Miguel, Amblard, Barceló & Madella (eds.) Advances in Computational Social Science and Social Simulation
Barcelona: Autònoma University of Barcelona, 2014, DDD repository <http://ddd.uab.cat/record/125597>

Social Scientists, Qualitative Data, and Agent-Based Modeling

Roman Seidl
ETH Zurich, Natural and Social Science Interface (NSSI), Universitaetstrasse 22, CH-8092 Zurich, Switzerland
Email: roman.seidl@env.ethz.ch

***Abstract*— Empirical data obtained with social science methods can be useful for informing agent-based models, for instance, to fix the profile of heterogeneous agents or to specify behavioral rules. For the latter in particular, qualitative methods that investigate the details of individual decision processes are an option. In this paper, I highlight the challenges for social scientists who investigate social/psychological phenomena but at the same time have to consider the properties of agent-based simulation. To illustrate these challenges and potential solutions, I present four examples in which qualitative data is acquired for subsequent use in agent-based simulations and discuss the examples in terms of the challenges.**

I. Introduction

THE functions of qualitative data for modeling and simulations vary [1]. [2] evaluated different empirical methods and distinguished among sample surveys, participant observation, experiments (field and laboratory), companion modeling, and GIS and remotely sensed spatial data. Thus, various methods are available, each more or less suitable for different functions. In this presentation, I ask what functions of (qualitative) empirical data can be identified in modeling projects and give examples from my research. I draw on my experience and other literature and propose four different functions of qualitative data in modeling and simulation. There may be other functions, such as extraction of causalities between environmental triggers and actor behavior, which are not discussed here.

For instance, with *explorative* investigations we want to identify where to look deeper with other qualitative or more quantitative measures (e.g., surveys, databases, and statistics). Another function is to distill details of the rules of agent interaction or decision making [3, 4]. A third function is validation, for instance, of a model's plausibility or its simulation results [5]. Fourth, for simulations of future system behavior, one often has to rely on scenario data [6]. This can be more quantitative but often is qualitative, denoting a relative increase or decrease in exogenous variables.

Acquiring the necessary data looks different in different cases [7], and challenges await the social scientist who usually is trained not in "feeding" agent-based models (ABM) but in conducting disciplinary state-of-the-art research. This issue is important if one considers the increasing interdisciplinarity of many research projects, where researchers from different disciplines often follow different rationales and apply different methods. In the remainder of this extended abstract, I present challenges for social science research in dealing with agent-based models and simulation, followed by examples of specific functions of qualitative methods and data for agent-based models, and last, discuss the relation between the challenges and these functions.

II. Challenges for the Social Researcher

In this section, I briefly discuss challenges for the social researcher; other disciplines may face different challenges.

One problem with including stakeholders for data acquisition is the potential for "stakeholder bias." The question regarding qualitative data obtained from a low number of participants is, how representative is the data informed by individuals or small groups? This means that models and thus the simulation results may be valid only for the specific case and not easily generalizable. This may not be perceived as a problem if the model is meant as a means for social learning [8]. If the model developed is meant to be a more generic type, then the available empirical data may have to be generalized [9].

A second challenge is translating empirical data into agent profiles or rules[1]. Some authors share procedures for this implementation step [10]. For instance, [3] reports on semi-structured personal interviews with several stakeholders from business (e-commerce), which were recorded and transcribed. The most important quotes were then related to agent properties, such as "technical competence" with a range of integers from 1 to 3. Other statements were interpreted as an exponential function (for the variable demand).

Third, another issue is translating social science models or concepts into ABM rationales. Social science models usually—if at all—are tested or corroborated with

[1] Moreover, a potential challenging issue is if not all data is available and model implementation has to be based on assumptions – an unusual situation for social scientists.

quantitative studies or experiments but seldom in a *dynamic* way. Social scientists often do not think in "simulation dynamics." From this tradition, it is not self-evident how to implement variables and temporal dynamics into ABM. For instance, the theory of planned behavior, including the conceptual model proposed by [11], has been applied in different multi-agent systems, for example, by [12]. However, the challenge is to consider the temporal aspects not explicit in this static (or at least short-term) concept. Here, methods have to be applied that investigate the underlying processes (e.g., longitudinal survey or experiments).

Fourth, regarding the dynamics of real-world phenomena, qualitative data may provide insights into single actors' intentions, rationales, and behaviors. However, to capture the underlying processes of behavior variability under different environmental conditions, the question is how to access often *implicit knowledge and information processes* that are not consciously available to the individual? Selected methods may help elicit such knowledge, such as participatory interviews [13, 14].

In more general terms, these challenges force social scientists to adapt their methods to comply with the needs for agent-based systems and simulation. In the following, I illustrate four issues by referring to the use of qualitative methods in past and current modeling projects. I then relate these issues to the challenges I discussed in the introduction.

III. DIFFERENT FUNCTIONS ILLUSTRATED BY FOUR CASES

A. Explorative function

The first example illustrates the explorative function, where to look more thoroughly with quantitative measures (survey or statistical data). In a large-scale interdisciplinary project on global change and the water cycle, a risk perception module was to be implemented in an agent-based system so that agents could respond to water-related risks [16, 17]. The first qualitative investigation used a visualization of the participants' mental models during semi-structured interviews. Results showed that participants were in principle aware of slowly evolving future risks regarding drinking water availability and quality on the one hand and sudden risks from flooding due to climate change on the other. However, participants did not relate these processes to their daily routines, their near future, or their own (spatial) situation.

Thus, the main message of these qualitative interviews was to focus on two risks: high water and drought (because water quality, although a concern of the participants, was not covered by the models in the project). Another outcome was that we assumed personal avoidance techniques, because participants admitted that currently they do not think about these risks and only after incidents may the relevance may increase. We thus assumed that a decay function of agents' alertness is appropriate. In a subsequent quantitative survey, we focused on basic awareness of water-related climate risks and avoidance. Both issues were then implemented in the agents' profiles. The explorative interviews were not perfectly planned to inform agent rules but inspired further social science methods.

B. Details of rules (agent interaction, decision making)

A single cross-section survey is not appropriate for informing individuals about a dynamic phenomenon such as changes in public opinion. A longitudinal survey is better, but cannot help in terms of the micro-processes of opinion formation and change. In this case, socio-psychological experiments are more suitable. However, experiments can be conducted more or less artificially and controlled. For instance, we currently face the trade-off of a more restricted but controlled experiment (interaction of a participant with computer-aided portrayals of arguments on a topic for which the valence and importance are varied) versus a more natural interaction between participants and an instructed person. To distill the rules of changes of argument (and thus opinion) for the agent-based simulation, a careful analysis of the real interaction would be sufficient, since we seek prototypical behavior instead of specific cases. However, for the social-psychological researcher, an experiment based on computer interaction is more convenient, controlled, and precise, and thus more easily published in high-ranking socio-psychological journals. Therefore: *How much emphasis should be placed on model development and simulation at the expense of the experimental setting*?

C. Validation function

As [1] observed, validation and model construction should be seen as a joint process rather than different stages or processes. However, expert judgments and stakeholder interviews regarding model results and project outcomes are seen as potential means for validation. Often, the emerging macro-patterns are considered more or less reasonable [18].

In a current transdisciplinary case study (http://www.tdlab.usys.ethz.ch/casestudy/cs_actual), we evaluate a major interdisciplinary project on land-use that also implemented stakeholder knowledge in a set of coupled models [19]. We investigate what impact the project had and has among local stakeholders in one of the study regions (some participated in the project; others did not).

Aggregated results from agent-based models may be discussed as well as if the implemented rules are valid. Preliminary results indicate that it strongly depends on the applied perspective: Simulated land-use by farmer agents can make sense from a farmer's perspective, but forestry stakeholders have a different perception of land-use. This illustrates that rules implemented in a data-driven model are case specific and may be constrained and prototypical, particularly when the data has been obtained from stakeholder input. Validation of the model results and project outcomes again depends on the background of the stakeholders asked.

D. Scenarios function

The project mentioned in section *C* includes scenario workshops with stakeholders in the study region. We developed multi-scale scenarios (from the local to the global scale) [20] that provide the frame for the land-use change agent-based model (thus far the implementation of the ABM uses broad global scenarios only; see http://www.openabm.org/model/2870/version/2/view).

We applied formative scenario analysis, a structured technique that integrates knowledge from a wide range of sources such as literature, statistics, and stakeholder workshops, to arrive at a coherent and robust set of scenarios. This technique combines qualitative and quantitative approaches (using specific software: http://www.systaim.ch/). We realized that if we relied on stakeholders' knowledge and judgments alone we would obtain a biased picture, depending on the specific group of stakeholders and the most important impact factors uttered. It is all the more important to complement this type of data with scientific literature and other informants for a more generic picture that is not limited to the details of the case.

A critical issue, though, remains, regarding scenarios of long-term developments of social and environmental conditions. They cannot be forecast exactly, and future individuals' behavior and decisions are not fully predictable based on current individuals' interview data. For instance, qualitative stakeholder judgments in workshops on future landscapes and land-use to inform a land-use ABM remain limited because it is based on current preferences and people are blind to changing norms as the cognitive shifting baselines show [15]. This problem is difficult to overcome even with methodological insight.

IV. Discussion

In this section I, will discuss the previous examples and relate them to the challenges I mentioned in the introduction (see Table I).

TABLE I.
FUNCTIONS OF EMPIRICAL QUALITATIVE DATA AND RELATED CHALLENGES

Functions	Challenges
Explorative	Translation into ABM rationale; usability (type of data)
Details of agent rules	Translation of socio-psychological models of behavior (and considering the temporal dynamics); translating implicit knowledge.
Validation	Stakeholder-bias
Scenarios	Stakeholder-bias vs. detailed anthropological; constraints on complexity

Exploring potential key variables for subsequent investigations and agent-based models is a valuable function, especially for qualitative methods. However, the challenge is translating these variables appropriately into agent properties. The explorative function may or may not yield relevant results for the model. Moreover, the results may be preliminary, sufficient for adjusting the social scientists' focus to specific phenomena but not readily usable for the agent-based approach. In addition, the data may hint at issues that cannot be covered by the models applied. For instance, even if water quality is emphasized as an important issue during interviews (and in the literature), if the models' aim is different or implementation is just too complex, one cannot pursue this issue within this project.

To investigate the micro-processes that inform *agents' rules*, an experimental setting may be appropriate, because one can focus on specific processes in controlled settings. However, due to these controlled situations, the experiments are artificial and constrained to narrowly defined phenomena. Thus, the precision is higher than necessary for many multi-agent simulations. Therefore, as in our example, the trade-off is between socio-psychological excellence (artificial situation and control) and application relevance (e.g., focus group discussion or interacting dyad).

Validating the model concept and/or simulation results is part of the modeler's responsibility during the modeling process. However, as illustrated by the case of evaluating stakeholders' judgments, this evaluation has to be critically reflected. Stakeholder bias is a challenge here, too. Another challenge is that evaluation usually targets the macro or aggregate level of agent simulations (whether expected patterns emerge). Social scientists such as psychologists, instead, tend to be interested in phenomena and processes on the micro level, and address questions of representativeness of the results for individuals.

Developing scenarios with stakeholders or based on the literature forces the analyst to identify the most important impact factors and their future states. The question arising is what is the limiting factor: the complexity of the topic, and thus the number of impact factors and future states, or the number of parameters in the model? To arrive at a coherent set of scenarios, the interaction between the impact factors must be considered, whether they are conflicting, neutral, or enhancing. However, the degree of detail is usually limited by practical constraints. Here, a trade-off similar to the one identified in case *B* (details of rules) may appear. A scenario analysis focusing only on the properties of the model may fall short in relevance as a standalone contribution. An idiosyncratic, anthropological perspective on the case study area at hand and its specific conditions may be extremely interesting for social researchers and an idiographic approach recommended from a qualitative social-science perspective. However, from a modeling perspective, one cannot implement every detail, and each model will be limited in terms of the phenomena it can address.

The problem of potential stakeholder bias is obvious in the *scenario* function described for the formative scenario analysis but also affects the *validation* function. A countermeasure is a thorough stakeholder analysis and a representative group of participants (actors) from different sectors to cover different factors [21]. However, in reality, a "convenience sample" is often used, because of the time constraints of researchers and stakeholders alike. In our case, for instance, we missed representatives of the tourism business (e.g., hotel owners), because they were not interested in and were unavailable for our workshops.

The type of data needed for agent-based models typically differs from data gathered in social-science research; thus, sometimes the methods differ or are applied differently. Researchers using modeling are more interested in process data (e.g., behavioral rules) or time-series data (of long-term environmental and social variables). In this presentation, I highlighted examples social scientists (primarily social psychologists) may use to serve functions in agent-based modeling and simulation. As shown, there are challenges ahead, and compromises have to be made and trade-offs solved. A deeper discussion of these issues may yield interesting insights and reveal additional challenges, relevant for researchers of other disciplines. Moreover, reflections by researchers from other disciplines should complement the perspective given in this presentation.

REFERENCES

1. R. Boero and F. Squazzoni, *Does Empirical Embeddedness Matter? Methodological Issues on Agent-Based Models for Analytical Social Science.* Journal of Artificial Societies and Social Simulation, 2005. **8**(4): p. 6.
2. D.T. Robinson, et al., *Comparison of empirical methods for building agent-based models in land use science.* Journal of Land Use Science, 2007. **2**(1): p. 31-55.
3. R.I. Taylor, *Agent-based Modelling Incorporating Qualitative and Quantitative Methods: A Case Study Investigating the Impact of E-commerce Upont the Value Chain*, 2003, Manchester Metropolitan University.
4. S. Moss, *Critical incident management: An empirically derived computational model.* Journal of Artificial Societies and Social Simulation, 1998. **1**(4): p. 1.
5. P. Windrum, G. Fagiolo, and A. Moneta, *Empirical validation of agent-based models: Alternatives and prospects.* Journal of Artificial Societies and Social Simulation, 2007. **10**(2): p. 8.
6. A. Soboll, et al., *Integrated regional modelling and scenario development to evaluate future water demand under global change conditions.* Mitigation and Adaptation Strategies for Global Change, 2011. **16**(4): p. 477-498.
7. A. Smajgl and O. Barreteau, *Empiricism and Agent-Based Modelling*, in *Empirical Agent-Based Modelling. Challenges and Solutions*, A. Smajgl and O. Barreteau, Editors. 2014, Springer: New York. p. 1-26.
8. N. Ruankaew, et al., *Companion modelling for integrated renewable resource management: a new collaborative approach to create common values for sustainable development.* International Journal of Sustainable Development and World Ecology, 2010. **17**(1): p. 15-23.
9. T.A. Schenk, *Using stakeholders' narratives to build an agent-based simulation of a political process.* Simulation, 2014. **90**(1): p. 85-102.
10. E. Norling and F.E. Ritter. *Towards supporting psychologically plausible variability in Agent-Based human modelling*. in *Autonomous Agents and Multiagent Systems, 2004. AAMAS 2004. Proceedings of the Third International Joint Conference on*. 2004. IEEE.
11. I. Ajzen, *The theory of planned behavior.* Organizational Behavior and Human Decision Processes, 1991. **50**: p. 179-211.
12. N. Schwarz and A. Ernst, *Agent-based modeling of the diffusion of environmental innovations -- An empirical approach.* Technological Forecasting and Social Change, 2009. **76**(4): p. 497-511.
13. S. Bharwani, et al., *Multi-agent modelling of climate outlooks and food security on a community garden scheme in Limpopo, South Africa.* Philosophical Transactions of the Royal Society B-Biological Sciences, 2005. **360**(1463): p. 2183-2194.
14. P. Angelstam, et al., *Solving Problems in Social-Ecological Systems: Definition, Practice and Barriers of Transdisciplinary Research.* Ambio, 2013. **42**(2): p. 254-265.
15. S. Papworth, J. Rist, L. Coad, and E. Milner-Gulland, *Evidence for shifting baseline syndrome in conservation.* Conservation Letters, 2009. **2**(2): p. 93-100.
16. R. Seidl, *Eine Multi-Agentensimulation der Wahrnehmung wasserbezogener Klimarisiken [A Multi-Agent Simulation of the Perception of Water related Climate Change Risks]*. 2009, Marburg: Metropolis.
17. A. Ernst, *Using spatially explicit marketing data to build social simulations*, in *Empirical Agent-Based Modeling. Challenges and Solutions* A. Smajgl and O. Barreteau, Editors. 2014, Springer: New York. p. 85 - 103.
18. J.D.A. Millington, D. Demeritt, and R. Romero-Calcerrada, *Participatory evaluation of agent-based land-use models.* Journal of Land Use Science, 2011. **6**(2-3): p. 195-210.
19. R. Huber, et al., *Modeling Social-Ecological Feedback Effects in the Implementation of Payments for Environmental Services in Pasture-Woodlands.* Ecology and Society, 2013. **18**(2).
20. F.S. Brand, R. Seidl, Q.B. Le, J.M. Brändle, and R.W. Scholz, *Constructing Consistent Multiscale Scenarios by Transdisciplinary Processes: the Case of Mountain Regions Facing Global Change.* Ecology and Society, 2013. **18**(2).
21. K. Kok, D.S. Rothman, and M. Patel, *Multi-scale narratives from an IA perspective: Part I. European and Mediterranean scenario development.* Futures, 2006. **38**(3): p. 261-284.