequential logical inferences. For this reason we are often led astray when we try to think through the logical implications of causal reasoning that involves more than one variable or multiple iterations. This is precisely where simulation models that strip down a complex reality to key causal variables come in very handy because they offer a powerful tool for checking whether the logic of a verbal causal account implies plausible results (Krackhardt 2001; Sastry 2001).

Third, simulations can generate the data for studying in detail the causal relationship between variables for which only a limited number of observations are available in the historical record (Malerba et al 1999). If the concepts that constitute a causal explanation can be expressed in terms of continuous variables, the entire ranges of the variables are frequently not observed in the historical record. Simulations are very useful because they can fill in data points along the entire range of the variables, making it possible to create information about the relationship between two variables that would otherwise not be available. In this regard, simulations constitute a methodology for inducing new knowledge, provided that the simulation model replicates reality except for the variables that are varied.

The purpose of our paper is to spell out for a broad audience of social scientists how simulation models can serve as a useful tool for examining causal theories about observed social phenomena. Since causal analysis deserves to be at the center of sociological inquiry (Abbott 1998), it is very important to develop tools that help us identify the driving causes of social processes. Our aim is to expand existing simulation methods in order to be able to conduct detailed studies of causal relations in complex social systems. As long as simulation models contain only a few parameters or the values of most parameters are known exactly from empirical studies, the methods that are typically found the literature can be used. Many social processes, however, are so complex that they can be only modeled satisfactorily by specifying multiple mechanisms that imply a large number of parameters. Empirical research typically cannot provide exact values for many parameters but only ranges in which the exact values will fall with certainty. In these cases, existing simulation methodologies do not allow researchers to examine in detail causal relationships. The simulation method that we propose is specifically designed to study causal relations in social processes.

The paper unfolds in two parts. First, we will discuss in detail how simulation models can be used to carry out sophisticated studies of causal relationships. We articulate the requirements that have to be met for such an analysis to be fruitful, identify key methodological issues, and describe how the results obtained in such simulation exercises should to be evaluated. Under the name of history-friendly simulation experiments, we develop a methodology that will enable researchers to create knowledge about causal relationships that cannot be ascertained from the study of history alone. Simulation experiments, therefore, can help researcher to test causal hypotheses and carry out inductive reasoning.

In the second part of the paper, we illustrate the usefulness of this simulation methodology by modeling the development of the synthetic dye industry in the first 57 years of its existence. Murmann (forthcoming) has put forward a verbal explanation for why Germany overtook the early leaders, Britain and France, by the early 1870s and then dominated the industry for the next 40 years with a world market share reaching approximately 80%. As one key causal factor, he identified the fact that German universities produced a much larger number of organic chemists than Britain, France, Switzerland, and the United States in the early years of the industry. We test this causal argument by constructing a simulation model that captures the key features of the industry when it started and then tries to recreate key developments that occurred in the industry between 1857 and 1913. Murmann's (forthcoming) detailed empirical work on the synthetic dye industry supplies us with the necessary information to specify key factors and their relations. (The details of our simulation model are described in the Appendix.)

Following the method of history-friendly simulation experiments that we developed in the first part of the paper, our analysis involves two steps. First, we conduct counterfactual analyses to test the validity of Murmann’s causal explanation. By assigning to Germany different initial numbers of organic chemists in 1856 and by changing the ease at which the German university system responds to demand for organic chemists, we should see different global market shares of the German synthetic dye industry in the simulation year 1913 than we observe in the historical record. Then we conduct history-friendly simulation experiments to study the functional form of the relationships between the market share of German firms at the end of the simulation and two important causal variables: (1) the number of organic chemists that exist at beginning of the industry (1856) and (2) the number of organic chemists educated by the university system concurrently with the development of the industry. This allows us to develop knowledge about causal relationships that cannot be deduced from the historical record.

2. The Method of History-Friendly Simulation Experiments

Simulations are an excellent tool for conducting counterfactual analyses and for studying causal relations between variables because they make it possible to run many different scenarios and to "repeat" history (Carley and Hill 2001). Any time, stochastic processes play a role in determining a social outcome, repeating a simulation experiment with the same starting values creates a large number of observations that are not available in the historical record. The outcomes of such a simulation experiment can subsequently be analyzed with conventional statistical tools. For social developments, multiple observations of the same starting point in a causal process cannot be obtained by other methods.

Simulation as a tool for inquiry in the social sciences has not yet become a widely practiced method because it comes with its own limitations and shortcomings. As critics have pointed out, almost any result can be "confirmed" using simulations by tinkering with the model and the parameters long enough until they produce the desired result. We believe that a wider acceptance of simulation models in the social sciences requires the development of clear rules on how and under what circumstances simulations should be used. For this reason, we propose in this section not only a method that enables the modeler to make counterfactual analyses and study causal relations on the basis of simulations that rebuild history, but we also develop guidelines for how such simulation studies should be carried out to be scientifically compelling. As mentioned above, our aim is to study causal relationships in social processes that are too complex to be adequately described by models that include only a few parameters. There are good reasons to believe, however, that complexity is the rule rather than the exception in the social sciences.

Over the last five decades, a variety of different simulation methods have been deployed in the social sciences (Gilbert 1996; Jacobsen and Bromson 1997; Liebrand 1998; Lomi and Larsen 2001; March 2001). It is our conviction that each hitherto developed simulation methods practiced by themselves are not adequate to make it possible to induce robust knowledge about the relationship among variables under conditions of great complexity. By robust knowledge we mean knowledge about a causal relationship that holds for a wide range of model specifications making it likely that the real world falls within this range of models. In order to obtain such knowledge, we propose an approach that is more elaborated than those currently available in the literature. The novelty of our approach lies in the fact that we combine existing methods in a new way that allows us to deal with the problems that modelers face when the real world can only be simulated using many parameters that cannot be measured exactly with the available empirical data.

2.1 Key Features of history-friendly simulation experiments

Our proposed simulation methodology consists of four steps: First, the model of a social process and the relevant parameters are chosen with the help of empirical data representing the real state or dynamics in the world. Second, separate groups of simulations are constructed, representing different parameter values that are randomly drawn for each parameter from predetermined ranges. Third, within each group of simulations, a simulation experiment (following Winer 1971) is carried out to study a particular causal relation. Forth, the results from these simulation runs are analyzed statistically. Let us describe each of these four steps in detail.

Parameters

An important aspect of any simulation methodology is the specification of the model and the choice of parameters. When a modeler merely tries to replicate stylized facts and relations between some theorized variables, the main task is to specify the model properly. In contrast, when a modeler wants to study causal relations in historically observed processes and generate information about the specific relationship between key causal factors, choosing the parameters is also an important task in the modeling exercise. Valuable knowledge about causal relations can only be achieved if empirical data is extensively used for the estimation of parameters (for examples of this approach, see Eliasson and Taymaz 2000 or Richards 2002). This means that history-friendly simulation experiments require detailed knowledge about the empirical context. Such knowledge can be obtained either through an extensive study of the specific historical development that the modeler tries to recreate, as we do later in this paper, or by using many separate empirical studies that each focus on one of the mechanisms built into the simulations. Brenner (2001), for example, modeled the local clustering of economic activity by drawing on separate empirical sources to obtain information about innovation processes, firm start-up frequencies, and knowledge spillovers among firms.

With the help of such empirical knowledge, the model has to be specified in accordance with reality and all parameters have to be estimated. In some instances the parameters can be observed with precision. In these cases it is possible to fix the parameter to one value. In other cases the available data may not be sufficient to determine the exact value of the parameter. Under these circumstances it is important to define a range of values so that one can be sure that the true parameter value falls within the range. To demarcate this range, a minimum and maximum value for each parameter have to be specified. The larger these ranges, the more heterogeneous and therefore the less robust the simulation results will be (remember that we defined robust findings as those that hold independent of the exact specification of the simulation model). By the same token, the more detailed is the empirical information used, the more precisely the parameters can be estimated and the more robust will the knowledge about the causal relations generated through this method be.

Groups of simulations

Since generally we do not know the exact values of all parameters, we run groups of simulations based on values for each parameter that are randomly drawn from predetermined ranges. This means that each group of simulations is characterized by a common parameter set except for the parameters that are explicitly studied. If, for example, the relation between the initial number of chemists in a country and the market share of all firms in this country at the end of the simulation is to be studied, the parameter that describes this initial number of chemists is varied within a group of simulations. All other parameter settings are the same for all runs within a group of simulations.

We call the parameters that are not part of the relationship that the modeler wants to examine peripheral parameters. In the above example, the size of the global market is an example for such a peripheral parameter. As we argued earlier, it is important to set all parameters based on empirical observations. Since for many of the peripheral parameters only ranges can be empirically ascertained, it is necessary to randomly draw the value of each peripheral parameter from their predefined ranges. Each draw forms the basis for one group of simulations. Each group of simulations reflects one model specification and, therefore, one simulation world. If a large number of such groups are randomly drawn from the parameter ranges, it can be expected that one group of simulations contains peripheral parameter values that represent the true historical values. If, furthermore, the overall structure of the model is adequately chosen, this particular group of simulations will represent the true description of reality.

Note that the method we use here is similar to the Monte-Carlo approach since we run simulations across the entire range of empirically possible parameter settings. The larger the number of randomly drawn groups of peripheral parameters, the more confident the modeler can be that one of the groups of simulations represents the true description of reality.

Causal relations

In contrast to the usual Monte-Carlo approach, history-friendly simulation experiments are not interested only in the distribution of possible *characteristics* of reality, such as the global market share of German firms in the synthetic dye market in 1913. Instead, they primarily attempt to study *causal relations* in social processes, such as the dependence of the German global market share in this industry on the number of chemists educated in Germany between 1856 and 1913. For us a causal relation exists when one state or process -- in the application that we present in the second part of this paper a smaller or larger value of a parameter in 1856 -- brings about a particular characteristic of a later state or process -- in our application a lower or higher market share of German firms in 1913 (see Heise 1975, especially p. 4, for a detailed discussion of causality). Just as in a controlled experiment, we are truly dealing with causality here because in studying the impact on an outcome variable we only vary one parameter while holding all other parameters constant. To study causal relations according to our proposed method, it is necessary to specify the relevant variables and parameter settings. One defines some parameters as independent variables (e.g. the initial number of chemists in a country) and some characteristics of the simulation results as dependent variables (e.g. the market share of all firms in a country). Simulation exercises are not fruitful unless they focus on a very restricted set of variables from the entire set that plays a role in the social process. After having chosen a few focal independent and dependent variables, one carries out a simulation experiment in which the independent variables are varied in a systematic way, for example, by changing their value from the lowest value to the highest value within the respective range by equidistant steps. For each parameter value of the independent variable (or combinations of parameter values in the case of multiple independent variables), one or several simulation runs have to be carried out. The number of simulation runs has to be numerous enough so that within a group of simulations a sufficient number of data points are generated to make statistical analyses feasible. Remember that within a group of simulations all peripheral parameters are the same for all simulation runs and that each group of simulations is analyzed separately.

Statistical Analysis

The result of the simulations within one group creates a data set of independent and dependent variables similar to those that are constructed in empirical studies. These data sets can then be analyzed using standard statistical tools. Below we employ regression analyses, but other statistical methods can be used as well.

Since the statistical analysis is done for each group of simulations separately, a set of results is obtained for the relation between the independent and the dependent variables. This set of results represents all possible types of causal relations that might exist in the real world. Since, often we can only estimate the ranges in which real parameters lie, we only obtain ranges for the real causal relations. To claim otherwise would be an exaggeration of what simulations can accomplish. If, however, the same causal relation is found for all groups of simulations and if the model is an adequate representation of reality, a very strong inference can be made that the causal process does indeed operate in the real world in the manner specified. If different causal relations emerge, the method we propose generates the option to analyze which of the peripheral parameters generated the varying results.

2.2 Advantages and Drawbacks of History-friendly Simulation Experiments

Our proposed simulation methodology possesses significant advantages compared to existing approaches, but it also has drawbacks. The major drawback is that extensive data requirements have to be met to carry out such a study. First, the method requires detailed and reliable empirical knowledge of the processes that are simulated. Typically such data is not readily available and gathering the data is very time consuming. In some cases it is even impossible to obtain the necessary information about particular parameters and the parameter values have to be determined through logical inferences from other pieces of data. As a result, the ranges in which real parameter settings will have fallen with virtual certainty are necessarily very large, much larger than in the case in which one is able to observe the parameter settings. Larger ranges, in turn, imply more ambiguity in the simulation results and less reliable information about the causal processes. Second, the method requires an enormous number of simulation runs, which are very costly in terms of computer time. Our illustrative study that we present below required 24000 simulation runs that took around 60 days of computer time. (Our results would have been even more reliable, if we had been able run more simulations. The deadline for this special issue of AJS restricted us to 24,000 runs. In a future version of this paper, we will increase the number of runs to enhance further the reliability of our results). As computer speeds continue to increase, this disadvantage will slowly disappear.

The chief advantage of our proposed simulation method is that it offers the possibility of inducing knowledge about causal relations in social processes that cannot be obtained through the study of history alone. Furthermore, in contrast to the traditional simulation approaches, it generates robust findings about causal relations even if the model contains many parameters that cannot be fix exactly on the basis of empirical data. If the actual world lies within the ranges that are defined for the parameters, if the model is adequately specified, and if the same results are found independent of the peripheral parameters, the findings truly represent reality. Furthermore, the method is up-front about the possibility that different conclusions may be supported by the simulation results. Either the simulations produce robust knowledge about causal relations or they do not. If the results are not robust across the groups of simulations, the method makes it possible to study the reasons why the nature of the causal relations varies.

3. An example: The impact of chemists on the development of the dye industry

3.1 Purpose of our Simulation Experiment

Sociologists (Merton 1938; Ben-David 1971&1977) and historians (Landes 1969, Mokyr 1990) have long recognized the import role of science in the development of modern societies. The dramatic increase in gross domestic product per capita in western societies over the last three hundred years has been linked by many analysts to the proliferation of scientific thought since the 15th century. Modern societies produce more material wealth largely because of the dramatic increase in innovative activity over the last 400 years. Sociologists such as Marx (1906) and Gifillian (1935/1970) identified technological innovations as prime forces that transform societies. But many economically crucial technological innovations such as the steam engine would not have been possible without the growth in scientific knowledge about the physical universe (Mokyr 2002). For this reason the relationship between science and technology constitutes a crucial driver of social change.

In his global theory of intellectual change, Collins (1998) highlights that the rise and fall of schools of thoughts in philosophy display a structural pattern that he claims is universal in human societies across time and space. For Collins new ideas are generated by rivaling networks of intellectuals who compete for limited attention. According to Collins, schools of philosophical thought fall when attention is drained away from them in an intellectual community and redirected to other sets of ideas. Collins notes that "nodes in the attention space are crescive, emergent; starting with small advantages among the first movers, they accelerate past thresholds, cumulatively monopolizing attention at the same time that attention is drained away from alternative modes" (p. 15). Collins acknowledges that while it is possible to identify the sociological structure of intellectual communities that monopolize attention, it is not possible to predict who and what kind of ideas will become the most successful.

Although Collins's account of conceptual change in philosophy cannot be applied without modification to the development of science and engineering because these disciplines are driven more by utilitarian considerations (Schmookler 1965; Rosenberg 1998), Collins's focus on the competition between communities of scholars provides, in our view, a proper starting point for conceptualizing the development of different disciplines within national systems of higher education. Science and engineering disciplines compete for the limited resources that society is willing to spend on research. The ability of individual disciplines to gain not only attention but also material resources depends a great deal on how well they are supported by people who are not scientists themselves. According to Murmann (forthcoming), Collins's view that the success of individual schools of thought depends on early mover advantages also applies to academic disciplines in particular countries: Based on the study of the organic chemistry and the synthetic dye industry between 1857 and 1913, Murmann suggests that the German organic chemistry and the German dye industry stimulated and supported each other's development and allowed Germany to become the dominant global player.

To examine parts of the causal claims in the co-evolutionary explanation for German dominance in the synthetic dye industry before World War I, we carry out a simulation experiment of the development of the industry in the five major producer countries: Britain, France, Germany, Switzerland, and the United States. In particular, we want to examine the influence of differences in scientific capabilities in the early years of the industry on the competitive position of national firms over the next five decades. Murmann's empirical analysis suggests that one of the key reasons why German firms overtook their British and French competitors was that Germany had a larger number of chemists available during the early years of the industry who could become founders of dye firms. Later the German university system was more responsive to the need for additional chemists who could staff the R&D labs of the large dye firms. Because the chief purpose of this paper is spell out the methodology of history friendly simulation experiments, we do not have space to examine all the different causal factors that influenced in Murmann's account the German market share in the synthetic dye industry in 1913. Patents, for example, are believed to have played a key role and for this reason are explicitly modeled in our simulation. In light of our limited space here, we cannot provide a full account of the relevant causal factors and have to restrict our illustration of the history friendly simulation experiment method to two important causal variables. We focus our analysis on Murmann's causal explanations associated with two variables that characterize the development of German human manpower in organic chemistry.

Based on Murmann's historical analysis, it is possible to formulate the following two historically specific hypotheses:

> **H1**: A larger (smaller) number of German organic chemists in the first year of the synthetic dye industry (1857) would have caused a larger (smaller) global market share of German firms in this industry in 1913.
>
> **H2**: A higher (lower) responsiveness to the demand for chemists in the German university system would have caused a larger (smaller) global market share of German firms in this industry in 1913.

We want to test these hypotheses first in terms of a counterfactual analysis (Section 3.3). In a second step (Sections 3.4 and 3.5), we will go beyond this approach by analysing in detail the functional form of the dependence among the variables. Besides examining the impact of the number of chemists on the long-term competitive position of German firms, we also want to study whether the influence of a higher number of chemists on the market share of German firms is causally mediated through the number of firm start-ups in the country. Murmann argued that the larger number of start-up firms in Germany was an important determinant of the long-term collective market share of German firms. But he identified two important causes for the larger number of start-up firms in Germany: the larger number of chemists in Germany and the absence of effective patent protection in Germany before 1877. Anytime two causal factors are hypothesized to have the same empirical consequences, the historical record does not allow us to discriminate with accuracy which cause was operating at what strength to bring about the historically observed outcome. The limitation of Murmann's historical evidence is that one cannot tell whether the larger number of firm start-ups was caused by the larger number of chemists or the absence of patents in Germany because we only have one data point for each of the causal variables. Simulations can overcome this data limitation because they can model the causal arguments and then generate the data that allows the analyst to ascertain the relative importance of each cause. For this reason in our simulation experiments we also want to examine the relationship between the number of chemists available in Germany and the number of start-ups that occurred in Germany during the whole simulation run (1857-1913) as well as the number of firms that existed in Germany in 1913.

3.2 Method

To study the impact of human capital creation on the development of the synthetic dye industry, we have developed a simulation model of the industry. Based on the detailed historical analysis provided in Murmann (forthcoming), we determined that our model of the industry needed to capture how firms, innovations, products, patents and chemists interacted in the development of the industry. In particular we wanted to examine Murmann's explanation for why German firms came to dominate the industry for 40 years before World War I. The details of the model are given in the Appendix.

To analyze this model we use the method of history-friendly simulation experiments developed above. Our model contains 89 parameters that determine the initial conditions and the dynamics in the model. Precise historical information made it possible to determine a specific value for 49 of these parameters. For the remaining 40 parameters we only had sufficient historical information to define ranges into which the actual value of each parameter will have fallen (see the Appendix for a detailed description of the historical information used in specifying all parameters).

Two of the 89 parameters are the focus of our present study. One concerns the initial number of chemists in Germany in 1856, denoted by $c_{init,G}$. The other concerns the responsiveness to the demand for chemists in the German university system, denoted by e_G. We study the influence of these two parameters separately, but our proposed method also makes it possible to examine them together if the researcher wants to key in on the interaction of focal parameters. In the present case this means that we need to run two separate history-friendly simulation experiments.

Historical records allow us to estimate the actual number for $c_{init,G}$. It is given by $c_{init,G}$=54 (see the appendix). To study the impact of this initial value, we vary the initial chemists parameter between 5 and 105, using 6 different values: $c_{init,G}$=5, 25, 45, 65, 85, 105. For each group of simulations, all other parameters are chosen randomly from their defined ranges. We carry out 100 groups of simulations (more groups of simulations would have increased the power of our results but were not feasible given the deadline for this version of the paper). In each group of simulations 120 simulation runs are done, 20 runs for each value of $c_{init,G}$. Although all peripheral parameters are the same for the 20 runs done for each value of $c_{init,G}$, the simulation outcomes may differ because several stochastic processes such as product and process innovations, firm start-ups and strategic decisions are included in the model. We run 120 simulations within each group of simulations to obtain sufficient data for a statistical analysis. This means that altogether 12,000 simulations are run in this particular experiment.

Specifically we study the impact of $c_{init,G}$ on the following properties of the synthetic dye industry in 1913: the global market share of German firms, denoted by m_G, the number of start-ups that occurred in Germany during the whole simulation run (1857-1913), denoted by f_G, and the number of firms that existed in Germany in 1913, denoted by n_G. These values are recorded for each simulation run. Through this we obtain 120 strings of values ($c_{init,G}$, m_G, f_G, n_G) for each group of simulations.

There are many different options to analyze the results from each group of simulations. Our methodology requires that we classify the results but how such classes of results are defined depends on the particular research question that is addressed. Robust knowledge is generated if the results for all different groups of simulations fall into some but not all classes of results. In our application of the methodology the data sets for each group of simulations are analyzed with the help of regressions. Each groups of simulations is analyzed separately. $c_{init,G}$ is used as the independent variable, while the values characterizing the conditions at the end of the simulation are used as dependent variables. Our main goal is to ascertain whether $c_{init,G}$ has a positive or negative impact on our focal dependent variables. Besides determining whether the relationship is positive or negative over the entire range of the independent variables, we also want to ascertain whether their impact decreases or increases over the range of $c_{init,G}$. A regression of the form

$$m_G = \beta_C + \beta_L \cdot \left(1 + \tilde{\beta}_Q \cdot c_{init,G}\right) \cdot c_{init,G} = \beta_C + \beta_L \cdot c_{init,G} + \beta_Q \cdot c_{init,G}^2 \qquad (1)$$

is used where β_C, β_L and β_Q are regression parameters. The impact of the independent variable, $c_{init,G}$, on the dependent variable, m_G, is positive if β_L is significantly positive and negative if β_L is significantly negative. It increases if the regression parameter β_Q is greater than zero and decreases if β_Q<0 holds. This regression is repeated for all other dependent variables (i.e. m_G is replaced successively by each of the other dependent variables).

These regression analyses then provide us with information about whether the effect of the parameter $c_{init,G}$, which describes the initial number of German organic chemists, on the outcome characteristics at the end of the simulation, m_G, f_G, or n_G, is (1) positive, negative, or not significant and (2) increasing in $c_{init,G}$, decreasing in $c_{init,G}$, or not significantly dependent on $c_{init,G}$. This allows us to distinguish between 7 different kinds of effects: a positive increasing, a positive decreasing, a positive constant, a negative increasing, a negative decreasing, a negative constant, and no significant impact. We selected this classification scheme because it allows us to answer our particular research questions, but other classifications of the results are clearly possible.

For each group of simulations we obtain a profile of effects of our independent variable on our dependent variable. As argued above, only if all groups of simulations display the same profile of effects, can we conclude that this effect is a true description of reality. If different results are obtained for the various groups of simulations, we can interpret the difference and study which of the other parameters determine this difference. To explore differences in such profiles of effects, we also carry out additional analyses that are described in detail as we present them below. We should also point out that the particular techniques that we employ to analyze the data generated by our simulation experiments is clearly not the only possible one. There are other statistical methods that can analyze data generated by history-friendly simulation experiments to induce knowledge about social processes and relations.

The analyses we just described are carried out in exactly the same way for our second independent variable, e_G. The parameter e_G determines how many new students begin to study organic chemistry in the German university system. The number of new students is calculated in the simulation so that in three years time (after they finished their study) e_G times the number of the currently employed organic chemists will be available on the job market. This means that e_G reflects the responsiveness of the university system towards current demand for organic chemists. It also determines the maximum rate at which the number of organic chemists in German industry can grow within three years without important labor from other countries. Historical information indicates that this parameter lies between 1.0 and 1.2 (for a more detailed discussion consult the Appendix). To examine the influence of different levels of this parameter on the simulation outcome, we run simulations with six values: e_G=1.0, 1.04, 1.08, 1.12, 1.16, 1.2. We carry out again 100 groups of simulations. As described above, for each group all peripheral parameters are held constant and only our focal parameter is varied. Again we conduct 20 simulation runs for each value within each group of simulations. This means that each group of simulations again comprises 120 simulation runs. Altogether, then, this experiment involves again 12000 simulation runs for this focal parameter. Both experiments together require 24000 simulation runs. We present and discuss all results in the next sections.

3.3 German university education and the dominance of German synthetic dye firms: A basic counterfactual analysis

Murmann (forthcoming) argued that German firms dominated the dye industry for 40 years before World War I because more organic chemists were educated in Germany than in other countries during the early years of the dye industry. With the help of our simulation model we are able to test this argument by carrying out a counterfactual analysis. To do so we remove the assumed causes for the German dominance in the simulations by changing the relevant parameter values and then examine whether the German dominance disappears as predicted.

We identified above two parameters that determine the number of organic chemists available in Germany: the initial number of chemists in Germany, $c_{init,G}$, and the responsiveness of German universities to the demand for chemists, e_G. We estimated the real value of the former to be $c_{init,G}$=54 (see the Appendix). It is not possible to estimate the latter parameter precisely based on the available historical records. We have been able, however, to determine the possible range of e_G as$1.0 \leq e_G \leq 1.2$ (see the Appendix).

We run 2000 simulations to reflect what we regard as the best estimate for the true historical parameter settings: $c_{init,G}$=54 and e_G=1.2. On average, the German firms obtain a market share of 30.7% at the end of these simulation runs. This figure is clearly below the historical outcome of around 80%. This difference does not suggest that our model is not properly specified. In these simulation runs the market share of German firms varies between 0 and 96.5% at the end of the simulations. This means that the historically observed outcome lies clearly within the range of possible outcomes for the simulations. There are reasons why the average simulation result differs from the historical outcome. First, since we lack exact information about many of the parameters, they are fixed randomly for each simulation within their predefined ranges. As a result, only a limited number of the 2000 simulation runs may reflect closely the actual historical parameter settings. Second, the development of the synthetic dye industry, in reality and in the simulations, depends crucially on random events. It is possible, therefore, that the German market share in 1913 is influenced considerable by chance.

Even though the average market share of German firms is considerably lower in the simulations, this does not invalidate the hypothesis that higher initial numbers of German chemists, $c_{init,G}$=54, and a higher responsiveness of the German university system, e_G=1.2, are responsible for the higher market share of German firms. To test the validity of this argument we need to carry out counterfactual simulations that set lower parameter values for Germany than the real ones and make them more similar to those of the other countries. If the two education parameters are truly responsible for the dominance of German firms, the market share of these firms should decrease significantly if $c_{init,G}$ or e_G is decreased. To study the impact of each of these parameters, we carry out two sets of counterfactual analyses. In each, one parameter is reduced while the other is held constant. For each value of the varied parameter, 2000 simulations are run. The results are presented in Tables 1 and 3.

Table 1: Results of the counterfactual analysis for the parameter $c_{init,G}$.

$c_{init,G}$	Market share of German firms at the end of the simulation			
	average	standard deviation	minimum	maximum
5	22.5	12.3	0	78.3
25	25.8	12.7	0	94.1
45	27.2	13.8	0	94.6
65	27.3	13.7	0	97.5
85	27.7	13.6	0	83.0
105	27.8	14.4	0	93.7

Table 1 shows that the market share of German firms decreases if $c_{init,G}$ is reduced. This holds especially for decreases of $c_{init,G}$ below its historical value of 54. Above this level further increases in the relative number of chemists do not matter much for the development of the synthetic dye industry. Below this level they have a significant impact, at least on average. These conclusions are based on the average findings for all 2000 simulation runs. In accordance with the methodology that we articulated in the first half of the paper, we cannot conclude that these findings represent reality unless we check whether they depend on the settings of the peripheral parameters. Since we run 100 groups simulations that differ in the values of the peripheral parameters, we can study separately for each group of simulations whether the market share of German firms depends on the initial number of organic chemists. The functional form of this relationship is examined in detail in the next section. For the purposes of a basic counterfactual analysis, it suffices to compare the results for the parameter settings $c_{init,G}$=5 and $c_{init,G}$=25 with the results of the setting $c_{init,G}$=45, which is relatively close to the historical value.

To study whether Hypothesis H1 is a true description of reality, each group of simulations has to be tested separately. 100 groups of simulations are run. The results obtained for the variation of $c_{init,G}$ for each group of simulations are reported in Table 2.

Table 2: Number of cases in which the difference in the German marke share ,m_G, is of a certain kind (significance level of 0.05).

$c_{ini,G}$	difference in m_G between the conditions with $c_{ini,G}$ =45 and $c_{ini,G}$ =5 or 25			
	significantly positive	positive (not significant)	negative (not significant)	significantly negative
5	51	16	9	24
25	42	18	15	25

Table 2 does not confirm Hypothesis H1. Although for 51 (42) groups of simulations the market share of German firms increases significantly if $c_{init,G}$ is increased from 5 (25) to 45, there are also many simulation runs in which the opposite relationship is found. Hence the relationship between the initial number of German chemists and the final market share of German firms seems to depend crucially on the peripheral parameters. The 51 cases of a significantly positive impact of an increase of $c_{init,G}$ from 5 to 45 on the market share of German firms are characterized by a low responsiveness of the German university system (correlation value: -0.41) and a low innovation rate of chemists (correlation value: -0.24). e_G is by far the most important peripheral parameter in this context. Only if the responsiveness of the German university system, e_G, is low, does the initial number of German chemists have a significant positive impact on the market share of German firms. Historical analyses (Haber 1958, McClelland 1980, Homburg 1993, Homburg, Travis, Schröter 1998) show that the responsiveness of the German university system was quite high. Of those 24 groups of simulations that are characterized by e_G>1.15 only 30% show a significantly positive relationship between the initial number of German chemists and the final market share of German firms, while 50% show a significantly negative relationship. Hence, the data generated from our simulations runs does not confirm Hypothesis H1. It provides evidence that at the expense of other causal factors Murmann (forthcoming) placed too much emphasis on the initial number of chemists in determining the long-term market share of German firms.

Table 3: Results of the counterfactual analysis for the parameter e_G.

e_G	market share, m_G, of German firms at the end of the simulation			
	average	standard deviation	minimum	maximum
1.00	13.0	7.8	0	62.0
1.04	23.9	12.5	0	82.8
1.08	28.0	14.6	0	92.2
1.12	29.4	15.4	0	89.6
1.16	30.5	16.3	0	90.8
1.20	30.7	17.0	0	96.5

Let us now examine the impact of our second independent variable, e_G. Table 3 clearly shows that the market share of German firms decreases if e_G is reduced. All differences in the average values are significant at the 0.01 level except the difference between the results for e_G=1.16 and the results for e_G=1.2. These results support Murmann's (forthcoming) causal argument that the responsiveness of the national university system had a significant impact on the development of national firms. But the methodology that we use here also allows us to go beyond Murmann's historical analyses and gain additional insights. The simulation results suggest that the impact is strongest for small values of e_G. Once the national university system displays a reasonable responsiveness to the demand for educated chemists by firms, further increases of this responsiveness do not matter much.

These conclusions are based on the average result of 2000 simulations that were run for each of the six parameters settings of e_G. As we have already argued above, we cannot conclude that the findings represent reality unless we check whether they depend on the setting of the peripheral parameters. Again, we can study whether the market share of German firms depends on the responsiveness of the German university system for each group of simulations separately. The functional form of this relationship is studied in detail in section 3.5. For the purposes of the basic counterfactual analysis, it suffices to compare the results for the parameter settings, e_G=1.04, e_G=1.12 and e_G=1.16, to the results for the setting e_G=1.2 and determine whether the above findings hold independent of the peripheral parameters. The results of the comparison are presented in Table 4.

Table 4: Number of cases in which the difference in m_G is of a certain kind (significance level of 0.05).

e_G	difference in m_G between the conditions with e_G=1.2 and e_G=1.04, 1.12 or 1.16			
	significantly positive	positive (not significant)	negative (not significant)	significantly negative
1.04	66	17	13	4
1.12	32	30	24	14
1.16	19	36	30	15

Table 4 appears to confirm the average finding: the market share of German firms reacts significantly to large decreases of e_G while it reacts little and not in a robust manner to small decreases. Let us first consider a decrease in the responsiveness of the education system in Germany to a value of 1.04. This causes a significant decrease of the dominance of German firms in 66% of the groups of simulations. In another 17% there is a decrease, although for 13 of these 17 % the decrease is not significant. But since there are exceptions to this pattern, Hypothesis H2 is not confirmed to be truly independent of all the settings for all peripheral parameters. The history-friendly simulation experiment methodology makes it possible to determine the reasons for these exceptions and examine whether there are certain peripheral parameters that systematically counterbalance the influence of the responsiveness of the German education system.

To answer this question we analyze which peripheral parameters are correlated with a significant decrease and which are correlated with an increase in the German market share. The 66 groups of simulations for which significantly larger market shares are found for e_G=1.2 compared to e_G=1.04 are characterized by small innovation rates for chemists (correlation value: -0.31), wider patent scope in Germany (correlation value: 0.30), wider patent scope in Great Britain (correlation value: 0.21), and larger demand in Great Britain at the beginning of the simulation (correlation value: 0.21). The 17 groups of simulations for which the market share of German firms decreases if e_G is decreased from 1.2 to 1.04 are characterized by a narrow patent scope in Great Britain (correlation value: -0.28) and smaller demand in Great Britain at the beginning (correlation value: -0.25).

Two conclusions can be drawn from this analysis. First, the impact of the responsiveness of the German university system on the development of the synthetic dye industry is significantly lowered if the conditions are very favorable for the British industry. This suggests that British firms are the main competitors of German firms in the simulations. If patents only cover a small range of the product space and the demand in Britain is high, many firms are founded there. Under these conditions British firms appear to become so strong that they severely restrict the growth of the market share of German firms. As a result, the responsiveness of the German university system becomes irrelevant because there are not enough German firms around to take advantage of the German system.

Second, a large impact of chemists on the innovation rate of firms also seems to destroy the impact of the responsiveness of the German university system. Although this finding initially appears counter-intuitive, a closer look at the data generated by the simulations makes the finding intelligible. A higher innovation rate of chemists (see the Appendix for a description of how chemists influence the innovation rate) increases the variance in the simulation outcomes, including the market share of German firms (correlation value: 0.15). A high innovation rate of chemists favors those firms that decide by chance to employ more chemists. This decision depends on the aspiration level of the firms, which is randomly drawn for each firm when it is founded (see the Appendix for a detailed description). Hence, whether many chemists are employed is, to a large extent, randomly determined. Therefore, a high importance of chemists for the success of firms makes this success random in our simulations and advantageous circumstances in Germany less important. A wide patent scope in Germany seems to have the opposite effect. Since patent laws were only introduced in 1877, they seem to stabilize early developments and therefore increase the impact of the German university system if the patents cover a wide range in the product space.

The differences in the market shares of German firms between the situations characterized by e_G=1.12, 1.16 and 1.2 are not sufficiently significant to warrant a detailed discussion. Once the responsiveness of the German university system reaches a certain value, further increases in this responsiveness do not matter. We therefore only found support for Hypothesis H2 for values of e_G between 1.0 and 1.12. The more detailed analysis of the functional form of this relationship, which is described below (see Section 3.5), confirms this finding.

3.4 A history-friendly simulation experiment with the initial number of German chemists

The history-friendly simulation experiment methodology allows us to study in more detail the impact of the initial number of German chemists. The analysis above has not confirmed Hypothesis H1. The results have shown that the impact of the initial number of chemists depends crucially on the values of the peripheral parameters that represent different possible worlds.

To obtain a clearer picture of the relationship between the initial number of German chemists and the final (1913) market share of German firms, we carried out a regression analysis for each group of simulations as described in Section 3.2. We use Equation (1) without the last term on its right-hand side to determine whether the relationship under investigation is positive (β_L>0) or negative (β_L<0). The whole regression equation (1) is used to determine whether such a positive or negative relationship is increasing with $c_{init,G}$ (β_Q>0) or decreasing with $c_{init,G}$(β_Q<0). The market share of German firms at the end of the simulation, m_G, the number of German start-ups, f_G, and the number of German firms at the end of the simulation, n_G, are studied as dependent variables. The results are presented in Table 5.

Table 5: Frequency of relationships between variables

	type of regression result for $c_{init,G}$						
dependent variable	*positive increasing*	*positive decreasing*	*positive constant*	*negative increasing*	*negative decreasing*	*negative constant*	*no sign. impact*
m_G	1	30	15	3	3	15	33
f_G	5	38	57	-	-	-	-
n_G	2	21	31	1	-	6	39

The results for the dependent variable m_G confirm the findings above. The relationship between the initial number of German chemists, $c_{init,G}$, and the final market share of German firms, m_G, depends crucially on the peripheral parameters. All kinds of relationships are found for some groups of simulations. In 33% of the groups of simulations no significant relationship is found. 46% of the groups of simulations show a positive relationship, while 21% of the groups of simulations show a negative relationship. Hence, the impact of the initial number of German chemists on the market share of German firms depends crucially on the circumstances. This makes it impossible to identify the actual relationship with the help of the simulation experiment without more historical information about the precise values of the peripheral parameters. We are not able to confirm or reject Hypothesis H1 on the basis of the analysis conducted here.

By contrast, a clear picture emerges for the relationship between the initial number of German chemists, $c_{init,G}$, and the number of German start-ups, f_G. This relationship is always positive, independent of the values of the peripheral parameters. Hence, we can conclude that a higher number of chemists at the beginning of the simulation *causes* a higher number of start-ups. This is partly a consequence of how we formulated the simulation model. We assume in the model that each chemist who is *not* employed in a dye firm has a certain probability of founding a firm. Since many of the firms are founded at the beginning of the simulation, the initial number of chemists has to have an impact on the number of start-ups.

But let us now examine the hypothetical causal chain that more initial chemists lead to more start-ups which, in turn, leads to more firms in 1913 that collectively have a larger market share. Table 5 shows that in many cases a higher number of chemists at the beginning of the simulation also leads to a higher number of firms at the end of the simulation. But this causal relationship depends clearly on the values of the peripheral parameters. Only in 54% of the groups of simulations does a higher initial number of German chemists lead to more German firms in 1913. In many of the other groups of simulations the relationship is not significant. The impact of the initial number of German chemists on the final market share of German firms is even less robust. Since only the first step in the causal chain above received confirmation, we carried out two additional linear regressions for each group of simulation: One with the final number of firms as dependent and the number of start-ups as independent variable and one with the final market share of German firms as dependent and the final number of German firms as independent variable. The results are given in Table 6.

Table 6: Frequency of relationships between the variables f_G, n_G and m_G

regression function	$r > 0$	$r < 0$	r not significant
$n_G = r f_G + c$	67	3	30
$m_G = r n_G + c$	49	9	42

Table 6 reveals that a higher number of start-ups does not always lead to a higher number of firms at the end of the simulation. We have even more evidence that a higher final number of German firms does not automatically imply a higher collective market share for these firms. Only the first causal relation of our hypothesized causal chain holds independent of the circumstances: More initial chemists lead to more start-ups. The other two causal relations, however, are only present for some settings of the peripheral parameters. Other mechanisms seem to interfere so that the entire causal chain holds only for a restricted number of circumstances. Whether these circumstances were realized in the actual development of the synthetic dye industry cannot be answered on the basis of the historical data that is available for the industry before 1914.

3.5 A history-friendly simulation experiment with the responsiveness of the national university system variable

The history-friendly simulation experiment methodology also allows us to study in more detail the impact of the responsiveness of the German university system on the market share of German firms and to gain knowledge about this relationship through inductive reasoning. Two questions are addressed in this context: what is the functional dependence of the market share of German firms, m_G, on the responsiveness of the German university system, e_G, and to what extent is the number of start-up firms responsible for this relationship.

To study the functional form of the relationship between the market share of German firms and the responsiveness of the German university system, we carry out a regression analysis for each group of simulations. A linear and a quadratic function, given by Equation (1), are employed in this analysis. First, a linear function is used to examine whether the relationship is positive or negative. Then, a quadratic term is added to examine whether the relationship is increasing or decreasing. Table 7 reports how often each kind of relationship is obtained within the 100 groups of simulations (significance level of 0.05).

Table 7: Frequency of relationships between variables

	type of regression result for e_G:						
dependent variable	*positive increasing*	*positive decreasing*	*positive constant*	*negative increasing*	*negative decreasing*	*negative constant*	*no sign. impact*
m_G	2	74	16	-	-	-	8
f_G	33	7	59	-	-	-	1
n_G	9	32	51	-	-	-	8

Table 7 shows clearly the positive relationship between the responsiveness of the German university system, e_G, and the market share of German firms, m_G. A negative relationship is not found in any of the 100 groups of simulations. This means that across all values of all peripheral parameters, the responsiveness of the German university system never has a negative impact on the market share of German firms. For 8 groups of simulations the results are not significant. By doing some additional simulations, we discerned that a significant positive relationship emerges as soon as we increase the number of simulation runs in these cases. This means that for some peripheral parameter settings, the relationship between e_G and m_G is so weak that more than 120 simulation runs are needed to detect it. These data clearly confirm Hypothesis H2.

This particular analysis, however, considers the whole range of e_G from 1 to 1.2. We already found in our basic counterfactual analysis that the positive relationship is mainly caused by a dependence that exists for smaller values of e_G. This conclusion is also confirmed here since in 74% of the groups of simulations the regression results indicate a positive decreasing relationship. Only in 2% of the cases does a positive increasing relationship exist. In summary, the impact of the German university system's responsiveness to the demand for researchers on the development of the German synthetic dye industry seems to be high as long as the responsiveness has not reached a certain level and decreases thereafter.

But what about the question of whether the impact of the responsiveness of the university system on the market share of national firms is mediated by a higher number of firms. Table 7 shows that, indeed, the responsiveness of the German university system has a positive impact on the number of start-ups that occur in Germany and the number of firms that exist in Germany in 1913. This dependence, however, does not display the same functional form as the relationship between e_G and m_G. In particular in the case of the number of start-ups that occur between 1857 and 1913, the positive relationship increases with the value of e_G. This suggests that there might be other mechanisms involved, although the responsiveness of the university system increases the number of firms.

To examine this in more detail, we try to explain the market shares of German firms by e_G and either the number of start-ups, f_G, or the number of firms in Germany in 1913, n_G, simultaneously. This means that we conduct regressions according to

$$m_G = \beta_C + \beta_E \cdot e_G + \beta_F \cdot f_G \tag{2}$$

and

$$m_G = \beta_C + \beta_E \cdot e_G + \beta_F \cdot n_G \,. \tag{3}$$

If the only effect of the responsiveness of the German university system is a higher number of firms, which then increases the market share of all German firms, the last term in these regression functions should explain much of the variance in the market share, m_G. If, instead, other mechanisms cause the positive relationship between e_G and m_G, the regression parameter, β_E, should be significantly positive while β_F might be insignificant. Again we carry out all regressions for each group of simulations separately. The results are given in Table 8.

Table 8: Frequency of results for the regression parameters, β_E and β_F, (the first value gives the results for Equation (2), the second value for Equation (3)).

		regression parameter β_F			**sum**
		positive significant	not significant	negative significant	
regression parameter β_E	positive significant	3 / 24	47 / 53	22 / 8	72 / 85
	not significant	7 / 9	16 / 5	0 / 0	23 / 14
	negative significant	5 / 1	0 / 0	0 / 0	5 / 1
sum		15 / 34	63 / 58	22 / 8	100

Table 8 shows that the responsiveness of the university system explains the market share, m_G, much better than the number of firms. Only in a few cases does the positive impact seem to result solely from a higher number of firms alone (those cases in which β_F is significantly positive and β_E is either not significant or significantly negative). There are 12 such cases for the number of start-ups and 10 such cases for the number of firms in Germany in 1913. There seem to be, therefore, a few peripheral parameter settings for which the only effect of the university system is a higher number of firms. For most settings of the peripheral parameters (72% and 85%, respectively), however, the responsiveness of the university system to the demand for chemists influences the development of the national dye industry in more ways than just increasing the number of firms.

4. Conclusions

Mathematical simulations represent an excellent tool for generating new knowledge about the social world that cannot be obtained in other ways. To overcome some of the valid criticisms about the ability of simulation methodologies to explain causal processes in the real world, we have proposed, under the label of "history-friendly simulation experiments," a stringent set of requirements for using simulations to generate knowledge about a causal relationship. Aside from determining the relevant variables for the model with the help of detailed historical knowledge about the social process, we highlighted that it is necessary to specify all parameters as precisely as possible based on historical data. Since for some parameters in the model only ranges can be defined into which the actual historical value of the parameter will have fallen with certainty, it is necessary to analyze different possible worlds that can emerge from the specified parameter ranges. Separate groups of simulations that represent the different possible worlds need to be constructed by drawing randomly for each parameter a value from the predetermined ranges. Within each group, a simulation experiment is then carried out that systemically varies the value of the causal variable and records the result for the relevant outcome variable. The data generated from these groups of simulations can subsequently be used to analyze, with statistical tools, the functional relationship between the causal variable and the outcome variable. If the causal relationship displays the same functional form across all possible worlds, we can induce with great confidence that the hypothesized causal relationship indeed represents a true description of the real world.

One of the key benefits of these history-friendly simulation experiments is precisely that they can be used to develop inductively information about the functional relationship among key variables. Our simulation of the development of the synthetic dye industry has confirmed Murmann's (forthcoming) proposition that the competitive success of national firms was strongly affected by the creation of human capital on the part of national universities. But our history-friendly simulation experiments also allowed us to induce much more specific knowledge about the nature of the causal relationship between the national university systems and the market share of German firms in 1913. Here we went substantially beyond the insights that Murmann's historical analysis could provide. We found that the impact of the initial number of German organic chemists depends crucially on the responsiveness of the German university system. Only for a low responsiveness of the German university system does the initial number matter for the long-term market share of German firms. This seems to suggest that as soon as a sufficient number of firms were active in the industry, the responsiveness of the German education system allowed existing firms to hire enough chemists and thereby secure a large global market share for the German industry in 1913. Furthermore, there is a threshold with respect to the impact of the responsiveness of the German university system. Once this threshold is reached, further increases in the responsiveness did not matter. This result seems to suggest that there are diminishing returns from educating more human capital than are presently needed in industry. It also provides evidence that a country can make up for a slow start in a science-based industry by having a university system that can respond quite rapidly to the changes in industrial demand. Although a higher level of responsiveness of the German university system caused a higher number of start-ups, we were also able to ascertain that the increase in the number of start-ups was not the main mechanisms by which the responsiveness of the German university system influenced the global market share of German firms in 1913.

In summary, by using a history-friendly simulation experiment, we were able to determine much more precisely than is possible through a standard historical analysis the causal pathways that created the German dominance in the synthetic dye industry before World War I. Nevertheless, it seems to us that the real bottleneck in the endeavor to develop robust knowledge about causal processes in social developments will not be the simulation technology. We already mentioned that cost of computing is decreasing each year while speed is increasing. Ever more user-friendly software packages are also becoming available (e.g. LSD and Stella; see Kwasnicki 1999 for a description of different platforms) that dramatically reduce the start-up costs of simulating causal processes. The real bottleneck in creating robust knowledge is the availability of good historical analyses that provide well-grounded conjectures about causal mechanisms driving a social process and give detailed information about the relevant parameters. Without the historical data and analysis that Murmann (forthcoming) put together through his multi-year study of the synthetic dye industry, we could not have carried out our history-friendly simulation experiments to gain inductively additional insights that cannot be logically inferred from the historical record. Hence it is only fitting to end this paper with a call for more careful historical research on the causal mechanisms that drive social change. Without such historical research, it is impossible to do history-friendly simulation experiments.

Appendix: Description of the Simulation Model

To capture the patterns of action observed in the historical record, we followed the Cyert, March and Simon tradition (March and Simon 1958; Cyert and March 1963/1992) in our basic behavioral assumptions about agents. The two kinds of agents that we model explicitly -- firms and chemists -- are assumed to display relatively simple rules of action: they search locally first and they exhibit considerable inertia in their choices. As a result, the agents in our model display a national bias in their decision-making. This renders countries meaningful units of analysis for analyzing industrial development.

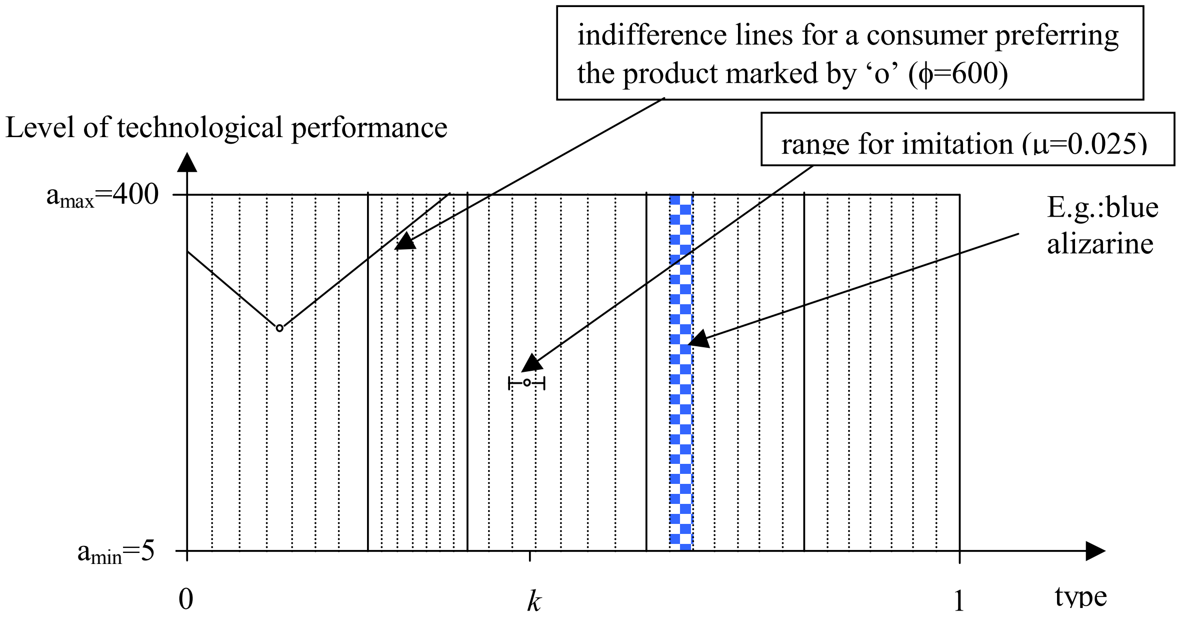
We assume the existence of six independent regions: United Kingdom (UK), Germany (G), France (F), Switzerland (SW), United States (USA), and the rest of the world. Every firm, chemist, and patent is assigned to one region.

We model the development of the dye industry from the year 1857 until the year 1913. The time period used in this simulation is the calendar month. This means that the simulations run 684 (12*57) periods. The behavior of each agent is updated every period. In addition, the demand function is calculated each period and patent frequencies for each country are updated. After presenting our approach to modeling the product space of the industry, we describe separately in detail below the updating procedure for the relevant parameters. The specification of the product space has ramifications for the modeling of many other processes and the setting of parameters. Therefore it deserves to be discussed first.

a. Product space

The synthetic dye products that firms offer in the market are defined in a two-dimensional space. This means that each product is characterized by two values. The first value, k, defines the type of the product. Dyes differ with respect to many characteristics. The most obvious is their color. Hence, not all dyes compete on the market with each other. We assume that each dye is of a certain type, k, where $0<k<1$ is assumed without restricting generality. Demand is modeled such that only dyes with similar values of k compete with each other. In our conception of the product space, the type of a product (k) summarizes in one dimension several aspects: The underlying chemical technology, color, textiles for which it can be used, durability, and other performance characteristics. The historical development of different kinds of synthetic dyes can be translated into our conception of the product space as follows: Historically there are around 5 product classes that differ in their underlying chemical technology (aniline, alizarin, azo dyes, sulfur dyes, and synthetic indigo). Our product space would be partitioned into 5 parts as it is done in Figure 1. The parts could be of different size reflecting the different market of the respective type of product. Furthermore, assuming that seven classes of color (red, orange, yellow, green, blue, violet, black) exist each part in turn would be further partitioned into 7 parts. Within these parts the synthetic dyes still differ in their durability, the textiles for which they can be used, and similar characteristics. As this discussion illustrates, using more than one dimension to characterize the type of dyes would be preferable. But for simplicity we collapse all these different dye characteristics into one dimension, k, which is a continuous measure ranging from 0 to 1.

Figure 1: Product space assuming five technological classes of synthetic dyes and seven colors.



Each product is also characterized by a level of performance, *a*. We assume that each process innovation of firms leads either to an improvement in the quality of a product or to a decrease of the cost of production. This is reflected in the model by the value *a*. The higher is *a*, the more competitive is the respective product in the market. This means that each innovation leads to an upward movement in the product space. The value of *a* is assumed to range between a_{min} and a_{max}. a_{min} determines the demand for the synthetic dye product of the first firm on the market, while it has little influence on the developments later on. For firms to be able to start their business, some demand is necessary. Therefore, a_{min}=5 is assumed. The first product that appears on the market is assumed to be characterized by a=5. a_{max} determines the greatest technological advance that is theoretically possible for a particular dye product. Whether this maximum technological potential is reached for a particular dye product depends on the innovations that occur during the simulation. For the results of the simulations the value a_{max} by itself is not important. What matters is the relation between the speed of innovations and the value of a_{max}. Thus, a_{max} has the characteristics of a normalizing factor and is set to a_{max}=400.

b. Firms

In each time period, firms are characterized by the products they offer in the market, the number of chemists they employ, the strategy they use, and their sales. They are located in one region and remain there forever.

Entry of firms

Start-ups occur in each country with a probability that is proportional to the number of chemists that are not employed in dye firms and the number of consumers in the respective national market. The product of these factors is multiplied by a factor that determines the overall start-up rate. Initial simulations show that a value of that factor of about 0.000005 leads to a number of start-ups that is similar to the historical record of 284 firms that appeared before 1913 (226 of the 284 start-ups failed by 1913, representing a failure rate of 79%). Hence, the factor is assumed to be between 0.000002 and 0.00002. Before firms are founded, the simulation program checks whether there is any demand for their products and whether the product they offer does not violate any patent. Only if both conditions are satisfied, are firms founded.

Start-up firms always employ one chemist at the beginning and offer only one product. They always try to imitate products of existing firms. The firm and the product that are imitated are chosen randomly. In imitating, start-ups change slightly both characteristics, k and a, of the product. As mentioned earlier, the type of a product includes several aspects: technology, color, durability, textiles for which it can be used, and other similar characteristics. If a product is imitated, it can be assumed that the new product will be in the same dye class and possess the same color as the imitated product. This restricts imitated products to a part of the product space that covers around 1/35 of the entire product space. Hence, a randomly drawn product differs, on average, by a value of between 0.025 and 0.05 in the k-space from the next product of another class or color. Thus, imitation can maximally lead to changes in k that lie between 0.025 and 0.05. The product that is created by imitation is characterized by a k-value that is between $k-\mu$ and $k+\mu$, where $0.025<\mu<0.05$ holds. Some start-up firms are not able to imitate exactly the product. For this reason we subtract randomly a value between 0 and 1 from the performance level of the imitated product.

Exit of firms

Firms disappear if the demand for their products decreases to zero.

Firm strategy

Firms pursue strategies that can differ along three dimensions that are captured by the following questions: (1) Do they try to carry out product innovations, (2) do they imitate products of other firms whose products are far away in the product characteristic space, and (3) how many chemists do they employ? At the beginning, every firm starts with the same strategy. We assume, however, that some firms change their strategy when they experience a decline in sales.

The strategy pursued by a firm is described by a variable $v(t)$ that ranges from 0 to 1. All start-ups are characterized by a strategy variable with a value of 0. The strategy variable is changed whenever the firm is not satisfied with the development of its sales. To check whether the aspiration level of the firm is met, the sales are calculated every month (each time step) and compared to the levels of the preceding month. The relative increase in sales is denoted by g. The historical record shows that not all firms react if their sales decrease. Following the Simon and March tradition, an aspiration level g_a is defined. This value is randomly drawn for each firm when it is started from a uniform distribution between -1 and 0.

Every month the value of g is compared to the aspiration level g_a. If $g>g_a$ holds, the firm is satisfied and the strategy of the firm remains constant. If $g<g_a$ holds, the firm increases its strategy by a certain amount as long as the strategy value $v(t)$ has not reached a value of 1. We

have no systematic knowledge of how the strategy variable of the average historical firm increased over time. Therefore, we have to define the value and its possible range following a different logic. Simulations show that a value of 0.01 causes only very few firms to have a strategy variable of $v(t)=1$ at the end of the simulation, while a value of 0.1 causes almost all existing firms at the end of the simulations to be characterized by $v(t)=1$. Hence, the increase in the strategy value is assumed to range between 0.01 and 0.1 in the simulations to represent what we know to be the approximate conditions in 1913. All firms that have a strategy variable of $v(t)>0$ attempt to carry out product innovations and product imitations (see below). Furthermore, the strategy variable determines the number of chemists employed by a firm as described below.

Chemists employed by firms

Each firm employs at least one chemist. Additional chemists are employed proportional to the size of firms. The higher the sales, $s(t)$, of a firm are, the more chemists it employs. Furthermore, the higher the strategy variable, $v(t)$, is, the more chemists are employed by a firm. The number of chemists, n_c, in a firm is given by

$$n_c = 1 + \left(c_{\min} + v(t) \bullet (c_{\max} - c_{\min})\right) \bullet s(t)$$

where c_{min} and c_{max} are parameters that determine the minimal and maximal number of chemists that firms employ per unit of sales. These parameters are set to c_{min}=0.0006 and c_{max}=0.01 in the simulations, to reflect historically realistic values of chemists in synthetic dye firms (these are the minimum and maximum share of chemists that were employed by the firms Bayer, BASF, Jäger and Levinstein at the end of the period of time studied. (For a detailed study of these firms, see Murmann, forthcoming). If a firm is not able to employ the number of chemists that it requires according to the above equation, it employs as many chemists as it can get and its sales are reduced according to the above equation. The remaining demand is distributed among the other firms in the market if they offer the respective types of products.

Innovations

Firms can offer many different products at the same time. They obtain new products through innovation and imitation. When they find new products they continue to offer the old products. Only when the demand for a product at some point goes to zero is it removed from the firm's product portfolio. Firms find new products in four ways: process innovations, product innovations, process imitations, and product imitations. All firms try to make process innovations. A process innovation means that a new product is created based on a product that is currently offered by the firm. The new product represents the same type, *k*, as the old one, but displays a level of performance, *a*, that is one unit greater. Only the most advanced product of each type can experience additional innovations.

The probability of an innovation is given by the addition of three values:

1. A basic innovation rate: This rate is the same for all firms independent of the variables that characterize their state. Empirical studies show that the average number of innovations by firms per year is around 1 (see Fritsch, Bröskamp and Schwirten 1996 and Evangelista 1999). If this would be caused by the basic innovation rate, this rate would be approximately 0.08. Since there are two other factors included in the innovation process here, we assume the basic innovation rate to be between 0.003 and 0.03 which means that the basic innovation rate accounts for, at most, one third of all innovations.
2. An innovation rate that increases linearly with the sales of a firm. The logic here is: the larger a firm, the higher its innovativeness. This relationship has been confirmed repeatedly in the

literature (see Audretsch and Acs 1991, Anselin, Varga and Acs 1997 and Blind and Grupp 1999). Typical numbers reported in the literature lie between 0.001 and 0.002 innovations per employee, per year (see Audretsch and Acs 1991 and Blind and Grupp 1999). In the dye industry, firms employ approximately 0.3 employees per ton of dyes they sell. Taken together, these two arguments imply that the innovation rate per sale, per month should be between 0.000025 and 0.00005. Since part of the increase might also be included in the third term contributing to the innovation rate, we have to reduce the lower boundary considerably if we assume that most of the innovativeness within the firms is created by the chemists employed. Therefore, we assume this innovation rate to lie between 0.0000001 and 0.000025.

3. An innovation rate that increases linearly with the number of chemists that are working on the products for the respective market. Based on the argument presented above, let us assume that all innovations of a firm that are due to size are caused by the chemists employed. Since between 0.0006 and 0.01 chemists are employed per unit of sales, this means there are between 0.002 and 0.03 chemists per employee. Hence the innovation rate per chemist, per month, has to range between 0.0008 and 0.025. Since it might be that not all innovations are caused by chemists, we assume an innovation rate between 0.0002 and 0.02.

In addition, the number of process innovations is assumed to depend on the highest level of performance that can still be reached. Hence, the sum of the above three values is multiplied by

$$\sqrt{\frac{10\cdot(a_{\max}-a)}{a_{\max}}}\ .$$

This value ranges between 0.3 and 3.2. On average it is 1, so that the innovation rate defined above is the average rate in the simulations. The square is used because otherwise the value would range between 0.1 and 10, which we assume to be a too strong influence. At the beginning the firms are small while at the end of the simulation they are large. The factor is introduced so that the innovation rate amounts to approximately 1 per year, per firm throughout the whole simulation.

Product innovations are not carried out by all firms. Only those firms that have a strategy variable $v(t)$ above zero (the strategy variable is defined below), try to invent new types of products. We assume that the probability for such a product innovation depends on sales and the number of chemists in the firm in the same way as in the case of process innovations. Product innovations, however, are less likely. The rate of product innovations is the rate of process innovations multiplied by a specific factor that is based on the following considerations. Patent statistics in the different countries show that there are between 1 and 5 times as many process as product innovations. Therefore, we assume the factor to range between 0.2 and 1. If a product innovation occurs, the type of the new product is determined randomly. The level of performance of the product is determined by identifying the product on the market that is most similar to the new product (smallest difference in k). The level of performance of this product is transferred to the new product, whereby up to one unit is randomly subtracted.

Firms may profit from spillovers from other firms. We distinguish two types of such spillovers. In the first type firms only profit from the level of performance that other firms have reached. We call these process imitations. Such imitations can be done by all firms. They require that the products of the two firms are similar. Similar means that the products are based on the same technological class and belong to the same class of color (see the above discussion on the product space and Figure 1). Therefore, we use the parameter, μ, that is defined above. We assume that process imitation is possible if the products differ with respect to their type by a value less than μ.

Firms can only profit from spillovers if the other product is technologically more advanced than their own. When firms profit from such spillovers, the level of performance of their own product is increased to the level of the imitated product. Since imitation is not perfect, up to one unit is subtracted randomly.

Historical records show that it generally took firms between half a year and 3 years to imitate a product that they wanted to offer themselves. We assume, therefore, that imitations within a country take place with a probability of 0.03 and 0.15 per month. This holds for process imitations as well as for the product innovations that are described next. Imitations of firms in other countries are more difficult and are assumed to take between 1 and 10 years. This implies an imitation probability of between 0.008 and 0.08 per month.

The second kind of spillovers are product imitations. In this case firms imitate products that are offered in the market by other firms. Similar to product innovations, product imitations are assumed to be carried out only by firms that are characterized by a strategy variable above zero. The probability of a product imitation is the same as the probability of a process imitation. We assume that firms that imitate the products of other firms do not create exactly the same product. The product deviates from the imitated product according to the rules that have been described for start-ups.

c. Demand

The demand for a product is calculated according to a demand function that is based on specific assumptions about the behavior of consumers. For this reason we first outline these assumptions before we formulate the demand function. We assume that each consumer is characterized by the two values k and a, analogous to the way we defined the products in product space. k defines the type of product that the consumer prefers most. a defines the minimal level of performance upon which the consumer switches from buying natural dyes to buying synthetic dyes. This means that at the beginning of the simulation, all consumers buy only natural dyes. The consumers are assumed to be uniformly distributed over the product space, which is given by $0<k<1$ and $0<a<400$. However, the fact that each consumer prefers one kind of product does not imply that the consumer does not buy any other type of product. We assume a Hotelling-type of model (see, e.g., Norman and Thisse 1999 and Foros and Hansen 2001 for a similar use of this modeling approach) according to which products are valued less the more they differ from the preferred good. The value of a good is given by its level of performance, a, (summarizing quality and price into one metric). The evaluation of a product, defined by k_p and a_p, by a consumer who prefers type k is given by

$$a_p - \phi \cdot |k_p - k|.$$

The parameter ϕ denotes the flexibility of the consumers with respect to the type of the product. The higher is ϕ, the less likely are consumers to buy products that differ from the preferred one. Historical price developments show that prices decreased by approximately a factor of 2000 from 1856 to 1913. In our simulation this relates to 395 steps of level of performance. Hence, in the simulation each step of performance advance amounts to a decrease in prices of approximately 1.7% if each step is assumed to cause the same relative price decrease. Historical records also show that buyers buy colors other than their preferred one if they cost only around one third of the price preferred good. Hence, another color is preferred if it is 59 steps technologically more advanced. Above we assumed that 7 different colors exist. Therefore, a product that differs from the preferred product by 1/7 in the k-space (one color) has to be characterized by an a-value that is 59 units higher than that of the preferred product to be bought. Thus, $\kappa \approx 400$. Since this is only an approximation for this parameter, we assume $200 < \phi < 1000$.

National demand

We assume to have at each time a certain number of consumers who either buy natural or synthetic dyes. In the year 1913 almost only synthetic dyes were consumed. We use the number of tons of synthetic dyes consumed in each country as a proxy for the demand at this time. Hence, the demand at the end of the simulation is given as 23,000 (UK), 20,000 (G), 9,000 (F), 3,000 (SW), 26,000 (USA) and 81,000 (rest of the world). This data is provided in Reader (1970, p. 258).

The number of consumers is assumed to increase linearly in each region. The number of consumers at the beginning of the simulation (1856) is estimated to be 59,000. This number is derived from the following considerations: In the UK 75,000 tons of natural dyes were consumed in 1856. This amounts to around 14% of the world consumption. Hence, around 535,000 tons of natural dyes were consumed worldwide. The literature reports that 1 ton of synthetic dye replaced around 9 tons of natural dye. Therefore, the consumption in 1856 would equal around 59,000 tons of synthetic dye.

To distribute these consumers across the different countries two assumptions are made: First, there is a worldwide increase in the demand for dyes. Second, in each country the demand for dyes increases linearly with the growth of the population. Both increases are assumed to be linear, so that only data for the population in each country in 1856 and 1913 has to be used to calculate the number of consumers in each region in 1856. This leads to the following numbers for 1856: 10,700 in UK, 7,800 in G, 6,300 in F, 1.400 in SW, 4.800 in USA and 28.000 in the rest of the world.

Since these numbers represent rough estimates of the real demand in 1856, we defined a range of values that could have represented the true historical demand. We determined the historical demand to range between 0.5 and 1.5 times the estimated demand based on the following considerations. The historical record makes it clear that the demand in 1856 in each country was not higher than demand in 1913. This implies for France that the demand in 1856 could not have been more than 1.5 times the value estimated above. This gives us the following range of estimates for the historical demand in 1857 by country: 5,300-16,100 for UK, 3,900-11,700 for G, 3,100-9,500 for F, 700-2,100 for SW, 2,400-7,200 for USA and 14,000-42,000 for the rest of the world.

Inter-regional transaction costs

Although consumers prefer to buy the dye products at their own location, they are also able to buy products in other countries via trading firms. Because traders have to be involved in these international sales, transaction costs arise. In the simulation, this additional cost of selling products in other countries is treated as being equivalent to offering products with a lower level of performance. As calculated above, in the simulation each step of technological advance amounts to a decrease in prices of approximately 1.7% if each step is assumed to cause the same relative price decrease. We assume that inter-regional transaction costs amount to between 5 and 25% of the domestic base price. Hence, products that are sold in other countries will automatically receive a score for their level of performance that is decreased by a value between 3 and 15.

Calculation of demand

To calculate demand, the product space is divided into 200 uniform parts (the number of parts only determine the precision of the calculation and the CPU time used by the simulations). For each part and each country, the demand is calculated separately. The total demand in each part is given by the respective share (1/200) of the national demand. All products on the market compete for this share if they are not excluded from this market by patent laws (patenting is described in detail below). Each part is assigned the type, k, that characterizes its average good (k=0.0025, 0.0075, 0.0125,...,0.9975). All products on the market are transferred to this type and the country under consideration by subtracting the inter-regional transaction costs and the devaluation caused by the difference in types from their level of performance. Through this calculation a re-scaled value of performance is assigned to each product. The product with the highest value, $\tilde{a}$, is identified. This value determines the share of demand in the part of the product space under consideration that is received by synthetic dye firms rather than natural dye makers. This specific market share is given by $(\tilde{a}_{max}/400)^{\gamma}$, where γ is a parameter that determines how fast natural dyes are replaced. The replacement of natural dyes seems to have been more rapid for initial rather than later steps in the technological improvement of dyes. Hence, the parameter is assumed to be below 1. $0.5<\gamma<0.9$ is used in the simulations.

The resulting demand is shared by all products on the market according to a linearly decreasing demand function (the usual way of modeling competition in the case of oligopolies). The slope of this demand function is set such that a product with $\tilde{a} < \tilde{a}_{max} - \delta$ does not receive any demand. δ is a parameter that determines how much lower the quality or higher the price of products can be to still be sold on the market. We assume $3<\delta<15$, which means that products with prices that are between 5 and 25% higher than other dyes receive only a minimal share of demand.

According to these rules the potential demand, d_{pot}, is calculated for each firm. However, we assume that consumers take some time to switch supplying firms even if better offers are already available in the current period. The sales, $s(t)$, are assumed to be given by

$$s(t) = (1-\sigma)\cdot s(t-1) + \sigma \cdot d_{pot}$$

where σ is a parameter that determines the speed of consumer adaptation. We assume $0.25<\sigma<0.75$, which implies, on average, that consumer take between 10 and 100 days to switch to a producer that offers a preferred good.

d. Chemists

Chemists are modeled explicitly in the simulations. Each chemist is characterized in each period of the simulation by three characteristics: age, region of origin and current location. Obviously, the chemist's age changes over time. Every 6 months it is updated. The region of origin is determined as described below and remains constant over time. The current location describes whether a chemist is available in the labor market or employed by a firm. This variable also contains information about which firm employs a particular chemist.

The initial numbers of chemists in each country, denoted by c_{init}, are set in accordance with the estimates given in Table 9. The ages of the chemists in these initial national pools of chemists are distributed uniformly among the possible ages. We assume that all additional chemists available to dye firms in subsequent years first have to be trained in universities. Training takes 3 years. Hence, the number of chemistry students deciding to start university education at a particular time influences the number of available chemists three years later. The study of chemistry can be started every 6 months with the beginning of a new semester.

Chemists remain in the labor market for 35 years before they retire.

Table 9: Estimates of chemists and estimates for 1857 assuming exponentially increasing numbers.

year	F	G	UK	SW	USA
1850	25	35	15	8	5
1900	225	750	150	75	75
estimate for 1857	34.0	53.8	20.7	10.9	7.3

Source: Ernst Homburg (E-mail, November, 2002)

Chemists who have been in the simulation for 38 years (3 years study and 35 years on the labor market) disappear. The number of new students is calculated as follows (the calculation is done for each country separately):

$$c_{new} = e \cdot c_{emp}(t) - c(t)$$

where $c_{emp}(t)$ is the number of chemists that are employed at time, t, $c(t)$ is the number of chemists available at that time and e is a factor that represents the ability of the national university system to provide the chemists that are needed by firms. If e=1, the national system provides exactly the number of chemists employed. The highest increase that can be observed in the studied history is an increase of chemists in Germany from 380 to 900 chemists during the period from 1851 to 1865. This increase is reached if e=1.2. Therefore, we assume $1<e<1.2$.

Migration of chemists

We assume that chemists who can find employment in a domestic dye company will join a domestic firm. Those chemists who are not employed by domestic dye companies will either stay in their home country and work for a different branch of the chemical industry or they will move to a dye firm in a foreign country. The relative frequency of moving abroad or staying in the home chemical industry enters the parameter matrix m_{ij} of our model: m_{ij} is the share of all chemists who originate from country i and would migrate to country j if they find employment in a dye company in country j but they do not find an employment in a domestic dye company and would have to work in a different branch of the domestic chemical industry.
The entries of the matrix are set to:

Table 10: Migration Probabilities for Chemists

Origin	Destination				
	UK	**G**	**F**	**SW**	**USA**
UK	-	0.005	0.005	0	0.02
G	0.025	-	0.01	0.045	0.005
F	0.01	0.005	-	0.05	0
SW	0.003	0.1	0.006	-	0.001
USA	0	0	0	0	-

e. Patents

Process innovation, product innovation, process imitation, product imitation, and the selling of products in other countries are restricted by patent laws. Each time a firm discovers a new product by any of the four processes, our simulation program checks whether a patent exists for this product in the country in which the firm is located. If such a patent exists and is held by another firm, the product is not produced. If no such patent exists but the home country offers patent protection, the firm patents the product.

Different patent laws existed in the countries modeled. First, patent laws did not exist in all countries during the entire time span from 1857 to 1913. Second, patent laws differed between the countries. All countries in our simulation except Germany granted patents for product innovations. After 1877, Germany granted process patents. We model this by defining ranges within the product space that patents cover. This means that each patent has a range with respect to the type of products and with respect to the level of performance of products. Patent counts show that there were between 1000 and 4000 different types of dyes patented in the different countries. If we take this as an approximation of the number of different types of products in our simulation of patents, each patent covers a range of between 0.00025 and 0.001 in the type dimension. In addition, in Germany process innovations were patented. These cover only one innovation step, so that their range is assumed to be 1 in the dimension of level of performance.

Because successful litigation often cut short the life-span of a patent, all patents are assumed to hold for between 10 and 20 years (in accordance to the historical record). All characteristics of patents are summarized in the following table.

Table 11: National Patent Characteristics

	existence	range in type dimension	range in techno-logical performance dimension	duration
UK	1856-1913	0.0025-0.001	∞	10-20 years
G	1877-1913	0.0025-0.001	1	10-20 years
F	1856-1913	0.0025-0.001	∞	10-20 years
SW	1908-1913	0.0025-0.001	∞	10-20 years
USA	1856-1913	0.0025-0.001	∞	10-20 years

Once patents expire, it is assumed in the simulation that the same product cannot be patented again. The respective products can then be produced and sold by all firms forever. Patents also restrict the selling of products. Firms are only allowed to sell products in other countries if these products are not covered by a patent there.

Bibliography

Abbott, Andrew. 1998. "Causal Devolution." *Sociological Methods & Research* 27:148-181.
Anselin, L., A. Varga, and Z.J. Acs. 1997. "Local Geographic Spillovers between University Research and High Technology Innovations." *Journal of Urban Economics* 42: 422-448.
Arthur, W. Brian. 1994. *Increasing Returns and Path Dependence in the Economy*. Ann Arbor: University of Michigan Press.
Audretsch, D.B. and Z.J. Acs. 1991. "Innovation and Size at the Firm Level." *Southern Economic Journal* 57: 739-744.
Ben-David, Joseph. 1971/1984. *The Scientist's Role in Society*. Chicago: University of Chicago Press.
Ben-David, Joseph. 1977. *Centers of Learning: Britain, France, Germany, United States*. New York: McGraw-Hill Book Company.
Blind, K. and H. Grupp. 1999. „Interdependencies between the Science and Technology Infrastructure and Innovation Activities in German Regions: Empirical Findings and Policy Consequences." *Research Policy* 28: 451-468.
Brenner, Thomas. 2001. "Simulating the Evolution of Localised Industrial Clusters -- An Identification of the Basic Mechanisms." *Journal of Artifical Social System Simulations* 4 (3): 4.
Carley, Kathleen M. 1995. "Computational and Mathematical Organization Theory: Perspectives and Directions." *Computational and Mathematical Organization Theory* 1:39-56.
Carley, Kathleen M., and Vanessa Hill. 2001. "Structural Change and Learning within Organizations." pp. 63-92 in *Dynamics of Organizations: Computational Modeling and Organization Theories*, edited by Alessandro Lomi and Erik R. Larsen. Menlo Park, CA and Cambridge, Massachusetts: AAAI Press/MIT Press.
Collins, Randall. 1998. *The sociology of philosophies: a global theory of intellectual change*. Cambridge, Mass.: Belknap Press of Harvard University Press.
Cyert, Richard M., and James G. March. 1963/1992. *A Behavioral Theory of the Firm*. Cambridge, Massachusetts: Blackwell Publishers.
David, Paul A. 1985. "Clio and the Economics of QWERTY." *American Economic Review* 75:332-337.
Eliasson, Gunnar, and Erol Taymaz. 2000. "Institutions, Entrepreneurship, Economic Flexibility and Growth – Experiments on an Evolutionary Micro-to-marco Model." pp. 265-286 in *Economic Evolution, Learning, and Complexity*, edited by Uwe Cantner, Horst Hanusch and Steven Klepper. Heidelberg: Springer-Verlag.
Evangelista, R. 1999. *Knowledge and Investment*. Cheltenham: Edward Elgar.
Foros, O., and B. Hansen. 2001. "Competition and Compatiblity Among Internet Service Providers." *Information Economics and Policy* 13: 411-425.
Fritsch, M., A. Bröskamp, and C. Schwirten. 1996. „Innovationen in der sächsischen Industrie - Erste empirische Ergebnisse." *Freiberg Working Papers 96/13*, Technische Universität Bergakademie Freiberg.
Gilbert, Nigel G. 1996. "Simulation as a Research Strategy." pp. 448-454 in *Social Science Microsimulation*, edited by G. Nigel Gilbert and Jim E. Doran. Berlin: Springer-Verlag.
Gilfillan, S.C. 1935/1970. *The Sociology of Invention*. Cambridge, Massachusetts: The M.I.T. Press.
Haber, L. F. 1958. *The Chemical Industry during the Nineteenth Century*. Oxford: Oxford

University Press.
Heise, David R. 1975. *Causal Analysis*. New York: John Wiley & Sons.
Homburg, Ernst. 1993. *'Chemiker' By Occupation: The Rise of the Industrial Chemist and Polytechnic Education in Germany, 1790-1850*. Delft, Netherlands: Delftse Universitaire Pers.
Homburg, Ernst, Anthony S. Travis, and Harm G. Schröter (Eds.). 1998. *The Chemical Industry in Europe, 1850-1914: Industrial Growth, Pollution, and Professionalization*. Dordrecht and Boston: Kluwer Academic.
Jacobsen, Chanoch and Richard Bromson. 1997. "Social Theory: Lessons from 15 Years' Experience." pp. 97-102 in *Simulating Social Phenomena*, edited by Rosana Conte, Rainer Hegselmann and Pietro Terna. Berlin: Springer-Verlag.
Krackhardt, David M. 2001. "Viscosity Models and the Diffusion of Controversial Innovation." pp. 273-268 in *Dynamics of Organizations: Computational Modeling and Organization Theories*, edited by Alessandro Lomi and Erik R. Larsen. Menlo Park, CA and Cambridge, Massachusetts: AAAI Press/MIT Press.
Kwasnicki, Witold. 1999. "Evolutionary Economics and Simulation." pp. 3-44 in *Computational Techniques for Modelling Learning in Economics*, edited by Thomas Brenner. Boston: Kluwer Academic Publisher.
Landes, David S. 1969. *The Unbound Prometheus: technological change and industrial development in Western Europe from 1750 to the present*. New York: Cambridge University Press.
Levinthal, Daniel. 2001. "Modeling Adaptation on Rugged Landscapes." pp. 377-404 in *Dynamics of Organizations: Computational Modeling and Organization Theories*, edited by Alessandro Lomi and Erik R. Larsen. Menlo Park, CA and Cambridge, Massachusetts: AAAI Press/MIT Press.
Liebrand, Wim B.G. 1998. "Computer Modeling and the Analysis of Complex Human Behavior: Retrospect and Prospect." pp. 1-14 in *Computer Modeling of Social Processes,* edited by Wim B.G. Liebrand, Andrzej Nowak and Rainer Hegselmann. .
Lomi, Alessandro, and Erik R. Larsen. 2001. "Introduction." pp. 3-36 in *Dynamics of Organizations: Computational Modeling and Organization Theories*, edited by Alessandro Lomi and Erik R. Larsen. Menlo Park, CA and Cambridge, Massachusetts: AAAI Press/MIT Press.
Macy, Michael, and David Strang. 2001. "Dedicated Followers of Success: A Computational Model of Fashionable Innovations." pp. 93-188 in *Dynamics of Organizations: Computational Modeling and Organization Theories*, edited by Alessandro Lomi and Erik R. Larsen. Menlo Park, CA and Cambridge, Massachusetts: AAAI Press/MIT Press.
Malerba, Franco, Richard Nelson, Luigi Orsenigo, and Sidney Winter. 1999. "'History-friendly' Models of Industry Evolution: The Computer Industry." *Industrial and Corporate Change* 8:3-40.
Malerba, Franco, and Luigi Orsenigo. 2002. "Innovation and Market Structure in the dynamics of the pharmaceutical industry and biotechnology: toward a History-Friendly Model." *Industrial and Corporate Change* 11:667-703.
March, James G. 2001. "Foreword." pp. ix-x in *Dynamics of Organizations: Computational Modeling and Organization Theories*, edited by Alessandro Lomi and Erik R. Larsen. Menlo Park, CA and Cambridge, Massachusetts: AAAI Press/MIT Press.
March, James G., and Herbert A. Simon. 1958. *Organizations*. New York: Wiley.

Marx, Karl. 1906. *Capital*. Chicago: Kerr.

McClelland, Charles E. 1980. *State, society, and university in Germany, 1700-1914*. Cambridge New York: Cambridge University Press.

Merton, Robert K. 1938. *Science, technology and society in seventeenth century England*. Bruges, Belgium,: Saint Catherine Press.

Mokyr, Joel. 1990. *The Lever of Riches: Technological Creativity and Economic Progress*. New York: Oxford University Press.

Murmann, Johann Peter. Forthcoming. *Knowledge and Competitive Advantage: The Coevolution of Firms, Technology, and National Institutions*. New York: Cambridge University Press.

Nelson, Richard R., and Sidney G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge: The Belknap Press of Harvard University Press.

Norman, G., and J.-F. Thisse. 1999. "Technology Choice and Market Structure: Strategic Aspects of Flexible Manufacturing." *Journal of Industrial Economics* 47: 345-372.

Richards, Gordon R. 2002. "A Simulation Model with Endogenous Technical Advance: Information Technology and Increasing Returns from Research." *Technological Forecasting & Social Change* 69: 29-51.

Rosenberg, Nathan. 1998. "Technological Change in Chemicals: The Role of the University-Industry Interface." pp. 193-230 in *Chemicals and Long-Term Economic Growth: Insights from the Chemical Industry*, edited by Ashish Arora, Ralph Landau, and Nathan Rosenberg. New York: John Wiley & Sons, Inc.

Sastry, M. Anjali. 2001. "Understanding Dynamic Complexity in Organization Evolution: A System Dynamics Approach." pp. 377-404 in *Dynamics of Organizations: Computational Modeling and Organization Theories*, edited by Alessandro Lomi and Erik R. Larsen. Menlo Park, CA and Cambridge, Massachusetts: AAAI Press/MIT Press.

Schmookler, J. 1965. "Catastrophe and Utilitarianism in the Development of Science." in *Economics of Research and Development*, edited by Richard A. Tybout. Columbus: Ohio State University.

Silverberg, Gerald, and Bart Verspagen. 2002. "A Percolation Model of Innovation in Complex Technology Spaces". MERIT-Infonomics Research Memorandum series 2002-025.

Tilly, Charles, and Arthur L. Stinchcombe. 1997. *Roads from past to future*. Lanham: Rowman & Littlefield.

Troitzsch, Klaus G. 1999. "Simulation as a Tool to Model Stochastic Processes in Complex Systems." pp. 45-69 in *Computational Techniques for Modelling Learning in Economics*, edited by Thomas Brenner. Boston: Kluwer Academic Publisher.

Winer, B.J. 1971. *Statistical Principles in Experimental Design*. NewYork: McGraw-Hill.