Petra Ahrweiler (2011)

Modelling Theory Communities in Science

Journal of Artificial Societies and Social Simulation **14** (4) 8
<http://jasss.soc.surrey.ac.uk/14/4/8.html>

Received: 28-Jun-2011 Accepted: 18-Jul-2011 Published: 31-Oct-2011

Abstract

This position paper presents a framework for modelling theory communities where theories interact as agents in a conceptual network. It starts with introducing the difficulties in integrating scientific theories by discussing some recent approaches, especially of structuralist theory of science. Theories might differ in reference, extension, scope, objectives, functions, architecture, language etc. To address these potential integration barriers, the paper employs a broad definition of "scientific theory", where a theory is a more or less complex description a describer puts forward in a context called science with the aim of making sense of the world. This definition opens up the agency dimension of theories: theories "do" something. They work on a — however ontologically interpreted — subject matter. They describe something, and most of them claim that their descriptions of this "something" are superior to those of others. For modelling purposes, the paper makes use of such description behaviour of scientific theories on two levels. The first is the level where theories describe the world in their terms. The second is a sub-case of the first: theories can of course describe the description behaviour of other theories concerning this world and compare with own description behaviour. From here, interaction and potential cooperation between theories could be potentially identified by each theory perspective individually. Generating inclusive theory communities and simulating their dynamics using an agent-based model means to implement theories as agents; to create an environment where the agents work as autonomous entities in a self-constituted universe of discourse; to observe what they do with this environment (they will try to apply their concepts, and instantiate their mechanisms of sense-making); and to let them mutually describe and analyse their behaviour and suggest areas for interaction. Some mechanisms for compatibility testing are discussed and the prototype of the model with preliminary applications is introduced.

Keywords:
Simulating Science, Theory Interaction, Agent-Based Modelling, Theory Network

Introduction

1.1 Scientific theory integration seems to imply the "impertinent plea to join others on the grounds of commonly interpreted reality" (Luebbe 1978: 125, own translation). As Steve Woolgar put it in a conference:

> The idea of cooperation or integration, … is that the objects of nature can be fixed by interdisciplinary triangulation. But this is the idea that there is a common object which *you* have discovered and which you now invite me and others to study by *our* methods. My own methods are placed in the service of studying "natural kinds" already identified by you, of working out someone else´s world picture (Nickles 1989: 226).

1.2 Theories seem to be less unwilling to work out somebody else's world picture in that sense, at least at first sight, than theorists. Therefore, the paper here addresses theories as agents. The final aim is to simulate interdisciplinary theory discourse in science where interaction is realised on the level of concepts. This might build some bridges for interaction between scientists who often refrain from cooperation due to in-commensurability or in-compatibility issues of their theoretical frameworks. Modelling theory communities might encourage scientists to follow the exemplary interaction and cooperation of their offspring and might help to integrate dispersed and fragmented scientific communities.

1.3 What is the integration problem regarding scientific theories? Why not put the pieces of disciplinary knowledge together like a puzzle in a cumulative way? Often, theories seem to relate to one another as "mutually supportive and, in the strongest cases, mutally necessary" (Nickles 1989: 229). However, looking at famous debates in science up to escalating ones such as the so-called Science Wars (Jardine and Frasca-Spada 1997), there definitely is a problem.

1.4 On a conceptual level, there are already several attempts of integrating scientific theories into a coherent picture. Logical Positivism with its programme of unifying science might be the best known and documented among them. In the eighties of the last century, Balzer, Moulines and Sneed (1987) offered with *An Architectonic for Science. The Structuralist Program* a theory network structure of all empirical sciences. To start with, the authors put forward their "essentially conservative and conventional views about the nature of knowledge" (Balzer, Moulines and Sneed 1987: xv): "this book is about the structure of knowledge. By 'knowledge' we mean 'reasoned, true belief'. This entails that knowledge consists of things like propositions or statements. By 'structure' we mean simply the way in which the truth values, or probability values of statements are interdependent" (Balzer, Moulines and Sneed 1987: xv). According to the authors, scientific, pre-scientific and non-scientific linguistic statements exist in an ordinal relation: just the first are incorporated as empirical knowledge in the theory architecture designed—explicitly excluded from this category are "normative political theory", and "the formative period of thermodynamics and contemporary sociological theory" (Balzer, Moulines and Sneed 1987: xix). Excluding these theory candidates for "ontological incommensurateness", the remaining theories of empirical sciences form a conceptual network where the relation between theories is established by so-called "inter-theoretical links" (relations of specialisation, theoricity, reduction and equivalence).

Worldviews, meaning and translation

2.1 What is with the theories *ante portas*, those which have not been granted access to the theory network? And even within this network: very often, there seems to be a missing mutual "connectivity" of disciplinary perspectives, a deep disagreement of theoretical orientations. More radically minded people even claim in-compatibility of all descriptions rejecting the possibility of all language-based understanding. This is certainly no route to go: rejecting the possibility of basic understanding and meaningful verbal intercourse is similar to the rejection of science.

2.2 However, just using the same language does not imply that there are no radical differences of perspectives possible verbalised in this idiom: Davidson has shown that we can describe very different things "using the equipment of a single language. Whorf, wanting to demonstrate that Hopi incorporates a metaphysics so alien to ours that Hopi and English cannot, as he puts it, 'be calibrated', uses English to convey the contents of sample Hopi sentences. Kuhn is brilliant at saying what things were like before the revolution using—what else?—our post-revolutionary idiom" (Davidson 1973/1974: 6).

2.3 Just using the same langugage certainly does not help with the problem how two theories with completely different concepts and logical claims could interact. In most examples of competing theories, neither explication of notions nor translation attempts can foster an approximation of different theoretical frameworks. Quite the opposite: trying to compare the descriptive frameworks of meaning which underly the respective concepts of theories, we as observers—if we are able to "switch" between theories—are forced to commit to a "drastic change in our worldview" (Balzer, Moulines and Sneed 1987: 318).

2.4 There are many statements concerning the problem of in-commensurability between theories (cf. Balzer 1985a and 1985b; see on the relation of commensurability, communicability and compatibility Kuhn 1983). Allowing participation for all theory candidates excluded in the structuralist theory network approach shows a huge variety to deal with. Current definitions of "theory" differ —not only between disciplines but also within. Just for the social sciences, there is a noticeable number in early theory systematics in Fijalkowski (1961) or Narr (1969). There are definitions of theory as "Philosophy of Weltanschauung" (Husserl 1965), the designation of theory as "attribute of society, not science" (Knorr-Cetina 1981), but also formal definitions like "a theory, T, is a structure (H,I) where H is a set of hypotheses and I is a relation in H called 'implication' or 'deducibility', so that H is weakly connected by I" (Galtung 1967: 451; cf. also Albert 1964). The various theory definitions cover statements differing in reference, extension, scope, objectives, functions, architecture, language etc. Not favouring one of the available definitions and excluding all the others, the broadest concept would be that a scientific theory is a more or less complex description a describer puts forward in a context called science with the aim of making sense of the world.

2.5 It is possible to start from such a broad understanding of theory for two reasons: The first is an endogeneous one. There is no position or definition available where completeness and self-sufficiency of a theory can be claimed: all definitions are open for critique of their "blind spots" (cf. Ahrweiler 2001). Theory definitions either rely more on empirical observations for generating scientific knowledge, and/or see logics and epistemics as basic. Whatever camp of these two, whatever interpretation of the terms in them, whatever combination of these two etc., this constitutes a fundamental openness of theories for "competing" alternatives with other choices. Theories might be under-determined by observation data[1] and/or suffer from the theory-ladenness of each observation: both has negative consequences for their completeness and self-sufficiency (cf. Carrier 1994: 1-19). If there are serious doubts about the self-sufficiency and completeness of any type of theory, there is no exempt position to exclude theory candidates from an attempt to model theory networks. The second reason for using a broad definition of "theory" is a pragmatic one connected to this, and has to do with curiousity. There might be some benefit in choosing this broad definition in the hope that there is something to learn from other theoretical perspectives, that there are contributions to our own insights from places we have not expected, by assuming that other perspectives can be generally considered as trustworthy and credible in their search for scientific knowledge. This would not prevent any critique of other positions and preferences for own ones but might offer new and relevant insights.

2.6 Letting the various theory types differ in reference, extension, scope, objectives, functions, architecture, language etc., means to accept that there is no "big translation" to integrate them into a theory network - even if sometimes "a philosopher who does not like to talk about ideas and propositions, often talks like a layman without care about the question of translation. He tends to take

the relation of a sentence and its translation as granted […]: a sentence is a translation of another if it expresses the same idea, the same thought, the same meaning, the same proposition" (Quine 1989: 61). To state translation functions convincingly for theories with different ontological, epistemological, architectural, extentional etc. scopes and objectives, would be hardly achievable. At best, this could be a replacement of one sentence by another; the exact relation between the two sentences would remain unclear. This also applies to the hopes around computer simulations aiming at the discovery of iso- or homomorphism of different theories by translating them into a common formal language. Introducing just another symbol system into the game hardly addresses the above mentioned fundamental translation problem.

Theories in interaction

3.1 How to preserve the characteristics of a theory in terms of reference, extension, and scientific claim? How to preserve different semantics and different ways to theorise? For modelling theory networks, the potentially radical diversity of theoretical frameworks has to be considered and left untouched, even though "communication" between in-compatible theory types needs to be enabled. Additionally, a systematic access for mutual completion and differentiation needs to be established for less problematic theory relations. Where to start? Is there any meaningful interpretation available that is productive for all theory types without "colonising" them by a translating super-theory? Is there any possibility to preserve all theory universes in their respective shape while offering an integration procedure acceptable for each position?

3.2 The suggestion here is to employ a socio-behavioural approach to theories (Ackoff 1957) based on their *agency*. Not imposing some postulated internal commonality on theories, a commonality can be used that derives from the broad definition above: a theory is a more or less complex description a describer puts forward in a context called science with the aim of making sense of the world. Theories "do" something. They work on a—however ontologically defined—subject matter. They describe something, and most of them claim that their descriptions of this "something" are superior to those of others (otherwise no scientific dispute).

3.3 This is why an ethno-methodological sociological approach to theory modelling is employed here. It uses descriptions on two levels. The first is the level where theories describe the world in their terms. The second is a sub-case of the first: theories describe the description behaviour of other theories concerning this world and compare with own description behaviour. Theories describe their mutual behaviours for suggesting ways of potential interaction followed from these descriptions. Thereby, it is not important how they define their subject, e.g. as "reality" or "socially constructed"; whether they aim at understanding, interpreting, explaining, predicting etc.; or what status they claim for their sense-making, be it causation, correlation or anything else. Everything is fine. They just apply their own descriptive concepts and their own mechanisms of sense-making to the subject under consideration, and judge from here about the quality of other theories trying the same.

3.4 This is looking at theories in the same way as at first-time-to-meet human tribes in far-away countries: they cannot talk to each other, maybe they cannot even understand the ways of others to think about the world even if they would speak the lingo, but they can describe what the others do, compare with their own ways, and in doing so learn about them, and maybe suggest cooperation. This approach complies to a species-approach to theories: species that act and interact in an environment (cf. evolutionary models of knowledge, which deal with terms, concepts, or sentences as units of analysis (Toulmin 1972, Blackmore 1999, Gilbert 1997). However, this approach also needs language to distinguish the different semantics, and the different ontological and epistemological claims of theories: this is to ensure that theories can stay in their own universe using their specific linguistic interpretation of objects, and their chosen reference between language and object. Therefore, both are needed: understanding language and observing behaviour.

Social simulation of theory communities

4.1 Generating inclusive theory communities and simulating their dynamics means

- To implement theories as agents

  There are many examples for computerised logical reconstructions of theories modelling conceptual architectures and their inference systems (Hanneman 1988). These models allow for testing completeness and consistency of theories. For modelling inclusive theory networks, the tricky challenges concern the difference-creating aspects of object formation, the different status of concepts, and the mechanisms for relating concepts (sense-making, interpreting, reasoning, inferring, deducing, abducing, explaining etc.). To help with this, many logical tools are available up to axiomatising whole theory structures (McDermott and Doyle 1980, Moore 1985, Gelfond 1987, Lifschitz 1989).

- To create an environment where the agents work as autonomous entities in a self-constituted universe of discourse.

  Here, the theory agents need to apply their concepts, and instantiate their mechanisms how to relate those concepts in attempts to explain, understand, make sense of the world etc. The workspace has to be something the theories are normally concerned with, a something to theorise about, an "issue". This requires to describe a theory's application area informally allowing the theories to make their choice of subject matters according to their own conceptual criteria using the "method of autodetermination", by which the theory´s formalism itself is left with the "responsibility" of deciding, in many cases, what its application is to be" (Balzer, Moulines and Sneed 1987: 39). To do this, all theories must speak

every-day language (see above section 1, paragraph 1 about the requirement that communication is possible). Here is the place to use ontologies, thesauri, parsing tools, syntax trees and lexica provided by computer linguistics: this does not impose a super-theory on the agents, it is their normal type of environment which they have to deal with. Interesting is how they handle and interpret it.

- To observe what they do with this environment (they will try to apply their concepts, and instantiate their mechanisms of sense-making)
- To let them mutually describe and analyse their behaviour and suggest areas for interaction

The fourth point now is not so easy. This procedure is about theories as *describers and interpreters of other theories*. The resulting interpretations are *uni-laterally presumed* by the interpreting theory, and approved neither by the interpreted theory itself nor by a binding external benchmark.

4.2 Taking seriously the discrepancies between theory types suggests at least two mechanisms for generating theory communities based on the theories' descriptions of others' behaviours: the first mechanism called "strong theory integration" leads to a conceptual network. Here, theories "decide" on potential relationships for inclusion, complementation, substitution and differentiation. For in-commensurable theories, the second mechanism called "weak theory integration" produces networks of loose coupling, which allow the external observer to switch between the different alternatives of the in-compatible theories via boundary concepts and their theory-specific reference structures.

4.3 The second mechanism, applied to two competing theories, will identify the units of the "issue" where the two theories have claimed to be competent by successfully applying their concepts, but where none could come up with a clever application of the first mechanism. This means there is a subject area, where the two theories will certainly raise a dispute but where the external observer would have the opportunity of a controlled switch between theory architectures and reference structures. The theories —like two dogs fighting for the same bone - talk about the same but offer completely different interpretations.

4.4 The first mechanism is more difficult. How can the interpreting theory decide whether some foreign behaviour is compatible? Would not that again super-impose a set of criteria not every theory(ist) would agree to share? There is again advice from ethno-methodology, which identified the most general cognitive categories as smallest common denominator for basic understanding of what others are doing (Lakoff, 1987, "Women, Fire, and dangerous Things. What Categories reveal about the Mind"). Our theories are allowed to use (i) identity: this behaviour relation is the same as mine; (ii) contradiction: this behaviour relation is the opposite of mine; (iii) container: this behaviour relation is part of mine, or mine is part of this. For the concept relations (sense-making, interpreting the world) theories can jointly use (iv) co-location: this concept always appears together with / before / after.

4.5 Here, it becomes immediately obvious, why an agency approach is useful, and has to be combined with a language-based approach. For example, sometimes it is hard to judge whether something is the same. The word "bank" e.g. might be identified as the same word on the surface by an economics theory and by an interior design theory. However, looking at their intepretation behaviour, the interpreting theory would quickly notice from the behaviour of the other theory around this word, from the "path" of the interpreted theory, its past behaviour around this word, and its behaviour concerning neighbouring areas etc. that the assumed identity on the surface level of words has been misleading. Theories might use same words but might mean different things. In this approach here, the word "bank" would become a boundary element for the second mechanism. Vice versa, the advantage of the behavioural approach also applies to the situation of a contradiction on the surface, i.e. where theories seem to offer radical alternatives. However, it is perfectly possible that they use the same definitorial and reasoning structures to describe something: they only come up with a different final concept to express this. The behavioural pattern and the working history, however, would reveal the large areas of similarity, and might point to cooperation possibilities. This option reminds of the structuralist approach concerning "empirical equivalent" theories, i.e. theories not sharing a common conceptual structure, but offering identical explanations for the same phenomena (Balzer, Moulines and Sneed 1987: 290).

4.6 In the model implementing this approach here, compatibility testing is performed by each agent for any other theory bi-laterally, which means that two theories can come up with different results concerning compatibility with each other. The interpreting theory conducts structural comparisons with the patterns of a foreign theory, on the surface level of single concepts and concept relations, which might pass or might not pass the test at a first glance, and on the deeper level of relating foreign concepts and interpretations to the broader "working path" of a foreign theory as demonstrated in past and present behaviour concerning the subject matter, which might confirm or change the test results from the surface level. In this extensive area of bi-laterally organised compatibility testing of behaviour, the agents follow the principle of "fertility" (they check whether there is a big enough common set of areas covered by both theories), prefer if possible their own interpretations ("egoism" principle), and have a potential for "charity" (i.e. they are open for the interpretations of others: they allow others to look at and use their concepts and structures; and they credit others with assuming sensible interpretations, i.e. that in applying others' procedures/worldviews, indeed same results would follow).

4.7 A simple prototype of the model has been implemented in TCL so far for two case studies. The first is a simulated theory debate in Science and Technology Studies (STS) between Karin Knorr-Cetina (1981) "The Manufacture of Knowledge. An Essay on the Constructivist and Contextual Nature of Science" for Sociology of Science, and Philip Kitcher (1993) "The Advancement of Science. Science without Legend, Objectivity without Illusions" for Philosophy of Science. In the model, both theories work on a fictive story about the formation of a new discipline. The model finds many boundary elements of course, because these theories are competing theories on the subject of scientific community formation. However, the model also comes up with some

suggestions for cooperation. For example, both theories work on an area of the issue dealing with the usage of experimental methods in a new discipline. They apply the same concept relations here, i.e. that scientists tend to apply methods to a new problem which have already proven useful in a similar situation. However, the Knorr agent can say a little bit more about this than the Kitcher agent, because it has some concepts about how scientists identify similarity. While Kitcher just says that similarity of situation would be "evident", the Knorr agent offers some criteria for similarity and differentiates therefore the Kitcher concept of evidence.

4.8 The second example is a theoretical debate in a case study of Innovation Research between an incremental and a radical perspective on innovation in the European Car Industry (see Ahrweiler 2001). The two theory agents work on an empirical study provided by Harbour et al. (1998). The model again suggests points for cooperation where a principally incremental evolution of the European car distribution system would be complemented by some radical innovations in segments (new dominant players, new production technologies, new production, consumption and distribution options via e-commerce). This application of the theory discourse model motivated Arthur D. Little International Inc. to test a prototype of the model for mediating between different positions in some client projects. A full paper will introduce the model in detail and report the results of these two case studies.

4.9 It would be most interesting to know whether this model only works for theories or as well for theorists. What if scientists would constantly and closely observe each other in a gigantic field study how they do their daily job using and developing their theoretical frameworks? What if they would be engaged in compatibility testing as described above looking for cooperation options based on their interpretations of each others' behaviour? Would this be a step towards a huge trans-disciplinary scientific community? Or would they stay apart finding other ways to distinguish the self from the foreign?

Notes

[1]If it is possible to construct two or more theories, which are incompatible relying on the same set of experimental data, the choice between these theories cannot depend on “empirical facts” (Quine 1975).

References

ACKOFF, R.: (1957): Toward a Behavioral Theory of Communication. *Management Science* 4, S. 218-234.

AHRWEILER, P. (2001): *Informationstechnik und Kommunikationsmanagement. Netzwerksimulation fuer die Wissenschafts- und Technikforschung*. Campus: Frankfurt.

ALBERT, H. (1964): Probleme der Theoriebildung. In: H. Albert (Hrsg.): *Theorie und Realitaet*. Ausgewaehlte Aufsaetze zur Wissenschaftslehre der Sozialwissenschaften. Tuebingen: Mohr, pp. 3-70.

BALZER, W. (1985a): Incommensurability, Reduction, and Translation. *Erkenntnis* 23, pp. 254-267.

BALZER, W. (1985b): Was ist Inkommensurabilität? *Kant Studien* 76, pp. 196-213.

BALZER, W., C.U. Moulines and J.D. Sneed (1987): *An Architectonic for Science: The structuralist Program*. Dordrecht etc.: Reidel.

BLACKMORE, S. (1999): *The Meme machine*. Oxford: Oxford University Press.

CARRIER, M. (1994): *The Completeness of scientific Theories. On the Derivation of empirical Indicators within a theoretical Framework: The Case of Physical Geometry*. Dordrecht etc.: Kluwer.

DAVIDSON, D. (1973/1974): On the Very Idea of a conceptual Scheme. *Proceedings of the American Philosophical Association* 47, pp. 5-20.

FIJALKOWSKI, J. (1961): Ueber einige Theorie-Begriffe in der deutschen Soziologie der Gegenwart. *KZfSS* 13, pp. 88-109.

GALTUNG, J. (1967): *Theory and Methods of Social Research*. Oslo: Univ.-Forl. etc.

GELFOND, M. (1987): On stratified autoepistemic theories. *AAAI* 87, 1, pp. 207-211.

GILBERT, N. (1997): A Simulation of the Structure of Academic Science. *Sociological Research Online*, vol.2, no.2, 3 http://www.socresonline.org.uk/2/2/3.html.

HANNEMAN, R. A. (1988): Com*puter-assisted Theory Building. Modeling dynamic social Systems*. London etc.: Sage.

HARBOUR, M., J. Brown, P. Wade and the ICDP Research Team (1998): Future Directions for European Car Distribution: Evolution or Revolution? Research Paper 11/98, International Car Distribution Programme Ltd, Solihull (United Kingdom).

HUSSERL, E. (1965): *Philosophie als strenge Wissenschaft*. Ed. by W. Szilasi. Frankfurt/M.: Klostermann.

JARDINE, N. and M. Frasca-Spada (1997): Splendors and Miseries of Science Wars. *Studies in the History and Philosophy of Science* 28, Sp 219-235.

KITCHER, P. (1993): *The advancement of science: Science without legend, objectivity without illusions*. New York etc.: Oxford University Press.

KNORR-CETINA, K. (1981): *The manufacture of knowledge. An essay on the constructivist and contextual nature of science*. Oxford etc.: Pergamon Press.

KUHN, T.S. (1983): Commensurability, Comparability, Communicability. In: P. Asquith and T. Nickles (eds.): Philosophy of Science Association (PSA) 1982. East Lansing.

LIFSCHITZ, V. (1989): Between Circumscription and autoepistemic Logic. In: R.J. Brachman (ed.): *International Conference on Principles of Knowledge Representation and Reasoning*. San Mateo, CAL: Morgan Kaufmann, pp. 235-244.

LUEBBE, H. (1978): Pragmatismus oder die Kunst der Diskursbegrenzung. In: W. Oelmueller (ed.): *Normenbegruendung. Normendurchsetzung*. Paderborn: UTB Schoeningh: 118-125.

MCDERMOTT, D. and J. Doyle (1980): Non-Monotonic Logic I. *Artificial Intelligence* 13. pp. 41-72. [doi:10.1016/0004-3702(80)90012-0]

MOORE, R. (1985): Semantical Considerations on Nonmonotonic Logic. *Artificial Intelligence* 25, pp. 75-94. [doi:10.1016/0004-3702(85)90042-6]

NARR, W.-D. (1969): *Theoriebegriffe und Systemtheorie*. Stuttgart etc.: Kohlhammer.

NICKLES, T. (1989): Integrating the Science Studies Disciplines. In: S. Fuller, M. de Mey, T. Shinn and S. Woolgar (eds.): *The Cognitive Turn. Sociological and Psychological Perspectives on Science*. Dordrecht, Boston and London: Kluwer, pp. 225-256.

QUINE, W. V.O. (1989): *Die Wurzeln der Referenz*. Frankfurt/M.: Suhrkamp.

TOULMIN, S. (1972): *Human Understanding*. Volume I. General Introduction and Part I: The collective Use and Evolution of Concepts. Oxford: Clarendon.