Abstract

This paper investigates the contribution of evidence-based modelling to grounded theory (GT). It is argued that evidence-based modelling provides additional sources to truly arrive at a theory through the inductive process of a Grounded Theory approach. This is shown by two examples. One example concerns the development of software ontologies of criminal organisations. The other example is a simulation model of escalation of ethno-nationalist conflicts. The first example concerns early to middle stages of the research process. The development of an ontology provides a tool for the process of theoretical coding in a GT approach. The second example shows stylised facts resulting from a simulation model of the escalation of ethno-nationalist conflicts in the former Yugoslavia. These reveal mechanisms of nationalist radicalisation. This provides additional credibility for the claim that evidence-based modelling assists to inductively generate a theory in a GT approach.

Keywords:
Grounded Theory, Evidence Based Modelling, Theoretical Coding, Ontologies, Stylized Facts, Theory Development

Introduction

1.1 In recent years, research in computational social science has evolved in a way that indicates a certain stage of scientific maturation. While an initial phase has been characterised by an experimental stance to explore the possibilities of a new methodological tool (Deffuant et al. 2006), attention has now shifted to questions of empirical credibility of simulation models (Lorscheid et al. 2012). The purpose of this paper is to contribute to the research area of a cross-fertilisation between simulation and the standard methods of empirical social research (Squazzoni 2012). However, while Lorscheid et al. (2012) draw on the design of experiments of the classical quantitative approach in order to enhance the credibility of the analysis of simulation models, this paper aims at exploring additional sources of empirical credibility, which evidence-based modelling approaches can provide to the qualitative account of grounded theory (GT).

1.2 GT is an inductive process of qualitative social research. It is often questioned whether or not such research might generate theories. The main thesis of this paper is that the use of simulation models in an evidence-based modelling approach contributes to arrival at a theory within a GT framework. For this purpose, the paper will focus on two main arguments: first, it will be shown that the research process of evidence-based modelling shares a number of parallels with the GT approach. This is a rather unproblematic thesis. However, the parallels can be strengthened even further if experiences and guidelines of GT approaches are taken into account explicitly in a research process of evidence-based modelling. This thesis addresses primarily ICT specialists working in the field of evidence-based modelling. Following the research programme outlined by Lorscheid et al. (2012), it should contribute to an integration of evidence-based modelling into the canonical methods of empirical social research. Second, a more surprising thesis may be that evidence-based simulation provides methodological tools to strengthen the theoretical element of a GT. The objective of this argument is to inform specialists in the field of GT about the possibilities of a methodological cross-fertilisation between GT and evidence-based simulation; that is, to illustrate that GT can benefit from utilising simulation models—in particular, their formal precision and explicit representation of social dynamics—in its research process, as opposed to merely adding another tool to the toolbox of empirical research methods.

1.3 The paper is organised as follows: firstly, an overview of element of theory is provided. This allows the assessment of whether or not GT research suggests a theory. Secondly, an overview of evidence-based modelling is provided. It follows an overview of the methodology of the GT approach. Thirdly, a comprehensive discussion of the contribution of simulation tools and evidence-based modelling to the GT approach is provided. The discussion builds on two examples. The first example, which concerns the study of
organised crime, discusses how so-called theoretical coding benefits from knowledge representation in software ontologies. The second example is an evidence-based model of nationalist radicalisation, which demonstrates how simulation results provide additional explanatory power to a GT. Particular emphasis is put on the notion of stylised facts. Finally, the paper ends with concluding remarks.

Elements of a theory

2.1 It often remains precarious whether any qualitative empirical study is merely a description of a certain phenomenon, or if it can be claimed as truly a theory (Corbin & Strauss 2008, Strübing 2004). Within a qualitative approach, developing a theory might not even be the aim of research: 'In fact, theory development these days seems to have fallen out of fashion, being replaced by description of "lived experience" and "narrative stories"' (Corbin & Strauss 2008, p. 55). This raises the question of what, if it exists at all, is a theory in a GT approach. Strauss and Corbin (1998) describe their notion of theory as follows: 'For us, theory denotes a set of well-developed categories ... that are systematically related through statements of relationship to form a theoretical framework that explains some relevant ... phenomenon.' (Strauss & Corbin 1998, p. 22). To this end, 'final integration is necessary. Without it, there might be some interesting descriptions ... but no theory' (Strauss & Corbin 1998, p. 155). However, the notion of integration needs further clarification. It remains ambiguous insofar as such integration achieves a generalisation from the actual data, or remains at a level of an 'organization of data in discrete categories' (Corbin & Strauss 2008, p. 54). Corbin and Strauss (2008, p. 56) discuss this issue by using the example of an examination of gay disclosure / nondisclosure of information about their sexual identity to physicians, which might be expanded to a more general theory of information management, which encompasses a certain ideal type of human interaction (Weber 1968). Thus an explanation involves some kind of generality. This can be identified as the first element that a theory has to fulfil:

a) **Generality.** The object of a theory needs to be more than the description of a single phenomenon. Rather, a theory should explain a certain set of phenomena. In order to not be overly restrictive, we will also include middle-range theories of a limited set of phenomena.

2.2 However, this first element of theory characterisation already implies a further condition: namely, that a theory should explain something. This refers to the question of what an explanation actually is. In the social sciences and particularly in qualitative research, typological classifications, such as the concept of an ideal type, have a prominent place in scientific research. Typologies have some kind of generality, insofar as they allow the subsumption of individual cases into a certain category. However, it remains contested whether a typological categorisation is sufficient for an explanation of a phenomenon (Hedström 2005). As highlighted also by Strauss and Corbin (1998) that a theory should explain a certain phenomenon, a further element of theory is that something (i.e., the explanans), explains something else (i.e., the explanandum). Thus the aim of theory development in social science is to identify social processes from subjective experiences, rather than only describing subjective experiences. To provide an explanans for an explanandum can be characterised as a further core element of a theory:

b) **Explanans and explanandum.** A phenomenon X (the explanans) should be identified that explains a different phenomenon Y (the explanandum), which is the subject of scientific inquiry.

2.3 However, identifying explanans and explanandum of an explanation still leaves space for a number of diverging accounts of how they are related. The classical concept of philosophy of science is the deductive-nomological (DN) model (Hempel & Oppenheim 1948). According to this scheme, a phenomenon (the explanandum) is explained by a hypothetical theory (the explanans), if the hypothesis, taken together with specific boundary conditions of the individual circumstances of a certain case, allows deducing the phenomenon (i.e., the explanandum). For instance, the Newtonian laws and the specific knowledge of the mass of the moon and its position on the day X at midnight, allows deducing its position at midnight the next day.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boundary conditions</td>
<td>C</td>
</tr>
<tr>
<td>Phenomenon</td>
<td>B</td>
</tr>
</tbody>
</table>

2.4 The above model of an explanation has the form of a logical deduction. Note that in this account a theory is only of hypothetical character, leaving room, for example, for Popper's falsificationism (Popper 1935). The philosophical debates about the DN model and its decline since the 1960s will not be pursued here. However, one element of the debate shall be highlighted because of its relevance for the analysis of social processes, namely the question of causal mechanisms. A logical implication need not capture the mechanisms that connect explanans and explanandum. The following example is based on Salmon's (1989) famous logical deduction:

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>All people taking the pill will not become pregnant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boundary Condition: Steve takes the pill</td>
<td></td>
</tr>
<tr>
<td>Phenomenon</td>
<td>Steve does not become pregnant</td>
</tr>
</tbody>
</table>

2.5 While this is a correct syllogism, taking the pill is obviously not the cause of why Steve does not become pregnant. It overlooks the intervening variable of sex. Thus simple logical deduction is not enough. One of the various proposals to achieve a meaningful explanation is the mechanisms approach. Following this account implies a further condition: that an explanation of an empirical phenomenon needs to include the mechanisms that connect explanans and explanandum (Hedström 2005). The
Based modelling will demonstrate that the core assumptions of evidence-based modelling reflect central tenets of GT, motivating approaches (Edmonds & Moss 2005; Lotzmann & Meyer 2011). A common feature of the various methodologies in this research field is that they follow a descriptive account, rather than being based on a priori theoretical assumptions.

The central assumption of this modelling strategy is that detailed common sense descriptions allow for more valid statements about the target systems than do analytic propositions. Simplicity is not an aim in evidence-based modelling. While simple models, such as game-theoretically inspired ones, can be analysed more easily, it might be questionable whether the results can provide meaningful information about a target system (Edmonds & Moss 2005). In complex systems, it remains ambiguous which details of the systems’ components might be relevant to the systems’ behaviour. For this reason, Edmonds and Moss (2005) suggest keeping models of systems as empirically descriptive as possible, rather than relying on a priori theoretical assumptions. To achieve a detailed description, every source of evidence that can be gathered from the empirical field should be considered in the model-building process. Empirical evidence might take the form of classical statistical data, but also includes field observations and qualitative interviews, stakeholder knowledge (Barreteau 2003; Funtowicz & Ravetz 1994), textual data, audio and video files, and anecdotal evidence (Edmonds & Moss 2005; Yang & Gilbert 2008). These methods for data collection, derived from classical qualitative research, provide additional sources of evidence. In particular, these methods enable the integration of evidence about mental concepts, i.e. the meanings that participants ascribe to events, which is outside the scope of purely quantitative data (Yang & Gilbert 2008). To build model assumption on this evidence allows for a micro-validation of behaviour rules of the agents already involved in the process of the model development (Moss & Edmonds 2005). Only on the basis of a complex model can it later be decided which kinds of simplification preserve the properties of the system (Edmonds & Moss 2005).

Moreover, in participatory accounts the processes of model building, data gathering and model analysis are not separated, as is suggested by classical hypothesis testing. In contrast, modelling is a ‘bottom-up process’, meaning that it is an iterative process, cycling between modelling and field work (Barreteau 2003). Data generation in participatory accounts includes stakeholders who are involved in the process of model development. The process commences with initial stakeholder meetings to arrive at a basic first model, which is then presented to the stakeholders a second time in order to gather additional information from the stakeholders. This additional information is then again input for a revision of the initial model.

The descriptive account is partly supported by the fact that, in contrast to analytical methods of mathematical modelling, agent-based modelling allows for a rule-based modelling approach (Yang & Gilbert 2008). Rule-based modelling enables the implementation of a detailed description of individual decisions and actions on a social micro-level, within the rules of the model code (Lotzmann & Meyer 2011). This technical feature allows that models can ‘get away from numbers’ (Yang & Gilbert 2008), by replacing numbers with verbal descriptions in the code. The translation of empirical field notes into a computer model enables discovery of a system of rules in the empirical data. This implies that the transformation of field notes within the rules of a model code is a process that involves increasing abstraction to gain a consistent and coherent representation of the most salient features of the target system. The rules provide a core concept of the mechanisms at work in the target system.

In summary, evidence-based modelling applies a research methodology both iterative and inductive, starting from an idiographic description of the field of investigation. Thus, the research process does not follow the distinction between logic of discovery and logic of confirmation. Moreover, it shall be noted that simulation models in general enable an analysis of processes, because simulation consists of the observation of the dynamics of the model in simulated time steps. A simulation run allows the generation of statistical patterns that can be observed in the empirical data. This enables a qualitative and quantitative cross-validation (Moss & Edmonds 2005): a qualitative validation in the process of model development, and a quantitative validation in the analysis of simulation results.

Grounded Theory

Next, a brief overview of GT will be provided. The objective of this paper is not to examine the diverging variants of GT approaches (Kelle 2005). For this reason, only central tenets will be highlighted briefly. A comparison between GT and evidence-based modelling will demonstrate that the core assumptions of evidence-based modelling reflect central tenets of GT, motivating
4.2 The term ‘Grounded Theory’ is slightly misleading, since it is not a classical ‘grand’ theory, for example, a sociological systems theory; nor is it a middle-range theory of a certain phenomenon, for example, a theory of deviant behaviour. Rather, it denotes a certain methodological advice the stimulates the generation of theories (Flick 2002). The aim is not to test hypotheses as in the classical design of experiments, but instead to develop new theory by revealing hidden structures and relations in the data. Thus the research does not start with a theory that is subject of hypothesis testing, but instead with a detailed description of the field, from which relevant insights do not emerge until later stages in the research process. This research is conducted through an iterative process, in which data collection and analysis exist in a reciprocal relationship. The analysis of data should stimulate new questions posed to the data, which in turn stimulates new collection of data. Thus data collection is not a random sample, but rather a so-called theoretical sampling that is already guided by questions that have emerged during analysis. The process of switching between data collection and data analysis should be iterated until a stage of theoretical saturation is reached (Struebing 2004). Theoretical saturation is reached when further data reveals no further insights, indicating that the existing categories, which have been uncovered during the research process, are comprehensive (Goulding 2002; Locke 2001). This first sketch enables to highlight the following parallel between GT and evidence-based modelling:

- Like evidence-based modelling, GT is an inductive approach to studying social phenomena.
- Both methodologies commence research with a detailed description of the field.
- The concept of theoretical saturation parallels the iterative account in companion or participatory modelling approaches (Barreteau 2003), in which the process of model development is constantly informed by the expertise of stakeholders and vice versa.

4.3 However, distinctive differences shall also be mentioned:

- Simulation enables a representation of the dynamics of social processes.
- Representation of concepts in a computer code for simulation requires a high degree of formal precision.

Theoretical coding

5.1 GT aims at inductively reaching a theory. In the process of theory development, the notion of theoretical coding is of central relevance, as it describes the process of building categories from the data. The following overview will show that this process corresponds to the process of developing model assumptions in an evidence-based modelling approach.

5.2 While extracting categories from data has also been denoted as an art that should not be reduced to a technical execution of concrete instructions (Corbin & Strauss 2008), typically elements are denoted as line coding, focused coding, axial coding and selective coding (Flick 2002). For the purpose of this paper, these will be characterised briefly. The first elements, line coding and focused coding, are closely oriented at the data. In line coding, single lines of text are assigned to a code that describes the characteristics of the data, whereas in focused coding, larger text units are assigned to a code. These two data-oriented stages are also denoted as open coding. GT is particularly well-known for the so-called in vivo coding method, which uses the direct words of the research participants to describe a category. For instance, in a current research on criminal organisations, a criminal described the conditions, in which he had found himself, as a ‘rule of terror’, which provides a vivid picture of a war in the underworld. However, beginning with open coding, a process of increasing abstraction is initiated to integrate the empirical detail into a coherent picture. Axial coding encompasses the dimensions of and relation between the codes, and selective coding aims to analyse the story line that explains the phenomenon, for instance by identifying the core category or contrasting cases (Corbin & Strauss 2008; Flick 2002). Thus theoretical coding involves the process of building categories from key terms and relations in the data, by an increasing abstraction from detail. Categories do not denote the individual phenomena, but instead relate certain groups of phenomena into a single concept, that is, they denote a set. ‘Concepts that reach the status of a category are abstractions. They represent the stories of many persons or groups reduced into … highly conceptual terms’ (Corbin & Strauss 2008, p. 103). Furthermore, the set of categories is embedded in a web of relations that describe the properties of the categories in various dimensions. In the end, the categories themselves might become rather abstract. For instance, in a study on Vietnam veterans, Corbin revealed that categories such as ‘culture of war’ or ‘changing self from the interviews (Corbin & Strauss 2008) comprised ‘physical, psychological, social and moral problems’ (Corbin & Strauss 2008, p. 266) inherent in the phenomenon of war. The research commenced with a detailed coding of a single interview with volunteers who worked in an evacuation hospital; it was then enriched by more interviews with war participants who had been involved in other wartime situations, for instance, direct combat on the battle fields. Collectively the set of interviews enabled the development of abstract categories such as ‘changing self’. This parallels evidence-based modelling, which begins with a detailed description (line coding) for dissecting rules as mechanisms of salient features of the domain (axial coding), and only later reaches abstraction (selective coding). Thus the process of generating model assumptions in evidence-based modelling can be denoted as a variant of theoretical coding.

5.3 The process of theoretical coding is the central element of how GT aims to embrace a theory, rather than merely describing a phenomenon. Representing groups of stories in abstract conceptual terms fulfils the criterion of generality in a theory, that is, the criterion ‘a’ in this paper. Nevertheless, the relation between theory and description remains ambiguous. It is not guaranteed that simply following rules will generate a theory. For instance, one might fail to identify the core categories, or stick too closely to the data and retain a more descriptive account. Relying on the parallel between evidence-based modelling and theoretical coding,
Software ontologies

5.4 Ontologies are used in information systems and knowledge engineering for purposes of communication, automated reasoning, and representation of knowledge. In particular, the emergence of the World Wide Web generated the need for methods to extract information from a huge body of data (Gruber 2009). An ontology is defined as a formal, explicit specification of a conceptualisation (Guarino et al. 2009; Studer et al. 1998). The conceptualisation represents concept classes, which might include a hierarchy of subclasses. A conceptualisation consists of a triple C: \(<D, W, R>\) with D as the universe of discourse, W the set of possible worlds, which is the maximal set of observable states of affairs, and ( as the set of conceptual relations on the domain space \(<D, W>\). The universe of discourse is the domain of the ontology. Possible worlds represent possible applications. The structure is a relational structure, as it includes the relationships between the objects in the domain. The notion of conceptualisation can be defined as a set of representational primitives, typically classes, attributes, and relationships, with which to model a domain of knowledge (Gruber 2009). Here we will draw attention to the fact that automated reasoning requires the knowledge to be edited with formal precision, so as to be manageable for computers. For this reason, the development of an ontology is the mediating step between data analysis and simulation, by identifying key terms and relations of the domain (Diesner & Carley 2005; Hoffmann 2013) to increase the transparency of the derivation of simulation results (Livet et al. 2010). Thus, ontologies concern early and middle stages of the research in the developmental process of model assumptions.

5.5 The following examples will show how the development of an ontology can support the processes of theoretical and open coding in a GT process. The formulation of an ontology enables first a formal precision and coherence of the description of the domain of study. Second, utilising software tools such as Protégé (http://protege.stanford.edu/download/ontologies.html) enables automatic reasoning to inspect the implications of the logic system. This in turn enables an examination of the relations that describe the mechanisms of the processes driving the system. Third, the reference to classes (Gruber 2009) parallels the development of categories in a GT approach as sets, which relates certain groups of phenomena. By defining the sets, the formal precision contributes to theory development by assisting the generalisation of empirical findings.

An example of the contribution of ontologies to theoretical coding

5.6 The following example will demonstrate how ontologies contribute to the process of theoretical coding. It is drawn from ongoing research, which aims to investigate the global dynamics of extortion racket systems (ERS). The purpose is to develop a simulation model for understanding the dynamics of ERS such as the South-Italian Mafia. A first step in the development of a simulation model is the development of an ontology that provides the key terms and relations. Thus we need to define a space of discourse (D) and a set of relations (R) for our conceptualisation of ERS. This is based on a detailed analysis of the operations of the Cosa Nostra (the Sicilian Mafia) in the Sicilian Society (Scaglione 2011). In return for extortion money requested from the entrepreneurs, the Mafia offered them private protection and established a monopoly of violence (Franchetti 1876; Gambetta 1993), owing to a weak state authority and a lack of civil society. With regard to the question of what ontology development contributes to the process of theoretical coding, it needs to be emphasised that ontology development is an inductive and iterative process of code refinement, starting from data analysis to ontology development and back to the data. Ontology development enables the identification of gaps in the data basis, which suggests the need to gather new data. Data analysis and ontology development are a recursive process.

![Diagram](http://protege.stanford.edu/)

Figure 1. Reciprocal relation between ontology development and data analysis

Ontologies for theoretical coding

5.7 The ontology (we might call it ‘pizzo ontology’, as pizzo is the Italian word for extortion money) had been developed using the Protégé development tool (http://protege.stanford.edu/). To show how ontology development supports theoretical coding, two examples from this ontology will be provided: first, the representation of relevant organisations in the domain. Organisations are represented as objects derived from the root class of ‘thing’. The domain of ERS is characterised by three types of organisations (see fig. 2): the criminal organisation, the private enterprises—which provide the resources for extortion—and the public administration. In the case of the Sicilian Cosa Nostra, the criminal organisation of the Mafia consists of three operational unit
classes: the family as the basic unit, the Mandamento as a regional coordination unit, and the Cupola (Sicilian Mafia Commission), which is the top echelon of the hierarchy. For the purpose of characterising the domain of ERS, enterprises need to be distinguished into three classes, dependent on their likelihood of becoming victims of extortion: for small shops the likelihood is very high, whereas for big companies the danger of extortion is reduced. Construction companies are in high danger, regardless of whether they are big or small. The sphere of administration is divided into four professional organisations: the court, the police, the public administration, and politics. It can be seen in figure 2 that the ontology is arranged in a set theoretical manner: the more specific objects are subsumed under more general classes of objects by the relation ‘is a’. The more general classes represent sets of objects or sub-classes. However, note that the analysis begins from the reverse direction, namely by identifying the most specific organisations first. Ontology development is the process of classifying these specific organisations within more general classes until they are finally subsumed under the abstract class, ‘thing’. In a GT process, the identification of these general classes can be regarded as the process of abstraction in theoretical coding. The specification of subsets and relations between sets can be described as axial coding.

5.8 While the objective of this example is to demonstrate the principle of the development of terminologies in software ontologies, the objective of the next example is to show how the formal precision of the set theoretical account of ontologies provides a thinking tool for the development of terminology in the theoretical coding process.

5.9 This example shows parts of the actions undertaken by the organisations in the domain; specifically, the elements included in the extortion of entrepreneurs, and the service of protection offered in return. Actions are represented as objects. Note that not all entrepreneurs make use of this service. This is indicated by the black triangle. Nevertheless, the subsumption of terms looks rather strange here; it seems counter-intuitive to subsume protection under the heading ‘extortion’. However, the entrepreneur may decide to cooperate with the Mafia. The service offered by the Mafia includes protection from other criminals. This is the classical domain of a private protection company. Additionally, though, the Mafia may also help the entrepreneur by organising a cartel to hinder competitors looking to enter the market. Likewise, the Mafia may support illegal activities of the entrepreneur by using its contacts within the public domain. Protection generates a win-win situation for both parties. In this case, extortion is perceived as a kind of taxation, and protection is a subset of the groups of actions implied by successful extortion. Since ‘extortion’ and ‘protection’ are classical terminology used in research on the Mafia (e.g., Gambetta 1993) we decided to retain this terminology. Likewise, it seems counter-intuitive to subsume ‘buy’ under the term ‘protection’. However, if the entrepreneur decides to cooperate, the Mafia gains a hold within the company, and the entrepreneur no longer possesses absolute power. He
may even find himself in a situation where the Mafia takes over the company. This conflicts with the intuitive meaning of protection. In a GT process, the purpose of theoretical coding is to achieve the most abstract and general terms to precisely describe the phenomenon under investigation, as shall be illustrated by the example of the concept of ‘changing self’ to describe a situation of exposure to war. Thus the formal precision required to build a hierarchy of set theoretic subsumptions of terms reveals possible inconsistencies, and implausible or unequivocal terms. The objective of figure 3 is to provide an example of how ontology development stimulates theoretical coding.

Ontologies for open coding

5.10 Following the method of contrasting cases in the GT approach, this case had been contrasted with data from another case of organised crime. Here, the CCD tool (Scherer et al. 2013) has been utilised for knowledge representation. This is used to achieve a conceptual model of the data ready for simulation. The paper will draw on this example to show how formal knowledge representation in software engineering assists the process of open coding in GT. The data had been analysed using MaxQDA (http://www.maxqda.com/) and then imported to CCD. The coding derived with MaxQDA served as the basis for identifying relations with the CCD tool. CCD provides an environment for a controlled identification of condition-action sequences, which represent the micro-mechanisms at work in the processes described in the data. Whereas the data describes individual instantiations, the condition-action sequences represent mechanisms insofar as they describe generalisable event classes. However, empirical traceability is ensured by tracing the individual elements of the action diagram that resulted from the identification of condition-action sequences in the CCD tool, back to text annotation in the data. These annotations are extracted from the coding derived with MaxQDA (Neumann & Lotzmann 2014). The advantage of this formal knowledge representation is, first, that it enables a detailed analysis of the dynamics of processes by the condition-action sequences. Second, whereas the set theoretical account of finding abstract classes of concepts supports the process of abstraction in theoretical coding, CCD enables disentangling mechanisms on a very micro level, derived by single line coding. This assists open coding. However, by identifying the mechanisms that connect conditions with actions, the conceptual modelling already infers an element of theory in the data-driven stage of research. As the following example will show (see fig. 4), it requires finding concepts in which similar input corresponds to similar output. This is condition c) of a theory.

5.11 In contrast to the well-established Cosa Nostra, this case investigates the collapse of a criminal network in relatively early stages of the network’s existence. The data is based on police interrogations in 2005 and 2006. The network lasted for circa 10 years; it collapsed when initial conflicts escalated to a degree of violence that has since been described in the police interrogations as a ‘rule of terror’ in which ‘old friends were killing each other’. This overall situation consists of several micro-elements. However, at the time these were not visible to the members of the group, leading to the nebulous assumption of a ‘rule of terror’ that could not be attributed to individual persons. This has been described in the police interrogations as a ‘corrupt chaos’. In a covert organisation, commitment to the organisation cannot be secured by formal contracts. Therefore trust is essential. The following diagram shows a part of the process that led to a cascading effect in which trust collapsed.

![Diagram of action sequence](http://www.maxqda.com/)

Figure 4. Example of action diagram (part) of the contrasting case

5.12 Figure 4 reveals parts of the escalation process of conflicts within the criminal group. The starting point is the condition that some member of the group recognises an act of aggression performed against him by other members of the group. This triggers the action to interpret the aggression. It could be a sanction (i.e., norm enforcement), or self-interested aggression (i.e., violation of the trust he has in this group member). Interpreting the aggression as norm violation is the condition for counteraction either in form of betrayal or in counteraggression. Note that this abstract condition-action sequence can be traced back to annotations derived from the MaxQDA coding. An example for aggressive action against member X is the following annotation from the coding: ‘An attack on the life of M.’ Moreover, the data includes the testimony of a member who states that ‘M told the newspapers [about my role in the network]¹ because he thought that I wanted to kill him to get the money.’ M had survived an attack on his
life, but he was wrong in the assumption that this particular member of the organisation was behind this attack. Thus M decided to interpret the aggression as a violation of his trust in V01, and reacted by betraying him (Neumann & Lotzmann 2014). The fact that he was wrong in the attribution of guilt caused further conflicts within the network. This is an example of how CCD assists the process of open coding by dissecting the micro-mechanisms in the data.

Summary: Ontologies as tool for Grounded Theory

5.13 In summary, ontologies provide a tool for thinking in the processes of theoretical and open coding. Ontologies assist coding by means of the following features:

- They ensure formal precision and coherence of coding. The precision allows for the detection of gaps in the data.
- Identifying sets contributes to theory building by generalising empirically derived concepts. This corresponds to condition a) of a theory.
- Formal precision allows to check the consistency of the generalisation and to infer if the generalisation is sufficient to subsume the cases.
- They support open coding by disentangling complex verbal concepts from their constituent micro-elements. Identifying condition-action sequences assists the inference of mechanisms that drive a system. This corresponds to condition c) of a theory.

The explanatory power of Grounded Theory: Theoretical sensitivity

6.1 The example of ontologies of criminal organisations demonstrated how these tools might be utilised in the process of theoretical coding. This concerns early and middle stages of the research process. Next, the second criterion to assess the theoretical quality of results will be addressed, namely the relation between explanans and explanandum. This is the criterion b) of a theory. A theory aims to explain something. However, the degree inasmuch such insights are achieved by a field study remains ambiguous in the GT account. In the methodological research on GT, terms such as 'theoretical saturation' and 'theoretical sensitivity' provide quality criteria for the development of a theory. In the literature on GT methodology in particular, the term 'theoretical sensitivity' is used to assess the theoretical quality of the research (Corbin & Strauss 2008). Briefly, theoretical saturation is the criterion for stopping the iterative process of data collection and analysis. This is indicated if no more additional categories or properties can be found any more. On the other hand, theoretical sensitivity indicates the meaningfulness of the results. Corbin and Strauss (2008) define sensitivity as 'the ability to pick up on subtle nuances and cues in the data that infer or point to meaning' (p. 19). It is claimed that GT cannot be reduced to a routine application of certain methods. For this reason, in the GT literature (Glaser 1978; Glaser & Strauss 1967, Strauss & Corbin 1998) theoretical sensitivity is specified as being the credibility of the researcher. Thus emphasis is put on the notion of ability in the definition (i.e., the ability of the individual researcher), rather than on the subtle nuances in the data. For instance, the imagination and creativity of the researcher may be highlighted (Strauss & Corbin 1998), an action that evaluates the quality of the researcher rather than the research itself. This is a very personal conception (Birks & Mills 2011) and lacks a more objectifiable criterion. While it can be asserted that creativity and sensitivity in the field of analysis are essential for the significance of science, the assessment of the creativity of a researcher is highly subjective, as it depends to a large degree on the person undertaking the assessment. Moreover, this does not provide in itself a criterion to determine if the research achieved an insightful description or explanation of a certain phenomenon.

6.2 The objective of the second example is to show how simulation contributes to the quality criteria of reaching a theoretical explanation of a phenomenon, by specifying the explanans and the explanandum (i.e., criterion b) of a theory). For this purpose the example will draw on the notion of stylised facts, as developed for the investigation of simulation results (Heine et al. 2005, 2007). It will be shown how the simulation of stylised facts can provide a means to develop criteria for evaluation of the quality of qualitative research. Admittedly, this is not the conception of theoretical sensitivity. Nevertheless, it will be argued that simulation provides a source of evidence that the inductive research process generated theoretical insights rather than merely describing the phenomenon under investigation, by clearly specifying how stylised facts provide the mechanisms to connect explanans and the explanandum (i.e., criterion c) of a theory). Arguably this is a criterion for theoretical sensitivity, as it indicates the meaningfulness of the insights in order to provide an explanation. This will be demonstrated by a second example to show how simulation results contribute to the formulas of theoretical statements.

Stylised facts

6.3 A simulation model provides a means to clearly specify that which explains something else; the model assumptions provide the explanans while simulation results provide the explanandum. Simulation results are implications of the model assumptions, even if they may be too complex to be analytically tractable. Thus the assumptions generate the simulation result. However, as the discussion of theories in section 2 demonstrated, this does not suffice for a valid explanation. The example of Steve taking the pill shows that deduction need not be meaningful. An abstract model might provide sound statistical figures without providing meaningful information about a target system. However, in an evidence-based modelling account, model assumptions can already rely on qualitative empirical evidence. Results of simulation runs are typically some kind of statistical figures, which can be compared to empirical data. These two stages of evidence in the development of the model assumptions and simulation analysis describe the process of cross-validation (Moss & Edmonds 2005). The question remains if the model assumptions and
the simulation results are connected by causal mechanisms. The following example will show that stylised facts might provide such explanatory mechanisms. It will demonstrate how stylised facts enable explanation of the results of the simulation runs, by dissecting the mechanisms that drive the dynamics. Stylised facts provide a middle-range theory of the domain under investigation. The relation between evidence-based model assumptions and simulated stylised facts can be described as the explanatory narrative of the field, by revealing the explanatory power of the qualitative evidence of model assumptions.

Figure 5. Structure of an explanation

6.4 However, first the notion of stylised facts will be explained in more detail. The term ‘stylised facts’ was coined by Kaldor (1961) in macroeconomic growth theory. Heine et al. (2005) demonstrated that it can be applied beyond macro-economic analysis to the evaluation of simulation results. The central tenet of stylised facts is ‘to offer a way to identify and communicate key observations that demanded scientific explanation’ (Heine et al. 2005, 2.2). For this purpose, ‘stylised facts’ denote stable patterns that can be found throughout many contexts. Stylised facts are defined as follows:

♦ ‘Broad, but not necessarily universal generalisations of empirical observations and describe essential characteristics of a phenomenon that call for an explanation’ (Heine et al. 2007, p. 583).

6.5 Thus details of concrete empirical cases are left out in favour of a description of tendencies that have been identified as robust patterns that can be discerned in a certain class of phenomena. The fact that they are not restricted to an idiosyncratic description of a single case, but instead are salient characteristics of a class of phenomena, enables a generalisation of a particular case. Robust patterns identified by broad generalisations reflect the characteristic of mechanisms to regularly reveal similar outputs Y° under similar circumstances X°. In the case of evidence-based modelling this can be used as a cross-validation, to check whether the micro assumptions put in the model assumptions reveal stylised facts characteristic of the field of investigation. How this may encompass social mechanisms will be shown in the example below. With regard to the question of what a theory is in a GT approach, this provides an additional source of credibility for a GT process: if a simulation reveals stylised facts of mechanisms, which connect the input of model assumptions with the output of simulation results, then this would indicate a theoretic insight generated by the inductive process of evidence-based modelling.

An example for the contribution of stylised facts to the explanatory power of Grounded Theory

6.6 How simulation reveals stylised facts of mechanisms will be demonstrated by an example of a simulation model of the escalation of ethno-nationalist. Moreover, in this example it will be possible to integrate the results in the framework for theories of ethnic conflict. Thus the stylised facts allow for a final integration as demanded by Corbin and Strauss (2008), not only of the data, but also of the resulting theory in the canonical theoretical discourse of the domain. The example is a model that investigates the escalation of ethno-nationalist tensions into open violence. The evidence has been drawn from the case of the former Yugoslavia. The puzzling question is how and why neighbourhood relations between people with different ethnic backgrounds changed from genial and peaceful relations to traumatic and violent ones.

6.7 A simulation model has been developed to study the dynamics of nationalist radicalisation. Model assumptions were based on the empirical evidence of historical narratives of this much-studied case (e.g. Bringa 1995; Gagnon 2004; Melcic 1999; Woodward 1995; Sieber-Egger 2011; Silber & Little 1997; Witmer 2002). Initially the conflict started as a power struggle within the Yugoslavian Communist Party. Formerly communist politicians took advantage of ethnic sentiments, which allowed them to organise party loyalty with an ethnic agenda. In the beginning, the degree of ethnic mobilisation in the population remained small (Calic 1995). However, very soon civilians were becoming involved in the battles, and some even took part in war crimes. Ethnic homogenisation was undertaken by a paramilitary militia of civilians who were not integrated into the command structure of the Yugoslavian army. These civilians were responsible for numerous ethnic cleansings. Moreover, normally the militia pre-war inhabitants of certain villages—inhabitants who were of the same ethnic origin as the militia—of the imminent attacks; often the villagers chose not to pass on the information to their Croatian neighbors, and also participated in looting afterwards (Bringa 1995; Drakulic 2005; Rathfelder 1999). Thus the empirical evidence suggests a theoretical mechanism of a recursive feedback relation, between dynamics on a political level and socio-cultural dynamics at the population level.
6.8 The simulation model cannot be explained here in detail (Markisic et al. 2012). The model is public in the OpenABM archive: https://www.openabm.org/model/4048/version/1/view). For the justification of how the model development reflects a GT approach, see Neumann (2014). The target of the model assumptions is the change of neighbourhood relations. The empirical evidence suggests a two-level design of the model, namely to specify the mechanisms of the escalation dynamics of ethno-political conflicts as a recursive feedback between political actors and social identities at the population level. While a focus purely on the population level (e.g. Horowitz 2001) masks the responsibility of political actors, explanations that focus purely on the political level (e.g. Gagnon 2004) need to explain why certain politics were successful. Integrating both accounts generates a self-organised feedback cycle of political actors and attitudes. The basic mechanism in the model is an enforcement of the population's value orientations through political actors. These may be civil values or national identities. On one hand, politicians mobilise value orientations in the population to get public support. On the other hand, politicians appeal to the most popular value orientations in order to maximise the support. In abstract terms, the feedback relation can be described as a recursive function. Thus it is a positive feedback cycle; however, this is damped by the fact that various politicians compete over different value orientations in the population. The model was calibrated at the population census of 1991 in Serbia, Croatia, and Bosnia-Herzegovina. Whereas Serbia and Croatia had a rather homogeneous population, the population of Bosnia-Herzegovina was highly ethnically mixed.
Simulation experiments were undertaken with the assumption of complete ignorance about the empirical distribution of the cognitive components, namely of the political attitudes of the citizens and the political agenda of the politicians. Initially both the political agenda and the value orientation of the citizens are determined by chance, for all republics. This allows for studying the pure effect of the feedback cycle.

Thus, here we have our *explanans*: differences in the simulation results are due to differential population distribution since all other features are the same for all republics. For sake of simplicity of the argument, we concentrate on a single *explanandum*: the change in the value orientations of the population. This reflects the research question of how neighbourhood relations changed in course of the conflicts escalation. In fact, all republics reach a stage of nationalist radicalisation during the simulation. However, the dynamics reveals a crucial difference: while the simulated 'Serbian' and 'Croatian' population quickly becomes radicalised (see fig. 8 for the Croatian population), in Bosnia-Herzegovina radicalisation is much slower. Only in the second half of the simulation can a push towards nationalist radicalisation be observed (see fig. 9). This difference in the dynamics is our *explanandum*, which is explained by the *explanans* of the difference in the population distribution.
6.11 However, what are the mechanisms connecting explanans and explanandum? The example of Steve and how he did not become pregnant, presumably not because he took the pill, shows that a purely logical deduction is not sufficient to dissect the mechanisms. However, the model assumptions are based on qualitative evidence from the field, which ensures that the basic elements correspond to an empirical relative. During simulation these model assumptions generated stylised facts of two basic general mechanisms of the escalation dynamics. These stylised facts then reveal the mechanisms that connect the explanans and the explanandum. The first mechanism concerns political processes; the second mechanism concerns micro processes of neighbourhood relations. The interpenetration of the processes reveals a sequential ordering:

6.12 First, on a political level, visibility of the political appeals plays an essential role in radicalisation, by stimulating counter-radicalisation in a republic B to initial radicalisation in a republic A. This is driven by the political level, and accounts for a rather homogeneous population and nations with a common or closely related cultural heritage, such as Serbia and Croatia. Ethnically mixed populations, such as in Bosnia-Herzegovina, provide more power of resistance against political radicalisation prior to the outbreak of actual violence. Political radicalisation can be achieved easier in ethnically homogeneous nations.

6.13 Second, refugees and rumours play an essential role for later radicalisation in Bosnia-Herzegovina. Here, radicalisation is imported from outside and is driven predominantly by the population level. Dense networks increase the likelihood of radicalisation spreading.

6.14 These are theoretical results, derived from a dense description of a particular case, which has been transformed into a more abstract code of a simulation model. However, the simulation results are not limited to an idiothetic description of this particular case, but also provide stylised facts of the mechanisms of nationalist (or value-driven) radicalisation. This is an example of how simulation provides a means to analyse an explanation from model assumptions based on a GT approach, by dissecting the mechanisms that connect explanans and explanandum. Thus if evidence-based model assumptions generate meaningful stylised facts, the simulation run indicates a theory in the sense that something explains something else.

6.15 Moreover, it is possible to achieve a final integration not only of the data, but also of these results, into the theoretical discourse on ethnic conflicts. Note that a main result has indicated that ethnically mixed populations provide more power of resistance against initial political radicalisation. This addresses current theoretical debates in conflict research (Cederman et al. 2010; Fearon & Laitin 2003; Rutherford et al. 2011; Wimmer et al. 2009). Whereas classical sociology of conflicts explained inner state violence via the theory of relative deprivation (Gurr 1970), in the times after the cold war a rise in the number of ethnic conflicts was observed. While it might remain ambiguous whether or inasmuch conflicts simply were perceived differently after the cold war (Wimmer et al. 2009), this nevertheless initiated a research programme in recent decades, in which ethnic conflicts have become a subject of investigation in their own right, as opposed to being subsumed under a broad theoretical umbrella such as the theory of relative deprivation (Neumann 2014). The clash of civilisations (Kaplan 1996) is a prominent catchphrase for these accounts. Various causes have been suggested (e.g. Sambanis 2001) to explain why ethnic groups might be tempted to fall into violent conflicts. Wimmer et al. denote these accounts as a diversity-breeds-conflict theory (Wimmer et al. 2009), according to which it is the diversity that explains the conflicts. Thus ‘diversity’ is the explanans and ‘conflict’ is the explanandum of this theoretical account. However, the simulation study of our example reveals a different result: whereas ‘diversity’ is the explanans, the explanandum is different, namely power of resistance against political radicalisation. This casts doubt over the clash of civilisation thesis. It may be true that self-perpetuating violence on the micro level might indeed be more severe in ethically heterogeneous territories, once the power of resistance has been broken. The case of Bosnia-Herzegovina perpetuates this statement. Nevertheless, the model shows that diversity in itself is not a sufficient cause to breed conflict. In fact, the outbreak of violence happened later in Bosnia-Herzegovina than in Croatia (Rathfelder 1999). Simple reference to diversity lacks a specification of the mechanisms of conflict escalation. The diversity-breeds-conflict theory is based on statistical data analyses of the large-N research in conflict research (see also Florea 2012). However, statistics cannot reveal mechanisms of social dynamics. Indeed, this model reveals mechanisms that point in a different direction, namely ‘power of resistance against political radicalisation’. At least next to diversity, additional mechanisms need to come into play to foster conflicts. In the model, this refers to the second mechanism of imported violence, driven by refugees on the micro level of neighbourhood relations. This discussion demonstrates that simulation of stylised facts allows the achievement of theoretical insights from a GT starting point. In terms of theoretical integration, the integration of the simulation results into the broad scope of theories on ethnic conflict can be regarded as the most abstract framework for describing theories.

Conclusion

7.1 The paper argues that the theoretical element in a GT approach can be strengthened by supplementing the methodology of GT with evidence-based simulation. This is demonstrated by two examples: first, it is shown that the development of an ontology of
criminal organisations refines the process of theoretical coding by providing additional precision, which allows to detect gaps in data and concepts and to specify the scope of the domain. Set theory contributes to criterion a) of a theory (“generality”). Identifying condition-action sequences supports open coding by a specification of mechanisms in a system, thereby contributing to criterion c) of a theory (“mechanisms”). Second, by using the example of the escalation of ethno-nationalist conflicts in the former Yugoslavia, it is shown how findings from simulation of an evidence-based model generate stylised facts. Simulation tools enable to derive an explanation from a narrative story by connecting the model assumptions with the simulation results. Model assumptions provide the explanans, and results provide the explanandum. This contributes to criterion b) of a theory (“explanation”). Stylised facts enfold the mechanisms connecting these two, thus contributing to criterion c) (“mechanisms”). These are the basic elements of a theory. If a model, based on the evidence of an empirical case, generates broad patterns that reveal mechanisms which connect the explanans and explanandum, then this shows how a process starting with an idiosyncratic description succeeds in generating a theory. This contributes to a clarification of the precarious relation between a mere description and a strictly theoretical GT.

Acknowledgements

The work on this paper is part of the GLODERS project, funded by the European Commission under the 7th Framework Programme. It is an extension and refinement of a paper presented at the 9th Conference of the European Social Simulation Association, Warsaw 2013, in the panel ‘Using qualitative data to inform behavioral rules’. The contribution of comments and discussions to improve the paper are greatly acknowledged. The author would like to thank two anonymous reviewers and a proof-reader for critical and constructive hints to improve the paper. All flaws and errors are sole responsibility of the author.

Notes

¹For reasons of protecting privacy the specification of the role has been replaced.

References


