Modelling Rules, Norms, and Institutional Change Using the Grammar of Institutions

Eva Ebenhöh

Institute of Environmental Systems Research
University of Osnabrück
49069 Osnabrück
Eva.Ebenhoeh@usf.uos.de

Abstract This paper reports on a formal theory to analyse institutional statements - rules and norms - and how this theory can be applied in agent-based models to simulate and investigate institutional change. The theory is the grammar of institutions by Crawford and Ostrom [1,2]. This paper describes a possible application of this theory including an agent decision making mechanism, in which rules and norms that have been adopted by an agent serve as first filters to reduce the number of possible actions. The paper also reports briefly on two examples of models in which the grammar of institutions has been applied. This way of modelling institutional statements may help to model institutions and to investigate institutional change by means of social simulation.

1 Introduction

When agent-based social simulation is used for investigating institutional change, institutions, their effectiveness, and their emergence need to be endogenous in the models. This paper reports on a way of conceptualising institutions in order to transfer them from real-world problems into computer code. This is a step to facilitate modelling of institutions and institutional change, as well as of comparing institutions and their impacts across models.

One view on institutions regards them as constraints on decision makers' behaviour. As such they can be integrated in agent-based models. However, institutions vary greatly in both the way they are described and in their impact. The Grammar of Institutions [1,2] is a way of describing institutional statements using five distinct grammatical building blocks. This stringent view of analysing institutions reduces ambiguity while transferring rules and norms from the real world into computer code [3].

This paper presents the theory (Section 2) and its application to agent-based models, including an exemplary decision making process in which institutional statements constrain agent behaviour (Section 3). It includes a brief description of exemplary models in which this approach has been used (Section 4) and concludes with a short discussion on the usefulness of this approach.

2 Grammar of Institutions

The grammar of institutions [1,2] is a logical syntax, in which institutional statements can be represented as well as compared and contrasted. Institutional statements "describe opportunities and constraints that create expectations about other actors' behavior" [2, 137]. This syntax is useful for implementing institutions as constraints for agent behaviour in social simulation models.

This grammar can represent both formal and informal institutions [4]. Whether they are efficient or not, does not depend on their formality, but is determined by whether they are followed and enforced. Rule following and enforcement can happen in different ways in the real world, which can be represented in this theory. It allows for investigating consequences arising from not following a rule. Rule compliance can be modelled on this basis, including extrinsic (sanctions) and intrinsic motivation.

2.1 Five Building Blocks

According to this grammar, a *rule* is an institutional statement consisting of five components (ADICO): Attributes, Deontic, aIm, Conditions, and an "Or else" statement.

Attributes define any individual attributes needed to distinguish actors from each other, in order to define to whom the institutional statement applies.
Deontic A deontic statement is either permission, obligation, or forbiddance, that is "may," "must," or "must not", respectively. Note, that it is possible to define each deontic by the other two.
Aim defines what it is that may, must, or must not be done.
Conditions define states of the world in which the rule applies.
Or Else defines consequences for not following the rule.

Some examples throughout this paper illustrate the building blocks and some are later used as examples in a modelling context. A sign in a train could read: *If you don't have a valid ticket, you are charged 60 Euro.* Attributes in this case are implicitly referring to all people travelling in a train. However, the law probably specifies some age restriction, say age seven and older. The deontic is "obligatory." The aim is to *have a valid ticket.* Deontic and aim can be exchanged: *It is forbidden to travel in a train without a valid ticket.* Conditions are *in trains of a specified railway company.* The "Or Else" part is that you are charged 60 Euro. However, if you do not get caught, you do not have to pay. So, there is a chance that the sanction is not excerted. Re-phrasing the statement according to the grammar leads to: *All persons from age seven on - must - have a valid ticket - when travelling in trains of a specified railway company - or else suffer a fine of 60 Euro.*

2.2 Rules, Norms, and Shared Strategies

Three subsets of institutional statements can be distinguished according to which of the five elements are defined.

Shared Strategies - AIC Shared strategies consist of attributes, aim, and conditions. Consider for example the shared strategy: *Customers tip in a restaurant.* The attributes *customers* states that the shared strategy applies to all who eat a meal. The aim *to tip* refers to what is to be done. The condition *in a restaurant* states, where tipping is usual, for example this shared strategy says nothing about tips in a drive thru. Consider also: *Players of repeated prisoner's dilemma games - cooperate - in first moves* [2, 157].

Norms - ADIC Norms consist of attributes, deontic, aim, and conditions. Consider for example: *If you use the microwave, you must clean up your own mess* [2, 139]. The attributes refer to microwave users. The aim is to clean up. Cleaning up is obligatory, as the deontic "must" states. Conditions are if they caused a mess. The statement can be rephrased: *Microwave users - must - clean up - if they caused a mess* in order to bring the elements in the corresponding order of attributes - deontic - aim - conditions. Consider also this norm: *Players of repeated prisoner's dilemma games - must not - cheat - unless being cheated before.*

Rules - ADICO Rules consist of all five elements. Rules are distinguished from norms by adding explicit consequences for violating a rule. The train-riding example given above is an example for a rule. Another example would be the grim trigger: *Players of repeated prisoner's dilemma games - must not - cheat - unless being cheated before - or else suffer cheating in all following rounds.*

3 Applications in Agent-Based Models

3.1 Agent Decision Making

The agent decision making mechanism proposed here uses rules, norms, and shared strategies as first filters in a two-step filter process as displayed in Figure 1. From a list of possible actions, only those are considered, that are permitted by the institutional statements applicable in the current situation and individually considered to be important. From the remaining possible actions, which I call "considered actions", other mechanisms can be used to choose a single remaining action. These other mechanisms can include evaluative criteria or simple decision heuristics. By evaluative criteria, expected outcomes are judged, for example, highest expected return or equality of outcomes. Examples for simple decision heuristics include closeness to a previously made decision or choosing a focal point. The difference between evaluative criteria and decision heuristics is that the former refers to expected effects of an action, while the latter refers to the action itself. In our models, they are used in a similar way, but note, that the outcome of evaluative criteria depends on the way in which expectations are formed. Expectation formation may also be guided by institutional statements as suggested in [2] and modelled, for instance, in [5].

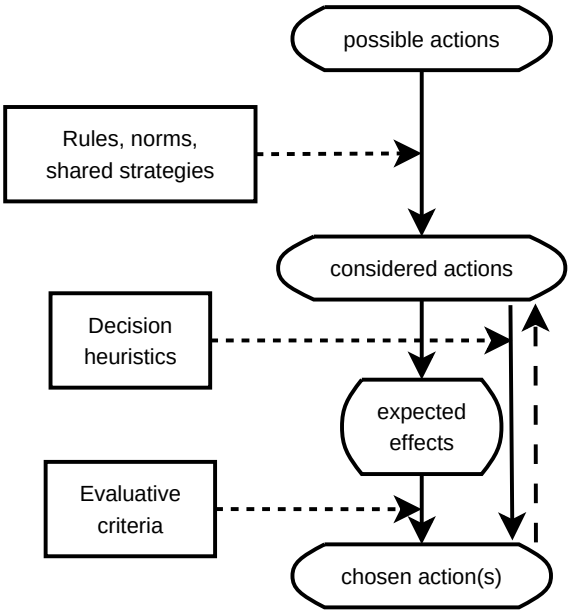

Figure 1. The decision making process suggested in this paper is a two-step process, which first filters out possible actions according to rules, norms, and shared strategies, and secondly determines the best option according to applied evaluative criteria or decision heuristics.

If a single action needs to be chosen, the second step of the process is repeated using a list of evaluative criteria and decision heuristics with decreasing priority to successively reduce the list of considered actions. This process stops, when there is only one choice left. In order to define an end-point to this process, in one model, I included a random choice decision heuristic with least priority to select a single action out of a list of actions, which could not be distinguished by an agents' evaluative criteria.

If more than one action can be chosen at the same time, the process gets more complicated. For now, in our models, only single decision processes have been modelled. Multi-decision processes have been modelled as sequences of single decision processes.

3.2 Reasons for Following Rules and Norms

As in the real world, the existence of a rule, does not automatically lead to it being followed. Different reasons exist for humans to follow rules. In order to use social simulation to investigate institutional change, these different reasons have to be taken into account in the respective models.

In the introduction I argued that this syntax reduces ambiguity. It can not, however, erase ambiguity completely. The prisoner's dilemma example for a rule, the grim trigger, can alternatively be stated as three shared strategies [2, 157]:

1. All players - cooperate - first round.
2. All players - cooperate - all rounds in which all players cooperated in the previous round.
3. All players - defect - all rounds after a defection.

Consequences of violating rules and norms, therefore, do not only include explicitly mentioned sanctions (which may or may not be excerted), but also behavioural changes in reaction to an agents' behaviour. Which of the two sets of institutional statements is more effective can not be seen, by merely looking at the corresponding formalisation. Rather, factors determining effectiveness are how many players follow the third shared strategy and how often the explicit sanction is actually exerted.

Sanctions One way of enforcing rules is by defining sanctions, monitoring rule compliance, and excerting defined sanctions after a rule violation has been perceived. Saam and Harrer model the Finder-Keeper norm as an institutionalised rule with sanctioning of observed violations [6, 4.15]. All three processes are of importance. Turning a norm into a rule by explicit "or else" statements, does not suffice if monitoring is not ensured or if there is no agent with the possibility and willingness to excert the sanction. The probability for being detected can be taken into account by agents, calculating expected returns of rule violations. In the train example, if tickets are rarely checked, the expected return is different than if tickets are checked on almost every ride. Nevertheless, pressure to follow

norms can be modelled without explicitly modelling monitoring, for instance as "peer pressure" in [7].

Sanctions are not always pecuniary. They can be excerted in specific choices for actions in the future, as the grim trigger in the prisoner's dilemma example.

I do not propose, however, that agents should calculate subjective expected utility on the basis of monitoring and sanctioning probabilities. Instead, it seems sufficient that rules are regarded as effective or not, and are followed (or not) accordingly. Thinking about the train example this approach seems more realistic than calculating expected returns.

Reputation In repeated interactions, trust building or more generally expectation formation of others' behaviour, is one way to enforce norms or shared strategies, which lack an explicit sanction, as for instance through normative reputation in [8]. In repeated prisoner's dilemma games, the threat of defecting does not have to be stated as an explicit sanction, since it is in the interests of players to cooperate on the long run. Once cooperation has been established, trust is increased. Trust is modelled as the expectation that the other follows the shared strategy of cooperation [9]. Trust loss through one defection, on the other hand, can be hard to build up again. Future interactions in which another agents' behaviour is bound to be based on the reputation of a decision making agent, may be incentive enough to follow group norms or shared strategies.

Conformity As humans are social beings and socially embedded [10], there seems to be an incentive for some to follow a norm which is followed by most. Agents can be programmed to do the same using the "majority rule" [11]. In our models, this agent trait is called conformity [9,12]. The higher their conformity, the lower the percentage of other agents following a norm needs to be, in order for them to follow it as well.

Norm adoption, however, may also be induced by mirroring peer behaviour and not general behaviour, see for instance [13] on imitation.

Emotions Due to internalisation of norms, emotions such as guilt, shame, and anger can be made responsible for norm following behaviour [14]. These would be modelled as intrinsic motivations to follow a given norm.

Habit To reduce the amount of decisions or thinking a human or software agent has to do, many norms and shared strategies are followed out of habit. In that case, stopping to follow a norm, needs to be triggered. We model this, by giving an institutional statement an individual strength. This strength may be reduced or increased by observations of others or outcomes of previous applications. Thresholds exist which would trigger a re-evaluation of an institutional statement.

3.3 Transferring Institutional Statements into Computer Code

In order to be applicable for social simulations, the building blocks need to be transferred to building blocks as used in a simulation model. The examples discussed in Section 4 draw on a modelling template [15] based on the Institutional Analysis and Development (IAD) framework by Elinor Ostrom [16,17]. However, the following findings are rather generic and maybe even obvious, but still it is worth noting that all meaningful elements of institutional statements need also be meaningful elements in the model, if their impact on actors' behaviour is to be investigated.

Attributes For attributes to provide a meaningful distinction between agents, they must be based on those elements, that constitute agent heterogeneity. If, for example, heterogeneity of an agent population is solely based on one attribute, say cooperativeness, values for cooperativeness could be distinguishing characteristics of institutional statements: *Players with a cooperativeness of greater than 0.5 - cooperate - in first moves.* In most models, however, even if agents are very much alike, they differ in more than one respect, for instance they can be located in specific points in space and time, have different experiences, and different connections in a social network and the like. A shared strategy in a prisoner's dilemma could be: *Players directly connected to a peer cooperator - cooperate - in first moves.*

Conditions Conditions refer to the state of the world. What has been said about attributes, also refers to conditions. They have to be meaningful in the model and, furthermore, agents need to be able to perceive them. If an agent has no way of distinguishing a cooperator from a defector, the shared strategy *All players - defect - when encountering a defector* is meaningless. If there is no way of knowing, the way in which expectations are formed by an agent is of high importance. *All players - defect - when expecting to encounter a defector* could be the same *All players - defect - when the other defected in the last game* if expectation is based on the latest experience.

The difference between conditions and attributes is that the former refers to the world outside the agent and the latter to the world within the agent. Experiences and expectation formation are somewhat in between, but refer to the outside world, even if they are processes or states inside an agent.

The distinction is clearer, for a tag directly perceivable by an agent. Consider the Finder-Keeper strategy [8,6] *Normative agents - must not - attack - agents who eat their own food.* It seems obvious that the food needs to be tagged so that the agents can see to whom it belongs. To make agents perceive if other agents are also normative was explicitly modelled as a model extension with communication of normative reputation. Only then, the norm turned into *Normative agents - must not - attack - agents who are normative and eat their own food.*

Deontic and Aims Aims need to be actions that are actual options for a decision making agent, or else the institutional statement is meaningless in a given situation.

In the decision making mechanism suggested above, institutional statements serve as filters to reduce the number of possible actions. Since each deontic statement can be defined by the other two, we could choose "must not" and "may" (see below) as the only implemented deontics. In prisoner's dilemma games, the norm *All players - must - cooperate* is equal to *All players - must not - defect.* More generally if an agent is obliged to do something, all other actions are forbidden. This, however, assumes that the institutional statement refers to the complete list of possible actions. This is not in all models the case. If the set of possible actions evolves and, for instance, walking away from a prisoner's dilemma situation is introduced [18], *All players - must - cooperate* does not allow walking away, while *All players - must not - defect* does. If the list of possible actions is not closed, "must" and "must not" need to be explicitly modelled.

A permission, on the other hand, does either not change the list of possible actions, or it is an extension to this list. In the latter case, however, either the environment or another institutional statement can contradict this extension. The environment is assumed to define physical impossibilities and therefore rules out a permission. Another institutional statement would have to be more important to the agent in order to rule out a permission.

Or Else The "or else" statement needs to be something that the agent can calculate in its decision making. Also, expectations of the probability of getting caught are important in this respect. Possible sanctions are not always carried out, either because the rule violation has not been monitored or because it appears to be of minor importance to or is costly for those who could excert a sanction.

4 Examples

This kind of decision making has been used in two different models. One is a model of a series of economic experiments with varying institutional settings including the possibility to communicate and adopt self-designed shared strategies as well as agree on sanctioning mechanisms. The second is a model of an upstream-downstream problem of flood protection, in which decentralised, centralised, and collective action regimes are contrasted using different institutional statements to create behavioural constraints for decision makers.

4.1 Institutional Statements in an Economic Experiment

In a series of economic experiments [16] eight participants have to make anonymous choices of how many tokens, out of their 25 per round, they want to invest

in a common pool resource. The return from the resource depends on total investment in a negatively quadratic way. In the parameterisation used in [16] group optimum is at 36 tokens total, 4.5 tokens per participant. In baseline experiments, no communication or sanctioning is allowed. Experiments have also been conducted with communication phases in which participants could decide on shared strategies, with sanctioning mechanisms, and with both communication and sanctioning, including the possibility to decide on adopting sanctioning mechanisms.

In the model of the baseline experiment, the only norm implemented is *All players - must not - cheat - always*. Cheating is defined as investing more than the symmetric Nash strategy, which is 9 tokens, in the parameterisation in [16]. Roughly half the agents start out following this norm, according to their individual conformity value.

The model allows suggesting and voicing support of shared strategies during communication rounds. Shared strategies are implemented as *All players - invest 4 (5, 6, 7 . . .) - as long as all others do so*. Individual complicance to these shared strategies depend on two things: The first is a combination of the individual conformity value and the perceived expressions of compliance to a shared strategy. The second is the expected returns of following the shared strategy and thus keep it going. Thus, if following a shared strategy pays, even agents with a low conformity can be drawn into mutual cooperation.

Shared strategies can turn into rules, even without explicit sanctioning mechanisms, for example as: *All players - must - invest 4 - as long as all others do - or else in the next round all invest 10.*

4.2 Institutional Statements in Transboundary Flood Protection

Three different institutional settings are contrasted in a model of flood protection as an upstream-downstream problem. A three bucket river is managed by three different administrative bodies.

The decentralised setting defines no institutional statement for taking the downstream effects into account, when deciding on flood protection plans.

The centralised setting defines a rule with varying "or else" parts: *All decision makers - must - take the whole basin into account - always - or else be relieved of office*; *or else pay a fine*; *or else do not receive money out of a joint fund for flood protection*. Different reasons for following these rules, are investigated in this setting, including but not limited to the sanctioning threat.

The third setting, a collective action setting, allows norms of other regard to emerge during communication phases, in which information is exchanged and trust is built. The norm says the same as the rule in the centralised setting without an "or else" part. The model is used to investigate circumstances under which a collective action situation can outperform a centralised regime.

5 Discussion

This paper describes the Grammar of Institutions, a theory which can be used to transfer institutional statements from the real world into computer code. The theory helps to analyse and compare institutions, by defining building blocks, by which institutions can be compared and their impacts analysed. This decomposition into building blocks helps to implement institutions and their impacts on agent behaviour in agent-based models. It further aids modelling institutional change by making explicit what elements need to be captured in the model.

Based on this conceptualisation, emergence of new rules as well as emergence of rule following may be modelled. With this kind of representation of institutional statements, agent-based models can be used to investigate institutional change.

However, as we have seen, institutions are not unambiguously represented in this theory. For instance, the grim trigger in prisoner's dilemma games can be modelled as a set of shared strategies or as an explicit sanctioning mechanism.

Also, the proposed decision making process does not pre-define all necessary processes. For instance, the adoption of norms and rules is so far modelled as a cue-based or learning direction mechanism. Even in the two examples, agent reactions on institutional change, have not been investigated thoroughly enough, to be modelled with confidence. Rather, the models mentioned in Section 4 were used to develop this technique. For social simulation models based on this template, careful data collection and qualitative confirmation with expert knowledge is necessary [19]. For the upstream-downstream model, this is currently done in close interaction with a case study.

References

1. Crawford, S.E.S., Ostrom, E.: A grammar of institutions. American Political Science Review **89** (1995) 582–600
2. Crawford, S., Ostrom, E.: A grammar of institutions. [17] chapter 5 137–174
3. Smajgl, A.: Advancing the capacity to simulate institutional change. [20] 189–196
4. Helmke, G., Levitsky, S.: Informal institutions and comparative politics: A research agenda. Perspectives on Politics **2** (2004) 725–740
5. Bhavnani, R.: Adaptive agents, political institutions and civic traditions in modern italy. Journal of Artificial Societies and Social Simulation **6** (2003) <http://jasss.soc.surrey.ac.uk/6/4/1.html> (March 22, 2007).
6. Saam, N.J., Harrer, A.: Simulating norms, social inequality, and functional change in artificial societies. Journal of Artificial Societies and Social Simulation **2** (1999) <http://jasss.soc.surrey.ac.uk/2/1/2.html> (March 22, 2007).
7. Thebaud, O., Locatelli, B.: Modelling the emergence of resource-sharing conventions: an agent-based approach. Journal of Artificial Societies and Social Simulation **4** (2001) <http://jasss.soc.surrey.ac.uk/4/2/3.html> (March 22, 2007).
8. Castelfranchi, C., Conte, R., Paolucci, M.: Normative reputation and the costs of compliance. Journal of Artificial Societies and Social Simulation **1** (1998) <http://jasss.soc.surrey.ac.uk/1/3/3.html> (March 22, 2007).

9. Ebenhöh, E., Pahl-Wostl, C.: Agent-based modelling with boundedly rational agents. In Rennard, J.P., ed.: Handbook of Research on Nature Inspired Computing for Economy and Management. Idea Group (2006)
10. Edmonds, B.: Capturing social embeddedness: A constructivist approach. Adaptive Behaviour **7** (1999) 323–348
11. Jaffe, K., Cipriani, R.: Culture outsmarts nature in the evolution of cooperation. Journal of Artificial Societies and Social Simulation **10** (2007) <http://jasss.soc.surrey.ac.uk/10/1/7.html> (March 22, 2007).
12. Ebenhöh, E.: A topology of agent attributes for modelling microbehaviour in economic experiments. [20]
13. Conte, R., Paolucci, M.: Intelligent social learning. Journal of Artificial Societies and Social Simulation **4** (2001) <http://jasss.soc.surrey.ac.uk/4/1/3.html> (March 22, 2007).
14. Staller, A., Petta, P.: Introducing emotions into the computational study of social norms: A first evaluation. Journal of Artificial Societies and Social Simulation **4** (2001) <http://jasss.soc.surrey.ac.uk/4/1/2.html> (June 10, 2006).
15. Ebenhöh, E.: Testing hypotheses on adaptive management using an agent-based model based on the institutional analysis and development framework. Deliverable 1.2.3 in the NeWater Project (2007) <www.usf.uos.de/∼eebenhoe/newater/Mileston1.2.3.pdf>.
16. Ostrom, E., Gardner, R., Walker, J., eds.: Rules, games, and common-pool resources. Ann Arbor, Michigan: The University of Michigan Press (1994)
17. Ostrom, E.: Understanding Institutional Diversity. Princeton University Press (2005)
18. Joyce, D., Kennison, J., Densmore, O., Guerin, S., Barr, S., Charles, E., Thompson, N.S.: My way or the highway: a more naturalistic model of altruism tested in an iterative prisoners' dilemma. Journal of Artificial Societies and Social Simulation **9** (2006) <http://jasss.soc.surrey.ac.uk/9/2/4.html> (March 22, 2007).
19. Moss, S., Edmonds, B.: Sociology and simulation: Statistical and qualitative cross-validation. American Journal of Sociology **110** (2005) 1095–1131
20. Terano, T., Takahashi, S., Sallach, D., Rouchier, J., eds.: Proceedings of the First World Congress on Social Simulation. Kyoto University (2006)