

Uchida, Satoshi; Yamamoto, Hitoshi; Okada, Isamu; Sasaki, Tatsuya

## Article

# Evolution of cooperation with peer punishment under prospect theory

Games

## Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

*Suggested Citation:* Uchida, Satoshi; Yamamoto, Hitoshi; Okada, Isamu; Sasaki, Tatsuya (2019) : Evolution of cooperation with peer punishment under prospect theory, Games, ISSN 2073-4336, MDPI, Basel, Vol. 10, Iss. 1, pp. 1-18, <https://doi.org/10.3390/g10010011>

This Version is available at:

<https://hdl.handle.net/10419/219234>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

Article

# Evolution of Cooperation with Peer Punishment under Prospect Theory

Satoshi Uchida <sup>1,\*</sup>, Hitoshi Yamamoto <sup>2</sup>, Isamu Okada <sup>3</sup> and Tatsuya Sasaki <sup>4,5</sup>

<sup>1</sup> Research Center for Etho-Culture Studies, RINRI Institute, Tokyo 102-0094, Japan

<sup>2</sup> Faculty of Business Administration, Rissho University, Tokyo 141-8602, Japan; hitoshi@ris.ac.jp

<sup>3</sup> Faculty of Business Administration, Soka University, Tokyo 192-8577, Japan; okada@soka.ac.jp

<sup>4</sup> F-Power Inc., Tokyo 106-6119, Japan; tatsuyak66@gmail.com

<sup>5</sup> Faculty of Mathematics, University of Vienna, 1090 Vienna, Austria

\* Correspondence: s-uchida@rinri-jpn.or.jp

Received: 4 December 2018; Accepted: 15 February 2019; Published: 21 February 2019



**Abstract:** Social dilemmas are among the most puzzling issues in the biological and social sciences. Extensive theoretical efforts have been made in various realms such as economics, biology, mathematics, and even physics to figure out solution mechanisms to the dilemma in recent decades. Although punishment is thought to be a key mechanism, evolutionary game theory has revealed that the simplest form of punishment called peer punishment is useless to solve the dilemma, since peer punishment itself is costly. In the literature, more complex types of punishment, such as pool punishment or institutional punishment, have been exploited as effective mechanisms. So far, mechanisms that enable peer punishment to function as a solution to the social dilemma remain unclear. In this paper, we propose a theoretical way for peer punishment to work as a solution mechanism for the dilemma by incorporating prospect theory into evolutionary game theory. Prospect theory models human beings as agents that estimate small probabilities and loss of profit as greater than they actually are; thus, those agents feel that punishments are more frequent and harsher than they really are. We show that this kind of cognitive distortion makes players decide to cooperate to avoid being punished and that the cooperative state achieved by this mechanism is globally stable as well as evolutionarily stable in a wide range of parameter values.

**Keywords:** evolution of cooperation; social dilemma; punishment; evolutionary games; prospect theory; nonlinear utility

**PACS:** 02.30.Hq; 02.70.-c; 05.45.-a; 87.15.Aa; 87.23.-n; 87.23.Kg

**JEL Classification:** C72; C73; D91

## 1. Introduction

Although cooperative relationships can be found in diverse systems ranging from microbiological communities to global economic spheres, cooperation frequently poses a scientific puzzle. Cooperation is clearly important to make biological and human societies effective and smooth, and evolutionary biologists and social scientists have long puzzled over the origin of cooperation. In recent decades, extensive theoretical efforts from various disciplines, such as economics, biology, mathematics, or even physics, have been made to figure out solution mechanisms to the cooperation dilemma [1–7].

In the literature, the cooperation puzzle is often called the social dilemma or the free rider problem. This can be described as follows: (1) individuals in a society have binary choices: cooperation (contributing to the community) or defection (refusing to contribute, i.e., free-riding), (2) a society

consisting of cooperators is more profitable than that with only defectors, (3) but within a society, individual defectors do better than individual cooperators, since cooperators must incur a cost for contribution, while defectors do not.

Thus, the defective strategy dominates the cooperative strategy and is the Nash equilibrium in the framework of game theory. In other words, it is more advantageous for individuals to choose defection regardless of what other individuals choose, which, by natural selection or social learning, leads to a society with only defectors. The theoretical issue therefore is to explain why cooperation is so ubiquitous in the real world.

In the literature on the evolution of cooperation, various mechanisms have been proposed and investigated. Representative examples include kin selection, direct reciprocity, indirect reciprocity, group selection, and spatial and network reciprocity [8]. Besides these well-studied mechanisms, punishment is thought to be one of the effective mechanisms and is extensively investigated in the literature [9–14]. If individuals have a possibility of being punished when they do not pay costs for cooperation and if the strength of the punishment is so harsh that it cancels the profit obtained by free-riding, the players are expected to choose cooperation.

The most elementary type of punishment is informal or peer punishment, typified by the “As you wronged me, I will punish you” attitude. However, the evolution of this type of punishment remains a theoretical challenging puzzle [15–17]: a simple theory on the evolution of peer punishment predicts that no one chooses to punish to avoid bearing the implementation cost of the punishment. This new type of dilemma caused by the introduction of punishment is called the second-order dilemma. In fact, punishment itself incurs a cost on punishers; consequently, those who perform the punishment earn less payoff than those who do not punish. Punishment would be effective if there were those who choose to punish in the population. In this sense, punishers can be interpreted as a different type of contributor than cooperators. However, rational individuals will never choose to punish others. Thus, the second-order dilemma occurs.

Peer punishment itself is interpreted as a cooperative act; thus, it could be used for the avoidance of the second-order dilemma. One way is to introduce the possibility of punishing those who do not punish defectors—the second-order punishment. However, this again raises the third-order dilemma that no one chooses to implement the second-order punishment since it is costly. In general, an introduction of the  $n$ -th order punishment raises the  $(n + 1)$ -th order dilemma, which results in an infinite regress of peer punishment [18–20]. The theoretical task is therefore to search for ways or mechanisms to cut the infinite chain.

In order to prevent the emergence of higher order dilemmas (which mean the  $n$ -th order dilemmas for  $n > 1$ ), additional and mostly complex mechanisms have been proposed in the literature on evolutionary game theory [21–31]. For instance, if players are asked whether to contribute to public goods that will be used for punishment in the future, the second-order dilemma can be avoided; thus, cooperation evolves. This type of punishment is called formal or pool punishment.

In this present paper, we show a way for peer punishment to work without additional mechanisms. To do this, we take cognitive distortions of human beings into account. Most game-theory studies on the evolution of cooperation are based on the linear expected utility theory, which asserts that individuals’ decisions rely on accurate calculations of expected payoffs. On the other hand, experimental researches in behavioral economics and experimental economics have revealed that real human beings are not so rational that they can objectively estimate probabilities and values without biases, both of which are elements of expected payoffs [32–37]. Thus, payoffs attributed to real human beings are subjective, and this fact plays an essential role in actual decision-making processes.

One of the most successful theories describing irrationality is prospect theory [38–40]. In this theory, distortions of probabilities and values are modeled as a weighted function and a value function, respectively. The nonlinearities and asymmetric properties of these functions represent the irrationality of human beings in prospect theory. What happens if we analyze peer punishment, not assuming the

linear expected utility theory but based on prospect theory to estimate expected payoffs? This is the main issue considered in this paper.

Thus, this paper gives the first opportunity to study the coevolution of cooperation and peer punishment in the framework of evolutionary game theory combined with prospect theory. Our main finding is that enlarging the effect of small probabilities makes it possible to avoid the second-order dilemma, which enables cooperation to evolve and to be sustained.

In the next section, we describe the game, strategies, and the model setting of this paper. Then, we derive the results and discuss them.

## 2. Materials and Methods

### 2.1. Game and Strategies

An infinitely large, well-mixed population of individuals (or players) is considered. From time to time, two players are selected at random from the population and made to engage in a “donation game” [5]: each player decides whether to support the opponent at a personal cost,  $c$ . If a player chooses to support the other, the opponent receives a benefit  $b > c$ ; otherwise, the player obtains nothing. Each individual in the population will experience such decision-making many times. From here on, we denote the action “support” by “C” and “refuse” by “D”. Table 1 shows the payoffs player A obtains when playing the donation game with player B.

Table 1. Payoffs player A obtains in donation game.

Player B’s Options Player A’s Options	Cooperate (C)	Defect (D)
Cooperate (C)	$b - c$	$-c$
Defect (D)	$b$	$0$

After they have played the donation game, both players consider whether to punish their opponents or not if the other player chose D in the donation game. If a player chooses to punish its opponent, the other player’s payoff is reduced by  $s$ . The punishment is not free but costly and therefore incurs a cost  $r$  on the punisher. We assume the strength of punishment  $s$  is greater than the punishment cost  $r$ . We denote the decisions “punish” and “not punish” by “P” and “N”, respectively.

Thus, the game considered in this paper consists of two phases, which we call “donation phase” and “punishment phase”. We call the combined game “donation–punishment game”. In the donation phase, individuals consider whether to cooperate (support the other) or not, and in the punishment phase, they consider whether to punish their opponents when they choose D in the donation phase.

As a result, individuals have 4 options in total, i.e., there are 4 types of strategies: cooperate punish (CP), cooperate not-punish (CN), defect punish (DP), and defect not-punish (DN). Each individual follows one of these strategies and makes decisions according to the strategy. We denote the payoff matrix of the donation–punishment game described in Table 2 by  $M$ .

Table 2. Payoffs player A obtains in donation-punishment game.

Player B’s Options Player A’s Options	Cooperate Punish (CP)	Cooperate Not-Punish (CN)	Defect Punish (DP)	Defect Not-Punish (DN)
Cooperate Punish (CP)	$b - c$	$b - c$	$-c - r$	$-c - r$
Cooperate Not-punish (CN)	$b - c$	$b - c$	$-c$	$-c$
Defect Punish (DP)	$b - s$	$b$	$-s - r$	$-r$
Defect Not-punish (DN)	$b - s$	$b$	$-s$	$0$

We assume that individuals sometimes deviate from their strategies due to error. By a small probability, those individuals who intend to cooperate do not cooperate and vice versa. In the same way, individuals who intend to punish may not punish and vice versa. Note that we assume symmetric deviations in double meanings to reduce parameters. Firstly, probabilities of deviations from C (P) to D (N) and from D (N) to C (P) are the same. Secondly, probabilities of deviations from C to D and from P to N are the same, which implies that error between C and D and error between P and N are independent. All probabilities of deviations are denoted by one parameter,  $\varepsilon$ .

## 2.2. Payoff and Strategy Switching

### 2.2.1. Linear Expected Utility Theory

The long-term payoff of an individual depends on its strategy and other individuals' strategies. For the sake of convenience, we denote the strategies by numbers: we call CP strategy 1, CN strategy 2, DP strategy 3, and DN strategy 4. Then  $x_i$  denotes the frequency of the  $i$ -th strategy in the population. These frequencies affect the expected payoffs that individuals obtain.

As mentioned above, we assume that individuals sometimes commit errors. This means that the strategy of an individual and its decisions may be different since actual actions can deviate from its strategies due to errors. We denote a strategy and an action actually chosen in a game by the same label. For instance,  $i = 1$  either means strategy CP or action CP, depending on the context.

Clearly, errors also affect the expected payoffs. Since an error occurs in decisions both for the donation and for the punishment, an individual with strategy  $i$  may actually choose any other action. In order to calculate the expected payoffs, considering these, we introduce error vectors. A player with strategy 1 actually chooses its actions according to the following error vector: ( $\bar{\varepsilon} = 1 - \varepsilon$  is the probability that an error does not occur.)

$$\vec{E}_1 = (\bar{\varepsilon}^2, \bar{\varepsilon}\varepsilon, \varepsilon\bar{\varepsilon}, \varepsilon^2)^T. \tag{1}$$

Here, the first element of the vector (the square of  $\bar{\varepsilon}$ ) gives the probability that the individual chooses action 1 (CP), because a player with strategy 1 actually chooses action C (cooperation) with probability  $\bar{\varepsilon}$  and P (punishment) with the same probability  $\bar{\varepsilon}$ . The second, the third, and the fourth elements are defined in the same way. Thus,  $\vec{E}_1$  provides the probability distribution of actions chosen in a game of individuals with strategy 1.

Similarly, the vectors  $\vec{E}_2 = (\bar{\varepsilon}\varepsilon, \bar{\varepsilon}^2, \varepsilon^2, \varepsilon\bar{\varepsilon})^T$ ,  $\vec{E}_3 = (\varepsilon\bar{\varepsilon}, \varepsilon^2, \bar{\varepsilon}^2, \bar{\varepsilon}\varepsilon)^T$ ,  $\vec{E}_4 = (\varepsilon^2, \varepsilon\bar{\varepsilon}, \bar{\varepsilon}\varepsilon, \bar{\varepsilon}^2)^T$  characterize the probability distributions on the action space of individuals with their corresponding strategies.

With these error vectors, the expected payoff of an individual with strategy  $i$  when playing a game with an individual with strategy  $j$  is given by

$$P_{ij} = \vec{E}_i^T M \vec{E}_j. \tag{2}$$

Since the probability that an individual with strategy  $i$  encounters an individual with strategy  $j$  is  $x_j$  by the definition, the expected payoff of an individual with strategy  $i$  is given by

$$P_i = \sum_{j=1}^4 P_{ij}x_j. \tag{3}$$

We obtain the same expected payoff with a different approach. This approach will be used to derive distorted expected payoffs in the framework of prospect theory. The vector

$$\vec{A} = \sum_{j=1}^4 \vec{E}_j x_j, \tag{4}$$

represents the probability distribution of actions chosen by an arbitrary player selected at random from the population. Now we define matrix  $B_i$  by

$$B_i = \overset{\rightarrow}{E}_i \overset{\rightarrow T}{A} . \tag{5}$$

The element  $(B_i)_{mn}$  of the matrix is interpreted as the probability that a player with strategy  $i$  obtains a payoff of  $(M)_{mn}$  of the payoff matrix of the donation–punishment game. Thus, matrix  $B_i$  is a probability distribution over payoff matrix  $M$  in the eyes of individuals with strategy  $i$ . The expected payoff is then expressed as a sum of values (payoffs) multiplied by probabilities.

$$P_i = \sum_{m=1}^4 \sum_{n=1}^4 (M)_{mn} (B_i)_{mn} . \tag{6}$$

Players adaptively switch their strategies, aiming at more expected payoffs, which results in gradual changes of strategy frequencies. We assume that the time evolution of the frequency of strategy  $i$  is governed by the replicator equation [41]:

$$\dot{x}_i = x_i \left( P_i - \sum_{j=1}^4 P_j x_j \right) . \tag{7}$$

### 2.2.2. Prospect Theory

The element  $(B_i)_{mn}$  of the matrix  $B_i$  represents the probability that the outcome  $(m, n)$  of the game is realized. Following prospect theory, we assume that this probability is subjectively calculated in a distorted way. The subjective probability is given by applying a nonlinear function called weighted function to the objective probability  $(B_i)_{mn}$  :

$$(W_i)_{mn} = w((B_i)_{mn}) , \tag{8}$$

with

$$w(x) = \frac{x^\gamma}{[x^\gamma + (1 - x)^\gamma]^{\frac{1}{\gamma}}} , \tag{9}$$

where  $x$  is an objectively given probability.

The function contains a parameter  $\gamma$ . If  $\gamma = 1$ , the function is linear, and this case corresponds to linear expected utility theory. The smaller  $\gamma$  is, the more distorted the subjective probability is. If  $\gamma$  is too small ( $\gamma < 0.28$ ), the function is not monotonically increasing anymore. Therefore  $\gamma$  must be equal to or greater than 0.28 theoretically. In the literature of prospect theory, it is reported that values of  $\gamma$  around 0.65 best fit experimental results [38–40]. We set  $\gamma$  to 0.65 in the following analysis. The shape of the function with this parameter value is shown in the left panel of Figure 1. According to this function, small objective probabilities are estimated to be greater than they are because of  $w(x) > x$  for small  $x$ . Analyses with other values of  $\gamma$  can be found in the supporting material.

Thus, the matrix  $W_i$  is interpreted as a subjective probability distribution on  $M$  in the eyes of individuals with strategy  $i$ . Note that the probability distribution is not normalized; thus, the sum of subjective probabilities is not one in general because there is no experimental evidence that people are so rational that they normalize probabilities. In fact, the weighted function given by Equation (9) is asymmetric with respect to  $x = 1/2$ , which implies that  $w(x) + w(1 - x)$  is not equal to one (see the left panel of Figure 1).

In the same way as in the case of probabilities, the payoff matrix  $M$  is also distorted by a nonlinear function

$$(V)_{mn} = v((M)_{mn}) . \tag{10}$$

The function is called the value function, which is given by

$$v(x) = \begin{cases} x^\alpha & (x \geq 0) \\ -\lambda(-x)^\alpha & (x \leq 0) \end{cases}, \tag{11}$$

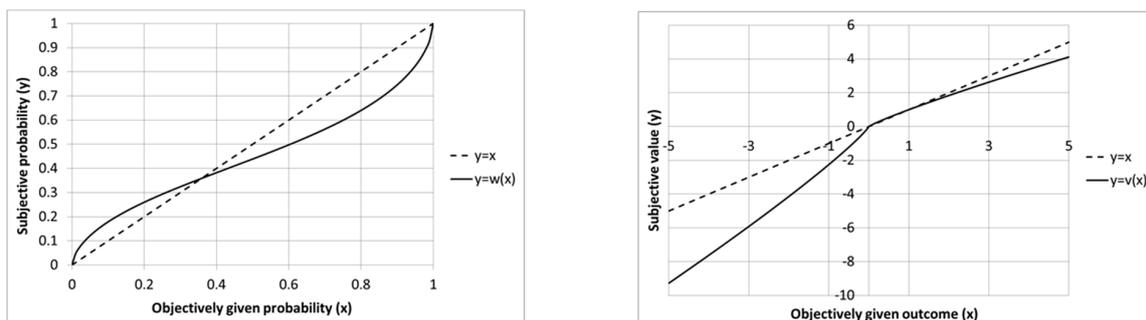
where  $x$  represents an objectively given outcome. The value function depends on two parameters  $\alpha$  and  $\lambda$ . If both parameters are one, the function is linear, which corresponds with the linear expected utility theory. In the original literature on the prospect theory [38],  $\alpha = 0.88$  and  $\lambda = 2.25$  are typical values that best fit experimental data. We adopt these parameter values in the following analysis. The right panel in Figure 1 shows the value function with these parameter values. Parameter  $\alpha < 1$  implies that the value function is concave and that a person obeying this function is risk-averse. Parameter  $\lambda > 1$  means that a person following this function is more sensitive to a loss than a gain (“loss-averse” so to say).

The subjectively distorted payoff matrix  $V$  does not depend on strategies. Note that since the function is nonlinear, the baseline of the payoff matrix or the reference point influences the results, which is different from the linear expected utility theory. We assume that the payoff obtained from the outcome (4, 4) in  $M$ , which is zero, is the reference point.

With these subjective probabilities and payoffs, the expected payoff is calculated by

$$P_i = \sum_{m=1}^4 \sum_{n=1}^4 (V)_{mn} (W_i)_{mn}. \tag{12}$$

The strategy change is described by the ordinary replicator dynamics mentioned in the last subsection (Equation (7)).



**Figure 1.** Left panel: The weighted function (solid curve) defined by Equation (9) with parameter  $\gamma = 0.65$ . The horizontal axis represents objectively given probabilities  $x$  and the vertical axis is subjective probability denoted by  $y$ . The linear function with  $\gamma = 1$  corresponding to the linear expected utility theory is also displayed (dashed line). Right panel: The value function (solid curve) given by Equation (11) with parameters  $\alpha = 0.88$ ,  $\lambda = 2.25$ . The  $x$ -axis represents objectively given outcomes and the  $y$ -axis subjective values. The linear function with  $\alpha = \lambda = 1$  is also shown (dashed line).

### 3. Results

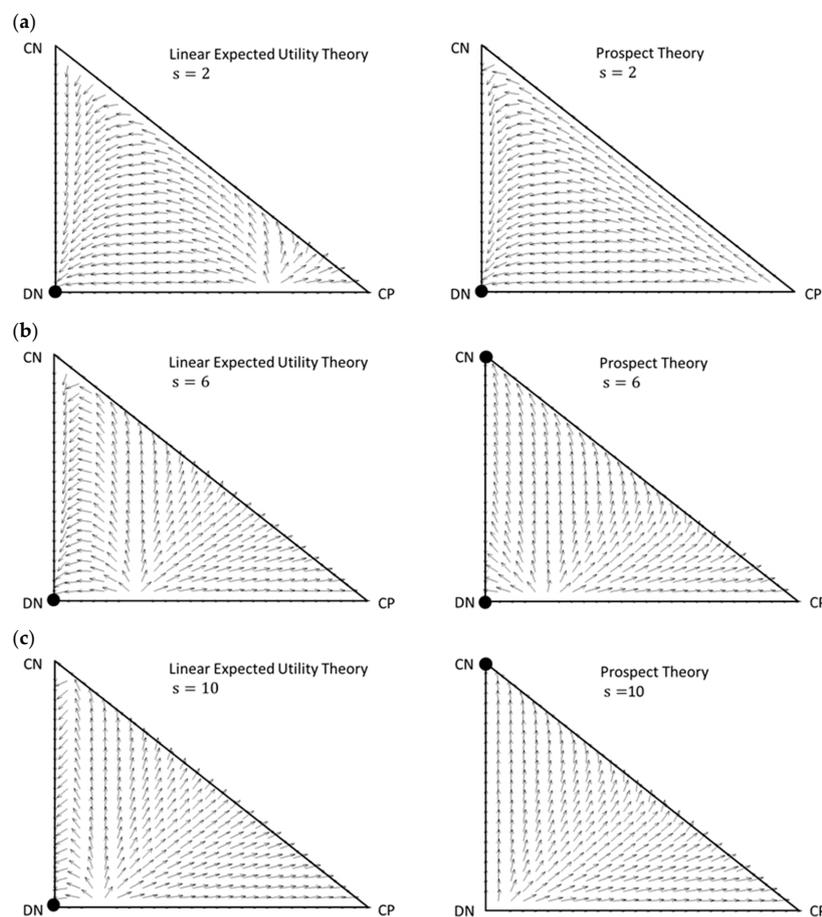
By calculating  $P_{ij}$  for all  $i$  and  $j$ , we find some basic characteristics of the model. Firstly, the contradictory strategy DP (strategy three) is dominated by DN. Since it never becomes evolutionarily stable, we eliminate the strategy from the analysis. We are interested in the time evolution of the vector  $(x_1, x_2, x_4)$ , which we call the state of the population.

Secondly, if the strength of punishment  $s$  is so large that it exceeds  $c/\epsilon$ ,  $P_2 > P_4$  holds for the linear expected utility theory. This means that strategy two (CN) dominates strategy four (DN). In other words, the second-order dilemma does not occur even with the linear expected utility theory. We exclude these trivial situations and focus on the parameter region  $s < \frac{c}{\epsilon} \stackrel{\text{def}}{=} s_{max}$ , in which the second-order dilemma (thus the first-order dilemma) occurs in the framework of the linear expected utility theory. The addressed question is whether the second-order dilemma will be solved in this parameter region under the prospect theory.

In the following analysis, we set  $c = r = 1$  and  $b = 4$  to reduce the dimension of the parameter space. Therefore, the variable parameters are the strength of punishment  $s$  and error rate  $\varepsilon$ .

### 3.1. Vector Fields

We show vector fields generated by the respective replicator dynamics derived from the linear expected utility theory and prospect theory, and we compare them in Figure 2. The state space is the simplex  $\{(x_1, x_2, x_4) | 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_4 \leq 1, x_1 + x_2 + x_4 = 1\}$ , which will be represented as a rectangular triangle in Figure 2. The vertex CP corresponds to  $(x_1, x_2, x_4) = (1, 0, 0)$ , CN to  $(0, 1, 0)$ , and DN to  $(0, 0, 1)$ . The state of the population  $(x_1, x_2, x_4)$  evolves along the vector fields in the state space. In the figure, stable fixed points are illustrated as solid circles. Note that the vector fields are normalized so that all vectors' lengths are one. The strength of punishment is varied from  $s = 2, s = 6$ , to  $s = 10$  and we set the error rate at 0.05 as an example here. As we will see, we find qualitatively different outcomes (vector fields) with these three parameter sets. The same changes in outcomes occur if we set the error rate at other values (see Figure 3 in the next subsection in which both parameters  $s$  and  $\varepsilon$  are varied).



**Figure 2.** The vector fields yielded by the replicator dynamics for the linear expected utility theory (left panel) and for the prospect theory (right panel). The state space is the simplex defined by  $\{(x_1, x_2, x_4) | 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_4 \leq 1, x_1 + x_2 + x_4 = 1\}$ , which is drawn as a rectangular triangle. The arrows in each triangle show in which direction the state  $(x_1, x_2, x_4)$  evolves in the rectangular triangle (including its edges). Parameters:  $c = r = 1$ ,  $b = 4$ ,  $\varepsilon = 0.05$ . The strength of punishment is varied: (a)  $s = 1$ , (b)  $s = 6$ , (c)  $s = 10$ . Stable fixed points are illustrated as solid circles. We see that CN becomes stable as  $s$  becomes larger for prospect theory, while DN is the unique stable fixed point in all cases for the linear expected utility theory.

If the strength of the punishment is low (panel (a)), the defective state DN is the only stable fixed point in both cases. On the edge CP–CN, CN dominates, and on the edge CN–DN, DN dominates. The difference of the two theories is found in the edge DN–CP. In the case of the linear expected utility theory, there is an unstable fixed point in the edge, and the system is bistable. On the other hand, strategy DN dominates strategy CP in the edge in case of prospect theory.

This tendency does not change if the strength of punishment is increased in the framework of the linear expected utility theory (the left picture of panel (b)). However, in the case of the prospect theory (the right picture of panel (b)), two unstable fixed points emerge in the edge CN–DN and in the edge DN–CP. As a result, the system becomes bistable. Thus, depending on initial conditions, trajectories approach either the cooperative state CN or defective state DN.

When the strength of punishment is increased further, there is no qualitative change for the linear expected utility theory (the left picture of panel (c)). Still, DN is the unique stable fixed point. In the case of prospect theory, however, DN becomes unstable and CN is the only stable fixed point in the state space; all trajectories (except for trajectories starting in the edge DN–CP) approach the cooperative state CN.

### 3.2. Stability Analysis of DN and CN

The above analysis clarified that the strength of punishment largely affects the system in the prospect theory. In fact, CN becomes “more” stable as the strength of punishment becomes large. Moreover, CP is always unstable regardless of the parameters. Taking these into account, we focus on the relation between CN and DN in the framework of prospect theory in this subsection. In order to investigate the effect of both parameters  $s$  and  $\varepsilon$ , we look for conditions for the parameters with which CN can invade DN and/or vice versa.

To do this, we define  $H_{ij}$  ( $i \in \{2, 4\}$ ,  $j \in \{2, 4\}$ ) as the expected payoff of individuals with strategy  $i$  in the situation where they only encounter individuals with strategy  $j$ . This expected payoff is given by

$$H_{ij} = \sum_{m=1}^4 \sum_{n=1}^4 v((M)_{mn})w(E_{im}E_{jn}), \quad (13)$$

where  $v$  and  $w$  are the value function and the weighted function, respectively,  $(M)_{mn}$  is the  $(m, n)$ -element of payoff matrix  $M$ , and  $E_{im}$  ( $E_{jn}$ ) is the  $m$ -th ( $n$ -th) element of the error vector  $\vec{E}_i$  ( $\vec{E}_j$ ). Or, we can obtain  $H_{ij}$  by substituting  $x_j = 1$  into  $P_i$  defined by Equation (12):  $H_{ij} = P_i$  with  $x_j = 1$ .

For instance,  $H_{42}$  represents the expected payoff in the situation where individuals with the DN strategy only encounter those with the CN strategy, and  $H_{24}$  is the expected payoff in the situation where individuals with the CN strategy only encounter those with the DN strategy.

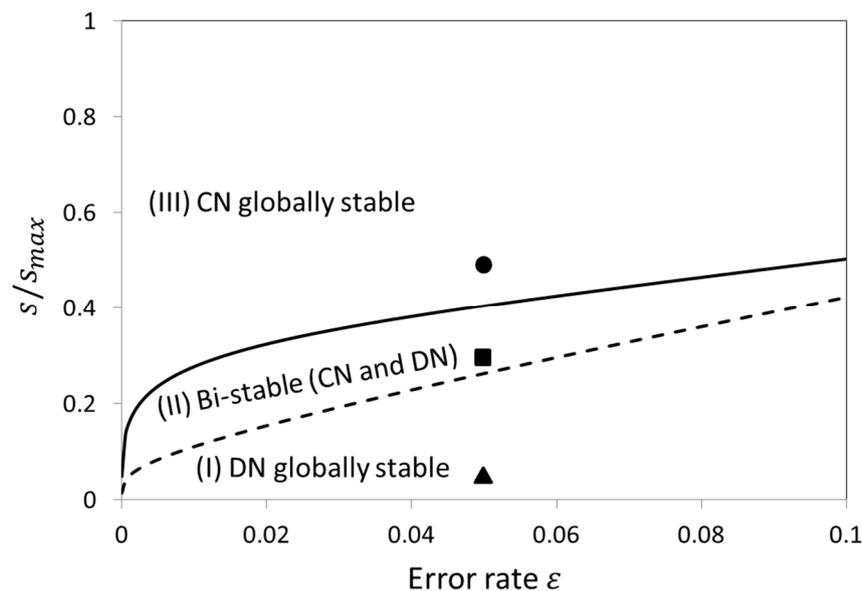
Then, the inequality  $H_{22} < H_{42}$  implies that a DN individual can invade into the population consisting of CN individuals. The inequality  $H_{44} > H_{24}$  means that the population consisting of DN individuals cannot be invaded into by a CN individual. If both inequalities hold, which corresponds to the right panel of Figure 2a), we see that DN is globally stable. Likewise, if both inequalities  $H_{22} > H_{42}$  and  $H_{44} < H_{24}$  are true, CN is globally stable, which corresponds to the right panel of Figure 2c. If both  $H_{22} > H_{42}$  and  $H_{44} > H_{24}$  hold, DN cannot invade into CN and vice versa. Therefore, the system is bistable (The right panel of Figure 2b). We look for regions in the parameter space  $(\varepsilon, s)$  in which the above inequalities hold.

The result is shown in Figure 3. Note that the strength of punishment  $s$  is normalized; thus, the vertical axis represents the value of  $s/s_{max}$ . If this parameter value exceeds 1, the second-order dilemma is resolved even in the case of the linear expected utility theory, and CN is globally stable. This means that in the parameter space given in the figure, the second-order dilemma occurs in the case of the linear expected utility theory; thus, DN is globally stable. The question is where the prospect theory can resolve the dilemma in this parameter region.

In the framework of the prospect theory, regardless of parameter  $\epsilon$ , DN is globally stable with small  $s$  (region (I)). In this region there is no qualitative difference from the linear expected utility theory. The solid triangle illustrated in this region in the figure corresponds to the parameter set ( $\epsilon = 0.05, s = 1$ ) used to generate Figure 2a. However, increasing the strength of punishment gradually destabilizes DN and alternatively stabilizes CN in the case of the prospect theory. As a result, the system becomes bistable (region (II)) with moderate  $\epsilon$ . The solid square in this region corresponds to ( $\epsilon = 0.05, s = 6$ ), which was used to generate Figure 2b. With high  $\epsilon$ , CN becomes able to invade DN populations, and CN becomes globally stable. We also see that the values of  $\frac{s}{s_{max}}$  that represent boundaries between the regions (illustrated by solid and dashed curves in the figure) become large as functions of error rate  $\epsilon$ . The solid circle put above the boundary in this region corresponds to ( $\epsilon = 0.05, s = 10$ ) used for Figure 2c.

The boundary drawn by the dashed curve in Figure 3 was found by numerically solving the equation  $H_{22} = H_{42}$  with respect to  $s$  for  $\epsilon = \epsilon_{min}, \epsilon_{min} + \delta, \epsilon_{min} + 2\delta, \dots, \epsilon_{max}$ , where  $\epsilon_{min} = 10^{-5}$ ,  $\epsilon_{min} = 0.1$ , and  $\delta = (\epsilon_{max} - \epsilon_{min})/200$ . Note that we excluded  $\epsilon = 0$ , since cooperation cannot evolve even under the prospect theory in this case. The inequality  $H_{22} < H_{42}$  ( $H_{22} > H_{42}$ ) holds under (above) this boundary. In order to find the boundary illustrated by the solid curve in Figure 3, the equation  $H_{24} = H_{44}$  with respect to  $s$  was solved. The inequality  $H_{24} < H_{44}$  ( $H_{24} > H_{44}$ ) holds under (above) this boundary.

As mentioned above, we used a specific set of values for parameters  $\gamma$ ,  $\alpha$ , and  $\lambda$  to produce Figure 3. We also generated figures with different parameter values, which are presented in the supporting material. No qualitative differences are found when the values of these parameters are varied. Nevertheless, we see that stable regions become larger as values of parameters  $\gamma$  and  $\alpha$  are smaller and that the parameter  $\lambda$  does not affect the results very much. This indicates that the system becomes more easily stable as the weighted function, and the value function gains “more nonlinearity”.



**Figure 3.** Different domains in the parameter space ( $\epsilon, \frac{s}{s_{max}}$ ) for prospect theory: (I) DN is globally stable (under the dashed line) in the case of the prospect theory, (II) both CN and DN are stable (the region sandwiched by the solid and dashed line). (III) CN is globally stable (the region above the solid line). In the parameter region shown in the figure, DN is globally stable in the case of the linear expected utility theory. The solid triangle in the figure corresponds to the parameter set ( $\epsilon = 0.05, s = 1$ ) used to generate Figure 2a, the solid square to ( $\epsilon = 0.05, s = 6$ ), which was used to generate Figure 2b and the solid circle to ( $\epsilon = 0.05, s = 10$ ) for Figure 2c. We see that CN becomes stable as  $s$  becomes larger for each error rate. However, the boundaries depicted by solid and dashed curves are monotonically increasing functions of error rate  $\epsilon$ .

#### 4. Discussion

We explored the effects of cognitive distortions described by the prospect theory in the context of the evolution of cooperation by peer punishment. Our main finding is that cognitive distortions make it possible for cooperation to evolve even in the parameter region in which cooperation cannot be achieved in the framework of the linear expected utility theory. We also found that the cooperative punitive strategy (CP) is dominated by cooperative nonpunitive strategy (CN). Thus, not CP but CN plays an essential role to sustain cooperation.

The CN players basically do not punish others, but they actually punish with a small probability due to error. This erroneous punishment effectively drives cooperation as far as the prospect theory is utilized. The same erroneous punishment cannot promote cooperation when the linear expected utility theory is adopted.

A key reason for why the error has a large impact on the population in the case of the prospect theory is found in the weighted function  $w$  and the value function  $v$ . According to the weighted function, individuals estimate small probabilities greater than they really are, and according to the value function, they assess the loss of benefit more than it actually is. Therefore, agents described by the prospect theory feel that punishment is more frequent and harsher than they really are.

In other words, individuals described by prospect theory are more sensitive to peer punishment than those described by the linear expected theory. Alternatively, we could say that individuals described by the prospect theory have the ability to imagine punishment, and this kind of imagination induces the fear to be punished. Such individuals who are afraid of punishment, even if it seldom occurs due to error, choose to cooperate to avoid being punished.

In spite of the findings mentioned so far, we have to remark that there remain several issues relevant for the coevolution of cooperation and peer punishment. The model studied in this present research especially has many limitations, which offers some tasks for future research.

Here, we mention the following two issues: the emergence problem and antisocial punishment. The emergence problem relates to the question about who starts giving punishment for the first time in the population [28]. Even in pool punishment, it is not easy to start costly punishment successfully, since punishing right and left in a sea of defectors imposes too much effort and cost on punishers. Several studies have proposed additional mechanisms or assumptions to overcome this emergence problem [42–46]. A similar problem holds for the peer punishment studied in this paper. We assumed that individuals punish others due to an error with a small probability. That is, individuals have an idea that they have the option to punish others from the beginning, even though the punishing activities are performed unintentionally. Under this assumption, we analyzed differences between the linear expected utility theory and the prospect theory.

Moreover, we assumed that only defectors are punished, and there is no chance for cooperators to be punished. Allowing punishment against prosocial behavior, such as cooperative actions, can offset the payoff advantage of the cooperators over free-riding. If antisocial punishment is included into our model, individuals afraid of counterpunishment might stop choosing cooperation. This problem caused by antisocial punishment occurs not only in our model, but is widely recognized as a serious issue in the literature of evolution of cooperation with punishment [47–49].

On the other hand, if we turn our attention to indirect reciprocity, which is known to be a powerful mechanism for the evolution of cooperation, the evolution of social norms is extensively investigated [50–55]. Social norms are defined as views on what is “good” or “bad”, and indirect reciprocity works in the way that bad individuals are discriminated in the population (bad individuals are not supported). However, there are many possibilities for the definitions of what is good or bad (thus social norms), and one of the main tasks in indirect reciprocity is to search for evolutionarily stable social norms that can maintain cooperation.

From the viewpoint of indirect reciprocity, the assumption that only defectors are punished is equivalent to assuming that the population has the unique social norm that prescribes to assessments of defectors as bad (and bad individuals are punished). This type of social norm is named “Scoring” in

the literature. Then, punishing cooperators (antisocial punishment) is equivalent to the social norm that regards cooperators as bad (“Antiscoring”). However, these are just a few examples of social norms. It is possible to consider other types of social norms, for instance, a norm that regards those as bad who cooperate with bad individuals.

A recent study using agent-based simulations revealed that prosocial norms such as “Scoring” can evolve, and antisocial norms such as “Antiscoring” become extinct in the melting pot of social norms if not a few but diverse norms coexist in the population [56]. In the model of this study, bad individuals are not punished, but they are not given help in the population. We can modify the model so that bad individuals are punished. Whether prosocial norms can also evolve in this modified model and thus whether the problem of antisocial punishment is solved is interesting and necessary research yet to be done. Recently, a paper was published which studies the coevolution of indirect reciprocity and punishment [57].

In extending our model with many social norms, agent-based simulations could be useful rather than an analytical approach taken in this paper. In this paper, we have discussed the evolution of cooperation in an analytical way under the assumption that there is a unique social norm in the population, and we have provided the first step to study the effects of cognitive distortions on the evolution of cooperation via peer punishment.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4336/10/1/11/s1>, Figure S1: Different domains in the parameter space  $(\epsilon, \frac{s}{s_{max}})$  for prospect theory (the upper panel) and the used weighted function and value function. The parameter  $\lambda$  is varied from 1.5 to 3.0 with  $\gamma$  and  $\alpha$  fixed (at 0.65 and 0.88 respectively). The blue curves correspond to  $\lambda = 1.5$ , red curves to  $\lambda = 2.25$  and green curves to  $\lambda = 3.0$ . Figure S2: Different domains in the parameter space  $(\epsilon, \frac{s}{s_{max}})$  for prospect theory (the upper panel) and the used weighted function and value function (the left-lower panel and the right lower panel). The parameter  $\alpha$  is varied from 0.8 to 0.96 with  $\gamma$  and  $\lambda$  fixed (at 0.65 and 2.25 respectively), Figure S3: Different domains in the parameter space  $(\epsilon, \frac{s}{s_{max}})$  for prospect theory (the upper panel) and the used weighted function and value function (the left-lower panel and the right lower panel). The parameter  $\lambda$  is varied from 1.5 to 3.0 with  $\gamma$  and  $\alpha$  fixed (at 0.65 and 0.88 respectively).

**Author Contributions:** All authors conceived and designed the project. S.U. and H.Y. built and analyzed the model. S.U. wrote the paper. All authors discussed the results, helped draft and revised the manuscript, and approved the submission.

**Funding:** Part of this work was supported by JSPS (Grants-in-Aid for Scientific Research) 15KT0133 (HY), 17KK0055(IO), 16H03120(HY and IO), 17H02044(IO and HY), 18H03498 (HY and IO) and the Austrian Science Fund (FWF) P27018-G11 (TS).

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. Nowak, M.A.; Highfield, R. *Super Cooperators*; Free Press: New York, NY, USA, 2011.
2. Ostrom, E. *Governing the Commons: The Evolution of Institutions for Collective Action*; Cambridge University Press: Cambridge, UK, 1990.
3. Bowls, S.; Gintis, H. *A Cooperative Species*; Princeton University Press: Princeton, UK; Oxford, UK, 2011.
4. Yamagishi, T. *Trust: The Evolutionary Game of Mind and Society*; Springer: New York, NY, USA, 2011.
5. Sigmund, K. *The Calculus of Selfishness*; Princeton University Press: Princeton, UK; Oxford, UK, 2010.
6. Perc, M.; Jordan, J.J.; Rand, D.G.; Wang, Z.; Boccaletti, S.; Szolnoki, A. Statistical physics of human cooperation. *Phys. Rep.* **2017**, *68*, 1–51. [[CrossRef](#)]
7. Nowak, M.A. *Evolutionary Dynamics*; Harvard University Press: Cambridge, MA, USA, 2006.
8. Nowak, M.A. Five rules for the evolution of cooperation. *Science* **2006**, *314*, 1560–1563. [[CrossRef](#)] [[PubMed](#)]
9. Balliet, D.; Mulder, L.B.; Van Lange, P.A. Reward, punishment, and cooperation: A meta-analysis. *Psychol. Bull.* **2011**, *137*, 594–615. [[CrossRef](#)] [[PubMed](#)]
10. Guala, F. Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behav. Brain Sci.* **2012**, *35*, 1–15. [[CrossRef](#)] [[PubMed](#)]
11. Axelrod, R. An evolutionary approach to norms. *Am. Political Sci. Rev.* **1986**, *80*, 1095–1111. [[CrossRef](#)]

12. Henrich, J.; McElreath, R.; Barr, A.; Ensminger, J.; Barrett, C.; Bolyanatz, A.; Cardenas, J.C.; Gurven, M.; Gwako, E.; Henrich, N.; et al. Costly punishment across human societies. *Science* **2006**, *312*, 1767–1770. [[CrossRef](#)] [[PubMed](#)]
13. Mathew, S.; Boyd, R. Punishment sustains large-scale cooperation in prestate warfare. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11375–11380. [[CrossRef](#)] [[PubMed](#)]
14. Casari, M.; Luini, L. Cooperation under alternative punishment institutions: An experiment. *J. Econ. Behav. Organ.* **2009**, *71*, 273–282. [[CrossRef](#)]
15. Fehr, E.; Gächter, S. Altruistic punishment in humans. *Nature* **2002**, *415*, 137–140. [[CrossRef](#)] [[PubMed](#)]
16. Boyd, R.; Gintis, H.; Bowles, S.; Richerson, P.J. The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3531–3535. [[CrossRef](#)] [[PubMed](#)]
17. Sigmund, K.; Hauert, C.; Nowak, M.A. Reward and punishment. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10757–10762. [[CrossRef](#)] [[PubMed](#)]
18. Milinski, M.; Rockenbach, B. Human behaviour: Punisher pays. *Nature* **2008**, *452*, 297–298. [[CrossRef](#)] [[PubMed](#)]
19. Kosfeld, M.; Okada, A.; Riedl, A. Institution formation in public goods games. *Am. Econ. Rev.* **2009**, *99*, 1335–1355. [[CrossRef](#)]
20. Boyd, R.; Richerson, P.J. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethol. Sociobiol.* **1992**, *13*, 171–195. [[CrossRef](#)]
21. Sigmund, K.; de Silva, H.; Traulsen, A.; Hauert, C. Social learning promotes institutions for governing the commons. *Nature* **2010**, *466*, 861–863. [[CrossRef](#)] [[PubMed](#)]
22. Yamagishi, T. The provision of a sanctioning system as a public good. *J. Personal. Soc. Psychol.* **1986**, *51*, 110–116. [[CrossRef](#)]
23. Traulsen, A.; Röhl, T.; Milinski, M. An economic experiment reveals that humans prefer pool punishment to maintain the commons. *Proc. Biol. Sci.* **2012**, *279*, 3716–3721. [[CrossRef](#)] [[PubMed](#)]
24. Andreoni, J.; Gee, L.K. Gun for hire: Delegated enforcement and peer punishment in public goods provision. *J. Public Econ.* **2012**, *96*, 1036–1046. [[CrossRef](#)]
25. Zhang, B.; Li, C.; De Silva, H.; Bednarik, P.; Sigmund, K. The evolution of sanctioning institutions: An experimental approach to the social contract. *Exp. Econ.* **2014**, *17*, 285–303. [[CrossRef](#)]
26. Schoenmakers, S.; Hilbe, C.; Blasius, B.; Traulsen, A. Sanctions as honest signals—The evolution of pool punishment by public sanctioning institutions. *J. Theor. Biol.* **2014**, *356*, 36–46. [[CrossRef](#)] [[PubMed](#)]
27. Okada, I.; Yamamoto, H.; Toriumi, F.; Sasaki, T. The effect of incentives and meta-incentives on the evolution of cooperation. *PLoS Comput. Biol.* **2015**, *11*, e1004232. [[CrossRef](#)] [[PubMed](#)]
28. Sasaki, T.; Uchida, S.; Chen, X. Voluntary rewards mediate the evolution of pool punishment for maintaining public goods in large populations. *Sci. Rep.* **2015**, *5*, 8917. [[CrossRef](#)] [[PubMed](#)]
29. Hilbe, C.; Traulsen, A.; Röhl, T.; Milinski, M. Democratic decisions establish stable authorities that overcome the paradox of second-order punishment. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 752–756. [[CrossRef](#)] [[PubMed](#)]
30. Sasaki, T.; Brännström, Å.; Dieckmann, U.; Sigmund, K. The take-it-or-leave-it option allows small penalties to overcome social dilemmas. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1165–1169. [[CrossRef](#)] [[PubMed](#)]
31. Sasaki, T.; Okada, I.; Uchida, S.; Chen, X. Commitment to cooperation and peer punishment: Its evolution. *Games* **2015**, *6*, 574. [[CrossRef](#)]
32. Tversky, A.; Kahneman, D. Judgement under uncertainty: Heuristics and biases. *Science* **1974**, *185*, 1124–1131. [[CrossRef](#)] [[PubMed](#)]
33. Tversky, A.; Kahneman, D. Extensional vs. intuitive reasoning: The conjunction fallacy in probability judging. *Psychol. Rev.* **1983**, *90*, 293–315. [[CrossRef](#)]
34. Schmeidler, D. Subjective probability and expected utility without additivity. *Econometrica* **1989**, *57*, 571–587. [[CrossRef](#)]
35. Gilboa, I.; Schmeidler, D. Maxmin expected utility with a non-unique prior. *J. Math. Econ.* **1989**, *18*, 141–153. [[CrossRef](#)]
36. Starmer, C. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *J. Econ. Lit.* **2000**, *38*, 332–382. [[CrossRef](#)]
37. Machina, M.J. Expected utility analysis without the independence axiom. *Econometrica* **1982**, *50*, 277–323. [[CrossRef](#)]

38. Kahneman, D.; Tversky, A. Prospect theory: Analysis of decision under risk. *Econometrica* **1979**, *47*, 263–291. [[CrossRef](#)]
39. Tversky, A.; Kahneman, D. Loss aversion in riskless choice: A reference-dependent model. *Q. J. Econ.* **1991**, *106*, 1039–1061. [[CrossRef](#)]
40. Wakker, P.P. *Prospect Theory: For Risk and Ambiguity*; Cambridge University Press: Cambridge, UK, 2010.
41. Hofbauer, J.; Sigmund, K. *Evolutionary Games and Population Dynamics*; Cambridge University Press: Cambridge, UK, 1998.
42. Boyd, R.; Gintis, H.; Bowles, S. Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* **2010**, *328*, 617–620. [[CrossRef](#)] [[PubMed](#)]
43. Raihani, N.J.; Bshary, R. The evolution of punishment in n-player public goods games: A volunteer's dilemma. *Evolution* **2011**, *65*, 2725–2728. [[CrossRef](#)] [[PubMed](#)]
44. Brandt, H.; Hauert, C.; Sigmund, K. Punishing and abstaining for public goods. *Proc. Natl Acad. Sci. USA* **2006**, *103*, 495–497. [[CrossRef](#)] [[PubMed](#)]
45. Dercole, F.; De Carli, M.; Della Rossa, F.; Papadopoulos, A.V. Overpunishing is not necessary to fix cooperation in voluntary public goods games. *J. Theor. Biol.* **2013**, *326*, 70–81. [[CrossRef](#)] [[PubMed](#)]
46. Hauert, C.; Traulsen, A.; Brandt, H.; Nowak, M.A.; Sigmund, K. Via freedom to coercion: The emergence of costly punishment. *Science* **2007**, *316*, 1905–1907. [[CrossRef](#)] [[PubMed](#)]
47. Nikiforakis, N. Punishment and counter-punishment in public good games: Can we really govern ourselves? *J. Public Econ.* **2008**, *92*, 91–112. [[CrossRef](#)]
48. Rand, D.G.; Nowak, M.A. The evolution of antisocial punishment in optional public goods games. *Nat. Commun.* **2011**, *2*, 434. [[CrossRef](#)] [[PubMed](#)]
49. García, J.; Traulsen, A. Leaving the loners alone: Evolution of cooperation in the presence of antisocial punishment. *J. Theor. Biol.* **2012**, *307*, 168–173. [[CrossRef](#)] [[PubMed](#)]
50. Ohtsuki, H.; Iwasa, Y. The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *J. Theor. Biol.* **2006**, *239*, 435–444. [[CrossRef](#)] [[PubMed](#)]
51. Nowak, M.A.; Sigmund, K. Evolution of indirect reciprocity. *Nature* **2005**, *437*, 1292–1298. [[CrossRef](#)] [[PubMed](#)]
52. Sasaki, T.; Okada, I.; Nakai, Y. The evolution of conditional moral assessment in indirect reciprocity. *Sci. Rep.* **2017**, *7*, 41870. [[CrossRef](#)] [[PubMed](#)]
53. Uchida, S.; Sigmund, K. The competition of assessment rules for indirect reciprocity. *J. Theor. Biol.* **2010**, *263*, 13–19. [[CrossRef](#)] [[PubMed](#)]
54. Chalub, F.; Santos, F.C.; Pacheco, J.M. The evolution of norms. *J. Theor. Biol.* **2006**, *241*, 233–240. [[CrossRef](#)] [[PubMed](#)]
55. Uchida, S.; Yamamoto, H.; Okada, I.; Sasaki, T. A Theoretical Approach to Norm Ecosystems: Two Adaptive Architectures of Indirect Reciprocity Show Different Paths to the Evolution of Cooperation. *Front. Phys.* **2018**, *6*, 14. [[CrossRef](#)]
56. Yamamoto, H.; Okada, I.; Uchida, S.; Sasaki, T. A norm knockout method on indirect reciprocity to reveal indispensable norms. *Sci. Rep.* **2017**, *7*, 44146. [[CrossRef](#)] [[PubMed](#)]
57. Schlaepfer, A. The emergence and selection of reputation systems that drive cooperative behaviour. *Proc. R. Soc. B Biol. Sci.* **2018**, *285*, 20181508. [[CrossRef](#)] [[PubMed](#)]

