Heuristics to characterise human behaviour in agent based models

Claudia Pahl-Wostl and **Eva Ebenhöh**

Institute of Environmental Systems Research, University of Osnabrück, Germany

Abstract: Human behaviour is one of the key factors to understand the causes for common pool resource problems and to develop policies to promote more sustainable resource management regimes. Agent based models can help to investigate the role of important processes in this respect such as factors determining the degree of trust and cooperation in a group. We have chosen a pragmatic approach to represent human behaviour by assuming that agents can be characterised by a set of attributes and their behaviour can be described by a set of simple decision heuristics. Individual agents differ in their importance of attributes (e.g. fairness, cooperativeness, trust), in their rules how to choose a heuristic, and in their responses to social interactions. The assumptions are tested by using data from experimental economics describing the behaviour of players in simple games dealing with resource allocation. A set of specific attributes and heuristics was derived by analysing data from different games. The plausibility and generality of the behavioural model is tested by applying it to different data sets from different games. We expect from these simulations insights into behavioural patterns that determine processes of social learning and negotiation. The modelling approach will be applied and tested with data from case studies where actors make decisions in a real world context of dealing with a resource management problem.

Keywords: Heuristics, modelling human behaviour, experimental economics, cooperation, fairness

1. INTRODUCTION

As early as 1968 Hardin in his famous article entitled "The Tragedy of the Commons" described a situation where villagers were using a common field to graze their cattle. The commons tended to be overgrazed since each villager would graze to a point where the private costs equalled the benefits, and social costs were neglected. In general, such a situation applies to the problem of 'common pool resources'. And the tragedy of the commons is a typical case of a 'social dilemma' where the maximization of the short-term self-interest of the individual, leaves all participants worse off than feasible alternatives. Each individual faces a trade-off between what is in his or her own short term interest and what is in the broader interest of the community in which he or she lives. A collective version of social dilemmas occurs frequently in the provision and management of public goods and may account for many environmental problems such as the overexploitation and pollution of water resources, arable land and the atmosphere. Hardin's analysis suggested that the only solution to preventing such social dilemmas would be regulation of the commons by a central entity. This would argue in favour of governmental regulation and control as the most promising strategy for dealing with environmental problems and managing public goods. However, in her influential book, Elinor Ostrom (1990) provided evidence that Hardin's analysis did not apply in general and that local communities have efficient ways of self-organizing and self-governance and may also prevent the degradation of resources on the base of voluntary cooperation. Hence, an alternative strategy would explore ways in which governmental intervention and the self-organizing capacity of communities interact and subtly reinforce themselves so as to develop more efficient and enduring resource management regimes. Such a strategy suggests as well a stronger role of participatory approaches in resources management to facilitate collective learning and choice processes. However, approaches to manage common pool resources are still hampered by a lack of understanding the nature of human behaviour.

The prisoners' dilemma and game theoretical approaches have mainly been used to analyse human behaviour in the context of common pool resource problems (Gintis, 2000). The basic underlying assumption has been the profit-maximizing, rational homo economicus. Such approaches provide little potential for analysing the possibilities of cooperation and self-governance. Empirical and experimental evidence show considerable deviations from theoretical assumptions based on homo economicus. In particular, the emerging field of experimental economics has developed a set of games and has collected a rich data base on human action in a number of experimental situations where fairness and equity issues in the allocation of rewards play an important role. The potential for innovation from these insights has not been fully exploited yet. Whereas most approaches to explain behaviour have been based on analytical mathematical approaches and extensions of game theory, more recently agent based modelling has been used to analyse and explain data from economic games (e.g. Duffy, 2001; Deadman, 2000; Ebenhöh and Pahl-Wostl, 2004). Agent based simulation offers a major methodological breakthrough in the ability to investigate the role of different processes determining human behaviour in more detail. The method is also very flexible since it is not linked to a specific disciplinary paradigm (Pahl-Wostl, 2002a).

We have decided to capture essential elements of human behaviour in resource allocation problems by assuming that agent behaviour is guided by heuristics (Gigerenzer and Selten 2001) and agents learn from experience.

2. MODELLING APPROACH

In the chosen modelling approach decisions are not necessarily based on elaborate calculations, but on heuristics, including simple rules humans may follow in making their decisions. Those heuristics depend on different attributes characterizing individual agents. Another very important aspect is agent diversity. Agents differ not only in their individual attribute values, but also in the heuristics they use.

We use data from economic experiments to support our heuristics approach (see section 2.2). In experimental economics various experiments are conducted with human subjects, placed in an artificial, laboratory environment in which they have to solve tasks or play games with or against each other. Usually those games are played anonymously, in order to prevent the formation of a social environment. In these situations, decision making of the human subjects is supposed to be abstracted from a specific context and social interactions. This may be seen as a breach with respect to actual human behaviour in day-to-day situations. However, this method's main advantage is that it produces comparable and reproducible data. The set of attributes presented in the next section were derived from a number of experiments from different games during the process of developing agent based models that can explain observed behaviour.

2. 1 Attributes

So far, our set of attributes includes nine different characteristic traits. All attributes are represented as real numbers in the interval between 0 and 1, where 0 implies that this trait is not important for the agent and 1 implies that it is very important. In our model, the numbers are random numbers, equally distributed, and the attributes are independent from each other.

Cooperativeness: How important is group utility for an agent? A high cooperativeness indicates the willingness to spend individual resources in order to further group resources. Mainly this is associated with an increase in efficiency.

Fairness (concerning others): How important is it for an agent that the other agents get roughly as much as it? This is a purely comparative attribute that does not refer to efficiency increases. Agents with a high fairness are willing to spend money in order to equalize the outcome. Note that this fairness does not yet consider equity considerations in reward allocations.

Conformity: How important is it for an agent to appear to be as others expect it to be? Agents with a high conformity may play fairly because they feel they are expected to, and not because of their own high fairness.

Fairness concerning me: How important is it that the agent's payoff is roughly as high as the other agents' payoffs? Agents with a high fairness concerning me are easily annoyed at being treated unfairly. However, whether or not this annoyance leads to retaliating actions is defined by negative reciprocity.

Positive reciprocity: How important is it to return behaviour that is perceived as fair and cooperative, with fair behaviour? An agent with a high positive reciprocity feels committed to play cooperatively in a second move when the other agents have played cooperatively before.

Negative reciprocity: How much is an agent willing to pay in order to make another agent pay (more)? An agent with a high negative reciprocity feels committed to punish in a second move when the other agents have defected before.

Risk aversion: How risk averse is an agent? An agent with a high risk aversion will not invest anything in a project that yields a high but uncertain return.

Commitment: How important are the decision to be made and previously made agreements for an agent? A player with a high commitment to a project will invest in group utility even when its cooperativeness is low.

Trustworthiness: To which degree does an agent respond to trust placed in him or her by other agents with expected behaviour instead of being opportunistic?

In addition to all these attributes, agents hold expectations about the attribute values of other agents. They learn from observed behaviour about the others' attribute values. These are referred to as expected cooperativeness, expected fairness etc. Note that trust is modelled as expected trustworthiness.

In literature on experimental economics attributes like the ones described above, often appear without being strictly defined. For example, fairness is rarely differentiated into fairness concerning me and fairness concerning others, although the distinction is quite apparent. Likewise, positive and negative reciprocity, are often considered as a single attribute reciprocity ("strong reciprocity" in Fehr and Rockenbach, 2003). Also, names can vary substantially. For example, what we refer to as "fairness concerning me" is essentially the same as "annoyance", expressed by subjects in post-experimental questionnaires in an experiment by Fehr and Gächter, and "negative reciprocity" corresponds to "willingness to punish" in the same experiment (Fehr and Gächter, 2002). Recently Cox especially designed and conducted an experiment to discriminate between trusting, positively reciprocating, and altruistic, other-regarding behaviour (Cox, 2004). Cox uses trust and positive reciprocity quite similar to the attributes presented in this paper. However, we differentiate other-regarding behaviour further into cooperative and fair.

In this attributes approach lies a major difference to other work combining experiments with agent-based models (Duffy 2001, Deadman 2000).

2.2 Heuristics

Heuristics usually make sense only in a concrete decision environment. This is why the heuristics presented here are given as examples for a specific game. For our modelling approach, however, it is also important that the heuristics chosen are more generic to be applicable to a large range of situations in experimental games and empirical case studies. Therefore, the heuristics are kept as simple as possible.

As example for a decision environment consider the following two-player game, taken from an experiment by Fehr and Rockenbach (2003). Both players receive 10 money units (MU). The first player is asked to give any number of his or her money units to the second player. This gift is tripled by the experimenter. Then, the second player may return any number of MU from 0 to the tripled gift to the first player. This is not tripled. When the first player gave the gift, he or she is also asked to indicate, how much he or she would like to receive.

A main result of this experiment is, that most first players do place trust in second players and most second players reciprocate trust with returns greater than 0. First players gave 6.5 MU on average to second players. Second players returned more than 40 percent of the tripled investment to first players. The higher the gift of the first player, the higher was the return by the second player (Fehr and Rockenbach 2003, 138f.). However, some first players gave less than 5 MU (26.5%) and some second players kept everything to themselves (16%).

The first player's first decision may depend on cooperativeness, because of efficiency considerations. Risk aversion may play a role, because a gift of 10 money units may be returned doubled or even tripled, but the return might also be 0. Finally, the expected trustworthiness is an indicator for perceived risk and thus may also be important. The first player's second decision, how much he or she would like to receive, is probably only influenced by fairness concerning others and fairness concerning me.

Table 1: Some heuristics for the 1st player's decision:

A1	gift = 0 MU
A2	gift = 5 MU
A3	gift = 10 MU
A4	gift = cooperativeness * 10 MU
A5	gift = (1-risk aversion) * 10 MU
A6	gift = expected trustworthiness * 10 MU
A7	if (exp. trustworthiness < low limit) gift = 0 MU else if (exp. trustworthiness > high limit) gift = 10 MU else gift = expected trustworthiness * 10 MU
A8	if (expected trustworthiness > some limit AND risk aversion < another limit) gift = cooperativeness * 10 MU
A9	calculate expected return with expected fairness, expected fairness concerning me and expected positive reciprocity for different gifts and take gift with highest expected return.

Although the second player's task is very similar to the first player's first decision (deciding on an amount of money to give to the other player), the attributes needed are different. The second player's decision is not influenced by risk aversion and expected trustworthiness. However, it may be influenced by fairness considerations, both concerning others and concerning me, as well as positive reciprocity and trustworthiness.

Table 2: Some heuristics for the 2nd player's decision:	
B1	return = 0 MU
B2	return = gift
B3	return = 2*gift
B4	return = fairness * 2*gift
B5	return = (1-fairness concerning me) * 2*gift
B6	if (fairness < low limit) return = 0 else if (fairness > high limit) return = 2*gift else return = fairness * 2*gift
B7	if (trustworthiness < low limit) return = 0 else if (trustworthiness > high limit) return = asked return else return = fairness * 2*gift

2.3 Choosing Heuristics

Heuristics should be as simple as possible. In order to do that, we give certain agents certain heuristics according to their attributes, thus moving the case differentiation out of the heuristic (as in A7, B6, and B7). A typical way of doing so would be:

```
if (agents cooperativeness > high limit)
        use cooperative heuristic (A3)
else if (agents cooperativeness < low limit)
        use maximizing heuristic (A9)
else use default heuristic (A7)
```

By this, the agents' attributes determine their decision making behaviour in two ways. The chosen heuristic as well as the actual decision made by the chosen heuristic both depend on the attribute values (see figure 1).

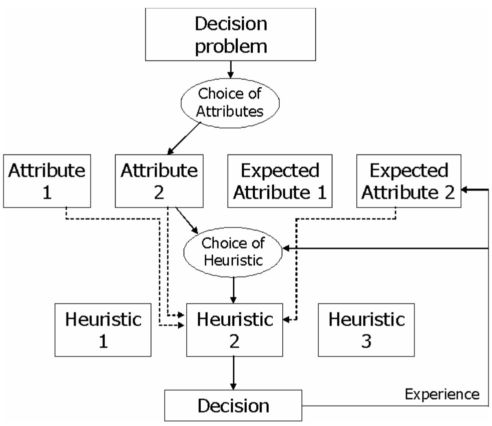

Figure 1: Role of attributes in the decision making process

2.4 Learning

Learning takes place in two different ways. The first and easier kind of learning is the adjustment of expected attributes to the experiences made by the agents. In our model, so far, the agents start with believing others to be exactly like themselves. If they make, for instance, cooperative experiences exceeding expectations the value of expected cooperativeness is increased. However, the adjustment is not necessarily exact, because the agents do not perceive the other agents' attributes directly, but only their decisions.

The second learning process affects heuristics. If an agent makes negative experiences using one heuristic, in some cases these experiences should lead to an exchange of that heuristic by another possible one. If, in the above case, an agent with a high cooperativeness uses *cooperative heuristic* and gets a return of 0, it might consider this heuristic to be inappropriate the next time a similar decision has to be made, although neither its own cooperativeness nor any expected attribute value has changed. In the example presented here, however, this kind of learning does not take place.

The attribute values themselves do not change over the time scale of the model simulations. Changes in attributes may occur over longer time scales, months or years.

3. TESTING ASSUMPTIONS WITH EXPERIMENTAL DATA

The example in the previous section shows that even with a simple decision task there may be a great number of possible heuristics that could explain human behaviour. In order to test model assumptions, we use data from experimental economics (Ebenhöh and Pahl-Wostl, 2004). Laboratory experiments provide us with a rich data base of individual human behaviour in simple controllable settings. By variations of the experimental settings, we can focus on different aspects of human behaviour. We base our choice of heuristics used in the model on the data and accompanying questionnaires. Furthermore, the process in which the agents choose between the heuristics is also derived from data. In order to be able to do this, we have to analyse individual and not only aggregated data. The representation of individual data is an important advantage of agent-based modelling compared to other modelling techniques (see also Duffy 2001, 309). However, it requires agent behaviour to be more psychologically plausible, than if only aggregated data were to be reproduced (Jager, Janssen, 2002, 99).

But even representation of individual data may not be sufficient to model the actual human behaviour. In the previous example, the game design does not allow us to distinguish trusting behaviour from genuinely cooperative behaviour. That is, we do not know if a first mover invested money, because he or she expected the other player to return a part of the tripled investment, or if the utility increase by tripling the investment was reason enough for him or her to give money away. Likewise we do not know, whether returned amounts are due to positive reciprocity or a sense of fairness (cf. Cox, 2004, p. 264 in a comment on a similar experiment by Berg, Dickhaut and McCabe, 1995). Reciprocity assumes that the player who trusts another player expects to trigger a social norm of fairness that is stronger than the possible desire to defect and maximize individual utility.

Such differences may be important to understand behaviour in real world settings. However, one question arises for all experimental approaches – how far can insights from such experiments be transferred to situations in real world settings?

4. PROSPECTS

The attribute approach presented in this paper is guided by empirical analysis and modelling practices rather than psychological theory. In psychology there are some “trait approaches” (Liebert and Spiegler, 1994) to explain human behaviour on the basis of dispositions, defined as “enduring, stable personality characteristic(s)” (Liebert and Spiegler, 1994, 156). For our modelling purpose, however, empirically tested traits, like the “big five” (neuroticism, extraversion, openness, agreeableness and conscientiousness) are too broad and general compared with the attributes chosen. In order to merge psychological theory with the model assumptions, one would have to conduct thorough empirical investigations on the correlation between the super traits and our attributes and how the super traits translate via attributes into observed behaviour.

To understand key behavioural phenomena as trust and cooperation requires an interdisciplinary approach in the social sciences combing insights from economics, psychology and sociology. Here data from experimental economics have their limitations. Social interactions are largely excluded. While experiments in economics often emphasize the generality of a situation and comprise monetary rewards and repeated trials, psychologists try to capture intrinsic motivations and the mental processes at work in a particular decision situation, what has been termed the *framing* of a decision problem. Hence it will be of major interest to test insights gained from experimental economics in real world settings where social interactions, context and framing play a major role.

The concepts developed in this modelling approach are currently tested in a number of case studies dealing with collective decision making processes and collaborative governance in the management of common pool resource problems. The typical methodological case study design includes participatory integrated assessment and modelling approaches. The decision context refers either to groups of stakeholders such as farmers engaging in collective action and cooperative governance of a common pool resource or to processes of social learning and negotiation in moderated group settings (Pahl-Wostl, 2002b).

Moderated group settings include an actors' platform with representatives from stakeholder groups who engage in processes of social learning and collective decision making over a period of 1-2 years. A typical platform is expected to involve the following sequence of steps (Pahl-Wostl, 2002b):

(1) Build up a shared problem perception in a group of actors, in particular when the problem is largely ill-defined (this does not imply consensus building).

(2) Build trust as base for a critical self-reflection, which implies recognition of individual mental frames and images and how they pertain to decision making.

(3) Recognize mutual dependencies and interactions in the actor network.

(4) Reflect on assumptions about the dynamics and cause-effect relationships in the system to be managed.

(5) Reflect on subjective valuation schemes.

(6) Engage in collective decision and learning processes (this may include the development of new management strategies and the introduction of new formal and informal rules and resource allocation schemes).

This sequence describes an idealized case. In reality quite a few obstacles may impede such cooperative learning and decision making processes. A crucial variable is the willingness to cooperate and trust between stakeholders participating in such a process (Panebianco and Pahl-Wostl, 2004).

In contrast to the experimental game settings the settings in case studies are determined by context and history. The actors know each other and hold expectations about the attributes of other individuals (in contrast to expectations about an average other agent in the game settings). They hold general expectations about expected degrees of cooperativeness and fairness that are determined by their prior experience and their cultural and social environment. At the same time the attributes of an individual actor are not assumed to be invariants but are assumed to be shaped in a long-term learning process determined by experience and the social and cultural environment. Hence one can conclude the social environment and the mutual expectations are partly socially constructed. And we can expect that attributes change over the time period of observation.

Role playing games have been used to generate trust in a stakeholder group and build a shared understanding of the problem (steps 1-3 in the above listed sequence). Such role playing games have been combined with agent based simulation models in an iterative fashion to elicit new insights about decision making strategies, and support the development of new strategies (e.g. Pahl-Wostl and Hare, in press, Barreteau et al 2001). The empirical investigations can be compared with the heuristics derived from experimental games.

Figure 2 shows an example of a bargaining routine that was elicited during a role playing game in a stakeholder platform, a situation where experimental results from dictator and ultimatum games can be quite useful (Ebenhöh and Pahl-Wostl, in review).

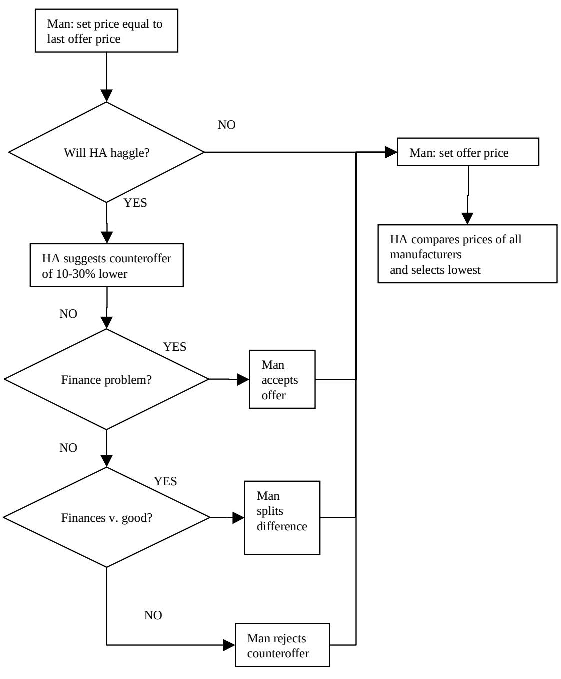

Figure 2 Bargaining algorithm between Housing Association and Manufacturer of Sanitary Technologies (Hare, Heeb and Pahl-Wostl, 2002)

Currently work is ongoing to test the applicability of the modelling approach described in section 3 in a number of case studies. Specific emphasis is given to:

- Applicability of attributes listed in section 3 to characterize actors in real world settings.
- Applicability of heuristics to characterize actor behaviour in real world settings.
- The dynamics of learning processes (individual attributes, expectations, and heuristics) over different time scales in real world settings.
- The importance of context and history in case studies for understanding human behaviour in contrast to the context free, mostly anonymous experimental game settings.

An improved understanding of the importance of trust and cooperation and the use of simple heuristics in learning and decision making processes will support the development of improved participatory approaches and decision support tools in the management of common pool resources.

5. REFERENCES

Barreteau, O., F. Bousquet, and J.-M. Attonaty, Role-playing games for opening the black box of multi-agent systems: method and lessons of its application to Senegal River Valley irrigated systems. *Journal of Artificial Societies and Social Simulation* 4, 2001 <http://jasss.soc.surrey.ac.uk/4/2/5.html>

Berg, J., J. Dickhaut, and K. McCabe, Trust, reciprocity, and social history, *Games and Economic Behaviour*, Vol 10, 122-142, 1995

Cox, J. C. How to identify trust and reciprocity, *Games and Economic Behaviour*, Vol 46, 260-281, 2004

Deadman, P.J., E. Schlager, and R. Gimblett, Simulating Common Pool Resource Management Experiments with Adaptive Agents Employing Alternate Communication Routines. *Journal of Artificial Societies and Social Simulation* 3(2), 2000 <http://jasss.soc.surrey.ac.uk/3/2/2.html>

Duffy, J. Learning to speculate: Experiments with artificial and real agents. *Journal of Economic Dynamics & Control*, 25, 295-319, 2001

Ebenhöh, E. and C. Pahl-Wostl, An adaptive toolbox model: a Pluralistic Modelling Approach for Human Behaviour based on Observation. *Journal of Artificial Societies and Social Simulation*, 7(1), 2004 <http://jasss.soc.surrey.ac.uk/7/1/3.html>

Fehr, E. and S. Gächter, Altruistic punishment in humans, *Nature*, 415, 137-140, 2002.

Fehr, E. and B. Rockenbach, Detrimental Effects of sanctions on human altruism, *Nature*, Vol 422, 137-139, 2003

Gigerenzer, G. and R. Selten, *Bounded Rationality, The Adaptive Toolbox*. Cambridge, Massachusetts: The MIT Press. 2001.

Gintis, H. Game Theory Evolving. Princeton University Press, Princeton, NJ, 2000.

Hardin, R.G. The Tragedy of the Commons. Science 162: 1243-1248, 1968.

Hare,M., J. Heeb and C. Pahl-Wostl, The Symbiotic Relationship between Role Playing Games and Model Development: A case study in participatory model building and social learning for sustainable urban water management. Proceedings of the Meeting of the International Society for Ecological Economics, Sousse, Tunisia, 2002.

Jager W. and M. A. Janssen, Using artificial agents to understand laboratory experiments of common-pool resources with real agents. In: M. A. Janssen (editor), Complexity and Ecosystem Management. The Theory and Practice of Multi-Agent Systems, Bodmin, Cornwall, 2002

Liebert, R. M. and M. D. Spiegler, *Personality: strategies and issues*, Wadsworth, Belmont, California, 1994.

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

Pahl-Wostl, C. Agent Based Simulation in Integrated Assessment and Resources Management. In: Rizzoli,A. and Jakeman,T. (eds). Integrated Assessment and Decision Support. Proceedings of the 1st biennial meeting of the International Environmental Modelling and Software Society, Vol 2, 239-250, 2002a.

Pahl-Wostl, C. Participative and Stakeholder-Based Policy Design, Evaluation and Modeling Processes, *Integrated Assessment*, 3(1), 3-14, 2002b.

Pahl-Wostl, C. and M. Hare, Processes of social learning in integrated resources management. *Journal of Community and Applied Social Psychology*, in press**.**

Panebianco, S. and C. Pahl-Wostl, Obstacles in launching a participatory modelling process. Submitted to iEMSs 2004.