A Game Theory Model for Manipulation Based on Machiavellianism: Moral and Ethical Behavior

Julio B. Clempner¹

¹*Instituto Politecnico Nacional, Lauro Aguirre 120, col. Agricultura, Del. Miguel Hidalgo, 11360, Mexico*
Correspondence should be addressed to julio@clempner.name

Journal of Artificial Societies and Social Simulation 20(2) 12, 2017
Doi: 10.18564/jasss.3301 Url: <http://jasss.soc.surrey.ac.uk/20/2/12.html>

Received: 04-07-2016 Accepted: 15-11-2016 Published: 31-03-2017

Abstract: This paper presents a new game theory approach for modeling manipulation behavior based on Machiavellianism (social conduct and intelligence theory). The Machiavellian game conceptualizes the Machiavellianism considering three concepts: views, tactics and immorality. For modeling the Machiavellian views and tactics we employ a Stackelberg/Nash game theory approach. For representing the concept of immorality, we consider that rational Machiavellian players employ a combination of the deontological and utilitarian moral rules, as well as, moral heuristics. We employ a reinforcement learning approach for the implementation of the immorality concept providing a computational mechanism, in which, its principle of error-driven adjustment of cost/reward predictions contributes to the players' acquisition of moral (immoral) behavior. The reinforcement learning algorithm is based on an actor-critic approach responsible for evaluating the new state of the system and it determines if the cost/rewards are better or worse than expected, supported by the Machiavellian game theory solution. The result of the model is the manipulation equilibrium point. We provide the details needed to implement the extraproximal method in an efficient and numerically stable way. Finally, we present a numerical example that validates the effectiveness of the manipulation model.

Keywords: Machiavellianism, Machiavellian Intelligence, Stackelberg/Nash Game, Machine Ethics, Moral, Markov Chains, Behavioral Games.

Introduction

Brief review

- 1.1 Game theory is a mathematical system for analyzing and predicting how humans behave in strategic situations. It uses three distinct concepts to make precise predictions of how people will, or should, interact strategically: strategic thinking, best-reply, and mutual consistency (equilibrium). Strategic thinking assumes that all players form beliefs based on an analysis of what other players may possibly do. The best-reply chooses the best reply given those beliefs. Mutual consistency adjust best responses and beliefs until they reach an equilibrium. In this sense, there is an understandable interest in the development and implementation of interaction social models which tries to understand, predict, manipulate and control the behavior of people, organizations, government, companies, etc. for instance, (see Schindler 2012; Xianyu 2010; Wijermans et al. 2013). Everyone that goes aboard on this kind of interaction models must consider the behavior observations of the fifteenth-century philosopher and politic, Niccolò Machiavelli (Wilson et al. 1996).
- 1.2 Machiavelli's primary contribution was his painfully honest observations about human nature (Machiavelli 1952). He distinguishes the natural laws that govern how effective leaders exercise power over the human resources and creates a new moral system, deeply rooted in Roman virtue (and vice). He develops his proposal against the conceptions of the Judeo-Christian self-contained moral systems. His ethical system works both as a limit to human possibilities and as the source of human virtue. Machiavelli says that human nature is aggressive and only in some measure able to be manipulated. In this sense, he observes that under competitive conditions the human bean pursue his main goals with increasing levels of ruthlessness (Machiavelli 1952, 1965, 2001).

- 1.3** It is important to note that Machiavelli showed consideration for moral individuals (Leary et al. 1986), and recognized that exist individuals able to sacrifice their own self-interest in order improve the interests of others. However, he did question the regular occurrence of self-sacrifice and ideal altruism behavior in the real world.
- 1.4** A Machiavellian individual (Calhoun 1969) is one who employs aggressive, manipulative, exploiting and devious moves in order to achieve personal and organizational objectives. Theses moves are undertaken according to perceived feasibility with secondary consideration (what is necessary under circumstances) to the feelings, needs and/or rights of others.
- 1.5** Although most of what Machiavelli had to say was intended to provide advice on how successful leaders exercise power over political organizations, his views can (and should) be applied to today's business executives and organizations, given that all organizations are subject to power politics. Machiavelli believed that all leaders deal with the fact that human beings compete for power. Therefore, it makes sense for leaders to see themselves as engaged in two different conflicts: internal and external with "followers" and "competitors" (Wilson et al. 1996).
- 1.6** Machiavellianism is defined as follows:
- Definition 1** *Machiavellianism is a social conduct strategy supposing that the world can be manipulated by applying (Machiavelli's) tactics with the purpose of achieving personal gains according (or not) to a conventional moral.*
- 1.7** For the purposes of this paper, we will consider the terms views, tactics and immorality defined as follows:
- *Views*: The belief that the world can be manipulated - the world consists of manipulators and manipulated
 - *Tactics*: The use of a manipulation strategies needed to achieve specific power situations (goals). Strategies are analyzed in Machiavelli's The Prince (Machiavelli 1965), The Discourses (Machiavelli 1965), The Art of War (Machiavelli 2001) and the psychological behavior patterns (described above).
 - *Immorality*: The disposition to not become attached to a conventional moral.
- 1.8** The *Machiavellian intelligence* is the capacity of an individual to be in a successful engagement with social groups (Byrne & Whiten 1988).

Machiavellianism and moral behavior

- 1.9** Machiavellianism has been associated with different variables, given a wide range of interpretations related to psychological components (Byrne & Whiten 1988). Smith (1979) argued that the descriptors of the psychopath and those of the Machiavellian must have common domains, because of their similarity (manipulative style, poor affect, low concern about conventional moral, low ideological compromise, and others). In agreement with Cleckley (1976), other Machiavellian tendencies coincide with some components of anomie (cynicism, low interpersonal credibility, external locus of control).
- 1.10** Researchers have investigated the relationship between locus of control and Machiavellianism. Solar & Bruehl (1971) were the first in establishing a relationship between Machiavellianism and locus of control, considering both as aspects of interpersonal power. Their study reported a significant relationship between Machiavellianism and external locus of control for males, but not for females. Prociuk & Breen (1976) supported this result. Mudrack (1989) conducted a meta-analytic review of 20 studies determining the relationship between Machiavellianism and external locus of control. Gable et al. (1992) sustain this result. They related locus of control, Machiavellianism and managerial achievement; their results did not show significant correlations between locus of control and achievement, but found a positive correlation between Machiavellianism and external control.
- 1.11** With respect to influence tactics, Falbo (1977) showed that persons with high Machiavellianism are associated with the use of rational indirect tactics (i.e., lies), while those with low Machiavellianism are associated with the rational use of direct tactics (i.e. rewards). Grams & Rogers (1990) research confirms this result and also shows that persons with high Machiavellianism are more flexible when it comes to breaking some ethical rules. Vecchio & Sussmann (1991) suggested that Machiavellianism and the tactics selection are related to sex and organizational hierarchy; the use of influence tactics is common in males and females with high-level positions.
- 1.12** In accordance with different studies of social psychology, manipulation is placed among the forms of social influence as part of the social conduct behavior. Raven (1993) argued that power can be psychologically studied

as a product of behavior, including personal attributes, with the possibility to affect others through interaction, and the environment structure. Dawkins (1976) proposed that, in terms of selfishness, altruism, cooperation, manipulation, lie and truth, genetically there exists a selfishness and manipulation gene. Dawkins & Krebs (1978) classified manipulation as a natural-selection state benefiting individuals able to manipulate others' behavior. Vleeming (1979) denoted a personality dimension in which people can be classified in terms of being more or less manipulated in different interpersonal situations. Wilson et al. (1996) defined Machiavellianism as a social strategy behavior involving the manipulation of others to obtain personal benefits, frequently against others' interests. They clarify that anybody is able to manipulate others to different degrees, and they also explain that selfishness and manipulation are behaviors widely studied in evolutionary biology. Hellriegel et al. (1997) defined Machiavellianism as a personal style of behavior in front of others, characterized by: the use of astuteness, tricks and opportunism in interpersonal relationships; cynicism towards other persons' nature; lack of concern with respect to conventional morals. Christie & Geis (1970) proposed three factors to evaluate high or low Machiavellianism: tactics, morality and views. Tactics are concerned with planned actions (or recommendations) to confront specific situations with the purpose of obtaining planned benefits at the expense of others. Morality is related to behavior that can be associated with some degree of "badness" with respect social conventions. Views involve the idea that the world consists of manipulators and manipulated. In this sense we introduce the following definition.

- 1.13 Immorality is a un-arrangement of customs. One of the best-known concepts is the immorality described by Nietzsche. Therefore, we consider the factor of morality proposed by Christie & Geis (1970) as inappropriate, because in the evaluation of the factor, immorality is considered the opposite to a "conventional moral."
- 1.14 There have been proposed several methods from different angles and application domains for solving the problem of moral and ethical decision making (Bales 1971; Cervantes et al. 2011; Dehghani et al. 2008; Indurkha & Misztal-Radecka 2016; Wallach et al. 2010; Hartog & Belschak 2012). However, there is still a lack of a fundamental and mathematical decision model and a rigorous cognitive process for decision-making. In this paper, we will propose a game theory solution combined with a reinforcement learning approach to acquire manipulation behavior.

Main results

- 1.15 This paper presents a new game theory approach for modeling manipulation behavior based on the Machiavellian social conduct theory. The Machiavellian game conforms a system that allows to analyze and predict how Machiavellian players behave in strategic situations combining all the following three fundamental features of game theory: a) formation of beliefs based on analysis of what others players might do (strategic thinking); b) choosing a best-reply given those beliefs (optimization); and c) adjustment of best-reply and beliefs until they are mutually consistent (equilibrium). The assumption of mutual consistency is justified by introducing a learning process. As a result, the Machiavellian equilibrium is the consequence of a strategic thinking, optimization, and equilibration (learning process).
- 1.16 A Machiavellian player conceptualizes the manipulation social conduct considering three concepts: views, tactics and immorality. For modeling the Machiavellian Views and Tactics we employ a Stackelberg/Nash game theory approach. For representing Machiavellian immorality we introduce utilitarian and deontological moral theories and we review psychological findings regarding moral decision making (Tobler et al. 2008). An advantage of the Machiavellian social conduct theory is that whereas other moral theories provide standards for how we should act, they do not describe how moral judgments and decisions are achieved in practice. Next, we establish a relationship between moral behavior and game theory (economic theories) and we suggest a reinforcement learning approach which provides evidences of how people acquire Machiavellian immoral behavior, considering its principle of error-driven adjustment of cost/reward predictions.
- 1.17 In summary, this paper makes the following contributions:
 - We represent the Machiavellian Views naturally as a Stackelberg game where the hierarchical organization consists of manipulators (leaders) and manipulated (followers) players. By definition of the Stackelberg game the leaders have commitment power.
 - The manipulators and manipulated players are themselves in a (non-cooperative) Nash game.
 - The Machiavellian Tactics correspond to the solution of Stackelberg/Nash game.
 - We employ the extraproximal method for solving the Machiavellian game (Antipin 2005; Trejo et al. 2015, 2016) (see the appendix).

- We suggest a reinforcement learning approach for modeling the Machiavellian immorality based-on an Actor-Critic architecture
- For representing the Machiavellian immorality, we consider that rational players employ deontological and utilitarian moral (depending the case)
- We suggest the immorality reinforcement learning rules needed for modeling immorality.
- We restrict our approach to a finite, ergodic and controllable Markov chains.
- The result of the model is the manipulation equilibrium point.
- Finally, we present a numerical example that validates the effectiveness of the proposed Machiavellian social conduct and intelligence approach.

Organization of the paper

- 1.18** The rest of the paper is organized as follows. In the next Section, we present the mathematical background needed for the understanding of the rest of the paper. Section 3 describes the general Machiavellianism social conduct architecture. We suggest a Machiavellian Stackelberg game theory model for representing Machiavellian Views and Tactics, in Section 4. Section 5 describes a reinforcement learning approach for modeling the Machiavellian immorality. In Section 6, we present a simulated experiments for Machiavellian social conduct theory. Finally, in Section 7 we present some conclusions and future work. To make the paper more accessible, the long details needed to implement the extraproximal method are placed in the appendix.

Preliminaries

- 2.1** Let S be a finite set, called the *state space*, consisting of finite set of states $\{s_{(1)}, \dots, s_{(N)}\}$, $N \in \mathbb{N}$. A *Stationary Markov chain* (Clempner & Poznyak 2014) is a sequence of S -valued random variables $s(n)$, $n \in \mathbb{N}$. The Markov chain can be represented by a complete graph whose nodes are the states, where each edge $(s_{(i)}, s_{(j)}) \in S^2$ is labeled by the transition probability. The matrix $\Pi = (\pi_{(ij)})_{(s_{(i)}, s_{(j)}) \in S} \in [0, 1]^{N \times N}$ determines the evolution of the chain: for each $k \in \mathbb{N}$, the power Π^k has in each entry $(s_{(i)}, s_{(j)})$ the probability of going from state $s_{(i)}$ to state $s_{(j)}$ in exactly k steps.

Definition 2 A *controllable Markov chain* (Poznyak et al. 2000) is a 4-tuple

$$MC = \{S, A, \mathbb{K}, \Pi\} \quad (1)$$

where:

- S is a finite set of states, $S \subset \mathbb{N}$, endowed with discrete a topology;
- A is the set of actions, which is a metric space. For each $s \in S$, $A(s) \subset A$ is the non-empty set of admissible actions at state $s \in S$. Without loss of generality we may take $A = \cup_{s \in S} A(s)$;
- $\mathbb{K} = \{(s, a) | s \in S, a \in A(s)\}$ is the set of admissible state-action pairs, which is a measurable subset of $S \times A$;
- $\Pi = [\pi_{(ij|k)}]$ is a stationary controlled transition matrix, where

$$\pi_{(ij|k)} \equiv P(s(n+1) = s_{(j)} | s(n) = s_{(i)}, a(n) = a_{(k)})$$

represents the probability associated with the transition from state $s_{(i)}$ to state $s_{(j)}$ under an action $a_{(k)} \in A(s_{(i)})$, $k = 1, \dots, M$;

Definition 3 A *Markov Decision Process* is a pair

$$MDP = \{MC, J\} \quad (2)$$

where:

- MC is a controllable Markov chain (1)
- $J : S \times \mathbb{K} \rightarrow \mathbb{R}$ is a cost function, associating to each state a real value.

2.2 The strategy (policy)

$$d_{(k|i)}(n) \equiv P(a(n) = a_{(k)} | s(n) = s_{(i)})$$

represents the probability measure associated with the occurrence of an action $a(n)$ from state $s(n) = s_{(i)}$.

2.3 The elements of the transition matrix for the controllable Markov chain can be expressed as

$$P(s(n+1) = s_{(j)} | s(n) = s_{(i)}) = \sum_{k=1}^M P(s(n+1) = s_{(j)} | s(n) = s_{(i)}, a(n) = a_{(k)}) d_{(k|i)}(n)$$

2.4 Let us denote the collection $\{d_{(k|i)}(n)\}$ by D_n as follows

$$D_n = \{d_{(k|i)}(n)\}_{k=\overline{1,M}, i=\overline{1,N}}$$

2.5 A policy $\{d_n^{loc}\}_{n \geq 0}$ is said to be *local optimal* if for each $n \geq 0$ it maximizes the conditional mathematical expectation of the utility function $J(s_{n+1})$ under the condition that the history of the process

$$F_n := \{D_0, P\{s_0 = s_{(j)}\}_{j=\overline{1,N}}; \dots; D_{n-1}, P\{s_n = s_{(j)}\}_{j=\overline{1,N}}\}$$

is fixed and can not be changed hereafter, i.e., it realizes the “one-step ahead” conditional optimization rule

$$d_n^{loc} := \arg \min_{d_n \in D_n} E\{J(s_{n+1}) | F_n\} \quad (3)$$

where $J(s_{n+1})$ is the utility function at the state s_{n+1} .

2.6 The dynamics of the Stackelberg game for Markov chains is described as follows. The game consists of $\iota = \overline{1, M+N}$ players and begins at the initial state $s^\iota(0)$ which (as well as the states further realized by the process) is assumed to be completely measurable. Each player ι is allowed to randomize, with distribution $d_{(k|\iota)}^\iota(n)$, over the pure action choices $a_{(k)}^\iota \in A^\iota(s_{(i)}^\iota)$, $i = \overline{1, N_\iota}$ and $k = \overline{1, M_\iota}$. The leaders correspond to $l = \overline{1, N}$ and followers to $m = \overline{1, M}$. At each fixed strategy of the leaders $d_{(k_l|i_l)}^l(n)$ the followers make the strategy selection $d_{(k_m|i_m)}^m(n)$ trying to realize a Nash-equilibrium. Below we will consider only stationary strategies $d_{(k|\iota)}^\iota(n) = d_{(k|\iota)}^\iota$. In the ergodic case when all Markov chains are ergodic for any stationary strategy $d_{(k|\iota)}^\iota$ the distributions $P^\iota(s^\iota(n+1) = s_{(j_\iota)})$ exponentially quickly converge to their limits $P^\iota(s = s_{(i_\iota)})$ satisfying

2.7

$$P^\iota(s_{(j_\iota)}) = \sum_{i_\iota=1}^{N_\iota} \left(\sum_{k_\iota=1}^{M_\iota} \pi_{(i_\iota, j_\iota | k_\iota)}^\iota d_{(k_\iota | i_\iota)}^\iota \right) P^\iota(s_{(i_\iota)}) \quad (4)$$

The cost function of each player, depending on the states and actions of all the other players, is given by the values $W_{(i_1, k_1; \dots; i_{M+N}, k_{M+N})}^\iota$, so that the “average cost function” $\mathbf{J}^\iota$ in the stationary regime can be expressed as

$$\mathbf{J}^\iota(c^1, \dots, c^{M+N}) := \sum_{i_1, k_1} \dots \sum_{i_{M+N}, k_{M+N}} W_{(i_1, k_1; \dots; i_{M+N}, k_{M+N})}^\iota \prod_{\iota=1}^{M+N} c_{(i_\iota, k_\iota)}^\iota \quad (5)$$

where $c^\iota := [c_{(i_\iota, k_\iota)}^\iota]_{i_\iota=\overline{1, N_\iota}; k_\iota=\overline{1, M_\iota}}$ is a matrix with elements

$$c_{(i_\iota, k_\iota)}^\iota = d_{(k_\iota | i_\iota)}^\iota P^\iota(s_{(i_\iota)}^\iota) \quad (6)$$

satisfying

$$c^{(\iota)} \in C_{adm}^{(\iota)} = \begin{cases} c^{(\iota)} : \sum_{i_\iota, k_\iota} c_{(i_\iota, k_\iota)}^\iota = 1, c_{(i_\iota, k_\iota)}^\iota \geq 0, \\ \sum_{k_\iota} c_{(j_\iota, k_\iota)}^\iota = \sum_{i_\iota, k_\iota} \pi_{(i_\iota, j_\iota | k_\iota)}^\iota c_{(i_\iota, k_\iota)}^\iota \end{cases} \quad (7)$$

where

$$W_{(i_1, k_1, \dots, i_{\mathcal{M}+\mathcal{N}}, k_{\mathcal{M}+\mathcal{N}})}^\ell = \sum_{j_1} \dots \sum_{j_{\mathcal{M}+\mathcal{N}}} J_{(i_1, j_1, k_1, \dots, i_{\mathcal{M}+\mathcal{N}}, j_{\mathcal{M}+\mathcal{N}}, k_{\mathcal{M}+\mathcal{N}})}^\ell \prod_{l=1}^{\mathcal{M}+\mathcal{N}} \pi_{(i_l, j_l | k_l)}^\ell$$

2.8 Notice that by Equation (6) it follows that

$$P^\ell(s_{(i_\ell)}) = \sum_{k_\ell} c_{(i_\ell, k_\ell)}^\ell \quad d_{(k_\ell | i_\ell)}^\ell = \frac{c_{(i_\ell, k_\ell)}^\ell}{\sum_{k_\ell} c_{(i_\ell, k_\ell)}^\ell} \quad (8)$$

In the ergodic case $\sum_{k_\ell} c_{(i_\ell, k_\ell)}^\ell > 0$ for all $\ell = \overline{1, \mathcal{M} + \mathcal{N}}$.

2.9 Let us introduce the variables

$$u^l := \text{col} \left(c^{(l)} \right), \quad U^l := C_{adm}^{(l)} \quad (l = \overline{1, \mathcal{N}}), \quad U := \bigotimes_{l=1}^{\mathcal{N}} U^l$$

for the leaders l where col is the column operator, and let us introduce the variables

$$v^m := \text{col} \, c^{(m)}, \quad V^m := C_{adm}^{(m)} \quad (m = \overline{1, \mathcal{M}}), \quad V := \bigotimes_{m=1}^{\mathcal{M}} V^m \quad (9)$$

for the m followers.

2.10 We consider a Stackelberg game where the leaders and the followers do not cooperate. Then, the definitions of the equilibrium point is as follows:

2.11 A Stackelberg/Nash equilibrium for the leaders is a strategy $u^{**} = (u^{1**}, \dots, u^{\mathcal{N}**})$ given the fixed strategy $v = (v^1, \dots, v^{\mathcal{M}})$ of the followers such that there does not exist any $u^l \in U, u^l \neq u^{l**}$

$$\mathbf{J}(u^{1**}, \dots, u^l, \dots, u^{\mathcal{N}**} | v) \leq \mathbf{J}(u^{1**}, \dots, u^{\mathcal{N}**} | v)$$

for $u^l \in U$ and $v \in V$.

2.12 A Stackelberg/Nash equilibrium for the followers is a strategy $v^* = (v^{1*}, \dots, v^{\mathcal{M}*})$ given the strategy $u = (u^1, \dots, u^{\mathcal{N}})$ for the followers such that

$$\mathbf{J}(v^{1*}, \dots, v^{\mathcal{M}*} | u) \geq \mathbf{J}(v^{1*}, \dots, v^m, \dots, v^{\mathcal{M}*} | u)$$

for any $v^m \in V$ and $u \in U$.

Machiavellianism architecture

3.1 This paper presents a new game theory model to represent the Machiavellianism (social conduct and intelligence theory). The Machiavellian social conduct is engaged with manipulating others for personal gain, even against the other's self-interest and the Machiavellian intelligence is the capacity of an individual to be in a successful engagement with social groups (Byrne & Whiten 1988). A Machiavellian agent is one who conceptualize the Machiavellianism considering three concepts: views, tactics and immorality. Views involve the idea that the world consists of manipulators and manipulated (restricting the model to two types of agents). The tactics are concerned the use of manipulation (Machiavelli's) strategies. The immorality is related to the natural behavior to not become restricted to a conventional moral (the Machiavellian agent's behavior is rationally bounded by the immorality). In order to (manipulate) acquire power, survive or sustain a particular position, Machiavellian agents make use of Machiavellian intelligence applying different selfish manipulation strategies, which include looking for control the changes taking place in the environment.

3.2 Figure 1 shows the schematic structure of the Machiavellian actor-critic algorithm involving a reinforcement learning process. The learning agent has been split into two separate entities: the actor (game theory) and the critic (value function). The actor coincides with a Stackelberg/Nash game (Stackelberg 2011) that implements the concepts of Views and Tactics of the Machiavellianism. The critic conceptualize the Machiavellian immorality represented by a value function process. The actor is responsible for computing a control strategy $d_{(i_\ell | k_\ell)}^{u^*}$,

Figure 1: Actor-Critic and the Machiavellian architecture.

for each player i , given the current state $\hat{i}_i$. The critic is responsible for evaluating the quality of the current strategy by adapting the value function estimate. After a number of strategy evaluation steps by the critic, the actor is updated by using information from the critic. The architecture is described as follows.

- 3.3 The structure of the game corresponds to a Stackelberg approach in response to the Machiavellian Views, which involves the idea that the world consists of manipulators and manipulated players. The dynamics of the Machiavellian game is as follows: the manipulators players consider the best-reply of the manipulated players, and then select the strategy that optimizes their utility, anticipating the response of the manipulated players. Subsequently, the manipulated players observe the strategy played by the manipulators players and select the best-reply strategy. The manipulators and manipulated players are themselves in a (non-cooperative) Nash game. Formally, the Stackelberg model is solved to find the subgame perfect Nash equilibrium, i.e. the strategy that serves best each player, given the strategies of the other player and that entails every player playing in a Nash equilibrium in every subgame. The Machiavellian Tactics are concerned with the use of manipulation strategies. In this sense, the equilibrium point of the game represents the strategies needed to achieve specific power situations. In this model, the manipulators have commitment power presenting a significant advantage over the manipulated players.
- 3.4 For representing the concept of immortality we employ a reinforcement learning approach providing a computational mechanism, in which, its principle of error-driven adjustment of cost/reward predictions contributes to the players' acquisition of moral/immoral behavior. The reinforcement learning algorithm is based on an actor-critic approach responsible for evaluating the new state of the system and determine if the cost/rewards are better or worse than expected, supported by the Machiavellian game theory solution. We view actor-critic algorithm as stochastic game algorithm on the parameter space of the actor. The game is solved employing the extraproximal method (see the appendix). The functional of the game is viewed as a regularized Lagrange function whose solution is given by a stochastic gradient algorithm.
- 3.5 The responsibility of the actor is computing the strategy solution $d_{(\hat{i}_i | \hat{k}_i)}^{l*}$ of the Machiavellian game. In the actor, the action selection follows the strategy solution $d_{(\hat{i}_i | \hat{k}_i)}^{l*}$ of the Machiavellian game: a) given a fixed $\hat{i}_i$, each player chooses randomly an action $a^i(t) = a_{(\hat{k}_i)}^i$ (for the estimated value $\hat{k}_i$) from the vector $d_{(\hat{i}_i | \hat{k}_i)}^{l*}$, b) then, players employ the transition matrix $\Pi^i = [\pi_{\hat{i}_i \hat{j}_i | \hat{k}_i}^i]$ to choose randomly the consecutive state $s^i(t+1) = s_{(\hat{j}_i)}^i$ (for the estimated value $\hat{j}_i$) from the vector $\pi_{(\hat{i}_i \hat{j}_i | \hat{k}_i)}^i$ (for a fixed $\hat{i}_i$ and $\hat{k}_i$), c) as soon as $s^i(t)$, $a^i(t)$ and $s^i(t+1)$

are selected they are sent to the Critic.

- 3.6** The role of the critic is to evaluate the current strategy prescribed by the actor and compute an approximation of the projection of $\hat{\pi}_{(i,j|\hat{k})}^{\ell}(t)$ and $\hat{J}_{(i,j|\hat{k})}^{\ell}(t)$. The actor uses this approximation to update its strategies in the direction of performance improvement of the Machiavellian game. In the Critic, we consider that rational Machiavellian players employ a combination of the deontological and utilitarian moral rules, as well as, moral heuristics, for representing the concept of immorality decision-making in the Machiavellianism social conduct theory. Most of the time a Machiavellian player considers a deontological moral. But the disposition to not become attached to a conventional moral is represented by the utilitarian moral which, in our case, promotes immoral values. The utilitarian moral establishes states where the moral quality of actions is determined by their consequences: the goal is to maximize the utility of all individuals in the society. On the other hand, deontological moral states that certain things are morally valuable in themselves. The moral quality of actions arises from the fact that the action is done to protect such moral values: consequences and outcomes of actions are secondary. Our approach considers that a Machiavellian player have the disposition to not become attached to a conventional moral as an utilitarian moral (immorality). This combination between utilitarian and deontological morals suggests a specific method for moral decision making.
- 3.7** The estimated values are updated employing the moral learning rules for computing $\hat{\pi}_{(i,j|\hat{k})}^{\ell}(t)$ and $\hat{J}_{(i,j|\hat{k})}^{\ell}(t)$. When the player's performance is compared to that of a player which acts optimally from the beginning, the difference in performance gives rise to the notion of regret. Then, the critic produces a reinforcement feedback for the actor by observing the consequences of the selected action. The critic takes a decision considering a TD-error e , in our case we consider the mean square error $\left(e \left[\hat{\Pi}^{\ell}(t-1) - \hat{\Pi}^{\ell}(t) \right] > 0\right)$, which determines if the cost/rewards are better or worse than expected with the preceding action. The TD-error e corresponds to the mean squared error of an estimator which in this case measures the difference between the estimator and what is estimated (the difference occurs because of randomness). The TD-error e in the reinforcement learning process is employed to evaluate the preceding action: if the error is positive the tendency to select this action should be strengthened or else, lessened. The value-minimizing/maximizing action at each state are taken whether the actor-critic learning rule $\hat{\pi}_{(i,j|\hat{k})}^{\ell}$ ensures convergence.
- 3.8** We note some minor differences with the common usage of reinforcement learning because the control strategy need to change as the actor computes the game. This need not present any problems, as long as the actor parameters are updated on a slower time scale. The Machiavellian architecture point towards that the cost/reward model may have benefits not only for individual survival by computing the Stackelberg/Nash game, but also by contributing to the computation of the moral value of actions employing a reinforcement learning approach.

Machiavellian Views and Tactics

The Machiavellian game

- 4.1** Let us consider a Stackelberg game (Trejo et al. 2015, 2016) with $\mathcal{N}$ manipulators whose strategies are denoted by $u^l \in U^l$ ($l = \overline{1, \mathcal{N}}$) where U is a convex and compact set. Denote by $u = (u^1, \dots, u^{\mathcal{N}})^{\top} \in U$ the joint strategy of the players and $u^{\hat{l}}$ is a strategy of the rest of the players adjoint to u^l , namely,

$$u^{\hat{l}} := (u^1, \dots, u^{l-1}, u^{l+1}, \dots, u^{\mathcal{N}})^{\top} \in U^{\hat{m}} := \bigotimes_{h=1, h \neq l}^{\mathcal{N}} U^h$$

such that $u = (u^l, u^{\hat{l}})$ ($l = \overline{1, \mathcal{N}}$). As well, let us consider $\mathcal{M}$ attackers with strategies $v^m \in V^m$ ($m = \overline{1, \mathcal{M}}$) and V is also a convex and compact set. Denote by $v = (v^1, \dots, v^{\mathcal{M}})^{\top} \in V := \bigotimes_{m=1}^{\mathcal{M}} V^m$ the joint strategy of the manipulated and $v^{\hat{m}}$ is a strategy of the rest of the players adjoint to v^m , namely,

$$v^{\hat{m}} := (v^1, \dots, v^{m-1}, v^{m+1}, \dots, v^{\mathcal{M}})^{\top} \in V^{\hat{m}} := \bigotimes_{q=1, q \neq m}^{\mathcal{M}} V^q$$

such that $v = (v^m, v^{\hat{m}})$ ($m = \overline{1, \mathcal{M}}$).

Remark 4 Considering the Machiavellian Views (see Subsection – Brief Review –) we represent manipulators and manipulated players in a Stackelberg game (Trejo et al. 2016): this is a multiplayer model involving a Nash game

for the manipulators and a Nash game for the manipulated players restricted by a Stackelberg game defined as follows.

Definition 5 A game with $\mathcal{N}$ manipulators and $\mathcal{M}$ manipulated players is said to be a **Machiavellian Stackelberg/Nash** game if

$$G(u, \hat{u}(u)|v) := \sum_{l=1}^{\mathcal{N}} [\varphi_l(\bar{u}^l, u^{\hat{l}}|v) - \varphi_l(u^l, u^{\hat{l}}|v)]$$

satisfies the properties:

- for the manipulators players

$$\max_{\hat{u}(u) \in \hat{U}} g(u, \hat{u}(u)|v) = \sum_{l=1}^{\mathcal{N}} \varphi_l(\bar{u}^l, u^{\hat{l}}|v) - \varphi_l(u^l, u^{\hat{l}}|v) \leq 0$$

where $u^{\hat{l}}$ is a strategy of the rest of the manipulators players adjoint to u^l , namely,

$$u^{\hat{l}} := (u^1, \dots, u^{l-1}, u^{l+1}, \dots, u^{\mathcal{N}}) \in U^{\hat{l}} := \bigotimes_{h=1, h \neq l}^{\mathcal{N}} U^h$$

and

$$\bar{u}^l := \arg \min_{u^l \in U^l} \varphi_l(u^l, u^{\hat{l}}|v)$$

- for the manipulated players

$$F(v, \hat{v}(v)|u) := \sum_{m=1}^{\mathcal{M}} [\psi_m(\bar{v}^m, v^{\hat{m}}|u) - \psi_m(v^m, v^{\hat{m}}|u)]$$

and

$$\max_{\hat{v}(v) \in \hat{V}} f(v, \hat{v}(v)|u) = \sum_{m=1}^{\mathcal{M}} \psi_m(\bar{v}^m, v^{\hat{m}}|u) - \psi_m(v^m, v^{\hat{m}}|u) \leq 0$$

where $v^{\hat{m}}$ is a strategy of the rest of the manipulated players adjoint to v^m , namely,

$$v^{\hat{m}} := (v^1, \dots, v^{m-1}, v^{m+1}, \dots, v^{\mathcal{M}}) \in V^{\hat{m}} := \bigotimes_{q=1, q \neq m}^{\mathcal{M}} V^q$$

and

$$\bar{v}^m := \arg \min_{v^m \in V^m} \psi_m(v^m, v^{\hat{m}}|u)$$

Definition 6 Let $G(u, \hat{u}(u)|v)$ be the cost functions of the manipulators players ($l = \overline{1, \mathcal{N}}$) and $F(v, \hat{v}(v)|u)$ the cost functions of the manipulated players ($m = \overline{1, \mathcal{M}}$). A strategy $u^* \in U$ of the manipulators players together with the collection $v^* \in V$ of the manipulated players is said to be a **Machiavellian Stackelberg/Nash equilibrium** if

$$(u^*, v^*) \in \arg \min_{u \in U, \hat{u}(u) \in \hat{U}} \max_{v \in V, \hat{v}(v) \in \hat{V}} \{$$

$$G(u, \hat{u}(u)|v) | g(u, \hat{u}(u)|v) \leq 0, f(v, \hat{v}(v)|u) \leq 0\}$$

(10)

Remark 7 If $G(u, \hat{u}(u)|v)$ and $F(v, \hat{v}(v)|u)$ are strictly convex then

$$(u^*, v^*) = \arg \min_{u \in U, \hat{u}(u) \in \hat{U}} \max_{v \in V, \hat{v}(v) \in \hat{V}} \{$$

$$G(u, \hat{u}(u)|v) | g(u, \hat{u}(u)|v) \leq 0, f(v, \hat{v}(v)|u) \leq 0\}$$

Remark 8 The strategy (u^*, v^*) is a solution of the Machiavellian game that corresponds to the Machiavellian Tactics needed to achieve specific power situations (see Subsection – Brief Review –).

4.2 Let $R = \sum_{l=1}^N N_l M_l$, then U_{adm} (U admissible) is defined as follows

$$U_{adm} = S^R \times S_1 \times \dots \times S_N$$

$$S_l = \left\{ \begin{array}{l} c^{(l)} : \sum_{i_l, k_l} c_{(i_l, k_l)}^l = 1, c_{(i_l, k_l)}^l \geq 0, \\ h_j^l(c) = \sum_{i_l, k_l} \pi_{(i_l, j_l | k_l)}^l c_{(i_l, k_l)}^l - \sum_{k_l} c_{(j_l, k_l)}^l = 0 \end{array} \right.$$

The regularized Lagrange principle

4.3 The Regularized Lagrange Function (RLF)

$$\begin{aligned} \mathcal{L}_{\theta, \delta}(u, \hat{u}(u), v, \hat{v}(v), \lambda, \omega, \xi) := & \theta G(u, \hat{u}(u)|v) + \\ & \lambda h_j^l(c) + \omega g(u, \hat{u}(u)|v) + \xi f(v, \hat{v}(v)|u) \\ & + \frac{\delta}{2} (\|u\|^2 + \|\hat{u}(u)\|^2 - \|v\|^2 - \|\hat{v}(v)\|^2 - \|\lambda\|^2 - \|\omega\|^2 - \|\xi\|^2) \end{aligned} \quad (11)$$

where the parameters θ, δ are positive and the Lagrange vector-multipliers $\omega \in \mathbb{R}^{M_\omega}$ and $\xi \in \mathbb{R}^{M_\xi}$ are non-negative and the components of $\lambda \in \mathbb{R}^{M_\lambda}$ may have any sign. Obviously, the optimization problem

$$\mathcal{L}_{\theta, \delta}(u, \hat{u}(u), v, \hat{v}(v), \lambda, \omega, \xi) \rightarrow \min_{u \in U, \hat{u}(u) \in \hat{U}} \max_{v \in V, \hat{v}(v) \in \hat{V}, \lambda, \omega \geq 0, \xi \geq 0} \quad (12)$$

has a unique saddle-point on u, v and λ since the optimized RLF (11) is *strongly convex* (Poznyak 2008) if the parameters θ and $\delta > 0$ provide the condition

$$\frac{\partial^2}{\partial u \partial u^\top} \mathcal{L}_{\theta, \delta}(u, \hat{u}(u), v, \hat{v}(v), \lambda, \omega, \xi) > 0, \forall u \in U_{adm} \subset \mathbb{R}^N$$

and is strongly concave on the Lagrange multipliers λ, ω, ξ for any $\delta > 0$. As well as,

$$\frac{\partial^2}{\partial v \partial v^\top} \mathcal{L}_{\theta, \delta}(u, \hat{u}(u), v, \hat{v}(v), \lambda, \omega, \xi) < 0, \forall v \in V_{adm} \subset \mathbb{R}^M$$

4.4 Based-on these properties the RLF given in Equation (11) has a unique saddle point

$$(u^*(\theta, \delta), \hat{u}^*(u(\theta, \delta)), v^*(\theta, \delta), \hat{v}^*(v(\theta, \delta)), \lambda^*(\theta, \delta), \omega^*(\theta, \delta), \xi^*(\theta, \delta))$$

(see The Kuhn-Tucker Theorem 21.13 in (Poznyak (2008))) for which the following inequalities hold: for any λ and ω, ξ with nonnegative components, $u \in \mathbb{R}^N$ and $v \in \mathbb{R}^M$

$$\begin{aligned} L_{\theta, \delta}(u^*(\theta, \delta), \hat{u}^*(u(\theta, \delta)), v(\theta, \delta), \hat{v}(v(\theta, \delta)), \lambda(\theta, \delta), \omega(\theta, \delta), \xi^*(\theta, \delta)) & \leq \\ L_{\theta, \delta}(u^*(\theta, \delta), \hat{u}^*(u(\theta, \delta)), v^*(\theta, \delta), \hat{v}^*(v(\theta, \delta)), \lambda^*(\theta, \delta), \omega^*(\theta, \delta), \xi^*(\theta, \delta)) & \leq \\ L_{\theta, \delta}(u(\theta, \delta), \hat{u}(u(\theta, \delta)), v^*(\theta, \delta), \hat{v}^*(v(\theta, \delta)), \lambda^*(\theta, \delta), \omega^*(\theta, \delta), \xi^*(\theta, \delta)) & \end{aligned} \quad (13)$$

As for the non-regularized function $\mathcal{L}_{1,0}(u, \hat{u}(u), v, \hat{v}(v), \lambda, \omega, \xi)$, it may have several (not obligatory unique) saddle points $(u^*, \hat{u}^*(u), v^*, \hat{v}^*(v), \lambda^*, \omega^*, \xi^*) \in U^* \otimes V^* \otimes \Lambda^*$.

Proposition 9 If the parameter θ and the regularizing parameter δ tend to zero by a particular manner, then we may expect that $u^*(\theta, \delta), \hat{u}^*(u(\theta, \delta)), v^*(\theta, \delta), \hat{v}^*(v(\theta, \delta))$ and $\lambda^*(\theta, \delta), \omega^*(\theta, \delta), \xi^*(\theta, \delta)$ which are the solutions of the min-max problem (12) tend to the set $U^* \otimes V^* \otimes \Lambda^*$ of all saddle point of the original game theory problem (11), that is,

$$\begin{aligned} \rho\{u^*(\theta, \delta), \hat{u}^*(u(\theta, \delta)), v^*(\theta, \delta), \hat{v}^*(v(\theta, \delta)), \lambda^*(\theta, \delta), \\ \omega^*(\theta, \delta), \xi^*(\theta, \delta); U^* \otimes V^* \otimes \Lambda^*\} \xrightarrow[\theta, \delta \downarrow 0]{} 0 \end{aligned} \quad (14)$$

where $\rho\{a; U^* \otimes V^* \otimes \Lambda^*\}$ is the Hausdorff distance defined as

$$\rho\{a; U^* \otimes V^* \otimes \Lambda^*\} = \min_{z^* \in U^* \otimes V^* \otimes \Lambda^*} \|a - z^*\|^2$$

Machiavellian Immorality

- 5.1** The designing of the Machiavellian immorality module involves the learning rule for estimating $\pi_{(i,j|k)}^\iota$ and the learning rule for estimating $J_{(i,j|k)}^\iota$. Both rules are defined as follows (Sánchez et al. 2015).
- 5.2** The learning rule for estimating $\pi_{(i,j|k)}^\iota$ is as follows

$$\hat{\pi}_{(\hat{i},\hat{j}|\hat{k})}^\iota(t) = \frac{f_{(\hat{i},\hat{j}|\hat{k})}^\iota(t-1) + 1}{f_{(\hat{i}|\hat{k})}^\iota(t-1) + 1}$$

given

$$f_{(i|k)}^\iota(t) = \sum_{t=1}^n \chi \left(s^\iota(t) = s_{(i)}^\iota, a^\iota(t) = a_{(k)}^\iota \right)$$

$$f_{(i,j|k)}^\iota(n) = \sum_{t=1}^n \chi \left(s^\iota(t+1) = s_{(j)}^\iota | s^\iota(t) = s_{(i)}^\iota, a^\iota(t) = a_{(k)}^\iota \right)$$

such that

$$\chi(\mathcal{E}_t) = \begin{cases} 1 & \text{if the event } \mathcal{E} \text{ occurs at interaction } t \\ 0 & \text{otherwise} \end{cases}$$

where $f_{(i|k)}^\iota(t)$ (the frequency of observed experiences) is the total number of times that the player ι evolves from state i applying action k in the RL process and, $f_{(i,j|k)}^\iota$ is the total number of times that player ι evolves from state i to state j applying action k in the RL process. For

$$f_{(\hat{i},\hat{j}|\hat{k})}^\iota(0) = 0, \quad \text{and} \quad f_{(\hat{i}|\hat{k})}^\iota(0) = 0.$$

Remark 10 The normalization of this processes is obtained by the projection on simplex. Given that $\sum_j \hat{\pi}_{\hat{i}j|\hat{k}}^\iota \neq 1$, at each step, the complete row $\hat{i}$ is projected to the simplex ($\mathcal{P} : \hat{\pi}_{\hat{i}|\hat{k}}^\iota(t) \rightarrow S^{N_\iota}$).

- 5.3** The learning rule for estimating $J_{(i,j|k)}^\iota$ is given by

$$\hat{J}_{(\hat{i},\hat{j}|\hat{k})}^\iota(t) = \frac{V_{(\hat{i},\hat{j}|\hat{k})}^\iota(t)}{f_{(\hat{i},\hat{j}|\hat{k})}^\iota(t)}$$

- 5.4** As well as, for the cost/reward model, we keep a running average of the rewards observed upon taking each action in each state as follows

$$V_{(\hat{i},\hat{j}|\hat{k})}^\iota(t) = \sum_{n=1}^t B_J^\iota(n) \chi \left(s(n+1) = s_{(\hat{j})} | s(n) = s_{(\hat{i})}, a(n) = a_{(\hat{k})} \right)$$

such that

$$B_J^\iota := J_{(\hat{i},\hat{j}|\hat{k})}^\iota + (\Delta_J) \eta, \quad \text{for } \Delta_J \leq J_{(\hat{i},\hat{j}|\hat{k})}^\iota \text{ and } \eta = \text{rand}([-1, 1])$$

where $V_{(\hat{i},\hat{j}|\hat{k})}^\iota(t)$ is the sum over all immediate costs/rewards, received after executing action a in state i and stepping to state j , incremented by Δ_J multiplied by a random value η , $-1 \leq \eta \leq 1$. The parameter η is a learning rate that can depend on the salience properties of both the unconditional and the conditional stimuli being associated.

Remark 11 The learning rules of the architecture are computed considering the maximum likelihood model where $\left(\frac{0}{0} := 0\right)$.

- 5.5** The designing of the adaptive module for the actor-critic architecture consists of the following learning rules. The definition involves the variable t_0 which is the time required to compute the corresponding matrices $\hat{\pi}_{(\hat{i},\hat{j}|\hat{k})}^\iota(t)$ and $\hat{J}_{(\hat{i},\hat{j}|\hat{k})}^\iota(t)$.

5.6 The learning rule for estimating $\pi_{(i,j|\hat{k})}^\ell$ is given by

$$\hat{\pi}_{(i,j|\hat{k})}^\ell(t) = \frac{f_{(i,j|\hat{k})}^\ell(t)}{f_{(i|\hat{k})}^\ell(t)} = \frac{f_{(i,j|\hat{k})}^\ell(t_0) + \xi_{(i,j|\hat{k})}^{\ell,1}(t)}{f_{(i|\hat{k})}^\ell(t_0) + \xi_{(i|\hat{k})}^{\ell,2}(t)} = \frac{\hat{\pi}_{(i,j|\hat{k})}^\ell(t_0) + \frac{1}{f_{(i|\hat{k})}^\ell(t_0)} \xi_{(i,j|\hat{k})}^{\ell,1}(t)}{1 + \frac{1}{f_{(i|\hat{k})}^\ell(t_0)} \xi_{(i|\hat{k})}^{\ell,2}(t)}$$

where

$$\xi_{(i,j|\hat{k})}^{\ell,1}(t) = \sum_{n=t_0+1}^t \chi \left(s^\ell(n+1) = s_{(j)}^\ell | s^\ell(n) = s_{(i)}^\ell, a^\ell(n) = a_{(\hat{k})}^\ell \right) \cdot$$

$$\chi \left(s^\ell(n) = s_{(i)}^\ell, a^\ell(n) = a_{(\hat{k})}^\ell \right)$$

and

$$\xi_{(i|\hat{k})}^{\ell,2}(t) = \sum_{n=t_0+1}^t \chi \left(s^\ell(n) = s_{(i)}^\ell, a^\ell(n) = a_{(\hat{k})}^\ell \right) \cdot$$

5.7 The estimation of the transition matrix Π for the random variables s_i , s_j and a_k is considered only if the estimated error e decreases for any predecessor estimated transition matrix $\hat{\Pi}(t-1)$ of $\Pi(t)$, i.e. $e(\hat{\Pi}(t-1) - \hat{\Pi}(t)) > 0$.

5.8 The learning rule for estimating $J_{(i,j|\hat{k})}^\ell$ is as follows

$$\hat{J}_{(i,j|\hat{k})}^\ell(t) = \frac{J_{(i,j|\hat{k})}^\ell(t_0) + \sum_{n=t_0+1}^t B_J^\ell(n) \chi \left(s^\ell(n+1) = s_{(j)}^\ell | s^\ell(n) = s_{(i)}^\ell, a^\ell(n) = a_{(\hat{k})}^\ell \right)}{f_{(i,j|\hat{k})}^\ell(t_0) + \sum_{n=t_0+1}^t \chi \left(s^\ell(n+1) = s_{(j)}^\ell | s^\ell(n) = s_{(i)}^\ell, a^\ell(n) = a_{(\hat{k})}^\ell \right)}$$

5.9 We have a RL process for representing the concept of Machiavellian immorality where if there exist changes in the system the players are able to learn and adapt to the environment. We can estimate the aleatory variables corresponding to the entry $(\hat{i}, \hat{j}, \hat{k})$ of both, the transition and the cost/reward matrices. The principle of error-driven adjustment of cost/reward predictions given by the estimated error e contributes to the players' acquisition of moral (immoral) behavior. The RL process converges because the ergodicity restriction imposed on the Markov game.

Numerical example

6.1 This example shows a proof-of-concept experimental result conducted using the proposed model. The environment has four player (two manipulators and two manipulated), four states and two actions. The dynamics of the process can be observed by studying the time varying estimated cost functions $\hat{J}_{ij|\hat{k}}$, estimated transition matrices $\hat{\pi}_{ij|\hat{k}}$, the estimated strategies $\hat{d}_{i|\hat{k}}$ and the error e .

6.2 For the purpose of implementation, the learning process can be conceptualized in two phases, where each phase corresponds to learning a single module. At the beginning of Phase 1, the players starts exploring by executing the actions considering a softmax action selection rules considering a Boltzmann distribution. After ten thousand algorithm iterations, the estimated transition matrices and cost functions of all the players begins to stabilize. In this sense, the error decreases and falls below a given threshold. In Phase 2 the players begins to explore by executing the actions based on the solution of the game solver. After ten hundred algorithm iterations,

(a) Manipulator 1

(b) Manipulator 2

Figure 2: Error estimated Transition Matrices $\pi_{(i,j|k)}^l$ of the Manipulators players.

(a) Manipulated 1

(b) Manipulated 2

Figure 3: Error estimated Transition Matrices $\pi_{(i,j|k)}^l$ of the Manipulated players.

the process converges.

6.3 The estimated values for $\hat{\pi}_{ij|k}$ are as follows:

$$\begin{aligned} \hat{\pi}_{ij|1}^{(1)} &= \begin{bmatrix} 0.2623 & 0.2475 & 0.2390 & 0.2512 \\ 0.2490 & 0.2448 & 0.2450 & 0.2612 \\ 0.2409 & 0.2665 & 0.2492 & 0.2434 \\ 0.2492 & 0.2314 & 0.2585 & 0.2609 \end{bmatrix} & \hat{\pi}_{ij|2}^{(1)} &= \begin{bmatrix} 0.3084 & 0.2327 & 0.2381 & 0.2208 \\ 0.2596 & 0.2489 & 0.2522 & 0.2393 \\ 0.2580 & 0.2587 & 0.2301 & 0.2532 \\ 0.2313 & 0.2595 & 0.2470 & 0.2622 \end{bmatrix} \\ \hat{\pi}_{ij|1}^{(2)} &= \begin{bmatrix} 0.2220 & 0.2560 & 0.2565 & 0.2656 \\ 0.2499 & 0.2619 & 0.2366 & 0.2516 \\ 0.2628 & 0.2463 & 0.2493 & 0.2416 \\ 0.2277 & 0.2674 & 0.2423 & 0.2625 \end{bmatrix} & \hat{\pi}_{ij|2}^{(2)} &= \begin{bmatrix} 0.2683 & 0.2496 & 0.2492 & 0.2329 \\ 0.2664 & 0.2296 & 0.2317 & 0.2723 \\ 0.2360 & 0.3116 & 0.1898 & 0.2626 \\ 0.2608 & 0.2532 & 0.2160 & 0.2700 \end{bmatrix} \\ \hat{\pi}_{ij|1}^{(3)} &= \begin{bmatrix} 0.2435 & 0.2383 & 0.2561 & 0.2621 \\ 0.2434 & 0.2591 & 0.2421 & 0.2553 \\ 0.2552 & 0.2568 & 0.2404 & 0.2476 \\ 0.2736 & 0.2435 & 0.2536 & 0.2293 \end{bmatrix} & \hat{\pi}_{ij|1}^{(3)} &= \begin{bmatrix} 0.2370 & 0.2758 & 0.2214 & 0.2659 \\ 0.2419 & 0.2612 & 0.2533 & 0.2436 \\ 0.2829 & 0.1994 & 0.2799 & 0.2379 \\ 0.2531 & 0.2396 & 0.2625 & 0.2448 \end{bmatrix} \\ \hat{\pi}_{ij|1}^{(4)} &= \begin{bmatrix} 0.2555 & 0.2537 & 0.2432 & 0.2476 \\ 0.2528 & 0.2508 & 0.2358 & 0.2606 \\ 0.2323 & 0.2601 & 0.2538 & 0.2537 \\ 0.2500 & 0.2448 & 0.2368 & 0.2684 \end{bmatrix} & \hat{\pi}_{ij|1}^{(4)} &= \begin{bmatrix} 0.2539 & 0.2787 & 0.2450 & 0.2225 \\ 0.2746 & 0.2268 & 0.2595 & 0.2392 \\ 0.2562 & 0.2359 & 0.2957 & 0.2122 \\ 0.2520 & 0.2399 & 0.2711 & 0.2370 \end{bmatrix} \end{aligned}$$

Figure 4: Error estimated Utilities $\hat{J}$ of the Manipulators players.

Figure 5: Error estimated Utilities $\hat{J}$ of the Manipulated players.

6.4 The estimated values for $\hat{J}_{ij|k}$ are as follows:

$$\begin{aligned} \hat{J}_{ij|1}^{(1)} &= \begin{bmatrix} 7857 & 3882 & 806 & 1883 \\ 6514 & 117 & 15826 & 1336 \\ 1218 & 1886 & 1154 & 2676 \\ 4739 & 5625 & 490 & 450 \end{bmatrix} & \hat{J}_{ij|2}^{(1)} &= \begin{bmatrix} 4944.6 & 4704.9 & 5008.7 & 2752.7 \\ 2546.5 & 193.6 & 3776 & 717.9 \\ 9263.7 & 4265.1 & 1327.2 & 2890 \\ 2874.2 & 7849.6 & 1352.2 & 65.1 \end{bmatrix} \\ \hat{J}_{ij|1}^{(2)} &= \begin{bmatrix} 1147 & 10792 & 1557 & 1067 \\ 76 & 2209 & 2071 & 1079 \\ 860 & 2085 & 5903 & 633 \\ 3462 & 243 & 1451 & 6401 \end{bmatrix} & \hat{J}_{ij|2}^{(2)} &= \begin{bmatrix} 1568.5 & 588.7 & 733.3 & 4931.7 \\ 90.3 & 970.9 & 755.2 & 1201.4 \\ 2482.6 & 977.9 & 4707.3 & 1577.2 \\ 8125.2 & 2228.8 & 1753.2 & 687 \end{bmatrix} \\ \hat{J}_{ij|1}^{(3)} &= \begin{bmatrix} 9536.9 & 648.8 & 1370.6 & 85.2 \\ 4720.3 & 1781.1 & 484.2 & 53.5 \\ 1961.6 & 6914.4 & 3302.1 & 3939.7 \\ 1939.2 & 1431.4 & 741.1 & 3511.8 \end{bmatrix} & \hat{J}_{ij|2}^{(3)} &= \begin{bmatrix} 2634.7 & 1766 & 3462.4 & 1701.7 \\ 4822.3 & 186.1 & 23.7 & 1645 \\ 5466 & 1294.1 & 898.1 & 3894.2 \\ 2345 & 940.1 & 529.1 & 988.1 \end{bmatrix} \\ \hat{J}_{ij|1}^{(4)} &= \begin{bmatrix} 996.4 & 1554.9 & 1066.6 & 5614 \\ 42.9 & 605.7 & 3933.7 & 5619.2 \\ 59.6 & 2357.1 & 1829.4 & 4045.1 \\ 3112 & 1260.3 & 1455.8 & 967.8 \end{bmatrix} & \hat{J}_{ij|2}^{(4)} &= \begin{bmatrix} 17.2 & 2887.8 & 193.9 & 4634.5 \\ 5704.7 & 5338.4 & 4689.8 & 3532.1 \\ 643.3 & 503 & 21.9 & 1066.5 \\ 9172 & 612.5 & 1626.1 & 3490.6 \end{bmatrix} \end{aligned}$$

6.5 Figures 2(a) and 2(b) show the error of the estimated transition matrix $\hat{\pi}_{ij|k}$ for the manipulator leader 1 and the error of the estimated transition matrix $\hat{\pi}_{ij|k}$ for the manipulator leader 2. Figures 3(a) and 2(b) show the error of the estimated transition matrix $\hat{\pi}_{ij|k}$ for the manipulated follower 1 and the manipulated follower 2. The order of the error e is 1×10^{-6} which is very small for our example. As well as, Figures 4 and 5 show the error of the estimated utility $\hat{J}_{ij|k}$.

6.6 Figures 6 and 7 show the convergence to the manipulation (Machiavellian) equilibrium point. The estimated

Figure 6: Convergence of the Strategies c of the Manipulators players.

Figure 7: Convergence of the Strategies c of the Manipulators players.

values for $\hat{d}_{i|k}$ are as follows:

$$\begin{aligned} \hat{d}_{i|1k}^{(1)*} &= \begin{bmatrix} 0.5355 & 0.4645 \\ 0.7965 & 0.2035 \\ 0.2105 & 0.7895 \\ 0.2904 & 0.7096 \end{bmatrix} & \hat{d}_{i|1k}^{(2)*} &= \begin{bmatrix} 0.7937 & 0.2063 \\ 0.5349 & 0.4651 \\ 0.6431 & 0.3569 \\ 0.3161 & 0.6839 \end{bmatrix} \\ \hat{d}_{i|1k}^{(3)*} &= \begin{bmatrix} 0.2757 & 0.7243 \\ 0.5183 & 0.4817 \\ 0.2780 & 0.7220 \\ 0.2812 & 0.7188 \end{bmatrix} & \hat{d}_{i|1k}^{(4)*} &= \begin{bmatrix} 0.2763 & 0.7237 \\ 0.7206 & 0.2794 \\ 0.2749 & 0.7251 \\ 0.7101 & 0.2899 \end{bmatrix} \end{aligned}$$

Conclusions and future work

- 7.1 In this paper we developed a manipulation model where the manipulator players can anticipate the predicted response of the manipulated players. The proposed Machiavellian game theory approach established a system that analyzes and predicts how players behave in strategic situations combining: strategic thinking, best-reply, and mutual consistency (equilibrium). We justified the mutual consistency to predict how players are likely to behave by introducing a reinforcement learning process. In fact, the equilibrium is the result of a strategic thinking, optimization, and equilibration (or learning).
- 7.2 In our proposed model the manipulators players play a Machiavellian strategy first and the manipulated players move sequentially in terms of the actions proposed by the manipulators. In the dynamics of the model the manipulators players consider the best-reply of the manipulated players, and then select the Machiavellian strategy that optimizes their utility, anticipating the responses of the manipulated players. Subsequently, the manipulated players observe the Machiavellian strategy played by the manipulators players and select their best-reply strategy. The Machiavellian behavior described above coincides with a Stackelberg game. Strategies are selected according to the Machiavellian social conduct theory.

- 7.3** We introduced the utilitarian (immoral behavior) and deontological moral theories for modeling the Machiavellian immorality and, we reviewed psychological findings regarding moral decision making. An advantage of the Machiavellianism is that whereas other moral theories provide standards for how we should act, they do not describe how moral judgments and decisions are achieved in practice. We established a relationship between moral and game theory (economic theories) and we suggested that the reinforcement learning theory, with its principle of error-driven adjustment of cost/reward predictions, provided evidences of how players acquire moral (immoral) behavior.
- 7.4** The introduction of the reinforcement learning theory suggested on the one hand, a mechanism by which utilitarian moral value may guide the Machiavellian behavior, as well as, the deontological moral obtain a value for certain moral principles or acts. We noted some minor differences with the common usage of classical reinforcement learning because the control strategy needs to change as the actor computes the game. This need not pose any problems, as long as the actor parameters are updated on a slower time scale. The main advantage of our approach is that it preserves the convergence properties of the extraproximal method. The result of the model is the manipulation equilibrium point. Validity of the proposed method was successfully demonstrated both theoretically and by a simulated numerical example.
- 7.5** There exist a number of challenges left to be addressed for future work. The obvious next step is to incorporate our model into an implemented explanatory system for testing it with real-life scenarios. An interesting technical challenge is that of finding an analytical solution for the manipulation equilibrium point. An additional interesting problem is to associate the manipulation equilibrium point proposed in this paper with the bagam-ing problem where players should cooperate when non-cooperation leads to Pareto-inefficient results. It will be interested also to examine the problem of the existence of a manipulation equilibrium point under conditions of moral hazard (when one player takes more risks because someone else bears the cost of those risks). These extensions should lead to a more complete account of complex manipulation understanding.

Appendix: Extraproximal

The individual cost-functions of the leaders are defined as follows:

Cost-functions for the leaders

$$\mathbf{J}^1 \left(c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(2)} | c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} \right) = \sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(1)} \left(c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} \right)$$

$$\mathbf{J}^2 \left(c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(2)} | c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} \right) = \sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(2)} \left(c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} \right)$$

Cost-functions for the followers

$$\mathbf{J}^3 \left(c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} | c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(2)} \right) = \sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(3)} \left(c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} \right) \left(c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(2)} \right)$$

$$\mathbf{J}^4 \left(c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} | c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(2)} \right) = \sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(4)} \left(c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} \right) \left(c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(2)} \right)$$

Considering that

$$\begin{aligned} \mathbf{J}^1 \left(\hat{c}_{(i_1, k_1)}^{(1)}, \hat{c}_{(i_2, k_2)}^{(2)} | c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} \right) &\leq \mathbf{J}^1 \left(c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(2)} | c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} \right) \\ \mathbf{J}^2 \left(\hat{c}_{(i_1, k_1)}^{(1)}, \hat{c}_{(i_2, k_2)}^{(2)} | c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} \right) &\leq \mathbf{J}^2 \left(c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(2)} | c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} \right) \end{aligned}$$

and

$$\begin{aligned} \mathbf{J}^3 \left(\hat{c}_{(i_3, k_3)}^{(3)}, \hat{c}_{(i_4, k_4)}^{(4)} | c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(2)} \right) &\leq \mathbf{J}^3 \left(c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} | c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(2)} \right) \\ \mathbf{J}^4 \left(\hat{c}_{(i_3, k_3)}^{(3)}, \hat{c}_{(i_4, k_4)}^{(4)} | c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(2)} \right) &\leq \mathbf{J}^4 \left(c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} | c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(2)} \right) \end{aligned}$$

We have that for the leaders that

$$\begin{aligned} \sum_{l=1}^N \varphi_l(\bar{u}^l, u^l | v) &= \mathbf{J}^1 \left(\hat{c}_{(i_1, k_1)}^{(1)}, \hat{c}_{(i_2, k_2)}^{(2)} | c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} \right) + \mathbf{J}^2 \left(\hat{c}_{(i_1, k_1)}^{(1)}, \hat{c}_{(i_2, k_2)}^{(2)} | c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} \right) = \\ &\sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(1)} \left(\hat{c}_{(i_1, k_1)}^{(1)}, \hat{c}_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} \right) + \\ &\sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(2)} \left(\hat{c}_{(i_1, k_1)}^{(1)}, \hat{c}_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)}, c_{(i_4, k_4)}^{(4)} \right) \end{aligned} \quad (15)$$

and

$$\sum_{l=1}^{\mathcal{N}} \varphi_l(u^l, u^{\hat{l}}|v) = \mathbf{J}^1 \left(c_{(i_1, k_1)}^{(1)}, c_{(i_2, k_2)}^{(\hat{2})} | c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) + \mathbf{J}^2 \left(c_{(i_1, k_1)}^{(\hat{1})}, c_{(i_2, k_2)}^{(2)} | c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) =$$

$$\sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(1)} \left(c_{(i_1, k_1)}^{(1)} c_{(i_2, k_2)}^{(\hat{2})} \right) \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) +$$

$$\sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(2)} \left(c_{(i_1, k_1)}^{(\hat{1})} c_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right)$$

For computing the Nash condition for the leaders we define the function $g(u, \hat{u}(u)|v)$ as follows

$$g_l(u^l, \hat{u}^l(u)|v) = \sum_{l=1}^{\mathcal{N}} \varphi_l(\bar{u}^l, u^{\hat{l}}|v) - \varphi_l(u^l, u^{\hat{l}}|v)$$

$$u_l = (\bar{u}^l, u^{\hat{l}}), \hat{u}_l(u) = (u^l, u^{\hat{l}})$$

$$g(u, \hat{u}(u)|v) = \sum_l g_l(u_l, \hat{u}_l(u)|v)$$

$$u = (u_1^T, \dots, u_n^T)^T, \hat{u}_l(u) = (\hat{u}_1^T(u), \dots, \hat{u}_n^T(u))^T$$

$$u = \begin{pmatrix} col \left\| \begin{matrix} \hat{c}_{(i_1, k_1)}^{(1)} \\ \hat{c}_{(i_2, k_2)}^{(\hat{2})} \\ \hat{c}_{(i_3, k_3)}^{(3)} \\ \hat{c}_{(i_4, k_4)}^{(4)} \\ \hat{c}_{(i_1, k_1)}^{(\hat{1})} \\ \hat{c}_{(i_2, k_2)}^{(2)} \\ \hat{c}_{(i_3, k_3)}^{(3)} \\ \hat{c}_{(i_4, k_4)}^{(4)} \end{matrix} \right\| \\ \end{pmatrix}, \hat{u}(u) = \begin{pmatrix} col \left\| \begin{matrix} c_{(i_1, k_1)}^{(1)} \\ c_{(i_2, k_2)}^{(2)} \\ c_{(i_3, k_3)}^{(3)} \\ c_{(i_4, k_4)}^{(4)} \\ c_{(i_1, k_1)}^{(\hat{1})} \\ c_{(i_2, k_2)}^{(\hat{2})} \\ c_{(i_3, k_3)}^{(\hat{3})} \\ c_{(i_4, k_4)}^{(\hat{4})} \end{matrix} \right\| \\ \end{pmatrix} \quad (16)$$

For $n = 0, 1, \dots$ let define the vectors

$$u_n = \begin{pmatrix} col \ c^{(1)}(n) \\ col \ c^{(2)}(n) \end{pmatrix}$$

$$\hat{u}_n(u) = u^{\hat{l}} = \begin{pmatrix} col \ c^{(\hat{1})}(n) \\ col \ c^{(\hat{2})}(n) \end{pmatrix}$$

$$\bar{u}^l = \begin{pmatrix} col \ \hat{c}^{(1)}(n) \\ col \ \hat{c}^{(2)}(n) \end{pmatrix} = \begin{pmatrix} col \ \arg \min_{z \in C_{adm}^{(1)}} \mathbf{J}^{(1)}(z, c^{(\hat{2})}(n), c^{(3)}(n), c^{(4)}(n)) \\ col \ \arg \min_{z \in C_{adm}^{(2)}} \mathbf{J}^{(1)}(c^{(\hat{1})}(n), z, c^{(3)}(n), c^{(4)}(n)) \end{pmatrix}$$

The regularized problem for $G(u, \hat{u}(u)|v)$ is defined as follows

$$G_{\delta}(u, \hat{u}(u)|v) = \sum_{l=1}^{\mathcal{N}} \varphi_l(\bar{u}^l, u^{\hat{l}}|v) - \varphi_l(u^l, u^{\hat{l}}|v) =$$

$$\sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(1)} \left(\hat{c}_{(i_1, k_1)}^{(1)} \hat{c}_{(i_2, k_2)}^{(\hat{2})} \right) \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) +$$

$$\sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(2)} \left(c_{(i_1, k_1)}^{(\hat{1})} \hat{c}_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) -$$

$$\sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(1)} \left(c_{(i_1, k_1)}^{(1)} c_{(i_2, k_2)}^{(\hat{2})} \right) \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) -$$

$$\sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(2)} \left(c_{(i_1, k_1)}^{(\hat{1})} c_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) +$$

$$\frac{\delta}{2} \left(\left\| \begin{matrix} c_{(i_1, k_1)}^{(1)} \\ c_{(i_2, k_2)}^{(2)} \end{matrix} \right\|^2 + \left\| \begin{matrix} c_{(i_1, k_1)}^{(\hat{1})} \\ c_{(i_2, k_2)}^{(\hat{2})} \end{matrix} \right\|^2 \right)$$

and for $g(u, \hat{u}(u))$ we have

$$\begin{aligned}
 g_{\delta}(u, \hat{u}(u)|v) = & \sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(1)} \left(\hat{c}_{(i_1, k_1)}^{(1)} \hat{c}_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) + \\
 & \sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(2)} \left(\hat{c}_{(i_1, k_1)}^{(1)} \hat{c}_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) - \\
 & \sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(1)} \left(c_{(i_1, k_1)}^{(1)} c_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) - \\
 & \sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(2)} \left(c_{(i_1, k_1)}^{(1)} c_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) + \\
 & \frac{\delta}{2} \left(\left\| \begin{pmatrix} c_{(i_1, k_1)}^{(1)} \\ c_{(i_2, k_2)}^{(2)} \end{pmatrix} \right\|^2 + \left\| \begin{pmatrix} \hat{c}_{(i_1, k_1)}^{(1)} \\ \hat{c}_{(i_2, k_2)}^{(2)} \end{pmatrix} \right\|^2 \right)
 \end{aligned}$$

Remark 12 We have a similar development for the followers

Let us apply the Lagrange principle, then we have

$$\begin{aligned}
 \mathcal{L}_{\delta}(u, \hat{u}(u), v, \hat{v}(v), \omega, \xi) := & \theta G_{L_p, \delta}(u(\lambda), \hat{u}(u)|v) + \omega g_{\delta}(u, \hat{u}(u)|v) + \xi f_{\delta}(v, \hat{v}(v)|u) - \frac{\delta}{2} (\xi^2 + \omega^2) = \\
 & (\theta + \omega) \left[\sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(1)} \left(\hat{c}_{(i_1, k_1)}^{(1)} \hat{c}_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) + \right. \\
 & \sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(2)} \left(\hat{c}_{(i_1, k_1)}^{(1)} \hat{c}_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) - \\
 & \sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(1)} \left(c_{(i_1, k_1)}^{(1)} c_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) - \\
 & \left. \sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(2)} \left(c_{(i_1, k_1)}^{(1)} c_{(i_2, k_2)}^{(2)} \right) \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) + \right. \\
 & \left. \frac{\delta}{2} \left(\left\| \begin{pmatrix} c_{(i_1, k_1)}^{(1)} \\ c_{(i_2, k_2)}^{(2)} \end{pmatrix} \right\|^2 + \left\| \begin{pmatrix} \hat{c}_{(i_1, k_1)}^{(1)} \\ \hat{c}_{(i_2, k_2)}^{(2)} \end{pmatrix} \right\|^2 \right) \right] + \\
 & \xi \left[\sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(3)} \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) \left(c_{(i_1, k_1)}^{(1)} c_{(i_2, k_2)}^{(2)} \right) + \right. \\
 & \sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(4)} \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) \left(c_{(i_1, k_1)}^{(1)} c_{(i_2, k_2)}^{(2)} \right) - \\
 & \sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(3)} \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) \left(c_{(i_1, k_1)}^{(1)} c_{(i_2, k_2)}^{(2)} \right) - \\
 & \left. \sum_{i_1, k_1} \sum_{i_2, k_2} \sum_{i_3, k_3} \sum_{i_4, k_4} W_{(i_1, k_1; i_2, k_2; i_3, k_3; i_4, k_4)}^{(4)} \left(c_{(i_3, k_3)}^{(3)} c_{(i_4, k_4)}^{(4)} \right) \left(c_{(i_1, k_1)}^{(1)} c_{(i_2, k_2)}^{(2)} \right) - \right. \\
 & \left. \frac{\delta}{2} \left(\left\| \begin{pmatrix} c_{(i_3, k_3)}^{(3)} \\ c_{(i_4, k_4)}^{(4)} \end{pmatrix} \right\|^2 + \left\| \begin{pmatrix} \hat{c}_{(i_3, k_3)}^{(3)} \\ \hat{c}_{(i_4, k_4)}^{(4)} \end{pmatrix} \right\|^2 \right) \right] - \frac{\delta}{2} (\xi^2 + \omega^2)
 \end{aligned}$$

The general format iterative version ($n = 0, 1, \dots$) of the extraproximal method (Antipin 2005; Trejo et al. 2015, 2016) with some fixed admissible initial values ($u_0 \in U, \hat{u}_0 \in U, v_0 \in V, \hat{v}_0 \in V, \omega_0 \geq 0, \xi_0 \geq 0$) is as follows:

1. The *first half-step* (prediction):

$$\begin{aligned}
 \bar{\omega}_n &= \arg \min_{\omega \geq 0} \left\{ \frac{1}{2} \|\omega - \omega_n\|^2 - \gamma \mathcal{L}_{\delta}(u_n, \hat{u}_n(u), v_n, \hat{v}_n(v), \omega, \bar{\xi}_n) \right\} \\
 \bar{\xi}_n &= \arg \min_{\xi \geq 0} \left\{ \frac{1}{2} \|\xi - \xi_n\|^2 - \gamma \mathcal{L}_{\delta}(u_n, \hat{u}_n(u), v_n, \hat{v}_n(v), \bar{\omega}_n, \xi) \right\} \\
 \bar{u}_n &= \arg \min_{u \in U} \left\{ \frac{1}{2} \|u - u_n\|^2 + \gamma \mathcal{L}_{\delta}(u, \hat{u}_n(u), v_n, \hat{v}_n(v), \bar{\omega}_n, \bar{\xi}_n) \right\} \\
 \bar{\hat{u}}_n &= \arg \min_{\hat{u} \in \hat{U}} \left\{ \frac{1}{2} \|\hat{u} - \hat{u}_n\|^2 + \gamma \mathcal{L}_{\delta}(u_n, \hat{u}(u), v_n, \hat{v}_n(v), \bar{\omega}_n, \bar{\xi}_n) \right\}
 \end{aligned} \tag{17}$$

$$\bar{v}_n = \arg \min_{v \in V} \left\{ \frac{1}{2} \|v - v_n\|^2 - \gamma \mathcal{L}_\delta(u_n, \hat{u}_n(u), v, \hat{v}_n(v), \bar{\omega}_n, \bar{\xi}_n) \right\}$$

$$\bar{\hat{v}}_n = \arg \min_{\hat{v} \in \hat{V}} \left\{ \frac{1}{2} \|\hat{v} - \hat{v}_n\|^2 - \gamma \mathcal{L}_\delta(u_n, \hat{u}_n(u), v_n, \hat{v}(v), \bar{\omega}_n, \bar{\xi}_n) \right\}$$

2. The second (basic) half-step

$$\begin{aligned}\omega_{n+1} &= \arg \min_{\omega \geq 0} \left\{ \frac{1}{2} \|\omega - \omega_n\|^2 - \gamma \mathcal{L}_\delta(\bar{u}_n, \bar{\hat{u}}_n(u), \bar{v}_n, \bar{\hat{v}}_n(v), \omega, \bar{\xi}_n) \right\} \\ \xi_{n+1} &= \arg \min_{\xi \geq 0} \left\{ \frac{1}{2} \|\xi - \xi_n\|^2 - \gamma \mathcal{L}_\delta(\bar{u}_n, \bar{\hat{u}}_n(u), \bar{v}_n, \bar{\hat{v}}_n(v), \bar{\omega}_n, \xi) \right\} \\ u_{n+1} &= \arg \min_{u \in U} \left\{ \frac{1}{2} \|u - u_n\|^2 + \gamma \mathcal{L}_\delta(u, \bar{\hat{u}}_n(u), \bar{v}_n, \bar{\hat{v}}_n(v), \bar{\omega}_n, \bar{\xi}_n) \right\} \\ \hat{u}_{n+1} &= \arg \min_{\hat{u} \in \hat{U}} \left\{ \frac{1}{2} \|\hat{u} - \hat{u}_n\|^2 + \gamma \mathcal{L}_\delta(\bar{u}_n, \hat{u}(u), \bar{v}_n, \bar{\hat{v}}_n(v), \bar{\omega}_n, \bar{\xi}_n) \right\} \\ v_{n+1} &= \arg \min_{v \in V} \left\{ \frac{1}{2} \|v - v_n\|^2 - \gamma \mathcal{L}_\delta(\bar{u}_n, \bar{\hat{u}}_n(u), v, \bar{\hat{v}}_n(v), \bar{\omega}_n, \bar{\xi}_n) \right\} \\ \hat{v}_{n+1} &= \arg \min_{\hat{v} \in \hat{V}} \left\{ \frac{1}{2} \|\hat{v} - \hat{v}_n\|^2 - \gamma \mathcal{L}_\delta(\bar{u}_n, \bar{\hat{u}}_n(u), \bar{v}_n, \hat{v}(v), \bar{\omega}_n, \bar{\xi}_n) \right\}\end{aligned} \tag{18}$$

References

- Antipin, A. S. (2005). An extraproximal method for solving equilibrium programming problems and games. *Computational Mathematics and Mathematical Physics*, 45(11), 1893–1914
- Bales, R. E. (1971). Act-utilitarianism: Account of right-making characteristics or decision-making procedures? *American Philosophical Quarterly*, 8, 257–265
- Byrne, R. & Whiten, A. (1988). *Machiavellianism Intelligence: The Evolution of the Intellect in Monkeys, Apes and Humans*. Clarendon Press
- Calhoun, R. P. (1969). Niccolò Machiavelli and the 20th century administrator. *Academy of Management Journal*, 12, 205–212
- Cervantes, J. A., Rodriguez, L. F., Lopez, S., Ramos, F. & Robles, F. (2011). Autonomous agents and ethical decision-making. *Cognitive Computation*, 8, 278–296
- Christie, R. & Geis, F. (1970). *Studies in Machiavellianism*. Academic Press
- Cleckley, H. (1976). *The Mask of sanity*. Mosby, 5th edn.
- Clempner, J. B. & Poznyak, A. S. (2014). Simple computing of the customer lifetime value: A fixed local-optimal policy approach. *Journal of Systems Science and Systems Engineering*, 23(4), 439–459
- Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press
- Dawkins, R. & Krebs, J. (1978). *Animal Signals: Information or Manipulation*. Blackwell
- Dehghani, M., Tomai, E., Forbus, K. & Klenk, M. (2008). An integrated reasoning approach to moral decision-making. In *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, (pp. 1280–1286). Chicago, Illinois, USA
- Falbo, T. (1977). Multidimensional scaling of power strategies. *Journal of Personality and Social Psychology*, 35, 537–547
- Gable, M., Hollon, C. & Dangello, F. (1992). Managerial structuring of work as a moderator of the machiavellianism and job performance relationship. *The Journal of Psychology*, 126(3), 317–325
- Grams, W. & Rogers, R. (1990). Power and personality: Effects of machiavellianism, need for approval, and motivation on use of influence tactics. *The Journal of General Psychology*, 117(1), 71–82
- Hartog, D. N. D. & Belschak, F. D. (2012). Work engagement and Machiavellianism in the ethical leadership process. *Journal of Business Ethics*, 107, 35–47

- Hellriegel, D., Slocum Jr., J. & Woodman, R. (1997). *Organizational Behavior*. South-Western Educational Publishing
- Indurkha, B. & Misztal-Radecka, J. (2016). Incorporating human dimension in autonomous decision-making on moral and ethical issues. In *AAAI Spring Symposium Series*, (pp. 226–230). Stanford University
- Leary, M., Knight, R. & Barns, B. (1986). Ethical ideologies of the Machiavellian. *Personality and Social Psychology Bulletin*, 12, 75–80
- Machiavelli, N. (1952). *The Prince*. Encyclopedia Britannica, Inc.
- Machiavelli, N. (1965). *Discourses in the first ten books of Titus Livius*. Duke University Press
- Machiavelli, N. (2001). *The Art of War*. Da Capo Press
- Mudrack, P. (1989). Machiavellianism and locus of control: A meta-analytic review. *The Journal of Social Psychology*, 130(1), 125–126
- Poznyak, A. S. (2008). *Advanced Mathematical tools for Automatic Control Engineers. Deterministic Technique*, vol. 1. Elsevier, Amsterdam, Oxford
- Poznyak, A. S., Najim, K. & Gomez-Ramirez, E. (2000). *Self-learning control of finite Markov chains*. Marcel Dekker, Inc., New York,
- Prociuk, T. J. & Breen, L. J. (1976). Machiavellianism and locus of control. *The Journal of Social Psychology*, 9, 141–142
- Raven, B. H. (1993). Machiavellianism and locus of control. *The Journal of Social Psychology*, 43(4), 227–251
- Sánchez, E. M., Clempner, J. B. & Poznyak, A. S. (2015). A priori-knowledge/actor-critic reinforcement learning architecture for computing the mean-variance customer portfolio: The case of bank marmarket campaigns. *Engineering Applications of Artificial Intelligence*, 46, 82–92
- Schindler, J. (2012). Rethinking the tragedy of the commons: The integration of socio-psychological dispositions. *Journal of Artificial Societies and Social Simulation*, 15(1), 4. doi:10.18564/jasss.1822
- Smith, R. (1979). *The Psychopath in Society*. Academic Press
- Solar, D. & Bruehl, D. (1971). Machiavellianism and locus of control: Two conceptions of interpersonal power. *Psychological Reports*, 29, 1079–1082
- Stackelberg, H. v. (2011). *Market Structure and Equilibrium*. Springer
- Tobler, P. N., Kalis, A. & Kalenscher, T. (2008). The role of moral utility in decision making: An interdisciplinary framework. *Cognitive, Affective, & Behavioral Neuroscience*, 8(4), 390–401
- Trejo, K. K., Clempner, J. B. & Poznyak, A. S. (2015). Computing the Stackelberg/Nash equilibria using the extraproximal method: Convergence analysis and implementation details for Markov chains games. *International Journal of Applied Mathematics and Computer Science*, 25(2), 337–351
- Trejo, K. K., Clempner, J. B. & Poznyak, A. S. (2016). An optimal strong equilibrium solution for cooperative multi-leader-follower Stackelberg Markov chains games. *Kibernetika*, 52(2), 258–279
- Vecchio, R. P. & Sussmann, M. (1991). Choice of influence tactics: Individual and organizational determinants. *Journal of Organizational Behavior*, 12, 73–80
- Vleeming, R. (1979). Machiavellianism and locus of control: Two conceptions of interpersonal power. *Psychological Reports*, 44, 295–310
- Wallach, W., Franklin, S. & Allen, C. (2010). A conceptual and computational model of moral decision making in human and artificial agents. *Topics in Cognitive Science*, 2(3), 454–485
- Wijermans, N., Jorna, R., Jager, W., van Vliet, T. & Adang, O. (2013). Cross: Modelling crowd behaviour with social-cognitive agents. *Journal of Artificial Societies and Social Simulation*, 16(4), 1. doi:10.18564/jasss.2114
- Wilson, D., Near, D. & Miller, R. (1996). Machiavellianism: A synthesis of the evolutionary and psychological literatures. *Psychological Bulletin*, 119(2), 285–299

Xianyu, B. (2010). Social preference, incomplete information, and the evolution of ultimatum game in the small world networks: An agent-based approach. *Journal of Artificial Societies and Social Simulation*, 13(2), 7. doi: 10.18564/jasss.1534