A Successful Strategy for Multichannel Iterated Prisoner's Dilemma

Zhen Wang[1,2,3], **Zhaoheng Cao**[1,3], **Juan Shi**[4], **Peican Zhu**[3], **Shuyue Hu**[5,*] and **Chen Chu**[6,3,*]
[1]School of Computer Science, Northwestern Polytechnical University
[2]School of Cybersecurity, Northwestern Polytechnical University
[3]School of Artificial Intelligence, OPtics and ElectroNics (iOPEN), Northwestern Polytechnical University
[4]School of Automation, Northwestern Polytechnical University
[5]Shanghai Artificial Intelligence Laboratory
[6]School of Statistics and Mathematics, Yunnan University of Finance and Economics
hushuyue@pjlab.org.cn, chuchenynufe@hotmail.com

Abstract

Iterated prisoner's dilemma (IPD) and its variants are fundamental models for understanding the evolution of cooperation in human society as well as AI systems. In this paper, we focus on multichannel IPD, and examine how an agent should behave to obtain generally high payoffs under this setting. We propose a novel strategy that chooses to cooperate or defect by considering the difference in the cumulative number of defections between two agents. We show that our proposed strategy is nice, retaliatory, and forgiving. Moreover, we analyze the performance of our proposed strategy across different scenarios, including the self-play settings with and without errors, as well as when facing various opponent strategies. In particular, we show that our proposed strategy is invincible and never loses to any opponent strategy in terms of the expected payoff. Last but not least, we empirically validate the evolutionary advantage of our strategy, and demonstrate its potential to serve as a catalyst for cooperation emergence.

1 Introduction

The Prisoner's Dilemma (PD) and its variants thereof are arguably the best-known concepts of game theory. They model the situations where agents can gain significant benefits from cooperation but each faces a temptation to free ride (or defect), and are fundamental models for understanding the evolution of cooperation. It is well-known that while mutual defection is the unique Nash equilibrium in a single play of PD, the folk theorems show that the *iterated* play of PD enables and sustains the emergence of cooperation [Mailath and Samuelson, 2006; Wang *et al.*, 2022; Chu *et al.*, 2022]. This naturally raises a question which has attracted much interest over the past decades—*what a successful strategy should be in the IPD* [Nowak and Sigmund, 1992; Press and Dyson, 2012; Wang and Lin, 2020; Baker, 2020; Zhao *et al.*, 2022].

*Corresponding Author

For a strategy to be successful, a typical criterion since the celebrated Axelord's tournaments is that it can result in *generally* high payoffs against other strategies [Axelrod and Hamilton, 1981]. Numerous strategies have been put forth towards this end (see reviews [Hilbe *et al.*, 2018] and references therein). However, on the downside, it has been shown that there is no universally optimal strategy in terms of evolutionary stability or robustness against indirect invasions [Selten and Hammerstein, 1984; Farrell and Maskin, 1989; Van Veelen, 2012; García and van Veelen, 2016]. That said, three principles have been identified to underlie various successful strategies [Miller, 1985; Lerer and Peysakhovich, 2017]: (i) be *nice*, i.e. never be the first to defect, (ii) be *retaliatory*, i.e. be able to retaliate when faced with an opponent's defection, and (iii) be *forgiving*, i.e. be able to resume cooperation after the opponent's defection.

This paper addresses the above challenge focusing on a notable recent development of the theory on IPD—the *multichannel* IPD. Donahue, Hauser, Nowak & Hilbe [2020] introduced this concept to extend IPD in order to better reflect real-world scenarios where agents can interact over multiple games (or channels) concurrently; for example, scientists collaborate on several projects, and negotiation agents buy/sell multiple products. In contrast to the traditional IPD, this multichannel extension enables agents to leverage their strategies in one channel to influence the outcomes in another. A timely example is the cooperation between Microsoft and OpenAI. While these two organizations can compete directly in multiple markets, Microsoft offers the data center infrastructure and cloud services to support OpenAI's research and development; in return, OpenAI's large language models GPT-series provide Microsoft with innovative tools that can be integrated into various applications, like Copilot and Office apps. Consequently, the multichannel IPD can increase agents' bargaining power to promote overall cooperation across multiple channels. However, as one can expect, such an interdependency across multiple channels also adds more complexity to the design of a successful strategy, which exacerbates the aforementioned challenge. In particular, as we shall show in our preliminary experiments (Figure 1), the two existing strategies for multichannel IPD, namely the multichannel-

Win-Stay-Lose-Shift (WSLS) and Cooperate if Coordinated (CIC) [Donahue *et al.*, 2020], are nice and forgiving though, they cannot sufficiently retaliate; specifically, they are vulnerable to long-lasting exploitation from the always defecting (ALLD) strategy, and can lose to other common strategies.

In light of this gap, this paper proposes a new, successful strategy for multichannel IPD, which is abbreviated as MCSUC for simplicity. Our proposed strategy tracks the cumulative number of times each agent has defected across all channels; then it responds with (i) full cooperation in all channels, (ii) partial cooperation in a single channel, or (iii) full defection in all channels. The rationale behind our strategy is as follows: if so far an agent's opponent has defected less often than the agent itself, the agent may choose to fully cooperate. On the contrary, if an agent's opponent has defected more frequently, surpassing certain tolerance thresholds, the agent may choose to defect partially in some channels or even completely in all channels. We show that MCSUC, embodying the concept of cumulative reciprocity [Li *et al.*, 2022], inherently adheres to the three principles of niceness, retaliation, and forgiveness.

To better understand the successfulness of MCSUC in the sense that whether it leads to generally high payoffs, we theoretically analyze its performance and stability under various scenarios. We show that if both agents adopt MCSUC (i.e. under the self-play setting), complete cooperation across all channels is achieved in the absence of any 'trembling hand' errors (Theorem 1); even when such errors are present, agents' expected cooperation rate and expected payoffs are nearly optimal (Proposition 1). We also show that when playing against the ALLD strategy, MCSUC readily retaliates and effectively avoids long-time exploitation (Theorem 2), as opposed to the existing strategies multichannel-WSLS and CIC. In contrast, when playing against the always cooperating (ALLC) strategy, MCSUC resists the temptation to exploit the opponent (Theorem 3). Additionally, we then show that under certain conditions, MCSUC is *fair* and is able to ensure the same expected payoff as that of *any* opponent strategy (Theorem 4). Perhaps most interestingly, we show that MCSUC is *invincible* in the sense that its resulting expected payoff is always equal to or greater than that of *any* opponent strategy, regardless of the presence of errors (Theorem 5 and Corollary 1). Put differently, this ensures that MCSUC will *never* lose to any opponent strategy and sometimes even outperform in terms of the expected payoff. Last but not least, we show that under the self-play setting, MCSUC is a Nash equilibrium strategy as well as a subgame perfect equilibrium strategy in the absence of errors (Theorem 6), and is an approximate Nash equilibrium strategy given a sufficiently small error rate (Theorem 7). In other words, MCSUC is *stable* under the self-play setting.

The above theoretical analyses consider games in which agents do not change their strategies. This kind of analysis is useful to explore a strategy's basic properties, but it does not take into account if agents have an incentive to adopt these strategies in the first place [Nowak and Sigmund, 1992]. To explore this latter question, we complement our theoretical analysis with two sets of evolutionary experiments: two-strategy populations as well as three-strategy populations, and let agents' strategies evolve. In each two-strategy population, we consider MCSUC and one of the ten strategies: multichannel-WSLS, CIC, ALLC, ALLD, Tit-for-Tat (TFT) [Axelrod and Hamilton, 1981], Generous Tit-for-Tat (GTFT) [Nowak and Sigmund, 1992], HardMajority [Mittal and Deb, 2009], the cumulative reciprocal strategy (CURE) [Li *et al.*, 2022], and the extortion as well as generous strategy [Press and Dyson, 2012]. While multichannel-WSLS and CIC were tailored for multichannel IPD, the rest were originally designed for traditional IPD. Thus, for comparison, we extend them to our setting by applying them to each channel. The population dynamics illustrate that MCSUC successfully invades nine out of the ten strategies, and thus is evolutionarily more advantaged than those nine strategies. The only exception that our strategy fails to invade is the ALLC strategy. However, such a failure to invade the ALLC strategy in evolutionary populations is common for IPD [Axelrod and Hamilton, 1981; Nowak and Sigmund, 1993], and the ALLC strategy itself is highly susceptible to the ALLD strategy. Motivated by this, we then consider the three-strategy population where initially there co-exist MCSUC, the ALLC strategy, and the ALLD strategy. We observe that over time, the presence of our strategy leads to the near disappearance of the ALLD strategy and the stable co-existence of MCSUC and the ALLC strategy. In other words, our strategy can protect the ALLC strategy from exploitation, thereby promoting and sustaining cooperation in evolutionary populations.

To summarize, our key contributions are as follows:

- Empirical analyses of the limitations of multichannel-WSLS and CIC, illustrating that they are vulnerable to long-lasting exploitation and can lose to multiple common strategies.
- A novel, nice, retaliatory, forgiving, and invincible strategy MCSUC for the multichannel IPD, and theoretical analyses that validate its successfulness and stability under multiple scenarios.
- Evolutionary experiments with two-strategy and three-strategy populations, demonstrating MCSUC's evolutionary advantage over nine other strategies, and its capability to promote and sustain cooperation in evolutionary populations.

2 Related Work

Cooperation in IPD and the variants thereof has attracted much interest in multiple areas, ranging from multi-agent systems to game theory and computational social science [Vinitsky *et al.*, 2023; Foerster *et al.*, 2018; Leibo *et al.*, 2017; Lu *et al.*, 2022]. The TFT strategy won the championship in the celebrated Axelrod's tournament [1981]. Later, numerous strategies were proposed, and in particular, GTFT and WSLS were shown to significantly outperform TFT [Nowak and Sigmund, 1992; Nowak and Sigmund, 1993]. These early works typically evaluate the strategies through evolutionary simulations. More recently, an emergent line of research targets strategies that allow for theoretical characterization of the payoff relationship between agents. Press and Dyson [2012] discovered the Zero-Determinant (ZD) strategy, showing that

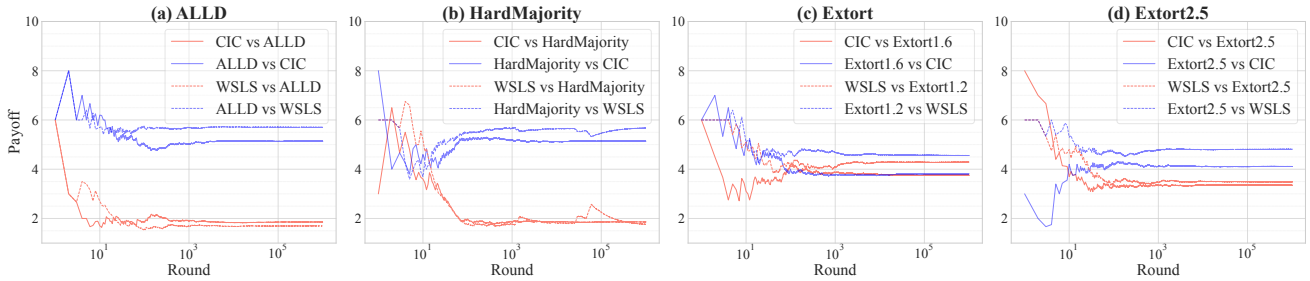

Figure 1: Two-player experiments, in which one agent employs the multichannel-WSLS/CIC strategy, while the opponent uses the ALLD, HardMajority, or the extortion strategies. The error rate is $\epsilon = 10\%$. Both the multichannel-WSLS and CIC tend to be exploited in the long term by the ALLD, HardMajority, and the extortion strategies.

in IPD, an agent can unilaterally enforce a linear relationship between its own payoff and the opponent's payoff. In particular, as a subset of ZD strategies, extortion strategies are able to enforce an extortionate linear relation between agents' payoffs [Lu *et al.*, 2022]. Akin [2016] moved beyond ZD strategies and proved that in the IPD, agents can fix the upper bound of the opponent's expected payoff through Akin's "good" strategy. Wang and Lin [2020] proposed the concept of invincibility and showed that in the absence of errors, the extortion strategies, TFT, as well as a proportion of Akin's "good" strategies, satisfy invincibility. Generalizing Akin's approach, Hao et al. [2018] and Li et al. [2019] established frameworks to allow agents to control their payoffs and their opponents' payoffs within a feasible region, thus enforcing the game towards mutual cooperation. Li et al. [2022] proposed the CURE strategy based on the principle of cumulative reciprocity, which can ensure fairness for any opponent strategy in terms of the same expected payoffs. However, all the aforementioned strategies were designed for traditional IPD. We summarize whether these strategies satisfy goodness, retaliation, forgiveness, and invincibility (with or without the presence of errors) in Appendix Section 6.

The multichannel-WSLS and CIC are the two existing strategies designed for multichannel IPD [Donahue *et al.*, 2020]. Our proposed strategy differs from these two strategies from two perspectives. First, these two strategies determine an agent's action choice in the current round based on the outcome of the last round, and do not have the cumulative characteristic of past interactions, as opposed to our proposed strategy. More importantly, these strategies are not invincible. When faced with exploitative strategies (like ALLD and most extortion strategies), they are periodically exploited without being able to retaliate effectively.

3 Background and Preliminary Experiments

In this section, we describe the multichannel IPD, multichannel-WSLS, and CIC. Moreover, we empirically illustrate the limitations of these strategies.

3.1 Multichannel IPD

In a multichannel IPD [Donahue *et al.*, 2020], agents interact with each other through various channels, and each channel is an *infinitely* IPD. For ease of presentation, we consider the two-channel IPD in this paper, however, the generalization of our theory to more than two channels is straightforward. In each channel (or game) $k \in \{1, 2\}$, the game consists of two agents X and Y, and each agent chooses to either cooperate (C) or defect (D). If both choose to cooperate, they receive a payoff of R. If one agent cooperates while the other defects, the defector gains a temptation payoff T while the cooperator receives a sucker's payoff S. If both defect, they face a punishment payoff P. This can be represented by the following payoff bi-matrix:

	C	D
C	(R_k, R_k)	(S_k, T_k)
D	(T_k, S_k)	(P_k, P_k)

(1)

where the payoff values must satisfy the conditions $T > R > P > S$ and $2R > T + S$. Under these conditions, mutual cooperation is superior to mutual defection, while defection is the dominant strategy for both agents. Therefore, the game embodies the tension between individual interest and collective interest.

The expected cooperation rate in the game is defined as the expected number of rounds in which an agent cooperates. Consider a focal agent X, let $a_X^k(t)$ represent X's action in channel k in round t. Its expected cooperation rate is

$$\rho_X = \lim_{T\to\infty} \frac{1}{2T} \sum_{t=1}^{T} \sum_{k=1}^{2} I(a_X^k(t) = C), \tag{2}$$

where $I(a_X^k(t) = C)$ is an indicator function such that $I(a_X^k(t) = C) = 1$ if agent X cooperates in game k at round t, and $I(a_X^k(t) = C) = 0$ otherwise.

The expected payoff can be defined similarly. Let $r_X^k(t)$ represent agent X's payoff in channel k at round t. Its expected payoff in channel k is given by

$$r_X^k = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} r_X^k(t). \tag{3}$$

Summing up the expected payoff in every channel yields its expected payoff in a two-channel PD, that is, $r_X = r_X^1 + r_X^2$. Likewise, we can obtain agent Y's expected cooperation rate ρ_Y and expected payoff r_Y in a similar fashion.

3.2 Multichannel-WSLS and CIC

By the multichannel-WSLS, an agent will cooperate in game k only if both the opponent agent and itself previously took

the same action in that game. By the CIC, an agent will fully cooperate in all channels if the agent and its opponent choose the same action in the last round, but will defect otherwise.

Consider an agent X that adopts the multichannel-WSLS or CIC. For a two-channel IPD, suppose that in a given round t, both the agent X and its opponent choose to defect; then in the next round $t+1$, agent X will choose to cooperate, no matter whether its opponent is exploitative or not. That is to say, even if its opponent always defects, agent X will immediately return to cooperation once it chooses to defect just like what its opponent consistently does; in this case, agent X will be periodically exploited by the always defecting opponent.

Inspired by this observation, we hypothesize that these two strategies are vulnerable to long-lasting exploitation, and may easily lose to exploitative strategies. To test this hypothesis, we conduct experiments and pair a multichannel-WSLS or CIC agent with an opponent that adopts one of the three types of exploitative strategies: the ALLD strategy, Hard-Majority, and the extortion strategy. Note that the extortion strategy involves a large class of strategies; we select two of them that stand for different levels of extortion. We consider $1,000,000$ rounds for each pair, and the average payoff per round is shown in Figure 1.

It is clear that under all these scenarios, the average payoff of the multichannel-WSLS or CIC agent is significantly lower than that of the opponent. This indicates that both the multichannel-WSLS and CIC lose to these strategies, as they can be exploited by these strategies over a long-lasting period. In addition to the scenarios with a error rate of $\epsilon = 10\%$ shown in Figure 1, we also conduct experiments with a lower error rate $\epsilon = 1\%$ and illustrate the results in Appendix Section 3. The results in both cases are broadly consistent.

4 A Successful Strategy for Multichannel IPD

In this section, we propose a successful strategy for multichannel IPD. We consider that for every round, the action choice of agent X in a two-channel IPD is influenced by the difference in the cumulative number of defections between two agents. Consider round t. Let $n_X^k(t)$ be the number of rounds, in which agent X chose to defect in channel k, out of the previous $t-1$ rounds, i.e.

$$n_X^k(t) = \sum_{t=1}^{n-1} I\left(a_X^k(t) = D\right), \tag{4}$$

where $I(a_X^k(t) = D)$ is an indicator function such that $I(a_X^k(t) = D) = 1$ if agent X chooses defect in game k at round t, otherwise, $I(a_X^k(t) = D) = 0$. For agent X, we define the cumulative number of defections to be $n_X(t) = n_X^1(t) + n_X^2(t)$. Likewise, agent Y's cumulative number of defections can be defined in a similar manner, and we denote this by $n_Y(t)$. Intuitively, for agent X, if its opponent Y has defected more frequently than itself, then agent X may be more prone to defection. Let $d_X(t) = n_Y(t) - n_X(t)$ represent the difference in the cumulative number of defections between two agents accumulated over the previous $t-1$ rounds. Formally, we define our strategy MCSUC for iterated multichannel PD as follows.

Definition 1. *In a two-channel PD, let $\Delta_{X,1}$ and $\Delta_{X,2}$ be agent X's tolerance thresholds, such that $\Delta_{X,1} \geq 1$, $\Delta_{X,2} \geq \Delta_{X,1} \geq 1$. At round t, agent X's choice of actions is determined by the following rule:*

- *If $d_X(t) \leq \Delta_{X,1}$, agent X cooperates on both channels.*
- *If $\Delta_{X,1} < d_X(t) \leq \Delta_{X,2}$, with probability p agent X defects in channel 1 but cooperates in channel 2, whereas with probability $1-p$ agent X defects in channel 1 but cooperates in channel 2.*
- *If $d_X(t) > \Delta_{X,2}$, agent X defects in both channels.*

Our proposed strategy MCSUC can be viewed as implementing the concept of cumulative reciprocity [Li *et al.*, 2022] in the context of multichannel IPD. The key idea of cumulative reciprocity is that it considers the entire history of interactions rather than just the most recent interaction. According to Definition 1, if an agent employs our strategy, they will conduct a cumulative evaluation based on all past actions of the opponent and themselves to determine whether to fully cooperate, partially cooperate, or fully defect. The values of $\Delta_{X,1}$ and $\Delta_{X,2}$ quantify the agent X's tolerance in two games—the larger the value, the higher the agent's tolerance level. Probability p represents the agent's preference for action choices in different channels. The larger p is, the greater the likelihood that agent X will defect in channel 1 and cooperate in channel 2. The smaller p is, the more likely agent X will defect in channel 2 and cooperate in channel 1.

As for the extension of our strategy to more channels. Every newly added channel will induce a new tolerance threshold. For example, consider three channels, an agent X will have three tolerance thresholds $\Delta_{X,1} \leq \Delta_{X,2} \leq \Delta_{X,3}$. In principle, the agent will defect in more channels should the cumulative number of defections $d_X(t)$ increase to exceed certain thresholds. That is, if $d_X(t) \leq \Delta_{X,1}$, the agent cooperates in all channels; if $\Delta_{X,1} < d_X(t) \leq \Delta_{X,2}$, the agent defects in one channel but cooperates in the other two, and so on so forth.

Niceness, retaliation, and forgiveness. Initially, MCSUC will cooperate in both games due to $d_X(t) = 0$. If the opponent defects so frequently that the difference in the cumulative number of defections between the two players exceeds MCSUC's tolerance threshold (i.e. $d_X(t) > \Delta_{X,1}$), then MCSUC will opt to retaliate by defecting. This shows that MCSUC is nice as it is never the first to defect; on the other hand, this also indicates that MCSUC is retaliatory in response to an opponent's continuous defection. Moreover, $\Delta_{X,1}$ reflects the degree of retaliation, with a smaller $\Delta_{X,1}$ suggesting a lower likelihood of retaliating against the opponent. However, MCSUC will not retaliate endlessly. Once the opponent opts for cooperation after frequent defections, the difference in the cumulative number of defections between the two players decreases. Once $d_X(t) < \Delta_{X,1}$, MCSUC will revert to cooperation, demonstrating its forgiveness.

5 Theoretical Analyses

In this section, we analyze the performance and stability of our strategy MCSUC in various scenarios; the proofs are presented in Section 4 of the Appendix, due to the lack of space.

5.1 Performance under Self-Play

To gain some first insights into the performance, we first analyze the use of our strategy under the self-play setting, i.e., the multichannel IPD between two agents that both adopt our strategy. In the following theorem, we show that our strategy leads to the desired state of mutual cooperation in every channel and every round:

Theorem 1. *In a two-channel IPD, agents X and Y, both using MCSUC, will always cooperate in both channels, leading to full cooperation. Their expected cooperation rates are $\rho_X = \rho_Y = 1$, and the expected payoffs are $r_X = r_Y = R_1 + R_2$.*

Note that this theorem makes no assumption about the tolerance thresholds $\Delta_{X,1}, \Delta_{X,2}, \Delta_{Y,1}, \Delta_{Y,2}$, or the payoff values as long as satisfying the mere requirement of being a PD in each channel. Thus, under self-play, our strategy can result in mutual cooperation.

This theorem, however, has not considered the presence of errors; in the real world, an agent's actions may be subject to 'trembling hand' errors. That is, when an agent chooses an action, there is a uniform probability ϵ $(0 \leq \epsilon < 0.5)$ that the agent takes a different action. We take the presence of errors into account in the following theorem:

Proposition 1. *In a two-channel IPD with an error rate ϵ, the expected cooperation rate for agents X and Y that both use MCSUC are*

$$\rho_X = \rho_Y \approx 1 - \epsilon + \frac{\epsilon - 2\epsilon^2}{(2\epsilon - 1)(\Delta_{X,1} + \Delta_{Y,1}) - 1}, \quad (5)$$

which monotonically increases as the error rate ϵ decreases, or as the tolerance thresholds $\Delta_{X,1}, \Delta_{Y,1}$ increase.

As the error becomes rare, the expected cooperation rate increases; in particular, as $\epsilon \to 0$, the expected cooperation rate tends to 1, which recovers the finding in Theorem 1. Moreover, if the tolerance thresholds are sufficiently large, the expected cooperation rates ρ_X, ρ_Y are approximately $1-\epsilon$, which represents the theoretical maximum—the highest achievable cooperation rate in the presence of error. We give an analytic form of the expected payoff under this setting in Section 4 of the Appendix.

Given that errors are typically unavoidable in the real world, our subsequent analysis assumes their presence unless otherwise stated.

5.2 Performance against ALLC or ALLD

Now, we consider the scenarios in which our proposed strategy interacts with two common strategies: ALLC and ALLD. The ALLD strategy manifests an extreme case of exploitation by invariably choosing defects. By interacting with the ALLD strategy, we can gauge the capability of our proposed strategy to counteract the exploitation from the opponent. On the other hand, the ALLC strategy represents another extreme case of always choosing cooperation. Thus, the ALLC strategy serves as a touchstone to test if our proposed strategy can resist the temptation of exploiting others.

In the following theorem, we characterize the expected cooperation rates and expected payoffs when our strategy interacts with the ALLD strategy.

Theorem 2. *In a two-channel IPD, suppose agent X uses MCSUC while the opponent agent Y uses the ALLD strategy. With error rate $\epsilon > 0$, both agents have the same expected cooperation rate $\rho_X = \rho_Y = \epsilon$, and have the same expected payoff*

$$r_X = r_Y = \epsilon^2(R_1 + R_2) + (1-\epsilon)^2(P_1 + P_2) + \epsilon(1-\epsilon)(T_1 + S_1 + T_2 + S_2). \quad (6)$$

The expected cooperation rate $\rho_X = \epsilon$ implies that our strategy mostly defects and effectively acts like the ALLD strategy, when it is confronted with the ALLD strategy. Moreover, the same expected payoff as the ALLD strategy, shown by $r_X = r_Y$ in Equation 7, suggests that our strategy can avoid lasting exploitation from the ALLD strategy, as opposed to the existing strategies (multichannel-WSLS and CIC) [Donahue *et al.*, 2020].

Likewise, we obtain the following result when our strategy interacts with the ALLC strategy:

Theorem 3. *In a two-channel IPD, suppose agent X uses MCSUC while the opponent agent Y uses the ALLC strategy. With error rate $\epsilon > 0$, both agents have the same expected cooperation rate, $\rho_X = \rho_Y = 1 - \epsilon$, and have the same expected payoff*

$$r_X = r_Y = (1-\epsilon)^2(R_1 + R_2) + \epsilon^2(P_1 + P_2) + \epsilon(1-\epsilon)(T_1 + S_1 + T_2 + S_2). \quad (7)$$

The expected cooperation rate $\rho_X = 1 - \epsilon$ implies that our strategy mostly cooperates, and effectively acts like the ALLC strategy, when it is confronted with the ALLC strategy. Thus, our strategy can resist the temptation to exploit the other's cooperation.

Putting Theorems 3 and 4 together, we conclude that our strategy demonstrates the characteristics of a "partner" strategy—returns cooperation for cooperation and returns defection for defection—which has been previously identified as the key to maintaining cooperation [Hilbe *et al.*, 2018].

5.3 Fairness and Invincibility

Based on the aforementioned results, we additionally notice an interesting finding: the expected cooperation rates of both agents are always the same, regardless of the specific scenario; this pattern also holds for the expected payoffs. This inspires us to ask whether our proposed strategy can ensure fairness in certain scenarios. We answer this question in the following theorem.

Theorem 4. *In a two-channel IPD, suppose agent X uses MCSUC while the opponent agent Y uses a strategy α. For any opponent strategy α, both agents always have the same expected cooperation rate $\rho_X = \rho_Y$. Moreover, if the two-channel IPD satisfies $(T_1 - S_1) - (T_2 - S_2) = 0$, both agents also have the same expected payoff $r_X = r_Y$.*

This theorem shows that under the condition $(T_1 - S_1) - (T_2 - S_2) = 0$, our proposed strategy ensures fairness in terms of the expected payoffs, no matter what strategy the other agent plays. This condition refers to the scenarios where the difference between the temptation to defect (denoted by

T) and the risk of being deceived (denoted by S) are the same across different channels.

It is then natural to ask what happens when the aforementioned condition $(T_1 - S_1) - (T_2 - S_2) = 0$ is not satisfied. Interestingly, we show in the following theorem that our proposed strategy can achieve invincibility in these scenarios. That is to say, even though fairness is not guaranteed, our strategy results in an expected payoff that is not smaller than that of any opponent strategy.

Theorem 5. *In a two-channel IPD, suppose agent X uses MCSUC while the opponent agent Y uses a strategy α. For any opponent strategy α, if $(T_1-S_1)-(T_2-S_2) > 0$, setting $p = 1$ will result in $r_X \geq r_Y$. If $(T_1 - S_1) - (T_2 - S_2) < 0$, setting $p = 0$ will result in $r_X \geq r_Y$.*

Therefore, when $(T_1 - S_1) - (T_2 - S_2) \neq 0$, by adjusting the probability p of cooperating in channel 1 but defecting in channel 2, our strategy never loses to any opponent strategy in terms of the expected payoff.

Putting Theorem 4 and Theorem 5 together, we conclude the performance of our proposed strategy against any opponent strategy as follows:

Corollary 1. *In a two-channel IPD, suppose agent X adopts our proposed strategy while the opponent agent Y uses a strategy α. For any opponent strategy α, there always exists a value of p (the probability that agent X defects in channel 1 but cooperates in channel 2) such that the expected payoff of our strategy is not smaller than that of the opponent strategy, i.e. $r_X \geq r_Y$.*

Simply put, our proposed strategy can always ensure invincibility, regardless of the opponent strategy. Note that this capability of ensuring invincibility is a distinct property of our proposed strategy. In contrast, the multichannel-WSLS and CIC typically fail to ensure invincibility or fairness as we have already shown in Figure 1.

5.4 Stability

Last but not least, we examine the stability of our proposed strategy through equilibrium analysis. We focus on the self-play settings, as the equilibrium analysis given any arbitrary opponent strategy is typically intractable in IPD [Press and Dyson, 2012; Akin, 2016].

Theorem 6. *In a two-channel IPD without the presence of errors, the strategy profile $(MCSUC, MCSUC)$ is a Nash equilibrium as well as a subgame perfect equilibrium.*

This theorem indicates that in the absence of errors, if both agents use our strategy, the strategy is stable in the sense that no player can increase the payoff in the two-channel IPD as well as in every subgame by unilaterally deviating from our strategy. However, as we show in the following proposition, this does not hold given the presence of errors:

Proposition 2. *In a two-channel IPD with a positive error rate $\epsilon > 0$, the strategy profile $(MCSUC, MCSUC)$ is not a Nash equilibrium.*

That said, under the self-play setting, even though the use of our strategy does not lead to a Nash equilibrium, we show in the following theorem that it results in an approximate Nash equilibrium given a sufficiently small error rate:

Theorem 7. *In a two-channel IPD with a positive error rate $\epsilon > 0$, given a non-negative $\hat{\varepsilon}$, the strategy profile $(MCSUC, MCSUC)$ is an $\hat{\varepsilon}$-equilibrium if*

$$\epsilon \leq \frac{\hat{\varepsilon}}{\max\{\max\left[(R_1 - P_1), (2R_1 - S_1 - T_1)\right], \max\left[(R_2 - P_2), (2R_2 - S_2 - T_2)\right]\}}. \tag{8}$$

For an $\hat{\varepsilon}$-equilibrium, the strategy profile approximately satisfies the condition of Nash equilibrium in the sense that no players can gain more than $\hat{\varepsilon}$ in the expected payoff by unilaterally deviating from the equilibrium strategy [Roughgarden, 2010]. Thus, this theorem shows that under the self-play setting, given a sufficiently small error rate, the use of our strategy is stable in the sense of satisfying the requirement of $\hat{\varepsilon}$-equilibrium.

6 Evolutionary Experiments

In this section, we present two sets of evolutionary experiments: two-strategy populations and three-strategy populations. More experiments that numerically validate our analytical results and examine the capability of our strategy in other multichannel games are presented in the Appendix.

6.1 Two-Strategy Populations

To investigate if our strategy holds an evolutionary advantage, we consider two-strategy populations and let agents' strategies evolve. One strategy is MCSUC with $\Delta_1 = 2, \Delta_2 = 4$, and the other strategy can be multichannel-WSLS, CIC, ALLC, ALLD, TFT, GTFT, HardMajority, CURE, the extortion strategy, or the generous strategy. We extend these strategies to multichannel IPD, and consider the extension in simulations. The details of these extensions are elaborated in the Appendix Section 1.

For each simulation, there are two steps. First, we obtain the average payoffs of two strategies in 1,000 simulation runs each lasting for 1,000,000 rounds. The payoff matrix is shown in Appendix Section 1. Then through a noisy 'survival of the fittest' environment with a mutation rate 10% [Antal *et al.*, 2009; Bodnar *et al.*, 2020], we calculate the strategies' frequencies based on the average payoffs obtained in the first step. Specifically, a noisy 'survival of the fittest' environment assumes a population of agents adopting each of the two strategies. Whether an agent changes its strategy is largely influenced by the fitness of each strategy, which is deduced from the average payoffs obtained in the first step. Our calculation of fitness follows Nowak and Sigmund's approach [Nowak and Sigmund, 1992]. A strategy is said to be able to invade if the strategy starts with a low initial frequency but finally dominates the population. A simulation ends if the frequency of each strategy no longer changes. This indicates two possible steady states, that is, either the full invasion of one strategy into the other or the stable coexistence of the two strategies. It is noteworthy that different from the experiments (shown in Figure 1) where one agent competes against another agent, here we consider a large population of agents, in which some agents employing the MCSUC compete against other agents employing a different strategy.

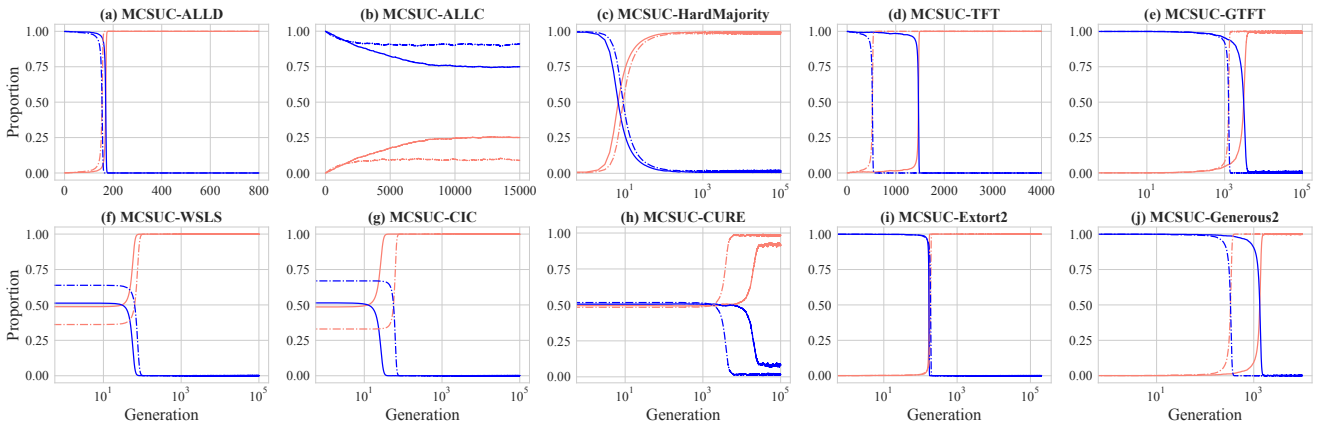

Figure 2: Two-strategy populations, where one strategy is MCSUC. The initial frequency of our strategy is 0.1% in most populations ((a)-(e),(i),(j)) and lower than 50% in other populations ((f),(g),(h)). There are two cases of error rates: $\epsilon = 1\%$ (solid line) and $\epsilon = 10\%$ (dashed line). The red line represents MCSUC, while the blue line represents the opponent's strategy. Our strategy is evolutionarily more advantaged than nine of the ten strategies except for the ALLC strategy.

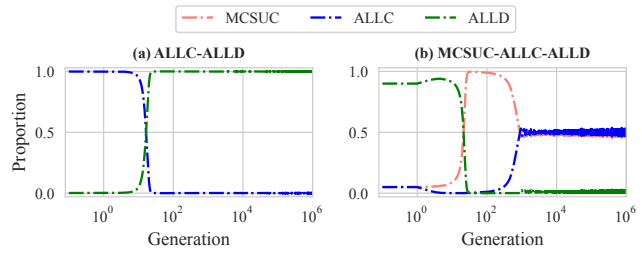

Figure 3: (a) A two-strategy population with only the ALLC and ALLD. The ALLD strategy invades the ALLC strategy, leading to the dominance of the ALLD strategy. (b) A three-strategy population with the ALLC, ALLD, and MCSUC. Eventually, the ALLD strategy almost disappears, and our strategy MCSUC and the ALLC strategy stably co-exist. In both subplots, the error rate is $\epsilon = 10\%$.

As shown in Figure 2(a),(c)-(j), our strategy successfully invades nine out of the ten strategies, except the ALLC strategy. This means our strategy has an evolutionary advantage against these strategies except the ALLC strategy. Remarkably, even with 0.1% of agents initially adopting our strategy in the population, our strategy fully invades the ALLD, HardMajority, TFT, GTFT, extortion, and generous strategies. Moreover, with less than 50% of agents initially using our strategy, our strategy fully invades the multichannel-WSLS, CIC, and CURE strategy. We also note that our strategy is risk-dominant compared to the multichannel-WSLS, CIC, and CURE strategies.

When playing against the ALLC strategy (Figure 2(b)), our strategy does not invade. Rather, at the end of the simulation, our strategy and the ALLC strategy stably co-exist. We remark that failures in an invasion against the ALLC strategy are common in the research on IPD. Moreover, the ALLC strategy itself can be invaded by the ALLD strategy (Figure 3(a)), the latter of which, in the other way around, can be invaded by our strategy. This motivates us to ask what if the ALLC, ALLD, and MCSUC coexist?

6.2 Three-Strategy Population

To answer the above question, we then consider a three-strategy population (with the ALLC, ALLD, and MCSUC) using a similar approach as in the two-strategy populations. As shown in Figure 3(b), after 70 generations, the ALLD strategy is almost eliminated from the population, while most agents in the population adopt our strategy. Later, as our strategy cannot invade the ALLC strategy, the number of agents adopting the ALLC strategy increases, which eventually leads to the stable co-existence of our strategy and the ALLC strategy. Because our strategy behaves like the ALLC strategy when facing the ALLC strategy, the eventual stable co-existence suggests that most of the agents will cooperate. Therefore, although our strategy cannot invade the ALLC strategy in the two-strategy population, our strategy can protect the ALLC strategy from exploitation in the three-strategy population with the presence of the ALLD strategy, thereby promoting and sustaining cooperation.

7 Conclusions

In this paper, we address the challenge of what a successful strategy should be in the multichannel IPD. We start with analyses of the existing strategies, showing that they are vulnerable to long-lasting exploitation and lose to some common strategies. Motivated by this, we propose a novel strategy that is nice, retaliatory, forgiving, and perhaps most interestingly, invincible. We present extensive analytical results regarding the performance and stability of our strategy under various scenarios. Our key results remain applicable in games with more channels. The expected cooperation rate will *not* change as the number of channels increases. However, as more channels will naturally induce additional payoff parameters to the games, the expected payoff will need to change by taking into account these additional payoff parameters; nevertheless, this involves just a re-calculation using our current approach. Likewise, our proofs for fairness, invincibility, and stability do *not* depend on the number of channels. Thus, even with more channels, MCSUC will still enjoy these properties.

Acknowledgements

This research was supported by The National Science Fund for Distinguished Young Scholars (no. 62025602), the National Natural Science Foundation of China (Nos. U22B2036, 62366058, 11937815, 62066045, 11971421, 62073263 and 11931015), Excellent Youths Project for Basic Research of Yunnan Province (No. 202101AW070015), Fok Ying-Tong Education Foundation, China (No.171105), the Fundamental Research Funds for the Central Universities (No. D5000230366), Yunnan Province XingDian Talent Support Program (YNWR-QNBJ-2020-041, YNWR-YLXZ-2018-020), the Foundation of Yunnan Key Laboratory of Service Computing (No. YNSC23117), Tencent Foundation and XPLORER PRIZE, Open Research Subject of State Key Laboratory of Intelligent Game (Grant no. ZBKF-24-02) and Shanghai Artificial Intelligence Laboratory.

References

[Akin, 2016] Ethan Akin. The iterated prisoner's dilemma: good strategies and their dynamics. *Ergodic Theory, Advances in Dynamical Systems*, pages 77–107, 2016.

[Antal *et al.*, 2009] Tibor Antal, Hisashi Ohtsuki, John Wakeley, Peter D Taylor, and Martin A Nowak. Evolution of cooperation by phenotypic similarity. *Proceedings of the National Academy of Sciences*, 106(21):8597–8600, 2009.

[Axelrod and Hamilton, 1981] Robert Axelrod and William D Hamilton. The evolution of cooperation. *science*, 211(4489):1390–1396, 1981.

[Baker, 2020] Bowen Baker. Emergent reciprocity and team formation from randomized uncertain social preferences. *Advances in neural information processing systems*, 33:15786–15799, 2020.

[Bodnar *et al.*, 2020] Cristian Bodnar, Ben Day, and Pietro Lió. Proximal distilled evolutionary reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3283–3290, 2020.

[Chu *et al.*, 2022] Chen Chu, Yong Li, Jinzhuo Liu, Shuyue Hu, Xuelong Li, and Zhen Wang. A formal model for multiagent q-learning dynamics on regular graphs. In *IJCAI*, pages 194–200, 2022.

[Donahue *et al.*, 2020] Kate Donahue, Oliver P Hauser, Martin A Nowak, and Christian Hilbe. Evolving cooperation in multichannel games. *Nature communications*, 11(1):3885, 2020.

[Farrell and Maskin, 1989] Joseph Farrell and Eric Maskin. Renegotiation in repeated games. *Games and economic behavior*, 1(4):327–360, 1989.

[Foerster *et al.*, 2018] Jakob Foerster, Richard Y Chen, Maruan Al-Shedivat, Shimon Whiteson, Pieter Abbeel, and Igor Mordatch. Learning with opponent-learning awareness. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 122–130, 2018.

[García and van Veelen, 2016] Julián García and Matthijs van Veelen. In and out of equilibrium i: Evolution of strategies in repeated games with discounting. *Journal of Economic Theory*, 161:161–189, 2016.

[Hao *et al.*, 2018] Dong Hao, Kai Li, and Tao Zhou. Payoff control in the iterated prisoner's dilemma. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 296–302, 2018.

[Hilbe *et al.*, 2018] Christian Hilbe, Krishnendu Chatterjee, and Martin A Nowak. Partners and rivals in direct reciprocity. *Nature human behaviour*, 2(7):469–477, 2018.

[Leibo *et al.*, 2017] Joel Z Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. Multi-agent reinforcement learning in sequential social dilemmas. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 464–473, 2017.

[Lerer and Peysakhovich, 2017] Adam Lerer and Alexander Peysakhovich. Maintaining cooperation in complex social dilemmas using deep reinforcement learning. *arXiv preprint arXiv:1707.01068*, 2017.

[Li and Hao, 2019] Kai Li and Dong Hao. Cooperation enforcement and collusion resistance in repeated public goods games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 2085–2092, 2019.

[Li *et al.*, 2022] Juan Li, Xiaowei Zhao, Bing Li, Charlotte SL Rossetti, Christian Hilbe, and Haoxiang Xia. Evolution of cooperation through cumulative reciprocity. *Nature Computational Science*, 2(10):677–686, 2022.

[Lu *et al.*, 2022] Christopher Lu, Timon Willi, Christian A Schroeder De Witt, and Jakob Foerster. Model-free opponent shaping. In *International Conference on Machine Learning*, pages 14398–14411. PMLR, 2022.

[Mailath and Samuelson, 2006] George J Mailath and Larry Samuelson. *Repeated games and reputations: long-run relationships*. Oxford university press, 2006.

[Miller, 1985] Nicholas R Miller. Nice strategies finish first: A review of the evolution of cooperation. *Politics and the Life Sciences*, 4(1):86–91, 1985.

[Mittal and Deb, 2009] Shashi Mittal and Kalyanmoy Deb. Optimal strategies of the iterated prisoner's dilemma problem for multiple conflicting objectives. *IEEE Transactions on Evolutionary Computation*, 13(3):554–565, 2009.

[Nowak and Sigmund, 1992] Martin A Nowak and Karl Sigmund. Tit for tat in heterogeneous populations. *Nature*, 355(6357):250–253, 1992.

[Nowak and Sigmund, 1993] Martin Nowak and Karl Sigmund. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, 364(6432):56–58, 1993.

[Press and Dyson, 2012] William H Press and Freeman J Dyson. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences*, 109(26):10409–10413, 2012.

[Roughgarden, 2010] Tim Roughgarden. Algorithmic game theory. *Communications of the ACM*, 53(7):78–86, 2010.

[Selten and Hammerstein, 1984] Reinhard Selten and Peter Hammerstein. Gaps in harley's argument on evolutionarily stable learning rules and in the logic of "tit for tat". *Behavioral and Brain Sciences*, 7(1):115–116, 1984.

[Van Veelen, 2012] Matthijs Van Veelen. Robustness against indirect invasions. *Games and Economic Behavior*, 74(1):382–393, 2012.

[Vinitsky *et al.*, 2023] Eugene Vinitsky, Raphael Köster, John P Agapiou, Edgar A Duéñez-Guzmán, Alexander S Vezhnevets, and Joel Z Leibo. A learning agent that acquires social norms from public sanctions in decentralized multi-agent settings. *Collective Intelligence*, 2(2):26339137231162025, 2023.

[Wang and Lin, 2020] Shiheng Wang and Fangzhen Lin. Nice invincible strategy for the average-payoff ipd. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2268–2275, 2020.

[Wang *et al.*, 2022] Zhen Wang, Chunjiang Mu, Shuyue Hu, Chen Chu, and Xuelong Li. Modelling the dynamics of regret minimization in large agent populations: a master equation approach. In *IJCAI*, pages 534–540, 2022.

[Zhao *et al.*, 2022] Stephen Zhao, Chris Lu, Roger B Grosse, and Jakob Foerster. Proximal learning with opponent-learning awareness. *Advances in Neural Information Processing Systems*, 35:26324–26336, 2022.