Artificial Morality and Artificial Law

LOTHAR PHILIPPS

Institute of Philosophy of Law and Computers and Law, University of Munich, Germany

Key words: artificial morality, prisoner’s dilemma, game of Chicken, contractual situation, competitive situation, common good, conditional cooperator, reciprocal cooperator, level of reflection, genetic algorithm.

Abstract. The article investigates the interplay of moral rules in computer simulation. The investigation is based on two situations which are well-known to game theory: the prisoner’s dilemma and the game of Chicken. The prisoner’s dilemma can be taken to represent contractual situations, the game of Chicken represents a competitive situation on the one hand and the provision for a common good on the other. Unlike the rules usually used in game theory, each player knows the other’s strategy. In that way, ever higher levels of reflection are reached reciprocally. Such strategies can be interpreted as ‘moral’ rules.

Artificial morality is related to the discipline of ‘Artificial Life’. As in artificial life, the use of genetic algorithms suggests itself. Rules of behaviour split and reunite as chromosome strings do.

1. Introduction

Artificial Intelligence and the law: this combination suggests a ‘liaison’: Artificial Law [Philipps 1989]. The discipline of Artificial Life exists already [Levy 1992]: artificial life forms, simulated by computer, adapt to an artificial environment, struggle for life, diversify, combine their natural assets and evolve. Could not the same be possible for morality and the law? Rules of behaviour, which guide imaginary people, struggle, cooperate, diversify and unify.

The Canadian philosopher Peter Danielson has recently published a book in which he explores these possibilities using computer programs: ‘Artificial Morality – Virtuous Robots for Virtual Games’ [Danielson 1992]. Some of the principles investigated by Danielson already extend into the realm of legal discourse.

2. The Prisoner’s Dilemma and Contractual Situations

One of the most important insights of modern moral philosophy is that many contractual relations share the structure of the prisoner’s dilemma. This insight is linked to the use of the computer as an instrument of philosophy.

The prisoner’s dilemma describes a scene set in the USA: After a robbery, two vagabonds are apprehended near the site of the crime. The sheriff is convinced that he has caught the perpetrators but he cannot prove it. He locks the suspects in separate cells and makes clear to them their situation:

1. Should one of them plead guilty, but not the other, the one who confesses will be released for giving State’s evidence. The other will face a long term of imprisonment.
2. Should both plead guilty, there will be no need for State’s evidence. Both will be sentenced to prison, but only to medium terms, since the confessions will constitute mitigating circumstances.

3. If neither of them should plead guilty, the court will have no choice but to sentence both to only short terms of imprisonment – for vagrancy.

The prisoner’s situation can be translated to a game theory matrix (a high number does not imply a long term in prison, on the contrary: the higher one’s number the better one’s situation):

<table>
<thead>
<tr>
<th></th>
<th>Denial (cooperation)</th>
<th>Confession (defection)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denial (cooperation)</td>
<td>2,2</td>
<td>0,3</td>
</tr>
<tr>
<td>Confession (defection)</td>
<td>3,0</td>
<td>1,1</td>
</tr>
</tbody>
</table>

This matrix shows what is likely to happen: both will confess. For each of them will say to himself: If my partner should confess, it will be better for me to confess as well, or my prison term will be long. If the other does not confess I will profit all the more: I will be released.

In terms of game theory this means that the strategy of confession dominates denial. Or perhaps the cautious maximin principle was applied. It states that the course of action should be chosen which offers in its worst case the comparatively best result. The worst result of confession is a medium prison term; in the case of denial it is a long prison term.

It is remarkable that the rational course of action for each individual prisoner is not prudent for both together. It would be better for both to keep silent; that would mean only short imprisonment for both.

The fundamental philosophical importance of the prisoner’s dilemma has long been realized. According to my knowledge, it was Canadian philosopher David Gauthier who first called attention to the fact that the situation is the same for contracts [Gauthier 1969]. This is definitely true for contracts in the ‘State of Nature’ where state power is not available for enforcement. Each party might reason this way: I would like to perform my part of the deal, but how can I be sure that the other will do the same? After all, he will consider the possibility that I might not perform. Therefore, to minimize the potential damage, neither will honour the agreement. This is true not only for outright breach of contract, but also for insufficient performance, which seems even more realistic.

Such contracts in a state of nature still exist in our present society. For example plea bargaining, the deal for the sentence struck between judge, defence, and state’s attorney. In Germany plea bargaining is not considered permissible but takes place again and again [Schünemann 1990]. It is possible for the judge to ignore the deal, or for a defendant to make a false confession and incriminate an innocent third party.

We all know the legal instruments of enforcement for contractual obligations. But perhaps many people honour their agreements regardless of the state’s threatening shadow. This is the case in long standing business relations, which can be simulated if the pris-