![JASSS logo]

Bruce Edmonds, Nigel Gilbert, Petra Ahrweiler and Andrea Scharnhorst (2011)

Simulating the Social Processes of Science

Abstract

Science is the result of a substantially social process. That is, science relies on many inter-personal processes, including: selection and communication of research findings, discussion of method, checking and judgement of others' research, development of norms of scientific behaviour, organisation of the application of specialist skills/tools, and the organisation of each field (e.g. allocation of funding). An isolated individual, however clever and well resourced, would not produce science as we know it today. Furthermore, science is full of the social phenomena that are observed elsewhere: fashions, concern with status and reputation, group-identification, collective judgements, social norms, competitive and defensive actions, to name a few. Science is centrally important to most societies in the world, not only in technical, military and economic ways, but also in the cultural impacts it has, providing ways of thinking about ourselves, our society and our environment. If we believe the following: simulation is a useful tool for understanding social phenomena, science is substantially a social phenomenon, and it is important to understand how science operates, then it follows that we should be attempting to build simulation models of the social aspects of science. This Special Section of *JASSS* presents a collection of position papers by philosophers, sociologists and others describing the features and issues the authors would like to see in social simulations of the many processes and aspects that we lump together as "science". It is intended that this collection will inform and motivate substantial simulation work as described in the last section of this introduction.

Keywords:
Simulation, Science, Science and Technology Studies, Philosophy, Sociology, Social Processes

Aim of the collection

1.1   The authors were invited to write position papers outlining what a simulation of the social processes of science should be like. This invitation was open to all viewpoints on the efficacy and nature of science. We sought to by-pass the debates on whether science is a special and/or uniquely effective social phenomenon. Rather, we simply wish to try and understand what happens in these processes. We did not expect agreement on the nature of science since it is a highly complex phenomenon, which includes a great variety of processes. However we did ask the authors to focus upon areas where they think simulation can contribute to our understanding rather than continuing the wider debate on the nature of science. Thus the purpose (though not necessarily the orientation) of this collection is *pragmatic* - to motivate the *building* of simulations of these social processes in science, in particular agent-based simulations. If this collection helps to stimulate the building of some new agent-based simulations of some aspects of science, it will have achieved its purpose.

Previous work

2.1   Previous models of science have often started from a desire to explain the 'stylised facts' about the growth of science that were noted in the last century. Lotka (1926) showed that the numbers of papers per author followed a power-law or scale-free distribution, while Price (1976) found such a distribution for citations per paper. Price (1963) had earlier observed exponential growth rates in papers and authors in the field of physics and reflected on the implications of this. Simon (1957) presented a simple stochastic-process model to generate a scale-free frequency distribution, and fitted it to Lotka's data.

2.2   Contributions to science modelling since Simon have explored the mathematical implications of such stochastic process models (Schubert and Glanzel 1984; Glanzel and Schubert 1990, 1995; Burrell 2001). For example Burrell (2001) relates the citation

process to the ageing and eventual obsolescence of papers. Burrell (2007) employs a stochastic model to estimate the behaviour under different conditions of Hirsch's h-index for measuring research output and impact based on citations.

2.3 Whereas Simon's (1957) urn model simply generated a frequency distribution for papers per author, Gilbert (1997) represented individual academic papers with references to past papers and their content. Using two continuous variables to represent paper topics, his model depicts an academic field as a two-dimensional plane. Subfields appear within this model as clusters of points. The *TARL* model ('Topics, Aging and Recursive Linking') of Börner et al. (2004) represents both authors and papers, including references and 'topics' for papers, and generates network data. The behaviour of scientists publishing within academic fields has been compared to heuristic search (Bruckner et al. 1990; Scharnhorst and Ebeling 2005; Chen et al. 2009). Weisberg and Muldoon (2009) also employ landscape search as a model for science.

2.4 While fairly simple, Gilbert's model inspired models in different areas (e.g.Boudourides and Antypas 2002) and using different approaches. Sun and Naveh (2009) extended the model by giving the scientist-agents a learning mechanism and the ability to select which areas they choose to work in. Watts and Gilbert (forthcoming) added a representation of the influence of academic journals and their referees to the mix.

2.5 Edmonds (2007) developed a model in which scientists are represented as theorem provers, generating new theorems by inference from existing premises. In this model, as in those of Ahrweiler (1998), Ahrweiler and Wolkenhauer (1998), Weisberg and Muldoon (2009) and Grim (2009), there is an attempt to model an explicit epistemic landscape in which some locations are harder to discover than others (see also Watts and Gilbert, forthcoming).

🌐 The current state of the art

3.1 Mathematical models of science that could form the basis for a simulation appear occasionally at conferences such as the annual meeting of the Society for Social Studies of Science (4S), or of the conference of the International Society of Scientometrics and Informetrics (ISSI). They can also be found in conferences on network science, statistical physics[1], sociology or even computational philosophy[2]. Few of these come with systematic simulation-based experiments - the community of quantitative science research has embraced simulation as an independent research method even less than it embraced dynamic mathematical models. It is therefore not surprising that simulations of the science system tend to appear at the interfaces between communities that do use simulation as a method, such as sociology and (somewhat unexpectedly) philosophy. Given the fact that current science studies, science of science, and science and technology studies are scattered among very different schools of thought and historic traditions, it is not surprising that the simulation research that occurs in one school has not been taken up in others. Thus we have the present situation where there is relatively little simulation work and what occurs is isolated.

3.2 A recent book, *Models of science dynamics* (Scharnhorst et al., forthcoming) collects review articles about different formal modelling techniques, including epidemic and opinion dynamics models of idea diffusion, evolutionary game theory on complex networks, and network analysis of co-authorship and citation networks. One chapter of this book (Lucio-Arias and Scharnhorst 2011) conducts an algorithmic historiography (historical research by means of bibliometrics) of mathematical models of science. It shows that despite a growth in publishing activities in this area, the recent threads of (mathematical) models do not refer to each other but remain isolated and this is also the case for existing simulations of the social processes of science. For the most part in this book analytic mathematical approaches predominate. Simulation appears, if at all, at the sidelines, and is not discussed as an independent research method.

3.3 However there is one chapter (Payette, forthcoming) that surveys agent-based models of science (including some of those mentioned above) and proposes a new one based on Hull (1988). The chapter also describes some general properties of agent-based models, focussing on those listed in Epstein (2006). He quotes Epstein: *"The main desideratum is that the notion of 'local' be well posed."* (2006, p. 6) and *"If you didn't grow it, you didn't explain it."* (2006, p. 51). These two principles can be seen as encapsulating the goal of explaining (growing) the macro from the micro (the local). In comparison there has been quite a lot of interest in either the micro (the individual scientist) *or* the macro.

3.4 On the macro side there is a growing request for mathematical models of science to infer initial and boundary conditions of "good" scientific activity; as well as to forecast the broad development of the science system. However the need is not for an abstract mathematical foundation but for very practical, empirical grounded scenario development that can inform policymaking. Such efforts could include a broad range of techniques including visual experiments (as for instance laying out evolving networks or overlapping knowledge diffusion processes to form global science maps as done, e.g., by Rafols et al. 2010), empirical validations and the simulation of theoretical assumptions. There is now a stream of research that seeks to capture aggregate patterns in the traces left by science (citation analysis, co-authorship patterns, and various other distributions observable in terms publishing or patents). These have the potential to aid the validation of simulation models, as demonstrated in Gilbert (1997) and Sun and Naveh (2009).

3.5 On the micro side there was a stream of research on the borders of artificial intelligence and philosophy of science to model how a single scientist might reason and induct new hypotheses (e.g. Holland et al. 1989, Thagard 1993). These did not consider social aspects of cognition or behaviour and since then the importance of modelling individual behaviour has lost out to the macro side. However, recently there has been a renewed focus on the individual. The "return of the actor" has been due to new ways to

trace scientific authors in bibliographic databases and on the web (e.g. Thompson Reuters' ResearcherID). This shadows a trend in social network analysis to elaborate the role and content of their nodes, allowing a more active behaviour on their part, with the dynamics of the interactions along network links becoming more evident as well as a new focus on how networks themselves might be changing.

3.6   The only technique currently available to link these micro and macro sides that does so in a precise and replicable manner, open to detailed critique and step-wise improvement, is agent-based simulation. The necessity for this approach can be seen as a result of the social embeddedness of interaction in science (Granovetter 1985): that if we reduce our models of science to only the macro (essentially reducing the interaction to some global relationships between factors plus noise) or only the individual (ignoring social effects), then we shall miss substantial parts of the story.

3.7   This turn back to the actor, combined with the turn towards time, dynamics and complexity in science and philosophy of science studies, along with the development of social simulation, provides a fertile academic background for this initiative. The possibilities of the semantic web to provide a systematic empirical basis for scholarly communication, and the increasingly sophisticated analyses of networks, are providing more ways to validate simulations. Agent-based simulations of social processes are able to incorporate lessons from qualitative social science studies of what scientists actually do on a day-to-day level as well as insights from the more naturalistic philosophers of science.

3.8   To summarise, there are relatively few existing simulations of the social processes that occur in science, but the ground is now ripe for these. We are aware of the start of a steady stream of papers on such simulations, including upcoming work by the following (in no particular order):

- Luna De Ferrari, Stuart Aitken, Jano van Hemert, and Igor Goryanin at the Computational Systems Biology group at the Centre for Systems Biology at Edinburgh (De Ferrari et al. 2009)
- Francisco Grimaldo, Mario Paolucci, and Rosaria Conte at LABS/ISTC at CNR Rome (e.g. Grimaldo et al, 2011)
- Giangiacomo Bravo (Torino, Italy), Flaminio Squazzoni (Brescia, Italy), Károly Takács (Corvinus University of Budapest, Hungary)
- Andre Martins (São Paulo, Brazil) (Martins 2010)
- Nicolas Payette (Département de philosophie, Université du Québec à Montréal) who is developing a simulation based on Hull (1988) (e.g. Payette 2011)
- Ron Sun (Rensselaer Polytechnic Institute, NY) and Isaac Naveh (University of Missouri, USA) (e.g. Sun and Naveh 2009)
- Nigel Gilbert (University of Surrey), Andreas Pyka (University of Hohenheim), and Petra Ahrweiler (University College Dublin) (e.g. Pyka, Gilbert and Ahrweiler 2007)
- Christopher Watts and Nigel Gilbert at the University of Surrey (e.g. Watts and Gilbert, Scientometrics, forthcoming)
- Paul Thagard and his team at Computational Epistemology Laboratory, University of Waterloo, who are extending their ECHO model of scientific inference to incorporate social aspects (following Thagard 1993, 2000)
- Petra Ahrweiler (University College Dublin) and Tyll Krueger (University of Bielefeld) who work on a project called "Semantic Landscapes" to model the language- and context-based features of science (http://abs-diffusion.univie.ac.at/program/).

The contributions

4.1   Answering the call for position papers, the following sixteen contributions (in alphabetical order) were submitted:

- To assist scientific discourse, *Ahrweiler* opts for a combined language- and behaviour-based framework for modelling theory networks in science, which looks at theories as competing and cooperating agents working on scientific domains.
- *Balzer and Manhart* emphasise the difference between scientific processes and processes in science, and explain how the incorporation of scientific theories in social simulations could lead to more united structural approaches.
- *Barreteau and Le Page* outline the complex dynamics, especially micro dynamics, involved in participatory research methodologies, and show how social simulation can help to address these issues.
- *Chattoe-Brown* identifies two challenges for simulating science: firstly to develop a "dynamic concept network" representation of scientific knowledge on which learning systems intended to model the scientific process can be compared; and secondly to develop an effective approach to providing data for a simulation of the scientific process.
- *Collins* starts from the demarcation problem, asking what science actually is, which leads to a range of difficulties for simulation, and puts forward three recommendations about how to deal with the issue.
- *Doran* suggests a generic long-term science model where science is a set of processes by which a community of individuals uses reliable methods to obtain reliable understanding (scientific knowledge) of itself and its environment over time.
- *Edmonds* surveys the observations and conclusions of some philosophers of science that might be relevant to a social simulation of science, observing that philosophers of science have not focussed much on the dynamic, social and complex aspects of science, which illustrates the need for simulations.
- Taking the example of Robotics as a domain, *Matthew Francisco, Staša Milojević and Selma Šabanović* model conferences as venues in which social, cognitive, and institutional practices of science are performed and which provide

a basis for analysis bridging local and system level features of science.

- *Meyer* addresses the question of how to design good social simulation models of science building on stylised facts of science derived from bibliometric studies.
- *Mölders, Fink and Weyer* combine a Luhmannian systems perspective with a model of decision making of individual actors embedded in a socio-political context ("new public management of science") to reconstruct and analyse how the science system works.
- *Parinov and Neylon* discuss how virtual research environments influence the social processes of science and how, building on social simulation insights, these systems could be designed to be more efficient and effective in supporting scientific communities.
- *Payette* conceptualises an agent-based model of the social processes of science that contains researchers who are organised in heterogeneous networks and who work on different domains communicating directly or through publications.
- *Squazzoni and Takács* argue for social simulation of the scientific peer review system, which is under increasing strain due to exploding demand, is under-investigated compared to its importance, and is in need of revision and innovation itself.
- *Thorngate, Liu and Chowdhury* apply a fundamental observation to the science field, namely that psychological factors such as competition for attention influence the social processes involved in the evolution of science such as the review process for journal papers.
- *Yilmaz* addresses general issues of workforce dynamics and applies them to science introducing various models while asking what produces successful scientists, and what identifies areas for additional research.
- *Zollman* points out that it is unknown how the imperfections of individual researchers impact upon the overall efficacy of science. He poses five key questions that have real and substantial bearing on the management and understanding of science, each of which could be the goal of a modelling programme.

Next steps

5.1 The aim of this collection of position papers is to motivate and challenge those in the social simulation community to attempt simulation models of the social aspects of science. The issues raised and the directions indicated in these papers should help inform and guide these attempts. We hope that any models developed in response will:

- Bridge the micro-macro gap in some way, that is establish explanations that link macro level outcomes from the micro level behaviour of individuals, and *vice versa*
- Be motivated in terms of their conception and design with respect to this collection of papers
- Include some indication of how and in what way they might be checked and/or validated.

5.2 After a suitable time, we (the authors of this introduction) will organise a workshop for the discussion of papers that respond to this collection. Responses which present credible simulations will be centre stage at this event, but others will also be involved. The idea is that it should be a forum to present and discuss these simulations in an extended manner, and thus motivate the production of more and better simulations in the future. We hope to eventually publish a set of papers that describe these.

5.3 Thus we call for contributions to this project from all fields, but especially those in social simulation and science studies, and look forward to the workshop in 1 to 2 years time.

Notes

[1] See as example the Focus Session: Science of Science at the Spring Conference of the German Physics Society 2010, http://www.dpg-verhandlungen.de/2010/regensburg/soe_en.html

[2] See for an recent example: The 2009 North American Conference on Computing and Philosophy, Indiana Bloomington, http://www.iacap.org/redirect.php?orig=na-cap09/program.htm

References

AHRWEILER, P. & Wolkenhauer, R. (1998). 'SiSiFOS´- Simulating Studies on the internal Formation and the Organization of Science'´ In: P. Ahrweiler and N. Gilbert (eds.). *Computer Simulations in Science and Technology Studies*. Berlin, New York: Springer: pp. 129-143. [doi:10.1007/978-3-642-58270-7_9]

AHRWEILER, P. (1998). Theories in (Inter)Action: A complex dynamic System for Theory Evaluation in Science. In Y. Bar-Yam (ed.). *Unifying Themes in Complex Systems*. Boston: Perseus Books: pp. 75-85.

BÖRNER K., Maru J. T. & Goldstone, R. L. (2004) .The simultaneous evolution of author and paper networks. *Proceedings of the National Academy of Sciences*, 101(S1), 5266-5273. [doi:10.1073/pnas.0307625100]

BRUCKNER, E., Ebeling, W. & Scharnhorst, A. (1990). The Application of Evolution Models in Scientometrics. *Scientometrics*, 18 (1-2), 21-41. [doi:10.1007/bf02019160]

BURRELL, Q. L. (2001). Stochastic modelling of the first-citation distribution. *Scientometrics*, 52(1), 3-12. [doi:10.1023/A:1012751509975]

BURRELL, Q. L. (2007) Hirsch's h-index: A stochastic model. *Journal of Informetrics*, 1, 16-25. [doi:10.1016/j.joi.2006.07.001]

CHEN C., Chen Y., Horowitz M., Hou H., Liu Z., & Pellegrino D. (2009) Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, 3(3), 191-209. [doi:10.1016/j.joi.2009.03.004]

DE FERRARI, L., Aitken, S., van Hemert, J. & Goryanin, I. (2009). A model of social collaboration in Molecular Biology knowledge bases. 6th European Social Simulation Association Conference in Guildford, UK.

EDMONDS, B. (2007). Artificial science: A simulation to study the social processes of science. In Edmonds, B. Troitzsch, K. G. & Iglesias, C. H. (Eds.) *Social Simulation: Technologies, Advances and New Discoveries*, pp. 61-67. IGI Global.

EPSTEIN, Joshua M. (2006) *Generative Social Science*. Princeton: Princeton University Press

GILBERT, N. (1997). A simulation of the structure of academic science. *Sociological Research Online*, 2(2), http://www.socresonline.org.uk/2/2/3.html [doi:10.5153/sro.85]

GLÄNZEL W., Schubert A. (1995). Predictive aspects of a stochastic model for citation processes. *Information processing & management*, 31(1), 69-80. [doi:10.1016/0306-4573(95)80007-G]

GLÄNZEL, W., Schubert, A. (1990). The cumulative advantage function. A mathematical formulation based on conditional expectations and its application to scientometric distributions. *Informetrics*, 89/90, 139-147.

GRANOVETTER, M. (1985). Economic Action and Social Structure: the Problem of Embeddedness., *American Journal of Sociology*, 91, 481-93. [doi:10.1086/228311]

GRIM, P. (2009). Threshold phenomena in epistemic networks. In Complex Adaptive Systems and the Threshold Effect: Views from the Natural and Social Sciences, Washington, DC, 2009. http://aaai.org/ocs/index.php/FSS/FSS09/paper/download/916/1234

GRIMALDO F., Paolucci, M., and Conte, R. (2011). Agent Simulation of Peer Review: The PR-1 Model. The 12th International Workshop on Multi-Agent-Based Simulation. (MABS 2011). Taipei (Taiwan), May 2011.[Accepted]

HOLLAND, J. H., Holyoak, K. J., Nisbett, R. E. & Thagard, P. R. (1989) *Induction - Processes of Inference, Learning, and Discovery*. MIT Press.

HULL D. (1988). *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*, University of Chicago Press. [doi:10.7208/chicago/9780226360492.001.0001]

LOTKA A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* 16(2), 317-323.

Lucio-Arias D. & Scharnhorst A. (forthcoming). Mathematical approaches to modelling science from an algorithmic-historiography perspective. In: Scharnhorst A, Börner K, Van den Besselaar P (eds.). *Models of Science Dynamics*. Springer.

MARTINS, A. C. R. (2010). Modeling Scientific Agents For A Better Science, *Advances in Complex Systems*, 13(4), 519-533. [doi:10.1142/S0219525910002694]

BOUDOURIDES, M. & Antypas, G. (2002) 'A Simulation of the Structure of the World-Wide Web' *Sociological Research Online*, 7(1), http://www.socresonline.org.uk/7/1/boudourides.html [doi:10.5153/sro.684]

PAYETTE N. (2011). Agent-based models of science. In: Scharnhorst A., Börner K., & Van den Besselaar P. (eds.) (forthcoming) *Models of Science Dynamics*. Springer. Ch 4. Springer.

PRICE, D. de Solla (1963). *Little Science, Big Science and Beyond*. Columbia University Press, New York.

PRICE, D. de Solla (1976) A General Theory of Bibliometric and Other Cumulative Advantage Processes. *Journal of the American Society for Information Science*, 27, 292-306 (1976). [doi:10.1002/asi.4630270505]

PYKA, A., Gilbert, N., & Ahrweiler, P. (2007). Simulating Knowledge-Generation and Distribution Processes in Innovation Collaborations and Networks. *Cybernetics and Systems*, 38(7), 667-693. [doi:10.1080/01969720701534059]

RAFOLS, I., Porter, A. L. & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology*, 61(9), 871-1887 [doi:10.1002/asi.21368]

SCHARNHORST A. & Ebeling, W. (2005). Evolutionary Search Agents in Complex Landscapes. A New Model for the Role of Competence and Meta-competence (EVOLINO and other simulation tools). http://arxiv.org/abs/physics/0511232

SCHARNHORST A., Börner K. & Van den Besselaar, P. (eds.) (forthcoming). *Models of Science Dynamics*. Springer. [doi:10.1007/978-3-642-23068-4]

SCHUBERT, A., & Glanzel, W. (1984). A dynamic look at a class of skew distributions. A model with scientometric applications. *Scientometrics*, 6 (3), 149-167. [doi:10.1007/BF02016759]

SIMON, H. A. (1957) *Models of Man, Social and Rational* New York: Wiley.

SUN R. & Naveh, I. (2009). Cognitive simulation of academic science. In Proceedings of the International Joint Conference on Neural Networks, pp. 3011-3017. [doi:10.1109/ijcnn.2009.5178638]

THAGARD, P. (1993), Societies of minds: Science as distributed computing. *Studies in History and Philosophy of Science*, 24, 49-67. [doi:10.1016/0039-3681(93)90024-E]

THAGARD, P. (2000). *Coherence in thought and action*. Bradford Books, MIT Press.

WATTS, C. & Gilbert, N. (forthcoming). Does cumulative advantage affect collective learning in science? An agent-based simulation. *Scientometrics*. Special issue "Modeling Science: Studying the Structure and Dynamics of Science"

WEISBERG M. & Muldoon, R. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of Science*, 76(2), 225-252. [doi:10.1086/644786]