Direct Reciprocity in Games of Choice

Direct Reciprocity in Games of Choice

An undergraduate thesis presented by
Rachel G. Gologorsky
to
The School of Engineering and Applied Sciences
in partial fulfillment of the requirements for
the joint degree of Bachelor of Arts in Computer Science and Mathematics with Honors

Harvard College
Cambridge, Massachusetts

March 29, 2019

Acknowledgements

I am incredibly fortunate to have had Dr. Christian Hilbe as my mentor. Thank you Dr. Hilbe for skillfully developing my understanding of evolutionary game theory and generously sharing your enthusiasm, insights, and knowledge with me. I would like to express my deep appreciation to Professor David Parkes and Professor Martin Nowak for being my teachers, mentors, and advisors. Your classes first aroused my curiosity to explore this field more deeply, and your advice and feedback have tremendously improved the clarity and content of this thesis. I am truly grateful for my wonderful friends: Thank you Aaron Klein, Joseph Hostyk, and Yael Goldstein for your warm friendship and for allowing my code to occasionally take over your computers from dusk to dawn. As it is said, "True friendship multiplies the good in life and divides its evils."[1] Most importantly, neither I nor this thesis would exist without my family. Thank you Mom and Dad for your unwavering love, support, and advice that have made this project possible.

[1] Baltasar Gracián

Abstract

How does cooperation evolve in a population of self-interested agents? One of the primary mechanisms is through direct reciprocity, which occurs when individuals choose to cooperate because they expect to be rewarded with cooperation in return. Prior work has mainly analyzed direct reciprocity through the lens of the iterated Prisoner's Dilemma (IPD). However, the iterated Prisoner's Dilemma is a simplification of real-world cooperation because participants always face the same payoff structure in every encounter. In the real world, participants exert some control over the reward structure of their interactions: that is the motivation behind gaining trust and investing in deeper relationships.

In order to capture the ability of future interactions to become more (or less) rewarding, I extend the IPD into a new model, games of choice. A game of choice is a stochastic game wherein the players directly choose between two Prisoner's Dilemma (PD) stage games each round. In a game-of-choice, each player specifies his preferred PD stage game, and a prespecified resolution rule decides which stage game is played. Cooperating provides a larger benefit to one's co-player in one stage game than in the other. In this thesis, I explore how different resolution rules and strategy spaces affect cooperation rates in games-of-choice, determine the stability condition under which reciprocal strategies become subgame-perfect equilibria (SPE), and interpret the mechanism promoting direct reciprocity.

I find that games of choice tangibly promote cooperation: cooperative SPE can exist in a game-of-choice between two PD stage games even when cooperative SPE do not exist in an IPD of either stage game alone. I discover that a single mechanism underlies all cooperative SPE strategies in the two most effective resolution rules. This fundamental mechanism stipulates that the only route back to mutual cooperation in the more-rewarding stage game is through mutual cooperation in the less-rewarding stage game. Requiring time to be spent in the less-rewarding game stabilizes cooperation by imposing an increased opportunity cost on defection. As a result of this work, we gain insight into the way that concepts such as "investing in deeper relationships" may manifest themselves in strategies and change the cost/benefit calculus of self-interested agents.

Contents

Chapter 1

Introduction

1.1 Overview

1.1.1 Evolutionary Game Theory

The purpose of evolutionary game theory is to make the mechanisms underlying biological evolution mathematically precise [1]. As the name suggests, evolutionary game theory combines game theory with evolution. Individuals interact in the context of game theory: each interaction between individuals is typically given by a normal-form game. Individuals adopt strategies for playing the game; a strategy is a decision rule that specifies the action a player will take after each possible sequence of moves. An individual's evolutionary "fitness" is his strategy's average payoff when playing against the strategies of other individuals in the population. The evolutionary process of selection pressure and mutation drives the macro-level process of strategy change and adaption.

Compared to traditional game theoretic analysis, the evolutionary approach has the advantage of modelling the behavior of self-interested agents without additionally assuming individual rationality and high levels of cognitive ability. This is thought to be a more realistic model of how self-interested organisms develop strategies in nature, and provides an additional perspective – that of evolutionary stability – from which to mathematically analyze strategy robustness [2]. Mathematically, the condition for evolutionary stability is in between that of strict and weak Nash equilibrium: a strict Nash equilibrium implies evolutionary stability, and evolutionary stability implies a weak Nash equilibrium [1]. In practice, the evolutionary perspective offers a conceptually simple way to probe the strategy space: one can directly simulate the stochastic mutation/selection process of evolution [3]. Simulating the process of evolution provides both (a) a method of impartially discovering successful strategies without relying on preconceptions of what those strategies look like and (b) insight into how certain natural phenomena, like cooperation among self-interested agents, may have arisen in nature in the first place [3]. Indeed, the study of the mechanisms underlying cooperation is an ongoing and vibrant area of research [4–9].

1.1.2 Direct Reciprocity

One of the main mechanisms studied in relation to the evolution of cooperation is direct reciprocity [3, 10–13]. The mechanism of direct reciprocity stipulates that individuals always have a positive probability of interacting again. The possibility of repeated interactions fundamentally changes the cost/benefit analysis of self-interested agents because cooperation now can be rewarded with the co-player's reciprocation next time. When the probability of future interactions is high enough, the temptation to not cooperate is offset by expected loss in the co-player's future cooperation; thus direct reciprocity can give rise to sustained mutual cooperation.

The study of direct reciprocity dates back to Robert Axelrod and William Hamilton's 1981 work on the evolution of cooperation, in which they proposed the iterated Prisoner's Dilemma (IPD) as a model. The iterated Prisoner's Dilemma is a repeated game between two players, where each individual interaction has the Prisoner's Dilemma reward structure. In a Prisoner's Dilemma (PD), each player simultaneously chooses between two actions: cooperation (C) and defection (D). Mutual cooperation (CC) is better than mutual defection (DD) for both players. Yet, from each player's point of view, the immediate payoff for choosing defection (resulting in a DC or a DD outcome) is always higher than the corresponding payoff for choosing cooperation (resulting in a CC or a CD outcome). A standard PD paradigm is that cooperation involves paying a cost c in order to provide a benefit $b > c$ to the co-player. On the other hand, defection is a neutral action that comes at no personal cost but provides no benefit to the co-player. Consequently, each player receives net payoff $b - c > 0$ for a mutual cooperation outcome and each player receives zero payoff for a mutual defection outcome. Yet, from each player's point of view, inaction always provides higher payoffs than cooperation:

- if the co-player cooperates, the choice is between receiving net payoff b (defection) or $b - c$ (cooperation).
- if the co-player defects, the choice is between receiving net payoff 0 (defection) or $-c$ (cooperation).

The iterated Prisoner's Dilemma thus captures the tension between the temptation to defect in any individual encounter and the potential benefit of sustained cooperation. Under the right mathematical conditions, Hamilton and Axelrod showed that cooperative strategies such as Tit-for-Tat could be initially viable and evolutionary stable. As the result of their influential work, the iterated Prisoner's Dilemma became the canonical way to model direct reciprocity [10, 11]. Moreover, the research framework of ecological simulations coupled with the analytical lens of evolutionary stability became the standard computational/analytical tools through which to explore and analyze the evolution of cooperation [3, 12].

Subsequent work in the field discovered other strategies of interest (such as Win-Stay-Lose-Shift and Zero-Determinant strategies) [12, 14], introduced more realistic features into the model (such as inherent noise and environmental feedback loops) [3, 15], probed the specific stages in the evolutionary development of cooperation and its catalysts [13], and developed conceptual paradigms through which to view the fundamental characteristics of discovered strategies [16–18].

1.1.3 Stochastic Games

Stochastic games, introduced by Shapley in 1953 [19], are a generalization of repeated games. In a repeated game, players play the same normal-form game repeatedly. In stochastic games, the stage game is allowed to vary. Conceptually, each possible stage game is a state, and players transition between states each round. A transition function, taking into account the players' current actions and the prior game state, governs the transition between game states.

The ability to dynamically transition between states makes stochastic games well-suited for modeling situations with dynamic reward structures, such as the ones I consider in my thesis. Stochastic games in the past have been used to model the classical "tragedy of the commons" paradigm as well as the topical scenario of individual agents interacting with, and contributing to, a fluctuating economy [20].

1.2 Thesis Contribution

In this thesis, I extend the standard iterated Prisoner's Dilemma into a new model, games of choice. A game of choice is a stochastic game wherein the players jointly choose between two Prisoner's Dilemma (PD) games each round. The benefit for cooperating is higher in the first PD game than it is in the second PD game. Conceptually, each round proceeds in three steps:

- First, each player specifies which PD stage game he prefers.
- Second, a pre-sepecified resolution rule decides the round's stage game.
- Third, each player chooses his action based on the stage game being played.

This process is illustrated in the figure below.

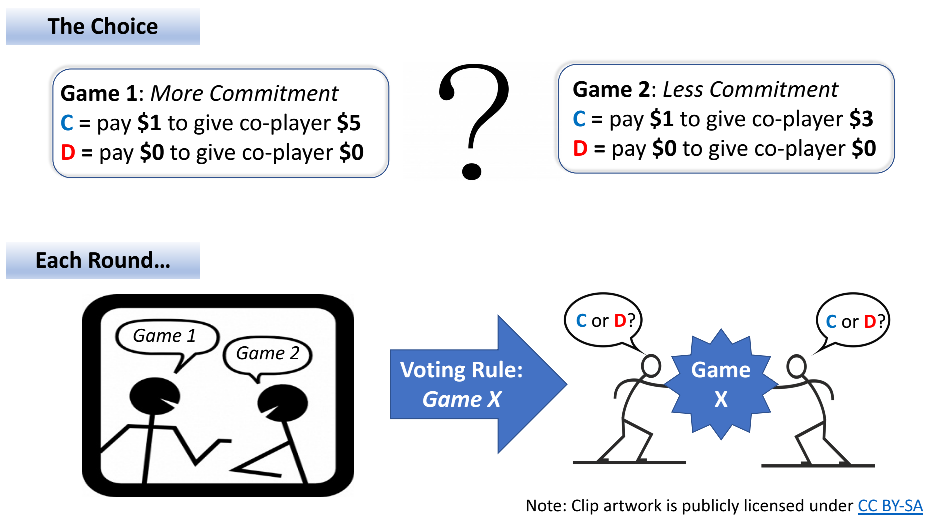


Figure 1.1: Game of Choice Illustration

In this thesis, I analyze the effect of resolution rules and strategy spaces on the evolution of cooperation, determine the fundamental mechanism promoting direct reciprocity in games-of-choice, and prove the mathematical conditions under which reciprocal strategies become subgame-perfect equilibria (SPE). Assuming that players condition their behavior only on the outcome of the previous round, I demonstrate the following results:

- Cooperative SPE strategies can exist in a game of choice between two PD games even when cooperative SPE do not exist in either iterated PD game alone. Specifically, the game-of-choice SPE condition is $2b_1 + b_2 \geq 2c$, which is always easier to achieve than that of either iterated PD game alone ($b_i \geq 2c$ for $i \in \{1, 2\}$). This can be seen by re-expressing the game-of-choice SPE condition as $b_1 + (b_1 - b_2) \geq 2c$ (easier to meet than $b_1 \geq 2c$) and $b_2 + 2(b_1 - b_2) \geq 2c$ (easier to meet than $b_2 \geq 2c$).
- Moreover, under the right resolution rule, this easier SPE condition remains achievable even when players are restricted to acting the same way in both the low- and high-reward PD game *and also* specifying their stage game preference without considering

whether the prior round outcome occurred in the low- or high-reward game.

- A single basic mechanism, operating across the two most effective resolution rules, stabilizes cooperation in a game-of-choice when it would be unstable in an IPD. This fundamental mechanism specifies that, following defection, the route to high-reward mutual cooperation must pass through low-reward mutual cooperation. In comparison to the standard IPD framework, spending time in the low-reward setting provides extra deterrence against defection, making it easier for cooperation to be a best-response.

As a result of this work, we gain insight into the way that game choice and resolution rules interact to promote cooperation, as well as the specific strategies, mathematical relationships, and underlying mechanisms that explain this result.

1.3 Related Work

To the best of my knowledge, no previous work has considered a stochastic game in which players directly choose between two reward settings each round.

1.3.1 Game Model

A one-shot version of this stochastic game is currently being explored by Momeni et al [21]. In this working paper, players "vote" on the PD game they wish to play. However, repeated games have fundamentally different cooperation dynamics than one-shot games because they allow players to engage in direct reciprocity. In addition, I consider the effect of resolution rules and strategy spaces other than those explored in Momeni et al [21].

Games with dynamic reward settings have also been studied before, such as in the context of the continuous Prisoner's Dilemma [22, 23]. In this framework, a player's action is the amount he chooses to invest in his co-player, where the invested amount can vary continuously. Players choose how much to invest independently of each other. My model differs from the continuous Prisoner's Dilemma in that players act in a shared (and discrete) reward setting: the symmetric Prisoner's Dilemma of their choice.

The main precedent for the model I study is Hilbe et al's 2018 work on the evolution of cooperation in stochastic games [15]. This paper also explores direct reciprocity in a stochastic game with two potential reward settings each round. However, in this model, the reward setting is governed by the exogenous environment state. Deviations from cooperation "degrade" the quality of the environmental resource, resulting in the "low-reward" setting; cooperation "restores" the quality of the resource, resulting in the "high-reward" setting. The reward setting is independent of player preference. My Game 1/Game 2 terminology for the high/low reward settings and choice of evolutionary parameter values is adopted from this work. However, my model differs in that players directly choose the reward structure of the game they play, thus modelling a different situation than the one resulting from exogenous feedback loops.

1.3.2 Strategy Spaces

The strategy spaces I consider are extensions of familiar strategy spaces in the iterated Prisoner's Dilemma.

In principle, strategies can be arbitrarily complex, making use of the full history of play in order to inform their next move. However, to make computational analysis tractable, the space of strategies is often restricted to specific subsets, such as pure (or almost pure) memory-1 strategies [11, 12, 14, 15]. Memory-1 strategies are strategies that only "remember" the outcome in the prior round. Pure strategies are "pure" because they choose to play a single action with probability 1 each round, rather than to mix over the available actions. Pure memory-1 strategies are of particular research interest because they are mathematically tractable and because studying a strategy's simplest form provides key insight into the way it works.

Incorporating a small amount of inherent noise into the strategy space is a refinement introduced by Sigmund & Nowak [3]. The addition of inherent "noise" causes actions to be taken with probability ϵ and $1-\epsilon$ instead of 0 and 1. Such "almost-pure" strategies better model biological reality, and also guarantee that, when two strategies meet, their long-term average payoff is independent of their initial start state.

In this thesis, I study pure and almost-pure memory-1 strategies. A memory-1 strategy s is often encoded in the form

$$s = (p_{cc}, p_{cd}, p_{dc}, p_{dd}) \in \{0, 1\}^4,$$

where p_{ab} is the player's probability to cooperate this round, given that the outcome of the prior round was ab.

I extend this type of representation to describe the strategy spaces in my model. Several fundamental IPD strategies have important analogues in the game of choice model as well. Four of these well-known IPD strategies are described in the table below.

Strategy Name	Description (This strategy...)	Representation
Always-Cooperate (ALLC)	... always cooperates, irrespective of prior round	(1,1,1,1)
Always-Defect (ALLD)	... always defects, irrespective of prior round	(0,0,0,0)
Tit-for-Tat (TFT)	...mirrors the co-player's prior move, reciprocating cooperation with cooperation.	(1,0,1,0)
Win-Stay-Lose-Shift (WSLS)	... repeats his own prior move if last round was rewarding (CC, DC); otherwise, changes his move.	(1,0,0,1)

Table 1.1: Fundamental Memory-1 Strategies in the iterated Prisoner's Dilemma

1.3.3 Evolutionary Model

I adopt Imhof & Nowak's finite-population evolutionary dynamics in the context of rare mutation events [13]. Finite populations are closer to biological reality than infinite populations and also offer the conceptual benefit that a single mutant, such as TFT in an ALLD population, has positive probability of taking root and fixating in the population. In contrast, in infinite population evolutionary dynamics, a mutant strategy must already compose a fraction of the population in order to be able to invade it [24]. Rare mutation events allow us to assume that there is only one host strategy in the population at any give time: if mutant events are sufficiently rare, then a mutant will either die out or fixate in the population before the next mutation event occurs. As a result, the set of host strategies in the population forms a Markov chain whose stationary distribution describes the proportion of time the population spends playing each strategy. The assumption of rare mutations in a finite population thus allows me to estimate the long-term proportion of time that the population spends in cooperative states. A key part of my research is exploring how the game-of-choice parameter values, strategy spaces, and resolution rules affect cooperate rates in a game-of-choice.

There are several ways of defining the transition probability in the Markov chain of host strategies, depending on the way one models strategy spread in a population. In this thesis, I use the pairwise comparison process outlined by Fudenberg & Imhof [25].

1.4 Thesis Outline

Chapter 2 formally defines my model, specifying the resolution rules and the strategy spaces under consideration. Chapter 3 presents the experimental results of evolutionary simulation. These experimental results suggest certain mathematical relationships, which Chapter 4 then sets out to prove. Chapter 5 suggests future directions and concludes this work. The

Appendix contains figures checking the robustness of my results to changes in evolutionary parameters, pseudocode for algorithms relevant to my research, and proofs of the theorems I state in the main text.

Chapter 2

Definitions

Informally, games of choice model situations wherein participants exert some influence over the reward structure of their interactions. Although informal language conveys the general idea, it fails to provide the particulars. This chapter dives into the model specifics, providing formal definitions for the game-of-choice and evolutionary model, describing the methods used in detail, and illustrating key concepts with examples and sample calculations.

2.1 Game-of-Choice Model

A simultaneous-move normal-form game $(\mathcal{N}, A, u)$ is defined [26] by:

- the set of players $\mathcal{N} = \{1, ..., n\}$,
- the set of actions A_i available to each player $i \in N$, and
- a utility function $u : A_1 \times \ldots \times A_n \rightarrow \mathbb{R}^n$ mapping action profiles to player playoffs.

In a repeated game, players play the same normal-form game repeatedly. Stochastic games generalize repeated games in that the stage game is allowed to vary. Consequently, stochastic games need to specify two additional objects: S, the set of possible stage games, and Q, a transition function mapping the current action profile and current round's stage game to the probability distribution over the stage game to be played next round [27]. Since payoffs can differ depending on the stage game that takes place, the utility function u is given the stage game as an additional argument.

In this thesis, I consider a game-of-choice in which players repeatedly choose between two Prisoner's Dilemma stage games. Cooperating in the first PD stage game provides a larger benefit to one's co-player then cooperating in the second PD stage game. These two PD matrices represent the ability for players to move between more-rewarding and less-rewarding game settings.

Formally, this game of choice $(\mathcal{N}, S, A, u, Q)$ is defined as follows:

- It is a two-player game; $\mathcal{N} = \{1, 2\}$.
- There are two stage games, $S = \{\text{Game 1, Game 2}\}$. Game 1 models the "high-reward" setting, and Game 2 models the "low-reward" setting.
- Each round, players first specify which stage game they wish to play. After the resolution rule determines stage game to be played that round, the players decide whether to cooperate (C) or defect (D); $A_i = \{1, 2\} \times \{C, D\}$.
- I present the utility function in the form of a payoff matrix for each stage game.

P1/P2	C	D
C	(b_1-c, b_1-c)	(-c, b_1)
D	(b_1, -c)	(0,0)

Table 2.1: Game 1 Payoff Matrix

P1/P2	C	D
C	(b_2-c, b-c)	(-c, b_2)
D	(b_2, -c)	(0,0)

Table 2.2: Game 2 Payoff Matrix

In both games, cooperation comes at personal cost c, while defection is a neutral action that carries no cost and provides no benefit. In Game 1, cooperation provides benefit b_1 to the other player. In Game 2, cooperation provides benefit b_2 to the other player. I let $b_1 > b_2$, so that Game 1 models the "high-reward" setting and Game 2 models the "low-reward" setting.

- I explore five different resolution rules, reflecting the different levels of authority and influence the players may have on the eventual outcome.

Random The stage game played is beyond the players' control. It is determined randomly, with equal probability of being Game 1 or Game 2. **Random** acts as a baseline from which to evaluate the efficacy of other resolution rules in promoting cooperation.

Equal-Say Game-2 default Players have equal say over which stage game is played next round. In the event of different preferences, the low-reward Game 2 is played by default.

Equal-Say Game-1 default Players have equal say over which stage game is played next round, but the default game is high-reward Game 1.

Unilateral Dictator One of the two players is randomly chosen to dictate the stage game in all of the encounters that occur between the two players.

Random Dictator Each round, one of the two players is randomly chosen to dictate the stage game.

2.1.1 Motivating Examples

For the sake of concreteness, I like to imagine:

- In Game 1, cooperation involves paying 1 unit to give the co-player 5 units,
- In Game 2, cooperation involves paying 1 unit to give the co-player 3 units, and
- In both games, defection is a "do nothing" action (no cost, no benefit).

Below are five examples, illustrating a use case for each of the five resolution rules.

A married couple. Whenever an issue comes up, each partner has a choice between compromising (cooperating) or demanding to get their way on an issue (defecting). Each issue could be important (Game 1) or not important (Game 2). *Under the **Random** resolution rule, each issue is intrinsically important (whether or not to have a baby) or unimportant (who takes out the garbage). The importance of the issue is independent of the players' preferences.*

Two researchers. Each time the researchers work together, they can either put in effort (cooperate) or coast on the other person's work (defect). They choose between collaborating on ambitious research projects (Game 1) or on mundane research papers (Game 2). *Under the **Equal-Say Game-2 default** resolution rule, both researchers need to agree in order to embark on a ambitious research project together.*

Two parties to a contract. The contract can be in effect (Game 1) or not in effect (Game 2). Whether or not the contract is in effect, each side can choose whether to act in accordance to its terms (cooperate) or not (defect). *Under the **Equal-Say Game 1 default** resolution rule, the contract remains in effect (Game 1) unless both parties agree to dissolve it.*

A CEO interacting with his employees. Each year, the CEO decides whether to promise a large end-of-year bonus for hard work (Game 1) or a small end-of-year bonus for hard work (Game 2). The employees decide whether to work hard (cooperate) or muddle through (defect); the CEO decides whether to follow through on the promise (cooperate) or say it was a bad year for the company (defect). *Under the **Unilateral Dictator** resolution rule, the CEO unilaterally decides, for all years, whether to promise a large or a small bonus.*

Two teenage siblings. On any issue, they can either compromise (cooperate) or demand to get their way (defect). *Under the **Random Dictator** resolution rule, on any specific issue, a random sibling can unilaterally make a big deal out of it (force Game 1) or act maturely (force Game 2).*

2.2 Defining Average Payoff in an Infinitely-Repeated Game

A game-of-choice is an infinitely repeated game. Each round generates a payoff, causing the players to accumulate an infinite stream of payoffs whose sum diverges to infinity. Yet, the players' payoffs must be compared to one another in order to assess strategy performance. There are several ways to solve this issue; in the following subsections, I define some useful conventions/notation and explain the two approaches I take.

2.2.1 Notation

In a game-of-choice between two players, each round can result in one of eight possible outcomes, $O = \{1CC, 1CD, \ldots, 2DC, 2DD\}$. The outcome consists of the stage game (Game 1 or Game 2) and the action profile (a_1, a_2), where a_1 corresponds to the first player's action and a_2 corresponds to the second player's action. For example, when individual i plays against individual j, 1DC represents a round that occurred in stage Game 1, in which individual i defected and individual j cooperated.

The utility function $u(o)$ maps a round outcome $o \in O$ to the players' respective payoffs. For example, $u_i(1DC) = b_1$ and $u_j(1DC) = -c$. For round outcomes $[1CC, 1CD, \ldots, 2DC, 2DD]$, player i's and player j's payoffs are given by:

$$\vec{u_i} = [b_1 - c, -c,\ \ldots,\ b_2, 0]$$
$$\vec{u_j} = [b_1 - c,\ b_1,\ \ldots, -c, 0]$$

When two strategies play each other, they progress through a sequence of round outcomes. Let o^k denote the kth round outcome in this sequence. I follow the convention of denoting the first round ($k = 1$) as Round 0.

2.2.2 Frequency-Weighted Average Payoff per Round

Suppose that when individual i with strategy s_i plays against individual j with strategy s_j,

- $\frac{1}{4}$ of their rounds result in outcome o_1, yielding payoffs of 4 and 1 to i and j respectively;
- $\frac{1}{4}$ of their rounds result in outcome o_2, yielding payoffs of 8 and 1;
- $\frac{1}{2}$ of their rounds result in outcome o_3, yielding payoffs of 12 and 1.

Over an infinite number of rounds, both players receive a cumulative payoff of ∞. However, a natural method of comparing their payoffs is to calculate the *frequency-weighted* average payoff per round. Let $E[u_i(s_i, s_j)]$ denote the frequency-weighted average payoff per round that individual i receives when strategy s_i plays against strategy s_j. In this case,

- individual i receives 4 with frequency $\frac{1}{4}$, 8 with frequency $\frac{1}{4}$, and 12 with frequency $\frac{1}{4}$, so $E[u_i(s_i, s_j)] = 4\frac{1}{4} + 8\frac{1}{4} + 12\frac{1}{2} = 1 + 2 + 6 = 9$.
- individual j receives 1 with frequency 1 (i.e. 100% of the time), so $E[u_j(s_i, s_j)] = 1 \times 1 = 1$.

In my evolutionary simulations, I consider strategies for which the probability of transitioning from round outcome o_1 to round outcome o_2 is positive for all round outcomes $o_1, o_2 \in O$. Consequently, the sequence of round outcomes is an irreducible and aperiodic Markov chain with a unique stationary distribution,

$$\vec{v} = [v_{1cc}, v_{1cd}, \ldots, v_{2dc}, v_{2dd}],$$

which describes the long-term proportion of round outcomes spent in each state [28, Chapter 11]. Given the stationary distribution, the frequency-weighted average payoff per round is:

$$\begin{aligned} E[u_i(s_i, s_j)] &= \sum_{o \in O} u_i(o) \cdot \text{Prob[game state } o] \\ &= \vec{u_i} \cdot \vec{v} \end{aligned}$$

Similarly, individual j's average payoff is: $E[u_j(s_i, s_j)] = \vec{u_j} \cdot \vec{v}$.

In my evolutionary simulations, I use this frequency-weighted average payoff per round as my payoff metric.

2.2.3 Discounted Average Payoff per Round

Deterministic, or pure, strategies always respond the same way to the same game state. These strategies are interesting from an analytical standpoint because their inner workings are easier understand and because their mathematical analysis is more tractable than that of probabilistic strategies.

When two pure strategies play each other, the initial state completely determines the subsequent sequence of round outcomes. Since this sequence may not contain all game states, the transition matrix may no longer be irreducible, and the stationary distribution may not be unique.

For example, consider the case of pure strategy Grim in the IPD. A Grim player cooperates on the first move. Afterwards, a Grim player only cooperates if the prior round outcome was CC; he defects otherwise. Proceeding from initial subgame state CC, two Grim players would continue cooperating with each other, resulting in an infinite CC loop. However, two Grim players proceeding from a non-CC initial subgame state would continue defecting against one other, resulting in an infinite DD loop. Hence the stationary distribution is not unique; it depends on the initial state.

Defining the frequency-weighted average payoff using the stationary distribution under the equilibrium path of play (both players cooperating on their first move) also contains a serious drawback. Consider the following two potential sequences of game states:

1. $CC \rightarrow DD \circlearrowright$, and
2. $CC \rightarrow CD \rightarrow DC \rightarrow DD \circlearrowright$

Both sequences result in an infinite DD loop; in the long-run, both spend 100% of their time in game state DD. If the average payoff is defined using the stationary distribution, then both sequences result in the same average payoff: DD weighted at 100%. In order to differentiate between the payoffs resulting from these two sequences, a different approach is needed.

One standard solution is to introduce a discount factor $\delta \in (0, 1)$ and calculate the cumulative discounted payoff. Individual i's payoffs generate the infinite series $\sum_{k=0}^{\infty} u_i(o^k)$. When future payoffs are discounted, $\sum_{k=0}^{\infty} u_i(o^k) \cdot \delta^k$, the cumulative (discounted) payoff converges to the finite value. Intuitively, δ can be thought as the probability of another round occurring (or as the decaying present value of future returns).

There are a 8 possible round outcomes in a game-of-choice. Since deterministic strategies always respond the same way to the same game state, the sequence of round outcomes must contain a cycle within the first 8 rounds when two deterministic strategies play one another, and this cycle must repeat itself *ad infinitum.* The value of the cumulative discounted payoff can therefore be calculated by:

- Identifying the cycle in the first 8 rounds,
- Computing the sum of the pre-cycle discounted payoffs,
- Computing the sum of the infinite tail of cyclic payoffs, which is a geometric series.

In my equilibrium analysis of pure game-of-choice strategies (Chapter 4), I define $E[u_i(s_i, s_j)]$ as individual i's *average* discounted payoff per round rather than as his *cumulative* discounted payoff. The average is simply a scalar adjustment - dividing the cumulative discounted payoff by the constant $\frac{1}{1-\delta}$ - that makes the mathematics for the equilibrium analysis slightly nicer. In addition, this choice results in a nice harmony between the two definitions of $E[u_i(s_i, s_j)]$: in the limit $\delta \to 1$ (i.e. as the continuation probability approaches 100%), both expressions represent the average payoff an individual receives per round in an infinitely-repeated game. Using the interpretation of δ as the game continuation probability, the constant factor $\frac{1}{1-\delta}$ represents the expected number of rounds:

$$\begin{aligned}
\text{E[num rounds]} &= \sum_{k=1}^{\infty} k \cdot \text{Prob[game lasts exactly k rounds]} \\
&= \sum_{k=1}^{\infty} k \cdot (\delta^{k-1}(1-\delta)) \\
&= \frac{1}{1-\delta}.
\end{aligned}$$

Note that the first round (Round 0) always occurs, so the number of rounds ranges from $k = 1$ to ∞. The probability that a game lasts exactly k rounds is the probability that it is continued $k-1$ times and is not continued the kth time, i.e. $\delta^{k-1}(1-\delta)$.

Individual i's cumulative discounted payoff is $\sum_{k=0}^{\infty} u_i(o^k) \cdot \delta^k$. As the average number of rounds is $\frac{1}{1-\delta}$, the average discounted payoff per round is:

$$E[u_i(s_i, s_j)] = (1-\delta) \sum_{k=0}^{\infty} u_i(o^k) \cdot \delta^k.$$

Sample Calculation. Suppose the first 8 payoff values are a, b, c, d, e, c, d, e. Then the cumulative discounted payoff is $a\delta^0 + b\delta^1 + (c\delta^2 + d\delta^3 + e\delta^4)\frac{1}{1-\delta^3}$, calculated as follows:

Round	0	1	2	3	4	5	6	7	8	9	…
Payoff	a	b	C	D	E	C	D	E	C	D	…

Figure 2.1: Sequence of Round Payoffs

Pre-cycle discounted payoffs: $a\delta^0 + b\delta^1$

Infinite tail discounted payoffs:

$$= (c\delta^2 + d\delta^3 + e\delta^4) + (c\delta^5 + d\delta^6 + e\delta^7) + \ldots$$
$$= [\delta^2(c + c\delta^3 + c\delta^6 + \ldots) + \delta^3(d + d\delta^3 + d\delta^6 + \ldots) + \delta^4(e + e\delta^3 + e\delta^6 + \ldots)]$$
$$= (c\delta^2 + d\delta^3 + e\delta^4)\,(1 + \delta^3 + \delta^6 + \ldots)$$
$$= (c\delta^2 + d\delta^3 + e\delta^4)\,(1/(1- \delta^3))$$

Figure 2.2: Calculation of the Cumulative Discounted Payoff

Thus, I take two complementary approaches to defining a strategy's average payoff per round in an infinitely-repeated game:

- in my evolutionary simulations of almost-pure strategies, I define $E[u_i(s_i, s_j)]$to be the frequency-weighted average payoff per round;
- in my equilibrium analysis of pure strategies, I define $E[u_i(s_i, s_j)]$ to be the average discounted payoff per round.

2.3 Strategy Space Definitions

A strategy in a game of choice is a decision rule, fixed in advance, that consists of two main components:

1. a choice rule, describing which stage game the player prefers to interact in, and
2. an action rule, describing the action the player will take in each stage game.

To gain intuition for the mechanisms at work in this model, I focus on some of the most fundamental strategy spaces: pure and almost-pure memory-1 strategies.

As mentioned in the previous section, pure strategies are deterministic. Almost-pure strategies are *almost* deterministic: actions that would have been certain to occur (probability 1) are now taken with probability $1 - \epsilon$; actions that would never have been taken (probability 0) are now taken with probability ϵ, where the noise level ϵ is a small positive value near zero. In addition to being a more realistic model of strategies in nature, incorporating $\epsilon > 0$ also causes the probability of taking any action to be intermediate between 0 and 1. As a result, there is always a positive probability of directly transitioning from round outcome o_1 to round outcome o_2 for all round outcomes $o_1, o_2 \in O$. Hence, incorporating inherent noise results in a unique stationary distribution and allows players' payoffs to be calculated in a manner independent of the initial subgame state.

Memory-1 strategies determine their next move based on the outcome of the prior round. For simplicity, I assume that a strategy's first move – before there is a prior round – is cooperation. Neither the evolutionary simulations nor the subgame-perfect equilibrium analysis is affected by the strategy's first move: the payoffs in the evolutionary simulations are independent of the initial game state, and SPE analysis explicitly considers the effect of all possible alternative moves.

In describing the strategy spaces in my model, it is useful to distinguish between the *simple* round outcome, as in CC, and the *compound* round outcome, as in $1CC$ or $2CC$. The simple round outcome specifies only the cooperation/defection result, while the compound round also specifies the stage game in which the result occurred.

I consider strategy spaces in which the action rules for Game 1 and Game 2 depend only on the simple round outcome. This enables a direct comparison to be made between game-of-choice strategies and memory-1 strategies in the IPD: in both, the action taken within a PD stage game depends only on the cooperation/defection outcome of the prior round.

2.3.1 S16 Strategy Space

I consider a 16 dimensional strategy space, in which strategies take the following form:

$$\begin{aligned} s = [&p_{1CC}, p_{1CD}, p_{1DC}, p_{1DD}, \quad p_{2CC}, p_{2CD}, p_{2DC}, p_{2DD}; \\ &x_{1CC}, x_{1CD}, x_{1DC}, x_{1DD}, \quad x_{2CC}, x_{2CD}, x_{2DC}, x_{2DD}] \in \{\epsilon, 1-\epsilon\}^{16}, \end{aligned}$$

where

$$\begin{aligned} p_{1ab} &= \text{action rule in Game 1} = \text{Prob[play C}|\text{in Game 1 and previous round's simple outcome} = ab], \\ p_{2ab} &= \text{action rule in Game 2} = \text{Prob[play C}|\text{in Game 2 and previous round's simple outcome} = ab], \\ x_{iab} &= \text{choice rule} = \text{Prob[prefer Game 1}|\text{previous round's compound outcome} = iab]. \end{aligned}$$

Strategies in this subspace may act differently in Game 1 and in Game 2, and may nuance their stage game choice depending on whether the previous cooperation/defection result occurred in Game 1 or in Game 2.

2.3.2 S12 Strategy Space

Strategies that earn high payoffs presumably spend most of their time playing in the high-reward Game 1. A natural question is: is nuancing the stage game choice an important element in maintaining high Game 1 payoffs?

To investigate this question, I consider a 12-dimensional strategy space in which the choice rule depends on the simple, rather than the compound, outcome of the previous round. In this strategy space, strategies take the following form:

$$\begin{aligned} s = [&p_{1CC}, p_{1CD}, p_{1DC}, p_{1DD}, \quad p_{2CC}, p_{2CD}, p_{2DC}, p_{2DD}; \\ &x_{CC}, x_{CD}, x_{DC}, x_{DD}] \in \{\epsilon, 1-\epsilon\}^{12}, \end{aligned}$$

where

$$\begin{aligned} p_{1ab} &= \text{action rule in Game 1} = \text{Prob[play C|in Game 1 and previous round's simple outcome} = ab], \\ p_{2ab} &= \text{action rule in Game 2} = \text{Prob[play C|in Game 2 and previous round's simple outcome} = ab], \\ x_{ab} &= \text{choice rule} = \text{Prob[prefer Game 1|previous round's simple outcome} = ab]. \end{aligned}$$

The only difference between S12 and S16 is that the choice rule in S12 depends on the simple, rather than compound, outcome of the previous round.

2.3.3 S8 Strategy Space

Further, we can ask whether strategies that spend most of their time in Game 1 would be negatively impacted if they must act the same way in Game 1 and in Game 2. In other words, is the action rule that is good for the goose (Game 1) also good for the gander (Game 2)?

To investigate this question, I consider an 8-dimensional strategy space in which there is only a single action rule for both Game 1 and Game 2. As in S12, the choice rule depends only on the simple outcome of the prior round. In this strategy space, strategies take the following form:

$$s = [p_{CC}, p_{CD}, p_{DC}, p_{DD}; \quad x_{CC}, x_{CD}, x_{DC}, x_{DD}] \in \{\epsilon, 1-\epsilon\}^{8},$$

where

- $p_{ab} = Prob[\text{play C}|\text{simple prior outcome} = ab]$, and

- $x_{ab} = Prob$[prefer Game 1|simple prior outcome $= ab$].

2.4 Game-of-Choice Transition Matrix

The game-of-choice transition matrix depends on both the strategies' action probabilities and on the game-of-choice resolution rule.

For strategy s_1, let

- p_{iab} = Prob[play C | in Game i and previous round's simple outcome $= ab$];
- x_{iab} = Prob[prefer Game 1 | previous round's compound outcome $= iab$].

In strategy spaces S8 and S12, the choice rule $x_{iab} = x_{1ab} = x_{2ab}$; in strategy space S8, the action rule $p_{iab} = p_{1ab} = p_{2ab}$ as well. Analogously, for strategy s_2, let

- q_{iab} = Prob[play C | in Game i and previous round's simple outcome $= ab$];
- y_{iab} = Prob[prefer Game 1 | previous round's compound outcome $= iab$].

Then the generic 8×8 game-of-choice transition matrix has the following form:

	1CC	1CD	⋯	2DC	2DD
1CC	f(x1cc, y1cc) ·p1cc·q1cc	f(x1cc, y1cc) ·p1cc·(1 - q1cc)		(1 - f(x1cc, y1cc)) ·(1 - p2cc)·q2cc	(1 - f(x1cc, y1cc)) ·(1 - p2cc)·(1 - q2cc)
1CD	f(x1cd, y1dc) ·p1cd·q1dc	f(x1cd, y1dc) ·p1cd·(1 - q1dc)		(1 - f(x1cd, y1dc)) ·(1 - p2cd)·q2dc	(1 - f(x1cd, y1dc)) ·(1 - p2cd)·(1 - q2dc)
⋮					
2DC	f(x2dc, y2cd) ·p1dc·q1cd	f(x2dc, y2cd) ·p1dc·(1 - q1cd)		(1 - f(x2dc, y2cd)) ·(1 - p2dc)·q2cd	(1 - f(x2dc, y2cd)) ·(1 - p2dc)·(1 - q2cd)
2DD	f(x2dd, y2dd) ·p1dd·q1dd	f(x2dd, y2dd) ·p1dd·(1 - q1dd)		(1 - f(x2dd, y2dd)) ·(1 - p2dd)·q2dd	(1 - f(x2dd, y2dd)) ·(1 - p2dd)·(1 - q2dd)

Figure 2.3: Game-of-Choice Transition Matrix

The function $f(a, b)$ determines the probability of the interaction taking place in Game 1, given that s_1 prefers Game 1 with probability a and s_2 prefers Game 1 with probability b.

The function f depends on the particular resolution rule:

Random	$f(a, b) = 0.50$
Equal-Say Game-2 default	$f(a, b) = \text{Prob[both prefer Game 1]} = a \cdot b$
Equal-Say Game-1 default	$f(a, b) = \text{Prob[at least one prefers Game 1]} = a + b - a \cdot b$
Unilateral Dictator	
When s_1 is the dictator ...	$f(a, b) = a$
When s_2 is the dictator ...	$f(a, b) = b$
Random Dictator	$f(a, b) = \text{Prob[the chosen dictator prefers Game 1]} = \frac{1}{2}a + \frac{1}{2}b$

The Unilateral Dictator dynamic presents a slight difficulty because the players are asymmetrical. To circumnavigate the issue, I define the stationary distribution in this case as the average of the stationary distribution when s_1 is dictator and that when s_2 is dictator: $v = \frac{1}{2}v_{s_1 \text{ dictator}} + \frac{1}{2}v_{s_2 \text{ dictator}}$. This stationary distribution represents the expected proportion of time spent in each game state when either player may be randomly chosen as the Unilateral Dictator.

2.5 Equilibrium Definitions

In large strategy spaces, it is unlikely that one of the strategies is always a better choice than all the other strategies in the strategy space [29]: a strategy's performance depends on its opponent. However, players are expected to reach an equilibrium state in which every individual is playing a best-response strategy with respect to the strategies everyone else is playing. In this equilibrium state, no individual has incentive to unilaterally deviate from his chosen strategy. If deviating to another strategy strictly lowers an individual's payoff, the equilibrium is called a strict Nash equilibrium. If deviating to another strategy cannot improve an individual's payoff, the equilibrium is called a weak Nash equilibrium. The formal definitions for strict and weak Nash equilibrium in a two-player game are as follows:

Definition 1. A strategy profile (s_i, s_j) is a *strict Nash equilibrium* if, and only if,

- for any alternative strategy $\tilde{s}$ that individual i adopts, $E[u_i(\tilde{s}, s_j)] < E[u_i(s_i, s_j)]$, and
- for any alternative strategy $\tilde{s}$ that individual j adopts, $E[u_j(s_i, \tilde{s})] < E[u_j(s_i, s_j)]$.

The strict Nash equilibrium condition is often considered to be too stringent, as multiple strategies can behave the same way on the equilibrium path of play, resulting in the same payoffs. As a result, a Nash equilibrium is typically understood to mean a weak Nash equilibrium:

Definition 2. A strategy profile (s_i, s_j) is a *(weak) Nash equilibrium* if, and only if,

- for any alternative strategy $\tilde{s}$ that individual i adopts, $E[u_i(\tilde{s}, s_j)] \leq E[u_i(s_i, s_j)]$, and
- for any alternative strategy $\tilde{s}$ that individual j adopts, $E[u_j(s_i, \tilde{s})] \leq E[u_j(s_i, s_j)]$.

Both the strict and the weak Nash equilibrium concept are defined in terms of their payoffs on the equilibrium path of play. A drawback of this approach is that strategies may specify non-optimal actions (or non-credible threats) for game states not on the equilibrium path. If those game states were to be reached, following through on the specified course of action would not be in the individual's best interest (some other course of action would result in higher expected payoffs).

Consequently, a subgame-perfect equilbrium is a refinement to the Nash equilibrium concept; it stipulates that strategies specify best-response actions in all games states (thus resolving the problem of non-credible threats).

Definition 3. A strategy profile (s_i, s_j) is a *subgame-perfect equilibrium* if, and only if, the strategy profile is a weak Nash equilibrium starting from any subgame state.

The Single-Deviation Principle provides a shortcut way to determine whether a strategy profile is subgame-perfect without considering the entire space of alternative strategies.

A *single deviation* from strategy s is a strategy $\tilde{s}$ that deviates from the action s_i would play only in a single round; for all other rounds, it acts as strategy s_i acts. The single deviation strategy $\tilde{s}$ is *useful* if there exists a subgame state such that, starting from that initial state, an individual can improve his payoff by playing a single deviation strategy $\tilde{s}$ instead of strategy s.

Theorem 1. ***Single Deviation Principle.*** *A strategy profile (s_i, s_j) is a subgame-perfect equilibrium if, and only if, there is no useful, single deviation.*

Subgame perfect equilibria are considered to be more sensible than Nash equilibria because they contain only credible threats.

Because the game-of-choice is symmetric from both players' perspectives, it is natural to focus on equilibria in which both players are assigned the same strategy. In Chapter 4, I focus my equilibrium analysis on symmetric subgame-perfect equilibria.

Another refinement to the Nash equilibrium concept is that of evolutionary stability. Weak Nash equilibria are unstable because they can be broken by individuals switching to neutral alternative strategies. However, if s is an evolutionary stable strategy, then (s, s) is a weak Nash equilibrium in which natural selection opposes the introduction of any alternative strategy. If there exists an alternative strategy $\tilde{s}$ that is also a best-response to strategy s, evolutionary stability requires the current strategy to perform better against the alternative than the alternative performs against itself. Formally:

Definition 4. A strategy s is evolutionary stable if, and only if, for any alternative strategy $\tilde{s}$,

- $E[u_i(\tilde{s}, s)] < E[u_i(s, s)]$, or
- $E[u_i(\tilde{s}, s)] = E[u_i(s, s)]$ and $E[u_i(\tilde{s}, \tilde{s})] < E[u_i(s, \tilde{s})]$

Note that the condition for evolutionary stability is between that of a strict and a weak Nash equilibrium: if (s, s) is a strict Nash equilibrium, then s is evolutionary stable; if s is evolutionary stable, then (s, s) is a weak Nash equilibrium.

The perspective of evolutionary stability also provides a practical approach – directly simulating the stochastic mutation/selection process of evolution – to finding the equilibria that most successfully balance:

1. Robustness to invasion attempts by alternative strategies, and
2. Selection pressure for the high payoffs resulting from mutual cooperation.

In the next section, I delve into specifics of the selection/mutation process I use.

2.6 Evolutionary Dynamics

I model evolution as a pairwise imitation process in which mutation events occur rarely [13, 25]. Conceptually, two timescales are at play in this model: the short-term timescale and the long-term timescale. The short-term perspective underlies the calculation of individual fitness; the long-term perspective underlies the calculus of population adaption.

2.6.1 Individual Fitness

In the short-term perspective every individual i in the population plays according to a fixed strategy s_i. An individual's fitness is defined relative to the current population: an individual's fitness is the average of his expected payoffs when playing against the other individuals in the population. Let's denote individual i's fitness by π_i. Then

$$\pi_i = \frac{1}{N-1} \sum_{j \in \mathcal{N} \setminus \{i\}} E(u_i(s_i, s_j)).$$

2.6.2 Individual Adaption

The short-term perspective enables individual fitness to be calculated. However, in the long-term, individuals can learn to play better strategies; selection pressure drives individuals toward strategies with higher fitness values (i.e. higher expected payoffs).

How do individuals adapt their strategies? In a pairwise imitation process, two individuals are sampled at random. One individual is the "learner" and the other individual is the "rolemodel." Let π_l denote the fitness of the learner's strategy and let π_r denote the fitness of the rolemodel's strategy. Then the learner imitates the rolemodel's strategy with probability $\sigma(\beta, \pi_r, \pi_l) = \frac{1}{1+e^{-\beta(\pi_r - \pi_l)}}$. The term $\beta > 0$ is the selection pressure; it plays

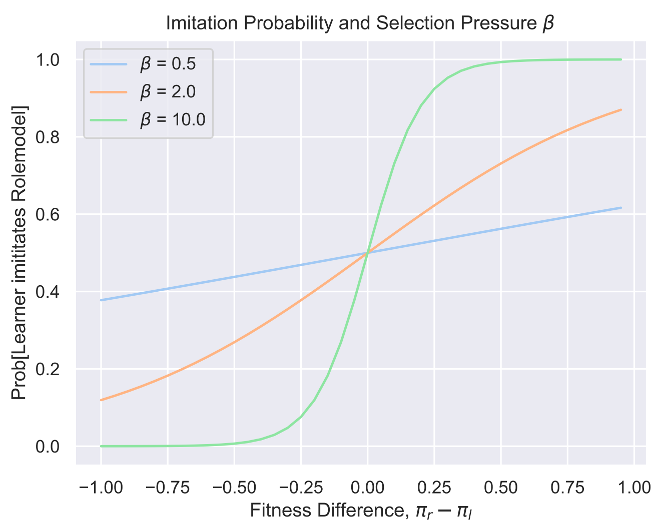


Figure 2.4: Effect of Selection Pressure on the Imitation Probability

an analogous role to temperature in the simulated annealing algorithm. High values of β make it unlikely for a learner to switch to a strategy with lower fitness; low values of β encourage exploration of more strategies. As $\beta \to 0$, the imitation probability simplifies to $\sigma(\beta, \pi_r, \pi_l) = \frac{1}{2}$, and strategy changes are purely random. On the other hand, as $\beta \to \infty$, the imitation probability simplifies to $\sigma(\beta, \pi_r, \pi_l) = \begin{cases} 1 & \text{if } \pi_r > \pi_l \\ 0 & \text{if } \pi_r < \pi_l \\ \frac{1}{2} & \text{if } \pi_r = \pi_l \end{cases}$, and learners always adopt strategies with higher fitness values.

2.6.3 Population Adaption

Suppose that individuals rarely mutate their strategies, and that mutations appear one at a time. In this case, when a mutant arises, it either goes extinct or fixates in the population before another mutant appears. Consequently, the population can be thought of as passing through a series of homogeneous states, where state s corresponds to a homogeneous population all playing strategy s. The transition probability from state s to state $\tilde{s}$ is the probability that a single mutant playing strategy $\tilde{s}$ fixates in a population in which the other players are all playing strategy s. The evolutionary trajectory of population states is thus a random walk on a Markov chain. In evolutionary dynamics with finite β, a mutant strategy always has some positive probability of fixating in the population, resulting in an irreducible Markov chain with a unique stationary distribution that is independent of the initial state. The stationary distribution describes the proportion of time the population spends playing each strategy, and thus can be used to determine the proportion of time the population spends in cooperative states.

How is the transition probability $Q(s, \tilde{s})$ from state s to state $\tilde{s}$ calculated?

- Let N be the population size,
- Let T_j^+ be the probability that the number of individuals playing mutant strategy $\tilde{s}$ increases from j to $j+1$, and
- Let T_j^- be the probability that the number of individuals playing mutant strategy $\tilde{s}$ decreases from j to $j-1$.

As Traulsen & Hauert derive in [2],

$$Q(s, \tilde{s}) = \frac{1}{1 + \sum_{k=1}^{N-1} \Pi_{j=1}^{k} \frac{T_j^-}{T_j^+}}. \tag{2.1}$$

In the pairwise imitation process,

$$\begin{aligned}
T_j^+ &= \text{Prob[a non-mutant is chosen as a learner;}\\
&\qquad \text{then a mutant is chosen as a rolemodel;}\\
&\qquad \text{and the learner imitates the rolemodel]}\\
&= \frac{N-j}{N}\frac{j}{N-1}\sigma(\beta, \pi_s, \pi_{\tilde{s}})\\
T_j^- &= \text{Prob[a mutant is chosen as a learner;}\\
&\qquad \text{then a non-mutant is chosen as a rolemodel;}\\
&\qquad \text{and the learner imitates the rolemodel]}\\
&= \frac{j}{N}\frac{N-j}{N-1}\sigma(\beta, \pi_{\tilde{s}}, \pi_s)
\end{aligned}$$

Let $u_{\tilde{s}s}$ be shorthand for $E[u_{\tilde{s}}(\tilde{s}, s)]$. Further manipulation of Eq. (2.1) yields:

$$Q(s, \tilde{s}) = \sum_{k=1}^{N-1} \frac{1}{e^{-\beta k[\frac{k+1}{2u}+v]}}, \text{ where} \tag{2.2}$$

$$\begin{aligned}
a, b, c, d &= u_{\tilde{s}\tilde{s}}, u_{\tilde{s}s}, u_{s\tilde{s}}, u_{ss}\\
u &= (a-b-c+d)/(N-1)\\
v &= (-a+bN-dN+d)/(N-1).
\end{aligned}$$

2.6.4 Stochastic Evolutionary Process

Since I am interested in how cooperation arises, my evolutionary simulation begins in state $ALLD_2 := [0\ldots0; 0\ldots0]$. The subscript refers to the default preference for Game 2. Individuals playing strategy $ALLD_2$ are minimally cooperative: cooperation and preferring Game 1 occur only because of inherent noise. Note that the initial state has little effect on the long-term average cooperation rates: the evolutionary process simulates an ergodic Markov chain, and so eventually converges to the unique stationary distribution regardless of the initial state.

At every time step, a mutant strategy is uniformly sampled from the discrete strategy space. (Since the strategy spaces I consider are memory-1 (8 game states) and pure/almost-pure (2 possibilities for each strategy probability), my strategy spaces have finite size. In particular, S8, S12, and S16 have size $2^8, 2^{12}$, and 2^{16}.)

With probability equal to the fixation probability, the randomly generated mutant fixates in the population and we transition to a state in which everyone in the population is playing the mutant strategy.

I simulate the above stochastic process for T timesteps, and obtain a sequence of population states, $s_1, ..., s_T$.

On the basis of this sequence of states, I calculate the average proportion of time that the population spends cooperating with one another.

2.6.5 Calculation Details

In each state s_i, all individuals play strategy s_i. Thus the proportion of time that the population spends in each game state is described by the stationary distribution $\vec{v} = [v_{1cc}, v_{1cd}, v_{1dc}, v_{1dd}, v_{2cc}, v_{2cd}, v_{2dc}, v_{2dd}]$ that occurs when strategy s_i plays s_i. Consequently, individuals in this population state cooperate with frequency

$$\text{C rate} = \frac{1}{2}(2v_{1cc} + v_{1cd} + v_{1dc} + 2v_{2cc} + v_{2cd} + v_{2dc}).$$

The average proportion of time that the population spends cooperating is obtained by averaging the cooperation rates over all the states s_i that occur in the stochastic process.

Chapter 3

Evolution of Cooperation

3.1 Experimental Variables

In the standard IPD model, the payoff matrix can be parameterized by b, the benefit cooperation provides to the co-player, and c, the personal cost of cooperation. In the game-of-choice model, the payoff matrix is parameterized by b_1, b_2 and c, where:

- b_1 is the benefit cooperating provides to the co-player in Game 1,
- b_2 is the benefit cooperating provides to the co-player in Game 2, and
- c is the personal cost of cooperating in both Game 1 and Game 2.

In the IPD, reciprocal strategies stabilize cooperation by punishing unilateral deviations from cooperation with defection. In games of choice, an additional possibility exists: unilateral deviations from cooperation can also be punished with time in Game 2, the low-reward setting. A natural hypothesis is that the "Game 2 as additional punishment" mechanism acts to promote cooperation. In this chapter, I put this hypothesis to the test.

3.2 Experimental Setup

I compare cooperation rates in games-of-choice across the five resolution rules and the three strategy spaces described in Chapter 2. Each strategy space/resolution rule combination describes a "meta" game-of-choice. For each "meta" game-of-choice I map how cooperation rates respond to the specific Game 1/Game 2 parameters b_1, b_2, and c. In particular, I vary b_1 and hold the other parameters $b_2 = 1.2$ and $c = 1.0$ fixed. I vary b_1 in intervals of 0.14

for a total of 16 data points ranging from $b_1 = 1.0$ to $b_1 = 3.10$. Within each particular game-of-choice, I simulate the stochastic process described in Section 2.6 on page 27 for $T = 5 \times 10^5$ timesteps and the following conventional evolutionary parameters ([13, 15]):

Parameter	Description	Value
N	Population size (large)	100
β	Selection Pressure (intermediate)	2.0
ε	Inherent Noise (low)	0.001

Figure 3.1: Key Parameters

To compare the game-of-choice with the IPD, I also map the evolution of cooperation in an IPD with $c = 1$ fixed and varying benefit $b_1 \in [1.0.3.10]$.

All evolutionary simulations were repeated five times. Across five runs, the standard deviation of the cooperation rate is generally less than 0.05. In Appendix A.1 on page 53, I test the robustness of my results with respect to the evolutionary parameters N, β, and ϵ.

In the following discussion I use the following abbreviations for the resolution rules:

- EqualSay Game 1 Default → Eq_G1
- EqualSay Game 2 Default → Eq_G2
- Unilateral Dictator → UniD
- Random Dictator → RandD

3.3 Resolution Rule/Strategy Space Hypotheses

I expect that resolution rule Eq_G2 Default will be most effective in promoting cooperation because it is the only resolution rule that enables individuals to directly punish defection with time in Game 2 (the low-reward setting). On the other hand, I predict that cooperation rates under Eq_G1 will be similar to those in the baseline IPD. I predict this because non-cooperative strategies can unilaterally force all interactions to occur in Game 1 in a game-of-choice with resolution rule Eq_G1. If all interactions happen in the same stage game, the result should be equivalent to an IPD of that stage game. I expect that Random will act analogously to an IPD in which the benefit of cooperation in the stage game is the average of b_1 and b_2. Consequently, Random should lag behind the baseline IPD with benefit of cooperation b_1.

Because resolution rules RandD and UniD have an attenuated ability to punish defection with time in Game 2 compared to Eq_G2, I expect their cooperation rates to be in between

Eq_G2 and the baseline IPD. However, it is unclear whether cooperation is better promoted by RandD, in which a dictator is randomly chosen each round, or by UniD, in which the randomly chosen dictator remains in power for all rounds.

Among the three strategy spaces, I expect that S16, being a superset of S12 and S8, will be most effective in punishing defection and promoting cooperation. However, will strategies in S12, whose choice rule is based on the simple round outcome, be as capable at promoting cooperation as strategies in S16? The comparison between S8 and S12 should also shed light on whether it is important to act differently in Game 1 and in Game 2.

While I expect the game-of-choice to promote cooperation rates, the experimental results will give concrete estimates as to the magnitude of the effect.

3.4 Experimental Results

Fig. 3.2 shows how cooperation rates evolve in each of the strategy spaces and under each of the described resolution rules.

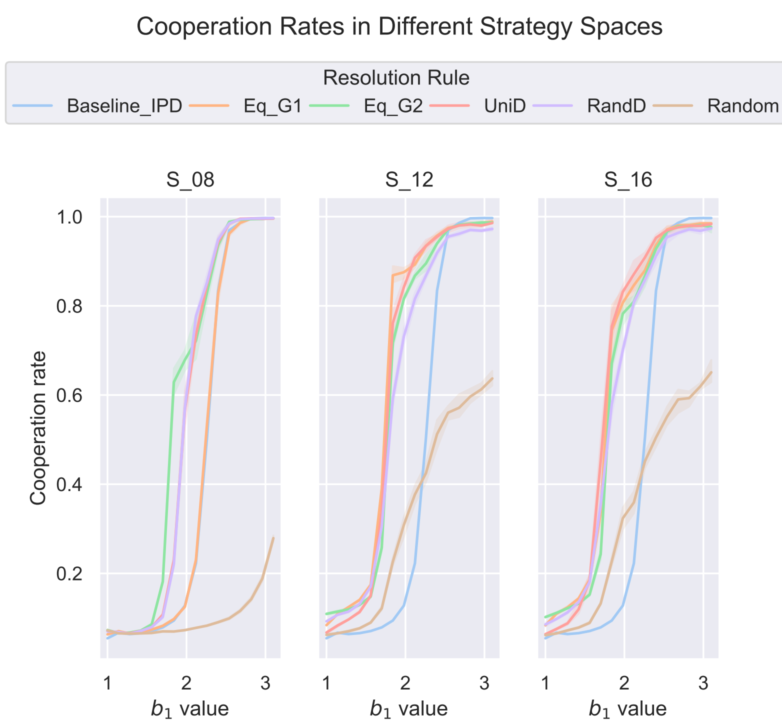


Figure 3.2: Cooperation Evolution

In all three strategy spaces and under most resolution rules, cooperation arises significantly earlier in games-of-choice when compared to the baseline IPD.

One universally observed result is that the Eq_G2 resolution rule, along with the two dictator rules, always outperforms the baseline IPD. In Fig. 3.3, I illustrate this enhanced cooperation phenomenon by comparing cooperation rates between:

- an IPD game with parameters $b = 1.8, c = 1.0$,
- an IPD game with parameters $b = 1.2, c = 1.0$,
- a Game-of-Choice between the above two PD games, $b_1 = 1.80, b_2 = 1.20, c = 1.0$, under resolution rule Eq_G2.

Fig. 3.3 shows the cooperation rate at each timestep, averaged over 50 independent runs. In the IPD games, cooperation rates are capped at 20%. In the games of choice, cooperation rates hover in the 40-60% range. A game-of-choice between two PD games clearly results in cooperation rates above and beyond what can be achieved by either IPD game alone.

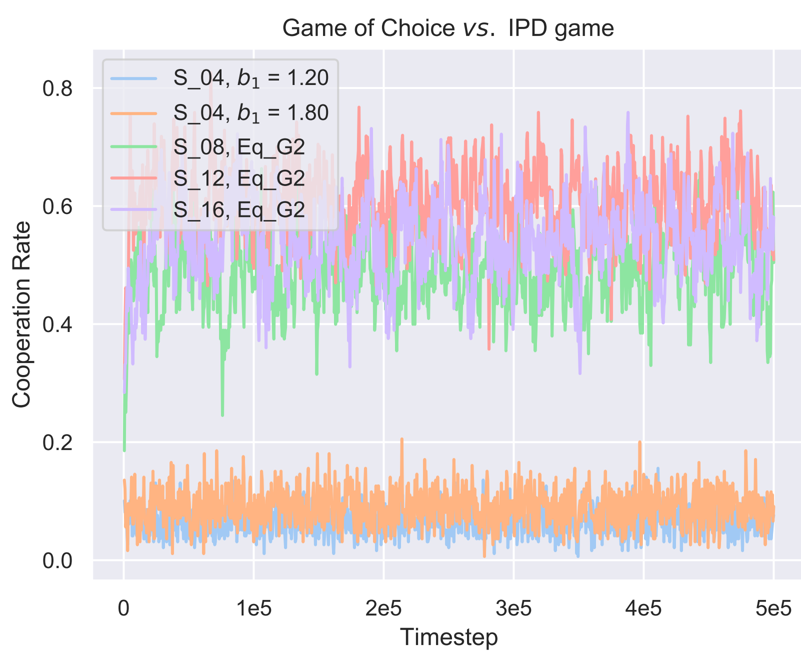


Figure 3.3: Game-of-Choice vs. IPD of Game 1 vs. IPD of Game 2.

The difference in cooperation rates between the games-of-choice and the baseline IPD is clear. However, is harder to discern the relative ranking of resolution rules within each strategy space and how resolution rule performance varies between strategy spaces. In order to assess these differences, I focus on a fixed b_1 slice in Fig. 3.2 on page 34. Fig. 3.4 on the following page shows the cooperation rates at the fixed value of $b_1 = 1.84$. The mean and standard deviation of the cooperation rate across the five repeat runs are displayed as well.

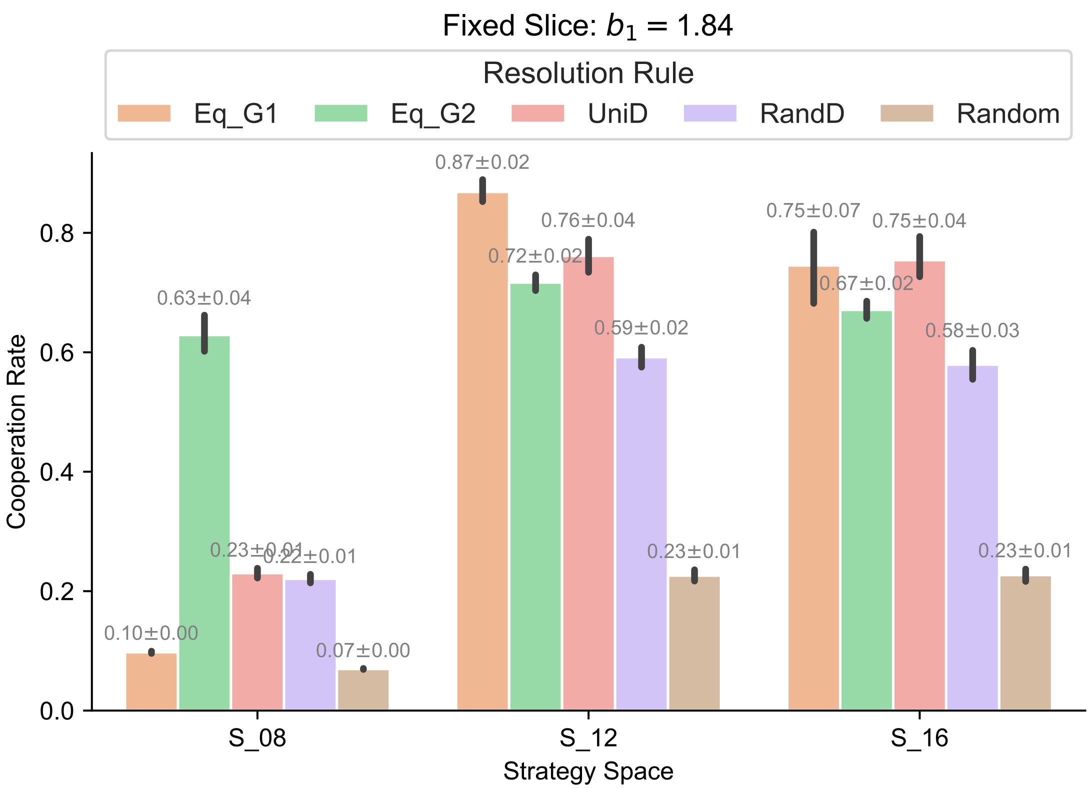


Figure 3.4: Cooperation Rate Comparison

3.5 Resolution Rule Performance

3.5.1 Resolution Rule Performance in S8

In S8, the resolution rules performed as predicted. Referring to Fig. 3.2 on page 34 and Fig. 3.4, we see that Eq_G2 was the best-performing resolution rule, and that Eq_G1 exhibited the same dynamics as the baseline IPD. UniD and RandD's performances fell in-between these two. As expected, Random performed worse than the baseline IPD. The Random curve appears to be stretched horizontally compared to the baseline IPD. This makes sense, since I expect the Random cooperation rates to follow an IPD curve with benefit of cooperation $\bar{b} = \frac{1}{2}b_1 + \frac{1}{2}b_2$. As b_1 increases, so does the gap between b_1 and $\bar{b}$. In the next chapter, I investigate the strategies contributing to Eq_G2's success in S8.

3.5.2 Resolution Rule Performance in S12/S16

First, S12 and S16 exhibit very similar dynamics between themselves. In both, the resolution rules share the same relative ranking (Eq_G1 > UniD > Eq_G2 > RandD > Random). Moreover, the precise value of the cooperation rates are generally within 0.01 of each other, with Eq_G1 being the sole exception. However, that exception may be due to the singularly high standard deviation observed for this resolution rule in S16. If my computing capacity were increased, I would replicate this experiment using 10 runs or more in order to check whether the anomaly disappears under less variable experimental conditions.

Although the resolution rules behave consistently in S12 and S16, their performances are unexpected. I hypothesized that resolution rules promote cooperation in relation to their ability to punish defection with Game 2. Under this hypothesis, I expected Eg_G2 > UniD/RandD > Eq_G1 = baseline IPD. Contrary to expectations, Eq_G1 outperformed Eq_G2 and all the other resolution rules. In the next chapter, I investigate the strategies contributing to Eq_G1's success in S12/S16.

Other surprises include Random's altered cooperation trajectory; for $b_1 \in [1, 2]$, the Random resolution rule even outperforms the baseline IPD. Reassuringly, however, Random performs the same between S12 and S16, as expected. S12 and S16 differ in the strategies' choice rules; however, under the choice rule is irrelevant the Random resolution rule as the stage game is chosen randomly.

Another interesting result is that UniD outperforms Equal-Say G2 default and RandD in S12/S16.

The drastic increase in Eq_G1's performance signals that a cooperation-promoting mechanism exists in S12/S16 that doesn't exist in S8. Hence, acting differently in Game 1 and Game 2 is critical to establishing cooperation under resolution rule Eq_G1. It is also worth noting that cooperation also skyrockets for the dictator resolution rules, suggesting that strategies under these resolution rules also rely on acting differently in Game 1 and Game 2. The increase in the Random resolution rule's performance is also likely due to the increased ability for strategies to act differently in Game 1 and in Game 2. However, it is surprising that players can achieve higher cooperation rates when they are randomly assigned to play a high-reward or low-reward PD than when they repeatedly play either PD game alone.

The modest increase in Eq_G2's performance suggests that it might not be caused by a radically different cooperation mechanism, but rather by factors related to the increase in the size of the strategy space. This hypothesis is also supported by evidence that I will present later.

To summarise, the key takeaways are:

- In all three strategy spaces, a game-of-choice equipped with an appropriate resolution rule can enhance cooperation rates relative to the baseline IPD. A game-of-choice in even the simplest strategy space (S8) is still able to enhance cooperation relative to the IPD baseline. It is worth reiterating that strategies in S8 lack all ability to distinguish between Game 1 and Game 2 outcomes: they have a single action rule for playing both Game 1 and Game 2, and they choose their preferred stage game based only on the previous round's simple outcome.

- In S8 the performance of the resolution rules is consistent with the hypothesis that these strategies use Game 2 to punish defection. In S12 and S16 the performance of the resolution rules suggests that a different mechanism may be at play. In particular, this mechanism relies on acting differently in Game 1 and in Game 2.

- Consequently, there is added value in being able to act differently in Game 1 and in Game 2. On the other hand, S12 and S16 share remarkably similar cooperation rates across the board, so it seems as though there is limited value in choosing the preferred stage game differently for Game 1 and Game 2 outcomes.

How do these experimental results relate to the motivating examples described in Section 2.1.1 on page 16? The results can be interpreted as follows:

There is a difference between cooperating on minor issues and on major issues (S8 vs. S12). But even if one doesn't distinguish between minor and major issues (S8), cooperation can still be enhanced relative to the baseline IPD thorough the introduction one of the Eq_G2, RandD, or UniD resolution rules.

Surprisingly, the experimental results show that if one can act differently in Game 1 and Game 2, and if the the benefit of cooperating on high-importance issues is not too high, then strategies exist for debating low-importance and high-importance issues beyond both players' control (i.e. the Random resolution rule) that actually promote cooperation relative to a baseline diet of either only low-important issues or high-important issues.

Chapter 4

Subgame-Perfect Equilibrium Analysis

Now we turn to the question of *why*: why does cooperation arise earlier in games of choice? What specific strategies are behind the observed cooperation rates, and how do they work? In this chapter, I offer an equilibrium analysis for two most successful resolution rules, Eq_G1 and Eq_G2. The analytical approach I use for analyzing these two resolution rules can be applied to analyzing the other resolution rules as well.

4.1 Hypotheses: Properties of Cooperation-Stabilizing Strategies

The strategies in the experiments are ϵ-noisy approximations of pure strategies. What are the underlying pure strategies that stabilize cooperation?

I posit that such cooperation-stabilizing strategies exhibit three key features:

- First, in order to qualify as cooperative, the natural result when this strategy plays itself should be repeated high-reward mutual cooperation: the equilibrium path of round outcomes should be 1CC repeated infinitely.
- Second, in order to be robust to inherent noise, these strategies should be able to return to cooperation even in the presence of mistakes: 1CC should be an absorbing state when play proceeds from any initial state.
- Third, in order to be enduring, these strategies should be robust to invasion attempts

by other strategies. As no pure strategy is evolutionary stable in the IPD [29], I approximate evolutionary stability with subgame-perfect stability.

I refer to strategies matching the above description as CoopSPE strategies.

4.2 Finding CoopSPE Strategies

Do CoopSPE strategies exist, and if so, what do these strategies look like? To investigate this question, I wrote a program $findCoopSPE(g, \delta)$ that outputs the set of cooperative SPE strategies for the given game of choice g and discount factor $\delta \in (0, 1)$. Since this analysis focuses on pure strategies, I use a discount factor to calculate the average payoff per round. In particular, I use $\delta = 1.0 - 10^{-5}$ in order to approximate the limit $\delta \to 1$. This limit, approaching the infinitely-repeated game, corresponds to the average payoffs per round calculated in the experimental setting. The pseudo-code for computing the CoopSPE strategy set is located in Appendix A.2 on page 56.

I ran my $findCoopSPE$ program on games-of-choice equipped with resolution rules Eq_G1 and Eq_G2. I tested the following, already familiar, stage game parameter combinations: $c = 1.0$, $b_2 = 1.2$, $b_1 \in [1.0; 3.0]$ (increments of 0.14). The following table shows the number of CoopSPE strategies in a game-of-choice with the given resolution rule and strategy space.

b_1 value	S8	S12	S16
[1.00, 1.56]	0	0	0
[1.70, 1.98]	0	18	86
2.12	4	54	186
[2.26-2.98]	4	54	210

(a) Equal-Say Game 1

b_1 value	S8	S12	S16
[1.00, 1.56]	0	0	0
[1.70, 1.98]	4	36	108
2.12	8	72	358
[2.26-2.98]	10	136	667

(b) Equal-Say Game 2

Figure 4.1: Number of CoopSPE strategies as a function of b_1

In the IPD, WSLS is the only memory-1 cooperative SPE [18], and it becomes stable only when $b \geq 2c$. The spike in the number of CoopSPE strategies for $b_1 > 2.0$ is therefore likely to be due to the appearance of WSLS analogues. Most interestingly, cooperative SPE strategies exist at $b_1 < 2.0$ in the game-of-choice even though no cooperative SPE strategies exist in a repeated Prisoner's Dilemma of either stage game. I focus my analysis on this set of “EarlyCoopSPE” strategies that exist for $b_1 < 2c$.

Under resolution rule Eq_G1, EarlyCoopSPE exist in S12/S16 but not in S8. This suggests that EarlyCoopSPE strategies under resolution rule Eq_G1 must act differently in Game 1 and Game 2. In addition, the acquisition of EarlyCoopSPE strategies in S12 and S16 provides a concrete reason for why, in Fig. 3.4 on page 36, cooperation rates under Eq_G1 drastically improve from $10\% \pm 0\%$ in S8 to $87\% \pm 2\%$ in S12 and $75\% \pm 7\%$ in S16.

The synchronous appearance of EarlyCoopSPE strategies at $b_1 = 1.70$ suggests that the stability threshold in S12/S16 under Eq_G1 is the same as it is in S8, S12, and S16 under Eq_G2. The SPE analysis of EarlyCoopSPE strategies that follows confirms this hypothesis and derives the unexpectedly generic equilibrium condition of $2b_1 - b_2 \geq 2c$.

4.3 EarlyCoopSPE under Resolution Rule Eq_G2

4.3.1 S8 EarlyCoopSPE

S8 contains 4 EarlyCoopSPE strategies under resolution rule Eq_G2. These EarlyCoopSPE strategies are:

Strategy No.	Action Rule	Choice Rule
1.	1,0,0,1	1,0,0,0
2.	1,0,0,1	1,0,1,0
3.	1,0,0,1	1,1,0,0
4.	1,0,0,1	1,1,1,0

Figure 4.2: EarlyCoopSPE strategies in S8/Equal-Say Game 2

These strategies share a common pattern: they take the form **(1001,1??0)**; the two question marks are free variables.

The action rule is the familiar Win-Stay-Lose-Shift from IPD. As expected, the choice rule specifies Game 1 after a CC result (in order to maintain high-reward mutual cooperation). This EarlyCoopSPE strategy set can be described as follows:

- If the prior round was CC, choose Game 1 as the preferred game and cooperate.
- If the prior round was CD or DC, choose any stage game and defect.
- If the prior round was DD, force Game 2 and cooperate.

Why does the addition of the choice rule **(1??0)** to WSLS in a game-of-choice make cooperation stable when WSLS isn't stable in an IPD of either stage game?

On the equilibrium path of play, two EarlyCoopSPE players sustain high-reward mutual cooperation indefinitely. If one of the players makes a mistake (resulting in CD or DC), then cooperation is re-established in the following rounds by passing through the sequence $?DD \to 2CC \to 1CC$. For WSLS in the IPD, the error-correction sequence is $DD \to CC$.

Consequently, EarlyCoopSPE strategies in the game-of-choice impose an extra punishment on defection relative to WSLS in the IPD: defection results in an extra round in game state 2CC, which imposes an increased opportunity cost on defection. The internal mechanics of the S8/Eq_G2 EarlyCoopSPE strategy set therefore support the "Game 2 as punishment" hypothesis described in Chapter 3.

Note that adding the choice rule **(1??1)** to WSLS instead of **(1??0)** results in an error-correction sequence of $?DD \to 1CC$, which is functionally equivalent to the error-correction sequence of WSLS in an IPD of stage Game 1. Indeed, the four WSLS analogues **(1001,1??1)** are the additional CoopSPE strategies that appear at $b_1 = 2.12$. The final two CoopSPE strategies that appear at $b_1 = 2.26$ are **(1111,100?)**. At high b_1 levels, CoopSPE can even get away with always cooperating. In this case, the error-correction sequence is $2CC \to 1CC$, and the payoff difference between 1CC and 2CC alone is enough to deter defection.

The stability condition for WSLS in the IPD is $b \geq 2c$. What is the corresponding stability condition for WSLS + choice rule **(1??0)** in the game-of-choice?

Theorem 2 (S8 Eq_G2 SPE Condition)**.** *In a game-of-choice equipped with resolution rule Equal-Say Game-2 Default, strategies of the form **(1001,1??0)** are subgame-perfect in the limit $\delta \to 1$, if, and only if:*

$$2b_1 - b_2 \geq 2c.$$

The proof of Theorem 2 can be found in Appendix A.3.1 on page 57.

As a sanity check, note that this stability condition simplifies to the WSLS stability condition $b_1 \geq 2c$ when $b_1 = b_2$. This makes sense because $b_1 = b_2$ means that there is no difference between Game 1 and Game 2; there is only one effective PD game, so the stability condition is the same as in the standard IPD.

Note that the game-of-choice stability condition, $2b_1 - b_2 \geq 2c$, is *always* easier to meet than that of either corresponding IPD:

- An IPD of Game 1 has stability condition $b_1 \geq 2c$. In comparison, the game-of-choice stability condition is $b_1 + (b_1 - b_2) \geq 2c$. As $b_1 > b_2$, the game-of-choice stability condition is easier to meet.

- An IPD of Game 2 has stability condition $b_2 \geq 2c$. In comparison, the game-of-choice stability condition is $b_2 + 2(b_1 - b_2) \geq 2c$. As $b_1 > b_2$, the game-of-choice stability condition is easier to meet.

These results can be summarized as follows:

In the iterated Prisoner's Dilemma, WSLS is an SPE when $b \geq 2c$; no other cooperative SPE exist in a PD game with $c = 1.0$ and $1.0 \leq b \leq 3.0$. On the other hand, testing discrete b_1 values in a game-of-choice with $c = 1.0, b_2 = 1.2$ shows that the number of cooperative SPE is monotonically increasing for $1.0 \leq b_1 \leq 3.0$; for $b_1 = 2.98$, there are 10 distinct CoopSPE strategies.

In particular, four cooperative SPE appear once $2b_1 - b_2 \geq 2c$. This stability condition is always easier to meet than the stability conditions corresponding to either the IPD of stage Game 1 ($b_1 \geq 2c$) or the IPD of stage Game 2 ($b_2 \geq 2c$).

These four cooperative SPE strategies are WSLS combined with choice rule **(1??0)**. Although WSLS is not stable in an IPD of either stage game when $b_2 < b_1 < 2c$, WSLS + choice rule **(1??0)** may be stable in the game-of-choice; this occurs when $b_1 + (b_1 - b_2) \geq 2c$. In this case, the ability to tarry an extra round in game state 2CC stabilizes cooperation by imposing an additional opportunity cost on defection.

4.3.2 S12 EarlyCoopSPE

S12 has 36 EarlyCoopsPE strategies under resolution rule Eq_G2. These strategies have the following form: **(1ab?, ?001; 1cd0)**, where ? are free variables, $(a = 1) \Rightarrow (c = 0)$, and $(b = 1) \Rightarrow (d = 0)$.

Note that the action rule for Game 1 after a DD result is irrelevant because the choice rule forces Game 2. The relationship between a, c and b, d is suggestive:

If one's action rule in Game 1 specifies cooperation after a CD result, then the choice rule must force Game 2 after CD, resulting in defection. If one's action rule in Game 1 specifies cooperation after a DC result, then the choice rule must force Game 2 after a DC result, similarly resulting in defection. The relationship between a, c and b, d ensures that CD and DC round outcomes are always met with defection.

Hence, the Eq_G2 EarlyCoopSPE strategy set in S12 can be described as follows:

As a sanity check, these relations between a, c and b, d result in 9 possibilities:

4 possibilities: $ab = 00 \iff cd = \textbf{??}$

2 possibilities: $ab = 01 \iff cd = \textbf{?}0$

2 possibilities: $ab = 10 \iff cd = 0\textbf{?}$

1 possibility: $ab = 11 \iff cd = 00$

The two question marks in (1ab?, **?**001, 1cd0) contribute a multiplicative factor of four, yielding 36 overall EarlyCoopSPE strategies.

- If the prior round was CC, choose Game 1 as the preferred game.
  - If the stage game to be played is Game 1 (the normal result when this strategy plays itself), cooperate.
  - If the stage game to be played is Game 2, both actions are possible.
- If the prior round was CD or DC, choose any stage game and make sure to defect.
- If the prior round was DD, force Game 2 and cooperate.

The key point is that, once again, if one of the players playing this strategy makes a mistake (resulting in a CD or DC round outcome), the error-correction sequence that follows is $?DD \to 2CC \to 1CC$. Because the fundamental deterrence method is the same in S8 and in S12, the stability condition resulting from this deterrence mechanism is also the same in S8 and in S12: $2b_1 - b_2 \geq 2c$.

Theorem 3 (S12 Eq_G2 SPE Condition)**.** *In a game-of-choice equipped with resolution rule Equal-Say Game-2 Default, strategies **(1ab?,?001,1cd0)** satisfying* $(a = 1) \Rightarrow (c = 0)$ *and* $(b = 1) \Rightarrow (d = 0)$ *are subgame-perfect in the limit* $\delta \to 1$ *if, and only if:*

$$2b_1 - b_2 \geq 2c.$$

This proof of this result can be found in Appendix A.3.2 on page 61.

4.3.3 S16 EarlyCoopSPE

S16 has 108 EarlyCoopsPE strategies under resolution rule Eq_G2. The large number of strategies (exactly triple that in S12) makes finding the strategy pattern more difficult. Consequently, I present a numerical analysis rather than an analytical analysis of EarlyCoopSPE strategies in S16. This numerical analysis supports the hypothesis that the stability condition in S16 is also $2b_1 - b_2 \geq 2c$.

Specifically, I generated 100 random (b_1, b_2, c) triples such that $2b_1 - b_2 - 2c = -10^{-5}$ and 100 random triples such that $2b_1 - b_2 - 2c = +10^{-5}$. In particular, each (b_1, b_2, c) triple was generated by:

- generating two random floats $x, y \in (0, 1)$,
- setting $b_1 = max(x, y)$ and $b_2 = min(x, y)$, and
- setting $c = \frac{1}{2}(2b_1 - b_2 \pm 10^{-5})$.

If the stability condition is $2b_1 - b_2 - 2c \geq 0$, then my $findCoopSPE(g, \delta)$ program should find no CoopSPE strategies for (b_1, b_2, c) satisfying $2b_1 - b_2 - 2c = -10^{-5}$. On the other hand, the program should output all 108 CoopSPE strategies for (b_1, b_2, c) satisfying $2b_1 - b_2 - 2c = +10^{-5}$. In this case, I used discount factor $\delta = 1 - 10^{-9}$ in order to better approximate the limit $\delta \to 1$.

Running $findCoopSPE$ on the randomly generated (b_1, b_2, c) triples yielded the expected results: no CoopSPE strategies exist for stage game values satisfying $2b_1 - b_2 - 2c = +10^{-5}$, but all 108 CoopSPE strategies exist for stage game values satisfying $2b_1 - b_2 - 2c = +10^{-5}$. Table 4.1 and Table 4.2 show the number of CoopSPE strategies for 5 randomly generated triples when $2b_1 - b_2 < 2c$ (left) and when $2b_1 - b_2 > 2c$ (right).

b_1	b_2	c	$\#CoopSPE$
0.163	0.054	0.136	0
0.288	0.090	0.243	0
0.687	0.471	0.451	0
0.457	0.055	0.430	0
0.601	0.124	0.539	0

Table 4.1: Just below: $2b_1 - b_2 - 2c = -10^{-5}$

b_1	b_2	c	$\#CoopSPE$
0.905	0.691	0.559	108
0.624	0.137	0.556	108
0.657	0.029	0.643	108
0.946	0.011	0.941	108
0.849	0.400	0.649	108

Table 4.2: Just above: $2b_1 - b_2 - 2c = +10^{-5}$

This numerical evidence suggests that the stability condition in S16 is the same as it is in S8 and S12, namely: $2b_1 - b_2 \geq 2c$.

4.4 EarlyCoopSPE under Resolution rule Eq_G1

4.4.1 S8 EarlyCoopSPE

Although EarlyCoopSPE strategies exist in S8 under resolution rule Eq_G2, no such strategies appear to exist in S8 with resolution rule Eq_G1.

Why do strategies of the form **(1001,1??0)** no longer qualify as EarlyCoopSPE? After all, if one of the players makes a mistake (a CD or DC outcome), the error-correcting sequence still remains $?DD \to 2CC \to 1CC$. Yet, the ability to single-handedly force Game 1 enables a single deviation strategy $\tilde{s}$ to bypass the intermediate step $2CC$ by deviating to action 1D in subgame state DD. In contrast, under resolution rule Eq_G2, no way to single-handedly bypass Game 2 exists.

4.4.2 S12/S16 EarlyCoopSPE

S12 contains 18 EarlyCoopSPE strategies under resolution rule Eq_G1. These EarlyCoopSPE strategies share the following pattern: $s = (\mathbf{1000}, \mathbf{?ab1}, \mathbf{1cd0})$, where **?** is a free variable, $a = 1 \Rightarrow c = 1$, and $b = 1 \Rightarrow d = 1$.

Note that the Game 2 action following a CC result is irrelevant because the choice rule forces Game 1 after a CC round outcome. The relationship between a, c and b, d is suggestive: if the action rule in Game 2 specifies cooperation after a CD result, then the choice rule must force Game 1 after CD. Similarly, if the action rule in Game 2 specifies cooperation after a DC result, then the choice rule must force Game 1 after DC. Thus, CoopSPE strategies guarantee that CD and DC are met with defection.

As a sanity check, the relations between a, b and c, d result in 9 possibilities:
4 possibilities: $ab = 00 \iff cd = \mathbf{??}$
2 possibilities: $ab = 01 \iff cd = \mathbf{?}1$
2 possibilities: $ab = 10 \iff cd = 1\mathbf{?}$
1 possibility: $ab = 11 \iff cd = 11$
The question mark in (**1000**, **?ab1**, **1cd0**) contributes a multiplicative factor of two, yielding 18 overall EarlyCoopSPE strategies.

Hence, the Eq_G1 EarlyCoopSPE strategy set in S12 can be described as follows:

- If the prior round was CC, force Game 1 and cooperate.

- If the prior round was CD or DC, choose any stage game and make sure to defect.
- If the prior round was DD, choose Game 2 as the preferred game.
  - If the stage game to be played is Game 2 (the normal result when this strategy plays itself), cooperate.
  - If the stage game to be played is Game 1, defect.

Although the resolution rule is different, the mechanism that deters defection is the same: if one of the players playing this strategy makes a mistake, the error-correction sequence that follows is $?DD \rightarrow 2CC \rightarrow 1CC$. In addition, the ability to act differently in Game 1 and Game 2 has stabilized this error-correction sequence: because the action rule in Game 1 is **(1000)**, one must go through state 2CC in order to re-establish cooperation after a mistake. Because EarlyCoopSPE strategies under Eq_G1 use the same deterrence mechanism as those under Eq_G2, the resulting stability condition should also be the same in Eq_G1 and Eq_G2: $2b_1 - b_2 \geq 2c$.

Because a formal SPE analysis via the Single-Deviation Principle is time-consuming (though certainly possible), I performed the same numerical analysis as I did for the S16/Eq_G2 game-of-choice. The results again suggest that the SPE condition in S12 and S16 under the Eq_G1 is $2b_2 - b_2 \geq 2c$.

In summary, my equilibrium analysis has shown that cooperative SPE can exist in a game-of-choice between two PD stage games even when they cannot exist in an IPD of either stage game alone. In particular, I have proven that the stability condition for resolution rule Eq_G2 in S8 and S12 is $2b_1 - b_2 \geq 2c$ (in the limit $\delta \rightarrow 1$). In addition, I have provided numerical evidence suggesting that this stability condition also holds more broadly, i.e.:

- EarlyCoopSPE strategies in S8, S12, and S16, acting in a game-of-choice equipped with resolution rule Equal-Say Game-2 Default, *and*
- EarlyCoopSPE strategies in S12 and S16, acting in a game-of-choice equipped with resolution rule Equal-Say Game-1 Default,

are subgame-perfect in the limit $\delta \rightarrow 1$, *if and only if,* $2b_1 - b_2 \geq 2c$.

Finally, I have identified a general mechanism - "Game 2 as punishment" - that underlies all EarlyCoopSPE strategies across two resolution rules and in different strategy spaces.

Thus far, I have analyzed the game-of-choice model using two independent approaches:

- First, through the experimental observation of cooperation rates, and

- Second, through the analytical derivation of subgame-perfect equilibrum conditions.

In the last two sections of this thesis, I connect the experimental and analytical results together, providing evidence that the observed cooperation rates directly result from the adoption of CoopSPE strategies.

4.5 Connecting CoopSPE Strategies to Cooperation Rates

So far, the link between the theoretically-defined CoopSPE strategy set and the experimentally observed cooperation rates rests on two observations:

- First, that the large improvement in cooperation rates from S8 to S12/S16 in games-of-choice with Eq_G1 coincides with the appearance of EarlyCoopSPE strategies.
- Second, that cooperation rates in Fig. 3.2 on page 34 sharply increase at about $b_1 > 1.6$, which corresponds to the threshold value at which EarlyCoopSPE strategies become stable: the parameter values in the evolutionary simulation were $b_2 = 1.2$ and $c = 1.0$, so the threshold condition $2b_1 - b_2 > 2c$ simplifies to $b_1 > 1.6$.

In this section, I demonstrate a more direct link between cooperation rates and the adoption of CoopSPE strategies.

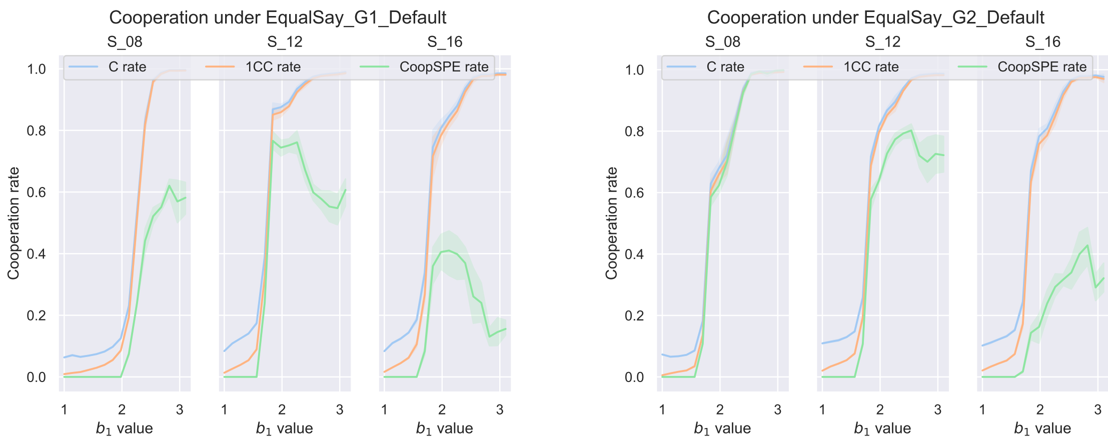
For each game-of-choice parameterized by b_1, I compute the proportion of time the population spends in CoopSPE strategy states. Section 4.5 on the next page below compares this CoopSPE rate to the rate of individual cooperation (C rate) and to the rate of mutual cooperation (1CC rate) in the three strategy spaces S8, S12, and S16.

First, a sanity check. Across all strategy spaces and resolution rules, the rate at which individuals cooperate (C rate) is closely linked to the proportion of time spent in the 1CC state (1CC rate). This is as it should be: individuals benefit from cooperating only if the cooperation is reciprocated.

4.5.1 Evaluating Goodness-of-Fit

In Section 4.5, cooperation rates and CoopSPE strategy frequency exhibit a nearly one-to-one correspondence in S8/Eq_G2. This one-to-one correspondence strongly supports the hypothesis that CoopSPE strategies cause cooperation inthe S8/Eq_G2 game-of-choice.

However, the correlation weakens as the dimension of strategy space increases. Why does



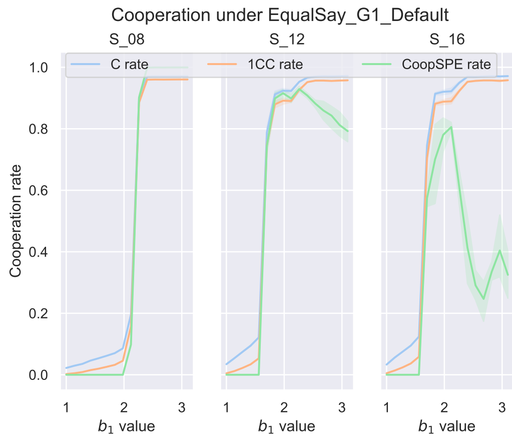
(a) Equal-Say Game 1

(b) Equal-Say Game 2

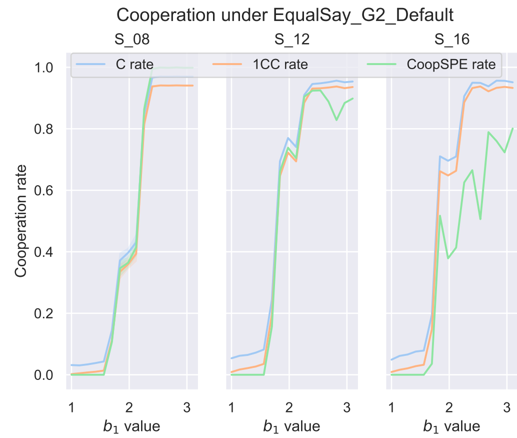
Figure 4.3: Correlation between Cooperation Rates and CoopSPE Frequency

this occur? One potential explanation is that there are more CoopSPE "look-alike" alternative strategies in higher-dimensional strategy spaces. These strategies act as CoopSPE does on the equilibrium path of play, but differ in how they act off the equilibrium path. Game states off the equilibrium path occur rarely when noise levels are low, and thus have minimal impact on strategy fitness. The addition of a large number of neutral alternatives both raises the cooperation rate and dilutes the frequencyof CoopSPE strategies.

If "look-alike" strategies are causing the discrepancy between cooperation rates and CoopSPE rates, then increasing the noise level and selection pressure should reduce this discrepancy. With more noise, off-equilibrium path games states are reached more frequently and have greater effect on strategy fitness; with higher selection pressure, strategies with greater fitness are more strongly selected for. However, increasing the selection pressure also increases the number of timesteps it takes for cooperative strategies to invade. Hence, I increase the number of timesteps in addition to increasing the noise level and selection pressure. Fig. 4.4 on the following page shows the goodness-of-fit when the noise level and selection pressure are increased.



(a) Equal-Say Game 1



(b) Equal-Say Game 2

Figure 4.4: Correlation between Cooperation and CoopSPE strategy frequency, under increased noise levels and selection pressure. Evolutionary parameters: $\epsilon = 0.01, \beta = 10.0, T = 10^6$.

The correlation in Fig. 4.4 dramatically improves when the noise level and selection pressure are increased. Under tougher evolutionary conditions, "look-alike" strategies disappear, and cooperation rates are nearly synonymous with CoopSPE strategy frequency for $1.6 \leq b_1 < 2.0$. This critical range of b_1 values corresponds to when the gap between the cooperation rates observed for the game-of-choice and the cooperation rates observed in the IPD is at its widest (Fig. 3.2 on page 34, Fig. 3.3 on page 35). This coalescence of cooperation rates and CoopSPE strategy frequency fuses together the experimental and analytical results: in the critical zone, EarlyCoopSPE strategies are clearly the driving force behind the game-of-choice's high cooperation rates.

Chapter 5

Conclusion

In this thesis, I extended the iterated Prisoner's Dilemma into a new model, games of choice. In a game of choice, players jointly choose between two Prisoner's Dilemma reward structures each round, and a "resolution rule" maps the players' preferences to the specific PD stage game played each round. This framework represents an initial step in incorporating the idea of "investing in deeper relationships" into the traditional theory of cooperation.

My investigation of this framework is comprised of three main layers. The first layer is that of experiment: I subjected the game-of-choice model to various strategy spaces and resolution rules, and observed its effect on the evolution of cooperation. The second layer is the analytical approach: I identified a promising strategy class based on its theoretical properties and proved the stability condition for this strategy class. The third layer is conceptual: I abstracted away from the specific form of the strategies in the two strategy spaces/resolution rules studied, and synthesized a understanding of the common mechanism by which they operate. These experimental, analytical, and conceptual layers mutually reinforce each other to provide a multi-dimensional understanding of direct reciprocity in games-of-choice.

The exploration process has revealed several interesting results. Across many different resolution rules, choosing between two Prisoner's Dilemma matrices each round promotes cooperation in comparison to playing a repeated game with either Prisoner's Dilemma alone. Under the right resolution rule, this result holds even when players are restricted to making decisions irrespective of whether the prior round was in the low-reward or high-reward PD game (S8/Eq_G2). It appears that cooperation can even promoted when the low-reward and high-reward PD games are chosen randomly.

In more concrete terms, my thesis has:

- linked the experimentally observed early-onset cooperation to a theoretically-defined strategy set (CoopSPE strategies) under the Equal-Say Game 1 default and Equal-say Game 2 default resolution rules,

- proven the SPE equilibrium condition for this strategy set ($2b_1 - b_2 \geq 2c$), demonstrated that it is always easier to meet than that the stability condition in either corresponding IPD, and provided numerical analysis suggesting that this equilibrium condition holds more broadly across other resolution rules as well,

- explained the mechanism resulting in this "easier" equilibrium condition. This basic mechanism specifies that, following unilateral defection, the route to high-reward mutual cooperation must pass through low-reward mutual cooperation. On an intuitive level, this makes sense because any additional time in the low-reward setting adds to the opportunity cost of defecting, thus making it easier for cooperation to become a best-response.

This thesis, however, has only scratched the surface of the game-of-choice model, and many unexplored questions remain. For example:

- In my game-of-choice parameterization, I held the cost of cooperation constant, $c = c_1 = c_2$. Conceivably, it may cost more to provide a larger benefit. How does $c_1 \neq c_2$ affect cooperation rates and the condition for when cooperation becomes stable?

- I considered cooperation in the context of pure and almost-pure memory-1 strategies. How does direct reciprocity evolve in the context of stochastic strategies (and in non memory-1 strategy spaces)?

- How does direct reciprocity evolve in a game of choice with more than two PD reward structures? How does direct reciprocity evolve in a game of choice between non-PD stage games, and in n-player interactions?

In conclusion, this thesis contributes an initial framework for capturing the ability of future interactions to become more (or less) rewarding. The results in my model show that cooperation can be easier to achieve than previously thought. Many research avenues remain unexplored, such as the effect of more complex strategy spaces, alternative reward structures, and multiplayer interactions. A full integration of this game-of-choice model into the august body of other work on the IPD remains an ambitious goal for future research.

Appendix

A.1 Robustness to Evolutionary Parameters

A major part of my thesis is the connection I show between CoopSPE strategy frequency and game-of-choice cooperation rates. The CoopSPE frequency and the cooperation rates are generated by an evolutionary process. In this section I test whether my results are specific to the evolutionary parameters I use, or whether the connection between cooperation rates and CoopSPE strategies holds more broadly.

As explained in Section 3.2 on page 32, three parameters govern the evolutionary process: the population size N, the selection pressure β, and the noise level ϵ. The key results in my thesis were obtained under evolutionary parameters $N = 100$, $\beta = 2.0$, and $\epsilon = 10^{-3}$. Fig. 1 on the next page below shows how the correlation between CoopSPE strategy frequency and cooperation rates are affected by changes to these evolutionary parameters. Specifically, I consider a fixed a S12/Eq_G2 game-of-choice with parameter values $b_1 = 1.8, b_2 = 1.2, c = 1.0$. I vary one evolutionary parameter at a time, holding the other evolutionary parameters fixed at $N = 100$, $\beta = 2.0$, and $\epsilon = 10^{-3}$. Each evolutionary simulation is run for $T = 5 \times 10^5$ timesteps. To control for variability between runs, I repeat each evolutionary simulation five times. The resulting individual cooperation rate (C rate), mutual cooperation rate (1CC rate), and CoopSPE strategy rate are plotted together in Fig. 1 on the following page.

A.1.1 Robustness to Selection Pressure, β

First, a sanity check. As selection pressure $\beta \to 0$, the individual cooperation rate approaches 50%. This is as expected: without selection for higher payoffs, all strategies have equal fitness. By symmetry, strategies that cooperate and defect are represented equally, resulting in a 50% individual cooperation rate.

When all strategies have equal fitness, CoopSPE strategies are adopted by chance. Hence

Figure 1: Cooperation Rate/SPE rate Robustness to Evolutionary Parameters.

the negligible CoopSPE rate as $\beta \to 0$ is to be expected because CoopSPE strategies represent only a negligible fraction of the strategy space (in this case, $36/2^{12}$, which is less than 1%).

When selection pressure is weak, individual cooperation rates are untethered from mutual cooperation rates because payoffs do not impact fitness: strategies that cooperate for no reason have the same fitness as the strategies that result in mutual cooperation.

However, at $\beta \approx 1.0$, individual and mutual cooperation rates converge. At this point, payoffs become relevant enough for direct reciprocity to overcome random chance as the main contributor of cooperation. At the same time, individual cooperation and CoopSPE rates also converge. This suggests that direct reciprocity, cooperation rates, and CoopSPE strategy frequency are all different faces of the same coin: CoopSPE strategies work through direct reciprocity; when direct reciprocity translates into a fitness advantage, CoopSPE strategies are adopted, resulting in both individual and mutual cooperation. My choice of $\beta = 2.0$ therefore represents a moderate level of selection that is slightly above the threshold needed for direct reciprocity to result in a tangible fitness advantage.

A.1.2 Robustness to Population Size, N

Sanity check: it makes sense that cooperation rates increase as the population size grows: smaller populations can promote spite rather than cooperation, and larger populations make it easier for cooperative strategies to fixate [24]. Once again, the individual cooperation rates, mutual cooperation rates, and CoopSPE rates all converge at approximately the same threshold value, approximately at $N = 64$. My choice of population size, $N = 100$, is in this convergence zone.

A.1.3 Robustness to Noise Level, ϵ

Sanity check: it makes sense that cooperation rates decrease at the same time as noise increases. Cooperation is only beneficial if it is reciprocated, and noise decreases the probability that a successful two-way reciprocal interaction will occur.

When the other evolutionary parameters are fixed at $N = 100$ and $\beta = 2.0$, the optimal level of noise for CoopSPE strategies appears to be $\epsilon = 10^{-3}$: cooperation rates and CoopSPE strategies are closest to each other at this value. My choice of precisely $\epsilon = 10^{-3}$ is particularly fortunate.

CoopSPE strategies and cooperation rates diverge the most at low noise levels. At low noise levels, off-equilibrium paths have minimal impact on strategy fitness; as a result, many neutral "look-alike" alternatives to CoopSPE strategies exist. The addition of a large number of neutral alternatives both raises the cooperation rate and dilutes the frequency of CoopSPE strategies.

At higher noise levels, cooperation rates and CoopSPE rates decrease in tandem. This is because noise disrupts all reciprocal strategies. CoopSPE strategies may be slightly negatively impacted than other cooperative strategies because CoopSPE strategies specifically delay returning to high-reward cooperation after mistakes.

Ultimately, the robustness experiment suggests that the individual cooperation and CoopSPE strategies converge when the selection pressure $\beta \geq 1.0$, the population size $N \geq 64$, and the noise level is in a Goldilocks zone around $\epsilon \approx 10^{-3}$. In particular, my choice of tried-and-true evolutionary parameters falls with these ranges. That being said, it appears that the connection between CoopSPE strategy frequency and individual cooperation rates emerges under the reasonable condition that direct reciprocity translates into a fitness advantage.

A.2 CoopSPE Pseudoalgorithms

This section contains the pseudocode for the algorithms referenced in Section 4.2 on page 40.

Algorithm 1 Is Strategy Cooperative?

Input strategy (binary tuple), game of choice (class instance)
Output True/False boolean

function IsStratCoop(strategy s, game g)
  for all $i \in \{1CC, 1CD, \dots, 2DC, 2DD\}$ **do**
    statesInCycle $\leftarrow$ findCycle($s_1 = s$, $s_2 = s$, prior state $= i$, game $= g$)
    if statesInCycle != [1CC] **then**
      return False
    end if
  end for
  return True
end function

Algorithm 2 Is Strategy SPE?

Input strategy (binary tuple), game of choice (class instance), discount factor $\delta \in (0, 1)$
Output True/False boolean

function IsStratSPE(strategy s, game g, discount factor δ)
  for all prior game state $i \in \{1CC, 1CD, \dots, 2DC, 2DD\}$ **do**
    baselineNextState = getNextState(s, s, prior state i)
    baselinePayoff = getAvgPayoffPerRound(s, s, baselineNextState, δ)

    possibleSingleDevs = getDevStates(s, s, i, g)
    possibleSingleDevs.discard(baselineNextState)

    singleDevPayoffs = getAvgPayoffsPerRound(possibleSingleDevs)

    if baselinePayoff $<$ **max** singleDevPayoffs **then**
      return False
    end if
  end for
  return True
end function

Algorithm 3 Find All CoopSPE

Input game (class instance), δ (float)
Output CoopSPE strategy list

```
function FindCooperativeSPE(game g, discount factor δ)
    coopSpeList = []
    numPureStrategies = 2^(game.strat_len)
    for all integers i ∈ {0, ..., numPureStrategies - 1} do
        s = toBinary(i, game.strat_len)            ▷ strategy repr. as binary tuple
        if isStratCoop(s, g) and isStratSPE(s, g, δ) then
            coopSpeList.append(s)
        end if
    end for
    return coopSpeList
end function
```

A.3 Proofs

A.3.1 S8 Equal-Say Game-2 SPE Condition Proof

The statement of Theorem 2 on page 42 is:

Theorem 2 (S8 Eq_G2 SPE Condition)**.** *In a game-of-choice equipped with resolution rule Equal-Say Game-2 Default, strategies of the form* ***(1001,1??0)*** *are subgame-perfect in the limit* $\delta \to 1$*, if, and only if:*

$$2b_1 - b_2 \geq 2c.$$

Proof. According to the Single Deviation Principle (Theorem 1 on page 26), it suffices to show that no useful single deviations exist in any subgame. Since strategies in S8 choose their actions based only on the prior round's *simple* (rather than compound) outcome, there are only 4 distinct subgame states: CC, CD, DC, and DD. The following case-by-case analysis shows that, in the limit of no discounting ($\delta \to 1$), meeting the condition $2b_1 - b_2 \geq 2c$ suffices for no useful single deviations to exist.

Notation: Let s denote a strategy of the form **(1001,1??0)**, and let $\tilde{s}$ denote a one-shot deviation strategy. That is, $\tilde{s}$ deviates from the action that strategy s would play in a single subgame state, and thereafter returns to acting in the way strategy s acts.

- Let $u_{ss} = E[u_s(s, s)]$ be shorthand for the baseline (discounted) average payoff per round when strategy s plays itself, and
- Let $u_{\tilde{s}s} = E[u_{\tilde{s}}(\tilde{s}, s)]$ be shorthand for the average (discounted) payoff per round $\tilde{s}$

receives when strategy $\tilde{s}$ plays strategy s.

- Let u_{err} be shorthand for the (discounted) average payoff per round resulting from the error-correcting sequence $?DD \to 2CC \to 1CC \circlearrowleft$.

It is helpful to remember from Chapter 2 that:

$$\text{Cumulative Discounted Payoff} = (\text{pre-cycle discounted payoffs}) + \frac{(\text{cycle discounted payoffs})}{1 - \delta^{\text{cycle len}}}$$

$$\text{Average Discounted Payoff} = (1 - \delta) \cdot (\text{Cumulative Discounted Payoff}).$$

Hence $u_{err} = (1-\delta) \cdot [0 + \delta(b_2 - c) + \delta^2(b_1 - c)\frac{1}{1-\delta}] = (1-\delta)\delta(b_2 - c) + \delta^2(b_1 - c)$.

Case 1, Subgame state CC:

Starting from subgame state CC, baseline play proceeds as follows:

Round	Prior Round	Game Pref	Result	s Payoff
0	CC/CC	G1/G1 = G1	1C/1C = 1CC	$b_1 - c \circlearrowleft$

Table 1: 1CC Subgame State: Baseline

Since s receives $b_1 - c$ each round, $u_{ss} = b_1 - c$.

Now consider a single-deviation strategy $\tilde{s}$ that deviates to action 1D (i.e. decides to choose Game 1 as the preferred stage game and defect)- and then returns to playing strategy s. When $\tilde{s}$ plays against s, this single deviation results in round outcome $1DC$; subsequent round outcomes proceed according to the error-correcting sequence $?DD \to 2CC \to 1CC \circlearrowleft$. In Table 2 below, the prior round outcome is presented from $\tilde{s}/s$'s point of view.

Round	Prior Round	Game Pref	Result	$\tilde{s}$ Payoff
0	CC/CC	G1/G1 = G1	1D*/1C = 1DC	b_1
1	1DC/1CD	G1 or G2 (depends on free var)	?DD	0
2	DD/DD	G2/G2 = G2	2C/2C = 2CC	$b_2 - c$
3	2CC/2CC	G1/G1 = G1	1C/1C = 1CC	$b_1 - c \circlearrowleft$

Table 2: 1CC Subgame State: 1D Single Deviation

1D Single-Deviation Avg. Payoff per Round:

$$\begin{aligned} u_{\tilde{s}s} &= (1-\delta) \cdot [b_1 + \delta^2(b_2 - c) + \delta^3(b_1 - c)\frac{1}{1-\delta}] \\ &= (1-\delta)[b_1 + \delta^2(b_2 - c)] + \delta^3(b_1 - c) \end{aligned}$$

When is this single-deviation strategy not useful?

$$
\begin{aligned}
(1-\delta)[b_1+\delta^2(b_2-c)]+\delta^3(b_1-c) &\leq b_1-c && \text{single-dev avg payoff} \leq \text{baseline avg payoff} \\
(1-\delta)[b_1+\delta^2(b_2-c)] &\leq (1-\delta^3)(b_1-c) && \text{subtract } \delta^3(b_1-c) \text{ from both sides} \\
b_1+\delta^2(b_2-c) &\leq (1+\delta+\delta^2)(b_1-c) && \text{for each } \delta \neq 1 \text{ we can divide by } 1-\delta>0 \\
b_1+(b_2-c) &\leq 3(b_1-c) && \text{take the limit as } \delta \to 1 \\
b_2 &\leq 2(b_1-c) \\
2c &\leq 2b_1-b_2
\end{aligned}
$$

Rather than playing the baseline action 1C in subgame state CC, $\tilde{s}$ deviated by playing action 1D. Two other deviation actions exist: 2D and 2C.

The payoffs resulting from single deviation 2D are the same as the payoffs resulting from single deviation 1D, except that the Round 0 payoff is now $b_2 < b_1$. The 2D single deviation payoff is therefore less than the 1D single deviation payoff. Since single deviation 1D is not useful when $2c \leq 2b_1 - b_2$, neither is single deviation 2D.

Similarly, the payoffs resulting from single deviation 2C are the same as the payoffs resulting from playing the baseline action 1C, except that the Round 0 payoff is now $b_2 - c < b_1 - c$. The 2C deviation payoff is therefore less than the baseline 1C baseline payoff, so the 2C deviation is also not useful.

Hence no useful single-deviation exist in subgame state CC.

Case 2, Subgame states CD and DC:

Baseline play follows the error-correction sequence: $u_{ss} = u_{err}$. Consider any single-deviation strategy $\tilde{s}$. Let o be the Round 0 outcome when $\tilde{s}$ plays against strategy s. Since EarlyCoopSPE strategies always defect after a CD or a DC outcome, there are four possibilities for round outcome o: 1CD, 1DD, 2CD, and 2DD. Note that $o = 1DD/2DD$ results in the error-correction sequence $?DD \to 2CC \to 1CC \circlearrowleft$, i.e. in $u_{\tilde{s}s} = u_{err}$. These two outcomes therefore do not represent a useful deviation. The other two outcomes, 1CD and 2CD, result in a payoff of $-c$ for strategy $\tilde{s}$, and are subsequently followed by the error-correction sequence. Consequently, for outcomes 1CD and 2CD,

$$\begin{aligned}
u_{\tilde{s}s} &= (1-\delta)(\text{Cumulative Discounted Payoff}) \\
&= (1-\delta)(-c + \delta(\text{the cumulative discounted payoff for the error-correction sequence})) \\
&= (1-\delta)(-c + \delta(u_{err}/(1-\delta)) \\
&= (1-\delta)(-c) + \delta u_{err} \\
&\leq u_{err}
\end{aligned}$$

Therefore no useful deviations exist in subgame states CD and DC.

Case 3, Subgame state DD:

In this case, the first step of the error-correction sequence (DD) has already been taken; the baseline play proceeds $2CC \to 1CC \circlearrowleft$. Compared to the payoffs in the error-correction sequence, these payoffs are missing the payoff in the first step (which was 0 anyways) and occur one round earlier (which results in less discounting). Hence the baseline payoff is $u_{ss} = u_{err}/\delta$. Since EarlyCoopSPE strategies force Game 2 after a DD outcome and cooperate, the possible round outcomes following a single-deviation action are 2CC and 2DC. 2CC is the same as the baseline play and therefore does not represent a useful deviation. 2DC results in payoff b_2 for $\tilde{s}$, and is followed by the error-correction sequence. In this case, $u_{\tilde{s}s} = (1-\delta)b_2 + \delta u_{err}$. When is this deviation not useful?

$$\begin{aligned}
u_{\tilde{s}s} &\leq u_{ss} \\
(1-\delta)b_2 + \delta u_{err} &\leq u_{err}/\delta \\
\delta(1-\delta)b_2 + \delta^2 u_{err} &\leq u_{err} \\
\delta(1-\delta)b_2 &\leq (1-\delta^2)u_{err} \\
\delta b_2 &\leq (1+\delta)u_{err} \quad \text{for each } \delta \neq 1 \text{ we can divide by } 1-\delta > 0 \\
\delta b_2 &\leq (1+\delta)[(1-\delta)\delta(b_2 - c) + \delta^2(b_1 - c)] \\
b_2 &\leq 2(b_1 - c) \quad \text{in the limit } \delta \to 1 \\
2c &\leq 2b_1 - b_2
\end{aligned}$$

Hence, no useful single deviations exist when $2b_1 - b_2 \geq 2c$. By the Single-Deviation Principle, in the limit of no discounting ($\delta \to 1$), EarlyCoopSPE strategies are subgame perfect for $2b_1 - b_2 \geq 2c$. $\square$

A.3.2 S12 Equal-Say Game-2 SPE Condition Proof

The statement of Theorem 3 on page 44 is:

Theorem 3 (S12 Eq_G2 SPE Condition)**.** *In a game-of-choice equipped with resolution rule Equal-Say Game-2 Default, strategies* ***(1ab?,?001,1cd0)*** *satisfying* $(a = 1) \Rightarrow (c = 0)$ *and* $(b = 1) \Rightarrow (d = 0)$ *are subgame-perfect in the limit* $\delta \to 1$ *if, and only if:*

$$2b_1 - b_2 \geq 2c.$$

Proof. This result can be proven by repeating the analysis detailed in Appendix A.3.1 on page 57. The main arguments are directly applicable to this strategy set, as the above strategies also choose their actions based only on the prior round's simple (rather than compound) outcome and have error-correction sequence $?DD \to 2CC \to 1CC$. □

Bibliography

[1] Martin A Nowak. *Evolutionary dynamics*. Harvard University Press, 2006.

[2] Arne Traulsen and Christoph Hauert. “Stochastic evolutionary game dynamics”. In: *Reviews of nonlinear dynamics and complexity* 2 (2009), pp. 25–61.

[3] Martin Nowak and Karl Sigmund. “A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner’s Dilemma game”. In: *Nature* 364.6432 (1993), p. 56.

[4] Martin A Nowak. “Five rules for the evolution of cooperation”. In: *science* 314.5805 (2006), pp. 1560–1563.

[5] Christoph Hauert et al. “Volunteering as red queen mechanism for cooperation in public goods games”. In: *Science* 296.5570 (2002), pp. 1129–1132.

[6] Jae C Oh. “Promoting cooperation using kin-biased conditional strategy in the iterated prisoner’s dilemma game”. In: *Information Sciences* 133.3-4 (2001), pp. 149–164.

[7] Zhen Wang, Attila Szolnoki, and Matjaž Perc. “Interdependent network reciprocity in evolutionary games”. In: *Scientific reports* 3 (2013), p. 1183.

[8] Attila Szolnoki and Matjaž Perc. “Conformity enhances network reciprocity in evolutionary social dilemmas”. In: *Journal of The Royal Society Interface* 12.103 (2015), p. 20141299.

[9] Karthik Panchanathan. “Two wrongs don’t make a right: The initial viability of different assessment rules in the evolution of indirect reciprocity”. eng. In: *Journal of Theoretical Biology* 277.1 (2011), pp. 48–54. ISSN: 0022-5193.

[10] Robert L Trivers. “The evolution of reciprocal altruism”. In: *The Quarterly review of biology* 46.1 (1971), pp. 35–57.

[11] Robert Axelrod and William Donald Hamilton. “The evolution of cooperation”. In: *science* 211.4489 (1981), pp. 1390–1396.

[12] K Sigmund and MA Nowak. "Tit for tat in heterogenous populations". In: *Nature* 355 (1992), pp. 250–253.

[13] Lorens A Imhof and Martin A Nowak. "Stochastic evolutionary dynamics of direct reciprocity". In: *Proceedings of the Royal Society B: Biological Sciences* 277.1680 (2009), pp. 463–468.

[14] William H Press and Freeman J Dyson. "Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent". In: *Proceedings of the National Academy of Sciences* 109.26 (2012), pp. 10409–10413.

[15] Christian Hilbe et al. "Evolution of cooperation in stochastic games". In: *Nature* 559.7713 (2018), p. 246.

[16] Ethan Akin. "What you gotta know to play good in the iterated prisoner's dilemma". In: *Games* 6.3 (2015), pp. 175–190.

[17] Christian Hilbe, Arne Traulsen, and Karl Sigmund. "Partners or rivals? Strategies for the iterated prisoner's dilemma". In: *Games and economic behavior* 92 (2015), pp. 41–52.

[18] Christian Hilbe et al. "Memory-n strategies of direct reciprocity". In: *Proceedings of the National Academy of Sciences* 114.18 (2017), pp. 4715–4720.

[19] Lloyd S Shapley. "Stochastic games". In: *Proceedings of the national academy of sciences* 39.10 (1953), pp. 1095–1100.

[20] Eilon Solan and Nicolas Vieille. "Stochastic games". In: *Proceedings of the National Academy of Sciences* 112.45 (2015), pp. 13743–13746. ISSN: 0027-8424. DOI: `10.1073/pnas.1513508112`. eprint: `https://www.pnas.org/content/112/45/13743.full.pdf`. URL: `https://www.pnas.org/content/112/45/13743`.

[21] N Momeni et al. "Evolutionary games with voting: Emergence of civil society". working paper.

[22] Lindi M Wahl and Martin A Nowak. "The continuous Prisoner's Dilemma: I. Linear reactive strategies". In: *Journal of Theoretical Biology* 200.3 (1999), pp. 307–321.

[23] Timothy Killingback and Michael Doebeli. "The continuous prisoner's dilemma and the evolution of cooperation through reciprocal altruism with variable investment". In: *The American Naturalist* 160.4 (2002), pp. 421–438.

[24] Martin A Nowak et al. "Emergence of cooperation and evolutionary stability in finite populations". In: *Nature* 428.6983 (2004), p. 646.

[25] Drew Fudenberg and Lorens A Imhof. "Imitation processes with small mutations". In: *Journal of Economic Theory* 131.1 (2006), pp. 251–262.

[26] David C Parkes and Sven Seuken. *Economics and Computation*. Cambridge University Press, 2019.

[27] Dana Nau. "Lecture 8: Stochastic Games". In: *Introduction to Game Theory, Lecture Notes*. 2010. URL: `http://www.cs.umd.edu/users/nau/game-theory/`.

[28] Joseph K. Blitzstein and Jessica Hwang. *Introduction to Probability*. CRC Press, Taylor & Francis Group, 2014.

[29] Robert Boyd and Jeffrey P Lorberbaum. "No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game". In: *Nature* 327.6117 (1987), p. 58.