Direct and Indirect Reciprocity in Public Goods Games

by

Emilie Zentner

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Agricultural and Resource Economics

Department of Resource Economics and Environmental Sociology

University of Alberta

Abstract

Cooperation and coordination are challenging to achieve in public goods games. As a result, public goods are often chronically under-provisioned due to free-riding. However, reciprocity has been increasingly associated with cooperative behaviour and may play an important role in driving individual contributions in public goods games. While reciprocity is often discussed as a two party interaction (*direct reciprocity*), reciprocity can also occur in interactions between three parties (*indirect reciprocity*), and capture effects of recent experiences (*indirect upstream reciprocity*) and reputation (*indirect downstream reciprocity*) on individual contributions. In this thesis, our objective is to examine the role of both direct and indirect reciprocity preferences in voluntary contributions in public goods games. Understanding the relevance of these preferences will have implications for many social dilemma situations impacting society. Using a psychological game theoretic approach, we apply and extend the utility framework introduced by Dufwenberg and Kirchsteiger (2004) to incorporate direct and indirect reciprocity preferences in a public goods setting. We identify several theoretical predictions which show the existence of cooperative equilibria. Furthermore, we identify several conditions under which the inclusion of direct and indirect reciprocity preferences can lead to more cooperative outcomes. This research furthers our understanding of the role of reciprocity preferences on individual contributions in public goods games and highlights the importance of both direct and indirect reciprocity in supporting cooperative behaviour.

Preface

Some of the research conducted for this thesis forms part of a collaborative research project led by Dr. Maik Kecinski at the University of Delaware, with Dr. Corinne Langinier and Dr. Wiktor (Vic) Adamowicz at the University of Alberta. The model described in Chapter 2 was a collaborative effort designed by myself and all members of the committee. The literature review, analysis, and interpretation of theoretical predictions are my original work. The research project received research ethics approval from the University of Alberta Research Ethics Board, "Decision-Making Games - Choices and Groups", No. Pro00091773, June 2020. The research project is funded, in part, by a Killam Cornerstone Operating Grant. Portions of this thesis have been presented at the Association of Environmental and Resource Economists sponsored session at the Western Economic Association International conference in June 2020 and at the Committee on Women in Agricultural Economics and Experimental Economics track session at the Agricultural and Applied Economics Association conference in August 2020.

Acknowledgements

I respectfully acknowledge that the University of Alberta is situated on Treaty 6 territory, traditional lands of First Nations and Métis people. I am grateful to the Department of Resource Economics and Environmental Sociology for providing a welcoming and supportive learning environment throughout both my undergraduate and graduate degrees. I am grateful for support and recognition through the University of Alberta Thesis-Based Master's Scholarship, the Social Sciences and Humanities Research Council Joseph Armand Bombardier Canada Graduate Scholarship (Master's), and the Faculty of Graduate Studies and Research Walter H. Johns Graduate Fellowship. The Resource Economics and Environmental Sociology Graduate Students' Association provided a welcome space for me to grow and learn, and I appreciate the skills that I was able to cultivate with the support of my peers. To my entire cohort, thank you for your support in many ways, big and small.

Dr. Vic Adamowicz - thank you for providing a home for me in the department with your research group and connecting me with your graduate students. To the Adamowicz research group - all of your feedback and encouragement is much appreciated. Thank you to Lusi Xie for being an incredible resource and mentor.

Dr. Corinne Langinier - thank you for the support that you have offered me. I have learned so much from you and feel privileged to have worked with you. Thank you for teaching me to accept non-linearity in the process.

Dr. Maik Kecinski - you have endeavoured to teach me many lessons over the years that we have worked together. I value your mentorship and the space you have created for me. Thank you. I hope that this work matters to some people for a brief period of time. Next beer is on me.

Contents

List of Figures

List of Tables

List of Symbols

n	group size, number of players, $n \geq 3$
h	history of the game
π	material payoff
y	endowment
c	contribution, $c \in [0, y]$
α	marginal per capita return (MPCR); $0 < \alpha < 1 < n\alpha$
U	utility
κ	kindness
λ	perceived kindness
$b_{i,j}(h)$	first-order beliefs; player i's belief about player j's contribution strategy conditional on history of the game
$b_{i,j,k}(h)$	second-order beliefs; player i's belief about player j's belief about player k's contribution strategy conditional on history of the game
R	direct reciprocity preference, $R \geq 0$
$R_{i,j}$	player i's direct reciprocity preference parameter; captures player i's direct reciprocity preferences towards player j, $R_{i,j} \geq 0$
R_i	player i's direct reciprocity preference parameter, $R_i \geq 0$
R_j	player j's direct reciprocity preference parameter, $R_j \geq 0$
R_J	player J's direct reciprocity preference parameter, $R_J \geq 0$
R_I	indirect reciprocity preference parameter, $R_I \geq 0$
γ	relative strength of indirect reciprocity preferences to direct reciprocity preferences; $\gamma = \frac{R_I}{R_J}$ and $\gamma > 0$. $\gamma > 1$ indicates stronger indirect reciprocity preferences compared to direct reciprocity preferences, and $\gamma < 1$ indicates weaker indirect reciprocity preferences compared to direct reciprocity preferences

List of Abbreviations

marginal per capita return (MPCR)

Sequential Reciprocity Equilibrium (SRE)

third-party beneficiary (*TPB*)

Chapter 1

Introduction

Individual contributions to public goods games have been increasingly shown to be associated with reciprocal behaviours (Chaudhuri, 2011; Croson, 2007). Cialdini (2009) defines reciprocity as returning in kind how one is treated by others. In the literature, reciprocity is mostly discussed in terms of direct reciprocity which involves conditional cooperation between two parties, such as tit-for-tat (Nowak and Sigmund, 2005; Axelrod and Hamilton, 1981). Less studied, indirect reciprocity addresses conditional cooperation between three parties, and as such incorporates elements of tit-for-tat but also depends on recent experiences and reputation of others (Nowak and Sigmund, 2005).

Indirect reciprocity can capture many social interactions that are excluded by the two party nature of direct reciprocity. Where direct reciprocity (Figure 1.1) captures an exchange of altruistic acts between two parties[1], three party interactions can provide richer environments to examine cooperative behaviour.

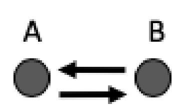

Figure 1.1: Direct Reciprocity occurs when A is kind to B and B is kind to A (Nowak and Sigmund, 2005).

[1] Trivers (1971) defines reciprocal altruism as acts that have a small cost to the giver and a large benefit to the receiver. Nowak and Sigmund (2005) state that reciprocal altruism involves the exchange of altruistic acts such that both parties obtain a net benefit. Pure altruism occurs when the giver incurs a small cost and does not benefit.

Indirect upstream reciprocity (Figure 1.2) focuses on the effects of recent experiences on an individual's behaviour. For example, if you were nice to me, I may be motivated to "pay-it-forward" and be nice to someone else.

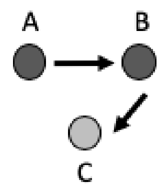

Figure 1.2: Indirect Upstream Reciprocity occurs when A is kind to B, then B is kind to C (Nowak and Sigmund, 2005).

In indirect downstream reciprocity (Figure 1.3), the focus is on the effect of reputation on an individual's behaviour. For example, if I saw you be nice to someone else, I may evaluate you as being a "kind" person and reward you by being nice to you. As well, indirect reciprocity preferences may be more

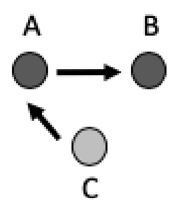

Figure 1.3: Indirect Downstream Reciprocity occurs when A is kind to B, then C is kind A (Nowak and Sigmund, 2005).

salient for individuals who place stronger value on societal appearances (see, for example, Nowak and Sigmund, 1998, for a discussion on indirect reciprocity and image scoring models).

The objective of this thesis is to examine the role of both direct and indirect reciprocity preferences in voluntary contributions to public goods. Understanding the relevance of these preferences will have implications for many social dilemma situations impacting society today – for example, climate change abatement, adaptation and mitigation efforts are global public goods that are chronically under-provisioned due to incentives to deviate from emissions reduction targets that would otherwise be welfare enhancing.

Free-riding in public goods games, in which an individual does not contribute to the public good yet may still benefit from its provision, undermines cooperation and is a key characteristic of collective action social dilemmas. Reaching the social optimum requires coordination, and in its absence, "[f]reedom in a commons brings ruin to all" (Hardin, 1968). However, individuals cooperate more than predicted (Ostrom, 2010) and conditional cooperation among individuals can support higher contributions to the public good (Chaudhuri, 2011). Accounting for reciprocity preferences can provide further insight into cooperative outcomes in public goods games and add new behavioural predictions.

Our theoretical framework builds on Dufwenberg and Kirchsteiger (2004)'s model of reciprocity, which we extend to a public goods game with direct and indirect reciprocity preferences. Deriving new theoretical predictions for behaviour in a public goods setting, we show that the inclusion of both direct and indirect reciprocity preferences can lead to more cooperative outcomes in public goods games. We specifically use a psychological game theory approach that incorporates first- and second-order beliefs when solving for simultaneous and sequential choice reciprocity models. The distinction between traditional game theory approaches and psychological game theory is important to note. When incorporating reciprocity preferences into a public goods game context, we are not just assessing whether an action is "kind" or "unkind", but also evaluating the intention behind the action, i.e., modelling reciprocity requires the consideration of beliefs, and considering beliefs and incorporating belief-dependent emotions into games moves away from traditional game theory and into psychological games. The general framework for psychological games was developed by Geanakoplos et al. (1989), who discuss the application of traditional game theory concepts to psychological games and provide a set of assumptions and appropriate solution concepts for this class of games. Many of Geanakoplos et al.'s assumptions address beliefs. For example, players are assumed to hold rational expectations (beliefs are correct in equilibrium) and coherent beliefs (first-order and second-order beliefs correlated). As well, there is collective coherency; that is, coherent beliefs are common knowledge among players (Geanakoplos et al., 1989). Finally, Geanakoplos et al. assume perfect recall among players. Perfect recall occurs when each player knows what they previously chose (action recall) and each player remembers whatever they knew at any previous move, including their previous choices as well as previous states of the game (memory of past knowledge) (Kuhn, 1953).

Furthermore, and relevant for our sequential choice reciprocity framework, traditional game theory concepts such as sequential rationality still hold in psychological game environments, but methods such as backward induction fail due to the inclusion of beliefs (Geanakoplos et al., 1989). In traditional game theory, sequential rationality occurs when, at every point in the game tree, a player's strategy specifies optimal actions (Mas-Collel et al., 1995). The key difference, according to Geanakoplos et al., lies in the fact that once players reach a decision node, this does not necessarily provide players with essential information needed for further decision making as nodes only identify histories or previous play and fail to incorporate the beliefs of the players. Furthermore, beliefs must be sequentially rational (strategies are rational given beliefs) and beliefs must be consistent (obtained through Bayesian updating given the strategies) - this encompasses information sets both on and off the equilibrium path. Hence, the sequential psychological equilibrium as outlined in Geanakoplos et al. (1989) differs from the sequential equilibrium concept (see for example, Kreps and Wilson, 1982) in that beliefs directly enter into players' utility functions and the Geanakoplos et al. (1989) sequential psychological equilibrium concept accounts for higher-order beliefs. An application of psychological game theory can be found in Rabin (1993), who incorporated kindness in players' utility functions to derive fairness equilibria. Similar to our reciprocity approach presented in this thesis, Rabin's work sought to incorporate social preferences through evaluating fairness considerations – Rabin showed that establishing a status quo of what everyone else thinks is "fair" and assessing "fair" action through the use of reference points adds new predictive power to economic models.

Two main classes of models have emerged in the literature on modelling reciprocity preferences: outcome-based models and intention-based models.

Outcome-based models focus on modelling distributional concerns and include work by Fehr and Schmidt (1999) who examine the roles of inequity-aversion and guilt-aversion, and Bolton and Ockenfels (2000) who incorporate both equity and reciprocity in their model. However, these models fall short of including higher-order beliefs, which are important in understanding the motivation behind belief-dependent emotions. On the other hand, intentions-based models incorporate first- and higher-order beliefs in players' utility functions. Dufwenberg and Kirchsteiger (2004)'s model of reciprocity incorporates concepts from both Geanakoplos et al. (1989) and Rabin (1993) – specifically, Dufwenberg and Kirchsteiger allow for belief updating as players reach decision nodes throughout the game, allowing players to consider both the history of the game and their beliefs at that decision node. Unlike Dufwenberg and Kirchsteiger (2004)'s model, Falk and Fischbacher (2006) present a blend of an outcome- and intention-based model, however, their model only allowed for initial beliefs and did not consider belief updating once players reach decision nodes. This is an important omission as belief updating and variations in belief updating (i.e., Jiang and Wu, 2019) are key to a player's assessment of what is "fair", or in Dufwenberg and Kirchsteiger (2004)'s model, what is "kind", and a player's belief about another player's kindness can be different depending on the decision node reached and the history of the game.

Most applications of reciprocity models in the literature are exclusively addressing direct reciprocity, or the interactions between two parties. The timing of these interactions may be simultaneous (Sugden, 1984) or sequential (Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006). Croson (2007) distinguishes between simultaneous reciprocity as "matching" behaviour (such as tit-for-tat strategies as outlined in Axelrod and Hamilton, 1981) and sequential reciprocity as "rewarding" behaviour. There have also been several applications of Dufwenberg and Kirchsteiger (2004)'s model of reciprocity to public goods games, including Dufwenberg et al. (2011), which examines guilt aversion and reciprocal motivations, and Dufwenberg and Patel (2017) which revisits Palfrey and Rosenthal (1984)'s participation problem in a discrete public goods game showing that both cost-sharing and reciprocity preferences can contribute to solving coordination problems in public goods games. These applications find that there is a positive correlation between beliefs and contributions, and that reciprocity increases coordination.

Ambrus and Pathak (2011) also examine cooperation in a repeated public goods game, allowing for both selfish and asymmetrically reciprocal players. They find that asymmetric reciprocity preferences can describe contribution decay behaviour observed in economic experiments. As well, Teyssier (2012) looks at inequity aversion and risk in a public goods game using the outcome-based model of Fehr and Schmidt (1999). While many public goods games are simultaneous, Teyssier (2012) uses an explicit sequential format to find that reciprocity is driven by advantageous-inequity aversion. This distributional concern arises depending on the player's role as first- or second-mover, where first-movers with stronger disadvantageous-inequity aversion will contribute less to public goods, and second-movers with sufficiently strong advantageous-inequity aversion will contribute more to public goods.

In another experiment, Bardsley and Sausgruber (2005) disentangle the impact of reciprocity versus conformity on the crowding-in effect in public goods game. Their design examines differences between own-group regarding behaviour (i.e., reciprocity) and other-group regarding behaviour (i.e., conformity), by modifying the information that players observe and when they make their contribution choices. While a subset of a group of players will simultaneously choose their contribution amounts, a randomly selected second-mover will observe a vector of contributions either from their own group of players or some other group. The player with this information then makes their contribution choice. Bardsley and Sausgruber find evidence that reciprocity accounts for the majority of crowing-in effects in public goods games. Clark et al. (2020) also examines the role of indirect reciprocity in driving cooperative behaviour. In their work, Clark et al. show that providing simple records of players' past behaviour can support cooperation in social dilemmas requiring coordination, including both prisoners' dilemma and public goods games. The observation of players' past behaviour aligns with the reputation-based aspect of indirect downstream reciprocity. However, indirect upstream reciprocity, or the role of recent experience in driving cooperation, is yet to be explored. While there are applications of reciprocity models in the literature, as far as we are aware, the work of this thesis to model indirect reciprocity as indirect upstream reciprocity and indirect downstream reciprocity is a unique application of Dufwenberg and Kirchsteiger (2004)'s model of reciprocity in a public goods game setting and provides a strong theoretical contribution to the literature.

Given the theoretical and experimental contributions presented above, this thesis adds to the literature by building on and extending the Dufwenberg and Kirchsteiger (2004) model. Specifically, we apply a theoretical framework of direct and indirect reciprocity to public goods games. The thesis further includes characteristics of the framework used in Bardsley and Sausgruber (2005)'s experimental design and follows the definitions of direct and indirect reciprocity as described by Nowak and Sigmund (2005).

The remainder of this thesis is structured as follows. In Chapter 2, we present a model setting that explicitly introduces both direct and indirect reciprocity preferences into a public goods game. In Section 2.1, we present a simultaneous choice public goods game with reciprocity preferences as a baseline case. In Section 2.2, we present a sequential choice public goods game with reciprocity preferences as a model of direct reciprocity. Indirect upstream reciprocity is presented in Section 2.3 and indirect downstream reciprocity is presented in Section 2.4. Chapter 3 concludes with a discussion of the theoretical findings and a discussion of the general implications from the model.

Chapter 2

The Model

We adapt and extend Dufwenberg and Kirchsteiger (2004)'s model of sequential reciprocity and develop an explicit model of indirect reciprocity following definitions provided by Nowak and Sigmund (2005). Dufwenberg and Kirchsteiger (2004)'s model of sequential reciprocity incorporates reciprocal motivations into a player's utility function and examines how these motivations influence a player's actions depending on the decision node reached and the history of the game. Belief formation and updating are important aspects of the Dufwenberg and Kirchsteiger model which drive perceptions of the intentions behind actions and motivations for reciprocity. Dufwenberg and Kirchsteiger present a utility function comprised of both a material payoff function and a reciprocity payoff function. The reciprocity payoff function evaluates kindness and perceived kindness, where utility is increased by reciprocity when these functions have matching signs (i.e., when a player is "kind" to others and believes that others intended to be "kind" to them, then their utility is increasing). In our application, the model focuses on how concerns for reciprocity drive individual contribution behaviour. In what follows, we present applications of Dufwenberg and Kirchsteiger (2004)'s model in a public goods game setting with simultaneous choice (Section 2.1) and sequential choice (Section 2.2). In Sections 2.3 and 2.4, we extend the model to include indirect reciprocity preferences.

Consider an n player public goods game with a voluntary contribution mechanism and $n \geq 3$. Each player i, where $i \in A = \{1, ..., n\}$, has endowment

y and chooses a contribution c_i to the group, with $c_i \in C_i = [0, y]$. Each player i can contribute nothing ($c_i = 0$), contribute something ($0 < c_i < y$), or contribute their[1] full endowment ($c_i = y$). A profile of contribution strategies is $c = (c_1, ..., c_{i-1}, c_i, c_{i+1}, ..., c_n)$.

The model setting includes both simultaneous and sequential play. Players may either choose their contribution amounts at the same time as the other players in their group, or a player may observe the average[2] contributions of the players in another group or their own group before making their contribution decision. The timing of each game will be described in greater detail in subsequent sections. If the game is sequential, then player i observes the average contribution of other players and has information about the history of the game; player i can consider the history of the game when choosing their contribution. That is, at every node in a sequential game, each player i might know the history of the game, h (i.e., if player i is the second-mover). The history of the game is defined as a vector of contributions made by other players. For instance, $h = (0, ..., 0)$ represents the history when player i knows with probability 1 that all the other players have contributed nothing, and when $h = (y, ..., y)$, player i knows with probability 1 that all the other players have contributed the full endowment, y. When $h = (c_1, ..., c_n)$, player i does not know with certainty what each player has contributed, but player i observes the average contribution $\frac{1}{n-1}\sum_{j \neq i} c_j = c$, where $0 < c < y$, so player i knows that each player j has contributed c_j, where $j = 1, ..., n$, $j \neq i$, and $0 \leq c_j \leq y$.

Following Dufwenberg and Kirchsteiger (2004), each player i has beliefs about the other players' contribution strategies, and each player i can assess probabilities on their beliefs about the other players' contributions, where first- and higher-order beliefs are contribution amounts. We denote $b_{i,j}(h)$ as

[1]Throughout the thesis and in the model exposition, we use gender-neutral language to discuss players (i.e., "they/them" pronouns). Every effort is made to distinguish between players when referring to a singular player versus plural players through additional descriptors such as set membership.

[2]We define average as the mean of the contributions of the other players in the group, where $\frac{1}{n-1}\sum_{j \neq i} c_j = c$. Chaudhuri (2011) highlights several papers that use average group contributions, including Croson (2007).

the first-order belief that player i holds about player j's contribution strategy conditional on history h, where player $i, j = 1, ..., n$, $i \neq j$, and $(b_{i,j}(h))_{j \neq i} \in B_{i,j}$. Each player i also holds higher-order beliefs. Player i holds beliefs about player j's belief about other players' contribution strategies conditional on history h, $(b_{i,j,k}(h))_{k \neq j} \in B_{i,j,k}$, where players $i, j, k \in A$ and $i \neq j$ and $j \neq k$. Note that $b_{i,j}(h)$ and $b_{i,j,k}(h)$ represent players' beliefs about contributions and that $B_{i,j,k} = B_{j,k} = C_k$. That is, we assume that beliefs are correct in equilibrium (i.e., rational expectations in Dufwenberg and Kirchsteiger, 2004), beliefs are coherent (first-order and higher-order beliefs correlated) and there is collective coherency among players (Geanakoplos et al., 1989). We also assume that players have perfect recall (Kuhn, 1953) such that they know the history of the game and all preceding choices up to the decision node reached. Finally, we assume that players are myopic and only care about how their decisions affect the current period's payoff not taking into account the future.

2.1 Baseline

As a baseline case, consider a simultaneous n player game with symmetric players, Figure 2.1. We define player $i \in A$, and player $j \in A_i$ such that

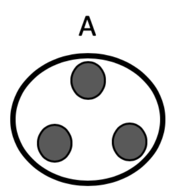

Figure 2.1: Baseline

$A_i = \{1, ..., i-1, i+1, ..., n\}$ where $A_i \subset A$. In other words, player i can be any player in set A while player j is a player in set A_i, which represents the set of all players excluding player i (i.e., a set of $(n-1)$ players).

2.1.1 No Reciprocity Preferences

Each player $i \in A$ simultaneously chooses $c_i \in [0, y]$, and receives the following material payoff

$$\pi_i = y - c_i + \alpha \sum_{l \in A} c_l, \tag{2.1.1}$$

where α is the marginal per capita return (MPCR) with $0 < \alpha < 1 < n\alpha$, and $\sum_{l \in A} c_l$ identifies the contributions of all n players in set A to the group account, where player $l \in A$[3]. The payoff for each player j is analogous. Player i chooses c_i that solves $\underset{c_i}{max}\ \pi_i$ where π_i is defined by (2.1.1). Given that $\alpha < 1$, the first-order condition is always negative,

$$-1 + \alpha < 0. \tag{2.1.2}$$

The only Nash equilibrium is thus $c_1^* = ... = c_n^* = 0$, such that each player i contributes nothing. This is the equilibrium that we expect to find in a classic public goods game. We define this equilibrium as the traditional public goods game Nash equilibrium.

2.1.2 Reciprocity Preferences

While reciprocity is typically modelled sequentially, and despite the simultaneous nature of this game, player i could have reciprocity preferences that influence their beliefs about the contribution strategies of other players. However, due to the structure of the game, players cannot discriminate - player i can only be “kind” to all the players in set A_i in order to be “kind” to any one individual player. We incorporate a reciprocity payoff into player i’s utility function following Dufwenberg and Kirchsteiger (2004). To do so, we define additional components such as beliefs and higher-order beliefs about the contribution strategies of the other players in the set. Given history h, let $\bar{b}_{i,A_i}$ be player i’s belief about the average contribution of the other players in set A_i, such that

[3] Player i is included in the set of players $l \in A$

$$\bar{b}_{i,A_i}(h) = \frac{1}{n-1}\sum_{j\in A_i} b_{i,j}(h), \tag{2.1.3}$$

where

$$\sum_{j\in A_i} b_{i,j}(h) = b_{i,1}(h) + ... + b_{i,i-1}(h) + b_{i,i+1}(h) + ... + b_{i,n}(h). \tag{2.1.3a}$$

Recall that A_i represents the set of all players but player i, and therefore, the set A_i contains $(n-1)$ players. Furthermore, $b_{i,j}(h)$ represents the belief that player i holds about player j's strategies. Because the $(n-1)$ players in set A_i are symmetric, we rewrite (2.1.3a) as

$$\sum_{j\in A_i} b_{i,j}(h) = (n-1)b_{i,j}(h). \tag{2.1.3b}$$

Therefore, we express (2.1.3) as

$$\bar{b}_{i,A_i}(h) = b_{i,j}(h). \tag{2.1.3c}$$

Let $\bar{b}_{i,A_i,i}(h)$ be player i's belief about what the other players in set A_i believe about player i's own contribution strategy, conditional on history h, such that

$$\bar{b}_{i,A_i,i}(h) = \frac{1}{n-1}\sum_{j\in A_i} b_{i,j,i}(h), \tag{2.1.4}$$

where

$$\sum_{j\in A_i} b_{i,j,i}(h) = b_{i,1,i}(h) + ... + b_{i,i-1,i}(h) + b_{i,i+1,i}(h) + ... + b_{i,n,i}(h). \tag{2.1.4a}$$

Because the $(n-1)$ players in set A_i are symmetric, we rewrite (2.1.4a) as

$$\sum_{j\in A_i} b_{i,j,i}(h) = (n-1)b_{i,j,i}(h). \tag{2.1.4b}$$

Therefore, we express (2.1.4) as

$$\bar{b}_{i,A_i,i}(h) = b_{i,j,i}(h). \tag{2.1.4c}$$

Due to symmetry, the belief that player i holds about player j's beliefs about player i's contribution strategy, as well as the strategies of the k other players in their set can be expressed as

$$b_{i,j,i}(h) = b_{i,j,k}(h), \tag{2.1.4d}$$

for $i \neq j$ and $j \neq k$. Following Dufwenberg and Kirchsteiger (2004), let $R_{i,j}$ be player i's reciprocity preference towards player j, where $R_{i,j} \geq 0$. However, because player i cannot discriminate between players and can only be "kind" to the group, we represent player i's reciprocity preferences as

$$R_i = \frac{1}{n-1}\sum_{j\in A_i} R_{i,j}. \tag{2.1.5}$$

Because the $(n-1)$ players in set A_i are symmetric,

$$\sum_{j\in A_i} R_{i,j} = (n-1)R_{i,j}, \tag{2.1.5a}$$

and, therefore, we express (2.1.5) as

$$R_i = R_{i,j}. \tag{2.1.5b}$$

Player i's utility function with reciprocity preferences is expressed by

$$
\begin{aligned}
U_i = &\underbrace{y - c_i + \alpha \sum_{l \in A} c_l}_{(i)} \\
&+ \underbrace{\sum_{j \in A_i} R_{i,j} \cdot \kappa_{i,j}(c_i, (b_{i,j}(h))_{j \neq i}) \cdot \lambda_{i,j,i}(b_{i,j}(h), (b_{i,j,k}(h)))_{k \neq j}}_{(ii)} .
\end{aligned}
\tag{2.1.6}
$$

The utility function, (2.1.6), now has two components: part (i), a material payoff as in the initial model, and part (ii), a reciprocity payoff. The reciprocity payoff function is also comprised of two parts

$$
\sum_{j \in A_i} R_{i,j} \cdot \underbrace{\kappa_{i,j}(c_i, (b_{i,j}(h))_{j \neq i})}_{(i)} \cdot \underbrace{\lambda_{i,j,i}(b_{i,j}(h), (b_{i,j,k}(h)))}_{(ii)}{}_{k \neq j}, \tag{2.1.7}
$$

where part (i) represents player i's kindness function and part (ii) represents player i's perceived kindness function. The kindness function, $\kappa_{i,j}(c_i, (b_{i,j}(h))_{j \neq i})$, is an evaluation of player i's kindness towards player j and depends on c_i, player i's own contribution, and $(b_{i,j}(h))_{j \neq i}$, a vector of player i's beliefs about the contributions of each player $j \in A_i$. The perceived kindness function, $\lambda_{i,j,i}(b_{i,j}(h), (b_{i,j,k}(h)))_{k \neq j}$, is an evaluation of player i's perception of the kindness of player j towards player i. Perceived kindness depends on player i's belief of the contribution of player j, $b_{i,j}(h)$, and a vector of their beliefs of what player j believes about player i's contribution strategy and the contribution strategies of the other players, $(b_{i,j,k}(h))_{k \neq j}$. First, we examine player i's kindness function, $\kappa_{i,j}(\cdot)$, part (i) in (2.1.7), where

$$
\begin{aligned}
\kappa_{i,j}(c_i, (b_{i,j}(h))_{j \neq i}) = &\underbrace{\pi_j(c_i, (b_{i,j}(h))_{j \neq i})}_{(i)} \\
&- \underbrace{\frac{1}{2}\left[\underset{c_i'}{max} \pi_j(c_i', (b_{i,j}(h))_{j \neq i}) + \underset{c_i'}{min} \pi_j(c_i', (b_{i,j}(h))_{j \neq i})\right]}_{(ii)} .
\end{aligned}
\tag{2.1.8}
$$

The payoff of each player j is represented by part (i) in (2.1.8). Here, player i chooses a contribution, c_i, given the beliefs that they hold about the contributions of others and the history of the game, $(b_{i,j}(h))_{j \neq i}$. Player i's kindness is evaluated against a reference point, part (ii) in (2.1.8). We suppose that player i could make a certain maximum or minimum payoff happen by choosing $c_i' = y$ or $c_i' = 0$, respectively. From (2.1.8), if part (i) is greater than part (ii), then $\kappa_{i,j}(\cdot) > 0$ which means that player i is "kind" to player j. If the converse is true and part (i) in (2.1.8) is less than part (ii), then $\kappa_{i,j}(\cdot) < 0$ which means that player i is "unkind" to player j. According to (2.1.1), the payoff of player j is thus

$$\pi_j(c_i, (b_{i,j}(h))_{j \neq i}) = y - b_{i,j}(h) + \alpha(c_i + (n-1)b_{i,j}(h)), \tag{2.1.8a}$$

as the sum of the contributions and beliefs about contributions is $c_i + (n-1)b_{i,j}(h)$.

If $c_i' = y$ and $c_i' = 0$, then the maximum and minimum payoffs of player j are

$$\pi_j(y, (b_{i,j}(h))_{j \neq i}) = y - b_{i,j}(h) + \alpha(y + (n-1)b_{i,j}(h)), \tag{2.1.8b}$$

and

$$\pi_j(0, (b_{i,j}(h))_{j \neq i}) = y - b_{i,j}(h) + \alpha(0 + (n-1)b_{i,j}(h)). \tag{2.1.8c}$$

Substituting these expressions into (2.1.8) and simplifying, we obtain

$$\kappa_{i,j}(c_i, (b_{i,j}(h))_{j \neq i}) = \alpha\left(c_i - \frac{1}{2}y\right). \tag{2.1.8d}$$

Next, we examine player i's perceived kindness function, part (ii) in (2.1.7), that we define as the level of player i's perceived kindness of player j towards player i, where

$$\lambda_{i,j,i}(b_{i,j}(h),(b_{i,j,k}(h))_{k\neq j}) = \underbrace{\pi_i(b_{i,j}(h),(b_{i,j,k}(h))_{k\neq j})}_{(i)} \\ \underbrace{-\frac{1}{2}\left[\underset{b'_{i,j}(h)}{max}\pi_i\left(b'_{i,j}(h),(b_{i,j,k}(h))_{k\neq j}\right) + \underset{b'_{i,j}(h)}{min}\pi_i\left(b'_{i,j}(h),(b_{i,j,k}(h))_{k\neq j}\right)\right]}_{(ii)}. \tag{2.1.9}$$

The payoff that player i believes they will receive is represented by part (i) in (2.1.9). This is evaluated against a reference point, part (ii) in (2.1.9). We suppose that player i can evaluate what they believe their payoff would be if they believe that player j could make a certain maximum or minimum payoff happen by choosing $b'_{i,j}(h) = y$ or $b'_{i,j}(h) = 0$, respectively. From (2.1.9), if part (i) is greater than part (ii), then $\lambda_{i,j,i}(\cdot) > 0$ and player i perceives player j to be "kind" towards them. If the converse is true and part i in (2.1.9) is less than part (ii), then $\lambda_{i,j,i}(\cdot) < 0$ and player i perceives player j to be "unkind" towards them. From (2.1.1), we thus expect payoff, part (i) in (2.1.9), to be

$$\pi_i(b_{i,j}(h),(b_{i,j,k}(h))_{k\neq j}) = y - b_{i,j,i}(h) \\ + \alpha(b_{i,j}(h) + b_{i,j,i}(h) + (n-2)b_{i,j,k}(h)), \tag{2.1.9a}$$

as the sum of the beliefs about contributions is now $b_{i,j}(h) + b_{i,j,i}(h) + (n-2)b_{i,j,k}(h)$. If $b'_{i,j}(h) = y$ and $b'_{i,j}(h) = 0$, then the maximum and minimum payoffs of player i are

$$\pi_i(y,(b_{i,j,k}(h))_{k\neq j}) = y - b_{i,j,i}(h) + \alpha(y + b_{i,j,i}(h) + (n-2)b_{i,j,k}(h)), \tag{2.1.9b}$$

and

$$\pi_i(0,(b_{i,j,k}(h))_{k\neq j}) = y - b_{i,j,i}(h) + \alpha(0 + b_{i,j,i}(h) + (n-2)b_{i,j,k}(h)). \tag{2.1.9c}$$

Since players are symmetric, $b_{i,j,i}(h) = b_{i,j,k}(h)$. Substituting these expressions into (2.1.9) and simplifying, we obtain

$$\lambda_{i,j,i}(b_{i,j}(h), (b_{i,j,k}(h))_{k \neq j}) = \alpha \left(b_{i,j}(h) - \frac{1}{2}y \right). \tag{2.1.9d}$$

We can now substitute these simplified expressions for kindness and perceived kindness, (2.1.8d) and (2.1.9d), respectively, into (2.1.6). The reciprocity payoff is thus

$$\sum_{j \in A_i} R_{i,j} \alpha^2 \left(c_i - \frac{1}{2}y \right) \left(b_{i,j}(h) - \frac{1}{2}y \right). \tag{2.1.10}$$

Since there are $(n-1)$ symmetric players in set A_i, and using (2.1.5b) we express the reciprocity payoff as

$$(n-1) R_i \alpha^2 \left(c_i - \frac{1}{2}y \right) \left(b_{i,j}(h) - \frac{1}{2}y \right). \tag{2.1.11}$$

Using (2.1.3b), the reciprocity payoff is thus

$$R_i \alpha^2 \left(c_i - \frac{1}{2}y \right) \left(\sum_{j \in A_i} b_{i,j}(h) - \frac{n-1}{2}y \right). \tag{2.1.12}$$

Note that, like the material payoff function, part (i) in (2.1.6), the reciprocity payoff function, (2.1.12), is increasing in the MPCR, α, which aligns with the literature (i.e., Ledyard, 1995). Further, findings by Ledyard (1995) and Chaudhuri (2011) and experimental evidence from Bagnoli and McKee (1991) suggest that group size does not have an effect on increasing cooperation in public goods games. However, as group size, n, increases, the reciprocity payoff function, (2.1.11), is increasing if the kindness and perceived kindness functions have matching signs.

The utility maximization problem represented by (2.1.6) for player i is now expressed as

$$\underset{c_i}{max}\left\{y - c_i + \alpha \sum_{l \in A} c_l + R_i \alpha^2 \left(c_i - \frac{1}{2} y\right) \left(\sum_{j \in A_i} b_{i,j}(h) - \frac{n-1}{2} y\right)\right\}. \tag{2.1.13}$$

This expression is similar to other applications of Dufwenberg and Kirchsteiger (2004)'s model, including Dufwenberg et al. (2011). The first-order condition is

$$-1 + \alpha + R_i \alpha^2 \left(\sum_{j \in A_i} b_{i,j}(h) - \frac{n-1}{2} y\right) = 0. \tag{2.1.14}$$

In the baseline case, we assume that players are symmetric in their reciprocity preferences, thus we can denote $R_i = R$. Furthermore, we assume rational expectations such that player i's beliefs about the contribution strategies of other players in set A_i are correct, $b_{i,j}(h) = c_j$. We can rewrite the first-order condition (2.1.14) as

$$-1 + \alpha + R\alpha^2 \left(\sum_{j \in A_i} c_j - \frac{n-1}{2} y\right) = 0. \tag{2.1.14a}$$

This first-order condition is different from the baseline case with no reciprocity preferences, first-order condition (2.1.2), and now depends on more factors. When we account for reciprocity preferences, the first-order condition depends on reciprocity preference, contributions of others, and group size. The function, (2.1.14a), can be strictly negative and downward sloping, in which case the maximum is found at $c_i = 0$. The function can be equal to zero (interior solution) such that the function is horizontal and the maximum exists anywhere along $c_i \in [0, y]$. The function can be strictly positive and upward sloping, in which case the maximum is found at $c_i = y$.

We derive players' best response functions to evaluate contribution strategies and determine potential equilibria. Given the contributions of the $(n-1)$

players in set A_i, (i.e., $\sum_{j \in A_i} c_j$), player i's best response function is

$$BR_i\left((c_j)_{j \neq i}\right) = \begin{cases} 0 & \text{if} \quad \sum_{j \in A_i} c_j < \frac{1-\alpha}{\alpha^2 R} + \frac{n-1}{2} y \\ [0, y] & \text{if} \quad \sum_{j \in A_i} c_j = \frac{1-\alpha}{\alpha^2 R} + \frac{n-1}{2} y \\ y & \text{if} \quad \sum_{j \in A_i} c_j > \frac{1-\alpha}{\alpha^2 R} + \frac{n-1}{2} y \end{cases} \tag{2.1.15}$$

We specify player j's utility maximization problem, where player j is symmetric to player i. Let $A_j = \{1, ..., j-1, j+1, ..., n\}$ and $k \in A_j$. Player j chooses c_j that solves

$$\begin{aligned} \underset{c_j}{max} \Bigg\{ y - c_j + \alpha \sum_{l \in A} c_l \\ + R_j \alpha^2 \left(c_j - \frac{1}{2} y \right) \left(\sum_{k \in A_j} b_{j,k}(h) - \frac{n-1}{2} y \right) \Bigg\}. \end{aligned} \tag{2.1.16}$$

As described above, players are symmetric in their reciprocity preferences, and thus we can denote $R_j = R$. The first-order condition is

$$-1 + \alpha + R\alpha^2 \left(\sum_{k \in A_j} b_{j,k}(h) - \frac{n-1}{2} y \right) = 0. \tag{2.1.17}$$

Beliefs are correct in equilibrium, $b_{j,k}(h) = c_k$. The best response function for each player j, $j \neq i$, is thus

$$BR_j\left((c_k)_{k \neq j}\right) = \begin{cases} 0 & \text{if} \quad \sum_{k \in A_j} c_k < \frac{1-\alpha}{\alpha^2 R} + \frac{n-1}{2} y \\ [0, y] & \text{if} \quad \sum_{k \in A_j} c_k = \frac{1-\alpha}{\alpha^2 R} + \frac{n-1}{2} y \\ y & \text{if} \quad \sum_{k \in A_j} c_k > \frac{1-\alpha}{\alpha^2 R} + \frac{n-1}{2} y \end{cases} \tag{2.1.18}$$

First, we determine if there exists an interior solution contribution choice,

$c \in [0, y]$, that satisfies player i's best response function, (2.1.15), and each player j's best response function, (2.1.18). If all players choose $c \in [0, y]$, then $\sum_{j \in A_i} c_j = (n-1)c$ and $\sum_{k \in A_j} c_k = (n-1)c$, which from (2.1.15) and (2.1.18) is possible if $(n-1)c = \frac{1-\alpha}{\alpha^2 R} + \frac{n-1}{2}y$. Thus, we find the interior solution

$$c^* = \frac{1-\alpha}{\alpha^2}\frac{1}{(n-1)R} + \frac{1}{2}y, \tag{2.1.19}$$

where $c_i = c_j = c^*$, and $0 \leq c^* \leq y$ if $R \geq \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}$.

Applying Dufwenberg and Kirchsteiger (2004)'s Sequential Reciprocity Equilibrium (SRE) concept, a profile of contribution strategies, $c^* = (c_i^*)_{i \in A}$ is a SRE if

$$U_i\left(c_i^*, (b_{i,j}(h), (b_{i,j,k}(h))_{k \neq j})_{j \neq i}\right) \geq U_i\left(c_i', (b_{i,j}(h), (b_{i,j,k}(h))_{k \neq j})_{j \neq i}\right) \;\; \forall \;\; c_i' \neq c_i^*,$$

and $b_{i,j} = c_j^* \;\forall\; j \neq i$ and $b_{i,j,k} = c_k^* \;\forall\; k \neq j, j \neq i$. The equilibrium concept states that given correct beliefs, a profile of contribution strategies is a SRE if, at a given history h, and given each player's beliefs, a player's contribution choice maximizes their utility.

From players' best response functions, (2.1.15) and (2.1.18), and Dufwenberg and Kirchsteiger (2004)'s SRE concept, there are three possible equilibria. The traditional public goods game Nash equilibrium $(0, ..., 0, ..., 0)$ is always satisfied because $0 < \frac{1-\alpha}{\alpha^2}\frac{1}{R} + \frac{n-1}{2}y$ is always satisfied as $R \geq 0$. The social optimum equilibrium, $(y, ..., y, ..., y)$, is satisfied if $R > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}$. The interior solution equilibrium, $(c^*, ..., c^*, ..., c^*)$, with $0 \leq c^* \leq y$, is satisfied if $R \geq \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}$.

There are now two additional equilibria compared with the baseline with no reciprocity preferences, where we only found the traditional public goods game outcome in equilibrium. When we account for reciprocity preferences, we also have the social optimum as an equilibrium, and an interior solution equilibrium at $(c^*, ..., c^*, ..., c^*)$ with $0 \leq c^* \leq y$. Note that the interior solution equilibrium is a specific contribution amount, c_i^* or c_j^*, and not a multiplicity of equilibria[4].

These additional equilibria suggest that reciprocity preferences play an important role in cooperative behaviour and can encourage socially optimal outcomes. The equilibria conditions presented above depend on several parameters, including α, the MPCR. The MPCR is a key feature in public goods games and ensures that players are faced with a social dilemma when making their contribution choice. In Figure 2.2, we represent the different equilibria in a graph (α, R). The function $f_1(\alpha)$ separates equilibria areas, where in *Area I* only the traditional public goods game Nash equilibrium exists and, in *Area II*, the additional cooperative equilibria are also present.

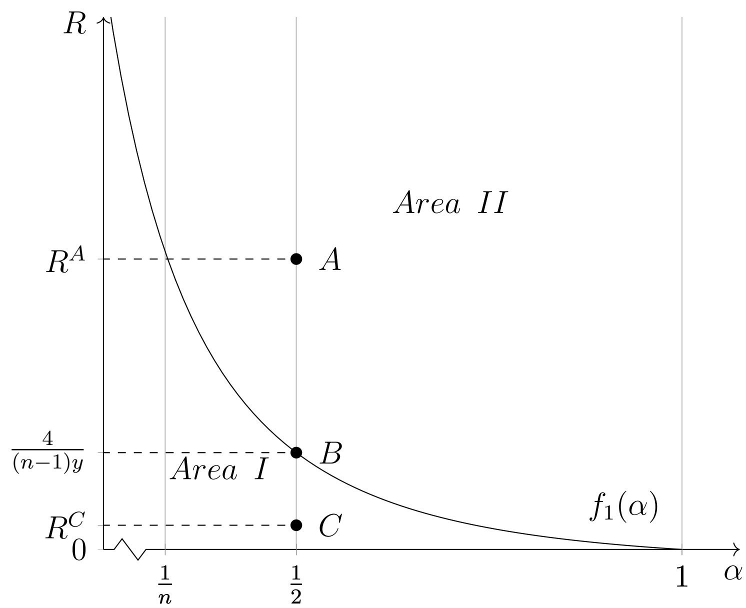

Figure 2.2: Baseline Model with Reciprocity Preferences

Let $f_1(\alpha) \equiv \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}$, where $0 < \alpha < 1 < n\alpha$ and $n \geq 3$. Area I denotes the equilibrium $(0, ..., 0, ..., 0)$. Area II denotes the equilibria $(0, ..., 0, ..., 0)$, $(y, ..., y, ..., y)$, and $(c^*, ..., c^*, ..., c^*)$. Along the curve $f_1(\alpha)$, the equilibrium $(c^*, ..., c^*, ..., c^*)$ also exists. To illustrate the importance of the MPCR on the

[4]Dufwenberg et al. (2011), a public goods game application of Dufwenberg and Kirchsteiger (2004), ignore the possibility of interior solutions. Our findings of a unique interior solution equilibrium presents a novel contribution on its own and highlights that there exists cooperative equilibria beyond the predictions of traditional game theory.

equilibria in a public goods game, we present two examples, when $\alpha = \frac{1}{2}$ and $\alpha = \frac{2}{3}$. When $\alpha = \frac{1}{2}$, then $f_1(\frac{1}{2}) = \frac{4}{(n-1)y}$. When a player's reciprocity preferences, R, are greater than $f_1(\alpha)$ (point A), then the player is in *Area II* and there are multiple possible equilibria. When a player's reciprocity preferences, R, are equal to $f_1(\alpha)$ (point B), then the equilibrium, $(c^*, ..., c^*, ..., c^*)$, represented along the curve of $f_1(\alpha)$ is satisfied, as is $(0, ..., 0, ..., 0)$ which is satisfied everywhere. When a player's reciprocity preferences, R, are less than $f_1(\alpha)$ (point C), then the player is in *Area I* and there only exists the traditional public goods game Nash equilibrium, $(0, ..., 0, ..., 0)$. Furthermore, as the MPCR, α, increases, the curve $f_1(\alpha)$ shifts inwards. For example, if $\alpha = \frac{2}{3}$, then $f_1(\frac{2}{3}) = \frac{3}{2(n-1)y}$. This inward shift of the threshold means that *Area II* expands and there exists a greater range of values for R which can satisfy the more cooperative outcomes. This aligns with findings from reviews by Ledyard (1995) and Chaudhuri (2011) on public goods games and mechanisms for cooperation in social dilemmas.

2.2 Direct Reciprocity

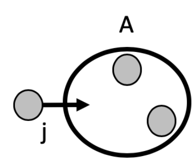

Figure 2.3: Direct Reciprocity

Consider a variation of the previous case where players now interact sequentially, Figure 2.3. In this sequential game, $(n-1)$ players play a simultaneous game in which each player $i \in A_j = \{1, ..., j-1, j+1, ...n\}$, where $A_j \subset A$, chooses a contribution, $c_i \in [0, y]$. Player j then observes the average contribution of these $(n-1)$ players, $\frac{1}{n-1}\sum_{i \in A_j} c_i$, and chooses their contribution, c_j. The rationale for this setting follows Bardsley and Sausgruber (2005)'s work on reciprocity and conformity. In Bardsley and Sausgruber's experiment design, the authors incorporate an explicit sequential structure to a public goods game, with opportunity for observation of other players' contribution amounts. While Bardsley and Sausgruber provide the selected player with the full vector of contributions from the other players in the set, in our model, the selected player, player j, observes the average contribution instead of a vector of contributions. In this way, player j may not know with certainty if one player is more or less "kind" than others in the set A_j. In line with the growing body of work on conditional cooperation discussed by Chaudhuri (2011), we hypothesize that as average contributions increase, player j will contribute more as well.

2.2.1 No Reciprocity Preferences

In the first period, each player i in set A_j simultaneously chooses $c_i \in [0, y]$. In the second period, player j observes $\frac{1}{n-1} \sum_{i \in A_j} c_i$ and then chooses $c_j \in [0, y]$. With no reciprocity preferences, each player i receives a material payoff (2.1.1), and the material payoff for player j is analogous, $\pi_j = y - c_j + \alpha \sum_{l \in A} c_l$. We assume that there is no discounting between periods. In the second period, player j chooses c_j that solves

$$\max_{c_j} \left\{ y - c_j + \alpha \sum_{l \in A} c_l \right\}. \tag{2.2.1}$$

The first-order condition is

$$-1 + \alpha < 0. \tag{2.2.2}$$

Therefore, we are at a corner solution where the optimal contribution choice is $c_j^* = 0$ because the maximization problem is linear.

In the first period, for the $(n-1)$ other players, each player $i \in A_j$ solves

$$\max_{c_i} \left\{ y - c_i + \alpha \sum_{l \in A} c_l \right\}. \tag{2.2.3}$$

The first-order condition is

$$-1+\alpha<0. \tag{2.2.4}$$

We are again at a corner solution where the optimal contribution choice is $c_i^* = 0 \ \forall \ i \in A_j$, such that $c_1^* = ... = c_{j-1}^* = c_{j+1}^* = ... = c_n^* = 0$. Not surprisingly, the subgame perfect Nash equilibrium is $(c_1^*, ..., c_n^*) = (0, ..., 0)$. Even though the game is sequential, in the absence of reciprocity preferences, none of the players have an incentive to contribute a positive amount to the group account and we are again at the traditional public goods game outcome.

2.2.2 Reciprocity Preferences

As in Section 2.1.2, incorporating reciprocity preferences into players' utility functions takes the same form in this sequential variation of the model. The equations and their components only differ in their subscripts, i.e., the player of interest and their beliefs about the other $(n-1)$ players. The history of the game is analogous to the one outlined in Section 2. The history of the game, $h = (c_1, ..., c_{j-1}, c_{j+1}, ..., c_n)$, is observed by player j in the form of the average contribution of others, $\frac{1}{n-1}\sum_{i \in A_j} c_i$, such that when player j observes $\frac{1}{n-1}\sum_{i \in A_j} c_i = 0$, player j knows with probability 1 that each player $i \in A_j$ chose $c_i = 0$. When player j observes $\frac{1}{n-1}\sum_{i \in A_j} c_i = y$, player j knows with probability 1 that each player $i \in A_j$ chose $c_i = y$. The history $h = (c, ..., c)$ represents a multiplicity of equilibria. When player j observes $\frac{1}{n-1}\sum_{i \in A_j} c_i = c$, where $c \in]0, y[$, player j does not know with certainty what each player $i \in A_j$ contributed. Thus, player j must form beliefs, $b_{j,i}(h)$, about the other players' contribution strategies. Player j's utility function with reciprocity preferences is expressed by (2.1.6) in Section 2.1.2.

Because the game is a psychological game with sequentiality, we cannot solve it by backward induction (Geanakoplos et al., 1989). Recall that backward induction fails in psychological games because once a decision node is reached it does not necessarily provide players with adequate information needed for further decision-making; the decision node only identifies the history or previous play and not the beliefs of the players (Geanakoplos et al., 1989). Instead, we examine players' best response functions, apply Dufwenberg and Kirchsteiger (2004)'s SRE concept, and evaluate potential equilibria.

In the second period, player j chooses c_j conditional on history h to solve

$$\begin{aligned} U_j = & \underbrace{y - c_j + \alpha \sum_{l \in A} c_l}_{(i)} \\ & + \underbrace{\sum_{i \in A_j} R_{j,i} \cdot \kappa_{j,i}(c_j, (b_{j,i}(h))_{i \neq j}) \cdot \lambda_{j,i,j}(b_{j,i}(h), (b_{j,i,k}(h))_{k \neq i})}_{(ii)} . \end{aligned} \tag{2.2.5}$$

Similar to (2.1.7) in Section 2.1.2, player j's reciprocity payoff, part (ii) in (2.2.5), takes the form

$$R_j \alpha^2 \left(c_j - \frac{1}{2}y\right) \left(\sum_{i \in A_j} b_{j,i}(h) - \frac{n-1}{2}y\right). \tag{2.2.6}$$

Player j's utility maximization problem can now be expressed as

$$\begin{aligned} \max_{c_j} \Bigg\{ & y - c_j + \alpha \sum_{l \in A} c_l \\ & + R_j \alpha^2 \left(c_j - \frac{1}{2}y\right) \left(\sum_{i \in A_j} b_{j,i}(h) - \frac{n-1}{2}y\right) \Bigg\}. \end{aligned} \tag{2.2.7}$$

The first-order condition is thus

$$-1 + \alpha + R_j \alpha^2 \left(\sum_{i \in A_j} b_{j,i}(h) - \frac{n-1}{2}y\right) = 0. \tag{2.2.8}$$

As we assume rational expectations in equilibrium, player j's beliefs about each player i's contributions are correct such that $b_{j,i}(h) = c_i$. We rewrite the first-order condition (2.2.8), as

$$-1 + \alpha + R_j\alpha^2 \left(\sum_{i \in A_j} c_i - \frac{n-1}{2} y \right) = 0. \tag{2.2.8a}$$

The choice of c_j depends on player j's beliefs about the average contribution of the other players, conditional on the history of the game, $b_{j,i}(h)$. Because the utility function is linear in the choice variable, c_j, the first-order condition can be positive, negative, or null. As described above, while player j observes the average contribution of other players through $\frac{1}{n-1} \sum_{i \in A_j} c_i$, there are only two cases in which player j knows with certainty what each player i chose: when $\frac{1}{n-1} \sum_{i \in A_j} c_i$ is 0 or y. When $\frac{1}{n-1} \sum_{i \in A_j} c_i = c \in \ ]0, y[$, player j does not know with certainty what each player i has contributed to the group account and must form beliefs about their contributions. The best response function for player j is thus

$$BR_j\left((c_i)_{i \neq j}\right) = \begin{cases} 0 & \text{if} \quad \sum_{i \in A_j} c_i < \dfrac{1-\alpha}{\alpha^2 R_j} + \dfrac{n-1}{2} y \\ [0, y] & \text{if} \quad \sum_{i \in A_j} c_i = \dfrac{1-\alpha}{\alpha^2 R_j} + \dfrac{n-1}{2} y \\ y & \text{if} \quad \sum_{i \in A_j} c_i > \dfrac{1-\alpha}{\alpha^2 R_j} + \dfrac{n-1}{2} y \end{cases} \tag{2.2.9}$$

In the first period, because first- and second-order beliefs are correct, each player i correctly anticipates what player j's best response will be in the second period given the history of the game. Therefore, each player i chooses c_i to solve

$$\max_{c_i} \Big\{ y - c_i + \alpha \sum_{l \in A} c_l + \sum_{k \in A_i} R_{i,k} \cdot \kappa_{i,k}(c_i, (b_{i,k}(h))_{k \neq i}) \cdot \lambda_{i,k,i}(b_{i,k}(h), (b_{i,k,l}(h))_{l \neq k}) \Big\}. \tag{2.2.10a}$$

As before, if we substitute the simplified expressions for player i's kindness and perceived kindness functions in (2.2.10), we obtain

$$\underset{c_i}{max}\left\{y - c_i + \alpha \sum_{l \in A} c_l + R_i \alpha^2 \left(c_i - \frac{1}{2}y\right)\left(\sum_{k \in A_i} b_{i,k} - \frac{n-1}{2}y\right)\right\}. \quad (2.2.10b)$$

The choice of c_i depends on player i's beliefs about the average contribution of the other players in the set $A_i = \{1, ..., i-1, i+1, ..., n\}$, where $A_i \subset A$. Because player i is making their choice at the same time as the other players in set A, player i does not know the history of the game.

The first-order condition is

$$-1 + \alpha + R_i \alpha^2 \left(\sum_{k \in A_i} b_{i,k}(h) - \frac{n-1}{2}y\right) = 0. \quad (2.2.11)$$

We assume rational expectations, therefore, in equilibrium player i's beliefs about each player k's contributions are correct such that $b_{i,k}(h) = c_k$. We can rewrite the first-order condition (2.2.11) as

$$-1 + \alpha + R_i \alpha^2 \left(\sum_{k \in A_i} c_k - \frac{n-1}{2}y\right) = 0. \quad (2.2.11a)$$

Given the contributions of the $(n-1)$ players in the set A_i, the best response function for player i is

$$BR_i\left((c_k)_{k \neq i}\right) = \begin{cases} 0 & \text{if} \quad \sum_{k \in A_i} c_k < \frac{1-\alpha}{\alpha^2 R_i} + \frac{n-1}{2}y \\ [0, y] & \text{if} \quad \sum_{k \in A_i} c_k = \frac{1-\alpha}{\alpha^2 R_i} + \frac{n-1}{2}y \\ y & \text{if} \quad \sum_{k \in A_i} c_k > \frac{1-\alpha}{\alpha^2 R_i} + \frac{n-1}{2}y \end{cases} \quad (2.2.12)$$

Because player j does not know what each player i has contributed to the group account when $\frac{1}{n-1}\sum_{i\in A_j} c_i = c \in]0,y[$, we now consider cases with different histories of the game: $h = (0,...,0)$, $h = (y,...,y)$, and $h = (c,...,c)$.

2.2.2.1 Case 1

First, if $h = (0,...,0)$, then player j believes that player i contributes nothing to the group account, $b_{j,i}(h) = 0$, and player j observes $\frac{1}{n-1}\sum_{i\in A_j} c_i = 0$. This means that with probability 1, player i contributes nothing to the group account, $c_i = 0$, and player j knows with certainty that in the first period, $c_1^* = 0,...,c_{j-1}^* = 0, c_{j+1}^* = 0,...,c_n^* = 0$. Therefore, the best response function of player j is

$$BR_j\left((c_i)_{i\neq j}\right) = \begin{cases} 0 & \text{if} \quad 0 < \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y \\ [0,y] & \text{if} \quad 0 = \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y \\ y & \text{if} \quad 0 > \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y \end{cases} \tag{2.2.13}$$

Note that $c_i = 0 \ \forall \ i \in A_j$, therefore the potential equilibria when the history of the game is $h = (0,...,0)$ occur at $(0,...,0,...,0)$, $(0,...,c_j^*,...,0)$, and $(0,...,y,...,0)$. However, an interior solution, c_j^*, does not exist because $R_j \geq 0$ and $n \geq 3$, so the condition $0 = \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y$ cannot be satisfied. From player j's best response function, we see that neither playing y nor $c_j \in [0,y]$ is possible. Therefore, when the history of the game is $h = (0,...,0)$, there only exists one equilibrium, $(0,...,0,...,0)$, as

$$\frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y > 0, \tag{2.2.14}$$

and

$$\frac{1-\alpha}{\alpha^2 R_i} + \frac{n-1}{2}y > 0 \tag{2.2.15}$$

are always satisfied for $\alpha > 0$, $R_j \geq 0$, $R_i \geq 0$, and $y \geq 0$. Even when reciprocity preferences are considered, if the other players in the set A_j do not contribute anything to the group account and players correctly anticipate that others are not contributing, then there exists only one equilibrium, the subgame perfect Nash equilibrium, as predicted by standard game theory as well.

2.2.2.2 Case 2

Second, if $h = (y, ..., y)$, then player j believes that player i is contributing their full endowment to the group account, $b_{j,i}(h) = y$, and player j observes $\frac{1}{n-1} \sum_{i \in A_j} c_i = y$. This means that with probability 1, $c_i = y$, and player j knows with certainty that in the first period, $c_1^* = y, ..., c_{j-1}^* = y, c_{j+1}^* = y, ..., c_n^* = y$. Therefore, the best response function of player j is

$$BR_j\left((c_i)_{i \neq j}\right) = \begin{cases} 0 & \text{if} \quad (n-1)y < \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y \\ [0, y] & \text{if} \quad (n-1)y = \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y \\ y & \text{if} \quad (n-1)y > \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y \end{cases} \tag{2.2.16}$$

Rearranging the terms, we express player j's best response function as

$$BR_j\left((c_i)_{i \neq j}\right) = \begin{cases} 0 & \text{if} \quad R_j < \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1)y} \\ [0, y] & \text{if} \quad R_j = \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1)y} \\ y & \text{if} \quad R_j > \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1)y} \end{cases} \tag{2.2.16a}$$

Since $c_i = y \;\; \forall \;\; i \in A_j$, the potential equilibria when the history of the game is $h = (y, ..., y)$ occur at $(y, ..., 0, ..., y)$, $(y, ..., c_j^*, ..., y)$, and $(y, ..., y, ..., y)$. First, to determine if c_j^* exists as an interior solution, we examine the best response function, (2.2.12). If all players but player j choose y, then $\sum_{k \in A_i} c_k = (n-2)y + c_j$, and from (2.2.12) $(n-2)y + c_j \geq \frac{1-\alpha}{\alpha^2 R_i} + \frac{n-1}{2}y$. The interior solution, c_j^*, exists at

$$c_j^* = \frac{1-\alpha}{\alpha^2 R_i} - \frac{n-3}{2}y, \tag{2.2.17}$$

where $0 \leq c_j^* \leq y$ is satisfied if

$$\frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y} \leq R_i \leq \frac{1-\alpha}{\alpha^2}\frac{2}{(n-3)y}, \tag{2.2.18}$$

and $n > 3$.

The free-riding equilibrium $(y, ..., 0, ..., y)$ is satisfied when for player j

$$0 \leq R_j < \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.2.19}$$

and for each player i

$$R_i > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-3)y}, \tag{2.2.20}$$

and $n > 3$. Free-riding is an optimal strategy for player j when their reciprocity preferences are sufficiently weak and when each player i has sufficiently strong reciprocity preferences. When this equilibrium occurs, each player i has correctly anticipated that player j will free-ride, but each player i's strong reciprocity preferences towards the group as a whole drive their cooperative behaviour of contributing the full endowment.

The equilibrium $(y, ..., c_j^*, ..., y)$ is not possible as it violates player i's best response function.

The social optimum equilibrium $(y, ..., y, ..., y)$ is satisfied when for player j

$$R_j > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.2.21}$$

and for each player i

$$R_i > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}. \tag{2.2.22}$$

The social optimum is a possible equilibrium if player j and each player i have sufficiently strong reciprocity preferences. Cooperation is possible in sequential play with reciprocity preferences.

2.2.2.3 Case 3

Finally, if $h = (c, ..., c)$, then player j cannot determine with certainty what the other players in set A_j have contributed. While $h = (c, ..., c)$ represents a multiplicity of interior contributions, here we examine the special case of the interior solution equilibrium, $h = (c^*, ..., c^*)$. Player j observes $\frac{1}{n-1}\sum_{i\in A_j} c_i = c \in]0, y[$, and thus player j's best response function is

$$BR_j\left((c_i)_{i\neq j}\right) = \begin{cases} 0 & \text{if } \sum_{i\in A_j} c_i < \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y \\ [0, y] & \text{if } \sum_{i\in A_j} c_i = \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y \\ y & \text{if } \sum_{i\in A_j} c_i > \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y \end{cases} \tag{2.2.23}$$

According to player j's best response function, (2.2.23), and given each player $i \in A_j$ are symmetric, $(n-1)c_i = \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y$. The interior solution, c_i^*, exists and is expressed as

$$c_i^* = \frac{1-\alpha}{\alpha^2}\frac{1}{(n-1)R_j} + \frac{1}{2}y, \tag{2.2.24}$$

where $0 \leq c_i^* \leq y$ if $R_j \geq \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}$. According to player i's best response function, (2.2.12), $(n-2)c_i + c_j = \frac{1-\alpha}{\alpha^2 R_i} + \frac{n-1}{2}y$. The interior solution, c_j^*, exists and is expressed as

$$c_j^* = \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_i} - \frac{n-2}{n-1}\frac{1}{R_j}\right) + \frac{1}{2}y, \tag{2.2.25}$$

where $c_j^* \leq y$ if

$$\frac{1}{R_j} \geq \underbrace{\frac{n-1}{n-2}\frac{1}{R_i} - \frac{\alpha^2}{1-\alpha}y\frac{n-1}{2(n-2)}}_{(i)}. \tag{2.2.26}$$

The right-hand side, part (i) in (2.2.26), can be positive or negative. If part $(i) \geq 0$, then

$$0 \leq R_i < \frac{1-\alpha}{\alpha^2}\frac{2}{y}, \tag{2.2.27}$$

and the condition (2.2.26) must be satisfied. If part $(i) < 0$, then

$$R_i > \frac{1-\alpha}{\alpha^2}\frac{2}{y}, \tag{2.2.28}$$

and the condition (2.2.26) is always satisfied. Furthermore, $c_j^* \geq 0$ if

$$\frac{1}{R_j} \leq \frac{n-1}{n-2}\frac{1}{R_i} + \frac{\alpha^2}{1-\alpha}y\frac{n-1}{2(n-2)}. \tag{2.2.29}$$

For the conditions $0 \leq c_j^* \leq y$ to be met at the same time, then we must have that $R_i > \frac{1-\alpha}{\alpha^2}\frac{2}{y}$. This means that player i's reciprocity preferences need to be sufficiently strong for c_j^* to exist as an eligible interior solution for player j (i.e., player j takes each player i's reciprocity preferences into account when choosing their contribution amount). Note that $c_j^* < c_i^*$ if $R_j < R_i$. When the history of the game is $h = (c, ..., c)$, then the possible equilibria are $(c_i^*, ..., 0, ..., c_i^*)$, $(c_i^*, ..., c_j^*, ..., c_i^*)$, and $(c_i^*, ..., y, ..., c_i^*)$. However, the equilibrium $(c_i^*, ..., 0, ..., c_i^*)$ is not satisfied as it violates the best response function for player j, (2.2.23). As well, the equilibrium $(c_i^*, ..., y, ..., c_i^*)$ is not satisfied as it violates the best response function of player j, (2.2.23).

The interior solution equilibrium, $(c_i^*, ..., c_j^*, ..., c_i^*)$, is satisfied when for player j

$$\frac{1}{R_j} \geq \frac{n-2}{n-1}R_i + \frac{\alpha^2}{1-\alpha}\frac{n-1}{2(n-2)}y, \tag{2.2.30}$$

and

$$R_i > \frac{1-\alpha}{\alpha^2}\frac{2}{y}, \tag{2.2.31}$$

such that the conditions are satisfied for $0 \leq c_j^* \leq y$. For player i

$$R_j \geq \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.2.32}$$

such that the conditions are met for $0 \leq c_i^* \leq y$. The interior solution equilibrium is possible if players have sufficiently strong reciprocity preferences.

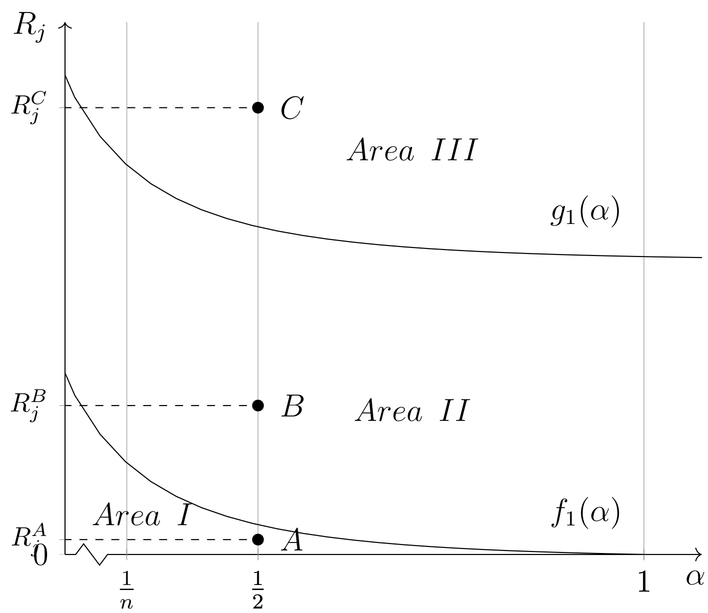

Figure 2.4: Direct Reciprocity

In Figure 2.4, we represent equilibria in all histories of the game in a graph (α, R_j). Let $f_1(\alpha) \equiv \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}$, and $g_1(\alpha) \equiv \frac{n-1}{n-2}R_i + \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}$, where $f_1(\alpha)$ and $g_1(\alpha)$ are equilibria conditions. In *Area I*, the equilibrium for Case 1 - $h = (0, ..., 0)$ is $(0, ..., 0, ..., 0)$, and the equilibrium for Case 2 - $h = (y, ..., y)$

is $(y, ..., 0, ..., y)$. In *Area II*, the equilibrium for Case 1 - $h = (0, ..., 0)$ is $(0, ..., 0, ..., 0)$, and the equilibrium for Case 2 - $h = (y, ..., y)$ is $(y, ..., y, ..., y)$. In *Area III*, the same equilibria hold as in *Area II*, and the equilibrium for Case 3 - $h = (c, ..., c)$ is $(c_i^*, ..., c_j^*, ..., c_i^*)$.

We examine how the equilibria solutions change as player j's reciprocity preferences shift depending on the history of the game. Let $\alpha = \frac{1}{2}$ such that $f_1(\frac{1}{2}) = \frac{4}{(n-1)y}$. The traditional public goods game subgame perfect Nash equilibrium is possible at all values of $R_j \geq 0$. At point A, player j's reciprocity preference $R_j^A < f_1(\alpha)$, and player j will contribute nothing as they have weak reciprocity preferences. The potential equilibria at point A are the traditional public goods game subgame perfect Nash equilibrium, and the free-riding equilibrium. As player j's reciprocity preferences increase to point B, the social optimum becomes a potential equilibrium. In *Area II*, $R_j^B > f_1(\alpha)$ which means that if player j has sufficiently strong reciprocity preferences, then player j will either be at the traditional public goods game subgame perfect Nash equilibrium or the social optimum equilibrium.

When player j's reciprocity preferences increase further to point C, then the interior solution for each player i and player j, $(c_i^*, ..., c_j^*, ..., c_i^*)$, also becomes a potential equilibrium, i.e., $R_j^C > g_1(\alpha)$. The requirements for the interior solution equilibrium to be satisfied are much more stringent than other equilibria as the conditions $0 \leq c_i^* \leq y$ and $0 \leq c_j^* \leq y$ must also be satisfied (i.e., condition $R_j \geq \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1)y}$ for $0 \leq c_i^* \leq y$, and conditions (2.2.26) and (2.2.29) for player j). A possible explanation for this shift in equilibria at a greater reciprocity preference level for player j is that the reciprocity preferences for each player i must also be large and each player i correctly anticipates that player j will not free-ride when observing $\frac{1}{n-1} \sum c_i = c$.

2.2.3 Comparative Statics

We examine several cases and compare how equilibria behaviour may be affected as certain parameters change. We use Mathematica (Wolfram Research, Inc, 2020), a technical computing system, to visualize the data. Mathematica code for reproduction of the interactive visualization is available in Appendix

A.1.1. Static snapshots and descriptions with parameter values are presented below.

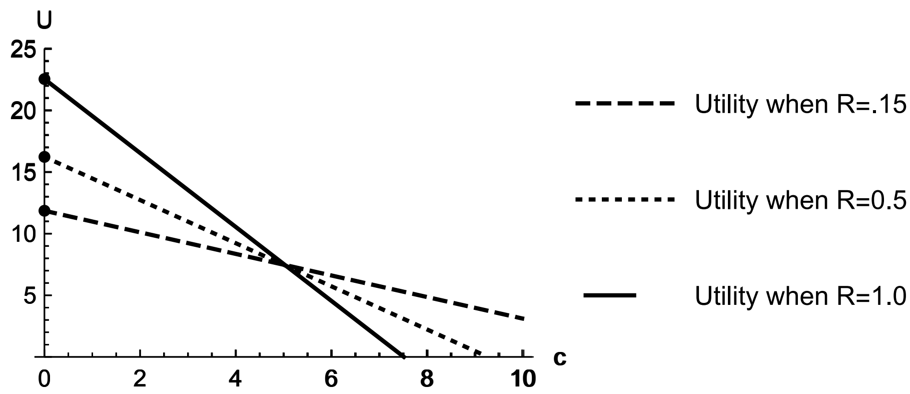

Figure 2.5: Direct Reciprocity - Case 1. Traditional Public Goods Game Nash Equilibrium.

In Figure 2.5, we represent player j's maximized utility as a function of their contribution, $U(c_j)$, in (c, U), where U represents player j's utility level and c represents player j's contribution choice. When $n = 3$, $y = 10$, $\alpha = \frac{1}{2}$, and player j has observed that the other players are contributing nothing (i.e., the history of the game is $h = (0, ..., 0)$ and $c_i = 0$), then player j's utility function is always maximized by contributing nothing given any value of $R_j \geq 0$. In other words, player j chooses $c_j = 0$ for any value of $R_j \geq 0$ when the history of the game is $h = (0, ..., 0)$. See Appendix A.1.1.1 for Mathematica code to reproduce the figure at these parameter values.

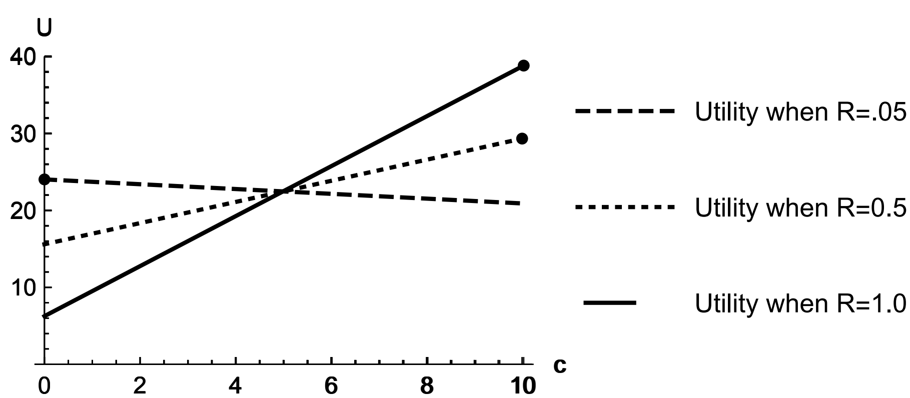

Figure 2.6: Direct Reciprocity - Case 2. Free-riding Equilibrium and Social Optimum.

Figure 2.6 represents player j's maximized utility as a function of their contribution choice, in (c, U), with the parameter values $n = 4$, $y = 10$, and $\alpha = \frac{1}{2}$. When player j observes that the other players are contributing the full endowment (i.e., $c_i = y$), then there are two potential equilibria, as identified in Figure 2.4. If player j has relatively weaker reciprocity preferences where $0 \leq R_j < f_1(\alpha)$, then free-riding is the optimal strategy for player j (i.e., $(y, ..., 0, ..., y)$ in *Area I* in Figure 2.4. If player j has relatively stronger reciprocity preferences where $R_j > f_1(\alpha)$, then the social optimum equilibrium results as in *Area II* in Figure 2.4. In Figure 2.6, $c_i = 10$ and $f_1(\alpha) = 0.133$, therefore, the function is downward sloping when $R_j < 0.133$ such that player j's utility is maximized when $c_j = 0$. When $R_j > 0.133$, the player j's utility upward sloping and is maximized when $c_j = y$. Figure 2.6 presents an intuitive finding: when player j does not have strong preferences for reciprocity, they will free-ride, but as player j's reciprocity preferences increase, they will fully contribute when the history of the game is $h = (y, ..., y)$.

See Appendix A.1.1.1 and A.1.1.2 for Mathematica code to reproduce the figures at these parameter values.

Comparative statics for $h = (c, ..., c)$ or interior solution equilibria are not presented here. Comparisons at the corner cases provide greater insight into thresholds for cooperative behaviour.

2.3 Indirect Upstream Reciprocity

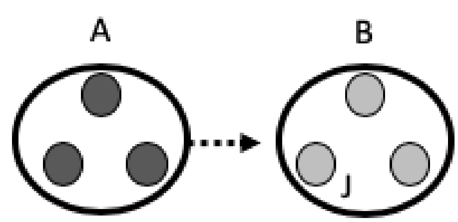

Figure 2.7: Indirect Upstream Reciprocity

Consider a variation of the baseline model, presented in Figure 2.7, where players interact simultaneously within their own group and there is an opportunity for one player, player J, to observe the outcome of a previous game with different players. In this multi-stage game, there are two groups of n players each: $A = \{1, ..., n\}$ and $B = \{n+1, ..., 2n\}$, with player $J \in B$. In the first period, each player $i \in A$ simultaneously chooses their contribution $c_i \in [0, y]$. This game has the same structure as the baseline case and the payoffs for each player $i \in A$ are outlined in Section 2.1. In the second period, the selected player, player $J \in B$, observes the average contribution of the n players in set A, $\frac{1}{n}\sum_{i\in A} c_i$, and receives a share of the group account, $\alpha \sum_{i\in A} c_i$. After player J has made this observation, all players in set B simultaneously choose their contribution amounts. Player J chooses their contribution, $c_J \in [0, y]$, and each player $j \in B_J$, where $B_J = \{n+1, ..., J-1, J+1, ..., 2n\}$ and $B_J \subset B$, chooses their contribution, c_j. The payoff for each player $j \in B_J$ is analogous to (2.1.1) in Section 2.1. The payoff for player $J \in B$ includes the additional share that they receive from the group account of players in set A, such that $\pi_J = y - c_J + \alpha \sum_{i\in A} c_i + \alpha \sum_{l\in B} c_l$, where l accounts for player J and each player $j \in B_J$. Indirect upstream reciprocity requires this game structure wherein player $J \in B$ plays simultaneously within their own group and cannot be "kind" to players in set A. Recall that indirect upstream reciprocity entails a "pay-it-forward" behaviour, therefore, the game structure requires that player J has no opportunity to directly reciprocate to players in set A. As before, we restrict the MPCR to $0 < \alpha < 1 < n\alpha$ to ensure that the game satisfies the conditions of a social dilemma in a public good game. Further, we again assume no discounting between periods.

2.3.1 No Reciprocity Preferences

If player $J \in B$ has no reciprocity preferences, then they will play a simultaneous game with $(n-1)$ symmetric players $j \in B_J$. In the second period, player J chooses c_J that solves

$$\max_{c_J}\{y - c_J + \alpha \sum_{i\in A} c_i + \alpha \sum_{l\in B} c_l\}. \tag{2.3.1}$$

The first-order condition is

$$-1 + \alpha < 0. \tag{2.3.2}$$

Therefore, we are again at a corner solution where the optimal contribution choice is $c_J^* = 0$ because the maximization problem is linear. At the same time, each player $j \in B_J$ chooses c_j to solve

$$\underset{c_j}{max}\{y - c_j + \alpha \sum_{l \in B} c_l\}. \tag{2.3.3}$$

The first-order condition is also

$$-1 + \alpha < 0. \tag{2.3.4}$$

We are again at a corner solution where the optimal contribution choice is $c_j^* = 0 \ \forall \ j \in B_J$, such that $c_{n+1}^* = ... = c_{J-1}^* = c_{J+1}^* = ... = c_{2n}^* = 0$. Once again, it is not surprising that the outcome of the game is the traditional public goods game Nash equilibrium where $(c_{n+1}^*, ..., c_{2n}^*) = (0, ..., 0)$. The first period of the game proceeds exactly as in the baseline case, Section 2.1. We have previously seen that the equilibrium in that game is also the traditional public goods game Nash equilibrium, $(c_1^*, ..., c_n^*) = (0, ..., 0)$. In the absence of reciprocity preferences, and even though player $J \in B$ may receive a share of the group account from players in set A, none of the players in set B have an incentive to contribute a positive amount to the group account.

2.3.2 Reciprocity Preferences

As in Section 2.1.2, incorporating reciprocity preferences into players' utility functions takes a similar form here. However, in the indirect upstream reciprocity case examined here, the reciprocity payoff function itself is expanded to account for both direct reciprocity preferences (as outlined in Section 2.1.2) and indirect reciprocity preferences. The reciprocity payoff function for player $J \in B$ will be described in detail below. The history of the game is analogous to the one outlined in Section 2, with the following exceptions. Since players in set B chose their contributions simultaneously, there is no history in that stage. However, players in set B still form beliefs about the contribution strategies of the others in their set. Player $J \in B$ observes the average contribution of players in set A and therefore knows something about the history of the game. The history of the game, $h = (c_1, ..., c_n)$, is observed by player $J \in B$ in the form of the average contribution of players in set A, $\frac{1}{n}\sum_{i \in A} c_i$, such that when player J observes $\frac{1}{n}\sum_{i \in A} c_i = 0$, then player J knows with probability 1 that each player $i \in A$ chose to contribute nothing to the group account, $c_i = 0$. When player J observes $\frac{1}{n}\sum_{i \in A} c_i = y$, then player J knows with probability 1 that each player $i \in A$ chose to fully contribute, $c_i = y$. However, when player J observes $\frac{1}{n}\sum_{i \in A} c_i = c$, where $c \in]0, y[$, then player J does not know with certainty what each player $i \in A$ contributed to the group account. Player J must then form beliefs, $b_{J,i}(h)$, about the other players' contribution strategies. Player J's utility function now consists of a material payoff function, a direct reciprocity payoff function, and an indirect reciprocity payoff function. Player J chooses c_J to solve

$$\begin{aligned} U_J = y - c_J + \alpha \sum_{i \in A} c_i + \alpha \sum_{l \in B} c_l \\ + \underbrace{\sum_{j \in B_J} R_{J,j} \cdot \kappa_{J,j}(c_J, (b_{J,j}(h))_{j \neq J}) \cdot \lambda_{J,j,k}(b_{J,j}(h), (b_{J,j,k}(h))_{k \neq j})}_{(i)} \\ + \underbrace{\sum_{i \in A} R_{J,i} \cdot \kappa_{J,j}(c_J, (b_{J,j}(h))_{j \neq J}) \cdot \lambda_{J,i,k}(b_{J,i}(h), (b_{J,i,k}(h))_{k \neq i})}_{(ii)}, \end{aligned} \tag{2.3.5}$$

where $i \in A$ and $j \in B_J$ where $B_J \subset B$. Unlike the reciprocity payoff function described in Section 2.1.2, player J's reciprocity payoff function is now comprised of two parts. First, player J's direct reciprocity payoff function, part (i) in (2.3.5), which is analogous to part (ii) in (2.1.6), takes the form

$$\sum_{j \in B_J} R_{J,j} \cdot \kappa_{J,j}(c_J, (b_{J,j}(h))_{j \neq J}) \cdot \lambda_{J,j,k}(b_{J,j}(h), (b_{J,j,k}(h))_{k \neq j}). \tag{2.3.6}$$

Following Section 2.1.2 and the direct reciprocity payoff, (2.1.10), player J's direct reciprocity payoff is

$$R_J \alpha^2 \left(c_J - \frac{1}{2} y \right) \left(\sum_{j \in B_J} b_{J,j}(h) - \frac{n-1}{2} y \right), \tag{2.3.7}$$

where R_J represents player J's direct reciprocity preference. Next, player J's indirect reciprocity payoff function, part (ii) in (2.3.5), takes the form

$$\sum_{i \in A} R_{J,i} \cdot \kappa_{J,j}(c_J, (b_{J,j}(h))_{j \neq J}) \cdot \lambda_{J,i,k}(b_{J,i}(h), (b_{J,i,k}(h))_{k \neq i}), \tag{2.3.8}$$

where $i \in A$ and $j \in B_J$. Since player J cannot be "kind" to players in set A, the indirect reciprocity payoff function captures the effect of player J's perceived kindness of players in set A on player J's own kindness towards players in set B_J. We examine player J's kindness function, $\kappa_{J,j}(\cdot)$, which evaluates player J's kindness towards player $j \in B_J$. From (2.1.3c), we know that

$$\kappa_{J,j}(c_J, (b_{J,j}(h))_{j \neq J}) = \alpha \left(c_J - \frac{1}{2} y \right), \tag{2.3.9}$$

where $j \in B_J$. In the indirect reciprocity payoff function, part (ii) in (2.3.5), player J's perceived kindness function, $\lambda_{J,i,k}(b_{J,i}(h), (b_{J,i,k}(h))_{k \neq i})$, evaluates player J's perception of the kindness of player $i \in A$ towards player J. Due to the structure of the game, each player $i \in A$ cannot directly be "kind" to player J. Therefore, player J must evaluate the perceived kindness of each player $i \in A$. Following (2.1.4d), we know that

$$\lambda_{J,i,k}(b_{J,i}(h),(b_{J,i,k}(h))_{k\neq i}) = \alpha\left(b_{J,i}(h) - \frac{1}{2}y\right), \tag{2.3.10}$$

where $i \in A$, $k \in A$, and $k \neq i$.

We denote R_I as player J's indirect reciprocity preference and we assume that $R_I = R_{J,i}\ \forall\ i \in A$ from part (ii) in (2.3.5). We express (2.3.5) as

$$\begin{aligned} \underset{c_J}{max}\Bigg\{ & y - c_J + \alpha\sum_{i\in A} c_i + \alpha\sum_{l\in B} c_l \\ & + R_J\alpha^2\left(c_J - \frac{1}{2}y\right)\left(\sum_{j\in B_J} b_{J,j}(h) - \frac{n-1}{2}y\right) \\ & + R_I\alpha^2\left(c_J - \frac{1}{2}y\right)\left(\sum_{i\in A} b_{J,i}(h) - \frac{n}{2}y\right)\Bigg\}. \end{aligned} \tag{2.3.11}$$

Note that the sum of beliefs about contributions relative to some reference point is different in the indirect reciprocity payoff compared to the direct reciprocity payoff in this equation and previous cases. Player J's utility maximization problem in (2.3.11) is different from the utility maximization problem in the baseline model, (2.1.13), and the utility maximization problem in the direct reciprocity model, (2.2.7). This is due to accounting for player J's indirect reciprocity preferences and the inclusion of an indirect reciprocity payoff function in addition to a direct reciprocity payoff function.

To simplify this expression, let $R_I = \gamma R_J$, where γ represents the relative strength of indirect reciprocity preferences to direct reciprocity preferences and $\gamma > 0$. When $\gamma > 1$, then player J's indirect reciprocity preferences are stronger than their direct reciprocity preferences. When $\gamma < 1$, then player J's indirect reciprocity preferences are weaker than their direct reciprocity preferences. We rewrite equation (2.3.11) as

$$\max_{c_J}\Bigg\{ y - c_J + \alpha \sum_{i\in A} c_i + \alpha \sum_{l\in B} c_l + R_J\alpha^2 \left(c_J - \frac{1}{2}y\right)\left(\sum_{j\in B_J} b_{J,j}(h) - \frac{n-1}{2}y\right) + R_J\alpha^2\gamma \left(c_J - \frac{1}{2}y\right)\left(\sum_{i\in A} b_{J,i}(h) - \frac{n}{2}y\right)\Bigg\}, \tag{2.3.11a}$$

or

$$\max_{c_J}\Bigg\{ y - c_J + \alpha \sum_{i\in A} c_i + \alpha \sum_{l\in B} c_l + R_J\alpha^2 \left(c_J - \frac{1}{2}y\right)\left(\sum_{j\in B_J} b_{J,j}(h) + \gamma \sum_{i\in A} b_{J,i}(h) - \frac{n-1+n\gamma}{2}y\right)\Bigg\}. \tag{2.3.11b}$$

The first-order condition is

$$-1 + \alpha + R_J\alpha^2 \left(\sum_{j\in B_J} b_{J,j}(h) + \gamma \sum_{i\in A} b_{J,i}(h) - \frac{n-1+n\gamma}{2}y\right) = 0. \tag{2.3.12}$$

We assume rational expectations in equilibrium; player J's beliefs about each player i's and player j's contributions are correct such that $b_{J,i}(h) = c_i$. and $b_{J,j}(h) = c_j$. We rewrite the first-order condition, (2.3.12), as

$$-1 + \alpha + R_J\alpha^2 \left(\sum_{j\in B_J} c_j + \gamma \sum_{i\in A} c_i - \frac{n-1+n\gamma}{2}y\right) = 0. \tag{2.3.12a}$$

The choice of c_J depends on player J's beliefs about the average contribution of the other players conditional on the history of the game, $b_{J,i}(h)$ and $b_{J,j}(h)$. Because the function is linear in the choice variable, c_J, the first-order condition can be positive, negative, or null. As described above, while player J

observes the average contribution of n players in set A through $\frac{1}{n}\sum_{i\in A} c_i$, there are only two cases in which player J knows with certainty what each player i chose to contribute to the group account; i.e., when $\frac{1}{n}\sum_{i\in A} c_i$ is 0 or y. When $\frac{1}{n}\sum_{i\in A} c_i = c \in \,]0,y[$, then player J does not know with certainty what each player i contributed to the group account and must form beliefs about the contributions of each player $i \in A$. As before, we examine different cases when the history of the game is $h = (0, ..., 0)$, $h = (y, ..., y)$, and $h = (c, ..., c)$.

The best response function for player J is thus

$$BR_J\left((b_{J,j}(h)),(b_{J,i}(h))\right)_{i\neq j\neq J} = \begin{cases} 0 & \text{if} \quad \sum_{j\in B_J} b_{J,j}(h) < \quad \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1+n\gamma}{2}y \\ & \qquad\qquad\qquad -\gamma\sum_{i\in A} b_{J,i}(h) \\ [0,y] & \text{if} \quad \sum_{j\in B_J} b_{J,j}(h) = \quad \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1+n\gamma}{2}y \\ & \qquad\qquad\qquad -\gamma\sum_{i\in A} b_{J,i}(h) \\ y & \text{if} \quad \sum_{j\in B_J} b_{J,j}(h) > \quad \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1+n\gamma}{2}y \\ & \qquad\qquad\qquad -\gamma\sum_{i\in A} b_{J,i}(h) \end{cases} \tag{2.3.13}$$

Next, we determine the utility maximization problem for each player $j \in B_J$. Similar to Section 2.1.2, each player j chooses c_j to solve

$$\max_{c_j}\Bigg\{ y - c_j + \alpha\sum_{l\in B} c_l + R_j\alpha^2\left(c_j - \frac{1}{2}y\right)\left(\sum_{k\in B_j} b_{j,k}(h) - \frac{n-1}{2}y\right)\Bigg\}, \tag{2.3.14}$$

where $B_j = \{n+1, ..., j-1, j+1, ..., 2n\}$ and $B_j \subset B$. Player j does not know the history of the game because player j is making their contribution choice at the same time as the other players in set B.

The first-order condition is

$$-1 + \alpha + R_j \alpha^2 \left(\sum_{k \in B_j} b_{j,k}(h) - \frac{n-1}{2} y \right) = 0. \tag{2.3.15}$$

We assume rational expectations in equilibrium; player j's beliefs about each player k's contributions are correct such that $b_{j,k}(h) = c_k$. We rewrite the first-order condition, (2.3.15) as

$$-1 + \alpha + R_j \alpha^2 \left(\sum_{k \in B_j} c_k - \frac{n-1}{2} y \right) = 0. \tag{2.3.15a}$$

Given the contributions of the $(n-1)$ players in set B_j, the best response function for player j is

$$BR_j\left((b_{j,k}(h))_{k \neq j}\right) = \begin{cases} 0 & \text{if } \sum_{k \in B_j} b_{j,k}(h) < \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2} y \\ [0, y] & \text{if } \sum_{k \in B_j} b_{j,k}(h) = \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2} y \\ y & \text{if } \sum_{k \in B_j} b_{j,k}(h) > \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2} y \end{cases} \tag{2.3.16}$$

We need to consider cases depending on the history of the game. Recall from 2.3.2 that the history of the game can be $h = (0, ..., 0)$, $h = (y, ..., y)$, or $h = (c, ..., c)$.

2.3.2.1 Case 1

First, if the history of the game is $h = (0, ..., 0)$, then player J has observed $\frac{1}{n} \sum_{i \in A} c_i = 0$. This means that with probability 1, $c_i = 0 \ \forall \ i \in A$, thus player

J knows with certainty that $c_1^* = 0, ..., c_n^* = 0$ (i.e., players in set A are at the traditional public goods game Nash equilibrium). Since player J is choosing their contribution amount at the same time as each player $j \in B_J$, player J must form beliefs about the contribution strategy of each player $j \in B_J$. Player J's best response function is

$$BR_J\left((b_{J,j}(0,...,0))_{j\neq J}\right) = \begin{cases} 0 & \text{if} \quad \sum_{j\in B_J} b_{J,j}(h) < \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1+n\gamma}{2}y \\ [0,y] & \text{if} \quad \sum_{j\in B_J} b_{J,j}(h) = \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1+n\gamma}{2}y \\ y & \text{if} \quad \sum_{j\in B_J} b_{J,j}(h) > \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1+n\gamma}{2}y \end{cases} \tag{2.3.17}$$

First, to determine if there exists interior solutions for players in set B, we examine the players' best response functions. From each player j's best response function, (2.3.16), and player J's best response function, (2.3.17),

$$(n-2)c_j + c_J = \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y, \tag{2.3.18}$$

and

$$(n-1)c_j = \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1+n\gamma}{2}y. \tag{2.3.19}$$

Solving for an interior solution, we find

$$c_J^* = \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_j} - \frac{n-2}{n-1}\frac{1}{R_J}\right) + \frac{n-1-n\gamma(n-2)}{2(n-1)}y, \tag{2.3.20}$$

where $R_J \geq 0$ and $R_j \geq 0$. To be an eligible interior solution, $0 \leq c_J^* \leq y$. We verify that $c_J^* \leq y$ if

$$\frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_j} - \frac{n-2}{n-1}\frac{1}{R_J}\right) + \frac{n-1-n\gamma(n-2)}{2(n-1)}y \leq y. \tag{2.3.21}$$

Rearranging the terms, we obtain

$$\frac{n-2}{n-1}\frac{1}{R_J} \geq \frac{1}{R_j} - \frac{\alpha^2}{1-\alpha}\frac{n-1+n\gamma(n-2)}{2(n-1)}y, \tag{2.3.21a}$$

or

$$\frac{n-2}{n-1}\frac{1}{R_J} \geq \frac{\overbrace{1 - \frac{\alpha^2}{1-\alpha}\frac{n-1+n\gamma(n-2)}{2(n-1)}R_j y}^{(i)}}{R_j}. \tag{2.3.21b}$$

If part $(i) > 0$, then (2.3.21) is satisfied if

$$R_j < \frac{1-\alpha}{\alpha^2}\frac{2(n-1)}{(n-1+n\gamma(n-2))y}, \tag{2.3.22}$$

and

$$R_J \leq \frac{n-2}{n-1}\frac{R_j}{1 - \frac{\alpha^2}{1-\alpha}\frac{n-1+n\gamma(n-2)}{2(n-1)}R_j y}. \tag{2.3.23}$$

If part $(i) \leq 0$, then (2.3.21) is satisfied if

$$R_j \geq \frac{1-\alpha}{\alpha^2}\frac{2(n-1)}{(n-1+n\gamma(n-2))y}, \tag{2.3.24}$$

and

$$R_J \geq \frac{n-2}{n-1}\frac{R_j}{1 - \frac{\alpha^2}{1-\alpha}\frac{n-1+n\gamma(n-2)}{2(n-1)}R_j}, \tag{2.3.25}$$

which is always satisfied because $R_J \geq 0$.

We check that $c_J^* \geq 0$ if

$$\frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_j} - \frac{n-2}{n-1}\frac{1}{R_J}\right) + \frac{n-1-n\gamma(n-2)}{2(n-1)}y \geq 0. \tag{2.3.26}$$

Rearranging the terms, we obtain

$$\frac{n-2}{n-1}\frac{1}{R_J} \leq \frac{1+\frac{\alpha^2}{1-\alpha}\overbrace{\left(\frac{n-1-n\gamma(n-2)}{2(n-1)}\right)}^{(i)}R_jy}{R_j}. \tag{2.3.26a}$$

If part $(i) > 0$, then $\gamma < \frac{n-1}{n(n-2)}$, and (2.3.26) is satisfied if

$$R_J \geq \frac{n-2}{n-1}\frac{R_j}{1+\frac{\alpha^2}{1-\alpha}\left(\frac{n-1-n\gamma(n-2)}{2(n-1)}\right)R_jy}. \tag{2.3.27}$$

If part $(i) < 0$, then $\gamma > \frac{n-1}{n(n-2)}$ and (2.3.26) is satisfied when

$$1+\frac{\alpha^2}{1-\alpha}\left(\frac{n-1-n\gamma(n-2)}{2(n-1)}\right)R_jy > 0, \tag{2.3.28}$$

or

$$0 \leq R_j < \frac{1-\alpha}{\alpha^2}\frac{2(n-1)}{(n\gamma(n-2)-n+1)y}, \tag{2.3.29}$$

then

$$R_J \geq \frac{n-2}{n-1}\frac{R_j}{1+\frac{\alpha^2}{1-\alpha}\left(\frac{n-1-n\gamma(n-2)}{2(n-1)}\right)R_jy}, \tag{2.3.30}$$

or when

$$1+\frac{\alpha^2}{1-\alpha}\left(\frac{n-1-n\gamma(n-2)}{2(n-1)}\right)R_jy < 0, \tag{2.3.31}$$

or

$$R_j > \frac{1-\alpha}{\alpha^2}\left(\frac{2(n-1)}{(n\gamma(n-2)-n+1)y}\right), \tag{2.3.32}$$

and $R_j \geq 0$, then

$$R_J \leq \frac{n-2}{n-1} \frac{R_j}{1 + \frac{\alpha^2}{1-\alpha}\left(\frac{n-1-n\gamma(n-2)}{2(n-1)}\right) R_j y}, \tag{2.3.33}$$

which is not possible as R_J cannot be less than zero. Depending on R_j, the conditions for $0 \leq c_J^* \leq y$ can be simultaneously satisfied, therefore c_J^* exists and is an eligible interior solution.

Solving for an interior solution for each player j, we find

$$c_j^* = \frac{1-\alpha}{\alpha^2} \frac{1}{(n-1)R_J} + \frac{n-1+n\gamma}{2(n-1)} y, \tag{2.3.34}$$

where $R_J \geq 0$. It is clear from (2.3.34) that $c_j^* \geq 0$, however to be an eligible interior solution, we also require $c_j^* \leq y$, which is satisfied if

$$\frac{1-\alpha}{\alpha^2} \frac{1}{(n-1)R_J} + \frac{n-1+n\gamma}{2(n-1)} y \leq y. \tag{2.3.35}$$

Rearranging the terms, we get

$$\frac{1}{(n-1)R_J} \leq -\frac{\alpha^2}{1-\alpha} y \overbrace{\frac{n-1+n\gamma}{2(n-1)}}^{(i)}. \tag{2.3.35a}$$

Since part $(i) > 0$, the condition (2.3.35) is satisfied if

$$R_J \leq -\frac{1-\alpha}{\alpha^2} \frac{2}{(n-1+n\gamma)y}, \tag{2.3.36}$$

which is not possible as R_J cannot be less than zero. Therefore, c_j^* exists and $c_j^* \geq 0$, but because c_j^* is not less than the full endowment, y, for any $\gamma > 0$, c_j^* is not an eligible interior solution.

When the history of the game is $h = (0, ..., 0)$, there are four possible equilibria, expressed as $(c_{n+1}, ..., c_J, ..., c_{2n})$. The traditional public goods game Nash equilibrium, $(0, ..., 0, ..., 0)$, is always satisfied because the condition $0 < \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1+n\gamma}{2} y$ is satisfied for player J for any value of $R_J \geq 0$, and the condition $0 < \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y$ is satisfied for each player j for any value of $R_j \geq 0$. This equilibrium is predicted by standard game theory for public goods games with no reciprocity preferences and remains an equilibrium even when accounting for reciprocity preferences.

The social optimum equilibrium, $(y, ..., y, ..., y)$, is satisfied when the following conditions are satisfied for player J and each player j. For player J, $\frac{1}{R_J} < \frac{n-1-n\gamma}{2} y \frac{\alpha^2}{1-\alpha}$ where if $n-1-n\gamma > 0$, or $\gamma < \frac{n-1}{n}$, then the equilibrium exists when

$$R_J > \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1-n\gamma)y}, \tag{2.3.37}$$

or if $n-1-n\gamma \leq 0$, or $\gamma \geq \frac{n-1}{n}$, then the equilibrium exists when

$$R_J \leq \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1-n\gamma)y}, \tag{2.3.38}$$

which is not satisfied since R_J cannot be less than zero. For player j

$$R_j > \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1)y}, \tag{2.3.39}$$

which is always satisfied if R_j is strictly greater than zero. By accounting for reciprocity preferences, the social optimum equilibrium can be reached when player J's direct reciprocity preferences are sufficiently strong and greater than their indirect reciprocity preferences.

The equilibrium with some free-riding, $(y, ..., c_J^*, ..., y)$, is satisfied when the following conditions are met for player J and each player j. For player J

$$R_J = \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1-n\gamma)y}, \tag{2.3.40}$$

where if $n-1-n\gamma > 0$, or $\gamma < \frac{n-1}{n}$, then $R_J > 0$ and the condition is satisfied, and if $n-1-n\gamma \leq 0$, or $\gamma \geq \frac{n-1}{n}$, then the condition is not satisfied as R_J cannot be less than zero. For player j

$$\frac{n-2}{n-1}\frac{1}{R_J} < \frac{\alpha^2}{1-\alpha}y\frac{n(n-1-\gamma(n-2))}{2(n-1)}, \tag{2.3.41}$$

where if $n-1-\gamma(n-2) > 0$, or $0 < \gamma < \frac{n-1}{n-2}$, then

$$R_J > \frac{1-\alpha}{\alpha^2}\frac{2(n-2)}{ny(n-1-\gamma(n-2))}, \tag{2.3.42}$$

and the condition is met, and if $n-1-\gamma(n-2) < 0$, or $\gamma > \frac{n-1}{n-2}$, then

$$R_J < \frac{1-\alpha}{\alpha^2}\frac{2(n-2)}{ny(n-1-\gamma(n-2))}, \tag{2.3.43}$$

and the condition is not met because R_J cannot be less than zero. The equilibrium with some free-riding is possible if $0 < \gamma < \frac{n-1}{n}$ where player J's indirect reciprocity preferences are weaker than their direct reciprocity preferences.

The equilibrium with total free-riding, $(y, ..., 0, ..., y)$, is satisfied when conditions on player J and each player j are met. For player J

$$\frac{1}{R_J} > \frac{\alpha^2}{1-\alpha}y\frac{n-1-n\gamma}{2}, \tag{2.3.44}$$

where if $n-1-n\gamma > 0$, or $\gamma < \frac{n-1}{n}$, then

$$0 \leq R_J < \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1-n\gamma)y}, \tag{2.3.45}$$

and if $n-1-n\gamma < 0$, or $\gamma > \frac{n-1}{n}$, then

$$R_J > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1-n\gamma)y}, \tag{2.3.46}$$

which is always satisfied because $R_J \geq 0$, and the condition is met. If $n-1-n\gamma = 0$, or $\gamma = \frac{n-1}{n}$, then $0 < \frac{1-\alpha}{\alpha^2}\frac{1}{R_j}$ is always satisfied because $R_J \geq 0$. For player j

$$R_j > \frac{1-\alpha}{\alpha^2} \frac{2}{(n-3)y}, \tag{2.3.47}$$

which is satisfied when $n > 3$. The existence of the equilibrium with total free-riding suggests that free-riding is an optimal strategy when player J has observed and been impacted by free-riding among players in set A. This further supports our hypothesis that indirect reciprocity preferences are important in determining player J's contribution choice. From the above conditions, it is clear that the barriers to free-riding are much lower when player J's indirect reciprocity preferences become greater than a certain point, $\gamma \geq \frac{n-1}{n}$.

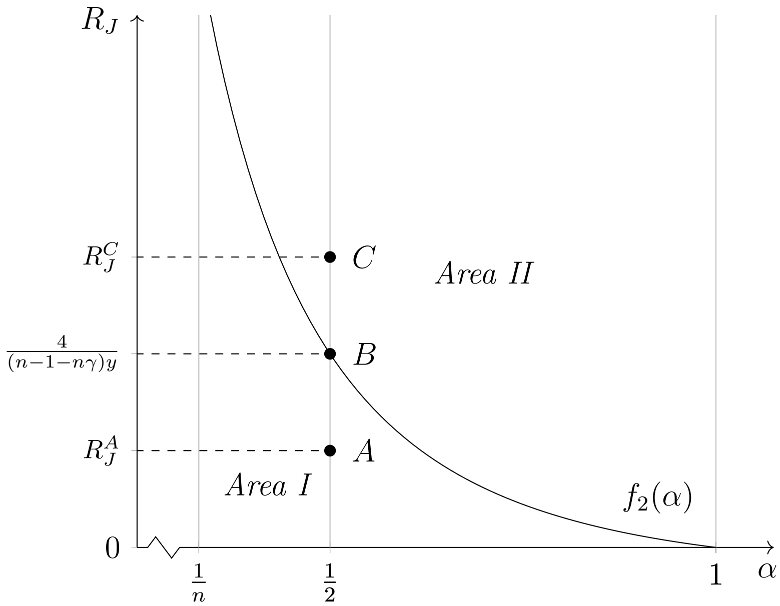

Figure 2.8: Indirect Upstream Reciprocity - Case 1: $h = (0, ..., 0)$

We represent these different equilibria in a graph (α, R_J) in Figure 2.8. In Figure 2.8, $f_2(\alpha) \equiv \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1-n\gamma)y}$ when $\gamma < \frac{n-1}{n}$. Recall that $\gamma > 0$ and $R_I = \gamma R_J$. Note that when $\gamma > \frac{n-1}{n}$, the function exists for negative values of R_J. However, since we impose the restriction $R_J \geq 0$, that portion of the function does not appear in the figure above (i.e., it exists in the quadrant below the α-axis). In *Area I*, the equilibria are the traditional public goods game Nash equilibrium, $(0, ..., 0, ..., 0)$, and the total free-riding equilibrium, $(y, ..., 0, ..., y)$, when $\gamma \lesseqgtr \frac{n-1}{n}$. In *Area II*, the equilibria are the traditional public goods game Nash equilibrium, the total free-riding equilibrium when $\gamma > \frac{n-1}{n}$, and the social optimum equilibrium, $(y, ..., y, ..., y)$. Along the curve $f_2(\alpha)$, the equilibrium with some free-riding, $(y, ..., c_J^*, ..., y)$, exists as does the traditional public goods game Nash equilibrium (which is satisfied everywhere). In Case 1, when $\gamma < \frac{n-1}{n}$ and is increasing, or as player J's indirect reciprocity preferences become stronger, then the threshold increases (i.e., $f_2(\alpha)$ shifts out) and it becomes harder to reach cooperative equilibria. In other words, when player J cares more about indirect reciprocity, then we expect free-riding behaviour to dominate when the history of the game is $h = (0, ..., 0)$.

To illustrate, let $\alpha = \frac{1}{2}$ such that $f_2(\frac{1}{2}) = \frac{4}{(n-1-n\gamma)y}$. At point A, player J's reciprocity preference is weaker, $R_J^A < f_2(\alpha)$, and player J will contribute nothing to the group account, irrespective of indirect reciprocity preferences, γ. At point B, player J's reciprocity preference is stronger, $R_J^B = f_2(\alpha)$, and cooperating with their group by contributing a positive amount to the public good becomes a possibility for player J (i.e., $c_J > 0$). This equilibrium suggests that the history of the game, $h = (0, ..., 0)$, tempers player J's cooperation such that player J does not contribute their full endowment, despite others in their group doing so. At point C, player J's reciprocity preference is strong, $R_J^C > f_2(\alpha)$, and player J will contribute nothing if player J believes that others in their group are also not contributing to the group account (i.e., the traditional public goods game Nash equilibrium), or if player J has relatively strong indirect reciprocity preferences, $\gamma > \frac{n-1}{n}$, then player J will free-ride when others are fully contributing. At point C, player J can reach a cooperative equilibrium as well, and the group can be at the social optimum equilibrium.

2.3.2.1.1 Comparative Statics We examine the case of free-riding behaviour in *Area II* in Figure 2.8, where $R_J > f_2(\alpha)$, and compare how equilibria behaviour may be affected as certain parameters change. We use Mathematica (Wolfram Research, Inc, 2020), a technical computing system, to visualize the data. Mathematica code for reproduction of the interactive visualization is available in Appendix A.1.2. Static snapshots and descriptions with parameter values are presented below.

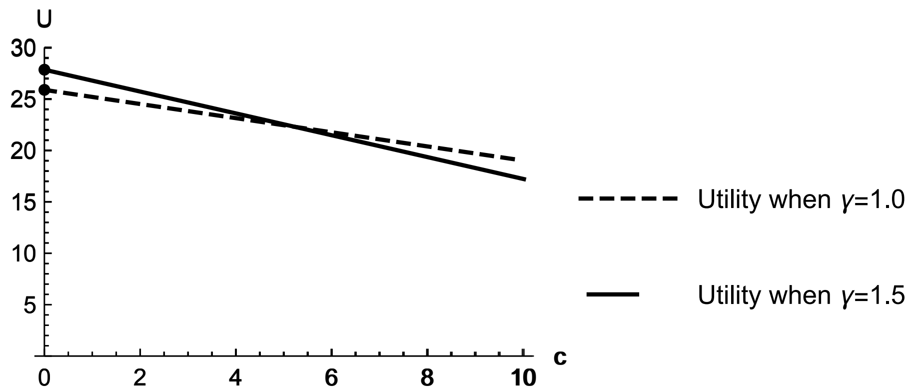

Figure 2.9: Indirect Upstream Reciprocity - Case 1. Player J's utility when $R_J = 0.15$.

In Figure 2.9, we represent player J's maximized utility as a function of their contribution choice, $U(c)$, where U represents player J's utility level and c represents player J's contribution choice. When the history of the game is $h = (0, ..., 0)$, $n = 4$, $y = 10$, $\alpha = \frac{1}{2}$, player J believes that the other players are contributing fully (i.e., $b_{J,j} = y$), and player J has relatively weak reciprocity preferences ($R_J = 0.15$), we examine three values of player J's indirect reciprocity preference parameter, γ, when $\gamma \lesseqqgtr 1$. Recall that γ represents the relative strength of player J's direct and indirect reciprocity preferences. Furthermore, recall that the total free-riding equilibrium is optimal if $R_J > f_2(\alpha)$ and $\gamma > \frac{n-1}{n}$, and the social optimum equilibrium occurs if $R_J > f_2(\alpha)$ and $\gamma < \frac{n-1}{n}$. When $\gamma < 1$, player J's indirect reciprocity preferences are weaker than their direct reciprocity preferences. We predict that when player J has weaker indirect reciprocity preferences (i.e., $\gamma < 1$) the social optimum equilibrium will result. We examine the following scenarios: $\gamma = 0.5$, $\gamma = 1$, and $\gamma = 1.5$. When $\gamma = 0.5$, then R_J is not greater than $f_2(\alpha)$ and the social optimum equilibrium is not possible. The social optimum equilibrium can only be reached for values of $\gamma < 0.083$ when $R_J = 0.15$. This outcome would require that player J's indirect reciprocity preferences are very weak, in addition to weak direct reciprocity preferences. When $\gamma = 1$, all else constant, then $R_J > f_2(\alpha)$ is satisfied. Player J's indirect reciprocity preferences are as strong as their direct reciprocity preferences. Player J's utility function is downward sloping and is maximized when $c_J = 0$. The results hold when $\gamma = 1.5$. Free-riding is an optimal behaviour for player J when their direct reciprocity preferences are weak.

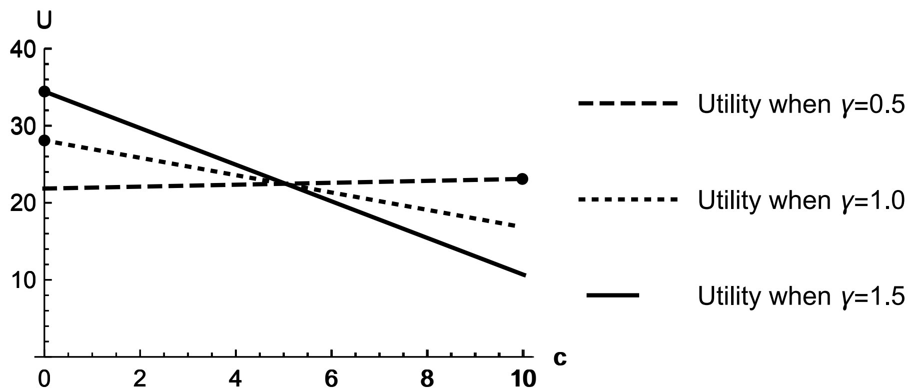

Figure 2.10: Indirect Upstream Reciprocity - Case 1. Player J's utility when $R_J = 0.5$.

We examine player J's utility for the same values for γ as R_J increases. In Figure 2.10, direct reciprocity preferences become stronger, $R_J = 0.5$, and as γ increases, we observe a shift from cooperative behaviour to free-riding. Player J will fully contribute when $\gamma = 0.5$ and their utility function is upward sloping and maximized when $c_J^* = y$. When γ increases and indirect reciprocity preferences become relatively stronger, there is a switch towards free-riding behaviour. Player J will free-ride when $\gamma = 1$ and their utility function is downward sloping and maximized when $c_J^* = 0$. When player J's indirect reciprocity preferences are stronger than their direct reciprocity preferences, $\gamma = 1.5$, then the total free-riding equilibrium, $(y, ..., 0, ..., y)$, is optimal.

In Figure 2.11, direct reciprocity preferences are stronger, $R_J = 1$, and we observe a similar behavioural shift as γ increases. When indirect reciprocity preferences are weaker, $\gamma = 0.5$, then the social optimum equilibrium, $(y, ..., y, ..., y)$, results. However, as player J's indirect reciprocity preferences become stronger ($\gamma \geq 1$), all else constant, then free-riding becomes the optimal strategy.

This comparison highlights the importance of indirect reciprocity prefer-

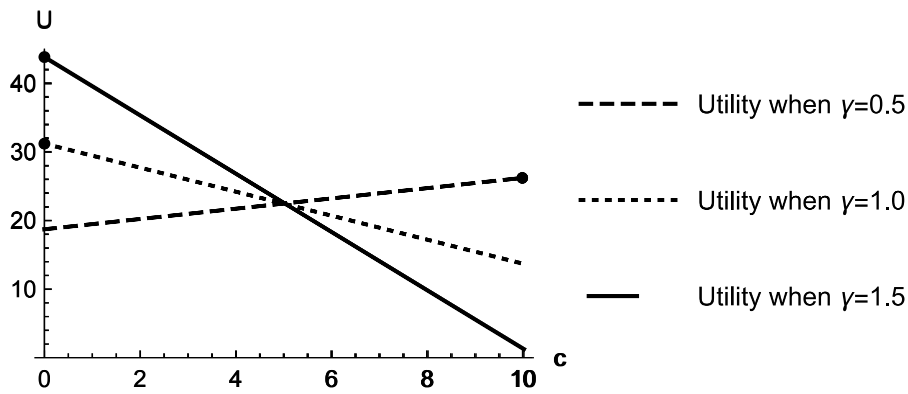

Figure 2.11: Indirect Upstream Reciprocity - Case 1. Player J's utility when $R_J = 1$.

ences and the role of the history of the game $h = (0, ..., 0)$ in tempering player J's cooperation. When player J has had a recent experience (i.e., players in set A were "unkind" to player J), and player J has strong indirect reciprocity preferences, then player J will free-ride. If player J has perceived each player $i \in A$ to be "unkind", then player J "pays forward" being "unkind" and behaves non-cooperatively with players in set B. However, if player J has weak indirect reciprocity preferences, then they will fully cooperate. Player J's recent experience does not drive their contribute decision as much as their direct reciprocity preference and their beliefs about the other players' contribution strategies.

See Appendix A.1.2.1 for Mathematica code to reproduce the figures at the specified parameter values.

2.3.2.2 Case 2

Second, when the history of the game is $h = (y, ..., y)$, then player J has observed $\frac{1}{n}\sum_{i \in A} c_i = y$. This means that with probability 1, each player $i \in A$ is fully contributing to the group account, $c_i = y \ \forall \ i \in A$, and thus player J knows with certainty that $c_1^* = y, ..., c_n^* = y$ (i.e., players in set A are at the social optimum equilibrium, $(y, ..., y, ..., y)$). However, because player J is choosing their contribution amount at the same time as each player $j \in B_J$, player J must form beliefs about the contribution strategy of each player

$j \in B_J$. Player J's best response function is

$$BR_J\left((b_{J,j}(y,...,y))_{j\neq J}\right) = \begin{cases} 0 & \text{if} \quad \sum_{j\in B_J} b_{J,j} < \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1-n\gamma}{2}y \\ [0,y] & \text{if} \quad \sum_{j\in B_J} b_{J,j} = \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1-n\gamma}{2}y \\ y & \text{if} \quad \sum_{j\in B_J} b_{J,j} > \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1-n\gamma}{2}y \end{cases} \tag{2.3.48}$$

From the best response functions for players J and j, (2.3.48) and (2.3.16) respectively,

$$(n-1)c_j = \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1-n\gamma}{2}y, \tag{2.3.49}$$

and

$$(n-2)c_j + c_J = \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y. \tag{2.3.50}$$

Solving for an interior solution, we find

$$c_J^* = \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_j} - \frac{n-2}{n-1}\frac{1}{R_J}\right) + \frac{n-1+n\gamma(n-2)}{2(n-1)}y, \tag{2.3.51}$$

where $R_j \geq 0$ and $R_J \geq 0$. We verify that $c_J^* \leq y$ if

$$\frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_j} - \frac{n-2}{n-1}\frac{1}{R_J}\right) + \frac{n-1+n\gamma(n-2)}{2(n-1)}y \leq y. \tag{2.3.52}$$

Rearranging the terms, we get

$$\frac{n-2}{n-1}\frac{1}{R_J} \geq \frac{1}{R_j} - \frac{\alpha^2}{1-\alpha}y\left(\frac{n-1-n\gamma(n-2)}{2(n-1)}\right), \tag{2.3.52a}$$

or

$$\frac{n-2}{n-1}\frac{1}{R_J} \geq \frac{\overbrace{1 + \frac{\alpha^2}{1+\alpha}\left(\frac{n\gamma(n-2)-n+1}{2(n-1)}\right)yR_j}^{(i)}}{R_j}. \tag{2.3.52b}$$

Note that part $(i) > 0$ if $\gamma \geq \frac{n-1}{n(n-2)}$ and

$$R_j \geq -\frac{1-\alpha}{\alpha^2}\frac{2(n-1)}{(n\gamma(n-2)-n+1)y}, \tag{2.3.53}$$

which is always satisfied as $R_j \geq 0$. As well, part $(i) > 0$ if $\gamma < \frac{n-1}{n(n-2)}$ and

$$R_j < -\frac{1-\alpha}{\alpha^2}\frac{2(n-1)}{(n\gamma(n-2)-n+1)y}, \tag{2.3.54}$$

then

$$0 \leq R_J \leq \frac{n-2}{n-1}\left(\frac{R_j}{1+\frac{\alpha^2}{1-\alpha}\left(\frac{n\gamma(n-2)-n+1}{2(n-1)}\right)yR_j}\right). \tag{2.3.55}$$

However, part $(i) < 0$ if $\gamma < \frac{n-1}{n(n-2)}$ and

$$R_j > -\frac{1-\alpha}{\alpha^2}\frac{2(n-1)}{(n\gamma(n-2)-n+1)y}, \tag{2.3.56}$$

then

$$R_J \geq \frac{n-2}{n-1}\left(\frac{R_j}{1+\frac{\alpha^2}{1-\alpha}\left(\frac{n\gamma(n-2)-n+1}{2(n-1)}\right)yR_j}\right), \tag{2.3.57}$$

which is always met as $R_J \geq 0$. We check that $c_J^* \geq 0$ if

$$\frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_j}-\frac{n-2}{n-1}\frac{1}{R_J}\right)+\frac{n-1+n\gamma(n-2)}{2(n-1)}y \geq 0. \tag{2.3.58}$$

Rearranging the terms, we get

$$\frac{n-2}{n-1}\frac{1}{R_J} \leq \frac{1+\frac{\alpha^2}{1-\alpha}\frac{n-1+n\gamma(n-2)}{2(n-1)}yR_j}{R_j}, \tag{2.3.58a}$$

which is satisfied if

$$R_J \geq \frac{n-2}{n-1} \frac{R_j}{1 + \frac{\alpha^2}{1-\alpha} \frac{n-1+n\gamma(n-2)}{2(n-1)} y R_j}, \tag{2.3.58b}$$

for any value of $\gamma > 0$. The conditions for $0 \leq c_J^* \leq y$ can be simultaneously satisfied.

Solving for an interior solution for each player j, we find

$$c_j^* = \frac{1-\alpha}{\alpha^2} \frac{1}{(n-1)R_J} + \frac{n-1-n\gamma}{2(n-1)} y, \tag{2.3.59}$$

where $R_J \geq 0$.

We verify that $c_j^* \leq y$ if

$$\frac{1-\alpha}{\alpha^2} \frac{1}{(n-1)R_J} + \frac{n-1-n\gamma}{2(n-1)} y \leq y. \tag{2.3.60}$$

Rearranging the terms, we get

$$\frac{1}{(n-1)R_J} \leq \frac{\alpha^2}{1-\alpha} y \frac{n-1+n\gamma}{2(n-1)}, \tag{2.3.60a}$$

which is satisfied when

$$R_J \geq \frac{1-\alpha}{\alpha^2} \frac{2}{(n\gamma + n - 1)y}. \tag{2.3.61}$$

We check that $c_j^* \geq 0$ if

$$\frac{1-\alpha}{\alpha^2} \frac{1}{(n-1)R_J} + \frac{n-1-n\gamma}{2(n-1)} y \geq 0. \tag{2.3.62}$$

Rearranging the terms, we get

$$\frac{1}{R_J} \geq \frac{\alpha^2}{1-\alpha} y \frac{\overbrace{n\gamma - n + 1}^{(i)}}{2}. \tag{2.3.62a}$$

If part $(i) > 0$, then $\gamma > \frac{n-1}{n}$ and

$$R_J \leq \frac{1-\alpha}{\alpha^2} \frac{2}{(n\gamma - n + 1)y}, \tag{2.3.63}$$

and if part $(i) \leq 0$, then $\gamma \leq \frac{n-1}{n}$ and

$$R_J \geq \frac{1-\alpha}{\alpha^2} \frac{2}{(n\gamma - n + 1)y}, \tag{2.3.64}$$

which is always met as $R_J \geq 0$. The conditions for $0 \leq c_j^* \leq y$ can be simultaneously satisfied.

When the history of the game is $h = (y, ..., y)$, there are five possible equilibria. The traditional public goods game Nash equilibrium, $(0, ..., 0, ..., 0)$, is satisfied if the following conditions for players J and j are met. For player J

$$\frac{1}{R_J} > \frac{\alpha^2}{1-\alpha} \frac{n\gamma - n + 1}{2} y, \tag{2.3.65}$$

where if $n\gamma - n + 1 > 0$, or $\gamma > \frac{n-1}{n}$, then

$$0 \leq R_J < \frac{1-\alpha}{\alpha^2} \frac{2}{(n\gamma - n + 1)y}, \tag{2.3.66}$$

or if $n\gamma - n + 1 \leq 0$, or $\gamma \leq \frac{n-1}{n}$, then

$$R_J > \frac{1-\alpha}{\alpha^2} \frac{2}{(n\gamma - n + 1)y}, \tag{2.3.67}$$

which is always met as $R_J \geq 0$.

For player j

$$0 < \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2} y, \tag{2.3.68}$$

which is satisfied for any value of $R_j \geq 0$. Unlike the outcome of the previous case (Section 2.3.2.1), the inequality (2.3.66) implies that player J needs to have sufficiently weak reciprocity preferences to play the traditional public goods game Nash equilibrium when player J's indirect reciprocity preferences are relatively stronger (i.e., when $\gamma > \frac{n-1}{n}$), given the history of the game $h = (y, ..., y)$. When player J cares less about indirect reciprocity, $\gamma < \frac{n-1}{n}$, then contributing nothing is an optimal strategy for player J at any value of $R_J \geq 0$. However, if player J cares relatively more about indirect reciprocity preferences, (i.e., as $\gamma > \frac{n-1}{n}$), then there is a smaller range of values for R_J that make contributing nothing an optimal strategy. The more that indirect reciprocity preferences matter to player J, and given the history of the game $h = (y, ..., y)$, then the more player J cares about reciprocating the "kindness" they received from players in set A.

The social optimum equilibrium, $(y, ..., y, ..., y)$, is satisfied when for player J

$$R_J > \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1+n\gamma)y}, \tag{2.3.69}$$

and for player j

$$R_j > \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1)y}. \tag{2.3.70}$$

These conditions imply that if player J and each player j have sufficiently strong reciprocity preferences, the social optimum equilibrium can be reached irrespective of player J's indirect reciprocity preferences relative to their direct reciprocity preferences (i.e., irrespective of γ).

The altruistic[5] equilibrium, $(0, ..., y, ..., 0)$, is satisfied when the following

[5]This equilibrium is not pure altruism as defined by Trivers (1971). Pure altruism occurs when the giver incurs a (small) cost and the receiver benefits. Player J incurs a cost from fully contributing to the group account when others are not in the form of lower material payoff, but player J benefits through their reciprocity payoff function.

conditions for player J and each player j are met. For player J

$$\frac{1}{R_J} < \frac{\alpha^2}{1-\alpha}\frac{n\gamma - n + 1}{2}y, \tag{2.3.71}$$

where for $n\gamma - n + 1 \leq 0$, or $\gamma \leq \frac{n-1}{n}$, then the condition $R_J \geq 0$ is not satisfied, and for $n\gamma - n + 1 > 0$, or $\gamma > \frac{n-1}{n}$, then

$$R_J > \frac{1-\alpha}{\alpha^2}\frac{2}{(n\gamma - n + 1)y}, \tag{2.3.72}$$

and $R_J > 0$. For player j

$$\frac{1}{R_j} > -\left[\frac{\alpha^2}{1-\alpha}\left(\frac{n-3}{2}\right)y\right], \tag{2.3.73}$$

where $R_j \geq 0$ is always satisfied for $n = 3$, and for $n > 3$

$$R_j > -\left[\frac{1-\alpha}{\alpha^2}\frac{2}{(n-3)y}\right], \tag{2.3.74}$$

where $R_j \geq 0$. As player J's indirect reciprocity preferences matter more (i.e., as γ increases), contributing to their own group even when players in their own group are not contributing is an equilibrium. This finding suggests that if player J has strong reciprocity preferences and $\gamma > \frac{n-1}{n}$, then player J will contribute their full endowment given that the history of the game is $h = (y, ..., y)$ despite others in their group contributing nothing.

The total free-riding equilibrium, $(y, ..., 0, ..., y)$, is satisfied when for player J

$$R_J < \frac{1-\alpha}{\alpha^2}\frac{2}{(n - 1 + n\gamma)y}, \tag{2.3.75}$$

and $R_J \geq 0$, and for player j

$$\frac{1}{R_j} < \frac{\alpha^2}{1-\alpha}\left(\frac{n-3}{2}\right)y, \tag{2.3.76}$$

where the condition $R_j \geq 0$ is not satisfied if $n = 3$, but is satisfied if $n > 3$ and

$$R_j > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-3)y}. \tag{2.3.77}$$

The total free-riding equilibrium holds if player J has sufficiently weak reciprocity preferences, regardless of the value of γ, and if each player j has sufficiently strong reciprocity preferences. This makes intuitive sense, as each player j will be motivated to contribute their full endowment if they have very strong reciprocity preferences.

The interior solution equilibrium, $(c_j^*, ..., c_J^*, ..., c_j^*)$, is satisfied if the conditions for $0 \leq c_J^* \leq y$ and $0 \leq c_j^* \leq y$ can be simultaneously met (i.e., conditions (2.3.52) and (2.3.58) for player J and conditions (2.3.60) and (2.3.62) for each player j). For $0 \leq c_j^* \leq y$, the conditions can be simultaneously satisfied, but are harder to compare with the conditions expressed in a graphical representation because the conditions include R_j. For $c_j^* \leq y$,

$$R_J \geq \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1+n\gamma)y}, \tag{2.3.78}$$

and for $c_j^* \geq 0$, if $\gamma < \frac{n-1}{n}$, then $R_J \geq 0$, and if $\gamma > \frac{n-1}{n}$, then

$$R_J \leq \frac{1-\alpha}{\alpha^2}\frac{2}{(n\gamma-n+1)y}. \tag{2.3.79}$$

In Figure 2.12, we represent these different equilibria in a graph (α, R_J). In the figure, $f_3(\alpha) \equiv -\frac{1-\alpha}{\alpha^2}\frac{2}{(n-1-n\gamma)y}$ and $f_4(\alpha) \equiv \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1+n\gamma)y}$. In *Area I*, the equilibria are the traditional public goods game Nash equilibrium, $(0, ..., 0, ..., 0)$, if $\gamma > \frac{n-1}{n}$, the total free-riding equilibrium, $(y, ..., 0, ..., y)$, and the interior

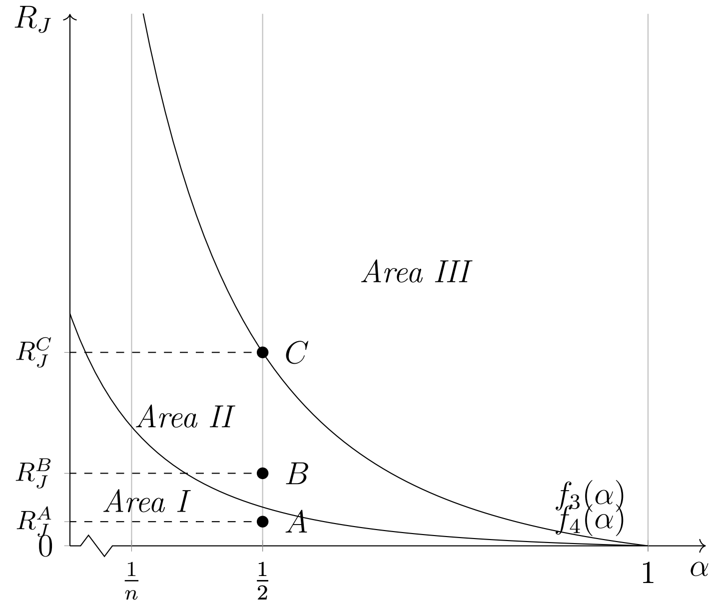

Figure 2.12: Indirect Upstream Reciprocity - Case 2: $h = (y, ..., y)$

solution equilibrium, $(c_j^*, ..., c_J^*, ..., c_j^*)$, if $\gamma > \frac{n-1}{n}$. In *Area II*, the equilibria are the traditional public goods game Nash equilibrium if $\gamma > \frac{n-1}{n}$, the interior solution equilibrium if $\gamma \lesseqgtr \frac{n-1}{n}$, and the social optimum equilibrium, $(y, ..., y, ..., y)$. In *Area III*, the equilibria are the traditional public goods game Nash equilibrium if $\gamma > \frac{n-1}{n}$, the interior solution equilibrium if $\gamma < \frac{n-1}{n}$, the altruistic equilibrium, $(0, ..., y, ..., 0)$, if $\gamma > \frac{n-1}{n}$, and the social optimum equilibrium. As γ increases, or as player J's indirect reciprocity preferences become stronger than their direct reciprocity preferences, the thresholds decrease, meaning that cooperative equilibira are "easier" to attain when indirect reciprocity preferences matter more. The converse happens when γ decreases. When player J's indirect reciprocity preferences are weaker, then the thresholds increase (i.e., $f_3(\alpha)$ and $f_4(\alpha)$ shift out).

To illustrate, let $\alpha = \frac{1}{2}$. If player J's reciprocity preference is $R_J^A < f_4(\alpha)$, then player J is in *Area I*. If player J's reciprocity preference is $f_4(\alpha) < R_J^B < f_3(\alpha)$, then player J is in *Area II*. At point C, player J has very strong reciprocity preferences, $R_J^C > f_3(\alpha)$, and player J is in *Area III*.

2.3.2.2.1 Comparative Statics We examine the case of cooperative behaviour and compare how equilibria may be affected as certain parameters change. We compare the two potential equilibria in *Area III* in Figure 2.12 when $\gamma > \frac{n-1}{n}$ - the traditional public goods game Nash equilibrium, $(0, ..., 0, ..., 0)$, and the altruistic equilibrium, $(0, ..., y, ..., 0)$. We use Mathematica (Wolfram Research, Inc, 2020), a technical computing system, to visualize the data. Mathematica code for reproduction of the interactive visualization is available in Appendix A.1.2. Static snapshots and descriptions with parameter values are presented below.

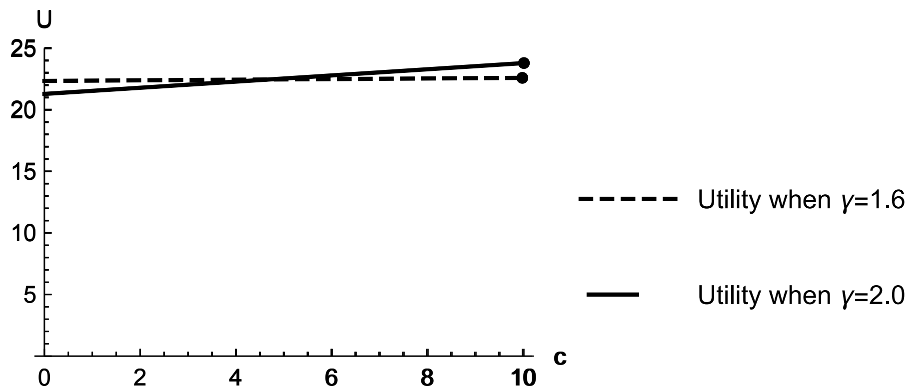

Figure 2.13: Indirect Upstream Reciprocity - Case 2. Player J's utility when $R_J = 0.15$.

In Figure 2.13, we represent player J's maximized utility as a function of contribution choice, where U represents player J's utility level and c represents player J's contribution choice. When the history of the game is $h = (y, ..., y)$, $n = 3$, $y = 10$, $\alpha = \frac{1}{2}$, player J believes that the other players are contributing nothing (i.e., $b_{J,j} = 0$), and player J has weak reciprocity preferences ($R_J = 0.15$), then player J will fully contribute to the group account when their indirect reciprocity preferences are stronger than their direct reciprocity preferences (i.e., $\gamma > 1.55$). Recall that the conditions $R_J > f_3(\alpha)$ and $\gamma > \frac{n-1}{n}$ are required for either the traditional public goods game equilibrium or the altruistic equilibrium to be potential equilibria in *Area III* in Figure 2.12. These conditions are satisfied for values of $\gamma > 1.55$ when $R_J = 0.15$. Player J's utility function is maximized when they are fully cooperating and contributing their full endowment to the group account, $c_J = y$, such that the optimal outcome of the game is the altruistic equilibrium. Player J will fully contribute to the group account even when their direct reciprocity preferences are weak, provided that they have strong indirect reciprocity preferences. This suggests that, when the history of the game is $h = (y, ..., y)$, recent experience can drive cooperative behaviour, even when others in the group are not contributing.

As direct reciprocity preferences become stronger, altruistic behaviour remains optimal for player J and the traditional public goods game Nash equilibrium prevails for all values of $\gamma > \frac{n-1}{n}$. In Figure 2.14, when $R_J = 0.5$, then player J's utility function is maximized when $c_J = y$ at the altruistic equilibrium, $(0, ..., y, ..., 0)$, for all values of $\gamma > 0.933$.

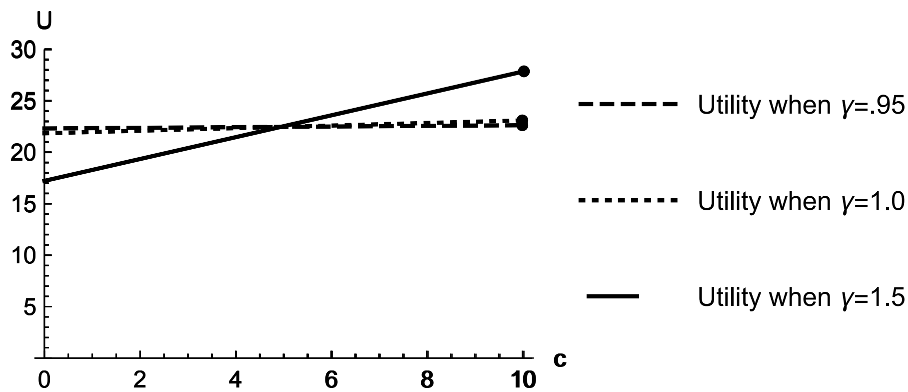

Figure 2.14: Indirect Upstream Reciprocity - Case 2. Player J's utility when $R_J = 0.5$.

In Figure 2.15, when $R_J = 1$, then player J's utility is maximized when $c_J = y$, and altruistic behaviour remains optimal for all values of $\gamma > 0.8$.

The altruistic equilibrium is driven by strong indirect reciprocity preferences. Indirect reciprocity preferences influence player J's optimal strategy to contribute fully, even when the other players in their group are not contributing. For example, when comparing across different values for player J's direct reciprocity preference from $R_J = 0.15$ to $R_J = 1$, this effect becomes even more pronounced. The slope of player J's utility function when $\gamma = 1.5$ is greatest when $R_J = 1$, compared to lower values for the direct reciprocity preference parameter. When player J's indirect reciprocity preference is relatively weaker

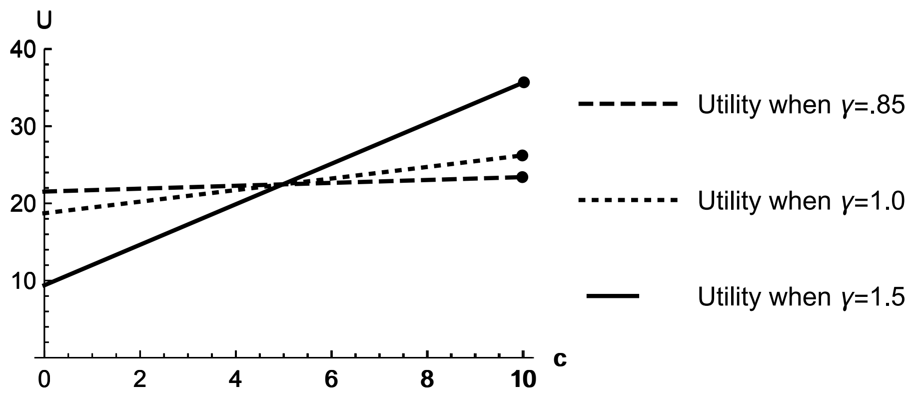

Figure 2.15: Indirect Upstream Reciprocity - Case 2. Player J's utility when $R_J = 1$.

(i.e., $\gamma < 1$) yet player J's direct reciprocity preference is sufficiently strong, then altruistic behaviour can still be optimal. Player J's indirect reciprocity preference needs to be "strong enough" in combination with strong direct reciprocity preferences to reach the altruistic equilibrium, $(0, ..., y, ..., 0)$. Given player J's beliefs about the contribution of each player $j \in B_J$ (i.e., $b_{J,j} = 0$), player J will be "kind" to players in set B_J, if players in set A were "kind" to player J (i.e., when the history of the game is $h = (y, ..., y)$) and indirect reciprocity matters to player J. That is, a recent experience can positively influence player J's contribution choice if indirect reciprocity matters to player J. Even if player J has stronger direct reciprocity preferences, $R_J = 1$, if player J's indirect reciprocity preferences are strong, $\gamma = 1.5$, then the altruistic equilibrium is optimal.

The prevalence of the altruistic equilibrium, $(0, ..., y, ..., 0)$, when $h = (y, ..., y)$ and player J believes that each player $j \in B_J$ will contribute nothing, $b_{J,j} = 0$, supports that recent experience and indirect reciprocity preferences drive cooperative behaviour for player J. As well, this result supports the "pay-it-forward" definition of indirect upstream reciprocity: player J perceived each player $i \in A$ to be "kind" to player J and player J is "kind" to players in set B_J.

See Appendix A.1.2.2 for Mathematica code to reproduce the figures at the specified parameter values.

2.3.2.3 Case 3

Finally, we examine the history of the game $h = (c, ..., c)$. While $h = (c, ..., c)$ represents a multiplicity of interior contributions, in Case 3 we examine the special case of the interior solution equilibrium, $h = (c^*, ..., c^*)$. If $h = (c^*, ..., c^*)$, then player J has observed $\frac{1}{n}\sum_{i\in A} c_i = c \in]0, y[$. This means that player J does not know with certainty the contribution amount of each player $i \in A$. Player J is choosing their contribution amount at the same time as each player $j \in B_J$ and player J must form beliefs about the contribution strategy of each player $j \in B_J$. From the baseline case in Section 2.1, we know that $c_i^* = \frac{1-\alpha}{\alpha^2}\frac{1}{(n-1)R_i} + \frac{1}{2}y \;\; \forall \;\; i \in A$, and $0 \leq c_i^* \leq y$ if $R_i \geq \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}$. Because we assume that players are symmetric in the baseline case (i.e., each player $i \in A$) and in set B_J (i.e., each player $j \in B_J$), we let $R_i = R_j = R$. Player J's best response function is

$$BR_J\left((b_{J,j}(c, ..., c))_{j\neq J}\right) = \begin{cases} 0 & \text{if} \quad \sum_{j\in B_J} b_{J,j} < \quad \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_J} - \frac{n\gamma}{n-1}\frac{1}{R}\right) \\ & \qquad\qquad\qquad + \frac{n-1}{2}y \\ [0, y] & \text{if} \quad \sum_{j\in B_J} b_{J,j} = \quad \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_J} - \frac{n\gamma}{n-1}\frac{1}{R}\right) \\ & \qquad\qquad\qquad + \frac{n-1}{2}y \\ y & \text{if} \quad \sum_{j\in B_J} b_{J,j} > \quad \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_J} - \frac{n\gamma}{n-1}\frac{1}{R}\right) \\ & \qquad\qquad\qquad + \frac{n-1}{2}y \end{cases} \tag{2.3.80}$$

From the best response functions for players J and j, (2.3.80) and (2.3.16) respectively,

$$(n-1)c_j = \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_J} - \frac{n\gamma}{n-1}\frac{1}{R}\right) + \frac{n-1}{2}y, \tag{2.3.81}$$

and

$$(n-2)c_j + c_J = \frac{1-\alpha}{\alpha^2 R_j} + \frac{n-1}{2}y. \tag{2.3.82}$$

Solving for an interior solution, we find

$$c_J^* = \frac{1-\alpha}{\alpha^2}\left(\frac{(n+n\gamma)(n-2)+1}{(n-1)^2}\frac{1}{R} - \frac{n-2}{n-1}\frac{1}{R_J}\right) + \frac{1}{2}y, \tag{2.3.83}$$

where $R_J \geq 0$ and $R \geq 0$. For c_J^* to be an eligible interior solution, it must satisfy $0 \leq c_J^* \leq y$. We verify that $c_J^* \leq y$ if

$$\frac{\overbrace{\frac{(n+n\gamma)(n-2)+1}{(n-1)^2} - \frac{\alpha^2}{1-\alpha}\frac{Ry}{2}}^{(i)}}{R} \leq \frac{n-2}{n-1}\frac{1}{R_J}, \tag{2.3.84}$$

where if part $(i) > 0$, then

$$0 \leq R \leq \frac{1-\alpha}{\alpha^2}\frac{2(n+n\gamma)(n-2)+2}{(n-1)^2 y}, \tag{2.3.85}$$

and

$$0 \leq R_J \leq \frac{n-2}{n-1}\left(\frac{R}{\frac{(n+n\gamma)(n-2)+1}{(n-1)^2} - \frac{\alpha^2}{1-\alpha}\frac{Ry}{2}}\right). \tag{2.3.86}$$

If part $(i) \leq 0$, then

$$R \geq \frac{1-\alpha}{\alpha^2}\frac{2(n+n\gamma)(n-2)+2}{(n-1)^2 y}, \tag{2.3.87}$$

and

$$R_J \geq \frac{n-2}{n-1}\left(\frac{R}{\frac{(n+n\gamma)(n-2)+1}{(n-1)^2} - \frac{\alpha^2}{1-\alpha}\frac{Ry}{2}}\right), \tag{2.3.88}$$

which is satisfied as $R_J \geq 0$. We check that $c_J^* \geq 0$ if

$$R_J \geq \frac{(n-1)(n-2)}{(n+n\gamma)(n-2)+1} R + \frac{1-\alpha}{\alpha^2} \frac{2(n-2)}{(n-1)y}. \tag{2.3.89}$$

The conditions for $0 \leq c_J^* \leq y$ can be simultaneously satisfied, therefore c_J^* exists as an eligible interior solution. From player J's best response function (2.3.80) and solving for an interior solution for each player j, we find

$$c_j^* = \frac{1-\alpha}{\alpha^2} \left(\frac{1}{n-1} \frac{1}{R_J} - \frac{n\gamma}{(n-1)^2} \frac{1}{R} \right) + \frac{1}{2} y, \tag{2.3.90}$$

where $R_J \geq 0$, $R \geq 0$, and $0 \leq c_j^* \leq y$. We verify that $c_j^* \leq y$ if

$$\frac{1}{R} \geq \overbrace{\frac{n-1}{n\gamma} \frac{1}{R_J} - \frac{\alpha^2}{1-\alpha} \frac{(n-1)^2}{2n\gamma} y}^{(i)}, \tag{2.3.91}$$

where if part $(i) > 0$, then

$$R_J > \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1)y}, \tag{2.3.92}$$

and

$$0 \leq R \leq \frac{n\gamma}{n-1} \frac{1}{R_J} - \frac{1-\alpha}{\alpha^2} \frac{2n\gamma}{(n-1)^2 y}. \tag{2.3.93}$$

If part $(i) < 0$, then

$$0 \leq R_J < \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1)y}, \tag{2.3.94}$$

and

$$R \geq \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1)y}, \tag{2.3.95}$$

which is always satisfied since $R \geq 0$. We check that $c_j^* \geq 0$ if

$$R \geq \frac{n\gamma}{n-1}R_J + \frac{1-\alpha}{\alpha^2}\frac{2n\gamma}{(n-1)^2 y}, \tag{2.3.96}$$

where $R \geq 0$. The conditions for $0 \leq c_j^* \leq y$ can be simultaneously satisfied when $0 \leq R_J \leq \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}$. If conditions are satisfied at the same time, then c_j^*, c_J^* can exist simultaneously.

When $h = (c, ..., c)$, then there exists seven possible equilibria. The traditional public goods game Nash equilibrium, $(0, ..., 0, ..., 0)$, is satisfied if the following conditions are met for players J and each player j. For player J

$$\frac{1}{R_J} > \overbrace{\frac{n\gamma}{n-1}\frac{1}{R} - \frac{\alpha^2}{1-\alpha}\frac{(n-1)y}{2}}^{(i)}, \tag{2.3.97}$$

where if part $(i) > 0$, then

$$0 \leq R < \frac{1-\alpha}{\alpha^2}\frac{2n\gamma}{(n-1)^2 y}, \tag{2.3.98}$$

and

$$0 \leq R_J < \frac{n-1}{n\gamma}R - \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}. \tag{2.3.99}$$

If part $(i) < 0$, then

$$R > \frac{1-\alpha}{\alpha^2}\frac{2n\gamma}{(n-1)^2 y}, \tag{2.3.100}$$

and

$$R_J > \frac{n-1}{n\gamma}R - \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.3.101}$$

which is always met as $R_J \geq 0$. Player j's best response is $c_j = 0$ if

$$R > -\left[\frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}\right], \tag{2.3.102}$$

which is always satisfied as $R \geq 0$. This equilibrium can be reached with or without reciprocity preferences (i.e., when $R \geq 0, R_J \geq 0$).

The social optimum equilibrium, $(y, ..., y, ..., y)$, is satisfied if the following conditions are met. For player J

$$R_J > \frac{n-1}{n\gamma}R + \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}. \tag{2.3.103}$$

For each player j

$$R > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}. \tag{2.3.104}$$

By incorporating reciprocity preferences, it is possible to reach the social optimum equilbrium if players have strong preferences for reciprocity.

The altruistic equilibrium, $(0, ..., y, ..., 0)$, is possible under certain conditions. For player J

$$\frac{1}{R_J} < \overbrace{\frac{n\gamma}{n-1}\frac{1}{R} - \frac{\alpha^2}{1-\alpha}y\frac{n-1}{2}}^{(i)}, \tag{2.3.105}$$

where if part $(i) > 0$, then

$$0 \leq R_J < \frac{1-\alpha}{\alpha^2}\frac{2n\gamma}{(n-1)^2 y}, \tag{2.3.106}$$

and

$$R_J > \frac{n-1}{n\gamma}R - \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.3.107}$$

and if part $(i) < 0$, then

$$R > \frac{1-\alpha}{\alpha^2} \frac{2n\gamma}{(n-1)^2 y}, \tag{2.3.108}$$

and the condition $R_J \geq 0$ is not met. For player j, the condition

$$R > -\left[\frac{1-\alpha}{\alpha^2} \frac{2}{(n-3)y}\right], \tag{2.3.109}$$

is always satisfied since $R \geq 0$. If player J's reciprocity preferences are sufficiently strong (i.e., condition (2.3.107)), player J will contribute their full endowment even when other players in set B are not contributing anything. As γ increases (i.e., as player J's indirect reciprocity preferences become stronger than their direct reciprocity preferences), the threshold for contributing their full endowment decreases.

The "kind" equilibrium, $(0, ..., c_J^*, ..., 0)$, is satisfied if for player J

$$R_J = \frac{n-1}{n\gamma} R - \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1)y}, \tag{2.3.110}$$

where $R_J \geq 0$ if

$$R \geq \frac{1-\alpha}{\alpha^2} \frac{2n\gamma}{(n-1)^2 y}, \tag{2.3.111}$$

and for player j

$$R > \frac{n-1}{n\gamma} R_J + \frac{1-\alpha}{\alpha^2} \frac{2n\gamma(n-2)}{(n-1)^2 y}, \tag{2.3.112}$$

which is possible if the conditions for $0 \leq c_J^* \leq y$ are also satisfied. Note that the conditions previously outlined for the existence of c_J^* must hold as well (i.e., conditions (2.3.84) and (2.3.89)). Each player j will free-ride despite having relatively higher reciprocity preferences in this case.

The equilibrium with some free-riding, $(y, ..., c_J^*, ..., y)$, is satisfied if the following conditions are met. For player J

$$R_J = \frac{n-1}{n\gamma}R + \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}. \tag{2.3.113}$$

For player j

$$\frac{1}{R} > \overbrace{\frac{n-1}{n\gamma}\frac{1}{R_J} - \frac{\alpha^2}{1-\alpha}\frac{(n-1)^2 y}{2n\gamma}}^{(i)}, \tag{2.3.114}$$

where if part $(i) > 0$, then

$$0 \leq R_J < \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.3.115}$$

which is not possible because of the condition expressed in condition (2.3.113), and if part $(i) < 0$, then

$$R_J > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.3.116}$$

and

$$R > \frac{n\gamma}{n-1}R_J - \frac{1-\alpha}{\alpha^2}\frac{2n\gamma}{(n-1)^2 y}, \tag{2.3.117}$$

and $R \geq 0$. Note that the conditions previously outlined for the existence of c_J^* must hold as well. In this equilibrium, each player j will contribute their full endowment if player J and each player j all have sufficiently strong reciprocity preferences.

The total free-riding equilibrium, $(y, ..., 0, ..., y)$, is satisfied if for player J

$$0 \leq R_J < \frac{n-1}{n\gamma}R + \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.3.118}$$

and for player j

$$R > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-3)y}, \tag{2.3.119}$$

which is possible when $n > 3$. Player J will free-ride when their reciprocity preferences are weaker. Each player j will contribute their full endowment if their reciprocity preferences are sufficiently strong.

The interior solution equilibrium, $(c_j^*, ..., c_J^*, ..., c_j^*)$, is satisfied if the conditions for $0 \leq c_J^* \leq y$ and $0 \leq c_j^* \leq y$ are simultaneously satisfied. The interior solution equilibrium does not represent a multiplicity of equilibria, but a very specific set of values for contributions which satisfy certain conditions (i.e., conditions (2.3.84) and (2.3.89) for player J and conditions (2.3.91) and (2.3.96) for each player j) and also vary depending on the values of R_J and R.

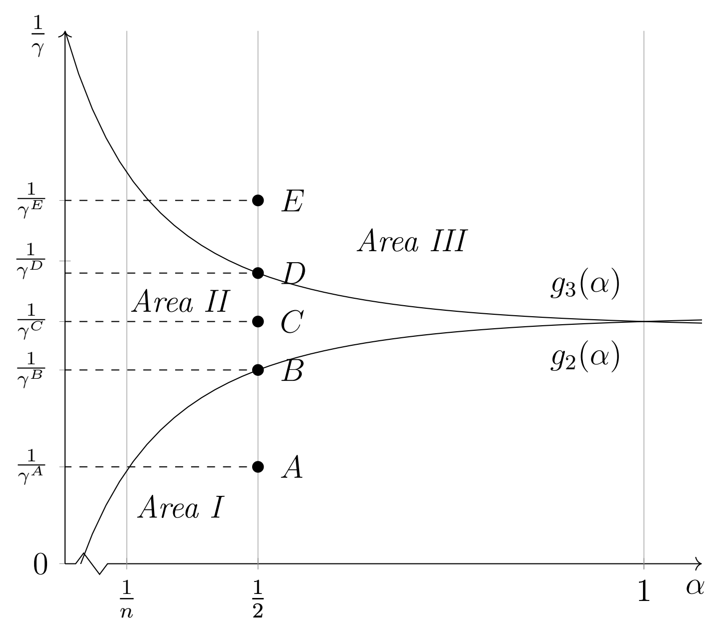

Figure 2.16: Indirect Upstream Reciprocity - Case 3: $h = (c, ..., c)$

In Figure 2.16, we represent the equilibria in a graph $(\alpha, \frac{1}{\gamma})$. Note that unlike Section 2.3.2.1 and Section 2.3.2.2, the equilibria areas are presented in the $(\alpha, \frac{1}{\gamma})$ space rather than the (α, R_J) space due to the structure of the equilibria conditions when the history of the game is $h = (c, ..., c)$ and to increase comparison across histories. Recall that $R_I = \gamma R_J$, therefore $\gamma = \frac{R_I}{R_J}$. It follows that $\frac{1}{\gamma} = \frac{R_J}{R_I}$. In Figure 2.16, $g_2(\alpha) \equiv \frac{n}{n-1}\frac{R_J}{R} - \frac{1-\alpha}{\alpha^2 R}\frac{2n}{(n-1)^2 y}$, and $g_3(\alpha) \equiv \frac{n}{n-1}\frac{R_J}{R} + \frac{1-\alpha}{\alpha^2 R}\frac{2n}{(n-1)^2 y}$. In *Area I*, the equilibria are the traditional public goods game Nash equilibrium, $(0, ..., 0, ..., 0)$, when $\frac{1}{\gamma} < \frac{1-\alpha}{\alpha^2 R}\frac{2n}{(n-1)^2 y}$, the social optimum equilibrium, $(y, ..., y, ..., y)$, and the total free-riding equilibrium, $(y, ..., 0, ..., y)$. In *Area II*, the only equilibrium is the traditional public goods game Nash equilibrium when $\frac{1}{\gamma} > \frac{1-\alpha}{\alpha^2 R}\frac{2n}{(n-1)^2 y}$. In *Area III*, the equilibria are the traditional public goods game Nash equilibrium when $\frac{1}{\gamma} > \frac{1-\alpha}{\alpha^2 R}\frac{2n}{(n-1)^2 y}$ and the altruistic equilibrium, $(0, ..., y, ..., 0)$.

The "kind" equilibrium, $(0, ..., c_J^*, ..., 0)$, exists along $g_3(\alpha)$ and the equilibrium with some free-riding, $(y, ..., c_J^*, ..., y)$, exists along $g_2(\alpha)$. Finally, the interior solution equilibrium, $(c_j^*, ..., c_J^*, ..., c_j^*)$, is met when the conditions for $0 \leq c_J^* \leq y$ (i.e., conditions (2.3.84) and (2.3.89)) and $0 \leq c_j^* \leq y$ (i.e., conditions (2.3.91) and (2.3.96)) are simultaneously satisfied.

When $\frac{1}{\gamma}$ is low (i.e., $\frac{1}{\gamma^A}$), then γ is relatively large and player J's indirect reciprocity preferences are stronger than their direct reciprocity preferences. When $\frac{1}{\gamma}$ is high (i.e., $\frac{1}{\gamma^C}$), then γ is relatively small and player J's indirect reciprocity preferences are weaker than their direct reciprocity preferences. As γ increases, or as player J's indirect reciprocity preferences become stronger than their direct reciprocity preferences (i.e., $R_I > R_J$), then $\frac{1}{\gamma}$ decreases and the equilibria thresholds decrease. As γ increases, or as player J's indirect reciprocity preferences become greater than their direct reciprocity preferences (i.e., $R_I > R_J$), then $\frac{1}{\gamma}$ decreases and the equilibria thresholds decrease. This makes some of the cooperative equilibria more accessible or "easier" to reach.

To illustrate this further, let $\alpha = \frac{1}{2}$ such that $g_2(\frac{1}{2}) \equiv \frac{n}{n-1}\frac{R_J}{R} - \frac{4n}{(n-1)^2 R y}$, and $g_3(\frac{1}{2}) \equiv \frac{n}{n-1}\frac{R_J}{R} + \frac{4n}{(n-1)^2 R y}$. At point A, γ is relatively larger and player J's indirect reciprocity preferences are stronger than their direct reciprocity preferences. At point A, player J is in *Area I* and faces three potential equilibria. As player J's indirect reciprocity preferences decrease and $\frac{1}{\gamma}$ increases, moving to point B suggests a shift from potentially total free-riding, $(y, ..., 0, ..., y)$, to under cooperating or some free-riding, $(y, ..., c_J^*, ..., y)$. This shift highlights that as player J has weaker preferences for indirect reciprocity, they will contribute more (i.e., $c_J > 0$) as player J's direct reciprocity preference increases given that the other players are contributing fully to the group account.

A similar shift occurs between point D and E, although with a less clear interpretation. At point D, player J shifts from the "kind" equilibrium, $(0, ..., c_J^*, ..., 0)$, to point E, either the altruistic equilibrium, $(0, ..., y, ..., 0)$, or the traditional public goods game Nash equilibrium, $(0, ..., 0, ..., 0)$, which is always satisfied. This shift, accompanied by a decrease in γ, suggests that direct reciprocity preferences matter more. However, if direct reciprocity preferences mattered more to player J, they would likely not contribute anything to their group account if player J correctly believes that the others are contribution nothing. The possibility of altruistic behaviour as optimal for player J indicates that the relationship between direct and indirect reciprocity preferences is influencing player J's contribution behaviour in more complex ways than predicted by traditional game theory. That is, there may be complexity due to interdependency between R, R_J, and γ. The traditional public goods game Nash equilibrium holds if $\frac{1}{\gamma} > \frac{1-\alpha}{\alpha^2 R}\frac{2n}{(n-1)^2 y}$, or if $R > \frac{1-\alpha}{\alpha^2}\frac{2n\gamma}{(n-1)^2 y}$. The altruistic equilibrium arises if R is weak and γ is small, suggesting that player J's altruistic behaviour is driven by strong direct reciprocity preferences, R_J.

As discussed in Section 2.2.3, comparative statics for $h = (c, ..., c)$ or interior solution equilibria are not presented here. Comparisons at the corner cases provide greater insight into thresholds for cooperative behaviour.

2.4 Indirect Downstream Reciprocity

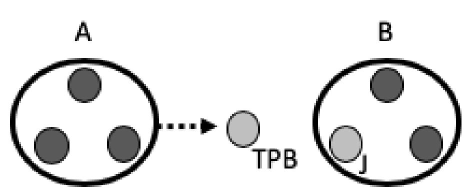

Figure 2.17: Indirect Downstream Reciprocity

Consider a variation of the baseline model, in Figure 2.17, where players interact simultaneously within their own group and there is an opportunity for one player, player J, to observe the outcome of a previous game. In this multi-stage game, there are two groups of n players each, where $A = \{1, ..., n\}$ and $B = \{1, ..., i-1, i+1, ..., n, J\}$ with player $J \in B$ and $n \geq 3$. Note that $B_J = \{1, ..., i-1, i+1, ..., n\}$, where $B_J \subset B$, is a group of $(n-1)$ players randomly selected from set A, therefore $B_J \subset A$. In the first period, each player $i \in A$ simultaneously chooses their contribution $c_i \in [0, y]$. This game has the same structure as the baseline case and the utility functions and payoffs for each player $i \in A$ are outlined in Section 2.1.

Similar to the model variation in Section 2.3 which discusses indirect upstream reciprocity, there is an outside player who is a beneficiary from the contributions to the group account from players in set A. In this variation with indirect downstream reciprocity, an outside player receives some monetary benefit without participating in the public goods game with players in set A; in Figure 2.17, this outside player is denoted as the third-party beneficiary (TPB). As in Section 2.3, in the second period player $J \in B$ observes the average contribution of n players in set A, $\frac{1}{n}\sum_{i \in A} c_i$, and player J knows that the TPB has benefited from the contributions of players in set A. After player J has made this observation, all players in set B, including player J, simultaneously choose their contribution amounts. Recall that players are myopic; that is, we assume that players are myopic in their decisions during each period and care only about the current period's payoff. This assumption allows us to focus on the reciprocal decision-making of player J and ignore dynamic interactions between periods[6]. The rewarding behaviour or reputation-based reciprocity that characterizes indirect downstream reciprocity can be captured in player J's contribution choice, c_J. Player J chooses $c_J \in [0, y]$, and each player $k \in B_J$ chooses $c_k \in [0, y]$. The payoff for all players in set B is analogous to (2.1.1) in Section 2.1. As before, we restrict MPCR such that

[6]While there may be a dynamic component of reciprocity, we are interested in how player J will be affected by observing first period behaviour, rather than understanding reciprocity in the first period or how reciprocity influences behaviour over time.

$0 < \alpha < 1 < n\alpha$ to ensure the conditions of a public goods game are met. We assume no discounting between periods.

2.4.1 No Reciprocity Preferences

In the second period, player $J \in B$ chooses c_J that solves

$$\underset{c_J}{max}\{y - c_J + \alpha \sum_{l \in B} c_l\}. \tag{2.4.1}$$

The first-order condition is

$$-1 + \alpha < 0. \tag{2.4.2}$$

We are again at a corner solution where the optimal contribution choice is $c_J^* = 0$ because the maximization problem is linear. At the same time, each player $k \in B_J$ simultaneously chooses c_k to solve

$$\underset{c_k}{max}\{y - c_k + \alpha \sum_{l \in B} c_l\}. \tag{2.4.3}$$

The first-order condition is

$$-1 + \alpha < 0. \tag{2.4.4}$$

We are again at a corner solution where the optimal contribution choice is $c_k^* = 0 \ \forall \ k \in B_J$, such that $c_1^* = ... = c_{i-1}^* = c_{i+1}^* = ... = c_n^* = c_J^* = 0$. The game results in the traditional public goods game Nash equilibrium, $(c_1^*, ..., c_{i-1}^*, c_{i+1}^*, ..., c_n^*, c_J^*) = (0, ..., 0)$. The first stage of the game proceeds exactly as in the baseline case presented in Section 2.1. We have previously seen that the outcome of that game is the traditional public goods game Nash equilibrium, $(c_1^*, ..., c_n^*) = (0, ..., 0)$, as well. In the absence of reciprocity preferences, and even though player $J \in B$ observes the average contribution to the group account from players in set A, none of the players in set B have an incentive to contribute to their own group account, and we remain at the traditional public goods game Nash equilibrium.

2.4.2 Reciprocity Preferences

Incorporating reciprocity preferences into players' utility functions takes a similar form as in the direct reciprocity model, Section 2.1.2, and the indirect upstream reciprocity model, Section 2.3.2. Player J's reciprocity payoff function accounts for both direct and indirect reciprocity preferences. Player J's direct reciprocity preference captures player J's reciprocity preferences towards the players in their own set, set B. Player J's indirect reciprocity preference captures the relative weight that player J places on player J's own evaluation of the behaviour (i.e., contributions) of each player $i \in A$ in the first stage, knowing that an $(n-1)$ random sample of these players are in subset B_J. Player J's indirect reciprocity preference is player J's reciprocity preference towards the interaction that occurred between players in set A and the TPB, where $R_{J,A|TPB} = R_I$.

We express player J's utility function as

$$
\begin{aligned}
U_J = \{ & \underbrace{y - c_J + \alpha \sum_{l \in B} c_l}_{(i)} \\
& + \underbrace{\sum_{k \in B_J} R_{J,k} \cdot \kappa_{J,k}(c_J, (b_{J,k}(h))_{k \neq J}) \cdot \lambda_{J,k,J}(b_{J,k}(h), (b_{J,k,l}(h))_{l \neq k})}_{(ii)} \\
& + \underbrace{R_I \cdot \bar{\kappa}_{J,k}(c_J, (b_{J,k}(h))_{k \neq J})_{k \in B_J} \cdot \bar{\lambda}_{J,i,TPB}(\bar{b}_{J,i}(h), (\bar{b}_{J,i,k}(h))_{k \neq i})_{i \in A}}_{(iii)} \},
\end{aligned}
\tag{2.4.5}
$$

where part (i) in (2.4.5) is the material payoff, part (ii) is the same direct reciprocity payoff as before, and part (iii) is player J's indirect reciprocity payoff. Note that in contrast to the indirect upstream reciprocity model in Section 2.3.2, player J's indirect reciprocity payoff has a different form to capture the reputation-based effects of indirect downstream reciprocity. Because player J is evaluating the action of players $i \in A$ towards the TPB, player J is evaluating average kindness and average perceived kindness. Consequently, player J holds only one indirect reciprocity preference, R_I. The kindness function evaluates how "kind" player J is, on average, to players in set B_J. In part (iii) in (2.4.5), the perceived kindness function is now an evaluation of how "kind" player J perceives each player $i \in A$ to be towards the TPB, on average.

The history of the game plays an important role in indirect downstream reciprocity. Indeed, the history of the game, $h = (c_1, ..., c_n)$, is observed by player J through the average contribution of players in set A. Player J observes $\frac{1}{n}\sum_{i \in A} c_i$ and knows with probability 1 that each player $i \in A$ chose $c_i = 0$ or $c_i = y$ when the average contribution is $\frac{1}{n}\sum_{i \in A} c_i = 0$ or y, respectively. When player J observes $\frac{1}{n}\sum_{i \in A} c_i = c \in \ ]0, y[$, then they do not know with certainty what each player $i \in A$ has contributed. Player J must form beliefs, $b_{J,i}(h)$, about the other players' contribution strategies. Following Section 2.1.2 and (2.1.10), player J's direct reciprocity payoff, part (ii) in (2.4.5) is

$$R_J \alpha^2 \left(c_J - \frac{1}{2}y\right) \left(\sum_{k \in B_J} b_{J,k}(h) - \frac{n-1}{2}y\right), \tag{2.4.6}$$

where R_J represents player J's direct reciprocity preference.

Next, player J's indirect reciprocity payoff function, part (iii) in (2.4.5) is represented as

$$R_I \cdot \bar{\kappa}_{J,k}(c_J, (b_{J,k}(h))_{k \neq J})_{k \in B_J} \cdot \bar{\lambda}_{J,i,TPB}(\bar{b}_{J,i}(h), (\bar{b}_{J,i,k}(h))_{k \neq i})_{i \in A}. \tag{2.4.7}$$

The indirect reciprocity payoff function captures the effect of player J's perceived kindness of players in set A on player J's own kindness towards players in set B_J. As a result, player J evaluates average kindness and average perceived kindness. Recall that player J does not know which players in set A

were "kind" and also does not know which players from set A are now in set B_J. First, player J's average kindness function, $\bar{\kappa}_{J,k}(\cdot)$, evaluates player J's kindness, on average, towards players in set B_J. From (2.1.8d), we know that

$$\kappa_{J,k}(c_J, (b_{J,k}(h))_{k \neq J}) = \alpha \left(c_J - \frac{1}{2} y \right), \tag{2.4.8}$$

where $k \in B_J$. Player J's average kindness function is

$$\bar{\kappa}_{J,k}(c_J, (b_{J,k}(h))_{k \neq J})_{k \in B_J} = \frac{1}{n-1} \sum_{k \in B_J} \kappa_{J,k}(c_J, (b_{J,k}(h))_{k \neq J}), \tag{2.4.8a}$$

or

$$\bar{\kappa}_{J,k}(\cdot) = \alpha \left(c_J - \frac{1}{2} y \right). \tag{2.4.8b}$$

Player J's average perceived kindness function, $\bar{\lambda}_{J,i,TPB}(\bar{b}_{J,i}(h), (\bar{b}_{J,i,k}(h))_{k \neq i})_{i \in A}$, evaluates player J's perception of the average kindness of players in set A towards the *TPB*. Following (2.1.9d) and (2.1.4d)

$$\bar{\lambda}_{J,i,TPB}(\bar{b}_{J,i}(h), (\bar{b}_{J,i,k}(h))_{k \neq i})_{i \in A} = \alpha \left(\bar{b}_{J,i}(h) - \frac{1}{2} y \right). \tag{2.4.9}$$

Player J observes $\frac{1}{n} \sum_{i \in A} c_i$ which is equivalent to $\bar{b}_{J,i}(h) \;\forall\; i \in A$. We denote R_I as player J's indirect reciprocity preference and let $R_I = \gamma R_J$, where γ represents the relative strength of indirect reciprocity preferences and $\gamma > 0$. Recall $\gamma \lesseqgtr 1$, where $\gamma > 1$ indicates relatively stronger indirect reciprocity preferences and suggests altruistic behaviour in some cases, and $\gamma < 1$ indicates relatively weaker indirect reciprocity preferences.

We express player J's utility function as

$$\begin{aligned} U_J = y - c_J + \alpha \sum_{l \in B} c_l \\ + R_J \alpha^2 \left(c_J - \frac{1}{2} y \right) \left(\sum_{k \in B_J} b_{J,k}(h) - \frac{n-1}{2} y \right) \\ + R_J \alpha^2 \gamma \left(c_J - \frac{1}{2} y \right) \left(\frac{1}{n} \right) \left(\sum_{i \in A} c_i - \frac{n}{2} y \right). \end{aligned} \tag{2.4.10}$$

Rearranging the terms in (2.4.10), we obtain player J's utility maximization problem

$$\begin{aligned} \max_{c_J} \Bigg\{ y - c_J + \alpha \sum_{l \in B} c_l \\ + R_J \alpha^2 \left(c_J - \frac{1}{2} y \right) \left(\sum_{k \in B_J} b_{J,k}(h) + \frac{\gamma}{n} \sum_{i \in A} c_i - \frac{n-1+\gamma}{2} y \right) \Bigg\}. \end{aligned} \tag{2.4.11}$$

The first-order condition is

$$-1 + \alpha + R_J \alpha^2 \left(\sum_{k \in B_J} b_{J,k}(h) + \frac{\gamma}{n} \sum_{i \in A} c_i - \frac{n-1+\gamma}{2} y \right) = 0. \tag{2.4.12}$$

Player J's best response function is

$$BR_J\left((b_{J,k}(h))_{k \neq J}\right) = \begin{cases} 0 & \text{if } \sum_{k \in B_J} b_{J,k}(h) < \frac{1-\alpha}{\alpha^2 R_J} - \frac{\gamma}{n} \sum_{i \in A} c_i + \frac{n-1+\gamma}{2} y \\ [0, y] & \text{if } \sum_{k \in B_J} b_{J,k}(h) = \frac{1-\alpha}{\alpha^2 R_J} - \frac{\gamma}{n} \sum_{i \in A} c_i + \frac{n-1+\gamma}{2} y \\ y & \text{if } \sum_{k \in B_J} b_{J,k}(h) > \frac{1-\alpha}{\alpha^2 R_J} - \frac{\gamma}{n} \sum_{i \in A} c_i + \frac{n-1+\gamma}{2} y \end{cases} \tag{2.4.13}$$

Note that history is captured in the term $\frac{1}{n}\sum_{i \in A} c_i$ and player J's best response function differs from the indirect upstream reciprocity case.

As before, the utility maximization problem for each player $k \in B_J$ takes a similar form to (2.1.6) in Section 2.1.2. Each player k chooses c_k to solve

$$\begin{aligned} \max_{c_k} \Bigg\{ y - c_k + \alpha \sum_{l \in B} c_l \\ + R_k \alpha^2 \left(c_k - \frac{1}{2} y \right) \left(\sum_{j \in B_k} b_{k,j}(h) - \frac{n-1}{2} y \right) \Bigg\}, \end{aligned} \tag{2.4.14}$$

where $B_k = \{1, ..., k-1, k+1, ..., n, J\}$ and $B_k \subset B$. Recall that player k is myopic and is making their contribution choice at the same time as the other players in set B. Player k does not know the history of the game in the second period and must form beliefs about the other players' contribution strategies during that period. The first-order condition is

$$-1 + \alpha + R_k \alpha^2 \left(\sum_{j \in B_k} b_{k,j}(h) - \frac{n-1}{2} y \right) = 0. \tag{2.4.15}$$

We assume rational expectations in equilibrium, therefore player k's beliefs about each player j's contributions are correct such that $b_{k,j}(h) = c_j$. We can rewrite the first-order condition, (2.4.15) as

$$-1 + \alpha + R_k \alpha^2 \left(\sum_{j \in B_k} c_j - \frac{n-1}{2} y \right) = 0. \tag{2.4.15a}$$

Given the contributions of the $(n-1)$ players in set B_k, the best response function for each player k is

$$BR_k\left((b_{k,j}(h))_{j\neq k}\right) = \begin{cases} 0 & \text{if } \sum\limits_{j\in B_k} b_{k,j}(h) < \dfrac{1-\alpha}{\alpha^2 R_k} + \dfrac{n-1}{2}y \\ [0,y] & \text{if } \sum\limits_{j\in B_k} b_{k,j}(h) = \dfrac{1-\alpha}{\alpha^2 R_k} + \dfrac{n-1}{2}y \\ y & \text{if } \sum\limits_{j\in B_k} b_{k,j}(h) > \dfrac{1-\alpha}{\alpha^2 R_k} + \dfrac{n-1}{2}y \end{cases} \qquad (2.4.16)$$

As before, we need to consider different cases depending on the history of the game.

2.4.2.1 Case 1

First, if the history of the game is $h = (0, ..., 0)$, then player J has observed $\frac{1}{n}\sum\limits_{i\in A} c_i = 0$. This means that with probability 1, $c_i^* = 0 \;\forall\; i \in A$, thus player J knows with certainty that $c_1^* = 0, ..., c_n^* = 0$ (i.e., players in set A are at the traditional public goods game Nash equilibrium). However, because player J is choosing their contribution amount at the same time as each player $k \in B_J$, player J must form beliefs about the contribution strategy of each player $k \in B_J$.

Player J's best response function is

$$BR_J\left((b_{J,k}(0, ..., 0))_{k\neq J}\right) = \begin{cases} 0 & \text{if } \sum\limits_{k\in B_J} b_{J,k}(h) < \dfrac{1-\alpha}{\alpha^2 R_J} + \dfrac{n-1+\gamma}{2}y \\ [0,y] & \text{if } \sum\limits_{k\in B_J} b_{J,k}(h) = \dfrac{1-\alpha}{\alpha^2 R_J} + \dfrac{n-1+\gamma}{2}y \\ y & \text{if } \sum\limits_{k\in B_J} b_{J,k}(h) > \dfrac{1-\alpha}{\alpha^2 R_J} + \dfrac{n-1+\gamma}{2}y \end{cases} \qquad (2.4.17)$$

From the best response functions for player J and and each player k, (2.4.17) and (2.4.16), respectively,

$$(n-1)c_k = \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1+\gamma}{2}y, \qquad (2.4.18)$$

and

$$(n-2)c_k + c_J = \frac{1-\alpha}{\alpha^2 R_k} + \frac{n-1}{2}y. \tag{2.4.19}$$

Solving for c_J, the interior solution is

$$c_J^* = \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_k} - \frac{n-2}{n-1}\frac{1}{R_J}\right) + \frac{n-1-\gamma(n-2)}{2(n-1)}y, \tag{2.4.20}$$

where $R_J \geq 0$, $R_k \geq 0$. To be an eligible interior solution, $0 \leq c_J^* \leq y$. We verify that $c_J^* \leq y$ if

$$\frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_k} - \frac{n-2}{n-1}\frac{1}{R_J}\right) + \frac{n-1-\gamma(n-2)}{2(n-1)}y \leq y. \tag{2.4.21}$$

Rearranging the terms, we obtain

$$\frac{1}{R_J} \geq \frac{n-1}{n-2}\left(\frac{1}{R_k} - \frac{\alpha^2}{1-\alpha}\left(\frac{n-1+\gamma(n-2)}{2(n-1)}\right)y\right), \tag{2.4.21a}$$

or

$$\frac{1}{R_J} \geq \frac{n-1}{n-2}\overbrace{\left(\frac{1-\frac{\alpha^2}{1-\alpha}\left(\frac{n-1+\gamma(n-2)}{2(n-1)}\right)yR_k}{R_k}\right)}^{(i)}. \tag{2.4.21b}$$

If part $(i) > 0$, then

$$0 \leq R_k < \frac{1-\alpha}{\alpha^2}\frac{2(n-1)}{(n-1+\gamma(n-2))y}, \tag{2.4.22}$$

and

$$0 \leq R_J \leq \frac{n-2}{n-1}\left(\frac{R_k}{1-\frac{\alpha^2}{1-\alpha}\left(\frac{n-1+\gamma(n-2)}{2(n-1)}\right)yR_k}\right). \tag{2.4.23}$$

If part $(i) < 0$, then

$$R_k > \frac{1-\alpha}{\alpha^2}\frac{2(n-1)}{(n-1+\gamma(n-2))y}, \tag{2.4.24}$$

and

$$R_J \geq \frac{n-2}{n-1}\left(\frac{R_k}{1-\frac{\alpha^2}{1-\alpha}\left(\frac{n-1+\gamma(n-2)}{2(n-1)}\right)yR_k}\right), \tag{2.4.25}$$

which is always met as $R_J \geq 0$. We check that $c_J^* \geq 0$ if

$$\frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_k}-\frac{n-2}{n-1}\frac{1}{R_J}\right)+\frac{n-1-\gamma(n-2)}{2(n-1)}y \geq 0. \tag{2.4.26}$$

Rearranging the terms, we obtain

$$\frac{1}{R_J} \leq \frac{n-1}{n-2}\left(\frac{1}{R_k}+\frac{\alpha^2}{1-\alpha}\left(\frac{n-1-\gamma(n-2)}{2(n-1)}\right)y\right), \tag{2.4.26a}$$

or

$$\frac{1}{R_J} \leq \left(\frac{n-1}{n-2}\right)\frac{1+\frac{\alpha^2}{1-\alpha}\overbrace{\left(\frac{n-1-\gamma(n-2)}{2(n-1)}\right)}^{(ii)}yR_k}{R_k}. \tag{2.4.26b}$$

If part $(ii) > 0$, then $\gamma < \frac{n-1}{n-2}$ and

$$R_J \geq \frac{n-2}{n-1}\left(\frac{R_k}{1+\frac{\alpha^2}{1-\alpha}\left(\frac{n-1-\gamma(n-2)}{2(n-1)}\right)yR_k}\right). \tag{2.4.27}$$

If part $(ii) < 0$, then $\gamma > \frac{n-1}{n-2}$. When $\gamma > \frac{n-1}{n-2}$, then the right-hand side of (2.4.26b) > 0 if

$$0 \leq R_k < \frac{1-\alpha}{\alpha^2} \frac{2(n-1)}{(n-1-\gamma(n-2))y}, \tag{2.4.28}$$

and

$$R_J \geq \frac{n-2}{n-1} \left(\frac{R_k}{1 + \frac{\alpha^2}{1-\alpha} \left(\frac{n-1-\gamma(n-2)}{2(n-1)} \right) y R_k} \right) \geq 0. \tag{2.4.29}$$

When $\gamma > \frac{n-1}{n-2}$, then the right-hand side of (2.4.26b) < 0 if

$$R_k > \frac{1-\alpha}{\alpha^2} \frac{2(n-1)}{(n-1-\gamma(n-2))y}, \tag{2.4.30}$$

which is always true as $R_k \geq 0$, and

$$R_J \leq \frac{n-2}{n-1} \left(\frac{R_k}{1 + \frac{\alpha^2}{1-\alpha} \left(\frac{n-1-\gamma(n-2)}{2(n-1)} \right) y R_k} \right), \tag{2.4.31}$$

which is not possible as R_J cannot be less than zero. Therefore, $0 \leq c_J^* \leq y$ is satisfied if

$$0 \leq R_k < \frac{1-\alpha}{\alpha^2} \frac{2(n-1)}{(n-1+\gamma(n-2))y}, \tag{2.4.32}$$

and

$$\frac{n-2}{n-1} \left(\frac{R_k}{1 + \frac{\alpha^2}{1-\alpha} \left(\frac{n-1-\gamma(n-2)}{2(n-1)} \right) y R_k} \right) \leq R_J \leq \frac{n-2}{n-1} \left(\frac{R_k}{1 - \frac{\alpha^2}{1-\alpha} \left(\frac{n-1+\gamma(n-2)}{2(n-1)} \right) y R_k} \right). \tag{2.4.33}$$

From the best response functions for player J and each player k, (2.4.17) and

(2.4.16), respectively, and solving for c_k, we find an interior solution for each player $k \in B_J$

$$c_k^* = \frac{1-\alpha}{\alpha^2}\frac{1}{(n-1)R_J} + \frac{n-1+\gamma}{2(n-1)}y, \tag{2.4.34}$$

where $R_J \geq 0$. To be an eligible interior solution, $0 \leq c_k^* \leq y$. We verify that $c_k^* \leq y$ if

$$\frac{1-\alpha}{\alpha^2}\frac{1}{(n-1)R_J} + \frac{n-1+\gamma}{2(n-1)}y \leq y. \tag{2.4.35}$$

Rearranging the terms, we obtain

$$\frac{1}{R_J} \leq \frac{\alpha^2}{1-\alpha}\frac{\overbrace{(n-1-\gamma)y}^{(i)}}{2}. \tag{2.4.35a}$$

If part $(i) > 0$, then $\gamma < n-1$ and

$$R_J \geq \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1-\gamma)y}. \tag{2.4.36}$$

If part $(i) < 0$, then $\gamma > n-1$ and

$$R_J \leq \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1-\gamma)y}, \tag{2.4.37}$$

which is not possible because R_J cannot be less than zero. We check that $c_k^* \geq 0$ if

$$\frac{1-\alpha}{\alpha^2}\frac{1}{(n-1)R_J} + \frac{n-1+\gamma}{2(n-1)}y \geq 0, \tag{2.4.38}$$

or

$$R_J \geq -\left[\frac{1-\alpha}{\alpha^2}\frac{2}{(n-1+\gamma)y}\right], \tag{2.4.38a}$$

which is always satisfied as $R_J \geq 0$. Taking the above conditions for $0 \leq c_J^* \leq y$ and $0 \leq c_k^* \leq y$, (2.4.21), (2.4.27), (2.4.29), (2.4.32), (2.4.35), c_J^* and c_k^* can exist simultaneously if

$$\gamma < \frac{2n-5}{3}\left(\frac{n-1}{n-2}\right). \tag{2.4.39}$$

When the history of the game is $h = (0, ..., 0)$, there are four possible equilibria, expressed as $(c_1, ..., c_J, ..., c_n)$.

The traditional public goods game Nash equilibrium, $(0, ..., 0, ..., 0)$, is always satisfied because for player J, $0 < \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1+\gamma}{2}y$, or

$$R_J > -\left(\frac{1-\alpha}{\alpha^2}\frac{2}{(n-1+\gamma)y}\right), \tag{2.4.40}$$

is always met because $R_J \geq 0$, and for each player k, $0 < \frac{1-\alpha}{\alpha^2 R_k} + \frac{n-1}{2}y$, or

$$R_k > -\left(\frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}\right), \tag{2.4.41}$$

which is also always met as $R_k \geq 0$. This equilibrium is predicted by traditional game theory for public goods games with no reciprocity preferences. As in the previous cases of the model, the traditional public goods game Nash equilibrium remains with the inclusion of reciprocity preferences.

The total free-riding equilibrium, $(y, ..., 0, ..., y)$, is satisfied when the following conditions are met for player J and each player k. For player J, $(n-1)y < \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1+\gamma}{2}y$, or

$$\frac{1}{R_J} > \frac{\alpha^2}{1-\alpha}\frac{\overbrace{(n-1-\gamma)y}^{(i)}}{2}, \tag{2.4.42}$$

where if part $(i) > 0$, or $\gamma < n-1$, then

$$0 \leq R_J < \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1-\gamma)y}, \tag{2.4.43}$$

or if part $(i) < 0$, or $\gamma > n-1$, then

$$R_J > \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1-\gamma)y}, \tag{2.4.44}$$

which is always met as $R_J \geq 0$. For each player k, $(n-2)y > \frac{1-\alpha}{\alpha^2 R_k} + \frac{n-1}{2}y$, or

$$\frac{1}{R_k} < \frac{\alpha^2}{1-\alpha} \frac{(n-3)y}{2}, \tag{2.4.45}$$

where if $n > 3$, then

$$R_k > \frac{1-\alpha}{\alpha^2} \frac{2}{(n-3)y}. \tag{2.4.46}$$

The existence of the total free-riding equilibrium suggests that if player J has observed non-cooperative behaviour and the history of the game is $h = (0, ..., 0)$, then player J will choose not to contribute to the group account. This suggests that indirect reciprocity preferences play a role in free-riding behaviour. Even when player J's indirect reciprocity preferences are relatively strong, $\gamma > n-1 > 1 \;\forall\; n \geq 3$, then player J's direct reciprocity preferences can be quite weak. Similarly, when $\gamma < n-1$, direct reciprocity preferences remain low. For γ, the threshold of $n-1 > 1$ suggests that indirect reciprocity preferences play a very strong role in determining this outcome. Note that the total free-riding equilibrium does not exist when $n = 3$.

The social optimum equilibrium, $(y, ..., y, ..., y)$, is satisfied when the following conditions are met for player J and each player k. For player J, $(n-1)y > \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1+\gamma}{2}y$, or

$$\frac{1}{R_J} < \frac{\alpha^2}{1-\alpha} \frac{\overbrace{(n-1-\gamma)y}^{(i)}}{2}, \tag{2.4.47}$$

where if part $(i) > 0$, or $\gamma < n-1$, then

$$R_J > \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1-\gamma)y}, \tag{2.4.48}$$

or if part $(i) < 0$, or $\gamma > n-1$, then

$$R_J < \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1-\gamma)y}, \tag{2.4.49}$$

which is not possible as R_J cannot be less than zero. For each player k, $(n-2)y + y > \frac{1-\alpha}{\alpha^2 R_k} + \frac{n-1}{2}y$, or

$$R_k > \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1)y}. \tag{2.4.50}$$

The social optimum equilibrium is possible if player J has sufficiently strong direct reciprocity preferences, and cares relatively less about indirect reciprocity (i.e., $\gamma < n-1$). So long as γ is not too great, or in other words, so long as the relatively role of indirect reciprocity is not too strong, player J will "forgive and forget" the history of of the game, $h = (0, ..., 0)$, and contribute their full endowment to the group account.

The interior solution equilibrium, $(c_k^*, ..., c_J^*, ..., c_k^*)$, is satisfied when the conditions outlined above are met for c_J^* and c_k^* (i.e., conditions (2.4.21) and (2.4.26) for player J and conditions (2.4.35) and (2.4.38) for each player k).

In Figure 2.18, we represent the equilibrium in a graph (α, R_J). Note that this figure differs from 2.18 in Case 1 of the indirect upstream reciprocity model in Section 2.4.2.1. In Figure 2.18, $f_5(\alpha) \equiv \frac{1-\alpha}{\alpha^2} \frac{2}{(n-1-\gamma)y}$. Recall that $\gamma > 0$ and $R_I = \gamma R_J$. In *Area I*, the equilibria are the traditional public

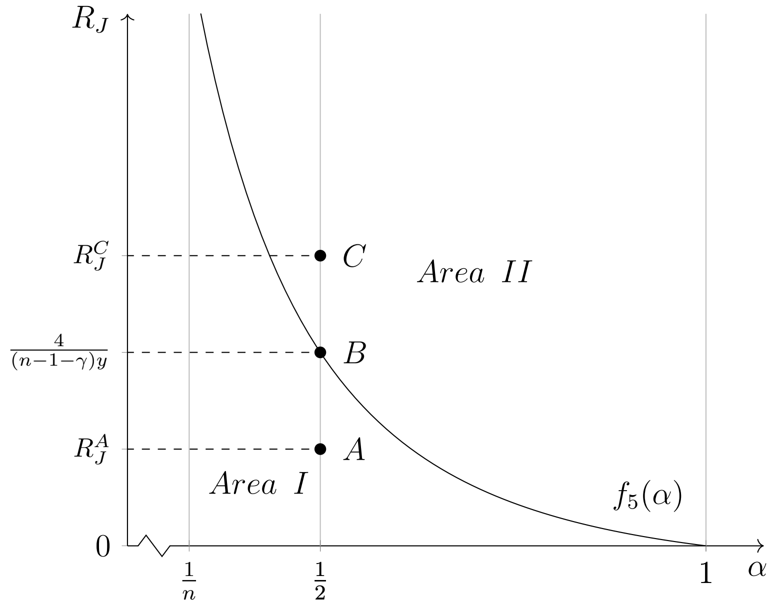

Figure 2.18: Indirect Downstream Reciprocity - Case 1: $h = (0, ..., 0)$

goods game Nash equilibrium, $(0, ..., 0, ..., 0)$, and the total free-riding equilibrium, $(y, ..., 0, ..., y)$, if $\gamma \lesseqgtr n - 1$ and $n > 3$. In *Area II*, the equilibria are the traditional public goods game Nash equilibrium, the total free-riding equilibrium if $\gamma > n - 1$ and $n > 3$, or the social optimum equilibrium, $(y, ..., y, ..., y)$, if $\gamma < n - 1$. Along the curve $f_5(\alpha)$, the interior solution equilibrium, $(c_k^*, ..., c_J^*, ..., c_k^*)$, also exists.

Let $\alpha = \frac{1}{2}$ such that $f_5(\frac{1}{2}) = \frac{4}{(n-1-\gamma)y}$. If player J's reciprocity preference is less than $f_5(\alpha)$, i.e., $R_J^A < \frac{4}{(n-1-\gamma)y}$, then player J is in *Area I*. If player J's reciprocity preference is equal to $f_5(\alpha)$, i.e., $R_J^B = \frac{4}{(n-1-\gamma)y}$, then player J is along the curve $f_5(\alpha)$. If player J's reciprocity preference is greater than $f_5(\alpha)$, i.e., $R_J^C > \frac{4}{(n-1-\gamma)y}$, then player J is in *Area II*.

Note that *Area II* showcases a tipping point in the threshold for γ, the relative strength of direct versus indirect reciprocity. If $\gamma < n - 1$, then the social optimum is the equilibrium. If $\gamma > n - 1$, then the total free-riding equilibrium prevails. In other words, if indirect reciprocity preferences are much stronger than direct reciprocity preferences, then we are at the total free-riding equilibrium, and the history of the game, $h = (0, ..., 0)$, has a strong impact on the outcome of the game. If indirect reciprocity preferences are slightly weaker (i.e., indirect reciprocity preferences may still be greater than direct reciprocity preferences, but are decreasing compared to the previous case), then the history of the game has less of a bearing on the game's outcome and the social optimum equilibrium results.

2.4.2.1.1 Comparative Statics We examine two potential equilibria that occur in *Area II* in Figure 2.18, where $R_J > f_5(\alpha)$, and illustrate the role of the indirect reciprocity preference parameter, γ, in determining the switching point between free-riding and cooperative behaviour. We use Mathematica (Wolfram Research, Inc, 2020), a technical computing system, to visualize the data. Mathematica code for reproduction of the interactive visualization is available in Appendix A.1.3. Static snapshots and descriptions with parameter values are presented below.

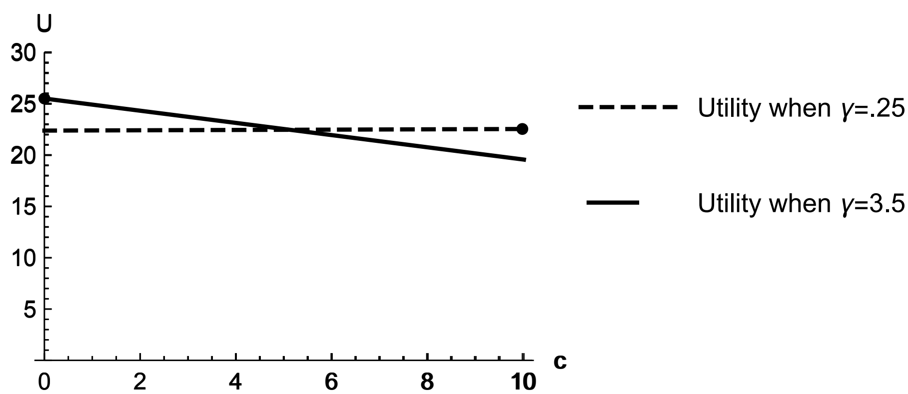

Figure 2.19: Indirect Downstream Reciprocity - Case 1. Player J's utility when $R = 0.15$.

In Figure 2.19, we represent player J's maximized utility as a function of contribution choice, $U(c)$, where U represents player J's utility level and c represents player J's contribution choice. When the history of the game is $h = (0, ..., 0)$, $n = 4$, $y = 10$, $\alpha = \frac{1}{2}$, player J believes that the other players are contributing fully (i.e., $b_{J,k} = y$), and player J's direct reciprocity preference parameter takes the value $R_J = 0.15$, then the social optimum equilibrium will result when player J's indirect reciprocity preferences are weak ($\gamma < 0.33$) and player J will free-ride only when their indirect reciprocity preferences are very high ($\gamma > n - 1$). Figure 2.19 represents these two scenarios with $\gamma = 0.25$

and $\gamma = 3.5$, respectively. Note that the condition $R_J > f_5(\alpha)$ is not satisfied for $0.33 < \gamma < n - 1$.

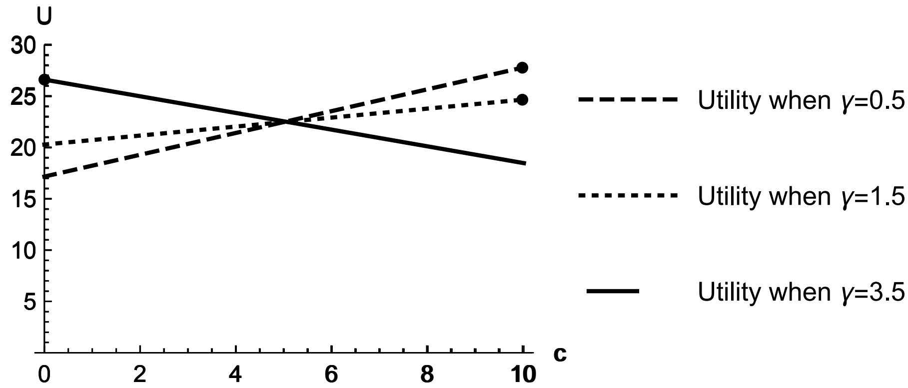

Figure 2.20: Indirect Downstream Reciprocity - Case 1. Player J's utility when $R = 0.5$.

As player J's direct reciprocity preference parameter increases, we observe a greater range of values for $\gamma < n - 1$ which satisfy the condition $R_J > f_5(\alpha)$ for the social optimum equilibrium to occur. In Figure 2.20, when $R_J = 0.5$ and player J's indirect reciprocity preferences are weak, $\gamma < 1$, then player J will contribute fully to the group account. However, when player J's indirect reciprocity preferences are relatively stronger, $1 < \gamma < 2.2$, all else constant, then player J's utility function is still upward sloping and maximized when they are contributing fully, $c_J^* = y$. The condition $R_J > f_5(\alpha)$ is not satisfied for values $2.2 < \gamma < n - 1$. Once player J's indirect reciprocity preference parameter becomes sufficiently strong, $\gamma > n - 1$, then player J's utility function is downward sloping and maximized when $c_J^* = 0$ such that total free-riding is optimal.

Finally, in Figure 2.21, when $R_J = 1$, we observe similar results. When player J's indirect reciprocity preferences are weak, $\gamma < 1$, then they will contribute fully to the group account. As player J's indirect reciprocity preferences become relatively stronger, $1 < \gamma < 2.6$, all else constant, then player J's utility function remains upward sloping and maximized when they are contributing fully, $c_J^* = y$. The social optimum equilibrium occurs for all values of $\gamma < 2.6$ when $R_J = 1$. Once player J's indirect reciprocity preference parameter becomes sufficiently strong, $\gamma > n-1$, then total free-riding becomes optimal.

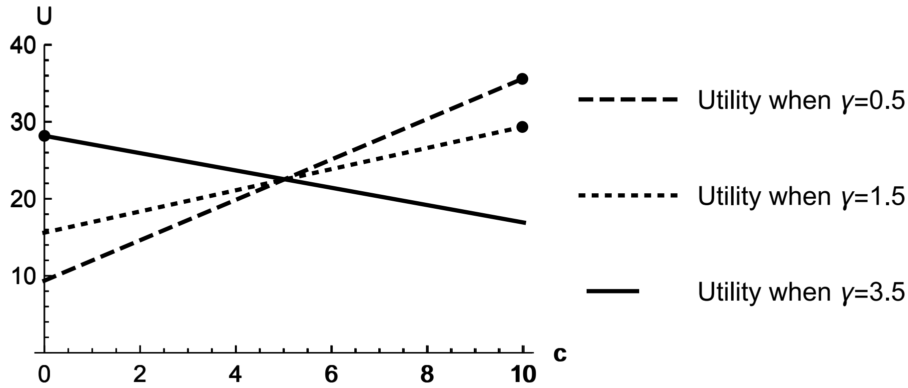

Figure 2.21: Indirect Downstream Reciprocity - Case 1. Player J's utility when $R = 1$.

This shift between non-cooperative free-riding behaviour towards cooperative behaviour can be attributed to the role of indirect reciprocity preferences. When player J has relatively strong indirect reciprocity preferences, $\gamma > n-1$, then total free-riding is the optimal behaviour. When player J's indirect reciprocity preferences are not as strong, $1 < \gamma < n-1$, then full cooperation is optimal, all else equal. When player J does not care as much about indirect reciprocity, $\gamma < 1$, then the history of the game, $h = (0, ..., 0)$ does not influence their contribution choice as much as when player J's indirect reciprocity preferences are very strong, $\gamma > n-1$.

This exercise shows that the reputation-based effects of indirect reciprocity only negatively affect cooperation when indirect reciprocity matters greatly to player J (i.e., when $\gamma > n-1$). Otherwise, even when player J has relatively strong indirect reciprocity preferences (i.e., $1 < \gamma < n-1$), if player J believes that each player $k \in B_J$ is contributing their full endowment, then player J will also contribute fully. Indirect reciprocity preferences only hamper cooperation when these preferences are very strong.

See Appendix A.1.3.1 for Mathematica code to reproduce the figures at the specified parameter values.

2.4.2.2 Case 2

Second, if the history of the game is $h = (y, ..., y)$, then player J has observed players in set A fully contributing, $\frac{1}{n}\sum_{i\in A} c_i = y$. This means that with probability 1, $c_i^* = y \;\forall\; i \in A$, thus player J knows with certainty that $c_1^* = y, ..., c_n^* = y$ (i.e., players in set A are at the social optimum equilibrium). However, because player J is choosing their contribution amount at the same time as each player $k \in B_J$, player J still must form beliefs about the contribution strategy of each player $k \in B_J$, $b_{J,k}(h)$. Player J's best response function is

$$BR_J\left((b_{J,k}(y,...,y))_{k\neq J}\right) = \begin{cases} 0 & \text{if } \sum_{k\in B_J} b_{J,k}(h) < \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1-\gamma}{2}y \\ [0,y] & \text{if } \sum_{k\in B_J} b_{J,k}(h) = \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1-\gamma}{2}y \\ y & \text{if } \sum_{k\in B_J} b_{J,k}(h) > \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1-\gamma}{2}y \end{cases} \tag{2.4.51}$$

Given the best response function of player J and each player k, (2.4.51) and (2.4.16), respectively, we find

$$(n-1)c_k = \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1-\gamma}{2}y, \tag{2.4.52}$$

and

$$(n-2)c_k + c_J = \frac{1-\alpha}{\alpha^2 R_k} + \frac{n-1}{2}y. \tag{2.4.53}$$

Rearranging the terms and solving for c_J, the interior solution for player J is

$$c_J^* = \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_k} - \frac{n-2}{n-1}\frac{1}{R_J}\right) + \frac{n-1+\gamma(n-2)}{2(n-1)}y, \tag{2.4.54}$$

where $R_J \geq 0$, and $R_k \geq 0$. To be an eligible interior solution, $0 \leq c_J^* \leq y$.

We verify that $c_J^* \leq y$ if

$$\frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_k} - \frac{n-2}{n-1}\frac{1}{R_J}\right) + \frac{n-1+\gamma(n-2)}{2(n-1)}y \leq y. \tag{2.4.55}$$

Rearranging the terms, we obtain

$$\frac{1}{R_J} \geq \frac{n-1}{n-2}\left(\frac{1}{R_k} - \frac{\alpha^2}{1-\alpha}\frac{(n-1-\gamma(n-2))y}{2(n-1)}\right), \tag{2.4.55a}$$

or

$$\frac{1}{R_J} \geq \left(\frac{n-1}{n-2}\right)\frac{\overbrace{1-\frac{\alpha^2}{1-\alpha}\frac{n-1-\gamma(n-2)}{2(n-1)}yR_k}^{(i)}}{R_k}. \tag{2.4.55b}$$

If part $(i) > 0$, then

$$\frac{1-\alpha}{\alpha^2}\frac{2(n-1)}{y} > R_k \underbrace{(n-1-\gamma(n-2))}_{(ii)}, \tag{2.4.56}$$

where if part $(ii) > 0$, or $\gamma < \frac{n-1}{n-2}$, then

$$0 \leq R_k < \frac{1-\alpha}{\alpha^2}\frac{2(n-1)}{(n-1-\gamma(n-2))y}, \tag{2.4.57}$$

and if part $(ii) < 0$, or $\gamma > \frac{n-1}{n-2}$, then

$$R_k > \frac{1-\alpha}{\alpha^2}\frac{2(n-1)}{(n-1-\gamma(n-2))y}, \tag{2.4.58}$$

where $R_k \geq 0$ is always met. In this case, where part $(i) > 0$, then

$$0 \leq R_J \leq \frac{n-2}{n-1}\left(\frac{R_k}{1-\frac{\alpha^2}{1-\alpha}\left(\frac{n-1-\gamma(n-2)}{2(n-1)}\right)yR_k}\right). \tag{2.4.59}$$

If part $(i) < 0$, then

$$\frac{1-\alpha}{\alpha^2}\frac{2(n-1)}{y} < R_k \underbrace{(n-1-\gamma(n-2))}_{(iii)}, \tag{2.4.60}$$

where if part $(iii) > 0$, or $\gamma < \frac{n-1}{n-2}$, then

$$R_k > \frac{1-\alpha}{\alpha^2}\frac{2(n-1)}{(n-1-\gamma(n-2))y}, \tag{2.4.61}$$

and if part $(iii) < 0$, or $\gamma > \frac{n-1}{n-2}$, then

$$R_k < \frac{1-\alpha}{\alpha^2}\frac{2(n-1)}{(n-1-\gamma(n-2))y}, \tag{2.4.62}$$

which is not possible because R_k cannot be less than zero. When part (i) in (2.4.55b) is less than 0, then

$$R_J > \frac{n-2}{n-1}\left(\frac{R_k}{1-\frac{\alpha^2}{1-\alpha}\left(\frac{n-1-\gamma(n-2)}{2(n-1)}\right)yR_k}\right), \tag{2.4.63}$$

and since part (i) in (2.4.55b) is less than 0, $R_J \geq 0$ is always satisfied. We check that $c_J^* \geq 0$ if

$$\frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_k}-\frac{n-2}{n-1}\frac{1}{R_J}\right)+\frac{n-1+\gamma(n-2)}{2(n-1)}y \geq 0. \tag{2.4.64}$$

Rearranging the terms, we get

$$\frac{1}{R_J} \leq \frac{n-1}{n-2}\left(\frac{1}{R_k}+\frac{\alpha^2}{1-\alpha}\frac{(n-1+\gamma(n-2))y}{2(n-1)}\right), \tag{2.4.64a}$$

or

$$R_J \geq \frac{n-2}{n-1}\left(\frac{R_k}{1+\frac{\alpha^2}{1-\alpha}\frac{n-1+\gamma(n-2)}{2(n-1)}yR_k}\right). \tag{2.4.64b}$$

The conditions for $0 \leq c_J^* \leq y$ can be satisfied simultaneously.

Next, we solve for the interior solution, c_k^*. From the best response functions for player J and each player k, (2.4.51) and (2.4.16), respectively, and solving for c_k, we obtain

$$c_k^* = \frac{1-\alpha}{\alpha^2}\frac{1}{(n-1)R_J} + \frac{n-1-\gamma}{2(n-1)}y, \tag{2.4.65}$$

where $R_J \geq 0$. To be an eligible interior solution, $0 \leq c_k^* \leq y$. We verify that $c_k^* \leq y$ if

$$\frac{1-\alpha}{\alpha^2}\frac{1}{(n-1)R_J} + \frac{n-1-\gamma}{2(n-1)}y \leq y. \tag{2.4.66}$$

Rearranging the terms,

$$\frac{1}{R_J} \leq \frac{\alpha^2}{1-\alpha}\frac{(n-1+\gamma)y}{2}, \tag{2.4.66a}$$

or

$$R_J \geq \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1+\gamma)y}. \tag{2.4.66b}$$

We check that $c_k^* \geq 0$ if

$$\frac{1-\alpha}{\alpha^2}\frac{1}{(n-1)R_J} + \frac{(n-1-\gamma)y}{2(n-1)} \geq 0. \tag{2.4.67}$$

Rearranging the terms,

$$\frac{1}{R_J} \geq -\left(\frac{\alpha^2}{1-\alpha} \frac{\overbrace{(n-1-\gamma)y}^{(i)}}{2} \right), \tag{2.4.67a}$$

where if part $(i) > 0$, or $\gamma < n - 1$, then

$$R_J \geq -\left(\frac{1-\alpha}{\alpha^2} \frac{2}{(n-1-\gamma)y} \right), \tag{2.4.68}$$

where $R_J \geq 0$ is always met, and if part $(i) < 0$, or $\gamma > n - 1$, then

$$R_J \leq -\left(\frac{1-\alpha}{\alpha^2} \frac{2}{(n-1-\gamma)y} \right), \tag{2.4.69}$$

where $R_J \geq 0$ is also possible. The conditions for $0 \leq c_k^* \leq y$ can be simultaneously satisfied. However, it is not possible for the interior solutions c_J^* and c_k^* to exist at the same time. This means that the interior solution equilibrium, $(c_k^*, ..., c_J^*, ..., c_k^*)$, is not a potential equilibrium.

When the history of the game is $h = (y, ..., y)$, there are four possible equilibria. This differs from the Case 2 in the indirect upstream reciprocity model in Section 2.3.2.2 which had five possible equilibria. However, this difference can be attributed to the fact that the conditions for the existence of the interior solutions, c_J^* and c_k^*, cannot be satisfied simultaneously in the indirect downstream reciprocity model. The following possible equilibria remain the same as in Section 2.3.2.2.

The traditional public goods game Nash equilibrium, $(0, ..., 0, ..., 0)$, is satisfied if the following conditions are met. For player J, $0 < \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1-\gamma}{2} y$, or

$$\frac{1}{R_J} > -\left(\frac{\alpha^2}{1-\alpha} \frac{\overbrace{(n-1-\gamma)y}^{(i)}}{2} \right), \tag{2.4.70}$$

where if part $(i) > 0$, or $\gamma < n-1$, then

$$R_J > -\left(\frac{1-\alpha}{\alpha^2}\frac{2}{(n-1-\gamma)y}\right), \tag{2.4.71}$$

or if part $(i) < 0$, or $\gamma > n-1$, then

$$R_J < -\left(\frac{1-\alpha}{\alpha^2}\frac{2}{(n-1-\gamma)y}\right), \tag{2.4.72}$$

which is possible if $R_J \geq 0$. For player k, $0 < \frac{1-\alpha}{\alpha^2 R_k} + \frac{n-1}{2}y$. Rearranging the terms, we get

$$\frac{1}{R_k} > -\left(\frac{\alpha^2}{1-\alpha}\frac{n-1}{2}y\right), \tag{2.4.73}$$

or

$$R_k > -\left(\frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}\right), \tag{2.4.73a}$$

which is always met as $R_k \geq 0$. The traditional public goods game Nash equilibrium is predicted by standard game theory and holds with or without accounting for reciprocity preferences. Even when player J cares about being indirectly reciprocal, or player J's indirect reciprocity preferences are quite strong, where $\gamma > n-1$, player J will contribute nothing if they have a relatively weak direct reciprocity preference. If player J's indirect reciprocity preference is relatively weaker, $\gamma < n-1$, then player J will contribute nothing to the group account at any value of $R_J \geq 0$.

The altruistic equilibrium, $(0, ..., y, ..., 0)$, is satisfied when the following conditions for player J and each player j are met. For player J, $0 > \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1-\gamma}{2}y$, or

$$\frac{1}{R_J} < -\left(\frac{\alpha^2}{1-\alpha} \frac{\overbrace{(n-1-\gamma)y}^{(i)}}{2} \right). \tag{2.4.74}$$

If part $(i) > 0$, or $\gamma < n-1$, then

$$R_J > -\left(\frac{1-\alpha}{\alpha^2} \frac{2}{(n-1-\gamma)y} \right), \tag{2.4.75}$$

which is always true since $R_J \geq 0$. If part $(i) < 0$, or $\gamma > n-1$, then

$$R_J < -\left(\frac{1-\alpha}{\alpha^2} \frac{2}{(n-1-\gamma)y} \right), \tag{2.4.76}$$

which is also possible, so long as $R_J \geq 0$. For each player k, $y < \frac{1-\alpha}{\alpha^2 R_k} + \frac{n-1}{2}y$, or

$$\frac{1}{R_k} > -\left(\frac{\alpha^2}{1-\alpha} \frac{n-3}{2} y \right). \tag{2.4.77}$$

If $n = 3$, then $\frac{1}{R_k} > 0$ is possible. If $n > 3$, then

$$R_k > -\left(\frac{1-\alpha}{\alpha^2} \frac{2}{(n-3)y} \right), \tag{2.4.78}$$

which is always met as $R_k \geq 0$. The conditions on R_J present a somewhat counter-intuitive result. If player J has relatively weaker indirect reciprocity preferences (i.e., $\gamma < n-1$), then any value of $R_J \geq 0$ can result in player J choosing to contribute their full endowment. However, if player J has relatively stronger indirect reciprocity preferences (i.e., $\gamma > n-1$), then player J will contribute their full endowment at lower levels of R_J. Note that in the former case, R_J can take any value, so long as $R_J \geq 0$, but in the latter, R_J is bounded from above. As well, the threshold value for γ is greater than one which implies that indirect reciprocity preferences are stronger than direct reciprocity preferences for at least the latter case and for some range in the first case. This finding is counter-intuitive because, as when the history of the game is $h = (0, ..., 0)$, we expect that player J would not be "kind" to their own group, especially when player J believes that their own group will be "unkind" to them as well. However, in this case, when the history of the game is $h = (y, ..., y)$, we find that player J will be "kind" to their own group depending on the strength of their indirect reciprocity preferences despite the other players in their group not contributing.

The total free-riding equilibrium, $(y, ..., 0, ..., y)$, is satisfied if for player J, $(n-1)y < \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1-\gamma}{2}y$, or

$$0 \leq R_J < \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1+\gamma)y}, \tag{2.4.79}$$

and for player k, $(n-2)y > \frac{1-\alpha}{\alpha^2 R_k} + \frac{n-1}{2}y$, or

$$\frac{1}{R_k} < \frac{\alpha^2}{1-\alpha}\frac{n-3}{2}y. \tag{2.4.80}$$

If $n = 3$, then $\frac{1}{R_k} < 0$, which is not possible since we restrict $R_k \geq 0$. If $n > 3$, then

$$R_k > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-3)y}. \tag{2.4.81}$$

In this equilibrium, player J will free-ride if their direct reciprocity preferences are weak enough, irrespective of their relative indirect reciprocity preferences. As γ gets larger, this threshold decreases to lower levels of R_J, which suggests that stronger indirect reciprocity preferences may push player J to one of the other different and more cooperative equilibrium.

The social optimum equilibrium, $(y, ..., y, ..., y)$, is satisfied if for player J, $(n-1)y > \frac{1-\alpha}{\alpha^2 R_J} + \frac{n-1-\gamma}{2}y$, or

$$R_J > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1+\gamma)y}, \tag{2.4.82}$$

and for player k, $(n-2)y + y > \frac{1-\alpha}{\alpha^2 R_k} + \frac{n-1}{2}y$, or

$$R_k > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}. \tag{2.4.83}$$

The social optimum equilibrium results when both player J and player k have sufficiently strong direct reciprocity preferences. Player J's threshold for contributing their full endowment decreases as γ, or the relative importance of player J's indirect reciprocity preference, increases. This suggests that the more player J cares about indirect reciprocity, the lower the threshold is for player J to contribute their full endowment to the group account (i.e., it is easier to contribute the full endowment).

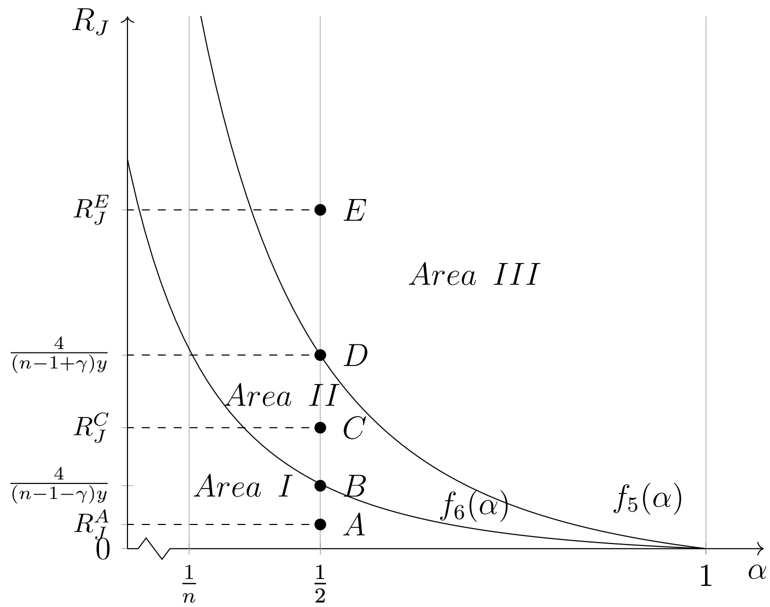

Figure 2.22: Indirect Downstream Reciprocity - Case 2: $h = (y, ..., y)$

In Figure 2.22, we represent the equilibria in a graph (α, R_J). In Figure 2.22, $f_5(\alpha) \equiv \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1-\gamma)y}$ and $f_6(\alpha) \equiv \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1+\gamma)y}$. In *Area I*, the equilibria are the traditional public goods game Nash equilibrium, $(0, ..., 0, ..., 0)$, when $\gamma \lesseqgtr n-1$, the altruistic equilibrium, $(0, ..., y, ..., 0)$, when $\gamma \lesseqgtr n-1$, or the total free-riding equilibrium, $(y, ..., 0, ..., y)$. In *Area II*, the equilibria are the traditional public goods game Nash equilibrium when $\gamma \lesseqgtr n-1$, the altruistic equilibrium when $\gamma \lesseqgtr n-1$, or the social optimum equilibrium, $(y, ..., y, ..., y)$. In *Area III*, the equilibria are the traditional public goods game Nash equilibrium when $\gamma < n-1$, the altruistic equilibrium when $\gamma < n-1$, or the social optimum equilibrium.

Let $\alpha = \frac{1}{2}$ such that $f_5(\frac{1}{2}) = \frac{4}{(n-1-\gamma)y}$ and $f_6(\frac{1}{2}) = \frac{4}{(n-1+\gamma)y}$. If player J's reciprocity preference is less than $f_6(\alpha)$, i.e., $R_J^A < \frac{4}{(n-1+\gamma)y}$, then player J is in *Area I*. If player J's reciprocity preference is between $f_6(\alpha)$ and $f_5(\alpha)$, i.e., $\frac{4}{(n-1+\gamma)y} < R_J^C < \frac{4}{(n-1-\gamma)y}$, then player J is in *Area II*. If player J's reciprocity preference is greater than $f_5(\alpha)$, i.e., $R_J^E > \frac{4}{(n-1-\gamma)y}$, then player J is in *Area III*. To illustrate, if player J has very strong reciprocity preferences, R_J^E, but also has relatively greater indirect reciprocity preferences, $\gamma > n-1$, then the social optimum equilibrium will result.

2.4.2.2.1 Comparative Statics We examine two potential equilibria that occur in *Area III* in Figure 2.22, where $R_J > f_5(\alpha)$, and illustrate the role of the indirect reciprocity preference parameter, γ, in determining the switching point between altruistic and non-cooperative behaviour. In *Area III* in Figure 2.22, when player J believes that each player $k \in B_J$ is not contributing, $b_{J,k} = 0$, then there are two potential equilibria: the traditional public goods game Nash equilibrium and the altruistic equilibrium. Both equilibria are satisfied if $R_J > -f_5(\alpha)$ when $\gamma < n-1$, and if $R_J < -f_5(\alpha)$ when $\gamma > n-1$ (see conditions (2.4.71) and (2.4.72) for the traditional public goods game Nash equilibrium and (2.4.75) and (2.4.76) for the altruistic equilibrium). We use Mathematica (Wolfram Research, Inc, 2020), a technical computing system, to visualize the data. Mathematica code for reproduction of the interactive visualization is available in Appendix A.1.3. Static snapshots and descriptions with parameter values are presented below.

In Figure 2.23, we represent player J's maximized utility as a function of contribution choice, $U(c)$, where U represents player J's utility level and c represents player J's contribution choice. When the history of the game is $h = (y, ..., y)$, $n = 3$, $y = 10$, $\alpha = \frac{1}{2}$, player J believes that the other players are contributing nothing (i.e., $b_{J,k} = 0$), and player J has relatively weak direct

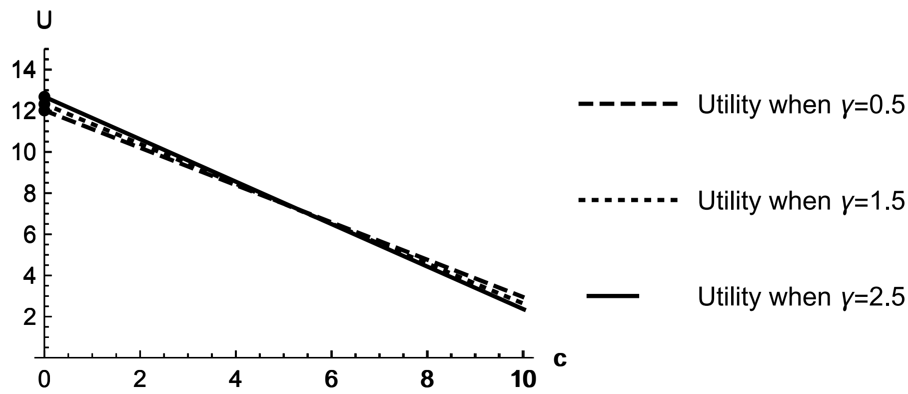

Figure 2.23: Indirect Downstream Reciprocity - Case 2. Player J's utility when $R = 0.15$.

reciprocity preferences, then not contributing to the group account is optimal for all values of $\gamma > 0$. Player J's utility function is downward sloping and maximized when player J is contributing nothing to the group account, $c_J^* = 0$. The traditional public goods game Nash equilibrium is optimal.

The traditional public goods game Nash equilibrium and the altruistic equilibrium are both potential equilibrium in *Area III* of Figure 2.22 where $R_J > f_5(\alpha)$ and when $b_{J,k} = 0$ and $\gamma < n - 1$. Yet comparative statics in Mathematica show that the traditional public goods game Nash equilibrium is always optimal for all values of γ, where $\gamma > 0$, compared to the other potential equilibria . Therefore, even when indirect reciprocity matters greatly to player J and player J has observed the history of the game where $c_i = y$, if player J believes that each player $k \in B_J$ will not contribute, $b_{J,k} = 0$, then player J will not contribute to the group account as well. The altruistic equilibrium does not present in the Mathematica data visualization for any parameter values that satisfy the conditions for the equilibrium (i.e., condition (2.4.75)).

To further illustrate this, the following figures depict player J's utility function at different values of R_J and γ. In Figure 2.24, we represent player J's utility function with low direct reciprocity preferences, $R_J = 0.5$, all else equal. Player J's utility function is maximized when $c_J^* = 0$ and the traditional public goods game equilibrium remains optimal for all values of γ. Figure 2.25 presents player J's utility function with strong direct reciprocity preferences,

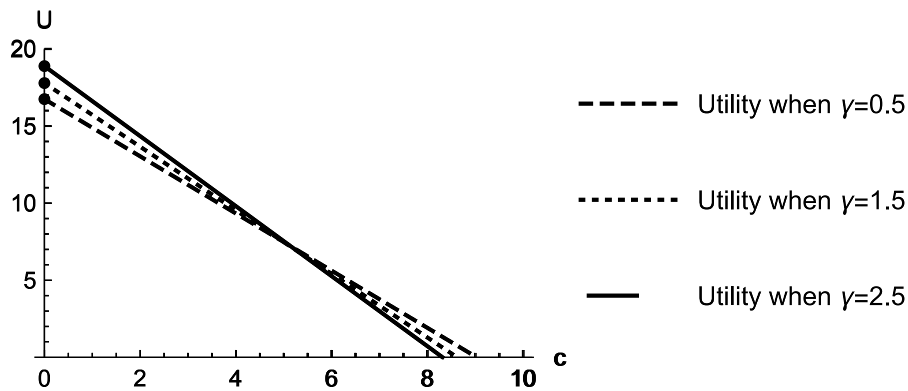

Figure 2.24: Indirect Downstream Reciprocity - Case 2. Player J's utility when $R_J = 0.5$.

$R_J = 1$, all else equal. Again the traditional public goods game Nash equilibrium is optimal and player J's utility is maximized when they are contributing nothing. Not contributing remains optimal. Note that player J's overall utility level is increasing, all else equal, as R_J increases from $R_J = 0.15$ to $R_J = 1$.

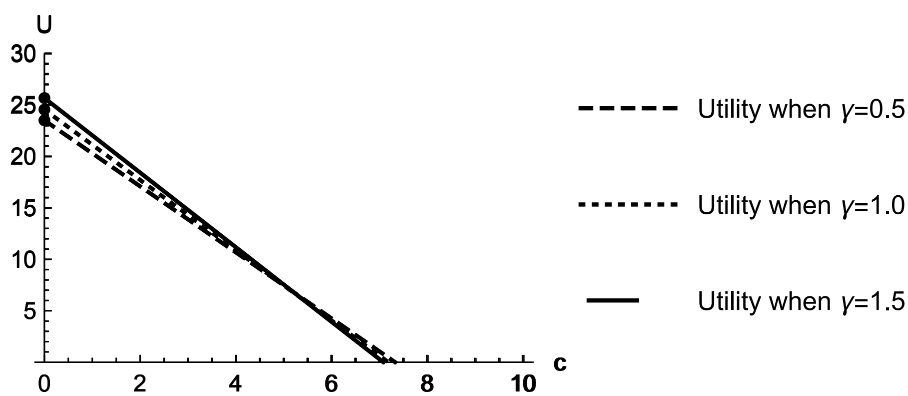

Figure 2.25: Indirect Downstream Reciprocity - Case 2. Player J's utility when $R_J = 1$.

When the history of the game is $h = (y, ..., y)$ and $R_J > f_5(\alpha)$ in Figure 2.22, and player J believes that each player $k \in B_J$ will contribute nothing, $b_{J,k} = 0$, then the traditional public goods game Nash equilibrium is optimal when $\gamma > 0$. In this case, reputation does not affect player J's contribution choice. If player J has observed each player $i \in A$ fully contribute, then when player J believes that each player $k \in B_J$, where $B_J \subset B$, will not contribute to the group account, player J will also not contribute, even when player J

has relatively strong indirect reciprocity preferences, $\gamma > 1$. Direct reciprocity preferences and beliefs about players' contribution strategies have a greater effect on player J's contribution choice than indirect reciprocity preferences in this case.

See Appendix A.1.3.2 for Mathematica code to reproduce the figures at the specified parameter values.

2.4.2.3 Case 3

Finally, if the history of the game is $h = (c, ..., c)$, then player J has observed $\frac{1}{n}\sum_{i \in A} c_i = c \in]0, y[$. This means that player J does not know with certainty the contribution amount of each player $i \in A$. As before, while $h = (c, ..., c)$ represents a multiplicity of interior contributions, in this case we examine the special case of the interior solution equilibrium is the history of the game, $h = (c^*, ..., c^*)$. Player J is choosing their contribution amount at the same time as each player $k \in B_J$ and player J must form beliefs about the contribution strategy of each player $k \in B_J$. From the baseline game, we know that $c_i^* = \frac{1-\alpha}{\alpha^2}\frac{1}{(n-1)R} + \frac{1}{2}y \;\; \forall \;\, i \in A$, and $0 \leq c_i^* \leq y$ if $R \geq \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}$. Because we assume that players are symmetric in the baseline case (i.e., each player $i \in A$) and in set B_J (i.e., each player $k \in B_J$ where $B_J \subset B$), we assume $R_i = R_k = R$. Player J's best response function is

$$BR_J\left((b_{J,k}(c,...,c))_{k\neq J}\right)=\begin{cases}0 & \text{if } \sum_{k\in B_J} b_{J,k}(h) < \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_J}-\frac{\gamma}{n-1}\frac{1}{R}\right) + \frac{n-1}{2}y \\ [0,y] & \text{if } \sum_{k\in B_J} b_{J,k}(h) = \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_J}-\frac{\gamma}{n-1}\frac{1}{R}\right) + \frac{n-1}{2}y \\ y & \text{if } \sum_{k\in B_J} b_{J,k}(h) > \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_J}-\frac{\gamma}{n-1}\frac{1}{R}\right) + \frac{n-1}{2}y\end{cases} \tag{2.4.84}$$

From the best response functions for player J and each player k, (2.4.84) and (2.4.16), respectively

$$(n-1)c_k = \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_J}-\frac{\gamma}{n-1}\frac{1}{R}\right)+\frac{n-1}{2}y, \tag{2.4.85}$$

and

$$(n-2)c_k + c_J = \frac{1-\alpha}{\alpha^2 R}+\frac{n-1}{2}y. \tag{2.4.86}$$

From these expressions, solving for c_J, the interior solution is

$$c_J^* = \frac{1-\alpha}{\alpha^2}\left(\frac{(n-1)^2+\gamma(n-2)}{(n-1)^2}\frac{1}{R}-\frac{n-2}{n-1}\frac{1}{R_J}\right)+\frac{1}{2}y, \tag{2.4.87}$$

where $R_J \geq 0$, and $R \geq 0$. To be an eligible interior solution, $0 \leq c_J^* \leq y$. We verify that $c_J^* \leq y$ if

$$\frac{1-\alpha}{\alpha^2}\left(\frac{(n-1)^2+\gamma(n-2)}{(n-1)^2}\frac{1}{R}-\frac{n-2}{n-1}\frac{1}{R_J}\right)+\frac{1}{2}y \leq y. \tag{2.4.88}$$

Rearranging the terms, we obtain

$$\frac{1}{R_J} \geq \frac{(n-1)^2 + \gamma(n-2)}{(n-1)(n-2)} \frac{1}{R} - \frac{\alpha^2}{1-\alpha} \frac{n-1}{2(n-2)} y, \tag{2.4.88a}$$

or

$$\frac{1}{R_J} \geq \frac{\overbrace{\frac{(n-1)^2+\gamma(n-2)}{(n-1)(n-2)} - \frac{\alpha^2}{1-\alpha}\frac{n-1}{2(n-2)} yR}^{(i)}}{R}. \tag{2.4.88b}$$

If part $(i) > 0$, then

$$0 \leq R < \frac{1-\alpha}{\alpha^2} \frac{2(n-1)^2 + 2\gamma(n-2)}{(n-1)^2 y}, \tag{2.4.89}$$

and

$$0 \leq R_J \leq \frac{R}{\frac{(n-1)^2+\gamma(n-2)}{(n-1)(n-2)} - \frac{\alpha^2}{1-\alpha}\frac{n-1}{2(n-2)} yR}. \tag{2.4.90}$$

If part $(i) < 0$, then

$$R > \frac{1-\alpha}{\alpha^2} \frac{2(n-1)^2 + 2\gamma(n-2)}{(n-1)^2 y}, \tag{2.4.91}$$

and

$$R_J \geq \frac{R}{\frac{(n-1)^2+\gamma(n-2)}{(n-1)(n-2)} - \frac{\alpha^2}{1-\alpha}\frac{n-1}{2(n-2)} yR}, \tag{2.4.92}$$

which is always met as $R_J \geq 0$. We check that $c_J^* \geq 0$ if

$$\frac{1-\alpha}{\alpha^2} \left(\frac{(n-1)^2 + \gamma(n-2)}{(n-1)^2} \frac{1}{R} - \frac{n-2}{n-1} \frac{1}{R_J} \right) + \frac{1}{2} y \geq 0. \tag{2.4.93}$$

Rearranging the terms, we obtain

$$\frac{1}{R_J} \leq \frac{(n-1)^2+\gamma(n-2)}{(n-1)(n-2)}\frac{1}{R} + \frac{\alpha^2}{1-\alpha}\frac{n-1}{2(n-2)}y, \tag{2.4.93a}$$

or

$$R_J \geq \frac{R}{\frac{(n-1)^2+\gamma(n-2)}{(n-1)(n-2)} + \frac{\alpha^2}{1-\alpha}\frac{n-1}{2(n-2)}yR}. \tag{2.4.93b}$$

The condition $0 \leq c_J^* \leq y$ can be satisfied.

Solving for the interior solution, c_k, we find

$$c_k^* = \frac{1-\alpha}{\alpha^2}\left(\frac{1}{n-1}\frac{1}{R_J} - \frac{\gamma}{(n-1)^2}\frac{1}{R}\right) + \frac{1}{2}y, \tag{2.4.94}$$

where $R_J \geq 0$ and $R \geq 0$. For c_k^* to be an eligible interior solution, $0 \leq c_k^* \leq y$. We verify that $c_k^* \leq y$ if

$$\frac{1-\alpha}{\alpha^2}\left(\frac{1}{n-1}\frac{1}{R_J} - \frac{\gamma}{(n-1)^2}\frac{1}{R}\right) + \frac{1}{2}y \leq y. \tag{2.4.95}$$

Rearranging the terms, we obtain

$$\frac{1}{R} \geq \frac{n-1}{\gamma}\frac{1}{R_J} - \frac{\alpha^2}{1-\alpha}\frac{(n-1)^2}{2\gamma}y, \tag{2.4.95a}$$

or

$$\frac{1}{R} \geq \frac{\overbrace{\frac{n-1}{\gamma} - \frac{\alpha^2}{1-\alpha}\frac{(n-1)^2}{2\gamma}yR_J}^{(i)}}{R_J}. \tag{2.4.95b}$$

If part $(i) > 0$, then

$$0 \leq R_J < \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.4.96}$$

and

$$0 \leq R \leq \frac{\gamma}{n-1}R_J - \frac{1-\alpha}{\alpha^2}\frac{2\gamma}{(n-1)^2 y}, \tag{2.4.97}$$

or

$$0 \leq R \leq \frac{R_J}{\frac{n-1}{\gamma} - \frac{\alpha^2}{1-\alpha}\frac{(n-1)^2}{2\gamma}yR_J}. \tag{2.4.98}$$

If part $(i) < 0$, then

$$R_J > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.4.99}$$

and

$$R \geq \frac{\gamma}{n-1}R_J - \frac{1-\alpha}{\alpha^2}\frac{2\gamma}{(n-1)^2 y}, \tag{2.4.100}$$

or

$$R \geq \frac{R_J}{\frac{n-1}{\gamma} - \frac{\alpha^2}{1-\alpha}\frac{(n-1)^2}{2\gamma}yR_J}, \tag{2.4.101}$$

and since $R \geq 0$, this condition is always met. We check that $c_k^* \geq 0$ if

$$\frac{1-\alpha}{\alpha^2}\left(\frac{1}{n-1}\frac{1}{R_J} - \frac{\gamma}{(n-1)^2}\frac{1}{R}\right) + \frac{1}{2}y \leq 0. \tag{2.4.102}$$

Rearranging the terms, we obtain

$$\frac{1}{R} \leq \frac{n-1}{\gamma}\frac{1}{R_J} + \frac{\alpha^2}{1-\alpha}\frac{(n-1)^2}{2\gamma}y, \tag{2.4.102a}$$

or

$$R \geq \frac{\gamma}{n-1} R_J + \frac{1-\alpha}{\alpha^2} \frac{2\gamma}{(n-1)^2 y}. \tag{2.4.102b}$$

However, the conditions for $0 \leq c_k^* \leq y$ cannot be simultaneously satisfied. Therefore, the interior solution, c_k^*, is not possible.

When the history of the game is $h = (c, ..., c)$, there are five possible equilibria. This differs from the Case 3 in the indirect upstream reciprocity model in Section 2.3.2.3 which had seven possible equilibria. However, this difference can be attributed to the fact that the conditions for the existence of the interior solution for player k, c_k^*, are not satisfied in the indirect downstream reciprocity model, and that the conditions required for the "kind" equilibrium, $(0, ..., c_J^*, ..., 0)$, are not satisfied. The following possible equilibria remain the same as in Section 2.3.2.3.

The traditional public goods game Nash equilibrium, $(0, ..., 0, ..., 0)$, is satisfied when the following conditions are met. For player J, $0 < \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_J} - \frac{\gamma}{n-1}\frac{1}{R}\right) + \frac{n-1}{2} y$. Rearranging the terms, we obtain

$$\frac{1}{R_J} > \frac{\gamma}{n-1}\frac{1}{R} - \frac{\alpha^2}{1-\alpha}\left(\frac{n-1}{2}\right) y, \tag{2.4.103}$$

or

$$\frac{1}{R_J} > \frac{\overbrace{\frac{\gamma}{n-1} - \frac{\alpha^2}{1-\alpha}\left(\frac{n-1}{2}\right) yR}^{(i)}}{R}. \tag{2.4.104}$$

If part $(i) > 0$, then

$$0 \leq R < \frac{1-\alpha}{\alpha^2}\frac{2\gamma}{(n-1)^2 y}, \tag{2.4.105}$$

and

$$R_J < \frac{R}{\frac{\gamma}{n-1} - \frac{\alpha^2}{1-\alpha}\left(\frac{n-1}{2}\right) yR}, \tag{2.4.106}$$

or

$$R_J < \frac{n-1}{\gamma} R - \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}. \tag{2.4.107}$$

If part $(i) < 0$, then

$$R > \frac{1-\alpha}{\alpha^2}\frac{2\gamma}{(n-1)^2 y}, \tag{2.4.108}$$

and

$$R_J > \frac{R}{\frac{\gamma}{n-1} - \frac{\alpha^2}{1-\alpha}\left(\frac{n-1}{2}\right) yR}, \tag{2.4.109}$$

or

$$R_J > \frac{n-1}{\gamma} R - \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.4.110}$$

and because $R_J \geq 0$, this condition is always satisfied. For player k, $0 < \frac{1-\alpha}{\alpha^2 R} + \frac{n-1}{2} y$, or

$$R > -\left(\frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}\right), \tag{2.4.111}$$

which is always satisfied as $R \geq 0$. The traditional public goods game Nash equilibrium is possible with or without accounting for reciprocity preferences.

The altruistic equilibrium, $(0, ..., y, ..., 0)$, is satisfied when the following conditions are met for player J and each player k. For player J, $0 > \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_J} - \frac{\gamma}{n-1}\frac{1}{R}\right) + \frac{n-1}{2} y$. Rearranging the terms, we obtain

$$\frac{1}{R_J} < \frac{\gamma}{(n-1)R} - \frac{\alpha^2}{1-\alpha}\left(\frac{n-1}{2}\right)y, \tag{2.4.112}$$

or

$$\frac{1}{R_J} < \frac{\overbrace{\frac{\gamma}{n-1} - \frac{\alpha^2}{1-\alpha}\left(\frac{n-1}{2}\right)yR}^{(i)}}{R}. \tag{2.4.113}$$

If part $(i) > 0$, then

$$0 \leq R < \frac{1-\alpha}{\alpha^2}\frac{2\gamma}{(n-1)^2 y}, \tag{2.4.114}$$

and

$$R_J > \frac{n-1}{\gamma}R - \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.4.115}$$

or

$$R_J > \frac{R}{\frac{\gamma}{n-1} - \frac{\alpha^2}{1-\alpha}\left(\frac{n-1}{2}\right)yR}. \tag{2.4.116}$$

From (2.4.113), if part $(i) < 0$, then

$$R > \frac{1-\alpha}{\alpha^2}\frac{2\gamma}{(n-1)^2 y}, \tag{2.4.117}$$

and

$$0 \leq R_J < \frac{n-1}{\gamma}R - \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.4.118}$$

or

$$0 \leq R_J < \frac{R}{\frac{\gamma}{n-1} - \frac{\alpha^2}{1-\alpha}\left(\frac{n-1}{2}\right) yR}. \tag{2.4.119}$$

However, R_J cannot be negative, so this equilibrium is not possible when $R > \frac{1-\alpha}{\alpha^2}\frac{2\gamma}{(n-1)^2 y}$. For player k, $y < \frac{1-\alpha}{\alpha^2 R} + \frac{n-1}{2}y$, or

$$\frac{1}{R} > -\left(\frac{\alpha^2}{1-\alpha}\frac{(n-3)y}{2}\right). \tag{2.4.120}$$

When $n = 3$, $\frac{1}{R} > 0$ is always met. When $n > 3$,

$$R > -\left(\frac{1-\alpha}{\alpha^2}\frac{2}{(n-3)y}\right), \tag{2.4.121}$$

which is always satisfied as $R \geq 0$. The altruistic equilibrium exists if player J has sufficiently strong reciprocity preferences. As the relative importance of player J's indirect reciprocity preference increases (i.e., as γ increases), then the threshold for R_J decreases. This means that as player J cares more about indirect reciprocity, then the threshold for contributing their full endowment is lowered. This suggests that the history of the game and player J's indirect reciprocity preference play roles in determining player J's cooperative behaviour, even when the players in player J's own set are not contributing to the group account.

The total free-riding equilibrium, $(y, ..., 0, ..., y)$, is satisfied when the following conditions for player J and each player k are met. For player J, $(n-1)y < \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_J} - \frac{\gamma}{n-1}\frac{1}{R}\right) + \frac{n-1}{2}y$. Rearranging the terms, we obtain

$$\frac{1}{R_J} > \frac{\gamma}{(n-1)R} + \frac{\alpha^2}{1-\alpha}\left(\frac{n-1}{2}\right)y, \tag{2.4.122}$$

or

$$0 \leq R_J < \frac{n-1}{\gamma}R + \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}. \tag{2.4.123}$$

For player k, $(n-2)y > \frac{1-\alpha}{\alpha^2 R} + \frac{n-1}{2}y$, or

$$\frac{1}{R} < \frac{\alpha^2}{1-\alpha}\left(\frac{n-3}{2}\right)y. \tag{2.4.124}$$

If $n = 3$, then $\frac{1}{R} < 0$ which is not possible because we restrict $R \geq 0$. If $n > 3$, then

$$R > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-3)y}. \tag{2.4.125}$$

The total free-riding equilibrium exists if player J's direct reciprocity preferences are sufficiently weak and each player k's reciprocity preferences are sufficiently strong. As player J's indirect reciprocity preference increases (i.e., as γ increases), then the threshold on R_J lowers. This means that player J has very weak reciprocity preferences towards players in their own set. While this finding may not be intuitive in isolation, in the context of the other equilibria, it highlights a switching point towards more cooperative outcomes.

The equilibrium with some free-riding, $(y, ..., c_J^*, ..., y)$, is satisfied when for player J, $(n-1)y = \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_J} - \frac{\gamma}{n-1}\frac{1}{R}\right) + \frac{n-1}{2}y$, or

$$R_J = \frac{n-1}{\gamma}R + \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.4.126}$$

and for player k, $(n-2)y + c_J > \frac{1-\alpha}{\alpha^2 R} + \frac{n-1}{2}y$. Rearranging the terms, we find

$$\frac{1}{R} > \frac{n-1}{\gamma R_J} - \frac{\alpha^2}{1-\alpha}\frac{(n-1)^2}{2\gamma}y, \tag{2.4.127}$$

or

$$\frac{1}{R} > \frac{\overbrace{\frac{n-1}{\gamma} - \frac{\alpha^2}{1-\alpha}\frac{(n-1)^2}{2\gamma}yR_J}^{(i)}}{R_J}. \tag{2.4.128}$$

If part $(i) > 0$, then

$$0 \leq R_J < \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}. \tag{2.4.129}$$

From the conditions outlined for player J, this is not possible (even when $R = 0$) as $R_J = \frac{n-1}{\gamma}R + \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}$. From (2.4.128), if part $(i) < 0$, then

$$R_J > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}, \tag{2.4.130}$$

which is possible if $R > 0$. Therefore,

$$R > \frac{\gamma}{n-1}R_J - \frac{1-\alpha}{\alpha^2}\frac{2\gamma}{(n-1)^2 y}, \tag{2.4.131}$$

which is always satisfied as $R \geq 0$. For the equilibrium with some free-riding to be satisfied, $R > 0$ and the conditions on $0 \leq c_J^* \leq y$ must be met (i.e., conditions (2.4.93) and (2.4.95)). As player J's indirect reciprocity preferences increase (i.e., as γ increases), the value of R_J decreases. In other words, when player J cares relatively more about indirect reciprocity, player J does not need to have as strong direct reciprocity preferences in order to contribute ($0 \leq c_J^* \leq y$) to the group account. Indeed, player J contributes less than their full endowment to the group account even when the other players are fully contributing. The strength of player J's indirect reciprocity preferences and the history of the game $h = (c, ..., c)$ may temper player J's contribution choice so that they are contributing less than the other players.

The social optimum equilibrium, $(y, ..., y, ..., y)$, is satisfied when the following conditions for player J and each player k are satisfied. For player J, $(n-1)y > \frac{1-\alpha}{\alpha^2}\left(\frac{1}{R_J} - \frac{\gamma}{n-1}\frac{1}{R}\right) + \frac{n-1}{2}y$. Rearranging the terms, we find

$$\frac{1}{R_J} < \frac{\gamma}{(n-1)R} + \frac{\alpha^2}{1-\alpha}\frac{n-1}{2}y, \tag{2.4.132}$$

or

$$R_J > \frac{n-1}{\gamma}R + \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}. \tag{2.4.133}$$

For player k, $(n-2)y + y > \frac{1-\alpha}{\alpha^2 R} + \frac{n-1}{2}y$. Rearranging the terms, we find

$$\frac{1}{R} < \frac{\alpha^2}{1-\alpha}\frac{n-1}{2}y, \tag{2.4.134}$$

or

$$R > \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}. \tag{2.4.135}$$

The social optimum equilibrium is satisfied when player J and player k have sufficiently strong reciprocity preferences.

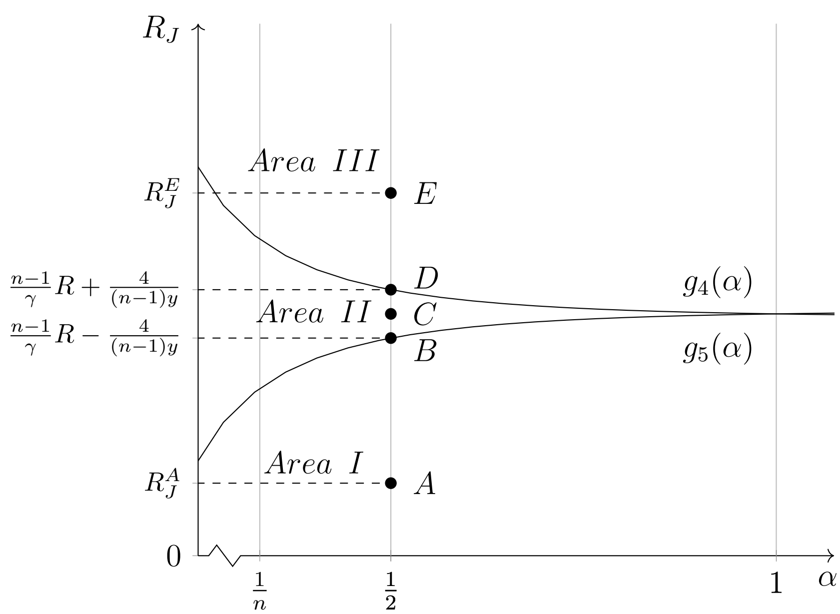

Figure 2.26: Indirect Downstream Reciprocity - Case 3: $h = (c, ..., c)$

In Figure 2.26, we represent the equilibria in a graph (α, R_J). In Figure 2.26, $g_4(\alpha) \equiv \frac{n-1}{\gamma}R + \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}$ and $g_5(\alpha) \equiv \frac{n-1}{\gamma}R - \frac{1-\alpha}{\alpha^2}\frac{2}{(n-1)y}$. In

Area I, the equilibria are the traditional public goods game Nash equilibrium, $(0, ..., 0, ..., 0)$, or the total free-riding equilibrium, $(y, ..., 0, ..., y)$. In *Area II*, the equilibria are the traditional public goods game Nash equilibrium, the total free-riding equilibrium, or the altruistic equilibrium, $(0, ..., y, ..., 0)$. In *Area III*, the equilibria are the traditional public goods game Nash equilibrium, the altruistic equilibrium, or the social optimum equilibrium, $(y, ..., y, ..., y)$. In addition to the traditional public goods game Nash equilibrium, which exists everywhere, the equilibrium with some free-riding, $(y, ..., c_J^*, ..., y)$, exists along $g_4(\alpha)$.

Let $\alpha = \frac{1}{2}$ such that $g_4(\frac{1}{2}) = \frac{n-1}{\gamma}R + \frac{4}{(n-1)y}$ and $g_5(\frac{1}{2}) = \frac{n-1}{\gamma}R - \frac{4}{(n-1)y}$. If player J's reciprocity preference is less than $g_5(\alpha)$, i.e., $R_J^A < \frac{n-1}{\gamma}R - \frac{4}{(n-1)y}$, then player J is at point A in *Area I*. If player J's reciprocity preference is between $g_5(\alpha)$ and $g_4(\alpha)$, i.e., $\frac{n-1}{\gamma}R - \frac{4}{(n-1)y} < R_J^C < \frac{n-1}{\gamma}R + \frac{4}{(n-1)y}$, then player J is at point C in *Area II*. If player J's reciprocity preference is greater than $g_4(\alpha)$, i.e., $R_J^E > \frac{n-1}{\gamma}R + \frac{4}{(n-1)y}$, then player J is at point E in *Area III*. If player J's reciprocity preference is equal to $g_4(\alpha)$, then player J is at point D along $g_4(\alpha)$.

As discussed in Section 2.2.3, comparative statics for $h = (c, ..., c)$ or interior solution equilibria are not presented here. Comparisons at the corner cases provide greater insight into thresholds for cooperative behaviour.

Chapter 3

Conclusion

3.1 Discussion

The inclusion of direct and indirect reciprocity preferences in public goods games can lead to outcomes not predicted by traditional game theory. In the baseline and direct reciprocity models, more cooperative equilibria can be reached if players have sufficiently strong reciprocity preferences, depending on their beliefs and the history of the game. In addition to direct reciprocity preferences, the relative strength of indirect reciprocity preferences also matters. Through the examples provided in the indirect upstream reciprocity model and indirect downstream reciprocity model, recent experience and reputation effects, respectively, can influence behaviour and push players towards more cooperative equilibria. In short, reciprocity preferences matter. In fact, among the different model specifications, reciprocity preferences can influence individual behaviour so profoundly that depending on the history of the game and players' beliefs, these preferences make the difference between contributing nothing (a traditional public goods game Nash equilibrium) and reaching the social optimum equilibrium. Furthermore, there are cases when reciprocity preferences matter more (i.e., the switching point between the total free-riding equilibrium or social optimum equilibrium in Case 1 of both indirect reciprocity models, Section 2.3.2.1.1 and Section 2.4.2.1.1, and the prevalence of the altruistic equilibrium in Case 2 of the indirect upstream reciprocity model, Section 2.3.2.2.1) and cases when they do not matter (i.e., the prevalence of the traditional public goods game Nash equilibrium in Case 2 of the indirect downstream reciprocity model, Section 2.4.2.2.1). Thus, under certain conditions, these preferences, in combination with either recent experiences or reputation, can further support cooperative outcomes and substantially change the outcomes of social dilemma situations.

Table 3.1 presents a summary of the potential equilibria in each history of the game for the different reciprocity model specifications outlined in this thesis. In the baseline model where contribution choices are made simultaneously, a player's own contribution "matches" their beliefs about other players' contributions. Further, the existence of a unique interior equilibrium represents a specific set of conditions for reciprocity preferences and is in itself a novel contribution. In the direct reciprocity model where contribution choices are made sequentially, the second-mover now "matches" their contribution choice to what they observe the first-moving players to contribute; this aligns with the "rewarding" behaviour of sequential reciprocity discussed by Croson (2007). However, in Case 2 when the second-mover has observed the other players fully contributing, the second-mover will totally free-ride if their reciprocity preferences are sufficiently weak but fully cooperate if their reciprocity preferences are strong enough. As shown in the comparative statics for the direct reciprocity model, Section 2.2.3, the traditional public goods game Nash equilibrium is prevalent in Case 1, but in Case 2 there exists a switching point between the social optimum equilibrium and the total free-riding equilibrium depending on the strength of reciprocity preferences.

Comparing the baseline and direct reciprocity models, the role of information arising from sequential play affects the potential equilibria in these two models. The differences between the number of potential equilibria in the baseline model (simultaneous play with contribution strategies dependent on beliefs) and direct reciprocity model (sequential play with contribution strategies dependent on beliefs and observed contributions) arises due to the certainty of information about others' contribution strategies. In sequential play, if the second-mover observes the first-moving players fully contributing, the second-mover will fully cooperate if they have strong reciprocity preferences, but they will totally free-ride if their reciprocity preferences are weak. In the simultaneous baseline model, the player will fully cooperate unconditionally if they believe the other players are also fully cooperating (i.e., free-riding is not optimal, nor an option, irrespective of the strength of reciprocity preferences).

In Table 3.1, both indirect reciprocity models have multiple potential equilibria for each history of the game. In the indirect upstream reciprocity model, the potential equilibria vary depending on the history of the game, players' beliefs about others' contribution strategies, and reciprocity preferences. Recall that relative strength of indirect reciprocity preferences matters. Compared to the baseline model, in Case 1 of the indirect upstream reciprocity model, the player of interest will not contribute if they believe that the other players are not contributing, but if they believe that the other players are fully contributing, then they may also fully cooperate or totally or partially free-ride depending on their reciprocity preferences. Similarly in Case 2 of the indirect upstream reciprocity model, if the player of interest believes that the other players are not contributing to the group account, then the player will either not contribute or fully contribute (i.e., the altruistic[1] equilibrium) depending on the strength of their reciprocity preferences. As shown in the comparative statics, Section 2.3.2.2.1, recent experience and indirect reciprocity preferences drive cooperative behaviour for the player of interest, compared to the baseline model. Further, in Case 3 of the indirect upstream reciprocity model, there are multiple potential equilibria compared to the baseline model. In Case 3, as indirect reciprocity preferences become relatively stronger, cooperative equilibria are easier to attain. The relationship between direct and indirect reciprocity preferences affects contribution behaviour in more complex ways than predicted by traditional game theory - whether the player of interest cooperates or not depends on the history of the game, their beliefs about others' contribution strategies, and their reciprocity preferences.

[1]Recall that "altruistic" behaviour here is not pure altruism as defined by Trivers (1971). Pure altruism occurs when the giver incurs a (small) cost and the receiver benefits. Player J incurs a cost from fully contributing to the group account when others are not in the form of lower material payoff, but player J benefits through their reciprocity payoff function.

Table 3.1: Equilibria Summary Table

		Traditional Public Goods Game $(0, ..., 0, ..., 0)$	Social Optimum $(y, ..., y, ..., y)$	Total Free-Riding $(y, ..., 0, ..., y)$	Altruistic $(0, ..., y, ..., 0)$	Interior $(c, ..., c, ..., c)$	Some Free-Riding $(y, ..., c, ..., y)$	"Kind" $(0, ..., c, ..., 0)$
Baseline	Case #1: $h = (0, ..., 0)$	×						
	Case #2: $h = (y, ..., y)$		×					
	Case #3: $h = (c, ..., c)$					×		
Direct Reciprocity	Case #1: $h = (0, ..., 0)$	×						
	Case #2: $h = (y, ..., y)$		×	×				
	Case #3: $h = (c, ..., c)$					×		
Indirect Upstream Reciprocity	Case #1: $h = (0, ..., 0)$	×	×	×			×	
	Case #2: $h = (y, ..., y)$	×	×	×	×	×		
	Case #3: $h = (c, ..., c)$	×	×	×	×	×	×	×
Indirect Downstream Reciprocity	Case #1: $h = (0, ..., 0)$	×	×	×		×		
	Case #2: $h = (y, ..., y)$	×	×	×	×			
	Case #3: $h = (c, ..., c)$	×	×	×	×		×	

Note: This table represents a summary of results. The "×" symbol indicates a potential equilibrium. In some cases, there are conditions that restrict the existence of certain equilibria. Each column represents a potential equilibrium; the traditional public goods game Nash equilibrium occurs when all players contribute nothing to the group account, the social optimum equilibrium occurs when all players contribute their full endowment to the group account, the total free-riding equilibrium is when one player contributes nothing while all other players are fully contributing, the altruistic equilibrium occurs when on player is fully contributing while the others are not, the interior equilibrium occurs when all players are contributing the unique interior solution contribution, the equilibrium with some free-riding occurs when one player contributes their unique interior solution contribution to the group account while the other players are fully contributing, and the "kind" equilibrium is when the player is contributing their unique interior solution contribution while the others are not contributing anything.

In the indirect downstream reciprocity model, Table 3.1 shows that there are multiple potential equilibria that vary depending on the history of the game, players' beliefs about others' contribution strategies, and reciprocity preferences. As before, compared to the baseline model, in Case 1 of the indirect downstream reciprocity model, the player of interest will not contribute if they believe that the other players are not contributing, but if they believe that the other players are fully contributing, then they may also fully cooperate or totally free-ride depending on their reciprocity preferences. Further, the interior solution equilibrium is also a potential equilibrium. This is similar to the baseline model where if the player believes that the other players are contributing the interior solution contribution amount to the group account, then it is optimal for the player to also contribute the interior solution contribution amount. In Case 2 and Case 3, the potential equilibria are similar to the indirect upstream reciprocity model as well. That is, the potential equilibria differ depending on reciprocity preferences. In Case 2, in addition to "matching" their own contribution with their beliefs about other players' contributions (i.e., the traditional public goods game Nash equilibrium or the social optimum equilibrium), the player of interest may free-ride when others are contributing fully or be reciprocally altruistic[2] when others are not contributing depending on their reciprocity preferences. Similar behaviour may occur in Case 3, however the player of interest may also partially free-ride as a result of "matching" the history of the game that they observed. The downstream indirect reciprocity model presents several potential equilibria which, conditional on indirect reciprocity preferences, may or may not depend on reputation effects.

How individuals respond to the kindness of others can support outcomes not predicted by traditional game theory and that are, in fact, counter-intuitive. In the direct reciprocity model, Section 2.2.3, when the history of the game is $h = (0, ..., 0)$, then not cooperating is optimal, as predicted by traditional game theory. However, when the history of the game is $h = (y, ..., y)$, we

[2]This is not pure altruism as defined by Trivers (1971), but aligns with the definition of reciprocal altruism from Nowak and Sigmund (2005).

see that total free-riding is an optimal strategy for very low values of the reciprocity preference parameter, but fully cooperating is optimal when reciprocity preferences are stronger.

When the history of the game is $h = (0, ..., 0)$, we observe switching points between cooperative and non-cooperative behaviour for both of the indirect reciprocity models. In Case 1 of the indirect upstream reciprocity model, Section 2.3.2.1.1, we show that stronger indirect reciprocity preferences drive non-cooperative behaviour, yet when direct reciprocity preferences are strong and indirect reciprocity preferences are weak, the social optimum equilibrium can result. In Case 1 of the indirect downstream reciprocity model, Section 2.4.2.1.1, we show that non-cooperative behaviour only holds when indirect reciprocity preferences are very strong, otherwise cooperation within the group can be achieved.

On the other hand, when the history of the game is $h = (y, ..., y)$, we see divergent outcomes in the indirect reciprocity models that are counter-intuitive. In Case 2 of the indirect upstream reciprocity model, Section 2.3.2.2.1, the prevalence of the altruistic equilibrium compared to the prevalence of the traditional public goods game Nash equilibrium in the indirect downstream reciprocity model, Section 2.4.2.2.1, suggests that both exposure to kindness and strength of reciprocity preferences are important in driving cooperative behaviour. Relative strength of direct versus indirect reciprocity preference parameters is also a driver of cooperation in these cases. For example, in the indirect upstream reciprocity model specification, recent experience when the history of the game is $h = (y, ..., y)$ leads to more pro-social behaviour than the reputation-based effects of indirect downstream reciprocity.

Further, the identification of interior solutions suggests that there exist scenarios in which corner solutions are not optimal, contrary to traditional theoretical predictions for public goods games. Recall that the interior solution equilibria represent special cases and not a multiplicity of equilibria. However, because the reciprocity preference parameter is exogenous and varies, the interior solution contribution choices that comprise the interior solution equilibrium vary as well. These special cases depend on the reciprocity preference parameters and satisfying the condition $0 \leq c^* \leq y$. For example, in the baseline model, Section 2.1.2, the interior solution, c^* from (2.1.19), suggests that for $R > 0$, $c^* > \frac{1}{2}y$. Therefore, the interior solution is always "kind" when evaluated against the reference points (i.e., part (*ii*) in (2.1.8)). By incorporating reciprocity preferences into players' utility functions, the interior solutions can be achieved.

3.2 Extensions

While the focus of the indirect reciprocity models presented here examines the effect of indirect reciprocity preferences on a player's contribution strategy, future extensions of this model could treat each player with the indirect reciprocity information either randomly or with common knowledge. For example, in the indirect reciprocity models, only one player of interest receives additional information and/or monetary benefit. However, the model could possibly be expanded to explore the effect of indirect reciprocity interactions across all players, either randomly or uniformly, such that all players know that some players will receive a benefit (i.e., information about average contributions and/or a monetary benefit), but players may not know which players or how many players receive the benefit. If all players have indirect reciprocity preferences, cooperative outcomes may be easier to achieve, however, adjustments to assumptions about symmetric players, information, and spillover effects, such as learning, norms, and belief formation, are needed.

Another extension of this model includes incorporating further higher-order beliefs and beliefs about other players' motivations. In particular, third-order beliefs, or what player i believes player j believes that player i believes about player j's contribution strategy, $b_{i,j,i,j}$, involve higher-order rationality (i.e., Kneeland, 2015) and allows for further specifications of perceived kindness. Where the perceived kindness function, $\lambda_{i,j,i}$, evaluates player i's perception of the kindness of player j towards player i, inclusion of further perceived kindness that evaluates player i's perception of player j's perception of player i's kindness towards player j, or in other words, how "kind" player i thinks player j thinks player i is to player j, could further parse out the effects of beliefs and histories on contribution choice.

However, there exists a tradeoff when incorporating odd-numbered higher-order beliefs. For example, the inclusion of a further perceived kindness function with third-order beliefs in player i's reciprocity payoff function means that all kindness functions need to have matching signs for the reciprocity payoff function to be additive to player i's utility. When there is a mismatch in signs, then player i's reciprocity payoff function can present disutility to player i. This could occur if player i is "kind" to player j (kindness), player i perceives that player j is "kind" to player i (perceived kindness), and player i perceives that player j perceives that player i is "unkind" to player j (further perceived kindness). In this case, player i receives disutility from being "kind" to player j, believing that player j is "kind" to player i, but believing that player j thinks player i is "unkind" to player j. However, the assumption of rational expectations ensures that beliefs are correct in equilibrium, such that second-order beliefs are correct, $b_{i,j,i} = c_i$, and first- and third-order beliefs are correlated and correct, $b_{i,j,i,j} = b_{i,j} = c_j$. This assumption would need to be relaxed for kindness and further perceived kindness functions to have mismatched signs. Incorporating further higher-order beliefs, particularly in the indirect downstream reciprocity model, may provide more insight into reputation-based effects and align with work on simple records (Clark et al., 2020), image-scoring (Nowak and Sigmund, 1998), and higher-order rationality (Kneeland, 2015).

Further extensions of the model could include adjustments to various assumptions. For example, relaxing assumptions on player symmetry would allow for asymmetric players and heterogeneous reciprocity preferences, similar to selfish and reciprocal player types in Ambrus and Pathak (2011). As well, allowing for the decay of kindness and perceived kindness through a discount factor on the kindness and perceived kindness functions as higher-order beliefs are incorporated could also be considered for future work.

The theoretical predictions derived from the reciprocity models presented in this thesis can be tested with economic experiments and provide further insight into voluntary individual contributions in public goods games. The explicit modelling of direct reciprocity, indirect upstream reciprocity, and indirect downstream reciprocity presented in this thesis provides foundations for an experimental design to test these predictions and analyze observed behaviour in public goods games experiments.

3.3 Final Thoughts

This research provides insight into cooperative and non-cooperative behaviour in collective action social dilemmas. While presented generally, these findings are particularly relevant in a climate change context. Climate change is global public good that requires coordination and cooperation to address. There exists incentives to free-ride in global environmental policy and pollution reduction efforts (Silva and Zhu, 2009), yet socially efficient outcomes regarding greenhouse gas emissions can be reached if countries behave reciprocally altruistic and voluntarily implement socially optimal allocations (Caplan et al., 1999). The reciprocity models that we present in this thesis suggest that there are certain conditions under which cooperative outcomes can be achieved. Accounting for direct and indirect reciprocity preferences can drive these cooperative outcomes in public goods games and have applications in global environmental policy. In particular, accounting for recent experience or reputation-based effects can identify the conditions under which cooperation can be reached in more complex three-party interactions.

For example, suppose that a country has directly benefited from the collective actions of a group of countries in mitigating climate change (i.e., pollution abatement). Suppose that this country is then coordinating with a group of countries (i.e., geographic neighbours or shared major industry) on climate change action as well. A recent experience may influence the country to "pay-it-forward" within their group of countries and push the group towards cooperative outcomes, including the interior solution equilibrium and the social optimum equilibrium. More cooperative outcomes on climate change action are welfare enhancing and globally beneficial.

Accounting for reciprocity preferences can address chronic under-provisioning of public goods, such as climate change abatement. Direct and indirect reciprocity preferences can influence cooperative outcomes in public goods games and may, in fact, reduce incentives to free-ride. The relative strength of direct versus indirect reciprocity preferences is also a contributing factor in reaching these outcomes, as are the history of the game and players' first- and higher-order beliefs. We explicitly model direct and indirect reciprocity and derive theoretical predictions that support the inclusion of reciprocity preferences in players' utility functions to reach cooperative outcomes.

Bibliography

Ambrus, A. and Pathak, P. A. 2011. Cooperation over finite horizons. *Journal of Public Economics* 95 (7/8): 500–512.

Axelrod, R. and Hamilton, W. D. 1981. The evolution of cooperation. *Science* (211).

Bagnoli, M. and McKee, M. 1991. Voluntary contribution games: efficient private provision of public goods. *Economic Inquiry* 29: 351–366.

Bardsley, N. and Sausgruber, R. 2005. Conformity and reciprocity in public good provision. *Journal of Economic Psychology* 26 (5): 664–681.

Bolton, G. E. and Ockenfels, A. 2000. Erc: A theory of equity, reciprocity, and competition. *American Economic Review* 90 (1): 166–193.

Caplan, A. J., Ellis, C. J., and Silva, E. C. 1999. Winners and losers in a world with global warming: noncooperation, altruism, and social welfare. *Journal of Environmental Economics and Management* 37 (3): 256–271.

Chaudhuri, A. 2011. Sustaining cooperation in laboratory public goods experiments. *Experimental Economics* 14 (1): 47–83.

Cialdini, R. B. 2009. Influence. Boston, Mass. [u.a.], Pearson, A and B, 5th edition.

Clark, D., Fudenberg, D., and Wolitzky, A. 2020. Indirect reciprocity with simple records. *Proceedings of the National Academy of Sciences* 117 (21): 11344–11349.

Croson, R. T. A. 2007. Theories of commitment, altruism and reciprocity: evidence from linear public goods games. *Economic Inquiry* 45 (2): 199–216.

Dufwenberg, M., Gächter, S., and Hennig-Schmidt, H. 2011. The framing of games and the psychology of play. *Games and Economic Behavior* 73 (2): 459–478.

Dufwenberg, M. and Kirchsteiger, G. 2004. A theory of sequential reciprocity. *Games and Economic Behavior* 47 (2): 268–298.

Dufwenberg, M. and Patel, A. 2017. Reciprocity networks and the participation problem. *Games and Economic Behavior* 101: 260–272.

Falk, A. and Fischbacher, U. 2006. A theory of reciprocity. *Games and Economic Behavior* 54 (2): 293–315.

Fehr, E. and Schmidt, K. M. 1999. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics* 114 (3): 817–868.

Geanakoplos, J., Pearce, D., and Stacchetti, E. 1989. Psychological games and sequential rationality. *Games and Economic Behavior* 1 (1): 60–79.

Hardin, G. 1968. The tragedy of the commons. *Science* 162: 1243–1248.

Jiang, L. and Wu, J. 2019. Belief-updating rule and sequential reciprocity. *Games and Economic Behavior* 113: 770–780.

Kneeland, T. 2015. Identifying higher-order rationality. *Econometrica* 83 (5): 2065–2079.

Kreps, D. M. and Wilson, R. 1982. Sequential equilibria. *Econometrica* 50 (4): 863–894.

Kuhn, H. W. 1953. Extensive games and the problem of information. In Kuhn, H. W. and Tucker, A. W. eds. Contributions to the theory of games, Volume II (Annals of Mathematics Studies, 28), Princeton: Princeton University Press.

Ledyard, O. 1995. Public goods: some experimental results. In Kagel, J. and Roth, A. eds. Handbook of Experimental Economics, Princeton: Princeton University Press.

Mas-Collel, A., Whinston, M. D., and Green, J. R. 1995. Microeconomic Theory. Oxford University Press.

Nowak, M. A. and Sigmund, K. 1998. Evolution of indirect reciprocity by image scoring. *Nature* 393 (6685): 573–577.

Nowak, M. A. and Sigmund, K. 2005. Evolution of indirect reciprocity. *Nature (London)* 437 (7063): 1291–1298.

Ostrom, E. 2010. Beyond markets and states: Polycentric governance of complex economic systems. *American Economic Review* 100 (3): 641–72.

Palfrey, T. R. and Rosenthal, H. 1984. Participation and the provision of discrete public goods: a strategic analysis. *Journal of Public Economics* 24 (2): 171–193.

Rabin, M. 1993. Incorporating fairness into game theory and economics. *The American Economic Review* 83 (5): 1281–1302.

Silva, E. C. and Zhu, X. 2009. Emissions trading of global and local pollutants, pollution havens and free riding. *Journal of Environmental Economics and Management* 58 (2): 169–182.

Sugden, R. 1984. Reciprocity: The supply of public goods through voluntary contributions. *The Economic Journal (London)* 94 (376): 772–787.

Teyssier, S. 2012. Inequity and risk aversion in sequential public good games. *Public Choice* 151 (1/2): 91–119.

Trivers, R. L. 1971. The evolution of reciprocal altruism. *The Quarterly Review of Biology* 46 (1): 35–57.

Wolfram Research, Inc. 2020. Mathematica, Version 12.1.0. Champaign, IL.

Appendix A

A.1 Mathematica Code

A.1.1 Direct Reciprocity

Let c represent c_j, b represent $b_{j,i} = c_i$, and R represent R_j.

```
Manipulate[ Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5 y) ((n
- 1) b - 0.5 (n - 1) y)}, c, 0, y,
PlotRange -> {0, 25},
PlotLegends -> {"Utility"},
AxesLabel -> {"c", "U"}, PlotStyle->Black],
{{α, 0.5}, 1/n, 1, Appearance -> "Labeled"},
{{y, 10}, 0, 20, 1, Appearance -> "Labeled"},
{{n, 3}, 3, 6, 1, Appearance -> "Labeled"},
{{b, 0}, 0, y, 1, Appearance -> "Labeled"},
{{R, 0}, 0, ((1 - α)/α^2)*(2/((n - 1)*y)) + 3, Appearance -> "Labeled"},
ControlPlacement -> Left, FrameLabel -> "Direct Reciprocity"]
```

A.1.1.1 Case 1

```
 p1 = DynamicModule[{b = 0, n = 3, R = 0.15, y = 10, α = 0.5`}, Plot[{y
- c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n - 1) b - 0.5` (n
- 1) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[0, 11.875]},
PlotRange -> {0, 25},
PlotLegends -> Placed[{"Utility when R=.15"}, {{1.35, 0.75}, {0.7,
0.5}}],
```

```
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dashed}]]
p2 = DynamicModule[{b = 0, n = 3, R = 0.5, y = 10, α = 0.5`}, Plot[{y
- c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n - 1) b - 0.5` (n
- 1) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 16.25}]},
PlotRange -> {0, 25},
PlotLegends -> Placed[{"Utility when R=0.5"}, {{1.35, 0.5}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dotted}]]
p3 = DynamicModule[{b = 0, n = 3, R = 1, y = 10, α = 0.5`}, Plot[{y
- c + α((n - 1) b + c) + R α^2 (c - 0.5` y) ((n - 1) b - 0.5` (n
- 1) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 22.5}]},
PlotRange -> {0, 25},
PlotLegends -> Placed[{" Utility when R=1.0"}, {{1.35, 0.25}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black}]]
Overlay[{p1, p2, p3}]
```

A.1.1.2 Case 2

```
 p1 = DynamicModule[{b = 10, n = 4, R = 0.05, y = 10, α = 0.5`}, Plot[{y
- c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n - 1) b - 0.5` (n
- 1) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 24.0625}]},
PlotRange -> {0, 40},
PlotLegends -> Placed[{"Utility when R=.05"}, {{1.35, 0.75}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dashed}]]
```

```
p2 = DynamicModule[{b = 10, n = 4, R = 0.5, y = 10, α = 0.5`}, Plot[{y
- c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n - 1) b - 0.5` (n
- 1) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{10, 29.375}]},
PlotRange -> {0, 40},
PlotLegends -> Placed[{"Utility when R=0.5"}, {{1.35, 0.5}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dotted}]]
p3 = DynamicModule[{b = 10, n = 4, R = 1, y = 10, α = 0.5`}, Plot[{y
- c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n - 1) b - 0.5` (n
- 1) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 38.75}]},
PlotRange -> {0, 40},
PlotLegends -> Placed[{" Utility when R=1.0"}, {{1.35, 0.25}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black}]]
Overlay[{p1, p2, p3}]
```

A.1.2 Indirect Upstream Reciprocity

Let c represent c_J, b represent $b_{J,j} = c_j$, d represent c_i, and R represent R_J.

```
Manipulate[ Plot[{y - c + α ((n - 1)*b + c) + α (n*d) + R α^2 (c
- 0.5 y) ((n - 1)*b + γ*n*d - 0.5 (n - 1 + n γ) y)}, {c, 0, y},
PlotRange -> {0, 25},
PlotLegends -> {"Utility"},
AxesLabel -> {"c", "U"},
PlotStyle -> Black]],
{{α, 0.5}, 1/n, 1, Appearance -> "Labeled"},
{{y, 10}, 0, 20, 1, Appearance -> "Labeled"},
{{n, 3}, 3, 6, 1, Appearance -> "Labeled"},
{{b, 0}, 0, y, 1, Appearance -> "Labeled"},
```

```
{{d, 0}, 0, y, 1, Appearance -> "Labeled"},
{{R, 0}, 0, ((1 - α/α^2)*(2/((n - 1)*y)) + 3, Appearance -> "Labeled"},
{{γ, 1}, 0.05, 10, 0.05, Appearance -> "Labeled"},
ControlPlacement -> Left, FrameLabel -> "Indirect Upstream Reciprocity"]
```

A.1.2.1 Case 1

When the history of the game is $h = (0, .., 0)$, player J believes that each player j will fully contribute (i.e., $b_{J,j} = c_j = y$),and player J has low reciprocity preferences, $R_J = 0.15$, then the free-riding equilibrium is optimal.

```
p1 = DynamicModule[{b = 10, d = 0, n = 4, R = 0.15‘, y = 10, α = 0.5‘,
γ = 1‘}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α^2 (c - 0.5‘
y) ((n - 1) b + γ n d - 0.5‘ (n - 1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 25.9375}]},
PlotRange -> {0, 30},
PlotLegends -> Placed[{"Utility when γ=1.0"}, {{1.35, 0.5}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"}, PlotStyle -> {Black, Dashed}]]
p2 = DynamicModule[{b = 10, d = 0, n = 4, R = 0.15‘, y = 10, α = 0.5‘,
γ = 1.5‘}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α2(c−0.5‘y)((n−
1)b+γ n d - 0.5‘ (n - 1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 27.8125}]},
PlotRange -> {0, 30},
PlotLegends -> Placed[{" Utility when γ=1.5"}, {{1.35, 0.25}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> Black]]
Overlay[{p1, p2}]
```

When $R_J = 0.5$, all else equal, then the social optimum results when γ is low, and the free-riding equilibrium results when γ increases.

```
p1 = DynamicModule[{b = 10, d = 0, n = 4, R = 0.5‘, y = 10, α = 0.5‘,
γ = 0.5‘}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α^2 (c -
```

```
0.5‘ y) ((n - 1) b + γ n d - 0.5‘ (n - 1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{10, 23.125}]},
PlotRange -> {0, 40},
PlotLegends -> Placed[{"Utility when γ=0.5"}, {{1.35, 0.75}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dashed}]]
p2 = DynamicModule[{b = 10, d = 0, n = 4, R = 0.5‘, y = 10, α = 0.5‘,
γ = 1‘}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α^2 (c - 0.5‘
y) ((n - 1) b + γ n d - 0.5‘ (n - 1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 28.125}]},
PlotRange -> {0, 40},
PlotLegends -> Placed[{"Utility when γ=1.0"}, {{1.35, 0.5}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dotted}]]
p3 = DynamicModule[{b = 10, d = 0, n = 4, R = 0.5‘, y = 10, α = 0.5‘,
γ = 1.5‘}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α^2(c−0.5‘y)((n−
1)b+γ n d - 0.5‘ (n - 1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 34.375}]},
PlotRange -> {0, 40},
PlotLegends -> Placed[{" Utility when γ=1.5"}, {{1.35, 0.25}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> Black]]
Overlay[{p1, p2, p3}]
```

When $R_J = 1$, all else equal, then the social optimum results when γ is low, and the free-riding equilibrium results when γ increases.

```
p1 = DynamicModule[{b = 10, d = 0, n = 4, R = 1‘, y = 10, α = 0.5‘,
γ = 0.5‘}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α^2 (c -
0.5‘ y) ((n - 1) b + γ n d - 0.5‘ (n - 1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{10, 26.25}]},
```

```
PlotRange -> {0, 45},
PlotLegends -> Placed[{"Utility when γ=0.5"}, {{1.35, 0.75}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dashed}]]
p2 = DynamicModule[{b = 10, d = 0, n = 4, R = 1‘, y = 10, α = 0.5‘,
γ = 1‘}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α^2(c−0.5‘y)((n−
1)b+γ n d - 0.5‘ (n - 1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 31.25}]},
PlotRange -> {0, 45},
PlotLegends -> Placed[{"Utility when γ=1.0"}, {{1.35, 0.5}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dotted}]]
p3 = DynamicModule[{b = 10, d = 0, n = 4, R = 1‘, y = 10, α = 0.5‘,
γ = 1.5‘}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α^2(c−0.5‘y)((n−
1)b+γ n d - 0.5‘ (n - 1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 43.75}]},
PlotRange -> {0, 45},
PlotLegends -> Placed[{" Utility when γ=1.5"}, {{1.35, 0.25}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> Black]]
Overlay[{p1, p2, p3}]
```

A.1.2.2 Case 2

When the history of the game is $h = (y, .., y)$ and player J believes that each player j will contribute nothing (i.e., $b_{J,j} = c_j = 0$), then when $R_J = 0.15$, $R_J > f(\alpha)$ and $\gamma > \frac{n-1}{n}$ are satisfied, the altruistic equilibrium is optimal

```
p1 = DynamicModule[{b = 0, d = 10, n = 3, R = 0.15‘, y = 10, α = 0.5‘,
γ = 1.6‘}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α^2 (c -
```

```
0.5‘ y) ((n - 1) b + γ n d - 0.5‘ (n - 1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{10, 22.625}]},
PlotRange -> {0, 25},
PlotLegends -> Placed[{"Utility when γ=1.6"}, {{1.35, 0.5}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dashed}]]
p2 = DynamicModule[{b = 0, d = 10, n = 3, R = 0.15‘, y = 10, α = 0.5‘,
γ = 2‘}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α^2(c−0.5‘y)((n−
1)b+γ n d - 0.5‘ (n - 1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{10, 23.75}]},
PlotRange -> {0, 25},
PlotLegends -> Placed[{" Utility when γ=2.0"}, {{1.35, 0.25}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> Black]]
Overlay[{p1, p2}]
```

When $R_J = 0.5$ and γ increases, all else constant, then player J's indirect reciprocity preferences become stronger and the altruistic equilibrium results, $(0, ..., y, ..., 0)$.

```
p1 = DynamicModule[{b = 0, d = 10, n = 3, R =
0.5‘, y = 10, α = 0.5‘, γ = 0.95‘}, Plot[{y - c + α ((n - 1) b +
c) + α (n d) + R α^2 (c - 0.5‘ y) ((n - 1) b + γ n d - 0.5‘ (n -
1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{10, 22.65625}]},
PlotRange -> {0, 30},
PlotLegends -> Placed[{"Utility when γ=.95"}, {{1.35, 0.75}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dashed}]]
p2 = DynamicModule[{b = 0, d = 10, n = 3, R = 0.5‘, y = 10, α = 0.5‘,
γ = 1‘}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α^2 (c - 0.5‘
y) ((n - 1) b + γ n d - 0.5‘ (n - 1 + n γ) y)}, {c, 0, y},
```

```
Epilog -> {PointSize[Medium], Point[{10, 23.125}]},
PlotRange -> {0, 30},
PlotLegends -> Placed[{"Utility when γ=1.0"}, {{1.35, 0.5}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dotted}]]
p3 = DynamicModule[{b = 0, d = 10, n = 3, R = 0.5`, y = 10, α = 0.5`,
γ = 1.5`}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α^2 (c -
0.5` y) ((n - 1) b + γ n d - 0.5` (n - 1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{10, 27.8125}]},
PlotRange -> {0, 30},
PlotLegends -> Placed[{" Utility when γ=1.5"}, {{1.35, 0.25}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> Black]]
Overlay[{p1, p2, p3}]
```

When player J has stronger direct reciprocity preferences, $R_J = 1$, the altruisitc equilibrium remains optimal for values of $\gamma > 0.8$.

```
p1 = p1 =
DynamicModule[{b = 0, d = 10, n = 3, R = 1`, y = 10, α = 0.5`, γ =
0.85`}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α^2 (c - 0.5`
y) ((n - 1) b + γ n d - 0.5` (n - 1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{10, 23.4375}]},
PlotRange -> {0, 40},
PlotLegends -> Placed[{"Utility when γ=.85"}, {{1.35, 0.75}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dashed}]]
p2 = DynamicModule[{b = 0, d = 10, n = 3, R = 1`, y = 10, α = 0.5`,
γ = 1`}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α^2 (c - 0.5`
y) ((n - 1) b + γ n d - 0.5` (n - 1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{10, 26.25}]},
PlotRange -> {0, 40},
```

```
PlotLegends -> Placed[{"Utility when γ=1.0"}, {{1.35, 0.5}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dotted}]]
p3 = DynamicModule[{b = 0, d = 10, n = 3, R = 1‘, y = 10, α = 0.5‘,
γ = 1.5‘}, Plot[{y - c + α ((n - 1) b + c) + α (n d) + R α^2 (c -
0.5‘ y) ((n - 1) b + γ n d - 0.5‘ (n - 1 + n γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{10, 35.625}]},
PlotRange -> {0, 40},
PlotLegends -> Placed[{" Utility when γ=1.5"}, {{1.35, 0.25}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> Black]]
Overlay[{p1, p2, p3}]
```

A.1.3 Indirect Downstream Reciprocity

Let c represent c_J, b represent $b_{J,k} = c_k$, d represent c_i, and R represent R_J.

```
Manipulate[ Plot[{y - c + α ((n - 1)*b + c) + R α^2 (c - 0.5 y) ((n
- 1)*b + (γ/n)*d - 0.5 (n - 1 + γ) y)}, {c, 0, y},
PlotRange -> {0, 25},
PlotLegends -> {"Utility"},
AxesLabel -> {"c", "U"},
PlotStyle->Black],
{{α, 0.5}, 1/n, 1, Appearance -> "Labeled"},
{{y, 10}, 0, 20, 1, Appearance -> "Labeled"},
{{n, 3}, 3, 6, 1, Appearance -> "Labeled"},
{{b, 0}, 0, y, 1, Appearance -> "Labeled"},
{{d, 0}, 0, y, 1, Appearance -> "Labeled"},
{{R, 0}, 0, ((1 - α)/α^2)*(2/((n - 1)*y)) + 3, 0.05, Appearance ->
"Labeled"},
{{γ, 1}, 0.05, 10, 0.05, Appearance -> "Labeled"},
```

```
ControlPlacement -> Left, FrameLabel -> "Indirect Downstream Reciprocity"]
```

A.1.3.1 Case 1

When the history of the game is $h = (0, .., 0)$ and player J believes that each player k will fully contribute (i.e., $b_{J,k} = c_k = y$), then when $R_J = 0.15$ and the social optimum can be achieved at low values of γ and free-riding is optimal when $\gamma > n - 1$.

```
p1 = DynamicModule[{b = 10, d = 0, n = 4, R = 0.15‘, y = 10, α = 0.5‘,
γ = 0.25‘}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5‘ y)
((n - 1) b + (γ d)/n - 0.5‘ (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{10, 22.58}]},
PlotRange -> {0, 30},
PlotLegends -> Placed[{"Utility when γ=.25"}, {{1.35, 0.75}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dashed}]]
p2 = DynamicModule[{b = 10, d = 0, n = 4, R = 0.15‘, y = 10, α = 0.5‘,
γ = 3.5‘}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5‘ y) ((n
- 1) b + (γ d)/n - 0.5‘ (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 25.47}]},
PlotRange -> {0, 30},
PlotLegends -> Placed[{" Utility when γ=3.5"}, {{1.35, 0.5}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black}]]
Overlay[{p1, p2}]
```

When $R_J = 0.5$ then the social optimum is stable for a greater range of values for γ.

```
p1 = DynamicModule[{b = 10, d = 0, n = 4, R = 0.5‘, y = 10, α = 0.5‘,
γ = 0.5‘}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5‘ y) ((n
- 1) b + (γ d)/n - 0.5‘ (n - 1 + γ) y)}, {c, 0, y},
```

```
Epilog -> {PointSize[Medium], Point[{10, 27.8125}]},
PlotRange -> {0, 30},
PlotLegends -> Placed[{"Utility when γ=0.5"}, {{1.35, 0.75}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dashed}]]
p2 = DynamicModule[{b = 10, d = 0, n = 4, R = 0.5`, y = 10, α = 0.5`,
γ = 1.5`}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n
- 1) b + (γ d)/n - 0.5` (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{10, 24.6875}]},
PlotRange -> {0, 30},
PlotLegends -> Placed[{"Utility when γ=1.5"}, {{1.35, 0.5}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dotted}]]
p3 = DynamicModule[{b = 10, d = 0, n = 4, R = 0.5`, y = 10, α = 0.5`,
γ = 3.5`}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n
- 1) b + (γ d)/n - 0.5` (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 26.5625}]},
PlotRange -> {0, 30},
PlotLegends -> Placed[{" Utility when γ=3.5"}, {{1.35, 0.25}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black}]]
Overlay[{p1, p2, p3}]
```

When player J has stronger direct reciprocity preferences, $R_J = 1$, the social optimum exists for values of $\gamma < 2.6$.

```
p1 = DynamicModule[{b = 10, d = 0, n = 4, R = 1`, y = 10, α = 0.5`,
γ = 0.5`}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n
- 1) b + (γ d)/n - 0.5` (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{10, 35.625}]},
PlotRange -> {0, 40},
```

```
PlotLegends -> Placed[{"Utility when γ=0.5"}, {{1.35, 0.75}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dashed}]]
p2 = DynamicModule[{b = 10, d = 0, n = 4, R = 1`, y = 10, α = 0.5`,
γ = 1.5`}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n
- 1) b + (γ d)/n - 0.5` (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{10, 29.375}]},
PlotRange -> {0, 40},
PlotLegends -> Placed[{"Utility when γ=1.5"}, {{1.35, 0.5}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dotted}]]
p3 = DynamicModule[{b = 10, d = 0, n = 4, R = 1`, y = 10, α = 0.5`,
γ = 3.5`}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n
- 1) b + (γ d)/n - 0.5` (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 28.125}]},
PlotRange -> {0, 40},
PlotLegends -> Placed[{" Utility when γ=3.5"}, {{1.35, 0.25}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black}]]
Overlay[{p1, p2, p3}]
```

A.1.3.2 Case 2

When the history of the game is $h = (y, .., y)$ and player J believes that each player k will not contribute (i.e., $b_{J,k} = c_k = 0$), then when $R_J = 0.15$ and the traditional public goods game Nash equilibrium is optimal.

```
p1 = DynamicModule[{b = 0, d = 10, n = 3, R = 0.15`, y = 10, α = 0.5`,
γ = 0.5`}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n
- 1) b + (γ d)/n - 0.5` (n - 1 + γ) y)}, {c, 0, y},
```

```
Epilog -> {PointSize[Medium], Point[{0, 12.03125}]},
PlotRange -> {0, 15},
PlotLegends -> Placed[{"Utility when γ=0.5"}, {{1.35, 0.75}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dashed}]]
p2 = DynamicModule[{b = 0, d = 10, n = 3, R = 0.15‘, y = 10, α = 0.5‘,
γ = 1.5‘}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5‘ y) ((n
- 1) b + (γ d)/n - 0.5‘ (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 12.34375}]},
PlotRange -> {0, 15},
PlotLegends -> Placed[{"Utility when γ=1.5"}, {{1.35, 0.5}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dotted}]]
p3 = DynamicModule[{b = 0, d = 10, n = 3, R = 0.15‘, y = 10, α = 0.5‘,
γ = 2.5‘}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5‘ y) ((n
- 1) b + (γ d)/n - 0.5‘ (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 12.65625}]},
PlotRange -> {0, 15},
PlotLegends -> Placed[{" Utility when γ=2.5"}, {{1.35, 0.25}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black}]]
Overlay[{p1, p2, p3}]
```

when $R_J = 0.5$, all else equal, the traditional public goods game Nash equilibrium is again optimal.

```
p1 = DynamicModule[{b = 0, d = 10, n = 3, R = 0.5‘, y = 10, α = 0.5‘,
γ = 0.5‘}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5‘ y) ((n
- 1) b + (γ d)/n - 0.5‘ (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 16.770833}]},
PlotRange -> {0, 20},
```

```
PlotLegends -> Placed[{"Utility when γ=0.5"}, {{1.35, 0.75}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dashed}]]
p2 = DynamicModule[{b = 0, d = 10, n = 3, R = 0.5`, y = 10, α = 0.5`,
γ = 1.5`}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n
- 1) b + (γ d)/n - 0.5` (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 17.8125}]},
PlotRange -> {0, 20},
PlotLegends -> Placed[{"Utility when γ=1.5"}, {{1.35, 0.5}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dotted}]]
p3 = DynamicModule[{b = 0, d = 10, n = 3, R = 0.5`, y = 10, α = 0.5`,
γ = 2.5`}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n
- 1) b + (γ d)/n - 0.5` (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 18.8541667}]},
PlotRange -> {0, 20},
PlotLegends -> Placed[{" Utility when γ=2.5"}, {{1.35, 0.25}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black}]]
Overlay[{p1, p2, p3}]
```

when $R_J = 1$, all else equal, the traditional public goods game Nash equilibrium remains optimal.

```
p1 = DynamicModule[{b = 0, d = 10, n = 3, R = 1`, y = 10, α = 0.5`,
γ = 0.5`}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n
- 1) b + (γ d)/n - 0.5` (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 23.5416667}]},
PlotRange -> {0, 30},
PlotLegends -> Placed[{"Utility when γ=0.5"}, {{1.35, 0.75}, {0.7,
0.5}}],
```

```
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dashed}]]
p2 = DynamicModule[{b = 0, d = 10, n = 3, R = 1`, y = 10, α = 0.5`,
γ = 1`}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n
- 1) b + (γ d)/n - 0.5` (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 24.583333}]},
PlotRange -> {0, 30},
PlotLegends -> Placed[{"Utility when γ=1.0"}, {{1.35, 0.5}, {0.7,
0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black, Dotted}]]
p3 = DynamicModule[{b = 0, d = 10, n = 3, R = 1`, y = 10, α = 0.5`,
γ = 1.5`}, Plot[{y - c + α ((n - 1) b + c) + R α^2 (c - 0.5` y) ((n
- 1) b + (γ d)/n - 0.5` (n - 1 + γ) y)}, {c, 0, y},
Epilog -> {PointSize[Medium], Point[{0, 25.625}]},
PlotRange -> {0, 30}, PlotLegends -> Placed[{" Utility when γ=1.5"},
{{1.35, 0.25}, {0.7, 0.5}}],
AxesLabel -> {"c", "U"},
PlotStyle -> {Black}]]
Overlay[{p1, p2, p3}]
```