Fundamenta Informaticae 156 (2017) 1–28
DOI 10.3233/FI-2015-1594
IOS Press

The Long-Term Benefits of Following Fairness Norms under Dynamics of Learning and Evolution

Emiliano Lorini*
Institut de Recherche en Informatique de Toulouse
Université Paul Sabatier, Toulouse, France
lorini@irit.fr

Roland Mühlenbernd†
Department of Linguistics
Eberhard Karls University, Tübingen, Germany
roland.muehlenbernd@uni-tuebingen.de

Abstract. In this study we present a game-theoretic model of guilt in relation to sensitivity to norms of fairness. We focus on a specific kind of fairness norm à la Rawls according to which a fair society should be organized so as to admit economic inequalities to the extent that they are beneficial to the less advantaged agents. In a first step, we analyze the impact of the sensitivity to this fairness norm on the behavior of agents who play a repeated Prisoner's Dilemma and learn via fictitious play. In a second step we transform the base game into a *meta-game* that represents a static description of a repeated game updated via fictitious play. We analyze such a meta-game under population dynamics by means of evolutionary game theory. Our results reveal two things: first of all, a great sensitivity to the fairness norm is beneficial in the long term when agents have the time to converge to mutual cooperation. Secondly, cooperativeness and fairness norm sensitivity can coevolve in a population of initially solely defectors.

Keywords: Fairness norm à la Rawls, Prisoner's Dilemma, fictitious play, evolutionary game theory

*Address for correspondence: Institut de Recherche en Informatique de Toulouse, Université Paul Sabatier, Toulouse, France
†I gratefully acknowledge the support of the ERC under project EVOLAEMP, http://www.evolaemp.uni-tuebingen.de

1. Introduction

Prototypical human and artificial societies (*e.g.*, a community, an organization) are populated by agents who have repeated encounters and can decide either to collaborate with the others thereby acting cooperatively, or to exploit the work of the others thereby acting selfishly. In a game-theoretic setting this kind of situations can be represented as an iterated Prisoner's Dilemma (PD) in which agents in the population have repeated one-to-one interactions with others (*i.e.*, at each round two agents in the population meet and play one-shot PD). The aim of this work is to study how fairness norms tend to emerge in this kind of societies in which agents are assumed (i) to be rational in the sense of being expected utility maximizers, and (ii) to learn from their past experiences. In the paper, we focus on a special kind of fairness norm à la Rawls [1] according to which a fair society should be organized so as to admit economic inequalities to the extent that they are beneficial to the less advantaged agents.

Our analysis is based on the general assumption that agents in the society are heterogeneous in the sense of being more or less sensitive to the fairness norm, where an agent's degree of norm sensitivity captures the extent to which the fairness norm has been internalized by the agent. The idea is that if a given norm is internalized by an agent then there is no need for an external sanction, a reward or punishment to ensure norm compliance. The agent is willing to comply with the norm because, if she does not do this, she will feel (morally) bad. We study the conditions under which an agent's disposition to follow the fairness norm à la Rawls (*i.e.*, the agent's sensitivity to the fairness norm) increases the agent's individual benefit in the long term. In other words, we aim at providing an utilitarian explanation of the internalization of the fairness norm à la Rawls, that is to say, we aim at explaining why rational agents with learning capabilities should become motivated to follow this fairness norm even without external enforcement (*e.g.*, external sanctions, punishment). Subsequently, we apply tools from *evolutionary game theory* (EGT) to study the conditions for which dispositions to follow the fairness norm à la Rawls can evolve in a population of agents lacking such dispositions at all. From an evolutionary perspective, an agent's degree of norm sensitivity should be conceived as the agent's morality level: the higher the agent's degree of sensitivity to the norm of fairness, the more the agent is disposed to sacrifice her well-being for the well-being of the others.

Norm internalization is a concept that has been widely discussed in the literature in social sciences [2, 3, 4]. As emphasized by Aronfreed [2, page 18], "...internalization occurs when a norm's maintenance has become independent of external outcomes, that is, to the extent that its reinforcing consequences are internally mediated, without the support of external events such as rewards or punishment". More recently, it has received interest from researchers in the field of multi-agent systems. For instance, Andrighetto et al. [5, 6] provide a simulation study of norm internalization that is able to account for experimental evidence showing that subjects playing PD or ultimatum games follow fairness consideration even without external enforcement [7]. They focus on the repeated PD with punishment played by a heterogenous population including three typologies of agents: strategic agents, normative agents and internalizers. Strategic agents are implemented as Q-learning agents. They are not aware about the existence of the norms and always choose the action that has provided them the maximum benefit in the past. Normative agents follow the norm of cooperation in order to avoid the punishment in case of violation. Finally, internalizer agents internalize the norm of cooperation when it has become *sufficiently salient* for them, where salience increases as a result of punishment.[1] Once the norm of cooperation is

[1]The authors distinguish two types of punishment both yielding a cost to the punisher: strategic and normative punishment. The

internalized, they follow it as an automatism, without any benefit-cost calculation. Before internalizing it, they behave as normative agents. Differently from our study, Andrighetto et al. do not provide any formal model or analytical study, as their approach is purely simulative. It is aimed at verifying hypotheses such as existence of a positive correlation between proportion of internalizers in the population (relative to strategic and normative agents) and cooperation rate, and between proportion of internalizers in the population and punishment rate. Moreover, differently from our study, Andrighetto et al. do not consider fairness norms and the conditions under which such norms emerge either through learning or through evolution.

Although the present work focuses on fairness norms à la Rawls and on the specific case of the repeated PD, it positions itself in the research area on the evolution of morality that has received increasing attention in the recent years [8, 9]. For example, Rand et al. [9] have recently provided an evolutionary explanation of why human agents are driven by fairness concerns in the one-shot anonymous ultimatum game. Their explanation is alternative to existing explanations based on the concept of reputation (see, e.g., [10, 11]) according to which fairness can be favored by natural selection if agents can recognize their partners' strategies or have reputations that carry from game to game. Rand et al.'s use stochastic evolutionary game theory to show that, if agents can make mistakes when judging the payoffs and strategies of others, then fairness can evolve.[2] The main difference between our approach and Rand et al.'s is that they model evolution of agents' *strategies* whereas we model evolution of agent's *degrees of sensitivity to fairness norm* or, more generally, evolution of agents' *strengths of morality*. This difference has important implications both at the conceptual level and at the technical level. In our evolutionary model every agent in the population is identified with a degree of sensitivity to the fairness norm and this is what passes down from generation to generation. In Rand et al.'s model every agent is identified with a strategy formalized by two parameters $p, q \in [0, 1]$, where p is the amount offered when acting as proposer in the ultimatum game and q is the minimum amount demanded when acting as responder. The idea of considering evolution of strengths of morality is something our model shares with some recent models of the evolution of morality in economics [12, 13].

Plan of the paper The rest of the paper is organized as follows. In Section 2 we present a game-theoretic model of guilt aversion which provides the static foundation of our analysis. The main idea of the model is that agents in a game are motivated both by their personal utilities and by the goal of avoiding guilt feeling. It is assumed that guilt feeling is triggered in case of the violation of an internalized norm. Specifically, the intensity of guilt feeling is proportional to the agent's sensitivity to the norm. In Section 3, we provide a dynamic extension of our model in order to formally specify repeated interactions and learning in a game-theoretic setting. The learning approach is the well-known fictitious play [14]. Section 4 provides some mathematical results about convergence for fictitious play in the case of iterated PD in which agents are assumed to be more or less sensitive to the fairness norm à la Rawls. Our mathematical analysis of convergence for fictitious play is partial, as it only covers a subset of the set of possible values of norm sensitivity for the agents in the population. Thus, in Section 5, we present some computational results about convergence for fictitious play which complement the analysis of Section 4. In Section 6, we present some experimental results in the case of iterated PD which highlight the relationship between

former only has a deterrent effect, while the latter has the additional effect of making explicit the norm existence and increasing its salience (it is accompanied with a deontic message).

[2] In their model of stochastic evolution, they introduce a parameter called the "intensity of selection". The idea is that the higher the value of this parameter, the more likely agents with higher payoffs are going to be reproduced (or imitated).

an agent's degree of sensitivity to the fairness norm à la Rawls and her individual benefit in the long term.

In Sections 7 and 8 we combine the learning approach with an evolutionary approach. We use tools from evolutionary game theory to analyze evolutionary aspects of sensitivity to the fairness norm à la Rawls in the Prisoner's Dilemma under population dynamics. The novel aspect of our analysis compared to existing analysis (cf. [15]) is that we focus on the evolution of different degrees of sensitivity to the fairness norm, while existing analysis focus on the evolution of pure or mixed strategies in the Prisoner's Dilemma, namely cooperation, defection, or any combination of the two. In other words, we are interested in identifying those degrees of fairness norm sensitivity that have a greater potential to survive (in EGT this potential can be formalized as *evolutionary stability*[3] [16]) under population dynamics, such as the well-known *replicator dynamics* [17]. To this aim, we first present in Section 7 an abstract concept of meta-game, which entails the accumulated reward of repeated interaction under fictitious play, and in which strategies correspond to the different degrees of sensitivity to the fairness norm an agent might have. Then, in Section 8 we present some computational results i) showing that cooperativeness and dispositions following the fairness norm à la Rawls can coevolve in a population of initially solely defectors under the replicator dynamics with mutation, and ii) indicating such a disposition to be evolutionary stable under the replicator dynamics. As a side note, we would like to remark that a preliminary version of this work has appeared [18], which does not contain any analysis of the evolutionary aspects of sensitivity to the fairness norm.

Motivating example The present work is mainly theoretical but, we believe, it can offer interesting insights for people working in the area of multi-agent systems (MAS) in which agents are supposed to be artificial entities such as a large team of robotic agents or a population of autonomous agents with learning capabilities, making decisions on the basis of expected utility maximization and having repeated encounters. One might wonder whether a fairness norm imposed by an external designer in order to promote cooperation will be followed by the agents even without external enforcement, as by following it an agent will get a benefit in the long run. By way of example, consider a future society with vehicles controlled by autonomous agents. These vehicles face different types of social dilemmas during their life cycles (see, e.g., [19], for a recent study of social dilemmas for autonomous vehicles). Some of these social dilemmas can be represented as repeated Prisoner's Dilemmas. For instance, suppose the owners of two autonomous vehicles, call them Ann and Bob, have to share a limited common space in the parking area in front of the hotel in which they both spend their summer holidays. Suppose also that in the late morning Bob regularly goes back to the hotel before Ann at around 12:00 am, while in the evening Ann regularly goes back to the hotel before Bob at around 6:30 pm. Each autonomous vehicle can decide to be either cooperative or non-cooperative. The cooperative option (C) consists in parking in the common space by taking care of leaving space available for the other vehicle, in case the other has not parked yet. The non-cooperative option (N) consists in parking in the common space without taking care of leaving space available for the other vehicle. The time cost of action N is significantly lower than the time cost of action C. Indeed, action C requires a careful maneuver by the vehicle and a considerable amount of time. In this sense, each vehicle has an incentive to play N. Nonetheless, if both vehicles choose option N, then this will be detrimental to both of them. Indeed, if both vehicles choose

[3] An evolutionary stable strategy (ESS) is a central concept in evolutionary game theory. An ESS has very insightful features, such as an invasion barrier against mutation/invasion as well as a close relation to the concept of a Nash equilibrium. For a formal definiton we refer to [16].

N, then Bob will be able to park his vehicle late in the morning but he will be unable to do it in the evening and, vice versa, Ann will be able to park her vehicle in the evening but she will be unable to do it late in the morning. Furthermore, the situation in which both vehicles choose option C is satisfactory for Bob and Ann, as they will be able to park their vehicles both late in the morning and in the evening. Finally, the situation in which one vehicle chooses C while the other chooses N is clearly the worst for the former and the best for the latter. In fact, by choosing C while the other chooses N, a vehicle will spend a considerable amount of time for properly parking but will not have space available for parking both late in the morning and in the evening. On the contrary, by choosing N while the other chooses C, a vehicle will spend a small amount of time for parking and will have space available for parking both late in the morning and in the evening.

This concrete scenario raises a number of general issues that the present paper aims to clarify. Suppose that Bob and Ann's autonomous vehicles are controlled by rational agents with learning capabilities. Under these assumptions, will the autonomous vehicles converge to the cooperative option in the long run after repeated encounters? If so, does this depend on the agents' sensitivities to the fairness norm? Note that the example can be seen as a finite repeated Prisoner's Dilemma game which lasts for the limited period of Bob and Ann's holidays. This excludes that the incentive to defect may evaporate because of the "shadow of the future" [30] and justifies the use of a fairness norm for promoting cooperation.

2. Game-theoretic model of guilt aversion

In this section, we present our game-theoretic model of guilt and of its influence on strategic decision making. We assume that guilt feeling originates from the agent's violation of a certain norm. Specifically, the intensity of an agent's guilt feeling depends on two parameters: (i) how much the agent is responsible for the violation of the norm, and (ii) how much the agent is sensitive to the norm. As emphasized in the introduction, in our model the agent's sensitivity to the norm captures the extent to which the norm is internalized by the agent.

Our model assumes that an agent has two different motivational systems: an endogenous motivational system determined by the agent's desires and an exogenous motivational system determined by the agent's internalized norms. Internalized norms make the agent capable of discerning what from his point of view is *good* (or *right*) from what is *bad* (or *wrong*). If an agent has internalized a certain norm, then she thinks that its realization ought to be promoted because it is *good* in itself. A similar distinction has also been made by philosophers and by social scientists. For instance, Searle [20] has recently proposed a theory of how an agent may want something without desiring it and on the problem of reasons for acting based on moral values and independent from desires. In his theory of morality [21], Harsanyi distinguishes a person's *ethical preferences* from her *personal preferences* and argues that a moral choice is a choice that is based on ethical preferences. A more recent account of ethical preferences in economics is given by Brekke et al. [22]. According to this study, a person can be intrinsically motivated to maintain a positive self-image of being morally responsible. Moral responsibility is defined as the difference between one's actual effort and the morally ideal effort, where the morally ideal effort is identified with the action that maximizes social welfare, given that everyone acts in the same way.[4]

[4] According to Brekke et al. [22] this corresponds to a simple version of Immanuel Kant's Categorical Imperative.

2.1. Normative game and guilt-dependent utility

Let us first introduce the standard notion of normal-form game.

Definition 2.1. (Normal-form game)
A normal-form game is a tuple $G = (N, (S_i)_{i \in N}, U)$ where:

- $N = \{1, \ldots, n\}$ is a finite set of agents or players;
- for every $i \in N$, S_i is agent i's finite set of strategies;
- $U : N \longrightarrow (\prod_{i \in N} S_i \longrightarrow \mathbb{R})$ is an utility function, with $U(i)$ being agent i's personal utility function mapping every strategy profile to a real number (*i.e.*, the personal utility of the strategy profile for agent i).

For every $i \in N$, elements of S_i are denoted by $s_i, s_i', \ldots$ Let $2^{Agt*} = 2^N \setminus \{\emptyset\}$ be the set of all non-empty sets of agents (*alias* coalitions). For notational convenience we write $-i$ instead of $N \setminus \{i\}$. For every $J \in 2^{Agt*}$, we define the set of strategies for the coalition J to be $S_J = \prod_{i \in J} S_i$. Elements of S_J are denoted by $s_J, s_J', \ldots$ We write S instead of S_N and we denote elements of S by $s, s', \ldots$ Every strategy s_J of coalition J can be seen as a tuple $(s_i)_{i \in J}$ where agent i chooses the individual strategy $s_i \in S_i$. For notational convenience we write $U_i(s)$ instead of $U(i)(s)$. As usual a mixed strategy for agent i is a probability distribution over S_i. Agent i's set of mixed strategies is denoted by Σ_i and elements of Σ_i are denoted by $\sigma_i, \sigma_i', \ldots$ The set of mixed strategy profiles is defined to be $\Sigma = \Sigma_1 \times \ldots \times \Sigma_n$ and its elements are denoted by $\sigma, \sigma', \ldots$ The utility function U_i reflects agent i's endogenous motivational system, *i.e.*, agent i's desires.

	C	D
C	R, R	S, T
D	T, S	P, P

Figure 1. Prisoner's dilemma (with player 1 being the row player and player 2 being the column player).

A well-known example of normal-form game is the Prisoner's Dilemma (PD) in which two agents face a social dilemma. The PD is represented in Figure 1.

Each agent in the game can decide either to cooperate (action C) or to defect (action D) and has an incentive to defect. Indeed, it is assumed that, if an agent defects, she gets a reward that is higher than the reward obtained in the case of cooperation, no matter what the other agent decides to do. In other words, cooperation is strongly dominated by defection. The social dilemma lies in the fact that mutual defection, the only Nash equilibrium of the game, ensures a payoff for each agent that is lower than the payoff obtained in the case of mutual cooperation. The Prisoner's Dilemma can be compactly represented as follows.

Definition 2.2. (Prisoner's Dilemma)
A Prisoner's Dilemma (PD) is a normal-form game $G = (N, (S_i)_{i \in N}, U)$ such that:

- $N = \{1, 2\}$;

- for all $i \in N$, $S_i = \{C, D\}$;
- $U_1(C,C) = R$, $U_1(D,D) = P$, $U_1(C,D) = S$ and $U_1(D,C) = T$;
- $U_2(C,C) = R$, $U_2(D,D) = P$, $U_2(C,D) = T$ and $U_2(D,C) = S$;

and which satisfies the following two conditions:

(**C1**) $\quad T > R > P > S$,

(**C2**) $\quad S = 0$.

Condition (**C1**) is the typical one in the definition of the Prisoner's Dilemma. Condition (**C2**) is an extra *normality* constraint which is not necessarily assumed in the definition of PD. It is assumed here to simplify the analysis of the evolution of fairness norms.

The following definition extends the definition of normal-form game with a *normative* component. Specifically, we assume that every outcome in a game is also evaluated with respect to its ideality degree, *i.e.*, how much an outcome in the game conforms to a certain norm. Moreover, as pointed above, we assume that an agent in the game can be more or less sensitive to the norm, depending on how much the norm is internalized by her.

Definition 2.3. (Normative game)
A normative game is a tuple $NG = (N, (S_i)_{i\in N}, U, I, \kappa)$ where:

- $(N, (S_i)_{i\in N}, U)$ is a normal-form game;
- $I : \prod_{i\in N} S_i \longrightarrow \mathbb{R}$ is a function mapping every strategy profile in S to a real number measuring the degree of ideality of the strategy profile;
- $\kappa : N \longrightarrow \mathbb{R}_{\geq 0}$ is a function mapping every agent in N to a non-negative real number measuring the agent's sensitivity to the norm.

For notational convenience we write κ_i instead of $\kappa(i)$ to denote agent i's sensitivity to the norm.

Following current psychological theories of guilt [23], we conceive guilt as the emotion which arises from an agent's self-attribution of responsibility for the violation of an internalized norm (*i.e.*, a norm to which the agent is sensitive). Specifically, intensity of guilt feeling is defined as *the difference between the ideality of the best alternative state that could have been achieved had the agent chosen a different action and the ideality of the current state, — capturing the agent's degree of responsibility for the violation of the norm —, weighted by the agent's sensitivity to the norm.* The general idea of our model is that the intensity of guilt feeling is a monotonically increasing function of the agent's degree of responsibility for norm violation and the agent's sensitivity to the norm.

Definition 2.4. (Guilt)
Let $NG = (N, (S_i)_{i\in N}, U, I, \kappa)$ be a normative game. Then, the guilt agent i will experience after the strategy profile s is played, denoted by $Guilt(i,s)$, is defined as follows:

$$Guilt(i,s) = \kappa_i \times (\max_{s'_i \in S_i} I(s'_i, s_{-i}) - I(s))$$

The following definition describes how an agent's utility function is transformed depending on the agent's feeling of guilt. In particular, the higher the intensity of guilt agent i will experience after the strategy profile s is played, the lower the (transformed) utility of the strategy profile s for agent i. Note indeed that the value $Guilt(i,s)$ is either positive or equal to 0. Guilt-dependent utility reflects both agent i's desires and agent i's moral considerations determined by her sensitivity to the norm.

Definition 2.5. (Guilt-dependent utility)
Let $NG = (N, (S_i)_{i\in N}, U, I, \kappa)$ be a normative game. Then, the guilt-dependent utility of the strategy profile s for agent i is defined as follows:

$$U^*_i(s) = U_i(s) - Guilt(i,s)$$

It is worth noting that the previous definition of guilt-dependent utility is similar to the definition of regret-dependent utility proposed in regret theory [24]. Specifically, similarly to Loomes & Sugden's regret theory, we assume that the utility of a certain outcome for an agent should be transformed by incorporating the emotion that the agent will experience if the outcome occurs.

2.2. Fairness norms

In the preceding definition of normative game an agent i's utility function U_i and ideality function I are taken as independent. There are different ways of linking the two notions.

For instance, Harsanyi's theory of morality provides support for an utilitarian interpretation of fairness norms which allows us to reduce an agent i's ideality function I to the utility functions of all agents [21]. Specifically, according to the Harsanyi's view, a fairness norm coincides with the goal of maximizing the collective utility represented by the weighted sum of the individual utilities.

Definition 2.6. (Normative game with fairness norm à la Harsanyi)
A normative game with fairness norm à la Harsanyi is a normative game $NG = (N, (S_i)_{i\in N}, U, I, \kappa)$ such that for all $s \in S$:

$$I(s) = \sum_{i\in N} U_i(s)$$

An alternative to Harsanyi's utilitarian view of fairness norms is Rawls' view [1]. In response to Harsanyi, Rawls proposed the *maximin* criterion of making the least happy agent as happy as possible: for all alternatives s and s', if the level of well-being in the worst-off position is strictly higher in s than in s', then s is better than s'. According to this well-known criterion of distributive justice, a fair society should be organized so as to admit economic inequalities to the extent that they are beneficial to the less advantaged agents. Following Rawls' interpretation, a fairness norm should coincide with the goal of maximizing the collective utility represented by the individual utility of the less advantaged agent.

Definition 2.7. (Normative game with fairness norm à la Rawls)
A normative game with fairness norm à la Rawls is a normative game $NG = (N, (S_i)_{i\in N}, U, I, \kappa)$ such that for all $s \in S$:

$$I(s) = \min_{i\in N} U_i(s)$$

In this paper we focus on fairness norm à la Rawls. In particular, we are interested in studying the relationship between the agents' sensitivities to this kind of norm and their behaviors in a repeated game such as the Prisoner's Dilemma in which the agents learn from their past experiences. To this aim, in the next section, we provide a dynamic extension of our model of guilt aversion.

3. Dynamic extension

In the dynamic version of our model, we assume that every agent in a given normative game has probabilistic expectations about the choices of the other agents. These expectations evolve over time. The following concept of history captures this idea.

Definition 3.1. (History)
Let $NG = (N, (S_i)_{i\in N}, U, I, \kappa)$ be a normative game. A *history* (for NG) is a tuple $H = ((\omega_{i,j})_{i,j\in N}, (c_i)_{i\in N})$ such that, for all $i, j \in N$:

- $\omega_{i,j} : \mathbb{N} \longrightarrow \Delta(S_j)$ is a function assigning to every time $t \in \mathbb{N}$ a probability distribution on S_j,
- $c_i : \mathbb{N} \longrightarrow S_i$ is a choice function specifying the choice of agent i at each time point $t \in \mathbb{N}$.

For every $t \in \mathbb{N}$ and $s_j \in S_j$, $\omega_{i,j}(t)(s_j)$ denotes agent i's subjective probability at time t about the fact that agent j will choose action $s_j \in S_j$. For notational convenience, we write $\omega^t_{i,j}(s_j)$ instead of $\omega_{i,j}(t)(s_j)$. For all $i, j \in N$, $t \in \mathbb{N}$ and $s_{-i} \in S_{-i}$ we moreover define:

$$\omega^t_i(s_{-i}) = \prod_{j\in N\setminus\{i\}} \omega^t_{i,j}(s_j)$$

$\omega^t_i(s_{-i})$ denotes agent i's subjective probability at time t about the fact that the other agents will choose the joint action s_{-i}.

The following definition introduces the concept of agent i's expected utility at time t. Notice that the concept of utility used in the definition is the one of guilt-dependent utility of Definition 2.5. Indeed, we assume a rational agent is an agent who maximizes her expected guilt-dependent utility reflecting both the agent's desires and the agent's moral considerations determined by her sensitivity to the norm.

Definition 3.2. (Expected utility at time t)
Let $NG = (N, (S_i)_{i\in N}, U, I, \kappa)$ be a normative game, let $H = ((\omega_{i,j})_{i,j\in N}, (c_i)_{i\in N})$ be a history for NG and let $t \in \mathbb{N}$. Then, the expected utility of action $s_i \in S_i$ for the agent i at time t, denoted by $EU_i^t(s_i)$, is defined as follows:

$$EU_i^t(s_i) = \sum_{s'_{-i}\in S_{-i}} \omega_i^t(s'_{-i}) \times U_i^*(s_i, s'_{-i})$$

As the following definition highlights, an agent is rational at a given time point t, if and only if her choice at time t maximizes expected utility.

Definition 3.3. (Rationality at time t)
Let $NG = (N, (S_i)_{i\in N}, U, I, \kappa)$ be a normative game, let $H = ((\omega_{i,j})_{i,j\in N}, (c_i)_{i\in N})$ be a history for NG and let $t \in \mathbb{N}$. Then, agent i is rational at time t if and only if $EU_i^t(c_i(t)) \geq EU_i^t(s_i)$ for all $s_i \in S_i$.

We assume that agents learn via fictitious play [14], a learning algorithm introduced in the area of game theory and widely used in the area of multi-agent systems (see, *e.g.*, [25]). The idea of fictitious play is that each agent best responds to the empirical frequency of play of her opponents. The assumption underlying fictitious play is that each agent believes that her opponents are playing stationary strategies that do not depend from external factors such as the other agents' last moves.

Definition 3.4. (Learning via fictitious play)
Let $NG = (N, (S_i)_{i\in N}, U, I, \kappa)$ be a normative game and let $H = ((\omega_{i,j})_{i,j\in N}, (c_i)_{i\in N})$ be a history for NG. Then, agent i learns according to fictitious play (FP) along H, if and only if for all $j \in N \setminus \{i\}$, for all $s_j \in S_j$ and for all $t > 0$ we have:

$$\omega_{i,j}^t(s_j) = \frac{obs_{i,j}^t(s_j)}{\sum_{s'_j\in S_j} obs_{i,j}^t(s'_j)}$$

where $obs_{i,j}^0(s_j) = 0$ and for all $t > 0$:

$$obs_{i,j}^t(s_j) \quad = \quad \begin{cases} obs_{i,j}^{t-1}(s_j) + 1 & \text{if } c_j(t-1) = s_j \\ obs_{i,j}^{t-1}(s_j) & \text{if } c_j(t-1) \neq s_j \end{cases}$$

Note that $obs_{i,j}^t(s_j)$ in the previous definition denotes the number of agent i's past observations at time t of agent j's strategy s_j.

Two notions of convergence for fictitious play are given in the literature, one for pure strategies and one for mixed strategies. Let $H = ((\omega_{i,j})_{i,j\in N}, (c_i)_{i\in N})$ be a history. Then, H converges in the pure strategy sense if and only if there exists a pure strategy $s \in S$ and $\bar{t} \in \mathbb{N}$ such that for all $i \in N$:

$$c_i(t) = s_i \text{ for all } t \geq \bar{t}$$

On the contrary, H converges in the mixed strategy sense if and only if there exists a mixed strategy $\sigma \in \Sigma$ such that for all $i \in N$ and for all $s_i \in S_i$:

$$\lim_{\bar{t}\to\infty} \frac{|\{t \leq \bar{t} : c_i(t) = s_i\}|}{\bar{t}+1} = \sigma_i(s_i)$$

Clearly, convergence in the pure strategy sense is a special case of convergence in the mixed strategy sense.

It has been proved [26] that for every *non-degenerate* 2×2 game (*i.e.*, two-player game where each player has two strategies available) and for every history H for this game, if all agents are rational and learn according to fictitious play along H, then H converges in the mixed strategy sense. The fact that the game is *non-degenerate* just means that, for every strategy of the second player there are no different strategies of the first player which guarantee the same payoff to the first player, and for every strategy of the first player there are no different strategies of the second player which guarantee the same payoff to the second player.[5] A generalization of this result to $2 \times n$ games has been given by [28].

4. Mathematical analysis in the PD with fairness norm à la Rawls

In this section, we provide convergence results for fictitious play in the case of iterated Prisoner's Dilemma in which players are more or less sensitive to the fairness norm à la Rawls.

The first thing we can observe is that for any possible combination of norm sensitivity values for the two players, the behaviors of both players will converge to mixed strategies. In particular:

Theorem 4.1. Let $NG = (N, (S_i)_{i \in N}, U, I, \kappa)$ be a normative game with fairness norm à la Rawls such that $(N, (S_i)_{i \in N}, U)$ is the Prisoner's Dilemma and let $H = ((\omega_{i,j})_{i,j \in N}, (c_i)_{i \in N})$ be a history for NG. Moreover, assume that every agent in N learns according to fictitious play along H and is rational for all $t \geq 0$. Then, H converges in the mixed strategy sense.

Proof:
For all possible values of κ_1 and κ_2, the transformed PD in which the utility function U_i is replaced by U_i^* for all $i \in \{1, 2\}$ is non-degenerate. The transformed PD is represented in Figure 2. Hence, the theorem follows from the fact that, as observed in the previous section, fictitious play is guaranteed to converge in the class of non-degenerate 2×2 games.

	C	D
C	R, R	$-\kappa_1 P, T - \kappa_2 R$
D	$T - \kappa_1 R, -\kappa_2 P$	P, P

Figure 2. Prisoner's Dilemma with transformed utilities according to fairness norm à la Rawls.

□

Our second result is the following theorem about convergence in the pure strategy sense. The theorem highlights that if at the beginning of the learning process every player has a uniform probability distribution over the strategies of the other player and the value of norm sensitivity is lower than the following threshold for cooperativeness

$$\theta_{tc} = \frac{P + T - R}{R - P}$$

[5] Miyazawa [27] assumed a particular tie-breaking rule to prove convergence of fictitious play in 2×2 games.

for both players, then the two players will always play mutual defection. On the contrary, if at the beginning of the learning process every player has a uniform probability distribution over the strategies of the other player and the value of norm sensitivity is higher than the threshold θ_{tc} for both players, then the two players will always play mutual cooperation.

Theorem 4.2. Let $NG = (N, (S_i)_{i\in N}, U, I, \kappa)$ be a normative game with fairness norm à la Rawls such that $(N, (S_i)_{i\in N}, U)$ is the Prisoner's Dilemma and let $H = ((\omega_{i,j})_{i,j\in N}, (c_i)_{i\in N})$ be a history for NG. Moreover, assume that every agent in N learns according to fictitious play along H and is rational for all $t \geq 0$, and that $\omega^0_{i,j}(s_j) = 0.5$ for all $i, j \in \{1, 2\}$ and for all $s_j \in \{C, D\}$. Then:

- if $\kappa_1 < \theta_{tc}$ and $\kappa_2 < \theta_{tc}$ then $c_1(t) = c_2(t) = D$ for all $t \geq 0$,
- if $\kappa_1 > \theta_{tc}$ and $\kappa_2 > \theta_{tc}$ then $c_1(t) = c_2(t) = C$ for all $t \geq 0$.

Proof:
Assume that every agent in N learns according to fictitious play along H and is rational for all $t \geq 0$, and that $\omega^0_{i,j}(s_j) = 0.5$ for all $i, j \in \{1, 2\}$ and for all $s_j \in \{C, D\}$. We are going to prove that, for all $i, j \in \{1, 2\}$, if $\kappa_i > \frac{P+T-R}{R-P}$ then $EU^0_i(C) > EU^0_i(D)$ and that if $\kappa_i < \frac{P+T-R}{R-P}$ then $EU^0_i(D) > EU^0_i(C)$.

First of all, let us compute the values of $EU^0_i(D)$ and $EU^0_i(C)$:

$$\begin{aligned} EU^0_i(D) &= 0.5 \times P + 0.5 \times (T - \kappa_i \times (R - S)) \\ &= 0.5 \times P + 0.5 \times (T - \kappa_i \times R) \\ &= 0.5 \times (P + T - \kappa_i \times R) \end{aligned}$$

$$\begin{aligned} EU^0_i(C) &= 0.5 \times R + 0.5 \times (S - \kappa_i \times (P - S)) \\ &= 0.5 \times R + 0.5 \times (-\kappa_i \times P) \\ &= 0.5 \times (R - \kappa_i \times P) \end{aligned}$$

It follows that $EU^0_i(D) > EU^0_i(C)$ if and only if $P + T - \kappa_i \times R > R - \kappa_i \times P$. The latter is equivalent to $\kappa_i < \frac{P+T-R}{R-P}$. Therefore, we have $EU^0_i(D) > EU^0_i(C)$ if and only if $\kappa_i < \frac{P+T-R}{R-P}$. By analogous argument, we can prove that $EU^0_i(C) > EU^0_i(D)$ if and only if $\kappa_i > \frac{P+T-R}{R-P}$.

It is routine task to verify that, for all possible values of κ_1 and κ_2 in the original normative game NG, the strategy profile (D, D) is a *strict Nash equilibrium* in the transformed PD depicted in Figure 2 in which the utility function U_i is replaced by U^*_i for all $i \in \{1, 2\}$. Hence, by Proposition 2.1 in [29] and the fact that every agent is rational for all $t \geq 0$, it follows that if $\kappa_1 < \frac{P+T-R}{R-P}$ and $\kappa_2 < \frac{P+T-R}{R-P}$ then $c_1(t) = c_2(t) = D$ for all $t \geq 0$.

It is also a routine to verify that, if $\kappa_i > \frac{T-R}{R-S}$ for all $i \in \{1, 2\}$, then the strategy profile (C, C) is a *strict Nash equilibrium* in the transformed PD depicted in Figure 2. Hence, by Proposition 2.1 in [29], the fact that every agent is rational for all $t \geq 0$ and the fact that $\frac{P+T-R}{R-P} > \frac{T-R}{R-S}$, it follows that if $\kappa_1 > \frac{P+T-R}{R-P}$ and $\kappa_2 > \frac{P+T-R}{R-P}$ then $c_1(t) = c_2(t) = C$ for all $t \geq 0$. □

5. Computational results in the PD with fairness norm à la Rawls

Theorem 4.2 shows that if both κ-values are smaller than the threshold for cooperativeness θ_{tc}, both players converge to mutual defection, whereas if both κ-values are greater than this threshold, both players converge to mutual cooperation. Note that this does not cover the whole space of tuples of κ-values, cf. how do agents operate, if one value is smaller and the other value is greater that θ_{tc}? In these terms we are faced with the more general question: for which combination of κ-values do agents converge to mutual cooperation or to mutual defection under fictitious play?

To examine the convergence behavior of players under fictitious play for different κ-values, we conducted multiple computations of repeated interactions, for different game parameters and a large subset of the κ^2-space. We recorded the results and we managed to deduce the conditions determining the convergence behavior that pertain perfectly with the data. These conditions are as follows.

For all normative games with fairness norm à la Rawls $NG = (N, (S_i)_{i\in N}, U, I, \kappa)$ and history $H = ((\omega_{i,j})_{i,j\in N}, (c_i)_{i\in N})$ for NG such that $(N, (S_i)_{i\in N}, U)$ is the Prisoner's Dilemma, every agent in N learns according to fictitious play along H, is rational for all $t \geq 0$, and $\omega^0_{i,j}(s_j) = 0.5$ for all $i, j \in \{1, 2\}$ and for all $s_j \in \{C, D\}$, the following three facts are observed:

1. if $(\kappa_1 - lim_{mx}) \times (\kappa_2 - lim_{mx}) < curv_{mx}$ then $\exists t' \in \mathbb{N} : c_1(t) = c_2(t) = D$ for all $t \geq t'$,
2. if $(\kappa_1 - lim_{mx}) \times (\kappa_2 - lim_{mx}) > curv_{mx}$ then $\exists t' \in \mathbb{N} : c_1(t) = c_2(t) = C$ for all $t \geq t'$,
3. if $(\kappa_1 - lim_{mx}) \times (\kappa_2 - lim_{mx}) = curv_{mx}$ then both players converge to a mixed strategy,

whereby:

$$curv_{mx} = \left(\frac{PT}{(R+P)(R-P)}\right)^2$$

$$lim_{mx} = \frac{P^2 + R(T-R)}{(R+P)(R-P)}$$

Note that the equation $(\kappa_1 - lim_{mx}) \times (\kappa_2 - lim_{mx}) = curv_{mx}$ defines a separating curve between the convergence to mutual cooperation and mutual defection: for at least one of both κ-values being less than given, the first condition holds and fictitious play converges to mutual defection, whereas for at least one of both κ-values being greater than given, the second condition holds and fictitious play converges to mutual cooperation. For each pair of κ-values that fulfills the equation, the third condition holds and fictitious play converges to a mixed strategy for each player. This curve can be defined as a function for the convergence to a mixed strategy f_{mx} over κ_1-values[6]:

$$f_{mx}(\kappa_1) = \frac{curv_{mx}}{\kappa_1 - lim_{mx}} + lim_{mx}$$

The function f_{mx} is depicted in Figure 3.

A necessary condition of function f_{mx} to be correct is that it has an intersection point for $\kappa_1 = \kappa_2 = \theta_{tc}$, as proved in Theorem 5.1. An implication of function f_{mx} to be correct is the fact that the

[6]Note that the function forms an *anallagmatic* curve, cf. it inverts into itself.

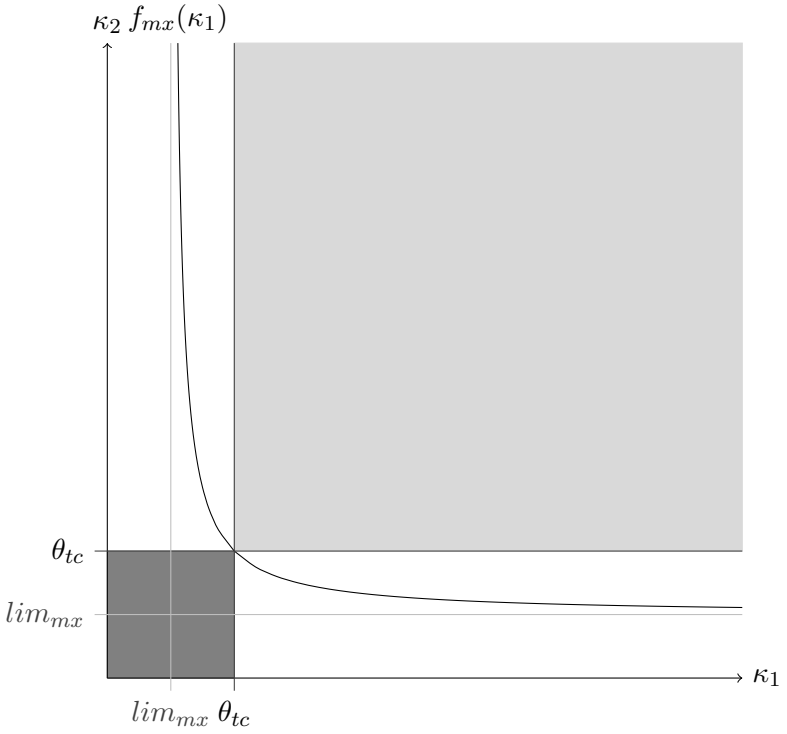


Figure 3. The dark gray/light gray area shows in accordance with Theorem 4.2 that if both κ-values are smaller than the threshold for cooperativeness $\theta_{tc} = \frac{P+T-R}{R-P}$, both players behave according to mutual defection (dark gray area), whereas if both κ-values are greater than this threshold, both players behave according to mutual cooperation (light gray area). The curve represents the function for non-convergence f_{mx} and defines for which combination of κ-values players converge to mutual cooperation (right of/above the curve), converge to mutual defection (left of/below the curve) or converge to a combination of mixed strategies (points of the curve). Note that i) lim_{mx} is the asymptote of the function f_{mx}, thus it defines a lower bound for mutual cooperation to emerge, and ii) $\kappa_1 = \kappa_2 = \theta_{tc}$ is an intersection point of the curve.

value lim_{mx} is the asymptote of the function f_{mx}, as proved in Theorem 5.2, and therefore determines a lower bound for κ-values that enable the convergence to mutual cooperation. Finally, note that the value $curv_{mx}$ determines the curvature of the function. Since lim_{mx} and $curv_{mx}$ both depend on the parameters of the PD game, the asymptote and curvature of a function f_{mx} can strongly differ among different games. Figure 4 shows the different curves of function f_{mx} for different game parameters.

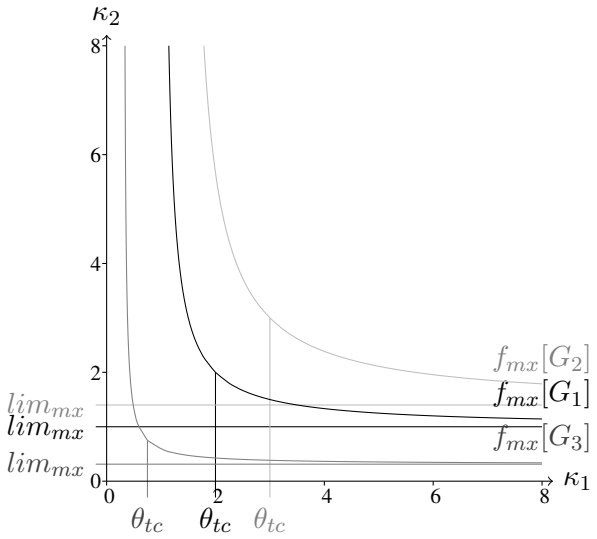
Theorem 5.1. $\kappa_1 = \kappa_2 = \theta_{tc}$ is an intersection point of function f_{mx}.

Proof:
We are going to show that $f_{mx}(\theta_{tc}) = \theta_{tc}$:

$$
\begin{aligned}
f_{mx}(\theta_{tc}) = & \frac{curv_{mx}}{\theta_{tc} - lim_{mx}} + lim_{mx} \\
= & \frac{\left(\frac{PT}{(R+P)(R-P)}\right)^2}{\frac{P+T-R}{R-P} - \frac{P^2+R(T-R)}{(R+P)(R-P)}} + lim_{mx} \\
= & \frac{\left(\frac{PT}{(R+P)(R-P)}\right)^2}{\frac{(R+P)(P+T-R)}{(R+P)(R-P)} - \frac{P^2+R(T-R)}{(R+P)(R-P)}} + lim_{mx} \\
= & \frac{\left(\frac{PT}{(R+P)(R-P)}\right)^2}{\frac{(R+P)(P+T-R)-(P^2+R(T-R))}{(R+P)(R-P)}} + lim_{mx} \\
= & \left(\frac{PT}{(R+P)(R-P)}\right)^2 \times \frac{(R+P)(R-P)}{(R+P)(P+T-R)-(P^2+R(T-R))} + lim_{mx} \\
= & \left(\frac{PT}{(R+P)(R-P)}\right)^2 \times \frac{(R+P)(R-P)}{PT} + lim_{mx}
\end{aligned}
$$

$$
\begin{aligned}
= & \frac{PT}{(R+P)(R-P)} + lim_{mx} \\
= & \frac{PT}{(R+P)(R-P)} + \frac{P^2+R(T-R)}{(R+P)(R-P)} \\
= & \frac{PT+P^2+R(T-R)}{(R+P)(R-P)} \\
= & \frac{PT+P^2+RT-R^2}{(R+P)(R-P)} \\
= & \frac{(P+T-R)(R+P)}{(R+P)(R-P)} \\
= & \frac{P+T-R}{R-P} = \theta_{tc}
\end{aligned}
$$

□



Game	G_1	G_2	G_3
T	3	4	7
R	2	3	6
P	1	2	2
θ_{tc}	2	3	.75
lim_{mx}	1	1.4	.125
$curv_{mx}$	1	2.56	$\approx .2$

Figure 4. Exemplary Prisoner's Dilemma games G_1 ($T = 3, R = 2, P = 1$), G_2 ($T = 4, R = 3, P = 2$) and G_3 ($T = 7, R = 6, P = 2$) and their corresponding values θ_{tc}, lim_{mx} and $curv_{mx}$ (right table). The graph shows the corresponding curves of the function f_{mx} for each game. Note that the value $curv_{mx}$ behaves anti-proportional to the curvature of the function.

Theorem 5.2. lim_{mx} is the asymptote of function f_{mx}.

Proof:
We are going to show that $\lim_{\kappa \to +\infty} f_{mx}(\kappa) = lim_{mx}$:

$$\begin{aligned}
\lim_{\kappa \to +\infty} f_{mx}(\kappa) &= \lim_{\kappa \to +\infty} \left(\frac{curv_{mx}}{\kappa - lim_{mx}} + lim_{mx} \right) \\
&= \lim_{\kappa \to +\infty} \left(\frac{curv_{mx}}{\kappa - lim_{mx}} \right) + lim_{mx} \\
&= lim_{mx}
\end{aligned}$$

□

6. Tournaments and experimental results

Let's assume we have a mixed population in terms of sensitivity to fairness norm κ. There might be individuals with high κ-values, with low κ-values or with no sensitivity to that norm at all. In such a setup it is reasonable to ask how beneficial fairness norm sensitivity might be. Is a low, middle or high sensitivity rather detrimental or profitable - especially in comparison with the outcome of the other individuals of the population? To get a general idea of how beneficial a particular degree of sensitivity to the fairness norm might be, we tested the performance of agents with different κ-values in a tournament. Such a tournament was inspired by Axelrod's tournament of the repeated Prisoner's Dilemma [30]. In Axelrod's tournament a number of agents play the repeated Prisoner's Dilemma - each agent against every other agent - for a particular number of repetitions. Each agent updates her behavior according to

a rule defined by her creator. The score of each encounter is recorded, and the agent with the highest average utility over all encounters wins the tournament.

In our tournament we also define n agents $0, 1, 2, ..., n-1$, where each agent plays against each other agent for a number of repetitions t_{max}. Differently from Axelrod's tournament, all agents i) play the Prisoner's Dilemma as a normative game with fairness norm à la Rawls, and ii) have the same update rule, namely, fictitious play. Although the agents have the same update rule, they differ in another crucial aspect, namely, their sensitivity to the fairness norm. To keep things simple, we predefine that their sensitivity i) is bounded above by a value $\kappa_{max} \in \mathbb{R}_{>0}$, and ii) is equally distributed among the n agents, by ascribing sensitivity to the fairness norm $\kappa_i = \frac{i \times \kappa_{max}}{n-1}$ to agent i.[7] A tournament works as follows: for each pair of agents i, j we conducted a normative game with fairness norm à la Rawls based on the Prisoner's Dilemma for a number of t_{max} repetitions, whereby agents i and j learn according to fictitious play along their common history.[8] For each agent i her average utility TU_i - called *tournament utility* - is computed, which is the average utility value an agent scored over all interactions.

For a given set of agents A that participate in such a tournament, the winner is the agent $i \in A$ who obtains the maximal tournament utility TU_i. In case that a tournament has multiple agents with maximal tournament utility, the winner is the one of those who has minimum sensitivity to the fairness norm, or in other words, the minimal index i.

We refer to the winner's κ_i value as the optimal fairness norm sensitivity κ^*, with respect to her tournament utility. In formal terms, we have that $\kappa^* = \kappa_i$ if and only if $i \in \arg\max_{z \in A} TU_z$ and $\kappa_i < \kappa_j$ for all $j \in \arg\max_{z \in A} TU_n$ such that $i \neq j$.

We computed 4 tournaments, each with 200 agents playing a normative game NG with fairness norm à la Rawls based on a Prisoner's Dilemma with $T = 3$, $R = 2$, $P = 1$ and $S = 0$. For such a game θ_{tc} is 2, and to ensure an equal portion of cooperative and non-cooperative agents, we set $\kappa_{max} = 2 \times \theta_{tc} = 4$. The tournaments differed in the parameter for t_{max}, here we chose the values 6, 20, 50 and 200. Figure 5 shows the performance of each agent in the appropriate tournament and the appended table shows the κ^* value of each tournament's winner. The results of the tournaments indicate that the optimal sensitivity to the fairness norm is by any means dependent on t_{max} and θ_{tc}. To verify this indication we computed a great number of further tournaments with different values $t_{max} \in \{2k | k \in \mathbb{N}\}$ and κ_{max} values.[9] The results support – without any exception – the following two observations:

1. For two tournaments that differ solely in the parameters t_{max} and t'_{max}, whereby the tournaments' optimal values of sensitivity to the fairness norm are κ^* and κ'^*, respectively, it holds that:

$$\text{if } t_{max} > t'_{max} \text{ then } \kappa^* \geq \kappa'^*$$

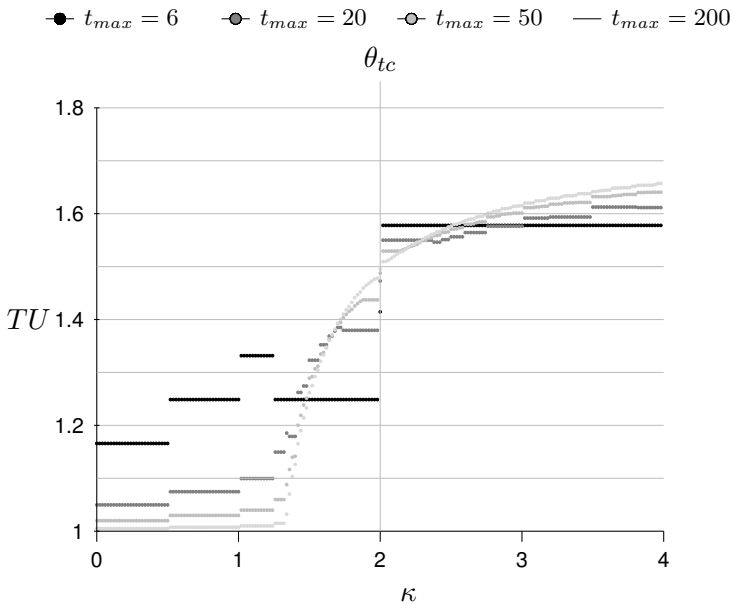
2. For every tournament it holds that:

$$\kappa^* > \theta_{tc}$$

[7] Note that to ascribe a value of fairness norm sensitivity $\kappa = \frac{i \times \kappa_{max}}{n-1}$ to agent i ensures that agent 0 has a sensitivity to the fairness norm of 0, agent $n-1$ has one of κ_{max}, and all other agents' sensitivity to the fairness norm are equally distributed between these boundaries.

[8] In case an agent has the situation where playing C or D both have the same expected utility, she uses a *tie-breaking rule* (see Section 7), here the rule is to play D.

[9] We restricted t_{max} to be an even number, since odd numbers of interactions can produce little bumps in the t_{max}-κ^* relationship due to alternating history behavior.



t_{max}	κ^*
6	2.02
20	3.5
50	3.92
200	3.96

Figure 5. The resulting tournament utilities of four different tournaments with 200 agents each, pairwise playing a normative game with fairness norm à la Rawls based on a Prisoners Dilemma game with $T = 3$, $R = 2$, $P = 1$ and $S = 0$. The right table shows for each t_{max} parameter the appropriate optimal sensitivity to the fairness norm κ^* of the tournament's winner.

The first observation unveils one condition for which a high sensitivity to the fairness norm à la Rawls is beneficial. It tells us that κ^* is monotonically increasing with respect to t_{max}, i.e. the value of the optimal sensitivity to the fairness norm increases with the number of repetitions of a repeated game in such a tournament. This result is in line with former insights, since i) we showed in Section 4 that a high value of fairness norm sensitivity supports cooperative behavior, and ii) we know from studies of repeated Prisoner's Dilemma that cooperative behavior is especially beneficial in combination with reputation [10], a virtue that needs repetition to be established. However, as we will delineate later, repetition can also be beneficial for cooperative behavior without attributing reputation to single agents. The second observation says that it is optimal to have a sensitivity to the fairness norm that ensures *preliminary cooperativeness*.[10] This stresses the fact that a great fairness norm sensitivity is only beneficial if it not only enables a line of mutual cooperation, but it also implies an initial disposition to start it.

7. Normative meta-games

Before we analyze normative games with fairness norms à la Rawls under evolutionary aspects, we want to convert the game to another game model that has two particular features: i) it indirectly entails the average payoffs for a particular number of repeated plays, and ii) its strategy set entails different levels

[10]As we have shown in Theorem 4.2, if agents have a sensitivity to fairness $\kappa > \theta_{tc}$, their first move is to cooperate. This behavioral characteristic can be seen as preliminary cooperativeness.

of sensitivity to the fairness norm. The basic idea is to convert a normative game NG into a normative *meta-game*. The average utility of such a number of encounters, called the *history utility*, is defined as follows. In order to keep the formal framework relatively simple, in this section, we only consider two-player games. Furthermore, we assume that every player i is associated with a linear ordering $\prec_i$ over the strategies in her strategy set. This linear ordering is used to define a tie-breaking rule that selects a strategy, in case two or more strategies have the same expected utility.

Definition 7.1. (History utility)
Let $NG = (\{1,2\}, (S_i)_{i\in\{1,2\}}, U, I, \kappa)$ be a two-player normative game and let $t \in \mathbb{N}$. Then, for every $i \in \{1,2\}$, player i's history utility $HU_i : \mathbb{N} \longrightarrow \mathbb{R}$ for NG and for a number of rounds t is defined as follows:

$$HU_i(t) = \sum_{0\leq t'<t} \frac{U_i(c_1(t'), c_2(t'))}{t}$$

where c_1 and c_2 are the choice functions of the unique history $H = ((\omega_{i,j})_{i,j\in\{1,2\}}, (c_i)_{i\in\{1,2\}})$ for NG such that:

- players 1 and 2 learn according to fictitious play and are rational along H,
- $\omega^0_{i,j}(s_j) = \frac{1}{|S_j|}$ for all $i, j \in \{1,2\}$ and for all $s_j \in S_j$,
- for all $t \in \mathbb{N}$ and for all $i \in \{1,2\}$, if $X = \arg\max_{s_i\in S_i} EU^t_i(s_i)$ and $|X| > 1$ then $c_i(t) = tie{-}break_i(X)$,

and $tie{-}break_i(X)$ is the maximal element in X according to the linear ordering $\prec_i$.

Player i's history utility for the normative game NG and for the number of rounds t is the average utility that player i will obtain after having played the game for t rounds under the following three conditions: (i) every player is rational and learns according to fictitious play during the game, (ii) every player has a uniform probability distribution over the strategies of the other at the beginning of the game, and (iii) every player uses her tie-breaking rule in order to select a strategy, in case two or more strategies have the same expected utility. Notice that conditions (i), (ii) and (iii) together guarantee the uniqueness of the history H.

Given the definition of history utility, the concept of normative meta-game is defined as follows.

Definition 7.2. (Normative meta-game)
Let $G = (\{1,2\}, (S_i)_{i\in N}, U)$ be a two-player normal-form game, let I be an ideality function over $\prod_{i\in N} S_i$, let $t_{max} \in \mathbb{N}_{\geq 1}$ be the number of rounds and let $K \subset \mathbb{R}_{\geq 0}$ be a finite set of values of norm sensivity. Then, the normative meta-game NMG induced by G, I, t_{max} and K is the two-player normal-form game $(\{1,2\}, (S'_i)_{i\in N}, U')$ such that:

- for every $i \in \{1,2\}$, $S'_i = K$;
- for every $x, y \in K$ and for every $i \in \{1,2\}$:

$$U'_i(x,y) = HU^{x,y}_i(t_{max})$$

	0	1	2
0	$1,1$	$1,1$	$3,0$
1	$1,1$	$1,1$	$3,0$
2	$0,3$	$0,3$	$2,2$

	0	1	2
0	$1,1$	$1,1$	$1.5,0.75$
1	$1,1$	$1,1$	$1.5,0.75$
2	$0.75,1.5$	$0.75,1.5$	$2,2$

Figure 6. Example of normative meta-games for the Prisoner's dilemma with payoffs $R = 2$, $S = 0$, $T = 3$ and $P = 1$, with fairness norm à la Rawls (with player 1 being the row player and player 2 being the column player), with set of norm sensitivity values $K = \{0, 1, 2\}$, and with numbers of rounds $t_{max} = 1$ (left figure) and $t_{max} = 4$ (right figure). It is assumed that the two players use the same tie-breaking rule which consists in playing C when playing C and D both have the same expected utility.

where $HU_i^{x,y}(t_{max})$ is player i's history utility for the number of rounds t_{max} and for the normative game $NG = (\{1, 2\}, (S_i)_{i\in N}, U, I, \kappa)$ such that $\kappa_1 = x$ and $\kappa_2 = y$.

In other words, the normative meta-game induced by the two-player normal-form game $G = (\{1, 2\}, (S_i)_{i\in N}, U)$, the ideality function I, the number of rounds t_{max} and the finite set K of norm sensitivity values is the two-player normal-form game such that: (i) the strategies of the players are the possible values of norm sensitivity in K, (ii) a strategy profile is a pair (x, y), where x is the value of norm sensitivity for player 1 and y is the value of norm sensitivity for player 2, and (iii) the utility of the strategy profile (x, y) for player $i \in \{1, 2\}$ is player i's history utility for the number of rounds t_{max} rounds and for the normative game $NG = (\{1, 2\}, (S_i)_{i\in N}, U, I, \kappa)$ in which player 1's value of norm sensitivity is x and player 2's value of norm sensitivity is y.

Let us consider the normal-form game $G = (\{1, 2\}, (S_i)_{i\in N}, U)$ corresponding to the Prisoner's Dilemma with payoffs $R = 2$, $S = 0$, $T = 3$ and $P = 1$. Moreover, let I be the ideality function corresponding to the fairness norm à la Rawls. Figure 6 represents the two normative meta-games induced by G and I, under the assumption that the set of norm sensitivity values is $K = \{0, 1, 2\}$ and that the number of rounds is $t_{max} = 1$ (left figure), or $t_{max} = 4$ (right figure), respectively. We assume that players 1 and 2 use the same tie-breaking rule which consists in playing C when playing C and D both have the same expected utility.

8. Normative meta-games under population dynamics

Note that in evolutionary game theory (EGT), i) the interpretation of a strategy is a specific feature or disposition, or – in sense of cultural evolution – a specific opinion, behavior or attitude, and ii) the interpretation of payoff is the fitness for reproduction. In this sense an EGT analysis of normative meta-games depicts the evolutionary dynamics (of a population) of repeatedly interacting and learning individuals with different attitudes or dispositions in terms of fairness norm sensitivity. Note that by combining learning dynamics (in form of fictitious play entailed as final utility values in the meta-game) and evolutionary dynamics in one model, we incorporate two essential forces of human evolution: ontogeny (individual life-time learning) and crossgenerational transmission (biological or cultural[11]). This dif-

[11]Note that evolutionary dynamics, such as the replicator dynamics, might represent biological or cultural evolution, dependent on the features that are transmitted. We abstract from the concrete type here and solely say that evolutionary dynamics represent the crossgenerational transmission of individual features.

ferentiation is important, since while the ontogenetic process (repeated fictitious play) represents how and why agents display individual features (fairness sensitivity, cooperativeness), the process of biological/cultural transmission factors in stability aspects of the whole populations that represent configurations of such features.[12]

To understand the evolutionary properties of normative meta-games, we tested them under a computational non-deterministic variant of the well-known *replicator dynamics* [17] that addresses unfaithful replication: *mutation* (see replicator-mutator equation [36]). We call this variant *replicator-mutator-similarity dynamics* (RMS dynamics), which considers a particular aspect of normative meta-games, namely that strategies are real numbers and therefore can be compared by *similarity*. The similarity between two strategies $x, x' \in S$ of a normative meta-game can be simply defined by a function $sim : S^2 \longrightarrow \mathbb{R}$, whereby $sim(x, x') = \frac{x_{max} - |x - x'|}{x_{max}}$ with $x_{max} = \max(S)$. Given a similarity function, the intuition of the RMS dynamics is as follows: we think that the evolutionary process is more realistic if mutation depends on the similarity between strategies. E.g. a parent with strategy $x = 0.6$ has more likely a more similar mutant offspring, such as $x = 0.5$ or $x = 0.7$, than a less similar one, such as $x = 2.1$. With this intuition we implemented a simple variant of the RMS dynamics, informally described by the following RMS algorithm:

Definition 8.1. (Simple RMS algorithm)
The simple replicator-mutator-similarity (RMS) algorithm is given by the following computational steps:

1. *Input*: population size $n \in \mathbb{N}_{>1}$, mutation rate $\epsilon \in \mathbb{R}$, normative meta-game $(\{1, 2\}, S, U)$;
2. *Initialize population*: create agent set $A = \{a_1, a_2, \ldots a_n\}$, whereby each agent is assigned with a strategy $x \in S$ by function $\alpha : A \longrightarrow S$. For notational convenience we write α_i instead of $\alpha(a_i)$;
3. *Repeat until breaking condition ϕ:*
   (a) *Compute agent's total utility:* Compute every agent's global utility value for pairwise interaction with any other agent by the function $GU : A \longrightarrow \mathbb{R}$, where:
   $$GU(a_i) = \sum_{a_j \in A, a_j \neq a_i} U(\alpha_i, \alpha_j)$$
   (b) *Compute strategy fitness:* Compute every strategy's fitness value by the function $F : S \longrightarrow \mathbb{R}$, where:
   $$F(x) = \sum_{a_i \in A, \alpha_i = x} GU(a_i)$$
   (c) *Compute strategy offspring:* Compute every strategy's number of offspring by the function $O : S \longrightarrow \mathbb{N}$, where:
   $$O(x) = \left\lfloor \frac{F(x) \times n}{\sum_{x' \in S} F(x')} \right\rfloor$$

[12]The differentiation between individual learning and crossgenerational evolution is essential, since both processes play different roles in human evolution and also in this model. This point is especially important concerning the fact that processes of learning dynamics and evolution dynamics show often very similar characteristics. For example, it has been shown for a range of different game classes that individual imitation/learning models converges to the replicator dynamics, such as shown for conditional imitation [31, 32], reinforcement learning [33, 34], or Q-learning [35].

(d) *Compute new population:* (i) empty set A, (ii) for each strategy $x \in S$: add $O(x)$ players of strategy x to A, (iii) while $|A| < n$: add another player with random strategy $x \in S$;[13]

(e) *Mutation:* For each $a_i \in A$ compute a random value $\phi \in [0.0, 1.0)$:

- If $\phi \leq \frac{\epsilon}{2}$:
  change α_i to $\alpha_i' \in S$, where $\alpha_i' < \alpha_i$ and $\forall x \in S : sim(\alpha_i', \alpha_i) < sim(x, \alpha_i)$,[14]
- If $\frac{\epsilon}{2} < \phi \leq \epsilon$:
  change α_i to $\alpha_i' \in S$, where $\alpha_i' > \alpha_i$ and $\forall x \in S : sim(\alpha_i', \alpha_i) < sim(x, \alpha_i)$.[15]

Note that steps 3(a-d) of the simple RMS algorithm approaches a non-deterministic version of the replicator dynamics [17] for a finite population.[16] Furthermore, step 3(e) realizes similarity mutation in its simplest way: with probability $\frac{\epsilon}{2}$ a strategy is replaced with the most similar strategy with a smaller value, and with probability $\frac{\epsilon}{2}$ a strategy is replaced with the most similar strategy with a greater value.

Let us consider the normal-form game $G = (\{1, 2\}, (S_i)_{i \in N}, U)$ corresponding to the Prisoner's Dilemma with payoffs $R = 2$, $S = 0$, $T = 3$ and $P = 1$. Moreover, let I be the ideality function corresponding to the fairness norm à la Rawls. Figure 6 represents the two normative meta-games induced by G and I, under the assumption that the set of norm sensitivity values is $K = \{0, 1, 2\}$ and that the number of rounds is $t_{max} = 1$ (left figure), or $t_{max} = 4$ (right figure), respectively. We assume that players 1 and 2 use the same tie-breaking rule which consists in playing C when playing C and D both have the same expected utility.

We tested the simple RMS algorithm for the following parameters and conditions:

- *population size*: $n = 100$
- *mutation rate*: $\epsilon = 0.1$
- *input game*: the two-player normative meta-game induced by:
  - the normal-form game $G = (\{1, 2\}, (S_i)_{i \in N}, U)$ corresponding to the Prisoner's Dilemma with payoffs $R = 2$, $S = 0$, $T = 3$ and $P = 1$,
  - the ideality function I corresponding to the fairness norm à la Rawls,
  - the maximal number of interactions $t_{max} = 10$,
  - the set of values of norm sensitivity $K = \{0.0, 0.25, 0.5, 0.75, 1.0, \ldots, 2.75, 3.0\}$,
  - whereby the utilities of the meta-game are computed with respect to a tie-breaking rule, which consists of playing C, when playing C and D both have the same expected utility.
- *initialization*: $\forall a_i \in A : \alpha_i = 0.0$
- *breaking condition*: number simulation steps > 10000

The result of an exemplary simulation run for the first 200 simulation steps is shown in Figure 7.

[13] It is routine to verify that $\sum_{x \in S} O(x) \leq n$. In case that $\sum_{x \in S} O(x) < n$ the new population would have less than n agents. In that case new agents with a random strategy are added to A until $|A| = n$.

[14] If $\alpha_i = \min(S)$, i.o.w. there is no strategy $x \in S$ with $x < \alpha_i$, then α_i won't be changed.

[15] If $\alpha_i = \max(S)$, i.o.w. there is no strategy $x \in S$ with $x > \alpha_i$, then α_i won't be changed.

[16] The algorithm is non-deterministic, since in step (d) additional players with a random strategy might be added.

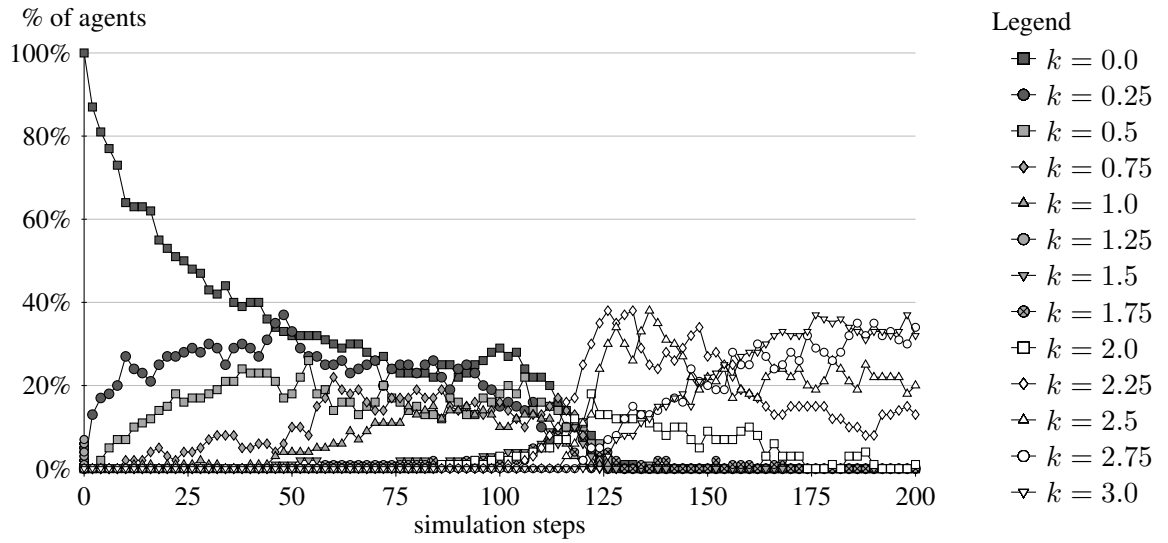


Figure 7. Result of an exemplary simulation run with the simple RMS algorithm for the normative meta-game NMG_1. Note that strategies of unconditional defectors ($k < \theta_{ud} = 0.5$) are colored black, strategies of initial defectors ($\theta_{ud} \leq k < \theta_{tc} = 2.0$) are colored gray, and strategies of initial cooperators ($k \geq \theta_{tc}$) are colored white.

Note first of all that the input game includes a threshold for cooperativeness $\theta_{tc} = 2.0$. This fact is represented in Figure 7 by marking all strategies with $k < \theta_{tc}$ as gray/black, and all strategies with $k \geq \theta_{tc}$ as white. Furthermore, the initialization of the simulation run produces an initial situation with all agents having a norm sensitivity strategy of $k = 0.0$. This exemplary run reveals that a population of unconditional defectors with no sensitivity to the fairness norm can evolve to a population of cooperators with a fairness norm sensitivity above θ_{tc}.[17] Thus, the evolutionary analysis of normative meta-games reveals a possible path for the coevolution of cooperation and fairness norm sensitivity.

As a further step we were interested in extracting conditions that are supportive to such an evolutionary path ρ. Our subsequent analysis is guided by the following intuition: the more supportive a particular condition to a specific evolutionary path ρ, the shorter the expected time that is needed for path ρ to be passed through. Since we apply an experimental approach, we rate expected time by the average number of simulation steps over multiple simulation runs. Note that in this way we can make comparative statements, such as 'for setting A the conditions are more supportive for the evolutionary path ρ than for setting B, since ρ has a shorter average runtime to be passed through.

With this approach we want to measure the influence of t_{max} on the average runtime of an evolutionary path leading from a population of solely defectors with $k = 0.0$ to a population of solely cooperators with $k \geq \theta_{tc}$. We used the RMS algorithm (Definition 8.1) with the following experimental settings:

- *population size*: $n = 50$
- *mutation rate*: $\epsilon = 0.1$
- *initialization*: $\forall a_i \in A : \alpha_i = 0.0$

[17] As shown with Theorem 4.2, every two agents with a fairness norm sensitivity above θ_{tc} do always cooperate.

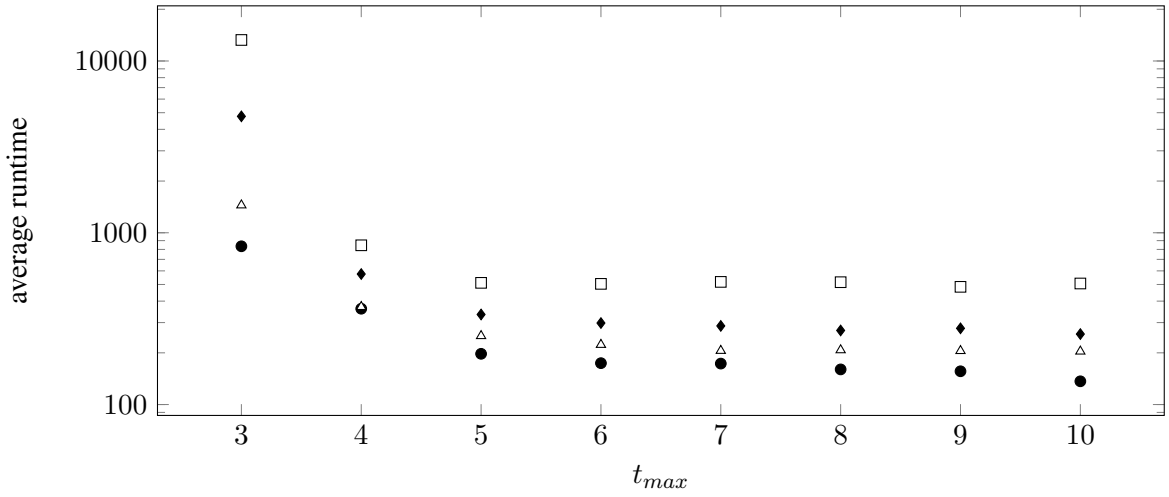


Figure 8. Average runtime until a population of cooperators has emerged over 100 simulation runs per experiment. Experiments for various normative meta-games induced by different t_{max}-parameters, varying from 3 to 10. Black cycles represent the results of the experimental setting as described in the text. In three further lines of experiments we made the same experiments by changing one of three parameters, either i) lowering ϵ to 0.05 (white triangles), ii) lowering n to 10 (black diamonds), or iii) changing to a more fine-grained segmentation of the strategy continuum K to $\{0.0, 0.15, 0.3, ..., 3.0\}$ (white squares). In comparison to the original setting, each of these changes was detrimental to a low expected runtime, whereby increasing t_{max} was supportive for a low expected runtime in each setting.

- *breaking condition*: $\forall a_i \in A : \alpha_i \geq \theta_{tc}$
- *input games*: different two-player normative meta-games, each induced by
  - the normal-form game $G = (\{1, 2\}, (S_i)_{i \in N}, U)$ corresponding to the Prisoner's Dilemma with payoffs $R = 2$, $S = 0$, $T = 3$ and $P = 1$,
  - the ideality function I corresponding to the fairness norm à la Rawls,
  - the set of values of norm sensitivity $K = \{0.0, 0.25, 0.5, 0.75, 1.0, \ldots, 2.75, 3.0\}$,
  - whereby the utilities of the meta-game are computed with respect to a tie-breaking rule, which consists of playing C, when playing C and D both have the same expected utility,

  with each game induced by a different maximal number of interactions t_{max}

With this setting we made 8 experiments, each with a different t_{max}-value from 3 to 10. For each experiment we conducted 100 runs and computed the average runtime. The results are depicted in Figure 8 (black circles). Note that the average runtime decreases with increasing t_{max}, e.g. for $t_{max} = 3$ the average runtime is more than 800, whereas for $t_{max} > 5$ the average runtime is around 200.

All in all, the results show that a great t_{max} value is supportive for a low expected runtime of the evolutionary path. This indicates the long-term benefits of sensitivity to fairness norm under evolutionary

dynamics: the shift from a defecting to a cooperating population is quickly established by agents who have enough time to develop a path of mutual cooperation, thereby signaling a high sensitivity to the fairness norm; note that this happens with agents not being able to distinguish between their interactors, a necessary condition for establishing reputation.[18] In other words, a high number of repetition is supportive for the emergence of fairness and cooperation without the establishment of reputation. The key mechanism for the transition here is *direct reciprocity*: cooperators help each other directly.

Finally, note that there are three further free parameters: the population size n, the mutation rate ϵ, and the definition of the set of strategies K. We made also a number of simulation runs by varying these parameters. The results are also exemplified in Figure 8: lowering ϵ (here to 0.05, white triangles) as well as a lowering n (here to 10, black diamonds), both are detrimental for a low expected runtime. Furthermore, also a more fine-grained segmentation of the strategy continuum K (here to $K = \{0.0, 0.15, 0.3, ..., 3.0\}$, white squares) is detrimental to a low expected runtime. All these simulation runs also reveal that the more supportive a parameter is for a low expected runtime, the more resistant is a population of cooperators with a fairness norm sensitivity above θ_{tc} against defectors with a fairness norm sensitivity below θ_{tc}. This resistance against mutation/invasion can be formally proven by studying the *evolutionary stability* of strategies with a fairness norm sensitivity above θ_{tc} for different parameter settings. Such a formal analysis is currently under development and will be part of a subsequent study.

9. Conclusion

Our study presents a game-theoretic model of guilt in relation to sensitivity to the norm of fairness à la Rawls. We employed this model – the *normative game* – on the Prisoner's Dilemma, and worked out the convergence behavior under fictitious play for any combination of the fairness norm sensitivity of both players. We found out that a particular threshold for cooperation θ_{tc} plays a crucial role: it defines for which combinations both agents cooperate or defect from the beginning, and for which combinations they might learn to cooperate or to defect.

In a second step we introduced *normative meta-games*: a static description of iterated normative games under fictitious play. With tools from evolutionary game theory we analyzed evolutionary aspects of normative meta-games by applying the *Simple RMS algorithm*. We found that there exists an evolutionary path from a population of only defectors with no fairness norm sensitivity to a population of only cooperators with a fairness norm sensitivity above θ_{tc}. Our computational results suggest that the probability of this evolutionary path to emerge is supported by a great number of repeated interactions t_{max}. All in all, our analysis reveals an *evolutionary path of the coevolution of cooperativeness and fairness norm sensitivity*. These results are in line with existing naturalistic theories of fairness according to which sensitivity to fairness norm might be the product of evolution (see, *e.g.*, [37]).

In the future, we plan to extend our learning-based and evolutionary analysis to a more general class of games including the public goods game, that can be seen as a multi-agent variant of the PD, and the ultimatum game. This generalization will allow us to compare our approach to existing literature on the evolution of fairness in the ultimatum game [9].

In our model of repeated PD, social interaction is unstructured: every agent is allowed to play with any other agent in the population. It has been shown that the structure of interaction, modeled as a social

[18] Note that reputation is a value that one would attribute to a specific individual: I help you, if you've helped others before (indirect reciprocity). Without the ability to distinguish between others, it is not possible to elaborate the concept of reputation.

network, can be relevant for the evolution of cooperation in the context of the repeated PD [15] and for the emergence of cooperation in the ultimatum game played by a population of learning agents [8]. For instance, in the model proposed by De Jong et al. [8], learning agents play a repeated continuous ultimatum game in a scale-free interaction network. Agents play together based on their connections in the network and may decide to break existing connections and to create new ones in a random way. Agents are motivated to disconnect themselves from (relative) defectors, as they prefer to play with relative cooperators. The authors show that the agents' capacity to rewire their connections in the network greatly enhances their ability to reach agreement in the ultimatum game. Following de Jong et al., in future work, we plan to extend both our learning model and our evolutionary model by a social network component in order to investigate whether rewiring plays a role (i) in increasing the long-term benefit of norm sensitivity and (ii) in promoting the selection of a relatively high degree of norm sensitivity.

References

[1] Rawls J. A theory of justice. Harvard University Press, Cambridge, 1971. ISBN-0674017722, 9780674017726.

[2] Aronfreed JM. Conduct and Conscience: The Socialization of Internalized Control Over Behavior. Academic Press,, New York, 1968. ISBN-10:1483210553, 13:978-1483210551.

[3] Gintis H. The hitchhiker's guide to altruism: Gene-culture co- evolution, and the internalization of norms. Journal of Theoretical Biology, 2003;220(4):407–418. https://doi.org/10.1006/jtbi.2003.3104.

[4] Gintis H. The genetic side of gene-culture coevolution: internalization of norms and prosocial emotions. Journal of Economic Behavior and Organization, 2004;53:57–67. http://www.sciencedirect.com/science/article/pii/S0167-2681(03)00104-5.

[5] Andrighetto G, Villatoro D, Conte R. Norm internalization in artificial societies. AI Communications. 2010;23(4):325–339. http://dl.acm.org/citation.cfm?id=1898063.1898066.

[6] Andrighetto G, Villatoro D, Conte R. The role of norm internalizers in mixed populations. In: R. Conte, G. Andrighetto, and M. Campenni, editors, Minding Norms: Mechanisms and dynamics of social order in agent societies, pages 153–174. Oxford University Press, Oxford, 2013. doi:10.1093/acprof:oso/9780199812677.003.0010

[7] Bicchieri C. The Grammar of Society: The Nature and Dynamics of Social Norms. Cambridge University Press, New York, 2006. ISBN-10:0521574900, 13:978-0521574907.

[8] De Jong S, Uyttendaele S, Tuyls K. Learning to reach agreement in a continuous ultimatum game. Journal of Artificial Intelligence Research, 2008;33:551. doi:10.1613/jair.2685.

[9] Rand DG, Tarnita CE, Ohtsuki H, Nowak, MA. Evolution of fairness in the one-shot anonymous ultimatum game. Proceedings of the National Academy of Sciences, 2013;110(7):2581–2586. doi:10.1073/pnas.1214167110.

[10] Nowak M, Sigmund K. Evolution of indirect reciprocity. Nature, 2005;437:1291–1298. doi:10.1038/ nature0413.

[11] Nowak M, Sigmund K. Evolution of indirect reciprocity by image scoring. Nature, 1998;393:573–577. doi:10.1038/31225.

[12] Alger I, Weibull J. Homo Moralis: Preference Evolution Under Incomplete Information and Assortative Matching. Econometrica, 2013;81(6):2269–2302. doi:10.3982/ECTA10637.

[13] Alger I, Weibull J. Evolution leads to Kantian morality. Games and Economic Behavior, 2016;98:56–67. https://doi.org/10.1016/j.geb.2016.05.006.

[14] Brown GW. Iterative solution of games by fictitious play. In T. C. Koopmans, editor, Activity Analysis of Production and Allocation,, pages 374–376. John Wiley, New York, 1951. ISBN-10:0300015399, 13:9780300015393.

[15] Alexander JM. The structural evolution of morality. Cambridge University Press, 2008. http://www.cambridge.org.

[16] Maynard Smith J. Evolution and the Theory of Games. Cambridge University Press; 1982. ISBN-10: 0521288843, 13:978-0521288842.

[17] Taylor P, Jonker L. Evolutionarily Stable Strategies and Game Dynamics. Mathematical Biosciences, 1978;40:145–156. https://doi.org/10.1016/0025-5564(78)90077-9.

[18] Lorini E, Mühlenbernd R. The long-term benefits of following fairness norms: A game-theoretic analysis. In Q. Chen, P. Torroni, S. Villata, J. Hsu & A. Omicini, edsitors, PRIMA 2015: Principles and Practice of Multi-agent Systems, volume 9387 of Lecture Notes in Artificial Intelligence. Springer. 2015, pp. 301–318. https://doi.org/10.1007/978-3-319-25524-8_19.

[19] Bonnefon J-F, Shariff A, Rahwan I. The social dilemma of autonomous vehicles. Science, 2016; 352(6293):1573–1576. doi:10.1126/science.aaf2654.

[20] Searle J. Rationality in Action. MIT Press, Cambridge, 2001. ISBN-9780262194631.

[21] Harsanyi J. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. Journal of Political Economy, 1955;63(4):309–321. http://www.jstor.org/stable/1827128.

[22] Brekke KA, Kverndokk S, Nyborg K. An economic model of moral motivation. Journal of Public Economics, 2003;87(9-10):1967–1983. https://doi.org/10.1016/S0047-2727(01)00222-5.

[23] Haidt J. The moral emotions. In R. J. Davidson, K. R. Scherer, and H. H. Goldsmith, editors, Handbook of Affective Sciences. Oxford University Press. 2003, pp. 852—870. ISBN-10:0195126017, 13:978-0195126013.

[24] Loomes G, Sugden R. Regret theory: An alternative theory of rational choice under uncertainty. Economic Journal, 1982;92(368):805–824. doi:10.2307/2232669.

[25] Vidal JM. Learning in multiagent systems: An introduction from a game-theoretic perspective. In E. Alonso, D. Kudenko, and D. Kazakov, editors, Adaptive Agents and Multi-agent Systems, volume 2636 of Lecture Notes in Computer Science. Springer-Verlag, Berlin. 2003, pp. 202–215. https://doi.org/10.1007/3-540-44826-8_13.

[26] Monderer D, Shapley LS. Fictitious play property for games with identical interests. Journal of Economic Theory, 1996;68(1):258–265. https://doi.org/10.1006/jeth.1996.0014.

[27] Miyazawa K. On the convergence of the learning process in a 2x2 non-zero-sum two-person game. Princeton University Econometric Research Program, 3, 1961. https://books.google.de/books?id=jNj8GgAACAAJ.

[28] Berger U. Fictitious play in 2xn games. Journal of Economic Theory, 2005;120:139–154. https://EconPapers.repec.org/RePEc:wpa:wuwpga:0303009.

[29] Fudenberg D, Levine DK. The theory of learning in games. MIT Press, Cambridge, 1998. ISBN-10:0262529246, 13:978-0262529242.

[30] Axelrod A. The Evolution of Cooperation. Basic books, 1984. ISBN-10:0465005640, 13:978-0465005642.

[31] Helbing D. A stochastic behavioral model and a 'microscopic' foundation of evolutionary game theory. Theory and Decision, 40(2):149–179, 1996. https://doi.org/10.1007/BF00133171.

[32] Schlag KH. Why imitaten and if so, how? Journal of Economic Theory, 1998;78(1):130–156. https://doi.org/10.1006/jeth.1997.2347.

[33] Börgers T, Sarin R. Learning through reinforcement and replicator dynamics. Journal of Economic Theory, 1997;77:1–14. https://doi.org/10.1006/jeth.1997.2319.

[34] Tuyls K, Nowé A. Evolutionary game theory and multi-agent reinforcement learning. The Knowledge Engineering Review, 2005;20(1):63–90. doi:10.1017/S026988890500041X.

[35] Panait L, Tuyls K, Luke S. Theoretical advantages of lenient learners: an evolutionary game theoretic perspective. Journal of Machine Learning Research, 2008;9:423–457. http://dl.acm.org/citation.cfm?id=1390681.1390694.

[36] Hofbauer J. The Selection Mutation Equation. Journal of Mathematical Biology, 1985;23:41–53. https://doi.org/10.1007/BF00276557.

[37] Binmore K. Natural justice. Oxford University Press, New York, 2005. doi:10.1093/acprof:oso/9780195178111.001.0001.