Adoption of a New Regulation for the Governance of Common-Pool Resources by a Heterogeneous Population

Marco A. Janssen

Center for the Study of Institutions, Population, and Environmental Change
Indiana University

Formerly:

Department of Spatial Economics
Vrije Universiteit, Amsterdam, the Netherlands

Elinor Ostrom
Center for the Study of Institutions, Population, and Environmental Change
Workshop in Political Theory and Policy Analysis
Indiana University

Acknowledgements

Thanks to Sam Bowles for his comments on an earlier version of this paper. Also thanks to the participants of seminars at Indiana University, Universite Libre de Bruxelles, and the Annual Resilience Alliance Meeting 2001 in Chiang Mau. Support of the Resilience Alliance, the National Science Foundation (Grant application SES0083511) and the European Union (contract nr. IST-2000-26016) is gratefully acknowledged.

Introduction

The rich, Lofoten cod fishery in Northern Norway has been successfully self-governed and managed for more than 100 years. The rules that regulate the use of this fishery – and make it likely that the fishery will be sustainable into the future – have been devised by the boat-owners themselves with minimal external assistance or coercion. Local regulatory committees are elected by the boat-owners, determine the rules for harvesting from the fishery, and effectively monitor this system (Princen, 1998). Once their own rules are in place, monitored, and sanctions for non-compliance are regularly applied, it is relatively easy to understand why the boat owners would comply with well-designed and enforced rules. How the users themselves developed their own rules and ways of monitoring and sanctioning non-compliance with these rules is much more difficult to understand given current accepted theories of collective action.

As discussed below, multiple examples exist of long-existing, self-governed systems for limiting resource use that increase the probability of sustainable resource systems (See Bromley, et al., 1992; Ostrom, et al., 2002). The design principles that characterize robust self-governed institutions have been identified and confirmed by multiple studies of successful and unsuccessful efforts at self-governance (see Ostrom, 1990; Morrow and Hull, 1996; Weinstein, 2000). The evidence from the field, however, constitute an anomaly. Existing theories of collective action do not yet provide a full explanation for how the appropriators (harvesters) from a common-pool resource can solve three nested social dilemmas:

1. The first dilemma is that of multiple appropriators harvesting from a single common-pool resource. This is the "tragedy of the commons" dilemma (Hardin, 1968). Most resource policy textbooks presume that appropriators will act individually and over-harvest from a common-pool resource.
2. The second dilemma is that of spending time and effort to create a new set of rules that jointly benefit many of those who rely on a resource whether or not they contribute time and effort to the process of devising regulations. Since rules are themselves public goods, this is a public good dilemma. The standard theoretical prediction is that rules will *not* emerge as the result of an endogenous process. Rather those involved must have rules imposed upon them from the outside.
3. The third dilemma is that of monitoring a set of rules and imposing sanctions on those who break the rules. Monitoring and sanctioning are costly activities. They generate rule conformance – which benefits most appropriators whether or not they contribute to the activity. This is also a public good dilemma. Most theories treat rule enforcement as an exogenous variable rather than something that the participants themselves undertake.

Given the extensive evidence that some local appropriators from common-pool resource do solve all three dilemmas, while others do not, a central question on our research agenda is: "How do rules emerge in complex common-pool resource systems used by multiple appropriators?" This paper is our first effort to use agent-based computational modeling[1] to address this question. We will explore the factors that enhance or detract from the possibility that individuals appropriating (harvesting) from a common-pool resource will first, impose

[1] Throughout this paper we refer to agent-based computational modeling, which is also known as agent-based modeling or multi-agent systems. A brief introduction is given in the sections on agent-based computational modeling. For a discussion on terminology we refer to Parker et al. (2002).

rules on themselves to limit their harvesting from a jointly used resource, and second, monitor and enforce their own rules. We will show that in order to adopt a rule to avoid a "tragedy of the commons," agents have to build up mutual trust relationships involving short-term costs. The degree of success of rule emergence depends, amongst other variables, on the heterogeneity within a population of agents and the type and frequency of interactions.

Why Use Agent-Based Computational Models

The use of agent-based computational models is still not widely accepted – especially by scholars using analytical theories, including game theory -- as a foundation for doing institutional analysis. Thus, before discussing what agent-based computational modeling is, it is important to address why we choose this modeling approach. Game theory is a useful tool for conducting studies of the choice of strategies *within* a given set of rules. Given its usefulness, game theory will continue to be applied extensively to analyze many questions of relevance to the study of institutions and common-pool resources (Ostrom, Gardner and Walker, 1994; McGinnis, 2000). To conduct a game theoretical analysis, however, one must assume that a fixed, commonly understood, and followed set of rules is already in place.

Evolutionary game theory is also an important tool for the institutional analyst, but strategies rather than rules are posited to evolve over time (for an excellent overview see Gintis, 2000). Further, to conduct a full analysis of a complex game usually requires that one assume a homogeneous set of players facing a homogeneous biophysical world. While one can never model the full extent of the complexity that exists in regard to the governance of common-pool resources, introducing some degree of heterogeneity in regard to the attributes of participants and of the bio-physical world they face is an important step in achieving a better theoretical understanding of these processes.

We are particularly interested in the resilience of institutional structures – the ability of a particular set of rules to absorb a disturbance. The resilience of an institutional structure, however, also depends on the heterogeneity of the biophysical environment and the participants. Thus, introducing heterogeneity is an essential aspect of our investigation. Further, we are not only interested in the possible existence of an equilibrium and what that equilibrium is, but also in the time-path towards an equilibrium state. Another aspect of analytical models in evolutionary game theory is the implicit assumption that each agent has an equal chance of interacting with every other agent. This implicit assumption results from the use of differential equations. Agent interactions, however, take place mainly within social networks. Recent investigations show that the structure of agents in a social network can have crucial consequences for social processes such as the evolution of cooperation (Ellison, 1993; Cohen et al., 2001). Agent-based computational models are able to examine such social relationships.

Tournaments between strategies for playing repeated prisoner dilemma games pioneered by Robert Axelrod (1984) can be considered as one of first applications of agent-based computational models in social science. Agent based computational models allow us to explore the impact of a variety of variables identified in the field as affecting the likelihood of successful self-organization including the size of the resource, the number of agents, the attributes that agents obviously carry (tags), and the presence or absence of trust. Using this modeling technique, it is possible to analyze how these important factors independently and interactively affect the likelihood of a rule being adopted and on the patterns of monitoring and sanctioning that may ensue. One can create an artificial, computational laboratory to examine how a more complex set of variables interacts than is possible with any fully analytical technique. Our approach in this paper is novel even for agent-based computational models. Many of the important works in this tradition have examined the change of strategies over time rather than the change of rules over time (see for example Axelrod, 1997; Miller, et al., 2002).

What Are Agent-Based Computational Models

During the last decade an increasing number of scholars in social and natural science use multi-agent systems, which consists of a number of interacting agents (Epstein and Axtell, 1996; Conte et al., 1997; Liebrand et al., 1998; Gilbert and Troitzsch, 1999). Agents can represents animals, people or organizations. Agents can be reactive, proactive, may sensor the environment, communicate with other agents, learn, remember and move.

Within economics this field is called agent-based computational economics and studies economies modeled as evolving systems of autonomous interacting agents (Tesfatsion, 2001). An important challenge of agent-based computational economics is studying how stylized facts at a macro level emerge from the bottom up. These approaches have been applied to financial markets, macroeconomics, innovation dynamics, economic districts, environmental management, labor economics, and other processes.

The two main components of agent-based computational systems are: 1) cellular automata (defined below) representing the relevant biophysical world, and 2) models of the agents representing how the agent is posited to process information, learn, and make decisions. Each agent is represented as a computerized independent entity capable of acting locally in response to stimuli or to communication with other agents. Therefore, the first task is to build an architecture for intelligent agents. The second task is to design a structure in which interacting agents may or may not accomplish a task.

John von Neumann and Stanislaw Ulam introduced the cellular automata (CA) approach at the end of the 1940s, mainly to give a reductionistic model of life and self-reproduction. The game of life, invented by John Conway in 1970, popularized the CA approach (Gardner, 1970). This game consists of cells on a checkerboard, which can have two states, 'alive' and 'dead'. Time goes by discrete steps. According to some deterministic rules, which are the same for each cell, the state of a cell in the next time step depends on its own present state and the states of all its surrounding cells in the present period. The resulting surprisingly complex dynamics that evolved from this simple game attracted the attention of many people. Since the early 1970s, CA have been used by many disciplines to study complex dynamic behavior of systems.

The basic features of a CA are (Hegselmann, 1998):

- A D-dimensional lattice.
- Time advances in *discrete* steps.
- A *finite* number of states. At each site of the lattice exists a cell, which is in one of the possible states.
- The cells change their states according to *local* rules, both in space and in time.
- The transition rules are usually deterministic, but non-deterministic rules are allowed as well.

- The system is *homogeneous* in the sense that the set of possible states is the same for each cell, and the same transition rule applies to each cell.
- The updating procedure can consist of applying the transition rule to all cells *synchronically* or *asynchronically*.

The architecture of agents in agent-based computational systems has been much influenced by work in Artificial Intelligence (AI). In this field a popular wave is the autonomous agents research or behavior-based AI, which studies the behavior of adaptive autonomous agents in the physical world (robots) or in the cyberspace (software agents). This field in AI is strongly inspired by biology. The phenomena of interest are those traditionally covered by ethnology and ecology (in the case of animals) or psychology and sociology (in the case of humans). The agents often consist of sensors to derive information from the environment and intelligent functions such as perception, planning, and learning that react to that information. Behavior is defined as regularities observed in the interaction dynamics between characteristics and processes of a system and the characteristics and processes of an environment. Examples are: a theory at the behavior level that explains the formation of paths in an ant society in terms of a set of behavioral rules without reference to how they are neurophysiologically implemented. Another example is the study of behavioral rules implemented in robots who have to survive (they need to reload energy every now and then) in a physical environment with other robots as a way to explore emergent behavior in such a group. An overview of this field can be found in, for example, Steels (1995) and Maes (1995).

Distributed artificial intelligence is a relatively recent development of artificial intelligence studies (Bond and Gasser, 1988). It concerns the properties of sets of intercommunicating agents coexisting in a common environment. The aim may be to study the properties of such systems in an abstract way, or to design systems of immediate practical use, or to use such a programmed agent-based computational system as a model of a human or other real-world system.

Agent-based computational systems consist often of interacting agents in a cellular automata environment. In this paper we study the emergence of collective action from the bottom up. This is related to the studies on the evolution of cooperation (see Gintis, 2000). This problem has been studied intensively during the last 20 years by agent-based computational models starting with the seminal work of Axelrod (1984). The original studies of iterated prisoners dilemmas considered repeated interactions of two players. However, when a game consists of more than two players, it becomes more complicated to assess the consequences of possible behavior of the other players for any of the individual player's own strategy. The more agents involved in a common-pool resource, the more sophisticated procedures are necessary.

We will provide specific details about our model below in the section on "Model description."

Common-Pool Resources

Common-pool resources share two attributes of importance for economic activities: (1) it is costly to exclude individuals from using a common-pool resource either through physical barriers or legal instruments, and (2) the benefits consumed by one individual subtract from the benefits available to others (V. and E. Ostrom, 1977; E. Ostrom, Gardner, and Walker, 1994). Recognizing this class of goods that share two important, theoretically relevant attributes enable scholars to identify the core problems facing individuals or groups who wish to utilize such resources for an extended period of time.

First, common-pool resources share with public goods the difficulty and cost of developing physical or institutional means of excluding beneficiaries. Second, the products or resource units from common-pool resources share with private goods the attribute that one person's consumption subtracts from the quantity available to others. Thus, common-pool resources are subject to problems of congestion, overuse, and potential destruction -- unless harvesting or use limits are devised, implemented, and enforced. In addition to sharing these two attributes, particular common-pool resources differ on many other attributes that affect their economic usefulness including their size, shape, and productivity and the value, timing, and regularity of the resource units produced.

Common-pool resources may be owned by national, regional, or local governments, by communal groups, by private individuals or corporations or used as open access resources by whoever can gain access. Each of the broad types of property regimes has different sets of advantages and disadvantages (Feeny et al., 1990). Examples exist of both successful and unsuccessful efforts to govern and manage common-pool resources by governments, communal groups, cooperatives, voluntary associations, and private individuals or firms (Bromley et al., 1992; Berkes, 1989; Singh, 1994; Singh and Ballabh, 1996).

In settings where a central authority has not already claimed and enforced ownership of a common-pool resource, an important theoretical question is how those who are appropriating resources from a common-pool are able to develop their *own* rules limiting the quantity of resource units harvested. Since Hardin (1968) used the strong metaphor of "The Tragedy of the Commons," many scholars have accepted the view that local appropriators (fishers, irrigators, pastoralists or others appropriating resource units from a common-pool resource) are trapped in an inevitable and tragic destruction of the resource upon which they were dependent. Extensive studies of local common-property systems, however, demonstrate that the "tragedy" is not inevitable (Ostrom, 1990; Bromley et al., 1992; Burger, et al., 2001; Pathak and Korhari, 2001). Successful self-organization is also not inevitable.

The core question is how can a group of appropriators independently create their own rules, monitor each other's conformance with rules, and sanction one another when rule infractions are discovered. As Gallarotti (2001) points out, extensive empirical research documents the capabilities of individuals to organize themselves, establish new rules, monitor rule conformance, and sanction rule breaking, but little theoretical work has explored how rules emerge in the first place (for an initial sketch of a theory, see, Chapter 6, E. Ostrom, 1990).

Thus, in this paper we address the puzzle of why some groups form and establish their own rules to regulate their resources, and why other do not. Closely related is the question why some groups fail after making an initial start. Some communal systems fail or limp along at the margin of effectiveness just as private firms fail or barely hang on to profitability over long periods of time.

Factors Associated with Self-Organization

Scholars who have conducted extensive field research on locally self-organized institutions for managing common-pool resources have identified a set of attributes of appropriators that the researcher considers to be conducive to the evolution of norms, rules, and property rights that improve the probabilities of sustainability of common-pool resources. While there is some controversy about particular variables - such as the size of group and the degree and kind of heterogeneity of the group - the following variables are frequently found in the list of factors considered by empirical researchers to be crucial foundations of self-governed resource use:

1. Accurate information about the condition of the resource and expected flow of benefits and costs are available at low cost to the participants leading to a common understanding of likely consequences of continuing the status quo (no regulation of use) as contrasted with feasible changes in rules (Blomquist, 1992; Gilles and Jamtgaard, 1981; Sethi and Somanathan, 1996; E. Ostrom, 1990).

2. Appropriators plan to live and work in the same area for a long time (and in some cases, expect their offspring to live there as well) and, thus, do not heavily discount the future (Grima and Berkes, 1989).

3. Appropriators are highly dependant on the resource (Gibson, 2001).

4. Appropriators use collective-choice rules that fall between the extremes of unanimity or control by a few (or even bare majority) and, thus, avoid high transaction or high deprivation costs (E. Ostrom, 1990).

5. The group using the resource is relatively stable (Seabright, 1993).

6. The size of the group is relatively small (Franzen, 1994; Fujita, Hayami, and Kikuchi, 1999; Wilson and Thompson, 1993; Kirch, 1997).

7. The group is relatively homogenous (Bardhan and Dayton-Smith, 2002, Bardhan, 1999; Libecap, 1995; Lam, 1998; but see Varughese and E. Ostrom, 2001).

8. Participants have developed generalized norms of reciprocity and trust that can be used as initial social capital (Bardhan, 1993; Cordell and McKean, 1992; E. Ostrom, 1990).

9. Participants can develop relatively accurate and low-cost monitoring and sanctioning arrangements (Berkes, 1992; Ostrom, 1990).

In our agent-based computational model that is described below, the first four factors are built into our model by assumption. We consider the eighth and ninth factors to be particularly important and we focus largely on the processes and consequences of establishing reciprocity and trust and on factors affecting monitoring and sanctioning arrangements. The fifth, sixth and seventh factors are varied to some extent in the simulations described below.

Establishing Trust

While many definitions of trust exist, all involve a trustor extending something of value to a trustee based on an assumption that the trustee will reciprocate the faith that the trustee has extended. A core aspect of most definitions of trust is the "intention to accept vulnerability based upon positive expectations of the intentions of the behavior of another" (Rousseau et al., 1998). Using this conception of trust, one can recognize that almost all economic transactions other than immediate spot exchanges involve greater or lesser levels of trust (Arrow, 1974). And, anyone willing to accept a rule restricting their actions in an economically important activity must trust that the others involved will follow that rule as well. If they do not trust others to cooperate and follow a rule, individuals are unlikely to follow risky strategies of cooperating with others. Once established, however, cooperation may almost be "habit-forming" based on the trust that others will continue to be trustworthy (Seabright, 1993).

In non-cooperative game theory, rational egoists do not trust other players voluntarily to perform a costly action unless either the other players expect greater benefits to occur than expected costs or unless some external agent enforces agreements. And yet, in many laboratory experiments of "The Trust Game," a large proportion of those individuals placed in the position of the trustor do make themselves vulnerable to others by investing resources in others that could be returned in a manner that makes all better off. And, a large proportion of those who are trusted, are actually trustworthy and do reciprocate the trust (see E. Ostrom and Walker, forthcoming, for an overview of these studies).

The degree of trusting behavior and trustworthiness, however, varies substantially in the experimental laboratory as well as in the field. Multiple contextual factors associated with the structure of experiments and field studies are posited to affect the variance in trusting and trustworthy behavior. The amount of information that individuals have about each other, the likelihood of repeated interactions, and the identity and homogeneity of the individuals involved all appear to affect the level of trust.

In the model simulations of the emergence and sustenance of a local rule, we thus place the development of trust as a key instrumental variable while examining how several other factors affect both the development of trust and the emergence and maintenance of a local rule. The model in this paper is based on the agent-based computational models for the study of cooperation, but will add some new elements endogenously to introduce a new rule set, and to monitor and enforce these rules.

Model Description

A stylized simulation model is developed in which agents share a common resource. If no rule is implemented, the resource and the appropriators will experience serious collapses, and the feasible cooperative optimum situation will not be reached. A possible rule is known to the participants and could be implemented. Our aim with this model is to investigate the critical conditions under which a heterogeneous population will implement this rule, and how successful it will be monitored and sanctions applied. The population of appropriators is heterogeneous in characteristics such as the intention to break rules, degree of monitoring, eating rates and the symbols they wear.

Many possible rules could of course be implemented to avoid over-harvesting, but only one particular set of rules is assumed to be a candidate in our model. Although we start this research avenue with the problem of adopting a candidate rule, our ambition is to understand how a variety of rules evolve. This requires, however, further conceptual developments on the evolution of rules (Janssen and Stow, 2002). The choice to model only one candidate rule is caused by the fact that leaving open the possibility of more than one set of rules adds immense complexity that is not directly related to the questions at the core of our interests.

Our interest is to identify the critical assumptions that lead to a timely implementation of a rule set, and an effective monitoring and sanctioning regime. As discussed above, the findings of field research and laboratory experiments show that the development of mutual trust is a crucial factor. Therefore, we are especially interested how mutual trust development can be influenced.

To avoid overwhelming the reader technicalities we present only the basic equations of the model. The rest is explained verbally. The programming code of the model can be derived upon request. The model is implemented in CORMAS (Bousquet et al., 1998; http://cormas.cirad.fr), which is a shell around the object-oriented language Smalltalk. CORMAS[2] is especially designed to simulate agents in a cellular automata environment for the study of common-pool resources.

The model developed in this paper differs from the standard models of a renewable resource since the resource is spatially explicit. This spatial element is included to restrict the knowledge of agents to available, local information. The resource is split up in NxN cells. To avoid edge effects, the corner cells are wrapped around to derive a torus, a donut shape system. Neighbors are defined by a so-called Moore neighborhood which includes cells to the north, south, east and west of the center cell as well as diagonal cells to the north east, north west, south east and south west. These always form a square pattern. So a cell has 8 neighbors, and a neighborhood contains of 9 cells.

Biomass on each cell grows according to the logistic function of equation (1). In this equation, a renewable resource (energy), *x,* is modeled according the standard models in bioeconomics (Clark, 1990) and develops according to a logistic growth equation,

$$x(t) = x(t-1) + r \cdot x(t-1) \cdot (1 - \frac{x(t-1)}{K}) - e_c \qquad (1)$$

where x is the biomass, K is the carrying capacity, r is the growth parameter and e_c is the energy consumption of an agent.

Each cell has the same carrying capacity, but the initial values of the stocks are drawn randomly. In this spatial landscape a population of mobile agents exists. Only one agent can occupy a cell for any one time period. Agents can move each time step to another unoccupied neighboring cell with the highest biomass level. If there is no neighboring cell with a higher biomass level than the present cell, the agent does not move. The agents harvest a fixed percentage q of the cell biomass, with a maximum of $e_{c,max}$. The embodied energy e can be defined as the embodied energy in the previous time step minus the metabolic rate c, plus the energy consumption e_c.

If agents are not successful in harvesting energy, and the level of embodied energy depletes, the agent will die (i.e. leave the system). Agents can also reproduce, asexually, when their embodied energy level exceeds a certain threshold f. Above this energy level, the agent splits into two agents with an initial energy level equal to $f/2$. Genetic information is copied to the offspring, although genetic information mutates with a low probability p_m for each property.

Each agent has a number of properties that can differ among the population. These properties can be unobservable (krypta) or observable (manifesta) (Bacharach and Gambetta, 2001). The unobservable properties are the fraction of biomass on a cell that is eaten by the agent (q), the

[2] CORMAS stands for Common-Pool Resources and Multi-Agent Systems

intention to break the rule ([illegible]), the probability that the agent will monitor and sanction others (p_s) and the weighting (w) assigned to the symbols. The observable features are the symbols (s). The agents have genetic symbols, which are unchangeable during the lifetime of the agent, and they have cultural symbols, which can be adapted by the agent. The evolution of the properties of the agents depends on genetic and cultural evolution as well as on learning.

The distribution of genetic information, the properties that remain fixed over the lifetime of an agent, may change over time as a consequence of mutation in the offspring. Cultural evolution influences the configuration of cultural symbols and depends on imitation processes (see Bowles, 1998). Learning leads to adaptations of weighting the symbols to estimate trustworthiness.

The fitness of an agent varies related to its energy use during different activities. Agents with an initial very low eating fraction will not derive enough energy to survive, while other agents with higher levels derive more energy and are able to produce offspring. When a rule is implemented, the energy budget is also influenced by the energy spent when the agent breaks the rule and is caught. Consequently, a co-evolutionary process exists between the probabilities that agents break rules and that they monitor each other. If there are many agents breaking rules, it benefits to monitor, but then the benefits for breaking rules decline. The embodied energy can now be defined as:

$$e(t) = e(t - 1) - c + e_c(t) - c_m - p + r \qquad (2)$$

where c_m is equal to the extra energy consumption for monitoring, p the penalty of being caught when breaking the rule, and r the reward for catching rule breakers.[3]

Evolution of Mutual Trust

The agents start in a situation where there is no regulation. There is one candidate rule, namely no harvesting is allowed when the energy level of the cell is below x_{min}[4]. Furthermore, if the rule is implemented an agreement exists on the sanction penalties, s_e, of rule breakers, and how the collected penalties are allocated among the group. The question is under which conditions the candidate rule will be activated, and how successful it will be monitored and sanctioned. Whether the rule will be implemented depends on the evolution of mutual trust.

The description of mutual trust relationship is based on the work on indirect reciprocity (Nowak and Sigmund, 1998). Within an indirect reciprocity approach agents keep track of image scores of individuals, where the image scores represent the degree of cooperative actions of the agent in the past. So, when an agent meets another agent it derives information about its past performance of cooperation. The drawback of such an image score is the assumption of perfect knowledge.

We propose another version of indirect reciprocity. All agents have symbols that are visible to each other. During interactions with others, agents learn to recognize those symbols and attach meaning to them to discriminate people they think are trustworthy from those who are viewed as untrustworthy people. This can have important consequences. Although two agents

[3] See Weissing and E. Ostrom (1991) for a game theoretic analysis of the costs and rewards of monitoring.
[4] One can think of the candidate rule as being the rule used by a successful neighboring community in achievable a sustainable harvesting regime over time.

may have a high image score, they may not trust each other. Another issue is whether symbols really represent discriminating factors. It is the belief of agents that they are able to recognize trustworthy others.

We assume that agents develop an expectation of the trustworthiness of others using observable information (the tags). Such an approach is in line with Macy and Skvoretz (1998). To update the relation between the tags and trustworthiness, agent interact with their neighbors. Each time step each agent interacts with a neighbor if this neighbor has an expected trust level, EY_{ij}, above a minimum, EY_{min}. During this interaction, agents exchange information. In formal terms, they play a basic trust game. One agent is the trustor and starts the interaction with another agent, the trustee. The probability that a trustor starts an interaction is stochastic and depends on the expected trust level EY_{ij}. When the trustor has started an interaction, C, representing cooperation, the trustee has to decide to perform action C or D. The probability of this decision depends on the expected trustworthiness of the trustor.

The relation between expected trustworthiness and the probability of choosing C or D differs for the trustor and the trustee. We assume that starting an interaction, which means a trustor chooses C, requires relatively more trust of the trustor in the trustee, than the trust that a trustee must have to respond to C with a choice of C. Starting the interaction involves more uncertainty, therefore a relative high level of expected trustworthiness is required. On the other hand, refusing to react positively to a proposed interaction is not very polite, and it reduces your perceived trustworthiness. The degree to which the probabilities to cooperate differ is represented by parameter □ ≥ 1. If □ is equal to one, the probabilities are equal for both trustee and trustor.

The pay-off matrix of the trust game is constructed as follows. If an agent is starting an interaction or reacting to a proposed interaction, it will costs an amount of energy equal to □. The receiver will derive an amount of energy equal to □, where □ < □. Therefore, the net reward depends on the response.

	trustee C	D
trustor C	(-□+□,-□+□)	(-□,□)
D	-	-

The relation between one's agent's assessment of the trustworthiness of the other agent with specific, observed tags is updated based on the actions of the other agent. For example, if a trustor chose "C" and the trustee react with "D", the assessed trustworthiness is updated in such a way that agents with similar tags as the trustee, are perceived as less trustworthy in the next round. If the trustee returns "C"," the perceived trustworthiness is reinforced. So due to repeated interaction between agents, mutual trust can be developed or destroyed. Note that trust games have negative pay-offs. The benefits of playing trust games are entirely indirect. By building up mutual trust relationships, the agents are more likely to benefit from other interactions in the future.

To calculate the expected trustworthiness of other agents, an agent makes an expectation based on observed symbols. This is implemented as a single layer neural network:

$$EY_{t,ij} = w_{t,i,0} + \sum_{k=1}^{ks} w_{t,i,k} \cdot y_{t,j,k} \tag{3}$$

where $EY_{t,i,j}$ is the level of trust agent i has in agent j, $y_{t,j,k}$ are the inputs of the neural network, the observed symbols (-1 or 1) of the other agent j. Finally, the inputs are weighted by $w_{t,i,k}$. A neural network is trained when new information about the input values is used to update the weights. The widely used Widrow-Hoff delta rule is used to train the neural network during the simulation (Menrota et al., 1997).

$$\Delta w_{t,i,j} = \eta \cdot \delta_{t,i} \cdot \frac{\mathbf{y}_i}{\|\mathbf{y}_i\|} \tag{4}$$

where

$$\delta_{t,i} = Y_{t,ij} - w_{t,i,0} - \sum_{k=1}^{ks} w_{t,i,k} \cdot y_{t,j,k} \tag{5}$$

where $\delta_{t,i}$ is the difference between the observed values and the expected value, and $Y_{t,ij}$ is the experienced trust during the interaction. The observed value is 1 in case of a positive interaction, and 0 in case of a negative interaction. The delta rule therefore updates the expectations according to the observed errors. The rate of updating depends on the value of η, which is suggested to lie between 0.1 and 1 (Gallant, 1993).

Motivation to Adopt the Rule

The agents can become motivated to adopt the candidate rule. This motivation depends on the observed depletion of the resource, and on the mutual trust among agents. The general idea is that when an agent has sufficient trust in other agents, and the observed resource condition falls below a threshold, the motivation to implement the rule increases, otherwise the motivation decays each time step by a fraction □. Reinforcement by observed depletion of the resource and trust in others can increase motivation.

The reinforcement of positive motivation, can only be derived when certain conditions are met. First, the observed biomass level in the neighborhood, x_n, should be below a level $x_{n,min}$ (equal to K/2). Second, the expected regrowth of the biomass in the neighborhood should not be sufficient to satisfy the metabolism of the number of agents a_n in the neighborhood. Third, if the resource is not expected to satisfy the metabolism of the agents, a maximum life expectancy of the agent is calculated. If this expectancy is lower than minimum value T_y and the expected trust in the population $O(Y)$ exceeds a minimum level $O(Y)_{min}$, then the motivation will be reinforced. The value of T_y can be interpreted as the maximum number of timesteps the agent wants to use for finding better opportunities. The expected trust in the population is defined as the moving average of expected trust in the last (1/_) neighbors. Thus if _ is equal to 1, the trust in the population is equal to the trust in the most recent neighbor looked at while a _ of 0.05 leads to an average expected trust of the last 20 neighbors.

A possibility exists that the observed resource is depleted, but the population of agents has been decreased to such a low level, that some agents do not meet others during a sequence of rounds. In such cases, it is not appropriate to reinforce motivation although the expected trust in the last 1/□ neighbors meets the conditions. Therefore, when an agent has not met another agent for the past 10 rounds, the motivation to implement a rule will not be reinforced.

An agent is motivated when m exceeds m_{min}. The regulation is implemented when a certain constitutional level of agents agrees with the adoption of the regulation. This could include several collective choice rules, including: (1) 50% of the population plus one, (2) 50% of the weighted population plus one where the votes are weighted with the welfare level (embodied energy) of the voters, or (3) a decision by the leader of the group.[5] In the simulation experiments of this paper we assume that the collective choice rule is the first one, namely, 50% of the population plus one.

When a regulation is implemented agents can follow rules, break rules, and can monitor others. When an agent wants to harvest cell biomass below the tolerated level, it breaks the rule with a certain probability. If an agent has neighbors it monitors the neighborhood with a fixed chance to check whether a neighbor is breaking a rule. When the agent is caught it has to pay a penalty equal to the amount below the tolerated level and a percentage s_e of the embodied energy. The level of s_e increases with the number of times an agent has been caught breaking the rule.

The penalty is split up. A fraction *(1-g)* of the penalty goes to reward the monitoring agent, which derives a. The remaining fraction *g* of the penalty will be allocated among the whole population. Furthermore, when a rule breaker is caught, the trust relationship is adjusted. The neural network of equation (5) is updated with an observation 0, a negative interaction with an agent.

Probability of breaking the rule is related to a fixed, genetic, habit to break rules □, the level of trust in the population O(Y) and the level of embodied energy. The higher the trust in the population, the lower the chance that the agent will break a rule. Furthermore, the higher the level of embodied energy, the lower the chance of breaking a rule.

The Completed Computer Experiments

The experiments focus on the determination of critical conditions that influence the ability of the agents to derive a cooperative solution to the sustainability of the resource system. The default set up is designed in such a way, that with plausible parameter values a cooperative solution is likely to be derived. Additional sensitivity experiments test whether changes in the assumptions lead to structurally different solutions.

The experiments are performed using N=20, thus 400 cells, and a Moore neighborhood, thus 8 neighbors for every cell. In Table 1 the parameter values of the default case are listed. To stimulate cooperation we assume in the default case no cost during the basic trust games and low monitoring costs.

In the first set of experiments all agents can use the whole lattice without limitations. We will test whether the type of neighborhood and the size of the lattice influence the outcomes of the default case. Furthermore, we will analyze the impact of the cost of monitoring, the sanction level, the share of the penalty allocated to the whole group, the costs of playing trust games, the chance of starting an interacting (parameter □), the threshold values that influence motivation (T_y, m_{min} and $O(Y)_{min}$), and the number of symbols. This set of experiments provides a comprehensive overview of the dynamics of the system.

[5]These three are only a small sample of the very large number of constitutional-level rules that could be used.

In the second set of experiments, the torus is split up in two. When agents stay in their own region, they will build up their own set of cultural symbols. When agents can enter another region, for example in times of scarcity, they are strangers to the existing population. The first genetic symbol is here assumed to indicate the region of birth. For simplicities sake, we assume that offspring of an agent in a foreign region is born in the original region of its parent. Agents move probabilistically to the other region when their embodied energy is low and they observe scarcity in the neighborhood.

Four different cases are distinguished in the analyses of the two groups. First, the carrying capacities can be equal in each region (K=10), or they can differ between the regions using K=5 for one region and K=15 for the other. Second, the agents can accept the rules in the foreign region, or they may not accept the foreign rules. Note that agents can only monitor and sanction when they are in their own region. Moreover, agents outside the region can vote for implementing the rule but they do not experience scarcity in their region of origin.

The theoretical cooperative optimum is equal to 100 agents and a biomass energy level of 2000. This can be derived as follows. The maximum output of a cell occurs at x=5. The regrowth is than equal to 0.2*5*0.5=0.5. Since each agent use 2 units of energy, and 400 cells produce 200 units of energy in the optimal case, 100 agents can be provided with energy. 400 cells of 5 units of energy equal 2000 units of energy.

For each experiment 50 runs are performed for 3000 time steps, which differ in the initial allocations of the agents, the initial level of energy at the cells, and the symbols and values of □, q, p_m of the initial population. The initial number of agents is equal to 50% of the theoretical equilibrium, which is 50 agents in the default case.

Results

To provide a basic understanding of the dynamics of the model, the default case will first be discussed in detail. Although 50 runs have been performed, we start with a discussion of a ‘typical’ run.

The system experienced two crises before the candidate rule was adopted. The population size fluctuated between 10 and 100 during the first 400 time steps. After the rule adoption around time step 250, the population size gradually increased toward the theoretical equilibrium of 100 agents (Figures 1 and 2). A similar development can be observed for the total amount of biomass energy on the torus (Figures 1 and 2). Note that the population change follows the biomass energy figures with some delay. When the resource starts to recover after a crisis, still a large part of the lattice does not produce enough energy to sustain the population, and population continues to decrease for a while (Figure 1).

In Figure 3 it is shown that during the first period of scarcity, the fraction of agents agreeing with implementing the rule set did not surpass 50%. At that time, the system started to recover, and the motivation decreased. During the second crisis, the average trust level was higher than during the first crisis, leading to a larger proportion of motivated agents. The voting agents reach about 52% around timestep 260. The reason for the decrease of the trust in their neighbors after the two periods of crises is caused by two factors. First, agents have a trust level of 0.5 when they are born, and have to meet people to build up trust in their neighborhoods. Second, the fast increase of the population is likely to include agents with new combinations of symbols. In this case, new agents have to learn which agents (characterized by which set of symbols) they can trust. Since the average trust in the neighborhoods increase after the rule set has been implemented, the amount of rule breaking decreased as well as the level of monitoring and the amount of energy harvesting, resulting in a steadily increase of the population size (Figures 1-2). During the periods of prosperity, a high amount of biomass energy and a low population size, the energy harvest peaks at 4 units per capital. The amount of energy spent on monitoring remains low compared to fixed metabolic rate.

Interestingly, the eating fraction q increases during the period without the candidate rule in place. In a 'cowboy' society (without any rules), those who consume the most energy, will get the highest number of offspring. Thus, evolutionary selection favors selfish agents. The higher eating rate causes a deeper collapse of the biomass energy during the period of scarcity. When the rule is implemented, the value of q decreases since it has lost its advantage in evolutionary selection.

The amount of monitoring is high during the period after rule adoption (Figure 4). This high level is caused by the fact that agents have a genetic value for monitoring which is drawn from a uniform distribution in the initial population. Although the energy use of monitoring seems to be rather low, it prevents agents with a high probability of monitoring to produce many offspring. The individual benefits of monitoring do not compensate the cost of extra energy use during monitoring. Therefore, the fraction of monitoring decreases in time but monitoring as well as rule breaking never disappears (see Weissing and E. Ostrom, 1991 for a similar finding derived from a game theoretical model of monitoring and rule breaking).

One of the main problems of agent-based computational models is the analysis and representation of the enormous amount of data from the simulation runs. In Figure 5, the population size for the 50 runs is depicted. In most cases, the rule is adopted and the system converges towards a cooperative optimum. The timing of adopting the rule differs between the simulation runs (Figure 6). In 15% of the cases, the rule is adopted in the first period of scarcity, while in about 75% of the cases, the rule is adopted in the second period of scarcity. In a few cases the system collapses entirely. Note from Figures 1-4 that the second period of scarcity is more severe than the first one. During this period the agents may not be able to survive the crises, even after a rule is adopted. The cure can come too late to heal the system. The regularity in the timing of the collapses is caused by the fixed mutation level of the agents and the logistic growth function of the resource. Whether a new collapse occurs, depends on the implementation of the candidate rule restricting energy consumption and therefore population size.

For a number of conditions and assumptions, we have examined the consequences of introducing the candidate rule, as well as the monitoring and reinforcement. For each experiment 50 simulations have been performed. Statistics of the system at time step 3000, as well as of the time step of adoptation of the rule, are the summary statistics characterizing an experiment and reported Table 2.

Configurations of the Torus

The size of the lattice influences the vulnerability of the system and the timing of adopting the rule. A smaller lattice has a smaller number of initial agents and a lower population before the situation becomes critical. Due to the smaller reservoir of heterogeneity among the agents, the system becomes vulnerable to disturbances, such as resource scarcity. If the rule is not adopted after the first crisis, the system more likely to collapse. A consequence of a small torus is the reduced probability that agents can survive due to little spots of biomass left over or because agents collected enough energy to survive a period of scarcity. When the rule is accepted, the monitoring rate decreases. Since monitoring rates are assumed to be fixed for an agent, a lower population size reduces the speed of evolution and thus of adjusting the monitoring rate. In case of a lattice with 900 cells, it may take a few crises before the rule is adopted. Due to the large initial variety of agents, agents are frequently meeting other agents with new combinations of symbols. Therefore, it takes longer to build up enough trust in other agents to support the proposed rule.

When agents only have 4 neighboring cells (south, north, east and west), the system becomes vulnerable. In none of the simulations is a rule is adopted. In 70% of the cases, the system collapses during the 3000 time steps. The reason for this vulnerability is the reduced social space, which limits the number of social contacts, and the ability to build up mutual trust. The opposite happens, when the agents have 24 neighboring cells (a Moore neighborhood with radius 2). Agents have the opportunity to meet many agents each time step, which accelerates mutual trust relationships. The rule is adopted in all cases directly at the beginning of the first crisis. The monitoring rate is reduced since an agent can monitor more agents per time step due to the larger social space.

Motivation

When the threshold T_y is lowered, the agent will tolerate a lower level of resource scarcity before becoming motivated to vote for regulation. This delay in motivation causes a decrease of the performance of the system. The percentage of collapses increases, and when the rule is adopted, it will be later than in the default case. But when the rule is adopted, the performance in terms of monitoring and sanctioning is similar to the default case, as would be expected.

An increased level of m_{min} will increase the time the agents have to build up motivation. This delays the moment of adopting the rule, and increases the chance of collapse. Similar arguments hold for increasing the value of $O(Y)_{min}$, the minimum level of trust in neighbors.

Trust

In the default case, the interactions between agents in order to build up mutual trust was costless. One could think of the act of saying hi or waving one's hand as an example of this type of costless interaction. When positive costs are included (□>0; □>0) agents who trust other agents lose energy in maintaining social contacts. One can think of this kind of interaction as involving the gift of small portions of "party" or "ritual" food that cost more than their energy benefit. Due to the extra energy use, the population growth is slower over time. Since trusting others reduces individual fitness, the average trust in others is lower than the default case. This slows down the introduction of the rule, and increases the probability of a collapse. In case □=1 and □=0.5, the probability of collapse is higher than 50%. When the system does not collapse, no rule is implemented in 40% of these cases, and the average population is low (around 30 agents). In 10% of these cases, a rule is implemented leading to a doubling of the population. The results do not significantly change for alternative values of □.

Rule Definition

When the candidate rule does not include a provision that agents have to pay an extra fee when caught for breaking a rule, thus $s_c = 0$, the percentage of rule breaking remains higher and the resource is more depleted than when these provisions are part of the rule. An increased percentage of the cases collapsed due to the high level of rule breaking just after the introduction of the rule. The population level and the mutual trust level of the runs where the system did not collapse are high which is caused by reduced probabilities of sanctioning rule breakers and the lower rate of introducing new agents.

Increasing the costs of monitoring, c_m, reduces the level of monitoring, and increases the level of rule breaking. Due to higher monitoring costs, the average population size at the end of the simulations drops by 10% compared to the default case.

In the default case, 70% of the penalty is shared among the population, and the monitoring agent receives 30% of the penalty. When the monitoring agent does not derive any personal reward for monitoring, the monitoring level is lower. When a monitoring agent can keep all of the penalty earnings, it stimulates monitoring. But due to a high level of monitoring, the agents become aware of more rule-breaking than observed in the default case. This reduces mutual trust, and increase the number of agents who do not survive after receiving a sanction (thus reducing the population size).

Carrying Capacity

A high carrying capacity (K=15) increases the population size and the number of mutual interactions. The faster increase of mutual trust leads to a timely adoption of the rule in all cases. The lower monitoring rate is caused due to the higher population size and selection pressure to reduce monitoring costs. However, a lower carrying capacity (K=5) reduces interactions between agents by such an amount that not enough mutual trust is built up to adopt the rule.

Number of Symbols

When agents have only one genetic and one cultural symbol instead of five of each type, the rule is adopted significantly faster. Due to less heterogeneity in the symbols the agents can learn faster who to trust.

Splitting the Torus

By splitting the torus into two, agents initially are allocated in two different regions and build up their own identity. This identity is the combination of cultural symbols that emerge in the two regions. When the agents from one region start to enter the other region, they are strangers and may not be trusted.

When the carrying capacities are equal in both regions, the rule is adopted in most cases (Table 3). However, the population sizes are lower than the default case of one unified torus. The reason is the behavior of the agents is constrained by the boundaries of the area, which reduces the potential available energy. Since the carrying capacities in both regions are the same, and in most cases the rule is accepted, the number of agents moving to the other region after 3000 time steps is low. There is a little difference between the cases where agents accept the rule in the other region of not. When agents do not accept a rule in another region, we found one case where the system collapses, compared with no collapse in the case of accepting the rule.

In some typical experiments we see that in periods of crises a lot of agents move to the other region. Not accepting the rule lead to sanctioning and starvation of agents outside their own region (Figures 7 and 8). But accepting the rule of the other region, reduces the average mutual trust level, due to the higher level of strangers, and this increases the level of breaking the rules.

When the carrying capacities differ, the region with the highest carrying capacities rapidly introduces the candidate rule, and the region with the lower carrying capacities does not. This results in a significantly lower total population. This result can be explained by the different performance of the system for different carrying capacities (Table 2). When a period of scarcity appears, the motivation grows rapidly, and the rule will be accepted. In the region with a low carrying capacity, the population grows slowly and thus remains low. Agents do not build much trust due to the low density and the few interactions. The period of scarcity starts later than in the neighboring region. Due to the low level of mutual trust, the motivation does not reach the threshold level. Many agents move to the other region. This improves the population size, but since they come in a region of strangers, their mutual trust relationship remains low. There is an important difference between accepting the rule of the foreign region of not (Figures 9 and 10). Although the absolute population figures are the same for both cases, the distribution differs substantially between the two groups.

Discussion

As can be seen in reviewing Tables 2 through 5, we are able to construct a large number of artificial worlds which vary in regard to important variables identified in the literature positing various explanations for why some groups of appropriators in are able to design their own rules and govern a common-pool resource sustainably over a long period of time and others do not.

When examining real cases, the number of relatively comparable cases which differ in regard to only one or two potentially significant variables are limited in number and hard to find. The Pacific islands Mangaia and Tikopia, for example, do share many attributes but yet had quite different histories. Mangaia collapsed in a typical overshoot and collapse, while the society at the much smaller Tikopia was able to develop a population control strategy that led to a sustainable development (although one may argue whether infanticide and expulsion are acceptable policies) (Kirch, 1997).

Kirch (1997) argues that the differences in social and geographical scale of these two islands are an important factor for explaining the different histories of the islands. On Tikopia, everyone knew everyone on a face-to-face basis. Mangaia was big enough to develop social clusters, different identities (sets of tags), and a mentality of "us" versus "them". Kirch asserts that it was perhaps the inability of island residents to recognize that what happened in the next tribe's valley was also of concern for all tribes on the Island. Kirch further recommends that the Pacific Islands are a wonderful "natural laboratory" for studying how various combinations of variables affect the likelihood of successful local initiatives to govern their own resource base successfully.

Use of agent-based computational models enables us to examine questions such as those raised by Kirch and others in environments that can be simulated multiple times so that one can assess the likely pattern of consequence of certain initial combination of variables in an evolutionary process over time. Our artificial worlds simulate different histories of agents living on an exotic torus. Like the Pacific islands, there is no possible escape. We are aware of the limitations and the exploratory nature of the current version of the model. The encouraging experimental results, however, provide us further insights to complement findings from a much smaller set of field studies about the possible critical factors associated with self-organization of institutions.

Let us briefly return to the list of factors that have been identified by scholars doing extensive fieldwork as being associated with self-organization (see page 6-7). As we indicated, we would build the first four factors listed into the basic structure of the model. In all of the runs summarized on Table 2-5, agents had accurate information about the condition of the resource, they did not discount the future at all, they were highly dependant on the resource for survival, and they used a collective choice rule (50% plus one) that falls between the extremes of unanimity or control by a few.

When these four factors are combined in the default condition with 1) a highly stable group, 2) zero costs of developing mutual trust, and 3) low monitoring and sanctioning costs, only 6% of the 50 artificial worlds collapsed and the average population size and energy biomass closely approached optimality. It did, however, take an average of 258 time steps before the agents on these artificial worlds voted on a rule to limit harvest levels on every cell. Further, most of these relatively "ideal" systems faced at least two resource crises before they implemented the candidate rule. If the relevant time step for a natural system is only a day, this is a relatively short time frame. But if the relevant time step is a year, that means that on average it take over 250 years before a group gets organized! The really good news, is that once a group is organized they do not suffer further crises and trust levels tend to grow over time as the percent breaking rules falls over time.

We explored the 5th factor – the stability of the group – by splitting the torus in half and creating in essence two populations which could move back and forth. In this setting, we did find that subtle changes in key parameters made substantial differences in outcomes. Immigration of agents from another region decrease the performance of the system, even if the immigrants accept the rule of their new region. If strangers do not accept the rules, more rule breaking happens, leading to higher monitoring costs. If strangers do accept the rules, the agents are confronted with new type of agents with whom they have no experience which will reduce the mutual trust level and increase rule breaking and monitoring costs.

In examining the effect of the size of the resource and group – the sixth factor -- our findings contradict some of the "conventional" wisdom. When a very small lattice is considered, the system is more vulnerable to collapse than when a larger lattice is assumed. If a rule is not adopted during the first resource crises, small systems are more likely to collapse than larger systems.

There are many forms of heterogeneity that can occur in regard to sets of agents using a common-pool resource. We constructed agents that varied in regard to the number of symbols they carried. We found that a high level of heterogeneity delays building up mutual trust relationships and the adoption of a rule regulating harvest. When agents carried only two symbols, none of the artificial worlds collapsed versus the range of collapse rates shown in the last column of Table 2 for more heterogeneous agents. Heterogeneity does not have to be a bottleneck in adopting a rule, however, as long as mutual trust relationships can be created. For building up these mutual trust relationships it is important to meet many agents. The larger the social space (neighborhood definition), and the larger the total carrying capacity of the system (size of torus, carrying capacity of a cells) the faster mutual trust relationship are built up, and the faster the rule is adopted.

When the torus is split in two regions, ‘us’ versus ‘them’ can evolve. The performance decreases, especially when the ecosystem dynamics differ between the two regions, which is in line with the histories of Mangaia and Tikopia.

The bottom line of our analysis in regard to heterogeneity is that unlike often argued, heterogeneity within a population does not have to be the bottleneck for self governance. Suitable conditions of the learning capacity of agents, and the structure and frequency of social interactions can stimulate creation of a trusting society. Within such trusting societies proposed regulation to solve common-pool resource dilemmas are more likely to be implemented and less costly to maintain.

Further, we find that levels of trust are important forms of social capital and groups that have made costly investments in building mutual trust are more likely to utilize their resource at close to optimal conditions over a long time frame. And, accurate and low-cost monitoring and sanctioning systems are also an important factor in the evolution of system performance over time.

References

Arrow, K. (1974) *The Limits of Organization.* New York: Norton Press.

Axelrod, R. (1984) *The Evolution of Cooperation*. New York: Basic Books.

Axelrod, R. (1997) *The Complexity of Cooperation. Agent-Based Models of Competition and Collaboration.* Princeton, NJ: Princeton University Press.

Bardhan, P. (1993) Rational Fools & Cooperation in a Poor Hydraulic Economy. In *Development, Welfare & Ethics: A Festschrift for Amartya Sen*, edited by K. Basu. Oxford: Clarendon Press.

Bardhan, P. and J. Dayton-Johnson (2002), Unequal Irrigators: Heterogeneity and Commons Management in Large-Scale Multivariate Research, in E. Ostrom et al. (eds.) *The Drama of the Commons*, Washington, D.C.: National Academy Press, 87-112.

Berkes, F. (ed.) (1989) *Common Property Resources; Ecology and Community-Based Sustainable Development*, London, Belhaven.

Berkes F. (1992), 'Success and Failure in Marine Coastal Fisheries of Turkey', in D.W. Bromley, et al. (eds.), *Making the Commons Work: Theory, Practice, and Policy,* Oakland, CA, ICS Press, 161-182.

Blomquist, W. (1992), *Dividing the Waters: Governing Groundwater in Southern California,* Oakland, CA, ICS Press.

Bond, A.H. and L. Gasser (eds.) (1988) *Readings in distributed artificial intelligence*, San Mateo, CA: Morgan Kaufmann.

Bousquet, F., I. Bakam, H. Proton and C. Le Page (1998) Cormas: Common-Pool Resources and Multi-Agent Systems. *Lecture Notes in Artificial Intelligence* **1416**: 826-837.

Bowles, S. (1998) Endogenous Preferences: The Cultural Consequences of Markets and other Economic Institutions, *Journal of Economic Literature*, 36: 75-111.

Bromley, D.W., D. Feeny, M. McKean, P. Peters, J. Gilles, R. Oakerson, C.F. Runge and J. Thomson (eds.) (1992), *Making the Commons Work: Theory, Practice, and Policy,* Oakland, CA, ICS Press.

Burger, J., E. Ostrom, R.B. Norgaard, D. Policansky and B.D. Goldstein (eds.) (2001) *Protecting the Commons*, Washington, D.C., Island Press.

Clark, C. (1990) *Mathematical Bioeconomics: The optimal management of renewable resources*. John Wiley & Sons Inc.

Cordell, J.C. and M.A. McKean (1992), 'Sea Tenure in Bahia, Brazil', in D.W. Bromley *et al.* (eds.), *Making the Commons Work: Theory, Practice, and Policy,* Oakland, CA, ICS Press, 183-205.

Cohen, M.D., R.L. Riolo and R. Axelrod (2001) The Role of Social Structure in the Maintenance of Cooperative Regimes, *Rationality and Society*, 13: 5-32.

Conte, R. R. Hegselmann, and P. Terna (Editors.) (1997) *Simulating Social Phenomena*. Lecture Notes in Economics and Mathematical Systems 456, Springer, Berlin.

Ellison, G. (1993) Learning, Local Interaction and Coordination, *Econometrica*, 61, 1047-1071.

Epstein J. and R. Axtell (1996), *Growing Artificial Societies: Social Science from the Bottom Up*, MIT Press/Brookings, MA.

Feeny, D., F. Berkes, B.J. McCay, and J.M. Acheson (1990), 'The Tragedy of the Commons: Twenty-Two Years Later', *Human Ecology* **18**(1), 1-19.

Ferber, J. (1999) *Multi-Agent Systems : an Introduction to Distributed Artificial Intelligence*. Reading, MA, Addison-Wesley.

Franzen, A. (1994) Group Size Effects in Social Dilemmas: A Review of the Experimental Literature and some New Results for One-Shot N-PD Games. In *Social Dilemmas and Cooperation*, edited by U. Schulz, W. Albers and U. Mueller. Berlin: Springer-Verlag.

Fujita, M., Y. Hayami and M. Kikuchi (1999). The Conditions of Farmer participation in Irrigation Management: A Cross-section Analysis for the Philippines. Paper prepared in the Social Science Division as part of the IRRI-Japan Shuttle Project

Gallant, S.I. (1993) *Neural Network Learning and Expert Systems*, MIT Press, Cambridge.

Gallarotti, G.M. (2001) The Rise of the Classical Gold Standard: The Role of Focal Points and Synergistic Effects of Spontaneous Order. *Humane Studies Review* 13 (3):1-21.

Gambetta, Diego and Bacharach, Michael (2001) 'Trust in Signs', in Cook, K. (ed.) *Social Structure and Trust* (New York, NY: Russell Sage Foundation).

Gardner, M. (1970), *The Fantastic Combinations of John Conways New Solitaire Game Life*, Scientific American, 223, 120-123.

Gibson, Clark C. (2001) "Forest Resources: Institutions for Local Governance in Guatemala." In Joanna Burger et al., eds. *Protecting the Commons: A Framework for Resource Management in the Americas,* 71-89. Washington, D.C.: Island Press.

Gilbert, N. and K.G. Troitzsch (1999) *Simulation for the Social Scientist.* Open University Press: London, UK.

Gilles, J.L. and K. Jamtgaard (1981), 'Overgrazing in Pastoral Areas: The Commons Reconsidered,' *Sociologia Ruralis* **2**, 335-358.

Gintis, Herbert. 2000. *Game Theory Evolving. A Problem-Centered Introduction to Modeling Strategic Interaction.* Princeton: Princeton University Press.

Grima, A.P.L..and F. Berkes (1989), 'Natural Resources: Access, Rights to Use and Management', in F. Berkes (ed.), *Common Property Resources; Ecology and Community-Based Sustainable Development*, London, Belhaven, 33-54.

Hardin, G. (1968) The tragedy of the commons, *Science*, 162: 1243-1248

Hegselmann, R. (1998) Modeling Social Dynamics by Cellular Automata, in: W.B.G. Liebrand, A. Nowak, and R. Hegselmann (Editors.) *Computer Modeling of Social Processes*. SAGE Publications, London, pp. 37-64.

Janssen, M.A. and D.W. Stow (2002) Ecosystem management and the evolution of rules: an immune system perspective, in review.

Kirch, P.V. (1997) Microcosmic Histories: Island Perspective on "Global" Change*, American Anthropologist*, 99(1): 30-42.

Lam, Wai Fung. 1998. *Governing Irrigation Systems in Nepal: Institutions, Infrastructure, and Collective Action.* Oakland, CA: ICS Press.

Libecap, G.D. (1995) The Conditions for Successful Collective Action. In *Local Commons and Global Interdependence. Heterogeneity and Cooperation in Two Domains*, edited by R. O. Keohane and E. Ostrom. London: Sage Publications.

Liebrand, W.B.G., A. Nowak and R. Hegselmann (Editors.) (1998) *Computer Modeling of Social Processes*. SAGE Publications, London.

Macy, M. and J. Skvoretz (1998) Evolution of Trust and Cooperation between Strangers: A Computational Model, American Sociological Review, 63: 638-660.

Maes, P. (1995), *Modelling Adaptive Autonomous Agents*, in Langton, C.G. (ed.), Artificial Life: an overview, The MIT Press, Cambridge/London.

McGinnis, Michael, ed. 2000. *Polycentric Games and Institutions: Readings from the Workshop in Political Theory and Policy Analysis.* Ann Arbor: University of Michigan Press.

Mehrota, K, C.K. Mohan and S. Ranka (1997) *Elements of Artificial Neural Networks*, MIT Press, Cambridge.

Miller, J. H., C. T. Butts, and D. Rode. 2002. "Communication and Cooperation," *Journal of Economic Behavior and Organization*. 47 (2) (February): 179-196.

Morrow, C.E., and R.W. Hull (1996) Donor-Initiated Common Pool Resource Institutions: The Case of the Yanesha Forestry Cooperative. *World Development* 24 (10):1641-1657.

Nowak, M.A. and K. Sigmund (1998) Evolution of indirect reciprocity by image scoring, *Nature*, 393: 573-577.

Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action*, New York, Cambridge University Press.

Ostrom, E., T. Dietz, N. Dolsak, P. Stern, S. Stonich, and E. Weber, eds. (Committee on the Human Dimensions of Global Change). 2002. *The Drama of the Commons.* Washington, D.C.: National Research Council, National Academy Press.

Ostrom, E., R. Gardner, and J.M. Walker (1994), *Rules, Games, and Common-Pool Resources*, Ann Arbor (MI), University of Michigan Press.

Ostrom, V. and E. Ostrom (1977), 'A Theory for Institutional Analysis of Common Pool Problems', in G. Hardin and J. Baden (eds.), *Managing the Commons*, San Francisco, W.H. Freeman, 157-172.

Ostrom E. and J. Walker (Forthcoming) "Introduction," in E. Ostrom and J. Walker (eds.), *Trust, Reciprocity, and Gains from Association: Interdisciplinary Lessons from Experimental Research*. New York: Russell Sage Foundation.

Parker, D.C, S.M. Manson, M.A. Janssen, M.J. Hoffmann, and P. Deadman (2002), Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review, in review.

Pathak, N. And A. Kothari. 2002. "Communities and Biodiversity: Lessons for Governance from South Asia," Pune, India: International Institute of Environment and Development, Working Paper.

Princen, Thomas. 1998. "From Property Regime to International Regime: An Ecosystems Perspective. *Global Governance*. 4: 395-413

Riolo, R.L., M.D. Cohen and R. Axelrod (2001) Evolution of cooperation without reciprocity, Nature, 414: 441-443.

Rousseau, D.M., S.B. Sitkin, R.S. Burt, and C. Camerer (1998) Not So Different After All: A Cross-Discipline View of Trust. *Academy of Management Review* 23:393-404.

Seabright, P. (1993) Managing Local Commons: Theoretical Issues in Incentive Design. *Journal of Economic Perspectives* 7 (4):113-134.

Sethi, R.and E. Somanathan (1996), 'The Evolution of Social Norms in Common Property Resource Use', *The American Economic Review* **86**(4), 766-788.

Singh, K. (1994), *Managing Common Pool Resources: Principles and Case Studies*, New Delhi, Oxford University Press.

Singh, K. and V. Ballabh (1996), *Cooperative Management of Natural Resources*, New Delhi, Sage.

Steels, L. (1995), *The Artifical Life Roots of Artifical Intelligence*, in Langton, C.G. (ed.), Artificial Life: an overview, The MIT Press, Cambridge/London.

Tesfatsion, L. (2001) Introduction to the special issue on agent-based computational economics, *Journal of Economic Dynamics & Control*, 25: 281-293.

Varughese, George, and Elinor Ostrom (2001) "The Contested Role of Heterogeneity in Collective Action: Some Evidence from Community Forestry in Nepal." *World Development* 29(5) (May): 747-765.

Weinstein, M.S. (2000) Pieces of the Puzzle: Solutions for Community-Based Fisheries Management from Native Canadians, Japanese Cooperatives, and Common Property Researchers. *Georgetown International Environmental Law Review* 12 (2):375-412.

Weissing, F. and E. Ostrom (1991) Irrigation Institutions and the Games Irrigators Play: Rule Enforcement without Guards, in Reinhard Selten (ed.) *Game Equilibrium Models II: Methods, Morals, and Markets*, Berlin: Springer-Verlag: 188-262.

Wilson, P.N., and G.D. Thompson (1993). Common Property and Uncertainty: Compensating Coalitions by Mexico's Pastoral *Ejidatorios*. *Economic Development and Cultural Change* 41 (2):299-318.

Table 1: Parameter values of the model for the default case

Parameter	Value	Explanation
Resource		
r	0.2	growth rate of the resource
K	10	carrying capacity of a cell
Agent		
q	[0,1]	fraction of a cell eaten by the agent
e_{max}	4	maximum absolute level of energy consumption per time step
c	2	energy metabolism per timestep
f	100	embodied energy level above which the agent will derive offspring
#symbols (genetic)	5	number of genetic symbols
#symbols (cultural)	5	number of cultural symbols
□	[0,1]	habit to break rules
p_s	[0,1]	probability of an agent to monitor and sanction
p_m	0.05	probability of mutation of a 'gene' during offspring generation
Trust		
□	2	parameter of skewness in starting to play basic trust games
□	0	cost of starting an interaction
□	0	reward of receiving an interaction signal
□	0.1	updating rate neural network of trust expectations
Rule-definition		
x_{min}	4	cell level of energy below which no harvesting is allowed
s_e	0.2	initial sanction rate
g	0.7	fraction of penalty allocated to the group
c_m	0.5	energy cost per timestep of monitoring a neighbor
Motivation		
T_y	10	expected life without no energy intake as a threshold level for motivation
□	0.1	decay rate of motivation
m_{min}	0.5	minimum level of motivation before an agent is motivated to vote
□	0.05	decay rate of trust in neighbors
$O(Y)_{min}$	0.5	minimum level of trust in neighbors before being motivated

Table 2: Statistics of the experiments when the torus is one region. The values are the average of non-collapsed runs in time step 3000. A run is collapsed when the population at time step 3000 is zero. The value between brackets is the standard deviation.

experiment	population size	energy biomass	trust	% monitoring	% breaking rules	time step adoption	% collapse
default	92.8 (3.7)	1915.1 (28.6)	98.4 (1.7)	9.2 (5.8)	2.7 (1.5)	257.8 (82.5)	6%
100 cells	22.3 (1.9)	484.9 (21.4)	97.6 (5.3)	17.9 (14.4)	3.1 (4.3)	189.3 (269.4)	16%
900 cells	210.6 (6.1)	4306.9 (39.3)	98.9 (0.7)	9.4 (6.3)	2.3 (1.1)	296.7 (57.4)	0%
4 neighbors	39.6 (30.7)	1695.0 (1242.6)	74.6 (16.7)	0	0	0	70%
25 neighbors	93.3 (3.7)	1910.6 (27.9)	99.7 (0.3)	3.7 (3.2)	2.0 (1.6)	63.7 (6.9)	0%
$T_y = 8$	93.2 (3.8)	1918.0 (28.7)	98.7 (1.2)	11.2 (6.3)	2.2 (1.4)	303.7 (80.9)	28%
$T_y = 6$	92.9 (3.1)	1917.2 (45.5)	98.0 (2.2)	9.5 (6.7)	3.1 (2.1)	434.7 (173.1)	64%
$m_{min} = 0.6$	92.8 (3.5)	1921.0 (30.1)	98.5 (1.3)	12.2 (8.4)	2.3 (1.5)	335.0 (105.8)	16%
$m_{min} = 0.7$	89.6 (12.0)	1985.5 (236.4)	98.1 (3.2)	9.7 (6.6)	2.7 (2.0)	478.6 (305.7)	64%
$O(Y)_{min} =0.6$	93.1 (3.9)	1914.5 (29.2)	98.5 (1.6)	9.7 (7.1)	2.3 (1.6)	311.7 (130.7)	18%
$O(Y)_{min} =0.7$	92.6 (4.0)	1913.1 (28.0)	97.2 (4.2)	10.7 (6.7)	2.9 (1.7)	330.8 (103.3)	34%
□=0.5; □=0.25	68.8 (2.2)	2094.6 (31.3)	98.1 (2.0)	12.6 (8.3)	2.5 (2.1)	393.0 (218.4)	18%
□=1; □=0.5	29.2 (22.7) 51.0 (5.9)	2316.8 (1348.7) 2483.5 (263.0)	74.2 (10.8) 86.4 (10.7)	0 31.3 (15.7)	0 3.1 (2.5)	0 1822.8 953.2	50% 10% rule adoption
□=1	93.7 (3.6)	1913.4 (23.8)	98.3 (1.5)	10.4 (7.9)	2.4 (1.7)	218.9 (90.6)	12%
□=4	93.7 (3.5)	1914.4 (27.4)	98.9 (1.3)	9.9 (8.0)	2.5 (1.5)	265.2 (58.5)	10%
$s_c = 0$	92.5 (3.0)	1884.0 (37.2)	99.3 (0.8)	11.2 (7.2)	4.6 (2.8)	247.8 (86.8)	20%
$c_m = 5$	84.8 (7.9)	1939.2 (92.0)	97.3 (5.9)	3.1 (3.2)	4.6 (5.1)	244.3 (79.4)	18%
g = 1	94.0 (2.5)	1909.7 (22.0)	98.7 (1.4)	8.6 (4.6)	2.8 (1.7)	248.5 (95.9)	4%
g = 0	84.3 (4.5)	1994.1 (37.5)	96.1 (2.7)	36.2 (13.0)	2.3 (1.8)	238.8 (84.4)	2%
K=5	19.2 (11.0)	593.6 (413.7)	73.3 (14.6)	0	0	0	12%
K=15	141.2 (4.1)	2664.7 (16.7)	99.2 (1.7)	6.5 (4.3)	2.2 (1.1)	80.9 (4.6)	0%
2 symbols	94.6 (2.5)	1908.2 (18.0)	99.8 (0.3)	7.5 (5.2)	2.5 (1.4)	105.4 (53.5)	0%

Table 3: Statistics of the experiments when the torus is split into two regions. The values are average of non-collapsed runs of the whole torus in time step 3000. A run is collapsed when the population at time step 3000 is zero. The value between brackets is the standard deviation. In the last column, the values between brackets denote the additional percentage of cases where the system did not collapsed, but where one of the two regions was not able to introduce the rule).

experiment	population size	energy biomass	trust	% monitoring	% breaking rules	% collapse
K=10 and accepting rules	84.8 (11.4)	1859.8 (193.0)	94.5 (2.5)	14.5 (10.3)	3.2 (2.4)	0% (10%/6%)
K=10 and not accepting rules	90.9 (3.9)	1950.7 (71.1)	97.6 (1.4)	16.0 (8.8)	2.9 (1.6)	2%
K=5/15 and accepting rules	79.9 (6.6)	1652.1 (263.0)	94.9 (3.0)	7.8 (6.0)	3.6 (2.8)	0%
K=5/15 and not accepting rules	80.1 (6.2)	1701.0 (287.3)	95.0 (2.6)	8.7 (5.7)	4.1 (2.7)	0%

Table 4: Statistics of the experiments when the torus is split into two regions. The values are average of non-collapsed runs of part A of the torus in time step 3000. A run is collapsed when the population at time step 3000 is zero. The value between brackets is the standard deviation.

experiment	Population size	population in other region	energy biomass	trust	% monitoring	% breaking rules	time step adoption
K=10 and accepting rules	42.9 (7.0)	0.8 (2.4)	922.6 (125.8)	94.2 (4.1)	12.8 (10.8)	2.8 (2.6)	412.9 (494.9)
K=10 and not accepting rules	45.3 (2.8)	0.2 (0.5)	974.9 (38.8)	97.3 (1.8)	17.4 (14.9)	2.8 (2.2)	406.0 (450.6)
K=5 and accepting rules	27.6 (18.2)	17.0 (15.8)	326.4 (253.3)	79.0 (2.7)	0	3.5 (3.9)	-
K=5 and not accepting rules	11.9 (5.6)	1.2 (1.7)	377.7 (274.8)	70.2 (27.7)	0	9.7 (14.7)	-

Table 5: Statistics of the experiments when the torus is split into two regions. The values are average of non-collapsed runs of part B of the torus in time step 3000. A run is collapsed when the population at time step 3000 is zero. The value between brackets is the standard deviation.

experiment	population size	population in other region	energy biomass	trust	% monitoring	% breaking rules	time step adoption
K=10 and accepting rules	41.9 (6.9)	0.9 (3.0)	937.2 (117.7)	94.5 (2.8)	15.9 (15.9)	3.9 (5.6)	349.4 (409.6)
K=10 and not accepting rules	45.6 (2.8)	0.1 (0.4)	975.8 (48.8)	97.8 (1.8)	15.4 (13.2)	3.0 (2.7)	328.7 (428.5)
K=15 and accepting rules	52.3 (14.9)	0	1325.8 (39.5)	97.9 (1.9)	10.6 (6.9)	3.8 (3.9)	77.5 (7.3)
K=15 and not accepting rules	68.2 (2.5)	0	1323.3 (32.1)	97.5 (2.2)	10.1 (6.4)	3.1 (2.2)	78.1 (7.2)

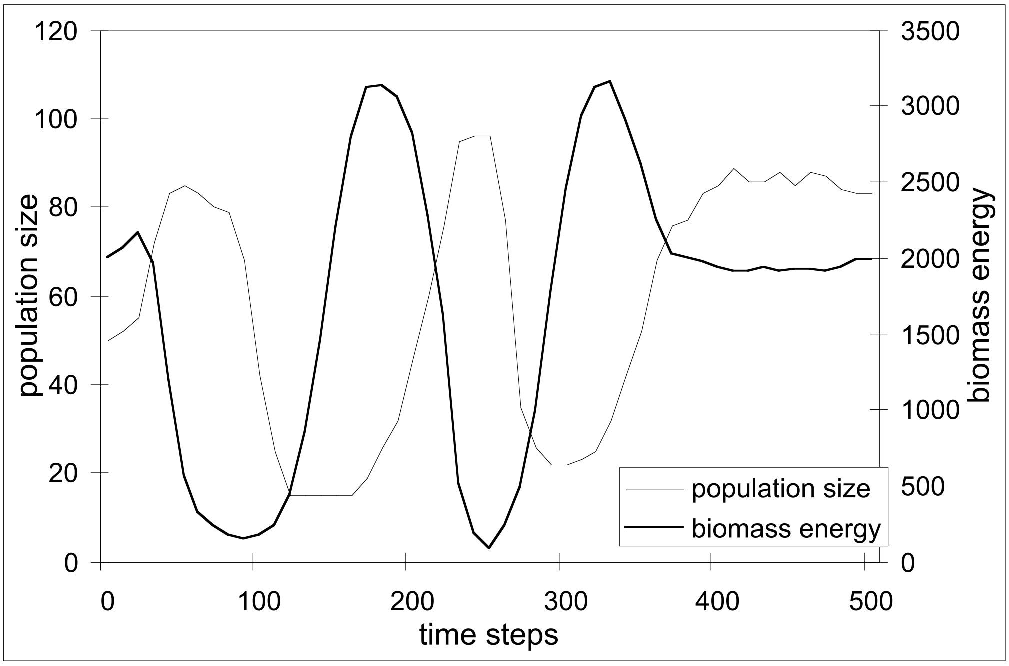

Figure 1: Population size and total biomass energy of a typical default run for the first 500 time steps.

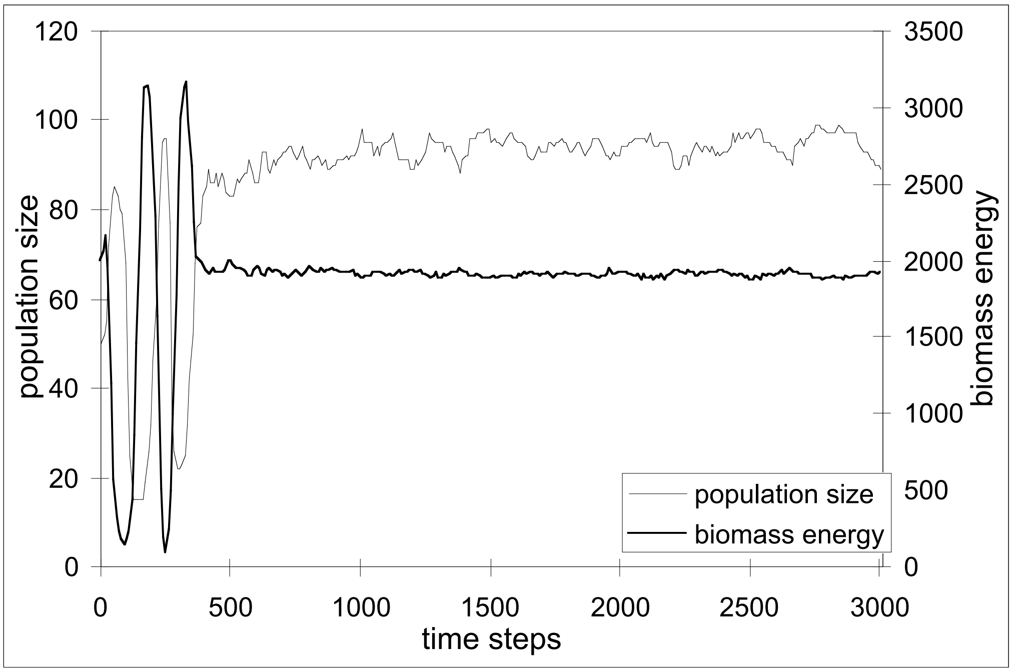

Figure 2: Population size and total biomass energy of a typical default run.

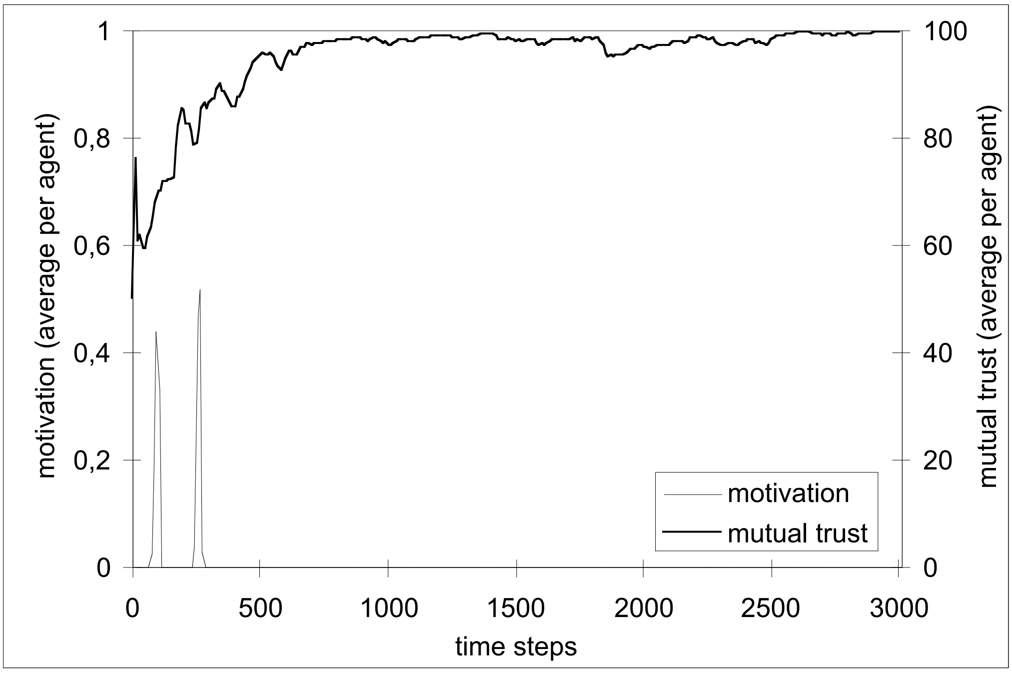

Figure 3: Motivation and mutual trust, both average per capita, of a typical default run.

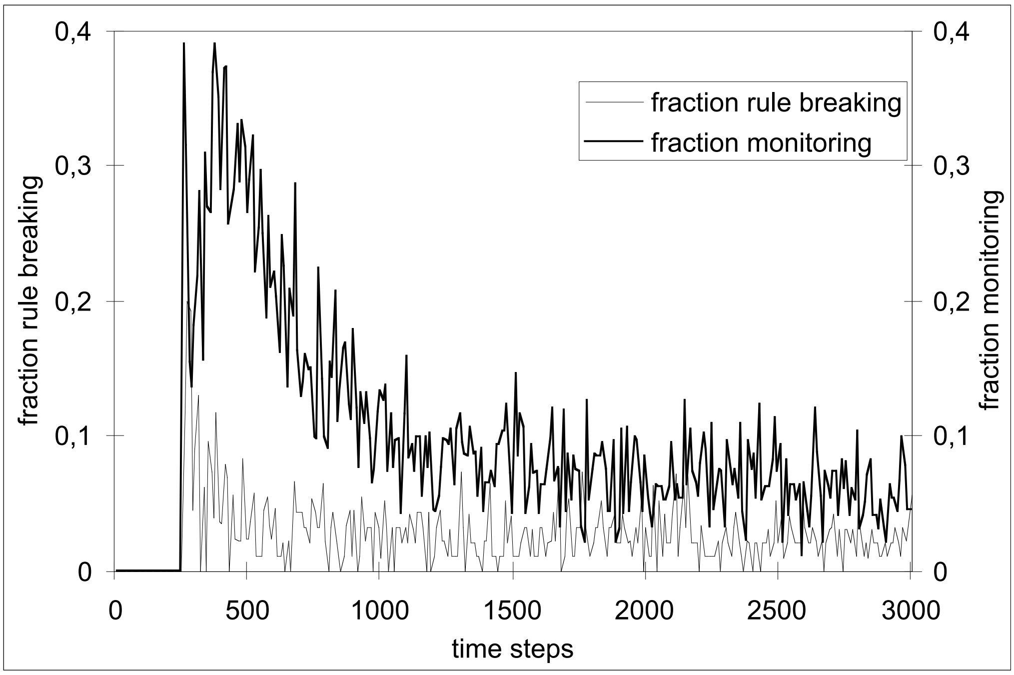

Figure 4: Fraction of the population breaking the rule, or monitoring the neighborhood of a typical default run.

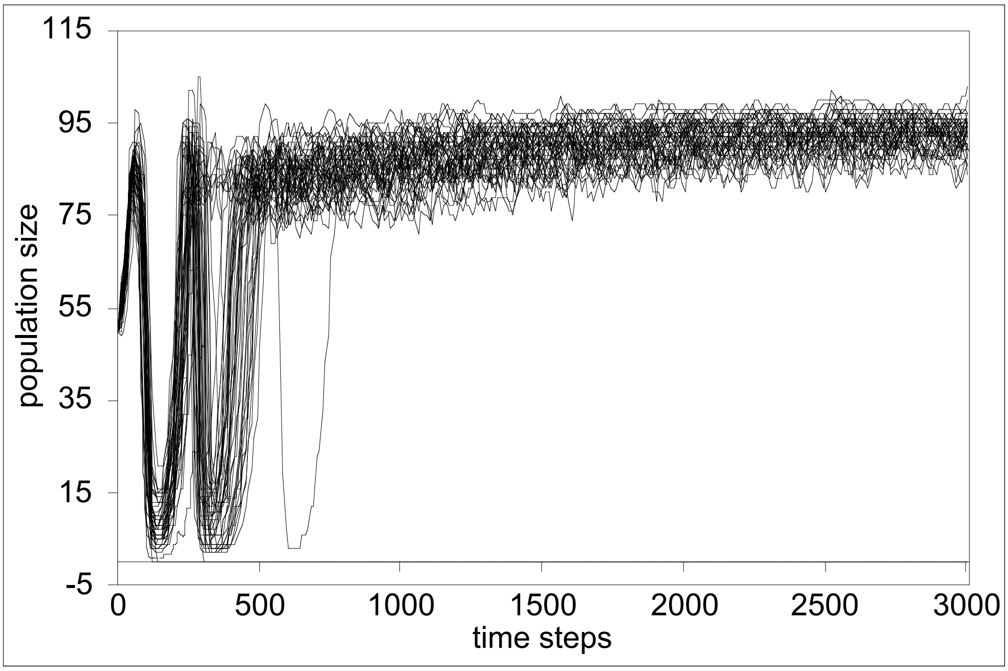

Figure 5: Population size for 50 different default runs.

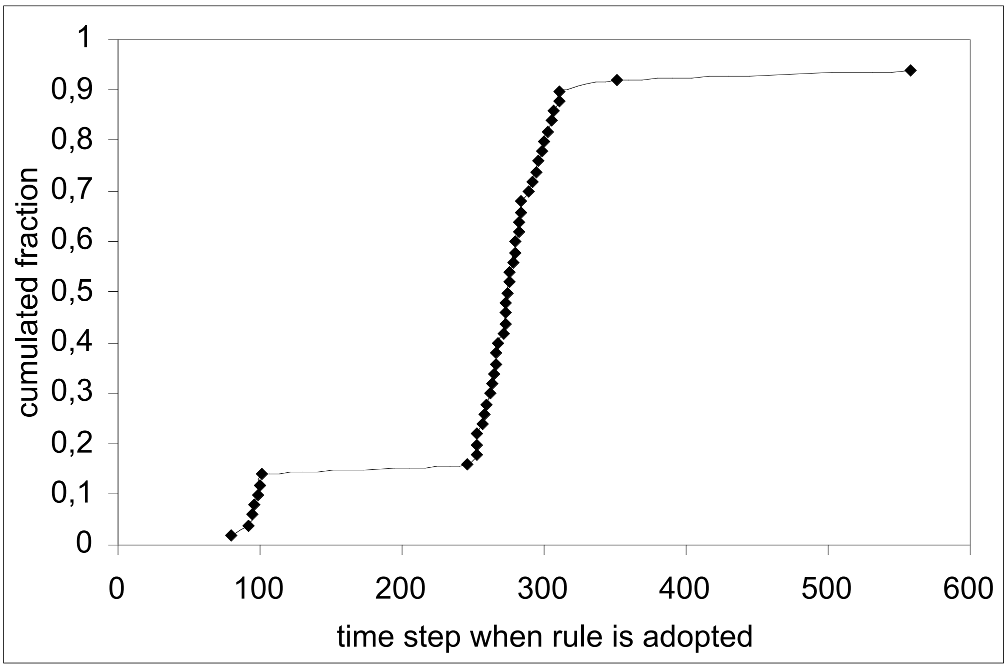

Figure 6: The time step at which the rule is adopted, starting with the fastest adoption. The figure shows different waves of adoption.

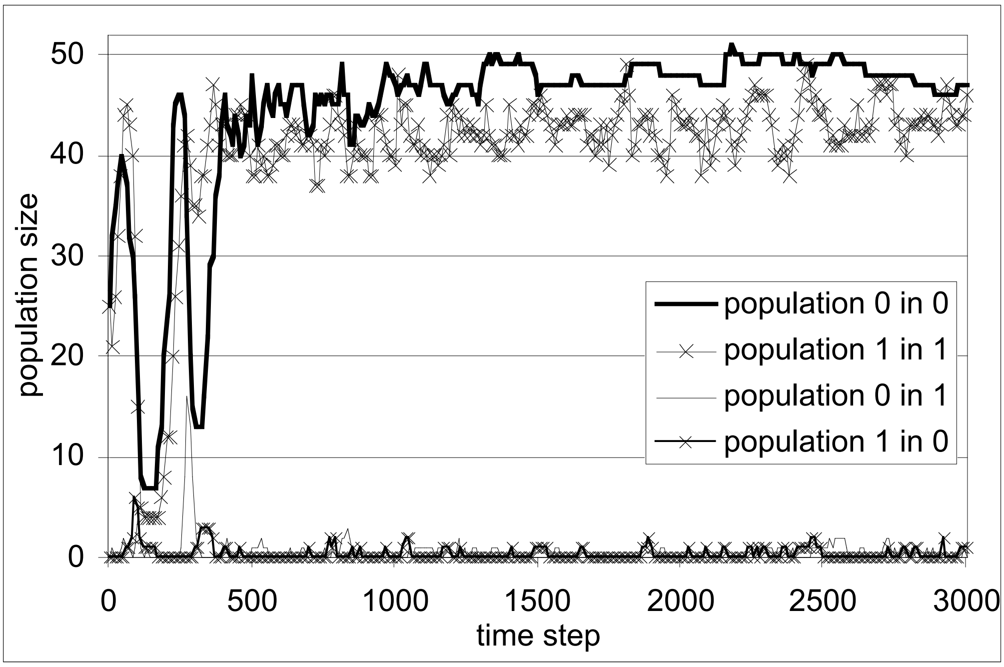

Figure 7: Population levels when torus is split in two regions with equal carrying capacities, and the agents accept the rules in other region. The different lines refer to population size of agents born in region i and located in region j.

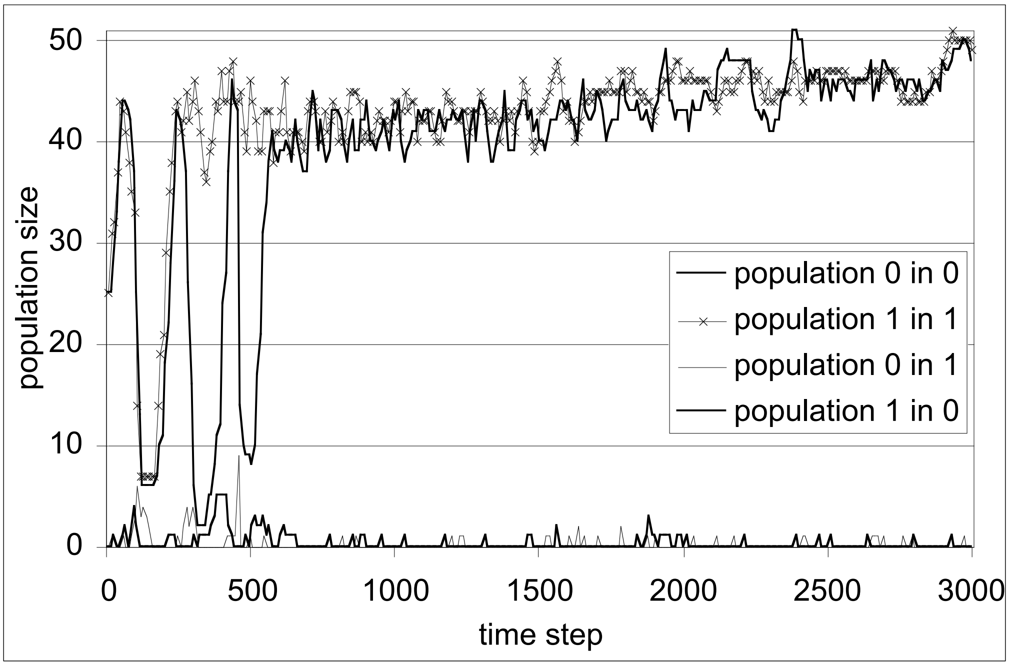

Figure 8: Population levels when torus is split in two regions with equal carrying capacities, and the agents do not accept the rules in the other region. The different lines refer to population size of agents born in region i and located in region j.

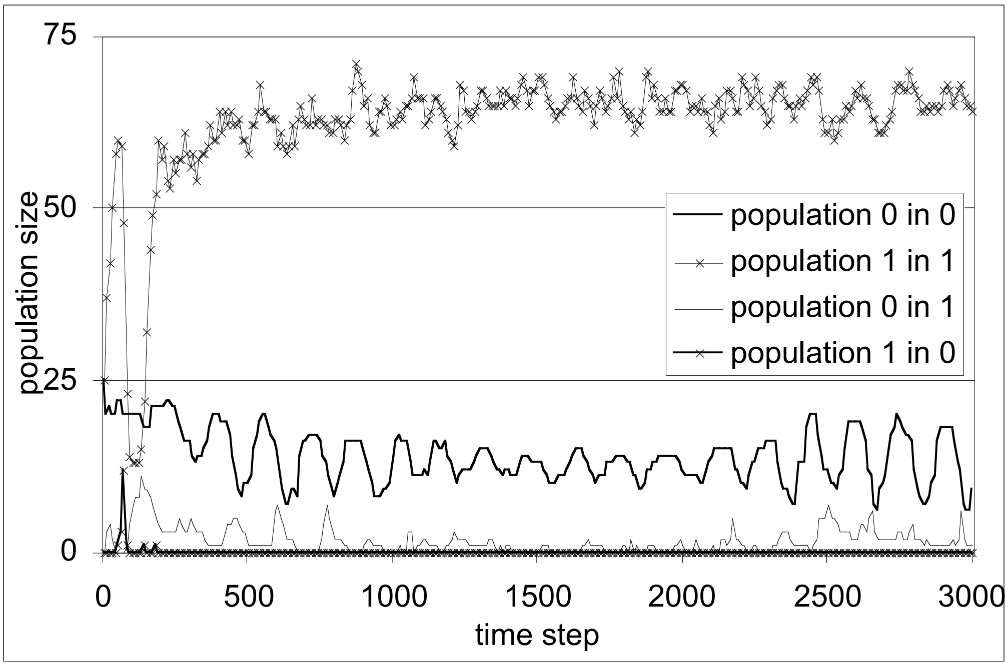

Figure 9: Population levels when torus is split in two regions with unequal carrying capacities (region 0: K=5; region 1: K=15), and the agents accept the rules in other regions. The different lines refer to population size of agents born in region i and located in region j.

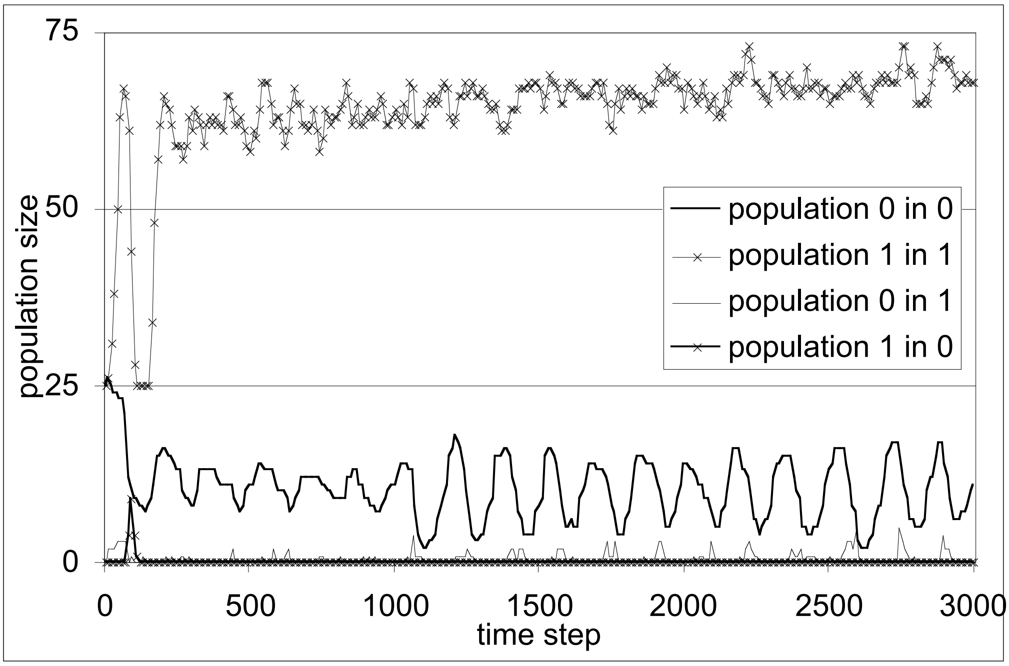

Figure 10: Population levels when torus is split in two regions with unequal carrying capacities (region 0: K=5; region 1: K=15), and the agents do not accept the rules in other regions. The different lines refer to population size of agents born in region i and located in region j.