A Review of Theoretical Studies on Indirect Reciprocity

Isamu Okada [1,2,†]

[1] Faculty of Business Administration, Soka University, Hachioji City, Tokyo 192-8577, Japan; okada@soka.ac.jp;
    Tel.: +81-42-691-8904
[2] Department of Information Systems and Operations, Vienna University of Economics, 1020 Vienna, Austria
† Current address: Tangi 1-236, Hachioji City, Tokyo 192-8577, Japan

Abstract: Despite the accumulation of research on indirect reciprocity over the past 30 years and the publication of over 100,000 related papers, there are still many issues to be addressed. Here, we look back on the research that has been done on indirect reciprocity and identify the issues that have been resolved and the ones that remain to be resolved. This manuscript introduces indirect reciprocity in the context of the evolution of cooperation, basic models of social dilemma situations, the path taken in the elaboration of mathematical analysis using evolutionary game theory, the discovery of image scoring norms, and the breakthroughs brought about by the analysis of the evolutionary instability of the norms. Moreover, it presents key results obtained by refining the assessment function, resolving the punishment dilemma, and presenting a complete solution to the social dilemma problem. Finally, it discusses the application of indirect reciprocity in various disciplines.

---

1. Introduction

While studies on indirect reciprocity over the past 30 years have led to the publication of more than 100,000 related papers, there are still many issues to be addressed. Indirect reciprocity is a major mechanism that answers the question of how cooperation evolves in social dilemma situations. Here, cooperation refers to the provision of a wide range of resources, including effort, time, and money, to the benefit of others or organizations, and by its definition is altruistic [1,2]. While classical game theory, which is based on the assumption of rational decision-making, teaches that cooperation does not emerge naturally, cooperative behavior is widespread not only in humans but also in animals. Mechanisms of cooperative behaviors include kinship, direct and indirect reciprocity, and group cohesiveness. Therefore, the evolution of cooperation has been studied not only in biology but also in a wide range of fields such as sociology, economics, politics, psychology, mathematics, and physics [3–7].

The analysis of social dilemma situations is an important task in the study of the evolution of cooperation. A situation constitutes a social dilemma if an individual rationally chooses not to cooperate with someone due to an economic or psychological incentive despite the fact that social welfare is maximized if all members of the group cooperate [8–13]. The cumulative effect of selfish behaviors is a society with low social welfare. Social dilemma situations are widely observed, from deciding whether to donate blood to deciding whether to take actions that protect the environment. The evolution of cooperation in social dilemma situations has attracted widespread social and academic interest because cooperation is maximized social welfare.

Indirect reciprocity is a well-known mechanism for maintaining cooperation among unrelated individuals in social dilemma situations. It involves cooperators requesting information about potential recipients. This cooperation mechanism prevents free riders, i.e., individuals who do not cooperate but receive benefits from naive cooperators, from invading the population of individuals. It works by imposing a discrimination criterion that prevents individuals from cooperating with individuals who do not have a good reputation.

Indirect reciprocity is considered to be one of five mechanisms by which cooperation evolves in the context of social dilemmas and is considered to be particularly versatile compared with the other mechanisms [6]. Because kin selection requires relative relationships, direct reciprocity requires repeated game play, while network reciprocity and group reciprocity require heterogeneous structures for interactions. If there are no such conditions, cooperation does not emerge. In contrast, indirect reciprocity does not require such conditions. It is thus a factor in the evolution of cooperation among individuals without direct interactions, so the breadth of its application area has attracted academic interest. Here, we look back on the studies of indirect reciprocity and identify the issues that have been resolved and the ones remaining to be resolved. Figure 1 presents an overview of indirect reciprocity research.
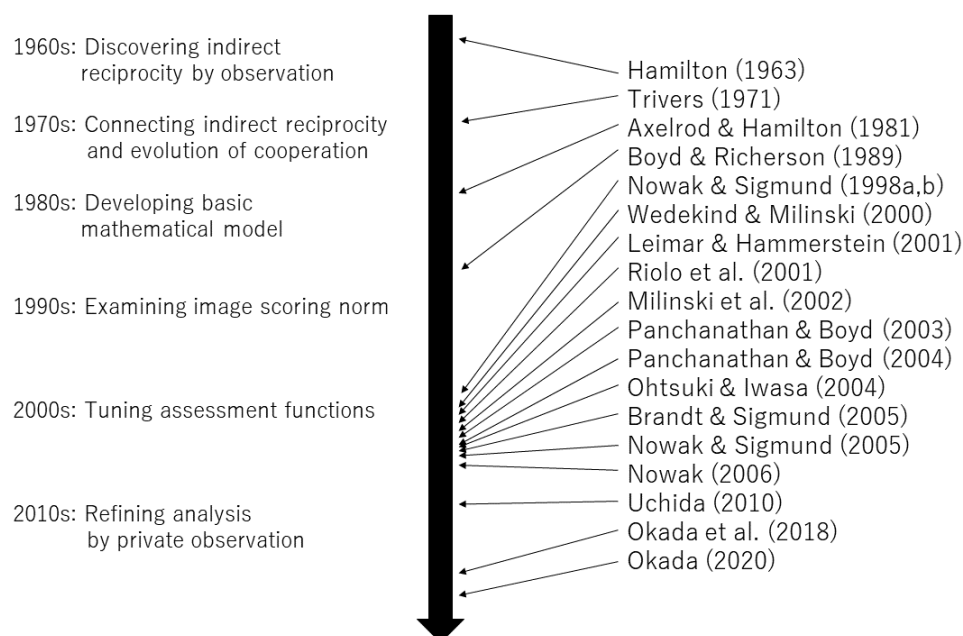
**Figure 1.** Historical overview of indirect reciprocity research and milestone papers.

2. Development of Research Areas and Research Methods

Indirect reciprocity was first identified in folklore research in the 1960s and was explicitly defined in relation to cooperative behavior in psychology in the late 1970s. For example, Eisenberg–Berg [14] showed through interviews with children who had faced moral judgment dilemmas that indirect reciprocity falls into the moral consideration category. Following a period of observing indirect reciprocity phenomena, efforts to understand the mechanism scientifically began in biology. Alexander [15–17] first discussed indirect reciprocity as a principle of human behavior and defined it as the provision of benefits among unrelated persons, using the framework of evolutionary game theory developed by Maynard Smith [18,19]. He focused on indirect reciprocity as a principle for explaining cooperative behaviors like altruism [1,2].

In direct reciprocity [2,20,21], the reciprocal cooperation attained by repeatedly playing with the same player maintains stable cooperation. Indirect reciprocity, in contrast, is based on the assumption of an indirect situation in which a player cooperates with a different player rather than the same

player. In other words, a player contributes to a player who was not the player from whom he or she received a contribution. There are two possible timings of the cooperative actions among these three players [22–24]. In upstream indirect reciprocity, Player X first cooperates with Player Y, and Player Y subsequently cooperates with Player Z. This timing is often observed in pay-it-forward situations. While economic experiments have attempted to reveal why upstream indirect reciprocity is often observed [25–27], explaining this timing theoretically is difficult because Player X does not receive any benefit from her or his cooperation [24,28–30].

In downstream indirect reciprocity [23,24], on the other hand, Player Y contributes to Player Z first and then Player X contributes to Player Y. This type of reciprocity can be theoretically rationalized because there is an economic incentive for Player Y to cooperate. Player Y expects to gain future cooperation by helping Player Z. Therefore, most theoretical analyses target downstream indirect reciprocity, and this manuscript reviews this type of indirect reciprocity. For the downstream type to work, cooperative actions must be communicated to a third party. In other words, as a precondition for establishing indirect reciprocity, society must have a system for distributing reputation information. This is because direct observation of all games is an extreme assumption, so it is necessary to assume a situation in which the behavior of a third party is instantly transmitted to society as a whole by reputation and gossip. Figure 2 summarizes the basic mechanisms of both direct and indirect reciprocity.
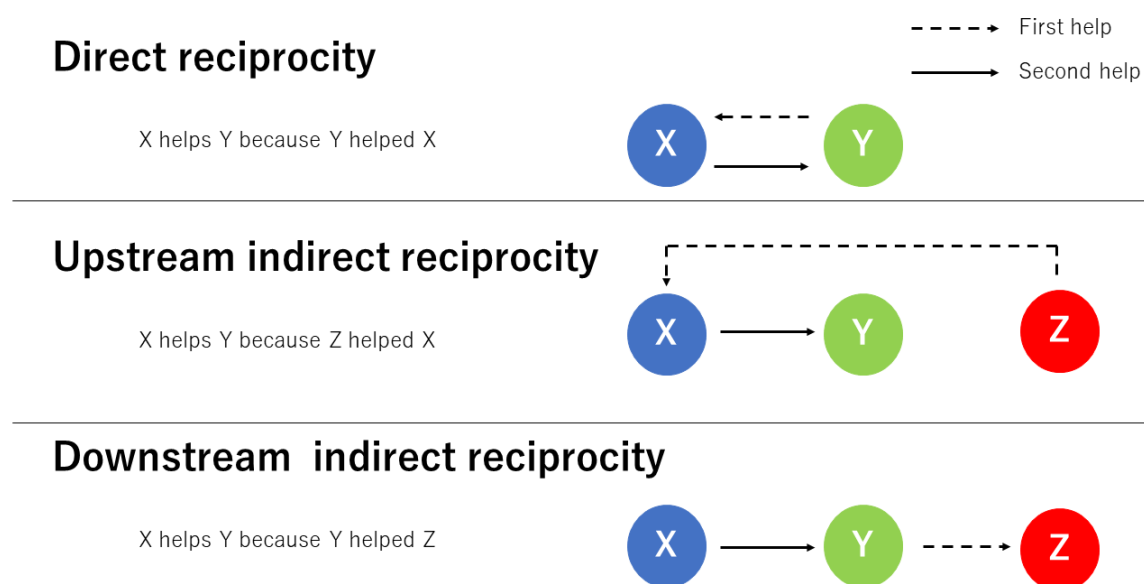
Figure 2. Basic mechanisms of direct and indirect reciprocity. This figure is based on Figure 1 of Nowak and Sigmund (2005) [23].

Boyd and Richerson [22] were the first to model indirect reciprocity in the framework of evolutionary games. They used indirect reciprocity as a principle to explain cooperation among large numbers of genetically unrelated individuals. Their idea was an extension on the TFT (tit-for-tat) strategy in the iterated prisoner's dilemma game introduced by Anatol Rapoport around 1980. The strategy is simple because a player who simply imitates the opponents latest behavior gains a relatively high payoff. Boyd and Richerson classified indirect reciprocity into two types, upstream and downstream, and asserted that downstream reciprocity has a cooperative equilibrium. However, they assumed a priori that information about all the other people in the group is available.

Nowak and Sigmund [31] introduced reputation information based on scores to explicitly utilize information about other people in the group. They assumed that each player is given an image score, the value of which increases when the player exhibits cooperative behavior and decreases when

the player does not cooperate when the opportunity to do so is presented. The players cooperate with those players whose scores exceed a threshold. That is, they explicitly stated that the players' reputation information, i.e., a summary of their previous behaviors, is necessary for cooperation to work [26,32–35]. This was an important breakthrough for indirect reciprocity research. Sigmund [24] later asserted that direct reciprocity requires repetition while indirect reciprocity requires reputation.

Nowak and Sigmund [36] also developed a method for analyzing stability by incorporating replicator dynamics [37] into their theoretical model of indirect reciprocity. They introduced into their model players who unconditionally do not cooperate (i.e., they defect) and players who cooperate unconditionally in addition to the players who adopt the norm[1] of indirect reciprocity. Nowak and Sigmund [36] developed a method for calculating the expected payoff for each player in any population. Moreover, they pioneeringly analyzed the population dynamics when the population ratio of the group was updated in accordance with the replicator dynamics from the expected payoffs. The development of this method enabled a rigorous mathematical analysis of stability.

3. Analytical Models of Indirect Reciprocity

Due to the pioneers' efforts, a basic analytical model of indirect reciprocity has been developed and shared. In the model, players play "giving games" (or "donation games"). A donor playing a game decides whether to cooperate (C) or defect (D). If the donor chooses C, she or he must pay a fixed amount c to a recipient. The recipient receives a fixed benefit b if and only if the donor contributes. In this game, the benefit cost ratio of a game satisfies $r \equiv b/c > 1$. If the donor chooses D, nothing happens. Self-interested myopic players choose D because contributors do not benefit from their own contribution, and thus the model reveals a social dilemma. Researchers in this field have typically focused on two types of errors. One is implementation error e_1, which means unintentional behavior. A player intending to cooperate may not actually cooperate due to this type of error. Strictly speaking, this type of implementation error is called a "unilateral implementation error". A "bilateral implementation error" is two types of errors: A player intending cooperation mistakenly defects and a player intending defect mistakenly cooperates. The second type of error is assessment error e_2, which means misperception of an assessment. This type of error can lead to a player being given an incorrect reputation.

Several developments have contributed to refined analysis of indirect reciprocity models. Nowak and Sigmund [31] introduced the use of the probability that a player knows (either through direct observation or via gossip) what a randomly assigned recipient did in the past. Brandt and Sigmund [38–40] developed a continuous-entry model to balance reputation values among the different types of groups. Ohtsuki et al. [41] established a two-time scale assumption, the analytical assumption that reputation dynamics are independent of strategy dynamics. It is based on the assumption that reputation information updating is much faster than strategy changing. Okada et al. [42] developed a solitary observation method that guarantees that a solution is obtained for reputation dynamics in a private assessment scheme. These developments have enabled a more systematic and comprehensive analysis of indirect reciprocity.

While there are many analytical models that deal with indirect reciprocity, here we introduce a typical replicator dynamics model consisting of three types of players. There are unconditional cooperators (X) who always cooperate, unconditional defectors (Y) who never cooperate, and discriminators (Z) who cooperate selectively on the basis of indirect reciprocity. To do so, they act in accordance with a rule that determines the action to be taken, either C or D. To analyze replicator dynamics, we must assume that

---

[1] "Norm" has often been used in theoretical studies of indirect reciprocity. In psychology and sociology, it means shared societal rules. In the studies of the evolution of cooperation, it means assessment rules for determining reputation and action rules for selecting cooperative behaviors.

an infinite number of well-mixed players play games repeatedly. Let x, y, and z be the proportions (population ratios) of X, Y, and Z, respectively, where $x + y + z = 1$ is always satisfied.

In a simple model, it is assumed that the discriminators cooperate only with those players who have good reputations. Let g be the fraction of good players. This fraction is decomposed into g_X, g_Y, and g_Z, where g_s is the fraction of good players with strategy s in the set $S = \{X, Y, Z\}$. $g = xg_X + yg_Y + zg_Z$ is always satisfied.

Let P_s be the expected payoff of an s type player, where $s \in S$. The expected payoff functions of the three strategists are:

$$
\begin{aligned}
P_X &= b(x + zg_x) - c \\
P_Y &= b(x + zg_y) \\
P_Z &= b(x + zg_z) - cg,
\end{aligned}
$$

where the factor $(1 - e_1)$ must be added if an implementation error is introduced.

We are ready to analyze a replicator dynamics model using these payoff functions. The dynamics are described as $\dot{x} = x(P_X - \bar{P}), \dot{y} = y(P_Y - \bar{P})$ and $\dot{z} = z(P_Z - \bar{P})$, where $\bar{P} = xP_X + yP_Y + zP_Z$ is the average payoff over the population. By drawing the dynamics completely, we can identify evolutionary stable equilibria in the population, locally asymptotic points, cyclical dynamics, neutral fixed points, neutral mutants, and so on. For example, an analysis shows that a discriminator (Z) is evolutionary stable because neither perfect cooperators (X) nor perfect defectors (Y) invade into a population consisting of only one type (Z) of player. Another analysis shows that the expected payoffs for a Z player and an X player in a group without any Y players are the same, so X and Z are able to invade to each other's population as neutral mutants.

4. Problems with Image Scoring Norm

Many analyses have been carried out on indirect reciprocity models by using evolutionary game theory. The first target was the image scoring norm [31,39,43–45], which follows very simple rules: Increase the score of a player who cooperates, decrease the score of a player who defects, and cooperate only with players with scores exceeding a certain threshold. This idea is simplified by using binary scoring based on the assumption that there are only two kinds of scores, good and bad, and by updating the scores in accordance with only the last behavior. A rigorous analysis of evolutionary stability revealed that the image scoring norm does not have evolutionary stability. This finding was important for the subsequent progress in research.

Rigorous analyses of the image scoring norm revealed three theoretical problems that prevent it from having evolutionary stability. The first is that a second-order free rider can become a neutral mutant [46–53], whose expected payoff is same as the other residents. The non-cooperators are called first-order free riders because they do not pay the costs of providing public goods while those who do not pay the costs of excluding first-order free riders are called second-order free riders. In the evolution of cooperation based on reciprocity, it is necessary to execute selective cooperation in order to exclude non-cooperators (first-order free riders). In other words, it is necessary to distinguish a cooperator from a defector and to cooperate with only the former. Perfect cooperators are free riders in this regard and are therefore called second-order free riders.

In a cooperative regime, players adopting the image scoring norm and those adopting the perfect cooperator norm behave in exactly the same way because a player who adopts the image scoring norm always cooperate in a cooperative regime, so they cannot be distinguished. Therefore, there is no difference in payoff between them, meaning that the perfect cooperators are neutral mutants against players who adopt the image scoring norm, and thus the perfect cooperators can invade the population of the image scoring norm players. In other words, even if image scoring eliminates the first-order free riders and results in the creation of a cooperative regime, the regime may still be invaded by second-order free riders (the perfect cooperators). The cooperative regime of the second-order free

riders is subsequently invaded by the first-order free riders, so cooperation cannot be maintained. This feature of the image scoring norm has been called its Achilles heel [54–56].

The second theoretical problem with the image scoring norm is the occurrence of bad reputation chains due to errors [57]. The third one is punisher downgrading. That is, with the image scoring norm, the score for a player who displays an uncooperative behavior because its partner is identified as having a bad reputation is downgraded [24,31]. This problem is called the scoring dilemma [23,24,44,45,55,58], a newly identified dilemma in trying to resolve the social dilemma in the framework of indirect reciprocity. Attempts to resolve this dilemma have brought a great breakthrough to indirect reciprocity research.

5. Refinement of Assessment Function

The solution to the scoring dilemma brought a new type of dilemma, and methods to overcome it have been sought. The direct solution is to identify non-cooperative behavior. How can we discriminate between non-cooperative behaviors that are selfish and those that are intended as punishment? Since the image scoring norm give scores to all players, it is natural to think that behaviors are discriminated on the basis of the recipient's score. To analyze this idea with a simple model, players are assigned binary reputation labels of good and bad[2]. Punishment is considered justified if it is non-cooperation with bad players and not non-cooperation with good players [44,58,61]. Many studies have analyzed norms in which justified punishment does not downgrade the punisher's reputation.

Many studies have looked for norms to replace the image scoring norm since refinement of the assessment function used to determine reputation became the main concern in the 2000s. Given a set of actions (Action $= \{C, D\}$) and a set of reputations (Reputation $= \{G, B\}$), the image scoring norm uses an assessment function, Assessment_{IS} : Donor's Action $\rightarrow$ Donor's Reputation, to assess the behaviors of each player. It can be assumed that $(\text{Assessment}_{IS}(C), \text{Assessment}_{IS}(D)) = (G, B)$. By extending this formulation, we can take into account justified punishment by defining a new assessment function: Donor's action $\times$ Recipient's reputation $\rightarrow$ Donor's reputation. Consideration was given to extending the domain of the assessment function in this way. In an exhaustive analysis, Ohtsuki and Iwasa [62] used Donor's action $\times$ Recipient's reputation $\times$ Donor's reputation $\rightarrow$ Donor's reputation as the assessment function. This formulation has been widely used in many studies. Santos et al. [63] conducted an analysis that included past information in the domain of the assessment function.

Consideration was also given to using an action function to extend the assessment function: $(\text{Action}_{IS}(G), \text{Action}_{IS}(B)) = (C, D)$ using the action function Action_{IS}: Recipient's reputation $\rightarrow$ Donor's Action. Expanding on this, Panchanathan and Boyd [58] analyzed two action functions in which the formula is Donor's reputation $\times$ Recipient's reputation $\rightarrow$ Donor's Action. Figure 3 shows the various configuration of assessment and action rules when binary reputation labels are used.

While many assessment and action functions have been formulated, determining how to maintain stable cooperation is an important task. Ohtsuki and Iwasa [62] provided a solution in their exhaustive analysis. By analyzing all combinations of 256 types of assessment functions and 16 types of action functions, they identified 8 types of norms, the "leading eight norms", that can maintain stable cooperation. A feature having all eight of these norms is justified punishment because, if the donor's reputation is Good, the donor's action is D, and if the recipient's reputation is Bad, the donor's reputation is updated to Good in the assessment functions of all eight norms. Moreover, if the donor's reputation is Good and the recipient's reputation is Bad, the donor's action is D in the action rules of all eight norms. These insights have been confirmed by their follow-up analyses [64,65] and other

---

[2] While many studies have dealt with only the binary reputation label, other studies have dealt with three or more reputation labels [59,60], as explained in detail by Rutte and Taborsly [25].

analyses [24,66,67] as well as by experimental validation [68,69] and analyses in other fields [70]. Studies on indirect reciprocity have made great progress due to their findings.
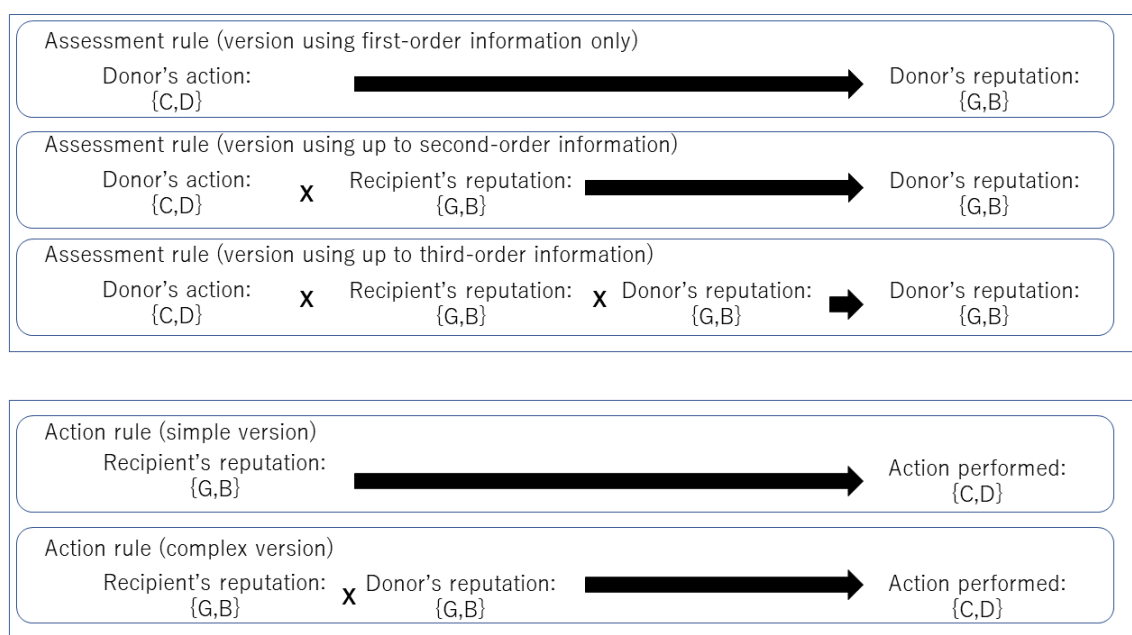
Figure 3. Assessment and action rules for binary reputation labeling. (G: "Good label"; B: "Bad label"; C: "Cooperation"; D: "Defect").

Several of the leading eight norms have been analyzed individually. The standing (or consistent-standing, L2) norm and the simple-standing (or L3) norm [71], which originated from Sugden(1986) [72], have been well analyzed. Their performance has been shown to be excellent [42,58,67,73,74]. The judging (or L8) norm and the stern-judging (or L6) norm [56,71,75–79], which was derived from Kandori(1992) [80], has also been well analyzed. While not included in the leading eight norms, the shunning norm [81], which is relatively simple, has also been analyzed individually.

6. Resolving the Punishment Dilemma

While Otsuki and Iwasa's exhaustive analysis [62] revealed the characteristics of the norms supporting indirect reciprocity, it has been pointed out by several authors over the past decade that the simple assumptions on which their model is based have revealed a new issue: The "punishment dilemma". Many theoretical analyses of indirect reciprocity have been, because of the ease of theoretical analysis, based on the assumption that all reputations are public information. However, it is unlikely that everyone will have the same impression of an individual.

If the assumption that all reputations are public information had not been adopted, the fact that "justified punishment" would not always be justified would have been overlooked in most studies. The punishment is justified if and only if both the punisher and the observer consider the recipient to be bad. If the observer considers the recipient to be good, the punishment is deemed unjustified. The issue has not been resolved yet in the evolution of cooperation based on indirect reciprocity, so the social dilemma problem has not been completely solved. Models of private assessment that addressed the punishment dilemma began to be reported a decade ago [42,67,82–88].

Analyses using these private assessment models clarified that several of the leading eight norms, which are considered to be cooperative norms in a public assessment scheme, cannot maintain stable

cooperation due to the punishment dilemma. In particular, the breakdown of the stern-judging norm[3] has been demonstrated in several studies [42,55,56,67,89,90]. In addition, the required infinite system of simultaneous equations makes the analysis quite difficult in the theoretical analysis of a private assessment scheme[4]. Okada et al. developed a method for deriving an analytical solution without any approximation that is based on the assumption of solitary observation [42]. An exhaustive analysis of a private assessment scheme [90] has clarified the features of norms that resolve all three dilemmas: Social, scoring, and punishment as shown in Figure 4.
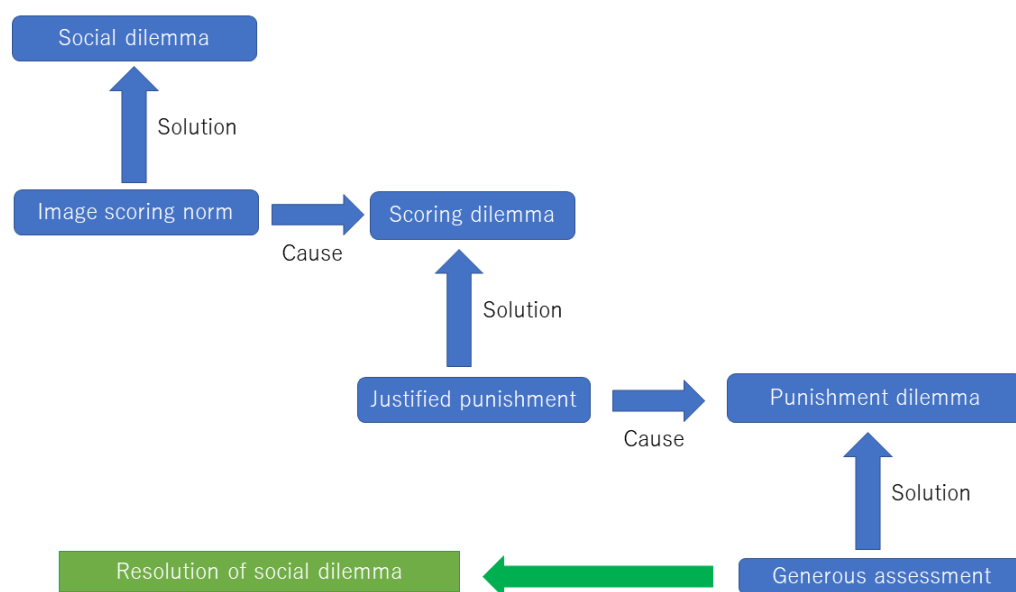
Figure 4. Resolving social dilemmas using indirect reciprocity. This figure is based on Figure 2 of Okada (2020) [90].

7. Remaining Issues in Study of Indirect Reciprocity Using Evolutionary Game Theory

There remain several important issues in the study of indirect reciprocity using evolutionary game theory. Generally, the greater the benefit cost ratio ($= r = b/c$) for cooperation, the greater the maintenance of cooperative regimes [91].Whitaker et al. (2016) [92] analyzed the relationship between the ratio and the emergence of cooperation and found that norms using second-order information such as the simple-standing and stern-judging norms require r to be around or greater than 1.2, whereas the image-scoring norm (using only first-order information) requires r to be around or greater than 5. Furthermore, analyses of the effects of errors have shown that the higher the error rate, the lower the level of cooperation [49,58,87].

According to the findings of studies that refined the assessment function, several norms maintain stable cooperation, but how the order of usage emerges in a situation with a mixture of norms remained unclear due to complexity of the required analytical methods. Yamamoto et al. [93] succeeded in revealing the role of the norms by comparing the state of activating a specific norm and the state of deactivating it by developing a norm knockout method. Uchida et al. [94] analyzed the norm

---

3   Of the leading eight norms, two (simple-standing and stern-judging) use up to second-order information and the other six use up to third-order information. Of the first two, only the stern-judging norm breaks cooperative regimes in a private assessment scheme.

4   This is because the definition of the conjunctive probability of v players whose images of a specified player are the same needs the conjunctive probability of $v + 1$ players. Therefore, the definition of the conjunctive probability of two players whose images of a specified player are the same requires an infinite system of simultaneous equations when the number of game observers is infinite.

ecosystem using equation systems representing the probabilities that a player adopting any norm has a good image of another player adopting any norm. Gaudeul et al. [95] fully characterized the evolutionary stable equilibria and analyzed their comparative statistics with respect to the benefit cost ratio.

Many studies have explored other mechanisms for establishing indirect reciprocity. One approach is to use a two-stage model in which punishment is administered separately from non-cooperative action; that is, the decision as to whether to administer punishment is made after playing a social dilemma game [96–98]. This approach to punishment has taken several branches. As well as peer punishment [99–103], researchers investigated methods for proactively collecting the costs of punishment [49] that overcome the evolutionary instability of peer punishment. Other extensions include institutional and pool punishment systems [104–110].

Separating the punishment system from the game opened the door to generalization of punishment to rewards and incentives [111]. Theoretical models of reward for cooperative behavior in contrast with punishment for non-cooperative behavior have been analyzed [100,112–117]. In addition, both anti-social punishment (punishment for cooperative behavior) [118,119] and anti-social reward (reward for non-cooperative behavior) [120] have been theoretically analyzed. However, Li et al. [121] argued that costly punishment has little effect on cooperation despite many the theoretical analyses that have shown that it does. This topic needs further discussion. The effects of incentive systems on the evolution of cooperation have been widely focused upon. A framework for introducing rewards and punishment is being actively researched in the field of social psychology [122–127]. The combination of punishment with indirect reciprocity has also been analyzed. Jordan et al. [128] conducted a game-theoretical analysis and their experiments showed that third-party punishment, i.e., punishment for behavior by unaffected observers, has positive effects on cooperation. Several groups have considered punishment to be a social norm and have explored conditions under which punishment emerges and its functions [122,124,129].

Another mechanism for establishing indirect reciprocity is to selectively exclude the second-order free riders from the population [46]. While indirect reciprocity drives selective cooperation, social exclusion is an alternative means [130]. A combination of reputation and ostracism supporting cooperation in groups of different sizes has been analyzed [131].

Clarifying effects of interaction structures on cooperation is also an important task. Theoretical analysis favors a well-mixed population due to analytical simplicity. However, network theory has revealed that real social networks have several features. Studies of the effect of the interaction structures on the evolution of indirect reciprocity have produced many findings [64,132–136].

Other tasks include clarifying the effects of mutation [137], the effects of large groups [61,138,139], and the effects of incomplete information [41,87,138,140]. Several studies compared cooperation rates based on indirect reciprocity with that based on direct reciprocity [61,141–143].

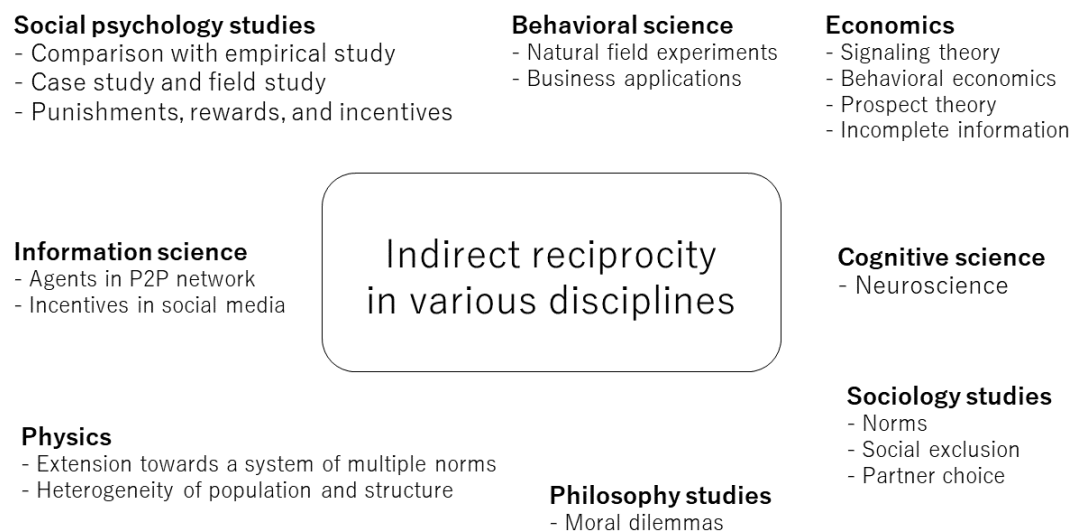## 8. Indirect Reciprocity in Various Disciplines

While we have focused on theoretical analysis of indirect reciprocity in game theory, its application has been widely studied in many disciplines. Figure 5 shows a bird's eye view of these disciplines.

In social psychology, several studies have revealed that the findings of theoretical research are inconsistent with those of experimental research. Milinski et al. (2001) [144] compared the image scoring norm, which takes into account only action information (first-order information), with a norm in which the reputation information (second-order information) of the recipient is also taken into account[5]. They showed that people prefer decision-making based on image scoring and thus concluded

---

[5] In this paper, we use the terms 'first-order' and 'second-order' in two different contexts: 'Free riders' and 'information used for rules assessment'. First-order free riders are players who do not cooperate while second-order free riders are players who do not punish first-order free riders. First-order information refers to donor action while second-order information refers to recipient reputation.

that people do not make decisions as complexly as theory suggests. With this as a starting point, theorists and experimentalists have argued about which domain of the assessment function should be adopted as the norm for the indirect reciprocity mechanism. Swakman et al. [68] presented results showing that people may actually take into account second-order information even of they must pay a cost for doing so. Okada et al. [69] demonstrated that there is a decision-making method in which second-order information is taken into account before first-order information. The debate continues.

Social psychology studies
- Comparison with empirical study
- Case study and field study
- Punishments, rewards, and incentives

Behavioral science
- Natural field experiments
- Business applications

Economics
- Signaling theory
- Behavioral economics
- Prospect theory
- Incomplete information

Information science
- Agents in P2P network
- Incentives in social media

Indirect reciprocity
in various disciplines

Cognitive science
- Neuroscience

Sociology studies
- Norms
- Social exclusion
- Partner choice

Physics
- Extension towards a system of multiple norms
- Heterogeneity of population and structure

Philosophy studies
- Moral dilemmas

Figure 5. Study of indirect reciprocity in various disciplines.

The origin of indirect reciprocity has also been studied in social psychology. A field study investigated the development of indirect reciprocity in 18-24-month-old toddlers and in infants aged around six months [145]. The results showed that toddlers reciprocated prosocial behavior indirectly. The behavior of adolescence has also been observed. In upstream indirect reciprocity, receiving an equal (vs. unequal) distribution led adolescents to become fairer to a third person. In downstream indirect reciprocity, older adolescents were more likely to devote their own resources to enforce fairness norms than younger adolescents [146].

Natural field experiments in behavioral science performed over the past decade have explored evidence of indirect reciprocity. Studies include investigation of indirect reciprocity and charitable work in a hair salon [147], the power of indirect reciprocity to promote prosocial behavior in one-shot interactions among drivers [148], large-scale cooperation in real world settings [149], and whether a service request is more likely to be rewarded for those with a profile history of offering the service (to third parties) in an online environment where members can repeatedly ask for and offer services to each other, free of charge [150].

In economics, several studies investigated the effects of conforming behavior [129] while others have investigated the connection to signaling theory [151,152]. A theoretical study used the mechanisms of indirect reciprocity to explain cognitive distortion predicted by prospect theory [153].

In sociology, the effect of indirect reciprocity on partner choice has been analyzed. Since indirect reciprocity separates a group of people into good people and bad people who cooperate selectively, it is a natural extension to play games with only good friends. One study examined game playing in combination with partner choice [54,154]. This direction led to a study of the connection of indirect reciprocity with group selection and with in-group favouritism [155]. A simulation analysis demonstrated the results of introducing an indirect reciprocity mechanism into partner selection in a group [156,157].

Application of indirect reciprocity to information science and the Internet is a hot topic. By modeling an agent's identity using traits, that can be shared with other agents, Bedewi et al. presented a basis for agents to change their identity [158]. They expect that indirect reciprocity will be effective for future technology and autonomous machines that need to function as a coalition. Tian et al. focused on how spatial reciprocity aids indirect reciprocity in sustaining cooperation in practical P2P (person-to-person) environments [159]. Their simulations showed that an incentive mechanism enhances cooperation better among structured peers than among unstructured peers. Toriumi et al. used simulation and the indirect reciprocity framework to investigate the relationships between social media posts and their responses[160]. Their findings would be useful for designing incentive systems in social media. Wang et al. used simulation of indirect group formation in a sparse network to demonstrate the group formation process [161].

Many disciplines have explored indirect reciprocity. In behavioral science, a statistical analysis of privately owned firms in China showed that philanthropic giving initiates and amplifies indirect reciprocity between visiting officials and local businesses, thereby increasing corporate investment [162]. In philosophy, indirect reciprocity has been explored in relation to moral dilemmas since indirect reciprocity research deals with the evolution of norms [63,163]. In cognitive science, the relationship between neuroscience and indirect reciprocity has been examined. One study used neuroimaging experiments to reveal the functional and anatomical neural bases of indirect reciprocity [74].

9. Conclusions

We have overviewed theoretical studies on indirect reciprocity. The history of these studies shows how indirect reciprocity came to be established as an academic field [6,23]. Moreover, we introduced many studies that refined the assessment function in an effort to enable a cooperative regime to be maintained. Academic efforts in this field have identified the features of assessment functions that resolve the scoring dilemma and the punishment dilemma that emerged simultaneously from the social dilemma [90]. While theoretical studies on indirect reciprocity have been ongoing, studies in many related disciplines have also investigated the mechanism of indirect reciprocity.

While indirect reciprocity requires the diffusion of reputation information, private assessment schemes that relax this requirement have been actively studied in recent years. In reality, selective cooperation is considered to be achievable through proper use of reputation information obtained through public assessment and impressions obtained through private assessment. Extension along these lines has begun [56], and future developments are expected.

The idea of selective cooperation as an explanation of how cooperation is rationalized in order to resolve social dilemmas is based on the concepts of direct reciprocity, indirect reciprocity, and network reciprocity. It comes from a kin selection framework, which is a common ancestor of these concepts. Therefore, they cannot be applied to public goods games as they are. This is because, in public goods games, the actions of individual players are equally distributed as public goods, so selective cooperation is impossible. Future extensions may bring a breakthrough that resolves this dilemma.

References

1. Hamilton, W.D. The evolution of altruistic behavior. *Am. Nat.* **1963**, *97*, 354–356. [CrossRef]
2. Trivers, R. The evolution of reciprocal altruism. *Q. Rev. Biol.* **1971**, *46*, 35–57. [CrossRef]
3. Axelrod, R.; Hamilton, W.D. The evolution of cooperation. *Science* **1981**, *211*, 1390–1396. [CrossRef] [PubMed]
4. Milinski, M. Tit for tat in sticklebacks and the evolution of cooperation. *Nature* **1987**, *325*, 433–435. [CrossRef] [PubMed]
5. Riolo, R.L.; Cohen, M.D.; Axelrod, R. Evolution of cooperation without reciprocity. *Nature* **2001**, *414*, 441–443. [CrossRef] [PubMed]
6. Nowak, M.A. Five rules for the evolution of cooperation. *Science* **2006**, *314*, 1560–1563. [CrossRef]
7. Traulsen, A.; Nowak, M.A. Evolution of cooperation by multilevel selection. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10952–10955. [CrossRef]
8. Dawes, R.M. Social Dilemmas. *Ann. Rev. Psychol.* **1980**, *31*, 169–193. [CrossRef]
9. Yamagishi, T. The provision of a sanctioning system as a public good. *J. Pers. Soc. Psychol.* **1986**, *51*, 110–116. [CrossRef]
10. Yamagishi, T.; Cook, K.S. Generalized exchange and social dilemmas. *Soc. Psychol. Q.* **1993**, *56*, 235–248. [CrossRef]
11. Macy, M.W.; Flache, A. Learning dynamics in social dilemmas. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7229–7236. [CrossRef] [PubMed]
12. Balliet, D.P.; Parks, C.D.; Joireman, J. Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Process. Interg.* **2009**, *12*, 533–547. [CrossRef]
13. Van Lange, P.A.M.; Joireman, J.; Parks, C.D.; Van Dijk, E. The psychology of social dilemmas: A review. *Organ. Behav. Hum. Dec.* **2013**, *120*, 125–141. [CrossRef]
14. Eisenberg–Berg, N. Development of children's prosocial moral judgment. *Dev. Psychol.* **1979**, *15*, 128–137. [CrossRef]
15. Alexander, R.D. The search for an evolutionary philosophy of man. *Proc. R. Soc. Vic.* **1971**, *84*, 99–120.
16. Alexander, R.D. The search for a general theory of behavior. *Behav. Sci.* **1975**, *20*, 77–100. [CrossRef]
17. Alexander, R.D. *The Biology of Moral Systems*; Aldine De Gruyter: New York, NY, USA, 1987.
18. Maynard Smith, J. The theory of games and the evolution of animal conflicts. *J. Theor. Biol.* **1974**, *47*, 209–221. [CrossRef]
19. Maynard Smith, J. *Evolution and the Theory of Games*; Cambridge University Press: Cambridge, UK, 1982.
20. Ohtsuki, H.; Nowakm, M.A. Direct reciprocity on graphs. *J. Theor. Biol.* **2007**, *247*, 462–470. [CrossRef]
21. Delton, A.W.; Krasnow, M.M.; Cosmides, L.; Tooby, J. Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13335–13340. [CrossRef]
22. Boyd, R.; Richerson, R.J. The evolution of indirect reciprocity. *Soc. Netw.* **1989**, *11*, 213–236. [CrossRef]
23. Nowak, M.A.; Sigmund, K. Evolution of indirect reciprocity. *Nature* **2005**, *437*, 1291–1298. [CrossRef] [PubMed]
24. Sigmund, K. *The Calculus of Selfishness*; Princeton University Press: Princeton, NJ, USA, 2010.
25. Rutte, C. Taborsky, Generalized reciprocity in rats. *PLoS Biol.* **2007**, *5*, e196. [CrossRef] [PubMed]
26. Engelmann, D.; Fischbauer, U. Indirect reciprocity and strategic reputation building in an experimental helping game. *Games Econ. Behav.* **2009**, *67*, 399–407. [CrossRef]
27. Horita, Y.; Takezawa, M.; Kinjo, T.; Nakawake, Y.; Masuda, N. Transient nature of cooperation by pay-it-forward reciprocity. *Sci. Rep.* **2016**, *6*, 19471. [CrossRef] [PubMed]
28. Pfeiffer, T.; Rutte, C.; Killingback, T.; Taborsky, M.; Bonhoeffer, S. Evolution of cooperation by generalized reciprocity. *Pract. R. Soc. B* **2005**, *272* 1115–1120. [CrossRef] [PubMed]
29. Bshary, R.; Grutter, A.S. Punishment and partner switching causes cooperative behavior in a cleaning mutualism. *Biol. Lett.* **2005**, *1*, 396–399. [CrossRef]
30. Nowak, M.A.; Roch, S. Upstream reciprocity and the evolution of gratitude. *Pract. R. Soc. B* **2007**, *274*, 605–609. [CrossRef]
31. Nowak, M.A.; Sigmund, K. Evolution of indirect reciprocity by image scoring. *Nature* **1998**, *393*, 573–577. [CrossRef]
32. Milinski, M.; Semmann, D.; Krambeck, H.J. Reputation helps solve the 'tragedy of the commons'. *Nature* **2002**, *415*, 424–426. [CrossRef]
33. Sommerfeld, R.D.; Krambeck, H.J.; Semmann, D.; Milinski, M. Gossip as an alternative for direct observation in games of indirect reciprocity. *Proc. Natl. Acad. Sci. USA* **2007**, *44*, 17435–17440. [CrossRef]

34.   Melis, A.P.; Semmann, D. How is human cooperation different? *Philos. Tans. R. Soc. B* **2010**, *365*, 2663–2674. [CrossRef] [PubMed]

35.   Rand, D.G.; Nowak, M.A. Human cooperation. *Trends Cogn. Sci.* **2012**, *117*, 413–425. [CrossRef] [PubMed]

36.   Nowak, M.A.; Sigmund, K. The dynamics of indirect reciprocity. *J. Theor. Biol.* **1998**, *194*, 561–574. [CrossRef] [PubMed]

37.   Hofbauer, J.; Sigmund, K. *Evolutionary Games and Population Dynamics*; Cambridge University Press: Cambridge, UK, 1998.

38.   Brandt, H.; Sigmund, K. The logic of reprobation: Action and assessment rules in indirect reciprocity. *J. Theor. Biol.* **2004**, *231*, 475–486. [CrossRef] [PubMed]

39.   Brandt, H.; Sigmund, K. Indirect reciprocity, image scoring, and moral hazard. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2666–2670. [CrossRef]

40.   Brandt, H.; Sigmund, K. The good, the bad and the discriminator ? Errors in direct and indirect reciprocity. *J. Theor. Biol.* **2006**, *239*, 183–194. [CrossRef] [PubMed]

41.   Ohtsuki, H.; Iwasa, Y.; Nowak, M.A. Reputation effects in public and private interactions. *PLoS Comput. Biol.* **2015**, *11*, e1004527. [CrossRef]

42.   Okada, I.; Sasaki, T.; Nakai, Y. A solution of private assessment in indirect reciprocity using solitary observation. *J. Theor. Biol.* **2018**, *455*, 7–15. [CrossRef]

43.   Wedekind, C.; Milinski, M. Cooperation through image scoring in humans. *Science* **2000**, *288*, 850–852. [CrossRef]

44.   Leimar, O.; Hammerstein, P. Evolution of cooperation through indirect reciprocity. *Proc. Natl. Acad. Sci. USA* **2001**, *268*, 745–753. [CrossRef]

45.   Berger, U. Learning to cooperate via indirect reciprocity. *Games Econ. Behav.* **2011**, *72*, 30–37. [CrossRef]

46.   Panchanathan, K.; Boyd, R. Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* **2004**, *432*, 499–502. [CrossRef] [PubMed]

47.   Helbing, D.; Szolnoki, A.; Perc, M.; Szabó, G. Evolutionary establishment of moral and double moral standards through spatial interactions. *PLoS Comput. Biol.* **2010**, *6*, e1000758. [CrossRef] [PubMed]

48.   Hilbe, C.; Traulsen, A. Emergence of responsible sanctions without second order free riders, antisocial punishment or spite. *Sci. Rep.* **2012**, *2*, 458. [CrossRef] [PubMed]

49.   Sasaki, T.; Okada, I.; Nakai Y. Indirect reciprocity can overcome free-rider problems on costly moral assessment. *Biol. Lett.* **2016**, *12*, 20160341. [CrossRef] [PubMed]

50.   Yamamoto, H.; Okada, I. How to keep punishment to maintain cooperation: Introducing social vaccine. *Physica A* **2016**, *443*, 526–536. [CrossRef]

51.   Ozono, H.; Kamijo, Y.; Shimizu, K. Punishing second-order free riders before first-order free riders: The effect of pool punishment priority on cooperation. *Sci. Rep.* **2017**, *7*, 14379. [CrossRef]

52.   Szolnoki, A.; Perc, M. Second-order free-riding on antisocial punishment restores the effectiveness of prosocial punishment. *Phys. Rev. X* **2017**, *7*, 041027. [CrossRef]

53.   Weber, T.O.; Weisel, O.; Gächter, S. Dispositional free riders do not free ride on punishment. *Nat. Commun.* **2018**, *9*, 2390. [CrossRef]

54.   Roberts, G. Partner choice drives the evolution of cooperation via indirect reciprocity. *PLoS ONE* **2015**, *10*, e0129442. [CrossRef]

55.   Sasaki, T.; Okada, I.; Nakai, Y. The evolution of conditional moral assessment in indirect reciprocity. *Sci. Rep.* **2017**, *7*, 41870. [CrossRef] [PubMed]

56.   Okada, I.; Yamamoto, H.; Uchida, S. Hybrid assessment scheme based on the stern-Judging rule for maintaining cooperation under indirect reciprocity. *Games* **2020**, *11*, 13. [CrossRef]

57.   Lotem, A.; Fishman, M.A.; Stone, L. Evolution of cooperation between individuals. *Nature* **1999**, *400*, 226–227. [CrossRef] [PubMed]

58.   Panchanathan, K.; Boyd, R. A tale of two defectors: The importance of standing for evolution of indirect reciprocity. *J. Theor. Biol.* **2003**, *224*, 115–126. [CrossRef]

59.   Tanabe, S.; Suzuki, H.; Masuda, N. Indirect Reciprocity With Trinary Reputations. *J. Theor. Biol.* **2013**, *317*, 338–347. [CrossRef] [PubMed]

60.   Berger, U.; Grune, A. On the stability of cooperation under indirect reciprocity with first-order information. *Games Econ. Behav.* **2016**, *98*, 19–33. [CrossRef]

61. Suzuki, S.; Akiyama, E. Evolution of indirect reciprocity in groups of various sizes and comparison with direct reciprocity. *J. Theor. Biol.* **2007**, *245*, 539–552. [CrossRef]

62. Ohtsuki, H.; Iwasa, Y. How should we define goodness? reputation dynamics in indirect reciprocity. *J. Theor. Biol.* **2004**, *231*, 107–120. [CrossRef]

63. Santos, F.P.; Santos, F.C.; Pacheco, J.M. Social norm complexity and past reputations in the evolution of cooperation. *Nature* **2018**, *555*, 242–245. [CrossRef]

64. Ohtsuki, H.; Iwasa, Y. The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *J. Theor. Biol.* **2006**, *239*, 435–444. [CrossRef]

65. Ohtsuki, H.; Iwasa, Y. Global analyses of evolutionary dynamics and exhaustive search for social norms that maintain cooperation by reputation. *J. Theor. Biol.* **2007**, *244*, 518–531. [CrossRef] [PubMed]

66. Suzuki, S.; Kimura, H. Indirect reciprocity is sensitive to costs of information transfer. *Sci. Rep.* **2013**, *3*, 1435. [CrossRef]

67. Okada, I.; Sasaki, T.; Nakai, Y. Tolerant indirect reciprocity can boost social welfare through solidarity with unconditional cooperators in private monitoring. *Sci. Rep.* **2017**, *7*, 9737. [CrossRef] [PubMed]

68. Swakman, V.; Molleman, L.; Ule, A.; Egas, M. Reputation-based cooperation: Empirical evidence for behavioral strategies. *Evol. Hum. Behav.* **2016**, *37*, 230–235. [CrossRef]

69. Okada, I.; Yamamoto, H.; Sato, Y.; Uchida, S.; Sasaki, T. Experimental evidence of selective inattention in reputation-based cooperation. *Sci. Rep.* **2018**, *8*, 14813. [CrossRef] [PubMed]

70. Tanaka, H.; Ohtsuki, H.; Ohtsubo, Y. The price of being seen to be just: An intention signalling strategy for indirect reciprocity. *Proc. R. Soc. B* **2016**, *283*, 20160694. [CrossRef] [PubMed]

71. Chalub, F.; Santos, F.C.; Pacheco, J.M. The evolution of norms. *J. Theor. Biol.* **2006**, *241*, 233–240. [CrossRef] [PubMed]

72. Sugden, R. *The Economics of Rights, Cooperation and Welfare*; Basil Blackwell: Oxford, UK, 1986.

73. Panchanathan, K. Two wrongs don't make a right: The initial viability of different assessment rules in the evolution of indirect reciprocity. *J. Theor. Biol.* **2011**, *277*, 48–54. [CrossRef]

74. Watanabe, T.; Takezawa, M.; Nakawake, Y.; Kunimatsu, A.; Yamasue, H.; Nakamura, M.; Miyashita, Y.; Masuda, N. Two distinct neural mechanisms underlying indirect reciprocity. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3990–3995. [CrossRef]

75. Uchida, S.; Sigmund, K. The competition of assessment rules for indirect reciprocity. *J. Theor. Biol.* **2010**, *263*, 13–19. [CrossRef]

76. Seinen, I.; Schram, A. Social status and group norms: Indirect reciprocity in a repeated helping experiment. *Eur. Econ. Rev.* **2006**, *50*, 581–602. [CrossRef]

77. Pacheco, J.M.; Santos, F.C.; Chalub, F.A.C.C. Stern-judging: A simple, successful norm which promotes cooperation under indirect reciprocity. *PLoS Comput. Biol.* **2006**, *2*, 1634–1638. [CrossRef] [PubMed]

78. Santos, F.P.; Pacheco, J.M.; Santos, F.C. Evolution of cooperation under indirect reciprocity and arbitrary exploration rates. *Sci. Rep.* **2016**, *6*, 37517. [CrossRef] [PubMed]

79. Inaba, M.; Takahashi, N. Linkage Based on the Kandori Norm Successfully Sustains Cooperation in Social Dilemmas. *Games* **2019**, *10*, 1–15. [CrossRef]

80. Kandori, M. Social norms and community enforcement. *Rev. Econ. Stud.* **1992**, *59*, 63–80. [CrossRef]

81. Takahashi, N.; Mashima, R. The importance of subjectivity in perceptual errors on the emergence of indirect reciprocity. *J. Theor. Biol.* **2006**, *243*, 418–436. [CrossRef] [PubMed]

82. Uchida, S. Effect of private information on indirect reciprocity. *Phy. Rev. E* **2010**, *82*, 036111. [CrossRef] [PubMed]

83. Sigmund, K. Moral assessment in indirect reciprocity. *J. Theor. Biol.* **2012**, *299*, 25–30. [CrossRef]

84. Martinez-Vaquero, L.A.; Cuesta, J.A. Evolutionary stability and resistance to cheating in an indirect reciprocity model based on reputation. *Phys. Rev. E* **2013**, *87*, 052810. [CrossRef] [PubMed]

85. Uchida, S.; Sasaki, T. Effect of assessment error and private information on stern-judging in indirect reciprocity. *Chaos Solitons Fract.* **2013**, *56*, 175–180. [CrossRef]

86. Olejarz, J.; Ghang, W.; Nowak, M.A. Indirect reciprocity with optional interactions and private information. *Games* **2015**, *6*, 438–457. [CrossRef]

87. Hilbe, C.; Schmid, L.; Tkadlec, J.; Chatterjee, K.; Nowak, M.A. Indirect reciprocity with private, noisy, and incomplete information. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 12241–12246. [CrossRef] [PubMed]

88. Clark, D.; Fudenberg, D.; Wolitzky, A. Indirect reciprocity with simple records. *Proc. Natl. Acad. Sci. USA* **2020**. [CrossRef] [PubMed]

89. Okada, I. Computational social science on adaptive norms in social dilemmas: Integrating theory, experiments, and simulations. *Socio-Informatics* **2019**, *8*, 19–33.

90. Okada, I. Two ways to overcome the three social dilemmas of indirect reciprocity. in review.

91. McNamara, J.M.; Doodson, P. Reputation can enhance or suppress cooperation through positive feedback. *Nat. Commun.* **2015**, *6*, 6134. [CrossRef]

92. Whitaker, R.M.; Colombo, G.B.; Allen, S.M.; Dunbar, R.I.; A dominant social comparison heuristic unites alternative mechanisms for the evolution of indirect reciprocity. *Sci. Rep.* **2016**, *6*, 1–12. [CrossRef]

93. 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]

94. 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]

95. Gaudeul, A.; Keser, C.; Muller, S. *The Evolution of Morals under Indirect Reciprocity*; Center for European, Governance and Economic Development Research: Wilhelmsplatz, Germany, 2019; p. 370.

96. Boyd, R.; Richerson, P. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethol. Sociobiol.* **1992**, *13*, 171–195. [CrossRef]

97. Rockenbach, B.; Milinski, M. The efficient interaction of indirect reciprocity and costly punishment. *Nature* **2006**, *444*, 718–723. [CrossRef]

98. Ohtsuki, H.; Iwasa, Y.; Nowak, M.A. Indirect reciprocity provides only a narrow margin of efficiency for costly punishment. *Nature* **2009**, *457*, 79–82. [CrossRef] [PubMed]

99. Fehr, E.; Gächter, S. Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **2000**, *90*, 980–994. [CrossRef]

100. 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]

101. Sasaki, T.; Okada, I.; Uchida, S.; Chen, X. Commitment to cooperation and peer punishment: Its evolution. *Games* **2015**, *6*, 574–587. [CrossRef]

102. Diekmann, A.; Przepiorka, W. Punitive preferences, monetary incentives and tacit coordination in the punishment of defectors promote cooperation in humans. *Sci. Rep.* **2015**, *5*, 10321. [CrossRef]

103. Grimalda, G.; Pondorfer, A.; Tracer, D.P. Social image concerns promote cooperation more than altruistic punishment. *Nat. Commun.* **2016**, *7*, 12288. [CrossRef]

104. Sigmund, K.; De Silva, H.; Traulsen, A.; Hauert, C. Social learning promotes institutions for governing the commons. *Nature* **2010**, *466*, 861–863. [CrossRef]

105. Szolnoki, A.; Szabó, G.; Perc, M. Phase diagrams for the spatial public goods game with pool punishment. *Phys. Rev. E* **2011**, *83*, 036101. [CrossRef]

106. Perc, M. Sustainable institutionalized punishment requires elimination of second-order free-riders. *Sci. Rep.* **2012**, *2*, 344. [CrossRef]

107. Traulsen, A.; Röhl, T.; Milinski, M. An economic experiment reveals that humans prefer pool punishment to maintain the commons. *Proc. Natl. Acad. Sci. USA* **2012**, *279*, 3716–3721. [CrossRef]

108. 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]

109. 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]

110. 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]

111. Balliet, D.; Mulder, L.B.; Van Lange, P.A.M. Reward, punishment, and cooperation: A meta-analysis. *Psychol. Bull.* **2011**, *137*, 594–615. [CrossRef] [PubMed]

112. Andreoni, J.; Harbaugh, W.; Versterlund, L. The carrot or the stick; Rewards, punishments, and cooperation. *Am. Econ. Rev.* **2003**, *93*, 893–902. [CrossRef]

113. Kendal, J.; Feldman, M.W.; Aoki, K. Cultural coevolution of norm adoption and enforcement when punishers are rewarded or non-punishers are punished. *Theor. Popul. Biol.* **2006**, *70*, 10–25. [CrossRef] [PubMed]

114. Sefton, M.; Shupp, R.; Walker, J.M. The effect of rewards and sanctions in provision of public goods. *Econ. Inq.* **2007**, *45*, 671–690. [CrossRef]

115. Sutter, M.; Haigner, S.; Kocher, M.G. Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Rev. Econ. Stud.* **2010**, *77*, 1540–1566. [CrossRef]

116. Hilbe, C.; Sigmund, K. Incentives and opportunism: From the carrot to the stick. *Pract. R. Soc. B* **2010**, *277*, 2427–2433. [CrossRef]

117. 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]

118. Herrmann, B.; Thoni, C.; Gächter, S. Antisocial Punishment Across Societies. *Science* **2008**, *319*, 1362–1367. [CrossRef] [PubMed]

119. Rand, D.G.; Nowak, M.A. The evolution of anti-social punishment in optional public goods games. *Nat. Commun.* **2011**, *2* 434. [CrossRef] [PubMed]

120. dos Santos, M.; Peña, J. Antisocial rewarding in structured populations. *Sci. Rep.* **2017**, *7*, 6212. [CrossRef] [PubMed]

121. Li, X.; Jusup, M.; Wang, Z.; Li, H.; Shi, L.; Podobnik, B.; Stanley, H.E.; Havlin, S.; Boccaletti, S. Punishment diminishes the benefits of network reciprocity in social dilemma experiments. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 30–35. [CrossRef]

122. Fehr, E.; Gächter, S. Altruistic punishment in humans. *Nature* **2002**, *415*, 137–140. [CrossRef]

123. Gürerk, Ö.; Irlenbusch, B.; Rockenbach, B. The competitive advantage of sanctioning institutions. *Science* **2006**, *312*, 108–111. [CrossRef]

124. Dreber, A.; Rand, D.G.; Fudenberg, D.; Nowak, M.A. Winners don't punish. *Nature* **2008**, *452*, 348–351. [CrossRef]

125. Egas, M.; Riedl, A. The economics of altruistic punishment and the maintenance of cooperation. *Proc. R. Soc. B* **2008**, *275*, 871–878. [CrossRef]

126. Wu, J.J.; Zhang, B.Y.; Zhou, Z.X.; He, Q.Q.; Zheng, X.D.; Cressman, R.; Tao, Y. Costly punishment does not always increase cooperation. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17448–17451. [CrossRef]

127. Milinski, M.; Rockenback, B. On the interaction of the stick and the carrot in social dilemmas. *J. Theor. Biol.* **2012**, *299*, 139–143. [CrossRef]

128. Jordan, J.J.; Hoffman, M.; Bloom, P.; Rand, D.G. Third-party punishment as a costly signal of trustworthiness. *Nature* **2016**, *530*, 473–476. [CrossRef] [PubMed]

129. Henrich, J.; Boyd, R. Why people punish defectors? Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *J. Theor. Biol.* **2001**, *208*, 79–89. [CrossRef] [PubMed]

130. Sasaki, T.; Uchida, S. The evolution of cooperation by social exclusion. *Proc. R. Soc. B* **2013**, *280*, 20122498. [CrossRef] [PubMed]

131. Giardini, F.; Vilone, D. Evolution of gossip-based indirect reciprocity on a bipartite network. *Sci. Rep.* **2016**, *6*, 37931. [CrossRef] [PubMed]

132. Hauert, C.; Doebeli, M. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* **2004**, *428*, 643–646. [CrossRef] [PubMed]

133. Heinrich, H.N.; Perc, M.; Szolnoki, A.; Helbing, D. Stability of cooperation under image scoring in group interactions. *Sci. Rep.* **2015**, *5*, 12145.

134. Ghang, W.; Nowak, M.A. Indirect reciprocity with optional interactions. *J. Theor. Biol.* **2015**, *365*, 1–11. [CrossRef]

135. Sasaki, T.; Yamamoto, H.; Okada, I.; Uchida, S. The evolution of reputation-based cooperation in regular networks. *Games* **2017**, *8*, 8. [CrossRef]

136. Gong, Y.; Liu, S.; Bai, Y. Reputation-based co-evolutionary model promotes cooperation in prisoner's dilemma game. *Phys. Lett. A* **2020**, *384*, 126233. [CrossRef]

137. Sethi, R.; Somanathan, E. The evolution of social norms in common property resource use. *Am. Econ. Rev.* **1996**, *86*, 766–788.

138. Masuda, N.; Ohtsuki, H. Tag-based indirect reciprocity by incomplete social information. *Proc. R. Soc. B* **2007**, *274*, 689–695. [CrossRef] [PubMed]

139. Suzuki, S.; Akiyama, E. Three-person game facilitates indirect reciprocity under image scoring. *J. Theor. Biol.* **2007**, *249*, 93–100. [CrossRef] [PubMed]

140. Fishman, M.A. Indirect reciprocity among imperfect. individuals. *J. Theor. Biol.* **2003**, *225*, 285–292. [CrossRef]

141. Pollock, G.B.; Dugatkin, L.A. Reciprocity and the evolution of reputation. *J. Theor. Biol.* **1992**, *159*, 25–37. [CrossRef]

142. Dufwenberg, M.; Gneezy, U.; Güth, W.; van Damme, E. Direct vs. indirect reciprocation: An experiment. *Homo Oeconomicus* **2001**, *18*, 19–30.

143. Roberts, G. Evoluation of direct and indirect reciprocity. *Pract. R. Soc. B* **2008**, *275*, 173–179. [CrossRef]

144. Milinski, M.; Semmann, D.; Bakker, T.C.M,Krambeck, H.J. Cooperation through indirect reciprocity: Image scoring or standing strategy? *Pract. R. Soc. B* **2001**, *268*, 2495–2501. [CrossRef]

145. Nava, E.; Croci, E.; Turati, C. 'I see you sharing, thus I share with you': Indirect reciprocity in toddlers but not infants. *Palgrave Commun.* **2019**, *5*, 66. [CrossRef]

146. Hu, Y.; Ma, J.; Luan, Z.; Dubas, J.S.; Xi, J. Adolescent indirect reciprocity: Evidence from incentivized economic paradigms. *J. Adolesc.* **2019**, *74*, 221–228. [CrossRef]

147. Khadjavi, M. Indirect reciprocity and charitable giving? Evidence from a field experiment. *Manag. Sci.* **2016**, *63*, 3708–3717. [CrossRef]

148. Mujcic, R.; Leibbrandt, A. Indirect reciprocity and prosocial behaviour: Evidence from a natural field experiment. *Econ. J.* **2017**, *128*, 1683–1699. [CrossRef]

149. Yoeli, E.; Hoffman, M.; Rand, D.G.; Nowak, M.A. Powering up with indirect reciprocity in a large-scale field experiment. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 10424–10429. [CrossRef]

150. van Apeldoorn, J.; Schram, A. Indirect reciprocity; a field experiment. *PLoS ONE* **2016**, *11*, e0152076. [CrossRef] [PubMed]

151. Roberts, G. Competitive altruism: From reciprocity to the handicap principle. *Proc. R. Soc. Lond. B* **1998**, *265*, 427–431. [CrossRef]

152. Gintis, H.; Smith, E.A.; Bowles, S. Costly signaling and cooperation. *J. Theor. Biol.* **2001**, *213*, 103–119. [CrossRef] [PubMed]

153. Uchida, S.; Yamamoto, H.; Okada, I.; Sasaki, T. Evolution of cooperation with peer punishment under prospect theory. *Games* **2019**, *10*, 11. [CrossRef]

154. Sylwester, K. Roberts, G.; Reputation-based partner choice is an effective alternative to indirect reciprocity in solving social dilemmas. *Evol. Hum. Behav.* **2013**, *34*, 201–206. [CrossRef]

155. Masuda, N. In-group favouritism and intergroup cooperation under indirect reciprocity based on group reputation. *J. Theor. Biol.* **2012**, *311*, 8–18. [CrossRef]

156. Nakai, Y.; Muto, M. Evolutionary simulation of peace with altruistic strategy for selected friends. *J. Socio-Inf. Stud.* **2005**, *9*, 59–71.

157. Nakai, Y.; Muto, M. Emergence and collapse of peace with friend selection strategies. *J. Artif. Soc. Simul.* **2008**, *11*, 6.

158. Bedewi, W.; Whitaker, R.M.; Colombo, G.B.; Allen, S.M.; Dunham, Y. The implications of shared identity on indirect reciprocity. *J. Inform. Telecommun.* **2020**, 1741858. [CrossRef]

159. Tian, L.L.; Li, M.C.; Wang, Z. Cooperation enhanced by indirect reciprocity in spatial prisoner's dilemma games for social P2P systems. *Physica A* **2016**, *462*, 1252–1260. [CrossRef]

160. Toriumi, F.; Yamamoto, H.; Okada, I. A belief in rewards accelerates cooperation on consumer-generated media. *J. Comput. Soc. Sci.* **2020**, *3*, 19–31. [CrossRef]

161. Wang, Z.; Szolnoki, A.; Perc, M. If players are sparse social dilemmas are too: Importance of percolation for evolution of cooperation. *Sci. Rep.* **2012**, *2*, 369. [CrossRef]

162. Jia, M.; Xiang, Y.; Zhang, Z. Indirect reciprocity and corporate philanthropic giving: How visiting officials influence investment in privately owned chinese firms. *J. Manag. Stud.* **2019**, *56*, 372–407. [CrossRef]

163. Nordin, A. Indirect reciprocity and reputation management in religious morality relating to concepts of supernatural agents. *J. Cogn. Sci. Relig.* **2015**, *3* 125–153. [CrossRef]