The Largesse Design Problem

David K. Levine[1]

Abstract

Largesse design asks what happens when players in a game, who have a limited willingness to sacrifice for the common good, can coordinate on a welfare optimal solution. This theory provides an alternative to psychological theories in explaining why players engage in costly punishments and rewards. It also says that players will engage in other low cost methods of encouraging pro-social behavior, for example, through reputation effects. In this paper, I study the properties and comparative statics of the largesse design problem. Despite a lack of concavity in the maximization problem, solutions are unique and continuous except on a lower dimensional bifurcation set. In examples, I study these bifurcations, in which a small change in the parameters can lead to a large changes in behavior or welfare.

[1] Department of Economics, RHUL

Acknowledgements: First version: December 10, 2025. I am particularly grateful to Drew Fudenberg for getting me started on this, and Klaus Ritzberger who pointed out the connection to semi-algebraic geometry. I would also like to thank Gary Charness, Pedro Dal Bo, Michalis Drouvelis, Rohan Dutta, John Mair, Andrea Mattozzi, Salvatore Modica, Catharina Tilmans, Myrna Wooders, and the Reputation Conference at the EUI. The Leverhulme Trust provided financial support for which I am grateful.

1. Introduction

Largesse design asks what happens in a game in which three assumptions are satisfied. First, some players are willing to sacrifice a limited amount of utility, their *largesse*, for the common good. Second, they agree on a common goal, social welfare measured *ex ante*, before their roles are determined. Third, they are good at solving coordination problems. Hence, largesse design asks what is the maximum welfare achievable, and how should players play to achieve that welfare, given that for each person there is an incentive constraint that they cannot sacrifice more utility than they are willing to. Harsanyi (1982) called this rule utilitarianism, and in the voting literature, Feddersen and Sandroni (2006) refer to it as the ethical voters model.

Largesse design problems have been studied by Feddersen and Sandroni (2006), in the context of voting, by Dutta, Levine and Modica (2021), in the context of public goods, and computationally by Levine (2025). In the voting literature the interaction between groups, each of which solves a largesse design problem, has been studied by Coate and Conlin (2004), Herrera, Morelli and Nunnari (2016), and by Levine and Mattozzi (2020). Comparative statics, the mapping from parameters to solutions, has not been extensively studied, and is the focus of this paper.

The largesse design problem is difficult, because the constraint set need not be convex, and the objective function need not be concave. Never-the-less all are defined by polynomials, so the relevant sets and mappings are semi-algebraic. In particular, I show that, except on a lower dimensional bifurcation set, the map from parameters to solutions is single valued and continuous. The bifurcation set is of particular importance, as bifurcations have been seen to occur in the laboratory. For example, in the centipede game, increasing the stakes, or changing the payoffs in the final round, generally causes the number of rounds before players grab to switch from relatively high to relatively low.[2] Outside the laboratory, bifurcations (or threshold effects, which are the same thing) have been used to explain phenomena such as mass protests and bank runs.

In addition to the generic single valuedness and continuity of the solution correspondence, I establish other general results about the largesse design problem. Two of these are straightforward, but useful. First, is establishing the utility transforms under which the feasible set, and those under which the solution set, remain un-

[2] See McKelvey and Palfrey (1992), Maniadis (2011), and Cox and James (2015).

changed. Second, is demonstrating the monotonicity of the feasible set and of welfare as largesse is increased. Finally, if each player has a distribution of largesse, I show that that replacing each type with a single type having the average largesse can only increase welfare. I then give conditions under which welfare and aggregate play remain unchanged, as the largesse distribution is altered. In particular if a player is playing a best response with positive probability, then there is a range of largesse distributions which do not change the solution.

I study several applications. I am particularly interested in games involving punishments (similar to ultimatum bargaining and public goods) and rewards (similar to trust and gift exchange). These are games which have been used in the laboratory to show that people are not entirely self-interested. These laboratory studies have been used as the basis for various psychological theories of preferences for fairness and reciprocity, such as Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Falk and Fischbacher (2006), and others. I show that largesse design provides an alternative explanation: that players strategically engage in punishments and rewards to provide incentives for pro-social behavior.

Largesse design theory is about deploying largesse strategically. As a motivating example, I consider a simple punisher's dilemma game, in which a second player can strategically punish a first player, for anti-social behavior. This game is designed to highlight the differences with existing theories. Selfish players, altruistic players, and Fehr and Schmidt (1999) fairness players never punish. In the solution to the largesse design problem, when largesse is low, the second player does not punish, and the first player engages in a low level of pro-social behavior, much as a mildly altruistic player would. As largesse increases, there is a bifurcation, and it becomes optimal for the second player to begin punishing the first player for anti-social behavior. This leads to a rapid increase in welfare as largesse increases. After the punishment is enough to induce the first player to act pro-socially with probability one, as largesse increases further, additional, smaller, gains take place as the second player reduces the socially costly punishments.

The second application is a family of reward games, similar to the trust game. These I study in greater detail. Unlike the punishment game, here there is no bifurcation, but rather as largesse increases, increased largesse by the receiver enables them to provide a higher rate of return, leading the sender to invest more. The third application considers a symmetric public goods game in which there are three choices:

to free ride, to contribute to the public good, or to contribute to the public good and punish free-riders. As in the case in the simple punisher's dilemma, there is a bifurcation. With three actions it is more dramatic. A small increase in largesse not only changes behavior from free riding to punishing, but welfare jumps up discontinuously when this occurs.

The final application looks at an alternative way in which largesse can be used strategically to encourage players to behave pro-socially. Largesse design is a theory of commitment with an *ex ante* constraint on the amount that players are willing to lose. Reputation theory is based on the existence of at least a small number of committed players. I show how, in largesse design, the endogenous commitment of a few players, with relatively small largesse, can be bootstrapped to give good welfare results in a game between a long-run player and a sequence of short-run opponents.

2. Literature Review

Before proceeding, I would like to convince you that the theory is relevant. I hope I do not have to convince you that some people are willing to make limited sacrifices for the common good.

With respect to social welfare as a criterion, I am going to observe that in laboratory studies (see in particular the review by Fehr and Charness (2024)) people care about both efficiency and fairness. With risk averse individuals, social welfare is a parsimonious model of people who care about both. Greater efficiency raises welfare, and, with risk aversion, greater fairness corresponds to better insurance, and also raises welfare. I am not going to argue that people necessarily agree that *ex ante* social welfare is the right goal. Rather, I am going to propose it is a reasonable criterion, and I am going to ask what happens they do. To this I will add that in Levine (2025) I show that it works well in explaining laboratory behavior.

I feel I do need to convince you that people are good at solving coordination problems, both inside and outside the laboratory. This is in light of the fact that, in the experimental lab, beginning with Van Huyck, Battalio and Beil (1990)'s study of the minimum game, coordination failures have been documented, and indeed there are a variety of theories, risk dominance and global games most prominently, that aim to predict when coordination failure will occur.

I am going to propose that these failures of coordination are not due to the inability of people to coordinate, but rather that, in the laboratory, as well as outside, the environment is noisy. Let me start with the Van Huyck, Battalio and Beil (1990) minimum game, with 15 or more players, in which it is observed that experienced players, rather than coordinating on the welfare optimal highest level of effort, instead only manage the least level of effort. My observation is this: with 15 or more players, in a game where only the minimum level of contribution to a public good matters, then if players tremble, it is highly likely that one will ruin things for everybody.

Specifically, (see Levine (2025)) if 1/3 of the players tremble uniformly then, in fact, the only equilibrium is for all non-trembling players to provide the least effort. Hence, in the relevant normal form, the one with trembling, they in fact coordinate on the welfare optimal equilibrium, as there is only one equilibrium. Now 1/3 of the players trembling uniformly may seem like a lot: 29% of the time should they be playing more than the least level of effort. In fact, in the Van Huyck, Battalio and Beil (1990) data in the final of ten periods, and despite the fact that in the previous six periods the minimum level of effort was the least possible, 29% are providing more than the least level and, indeed, 10% are providing the highest level.

I turn next to two meta-studies. The first is a meta-study of 2 × 2 stag hunt experiments by Dal Bo, Frechette and Kim (2021). These are strangers treatments in which players played at least eight times. In this game, the welfare optimum is for all to hunt stag, and the basin is defined as the greatest probability of hare that makes it optimal for all to hunt stag. The second is a meta-study of indefinitely repeated prisoner's dilemma experiments by Levine (2024). These are strangers treatments in which players played at least fifteen times. Here I follow the literature in defining the basin with respect to the 2×2 game in which the strategies or grim-trigger and always defect. I report welfare with utility normalized, as is standard in the literature, to zero for mutual defection and one for mutual cooperation. This is the same as the cooperation rate if play is perfectly correlated or if the off-diagonal welfare is 1/2 and is highly correlated with the cooperation rate.

Below, in Figure 2.1, I report the theory and the data for the different treatments. On the horizontal axis is the basin of the treatment, on the vertical axis is the fraction of time the players chose stag for the stag hunt games and the normalized welfare for the indefinitely repeated prisoner's dilemma games. The red correspondence is the best equilibrium when players tremble uniformly 1/3rd of the time. The blue correspondence is the predicted probability from both risk dominance and global games, which are the same in this setting, and both of which predict the "good" equilibrium of stag or grim-trigger if the basin is greater than 50%, and the bad equilibrium of hare or "always defect" if the basin is less than 50%.

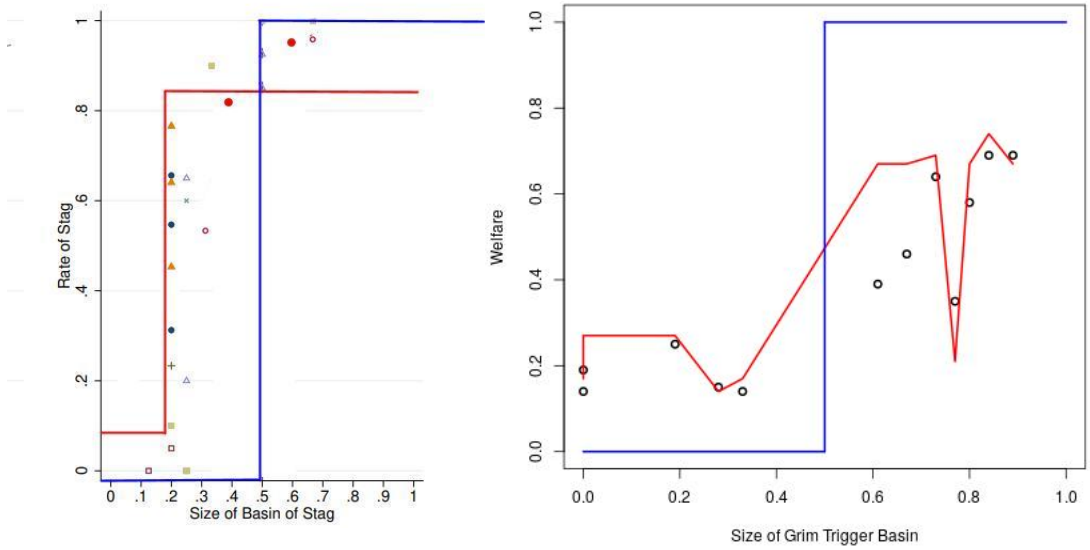

Figure 2.1: Stag Hunt and Indefinitely Repeated Prisoner's Dilemma Meta-studies

dots: data
blue: risk dominance/global games
red: best equilibrium with 1/3 probability of uniform tremble

left panel: stag hunt treatments, period 8
vertical axis: frequency of stag
right panel: indefinitely repeated prisoner's dilemma treatments, all periods starting in period 10
vertical axis: welfare with payoff to mutual defection normalized to zero and mutual cooperation to one

I believe that these two graphs show that best equilibrium with trembling is a viable alternative to risk dominance and global games for explaining what happens in coordination games.

Beyond this, I have documented in Levine (2025) that a simple theory of 1/3 uniform trembling in a largesse design problem in which half the population is selfish, and half have a largesse of roughly $1.00 over the course of play, does a good job in predicting the play of experimental participants across a wide variety of experiments. This built on earlier work by Dutta, Levine and Modica (2021), in which we examined the largesse design problem both inside and outside the laboratory.

Outside the Laboratory

Outside the laboratory, two Nobel prizes have been awarded for documenting the ability of people to coordinate on good equilibria: Coase (1960) and Ostrom (1990). I would add to this also the work of Townsend (1994) on insurance in rural societies. Indeed: the entire premise of contract theory is that people are pretty good at figuring out good solutions to social dilemmas. Outside the laboratory, of course more time is available and people can talk, but the problems tend to be more complex. The success of market design as a commercial enterprise convinces me of two things. First, people cannot necessarily solve hard problems, but they would like to.

In voting theory, the theory of ethical voters or rule utilitarianism, which give rises the largesse design problem have been studied theoretically by Feddersen and Sandroni (2006), and empirically by Coate and Conlin (2004). Subsequently the model has been widely used by other authors, such as Herrera, Morelli and Nunnari (2016).

As I indicated, the idea of largesse design is closely connected to the idea of rule utilitarianism. This is an idea dating back to Mill (1861), and is described by Garner and Rosen (1967) as "the rightness or wrongness of a particular action is a function of the correctness of the rule of which it is an instance." This is in contrast to act utilitarianism, a more commonly used concept in economics. The difference between the two ideas can be illustrated by the punishment of a free-rider in a public good game. As an act this is unambiguously bad, as it lowers welfare. However, used as a rule, it can be good, because it encourages people to contribute to the public good. The idea that rule utilitarianism is an individualistic way to solve coordination problems was studied in Harsanyi (1982).

Finally, the feasible set for the largesse design problem has a close connection to the idea of ϵ-equilibrium, introduced in Radner (1980). In largesse design each type of each player can have a different value of ϵ and it need not be small, but the relaxation of the constraints through largesse is exactly that from ϵ-equilibrium.

3. The Model

The setting is that of a normal form game. There are n players and each player has a finite strategy space $s^i \in S^i$ with payoffs $u^i(s^i, s^{-i})$. The space of pure strategy profiles is denoted by S. If there are trembles, then strategies represent intentions, and the normal form is the corresponding expected utility, that is, trembles are taken to already be incorporated into the normal form of the game.

Each player has a finite type space T^i with types denoted by τ. Denote the set of all types by $T = \cup_{i=1}^n T^i$. The fraction of type τ is denoted by $\phi_\tau \geq 0$ where the fractions of types of a single player add up to one: for each i it is the case that $\sum_{\tau \in T^i} \phi_\tau = 1$. Each type is characterized by a non-negative largesse $\gamma_\tau \geq 0$. Types are private information.

Let σ_τ^i denote a mixed strategy for type $\tau \in T^i$, and let σ be a vector of such mixed strategies. Denote by Σ the corresponding set of profiles. Define the induced mixture from all types of player i by $\overline{\sigma}^i = \sum_{\tau \in T^i} \phi_\tau \sigma_\tau^i$. For a given vector $\sigma \in \Sigma$ of mixed strategies for each player and type let $u^i(\sigma_\tau^i, \overline{\sigma}^{-i})$ be the expected utility of type τ of player i. Given σ each type $\tau \in T^i$ of player i faces an incentive constraint that the gain from deviating is no greater than γ_τ

$$g_\tau(\sigma) \equiv \max_{s^i \in S^i} u^i(s^i, \overline{\sigma}^{-i}) - u^i(\sigma_\tau^i, \overline{\sigma}^{-i}) \leq \gamma_\tau.$$

If this is satisfied for all players and types we say that σ is *incentive compatible.* If $\gamma_\tau = 0$ for all τ, then incentive compatibility is the same as Nash equilibrium.

Expected per capita *social welfare* from σ is given by

$$W(\sigma) \equiv \sum_{i=1}^n \sum_{\tau \in T^i} \phi_\tau u^i(\sigma_\tau^i, \overline{\sigma}^{-i})/n = \sum_{i=1}^n u^i(\overline{\sigma})/n.$$

The *largesse design problem* is to find the σ that maximizes social welfare subject to incentive compatibility. If $\gamma_\tau = 0$ for all τ this is the same as choosing the welfare optimal Nash equilibrium.

Several special cases are of particular interest. If there is only one type of each player we say that largesse is *unitary*, and replace the redundant type subscript τ in γ_τ with the player superscript, γ^i. If each player draws from the same distribution of types we say that largesse is *symmetric.* A type with largesse $\gamma_\tau = 0$ is called *selfish.*

4. The Punisher's Dilemma

The punisher's dilemma is a 2×2 game. The first player, called the US, chooses between (F)ighting and (C)onceding, and the second player, called the SU, chooses between (S)tanding aside, and (P)unishing. Punishing involves building a doomsday machine. Take first the case where the SU stands aside. If the US fights both get 4. If the US concedes they get less, 3, but the SU gets more, 7. Conceding is welfare superior, as it avoid the cost of fighting.

On the other hand, it costs 1 to build a doomsday machine. This machine will be activated only if the US fights, in which case it will explode causing 2 units of damage to the US (and none to the SU). This is summarized in the payoffs matrix in Table 4.1 below.

	S	P
F	$4, 4$	$2, 3$
C	$3, 7$	$3, 6$

Table 4.1: Punisher's Dilemma Game"

A key fact about this game is that a selfish, altruistic, or Fehr and Schmidt (1999) fairness SU will never build a doomsday machine. For selfish SU, standing aside, S strictly dominates building a doomsday machine. Moreover, it strictly dominates from a welfare point of view as well, so an altruistic SU would never would also never build a doomsday machine. For a Fehr and Schmidt (1999) fairness SU when the US plays F the choice S is strictly better from a fairness point of view. When the US plays C the gain in fairness to the US from the SU punishing is 1 and the cost to the SU is 1 and Fehr and Schmidt (1999) assume that no player would make such a trade-off.

Notice that when the SU is playing S neither the selfish nor Fehr and Schmidt (1999) US would play C so that the unique outcome is FS.

In contrast to the other theories, largesse design predicts that doomsday machines will sometimes be built. Specifically, I analyze the unitary symmetric case, in which there is one type of each player, and both have the same largesse: $\gamma^1 = \gamma^2 = \overline{\gamma}$. Figure 4.1 below summarizes Proposition 4.1 below, characterizing the solution of the largesse design problem.

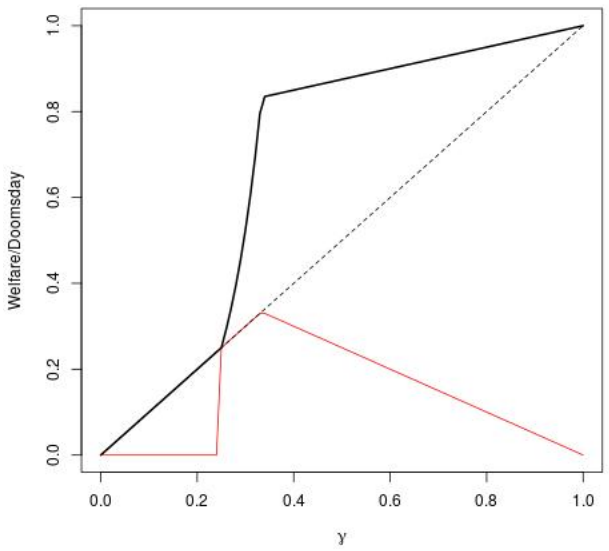

Figure 4.1: Punisher's Dilemma Game

welfare measured as fraction of difference between FS and CS
solid line: welfare from solution of largesse design problem
dotted line: welfare from largesse of player 1 only ($\gamma^2 = 0$)
red: σ_P^2, the probability of P in the solution of largesse design problem

Denote by σ_C^1 and σ_P^2 the probabilities of the (single types of) US conceding and the SU of building a doomsday machine. The cost of building a doomsday machine is 1, so the incentive constraint for the SU is $\sigma_P^2 \leq \overline{\gamma}$. For low levels of largesse, $\overline{\gamma}$, the cost from this low level of punishment is too great to justify the small improvement in incentives for the US. Hence the optimum is to take $\sigma_P^2 = 0$ and rely solely on the largesse of the US to get a low level of concession. When $\overline{\gamma} = 1/4$ this changes, and there is a bifurcation, two equally good solutions, one with $\sigma_P^2 = 0$ and one with $\sigma_P^2 = \overline{\gamma}$. Beyond this point it is better to use punishment, and σ_P^2 jumps up. As $\overline{\gamma}$ rises so does σ_P^2 and welfare rises rapidly. Always, σ_C^1 is chosen as large as is feasible. At $\overline{\gamma} = 1/3$, this maximum feasible σ_C^1 reaches 1. After this σ_P^2 should be reduced to avoid a costly and unnecessary punishment. Note that this "least possible punishment" is characteristic of the largesse design problem.

Notice in particular the subtle behavior of σ_P^2. It is initially zero. At $\overline{\gamma} = 1/4$ it jumps up, then rises linearly until $\overline{\gamma} = 1/3$, following which it declines linearly. I will show later that this behavior - a solution that is unique and continuous except on a lower dimensional bifurcation set (here $\overline{\gamma} = 1/4$) - is typical of solutions of the largesse design problem. Notice that in this case behavior, as measured by σ bifurcates, but welfare is continuous. I will refer to such a case as a *behavioral bifurcation.* As I will show later, there can also be *welfare bifurcations*, in which welfare jumps.

To understand a bit better why doomsday machines are useful for the largesse design problem, focus on the case in which $\overline{\gamma} = 1/2$. If no doomsday machine is built, the US must satisfy $\gamma_C^1 \leq 1/2$ so the highest possible welfare is 1/2 of 8 and 1/2 of 10, which is to say 9. By contrast, with $\overline{\gamma} = 1/2$ the SU could build a doomsday machine with probability $\sigma_P^2 = 1/2$ which would make the US indifferent between fighting and conceding. The welfare optimum is when they concede, in which case welfare is $50-50$ between CS and CP, which is to say 9.5, definitely better than the 9 that can be obtained using solely US largesse. Notice, though, that when $\overline{\gamma} = 1/2$ the optimal choice of $\sigma_P^2 \neq 1/2$. Because they have largesse, the US is willing to concede with probability 1 even if it involves a loss. That is, their constraint is strictly satisfied. Hence, σ_P^2 should be reduced until the US constraint is satisfied with equality, raising welfare by reducing the cost of punishment.

These results are summarized in Proposition 4.1 below.

Proposition 4.1. *There are five regions*

i. $\overline{\gamma} < 1/4$, *the solution is unique,* $\sigma_C^1 = \overline{\gamma}$, $\sigma_P^2 = 0$ *and* $2W = 8 + 2\gamma$

ii. $\overline{\gamma} = 1/4$, *there are two solutions* $\sigma_C^1 = 1/4$, $\sigma_P^2 = 0$ *and* $\sigma_C^1 = 1/2$, $\sigma_P^2 = 1/4$, *and both give welfare* $2W = 8.5$.

iii. $1/4 < \overline{\gamma} \leq 1/3$, *the solution is unique,* $\sigma_C^1 = \overline{\gamma}/(1 - 2\overline{\gamma})$, $\sigma_P^2 = \overline{\gamma}$ *and* $2W = 8 - 3\overline{\gamma} + 2\overline{\gamma}\left(1 + \overline{\gamma}\right)/\left(1 - 2\overline{\gamma}\right)$

iv. $1/3 \leq \overline{\gamma} < 1$, *the solution is unique,* $\sigma_C^1 = 1$, $\sigma_P^2 = (1 - \overline{\gamma})/2$ *and* $2W = 9.5 + \overline{\gamma}/2$

v. $\overline{\gamma} \geq 1$, *the solution is unique and first best,* $\sigma_C^1 = 1$, $\sigma_P^2 = 0$ *and* $2W = 10$

Proof. The objective function is

$$2W = 8 - 3\sigma_P^2 + 2\sigma_C^1\left(1 + \sigma_P^2\right).$$

The incentive constraint for 2 is $\sigma_P^2 \leq \overline{\gamma}$.

If in a solution C is a best response, then $\sigma_P^1 \geq 1/2$, and this implies $\overline{\gamma} \geq 1/2$. It follows that $\sigma_C^1 = 1$ since this certainly satisfies the incentive constraints, and maximizes welfare. Welfare is then $2W = 10 - \sigma_P^2$, so it is desirable to decrease σ_P^1.

If $\sigma_P^1 > 1/2$ this cannot be optimal. Moreover, since $\overline{\gamma} \geq 1/2$ it is possible to lower σ_P^1 further without violating the incentive constraint for player 1.

It follows that F is always a best response in any solution. Hence the incentive constraint for player 1 is

$$\sigma_C^1 \left(1 - 2\sigma_P^2\right) \leq \overline{\gamma}. \tag{4.1}$$

If the incentive constraint for player 1 does not bind, welfare can be improved by increasing σ_C^1 since this does not affect the constraint for player 2. Hence either the constraint binds or $\sigma_C^1 = 1$. If $\sigma_C^1 = 1$ welfare can be improved by decreasing σ_P^2, so if the constraint does not bind, then $\sigma_C^1 = 1$ and $\sigma_P^2 = 0$. This is feasible if and only if $\overline{\gamma} \geq 1$ which is case (v).

Assume, then, that the incentive constraint for player 1 does bind. Solve it to find

$$\sigma_C^1 = \frac{\overline{\gamma}}{1 - 2\sigma_P^2} \tag{4.2}$$

if $\sigma_P^2 < (1 - \overline{\gamma})/2$. In this case welfare is given by

$$2W = 8 - 3\sigma_P^2 + 2\overline{\gamma}\frac{1 + \sigma_P^2}{1 - 2\sigma_P^2}.$$

It can be checked that the second derivative is positive, so the solution is to either take $\sigma_P^2 = 0$ or as large as is feasible. The welfare difference between $\sigma_P^2 = \overline{\gamma}$ and $\sigma_P^2 = 0$ is then given by

$$-5\overline{\gamma} + 2\overline{\gamma}\frac{1 + \overline{\gamma}}{1 - 2\overline{\gamma}}.$$

This has a root at $\overline{\gamma} = 0$ and at $\overline{\gamma} = 1/4$, and is negative in between. This gives case (i) and half of case (ii).

For $\overline{\gamma} \geq 1/4$ and $\sigma_C^1 < 1$ it is then optimal to choose σ_P^2 as large as possible, that is, $\sigma_P^2 = \overline{\gamma}$. From 4.2 this gives

$$\sigma_C^1 = \frac{\overline{\gamma}}{1 - 2\overline{\gamma}}.$$

Observing that this equals 1 at $\overline{\gamma} = 1/3$ gives the rest of case (ii) and case (iii).

Once $\sigma_C^1 = 1$ welfare is $2W = 10 - \sigma_P^2$,so σ_P^2 should be chosen as small as possible, which, from 4.1 is $\sigma_P^2 = (1 - \overline{\gamma})/2$. As this is equal to 0 at $\overline{\gamma} = 1$, that completes case (iv).

□

5. Basic Theory

Existence of a solution is fundamental and fairly obvious. For completeness I state and prove this.

Proposition 5.1. *The incentive compatible set is compact and non-empty. A solution to the largesse design problem exists and the set of solutions is compact.*

Proof. The incentive compatible set is non-empty because it contains all Nash equilibria, and is closed because it is defined by continuous functions and weak inequalities. It is bounded because it lies in the simplex. The objective function is continuous. Hence there is a maximum, and the argmax set is compact. □

Notice, however, that the constraint set, while closed, is not generally convex and the objective function, while continuous, is not generally concave. Despite this, the solution correspondence has strong properties.

Fix the set of players and the spaces of pure strategies S^i. The largesse design problem is parametrized by the utility vector, any point in $\Re^{\#S}$, the type distribution, any point in the cartesian product of type simplices Φ, and the largesse vector, any point in $\Re_+^{\#T}$. The parameter space is some $Z \subseteq \Re^{\#S} \times \Re_+^{\#T} \times \Phi$.

For each $z \in Z$ let $F(z) \subseteq \Sigma$ denote the feasible set, that is, the subset of $\sigma \in \Sigma$ where $g(\sigma) \leq \gamma$, let $\hat{F}(z)$ denote the solutions to the largesse design problem, and let $\hat{W}(z)$ be the corresponding welfare.

Proposition 5.2. *The correspondence $z \rightrightarrows F(z)$ is upper-hemi continuous and compact valued.*

Proof. It is compact valued by Proposition 5.1, and upper-hemi continuous because it is defined by weak inequalities and continuous functions. □

Corollary 5.3. *If $z^n \to z$ then $\lim W(z^n) \leq W(z)$.*

This says that welfare can jump up in the limit, but not down.

Proof. Let $\hat{\sigma}^n \in \hat{F}(z^n)$ be a convergent sub-sequence. By Proposition 5.2 $\tilde{\sigma} = \lim \hat{\sigma}^n \in F(z)$. Hence, since the objective function is continuous, $\lim W(z^n) = \lim W(\hat{\sigma}^n, z^n) = W(\tilde{\sigma}, z) \leq W(z)$. □

5.1. Generic Continuity

Upper-hemi continuity of the feasibility correspondence might seem like the end of it: neither $\hat{F}(z)$ nor $\hat{W}(z)$ is upper-hemi continuous. Indeed, take the unitary symmetric case with $\gamma^1 = \gamma^2 = 0$ and consider the "prisoner's dilemma game" below in Table 5.1 where $B < 3$.

	D	C
D	$0, 0$	$B, -1$
C	$-1, B$	$1, 1$

Table 5.1: The "Prisoner's Dilemma Game"

For $B > 1$ this is indeed a prisoner's dilemma game, $F(B) = \hat{F}(B) = \{DD\}$, $\hat{W}(B) = 0$, and there is nothing more to be said. However, when $B = 1$ this is a coordination game, $F(B) = \{DD, CC\}$, and both $\hat{F}(B)$ and $\hat{W}(B)$ jump to CC and 1 respectively. When $B < 1$ the feasible set $F(B)$ has three points, $\{DD, CC\}$ and the mixed Nash equilibrium, while $\hat{F}(B)$ and $\hat{W}(B)$ remain unchanged. Notice that in the unitary symmetric case, with $\gamma^1 = \gamma^2 = \overline{\gamma} > 0$, the same jump occurs, albeit at $B - 1 = \overline{\gamma}$.

In contrast to the example of section 4, in this bifurcation welfare jumps along with behavior. As indicated, I refer to this as a welfare bifurcation. Notice that, like the bifurcation in section 4, the bifurcation is "rare" in the sense that it occurs only on a set of dimension 0, in this case the set $B = \overline{\gamma} + 1$. It turns out that the property that bifurcations occur only on lower dimensional sets is true in general. To understand why, it is important that the utility functions are not only continuous in the parameters and mixed strategies, but are polynomials.

A set defined by polynomial equalities and (strict) inequalities, and the unions of such sets are called semi-algebraic, and have strong properties.[3] Coste (2002) is a good reference, and Blume and Zame (1994) provide a good summary. Chief among these properties are that unions, intersections, complements, boundaries, interiors, and closures of semi-algebraic sets are semi-algebraic. Semi-algebraic sets are a finite union of connected real-analytic manifolds, and the dimension of a semi-algebraic set is the largest dimension of such a manifold. The boundary of a semi-algebraic set has lower dimension than the original set.

[3] I am grateful to Klaus Ritzberger for pointing this out to me.

A correspondence with semi-algebraic graph is called semi-algebraic. The values of a semi-algebraic correspondence are semi-algebraic sets. Semi-algebraic functions satisfy a variation on the implicit function theorem and Sard's theorem, called the Hardt Triviality Theorem.

From this point forward, I am going to assume that the parameter space Z is semi-algebraic. This implies that the constraint correspondence F is semi-algebraic. A detailed proof can be found in Blume and Zame (1994), who consider Nash equilibrium, but their proof applies if we replace 0 by γ^i in their constraints. Besides Blume and Zame (1994), a series of papers by Govindan and Wilson, for example, Govindan and Wilson (2009), also use the semi-algebraic properties of incentive constraints to prove results about refinements of Nash equilibrium, such as forward induction and strategic stability.

Write $W(\sigma, z)$ to make transparent the dependence of welfare on the parameters. In the current setting, consider that the solution correspondence can be written as

$$\sigma \in \hat{F}(z) \iff$$

$$\forall \sigma', ((W(\sigma, z) > W(\sigma', z)) \wedge (W(\sigma, z) = W(\sigma', z))) \vee (\sigma' \in F(z)) \vee (\sigma \in F(z))) \,.$$

This set is not obviously semi-algebraic because it uses the quantifier $\forall\sigma'$, but the Tarski-Seidenberg Theorem asserts that this does not matter, that, in fact, a set defined by polynomial equalities and inequalities, logical operations, and quantifiers, is semi-algebraic, that is, can be defined by (a different set of) polynomial equalities and inequalities without quantifiers. Hence the solution correspondences $\hat{F}(z), \hat{W}(z)$ are semi-algebraic.

Proposition 5.4. *The correspondence $z \rightrightarrows (F(z), \hat{F}(z), \hat{W}(z))$ is continuous (upper and lower) at every point of the complement of a (relatively) closed, lower-dimensional, semi-algebraic subset of Z.*

Proof. A lemma in Blume and Zame (1994), shows this to be true for any compact valued semi-algebraic correspondence on a semi-algebraic domain. □

5.2. Compressed Models

The basic idea of this subsection is that heterogeneity in largesse, that is, multiple types and divergence from the unitary largesse model, only matters when largesse is large. If largesse is small then players "should" be playing best responses most of the time, and if they do, the remaining small probability of playing something better can be distributed among different types in proportion to their largesse. This breaks down when players are not playing best responses most of the time, because there may be too many players with small largesse who are unable to play non-best responses frequently.

For any given model, we can define a unitary model, the *compressed model*, as the model with a single type of each player with the average largesse $\overline{\gamma}^i \equiv \sum_{\tau \in T^i} \phi_\tau \gamma_\tau$. If σ is incentive compatible in the original model then $\overline{\sigma}$ is clearly incentive compatible in the compressed model, so that $\overline{W} \geq W$. If in fact $\overline{W} = W$, I say that the model *compresses*. In this case, if $\hat{\sigma}$ solves the original problem, then $\overline{\hat{\sigma}}$ solves the compressed problem, as it is feasible there and gives the maximum possible welfare. In other words, the description of a player's play aggregated across types is the same for the original and compressed models.

As indicated, a model need not compress. Consider, for example, the two player dictator game with risk aversion in which the first player must choose between keeping an endowment giving a utility of $(8, 0)$ and splitting the endowment giving a utility of $(5, 5)$. Here, as will be the case when both players are risk averse, the unequal split provides less welfare than the equal split. Suppose that there are two equally likely types of player 1 with largesse 0 and 6, that is, one type is selfish. Then $\overline{\gamma} = 3$, so that the solution of the compressed problem is for player 1 to split the endowment with probability 1. However, this is not feasible in the original problem, since the selfish player will never split the endowment, and so there is only a fifty percent probability of player 1 splitting the endowment. On the other hand, if the two types have largesse 0 and 3 then the solution of the compressed model with $\overline{\gamma} = 1.5$ is to split half the time, and this is the same solution as the original model where the selfish players do not split, and the remaining players always do

Which case is relevant? The case where the model fails to compress as in the counter-example, or the case where it does compress as occurs with less largesse? Let $BR^i(\overline{\sigma})$ denote the set of best responses to $\overline{\sigma}^{-i}$ by player i. The next result gives a sufficient condition for a model to compress.

Proposition 5.5. *Let $\hat{\overline{\sigma}}$ be a solution of the compressed model. Let $\hat{\gamma}^i \equiv \max_{\tau \in T^i} \gamma_\tau$.*

If for each player i we have

$$\Sigma^i \equiv \sum_{s^i \notin BR(\hat{\overline{\sigma}})} \hat{\overline{\sigma}}^i(s^i) \leq \overline{\gamma}^i/\hat{\gamma}^i \tag{5.1}$$

then $\hat{\sigma}^i_\tau(s^i) \equiv (\gamma_\tau/\overline{\gamma})\hat{\overline{\sigma}}^i$ *for* $s^i \notin BR(\overline{\sigma})$ *and*

$$\hat{\sigma}^i_\tau(s^i) \equiv \left(1 - (\gamma_\tau/\overline{\gamma})\Sigma^i\right)\hat{\overline{\sigma}}^i(s^i)/\sum_{s^i \in BR(\hat{\overline{\sigma}})} \hat{\overline{\sigma}}^i(s^i)$$

for $s^i \in BR(\overline{\sigma})$ *is a solution to the original problem and* $W(\hat{\sigma}) = W(\hat{\overline{\sigma}})$*. That is, when 5.1 holds, the model compresses.*

What this says is that if the solution of the compressed problem has a sufficiently high probability of playing a best response for each player, then it is also a "solution" to the original problem. If the original problem was unitary, then the condition in equation 5.1 is always satisfied. More generally, equation 5.1 can be viewed as establishing the amount of heterogeneity in largesse that is consistent with an original model that has the "same" solution as the unitary model.

As an example, consider the punisher's dilemma game from section 4. In the range $\overline{\gamma}$ between 1/4 and 1/3 the largest probability of not playing a best response is by player 1 and is $\sigma^1_C = \overline{\gamma}/(1-2\overline{\gamma})$. Suppose there are two types, with the same distribution for both players, a selfish type with largesse 0 and an ethical type $\tau = E$ with probability ϕ_E and largesse $\overline{\gamma}/\phi_E$. Then from Proposition 5.5 the solution for the compressed problem is valid for the two type problem provided $\overline{\gamma}/\phi_E \leq \overline{\gamma}/(1-2\overline{\gamma})$. In other words, $\phi_E \geq 1 - 2\overline{\gamma}$. This ranges from $\phi_E \geq 1/2$ at $\overline{\gamma} = 1/4$ to $\phi_E \geq 2/3$ at $\overline{\gamma} = 1/3$.

Proof. Since $W(\hat{\sigma}) \leq W(\hat{\overline{\sigma}})$ it suffices to prove that $\hat{\sigma}$ is feasible in the original problem.

First, observe that $\overline{\hat{\sigma}} = \hat{\overline{\sigma}}$. For $s^i \notin BR^i(\overline{\sigma})$ we have $\overline{\hat{\sigma}}^i(s^i) = \sum_{\tau \in T^i} \phi_\tau(\gamma_\tau/\overline{\gamma}^i)\hat{\overline{\sigma}}^i(s^i) = \hat{\overline{\sigma}}^i(s^i)$. For $s^i \in BR(\overline{\sigma})$ we have

$$\overline{\hat{\sigma}}^i(s^i) = \sum_{\tau \in T^i} \phi_\tau \left(1 - (\gamma_\tau/\overline{\gamma}^i)\Sigma^i\right)\hat{\overline{\sigma}}^i(s^i)/\sum_{s^i \in BR(\hat{\overline{\sigma}})} \hat{\overline{\sigma}}^i(s^i) = \hat{\overline{\sigma}}^i(s^i).$$

The incentive constraints are always satisfied for $\sigma^i \in BR^i(\overline{\sigma})$ so in the original problem the incentive constraints are

$$\sum_{s^i \notin BR(\hat{\sigma})} \left(\max_{s^i \in S^i} u^i(s^i, \overline{\sigma}^{-i}) - u^i((\gamma_\tau/\overline{\gamma}^i)\hat{\overline{\sigma}}^i(s^i), \overline{\sigma}^{-i}) \right) \leq \gamma_\tau,$$

which are satisfied since the constraints are satisfied in the compressed model.

The crucial step is to check that $\hat{\sigma}^i_\tau$ is actually a probability distribution, that is, $(\gamma_\tau/\overline{\gamma}^i)\Sigma^i \leq 1$. Then

$$(\gamma_\tau/\overline{\gamma}^i)\Sigma^i = (\gamma_\tau/\overline{\gamma}) \sum_{s^i \notin BR(\hat{\sigma})} \hat{\overline{\sigma}}^i(s^i) \leq (\hat{\gamma}^i/\overline{\gamma}^i) \sum_{s^i \notin BR(\hat{\sigma})} \hat{\overline{\sigma}}^i(s^i).$$

By the assumption, equation 5.1, this final expression is less than or equal to one. □

It is tempting to think that when γ is small, a best response must be played with high probability. Because of mixing this might not be the case, but when the solution for $\gamma = 0$ is unique and a strict Nash equilibrium, if we bound the ratios of the γ_τ's, then, indeed, the model compresses.

Proposition 5.6. *Suppose that Z^B are z such that $\overline{\gamma}^i/\hat{\gamma}^i \geq B$ or $\gamma = 0$. Suppose at $\hat{z} \in Z^B$ that $\hat{\gamma} = 0$ and that the solution to the largesse design problem is unique and is a strict Nash equilibrium $\hat{\sigma}$. Then there is a neighborhood of $\hat{z}$ in Z^B in which the model compresses.*

In particular if we fix the payoffs, the types distribution and γ, and consider models $\lambda\gamma$ where λ is a positive scalar constant, and the solution to the largesse design problem is unique and a strict Nash equilibrium, then, for small enough λ, the model compresses.

Proof. Consider

$$\overline{\Sigma}(z) \equiv \max_i \sup_{\hat{\overline{\sigma}}^i | \hat{\overline{\sigma}} \in \hat{\overline{F}}(z)} \sum_{s^i \notin BR^i(\hat{\overline{\sigma}})} \hat{\overline{\sigma}}^i(s^i).$$

By Proposition 5.5 in Z^B if $\overline{\Sigma}(z) \leq B$ the model compresses. Suppose that in every neighborhood of $\hat{z}$ in Z^B there is some z^n such that the model does not compress, so that in particular $\overline{\Sigma}(z^n) > B$. We can choose a sequence $\hat{\overline{\sigma}}^n \in \hat{\overline{F}}(z)$ so that $\overline{\Sigma}(z^n) - \max_i \sum_{s^i \notin BR^i(\hat{\overline{\sigma}}^n)} \hat{\overline{\sigma}}(s^{in}) \leq B/2$. Hence $\max_i \sum_{s^i \notin BR^i(\hat{\overline{\sigma}}^n)} \hat{\overline{\sigma}}(s^{in}) \geq B/2$. Consider a convergent sub-sequence $\hat{\overline{\sigma}}^n \to \tilde{\sigma}$. If $\tilde{\sigma} = \hat{\sigma}$ then $\max_i \sum_{s^i \notin BR^i(\hat{\overline{\sigma}}^n)} \hat{\overline{\sigma}}(s^{in}) \to 0$ since $\hat{\sigma}$ is strict. This is a contradiction, so $\tilde{\sigma} \neq \hat{\sigma}$.

Since the solution to the largesse design problem at $\hat{z}$ is assumed to be unique, we conclude that $W(\tilde{\sigma}) < W(\hat{\sigma})$. However, $\hat{\gamma} = 0$, so $\hat{\sigma}$ is feasible for all z. Hence $W(\hat{\tilde{\sigma}}^n) \geq W(\hat{\sigma})$, and by continuity of W this implies $W(\tilde{\sigma}) \geq W(\hat{\sigma})$. This contradiction concludes the proof. □

5.3. Utility Transformations

It is useful to know when we can transform payoffs without changing the solution to the problem. Again, the next result is fairly obvious, but useful to have stated.

Proposition 5.7. *Consider the largesse design problem transformed by $A^i, \beta^i > 0, v^i(s^{-i})$ with transformed utility $\tilde{u}^i(s^i, s^{-i}) = A^i + \beta^i u^i(s^i, s^{-i}) + v^i(s^{-i})$ and largesse $\tilde{\gamma}_\tau = \beta^i \gamma_\tau$. The incentive compatible set of the transformed problem is the same as the original problem. If $\beta^i = \beta^j$ and $v^i(s^{-i}) = 0$, then the set of solutions in the transformed problem is that same as in the original problem.*

Proof. The transformed incentive constraints are

$$A^i + \beta u^i(\sigma^i_\tau, \tilde{\sigma}^{-i}) + \beta^i \gamma_\tau \geq A^i + \beta^i u^i(s^i, \tilde{\sigma}^{-i})$$

which is equivalent to the original. When $\beta^i = \beta^j$ and $v^i(s^{-i}) = 0$ the transformed objective function is

$$\sum_{i=1}^{n} \sum_{\tau \in T^i} \phi_\tau (A^i + \beta u(\sigma^i_\tau, \tilde{\sigma}^{-i}))/n = \sum_{i=1}^{n} \sum_{\tau \in T^i} \phi_\tau A^i/n + \beta W(\sigma)$$

equivalent to the original. □

Note the importance of $\beta^i = \beta^j$ for the solutions to remain the same. Using different multipliers for different players implicitly changes the utility weights on different players, and results in a different social welfare function. Equivalently, this is saying that utility for different players must be measured in compatible units.

5.4. Monotonicity

Another obvious, but useful, result is that increasing largesse can only increase welfare.

Proposition 5.8. *Fix a normal form game and suppose that $\gamma' \geq \gamma$. Then the welfare W' from the solution of the γ' problem satisfies $W' \geq W$, where W is the welfare from the solution of the γ problem.*

Proof. The solution of the γ' problem is feasible in the γ problem. □

Clearly, if largesse is large enough, the first best is attainable.

Proposition 5.9. *Generically there is a unique first best profile, which is in pure strategies. Denote the profile in which all types play the first best by $\hat{\sigma}$. Then the solution to the largesse design problem is first best if and only if $\gamma \geq g(\hat{\sigma})$.*

Proof. Generically each pure profile has a different level of welfare, so there is a unique welfare maximizing pure strategy. This is welfare maximizing also over mixed strategies, since mixed strategies induce a probability distribution over the entire profile space. Clear then, the first best is attainable if and only if $\hat{\sigma}$ is incentive compatible. □

5.5. Two Player Games

Proposition 5.10. *In a two player game either a constraint binds or the outcome is the first best.*

Proof. If no constraint binds then the outcome must be an interior local welfare maximum. Since the objective function is quadratic, an interior local maximum is a global maximum. □

5.6. Uniqueness

Say that Z allows *utility perturbations* if for $z = (u, \gamma, \phi) \in Z$ there is an open neighborhood O of u such that $(O, \gamma, \phi) \subseteq Z$. Let $\overline{\hat{F}}$ denote the projection of $\hat{F}$ onto the aggregate strategies $\overline{\sigma}$ derived from the type strategies σ.

Theorem 5.11. *Suppose that Z allows utility perturbations. Then $\overline{\hat{F}}(z)$ is continuous function (a singleton) at every point of the complement of a (relatively) closed, lower-dimensional, semi-algebraic subset of Z.*

This cannot be true in $\hat{F}$. Consider the punisher's dilemma game where $1 > \overline{\gamma} > 1/3$. Suppose that there are two types with generic γ_τ close to $\overline{\gamma}$. Since the constraint for player 2 is slack, we are free to shift the burden of punishment slightly between the two types without violating the incentive constraints. Perturbing payoffs, and so forth, a small amount will not change this. In other words, the behavior of a player is generically unique, but behavior of each type need not be.

Proof. The idea of the proof is to prove that if there are multiple solutions in $\overline{\hat{F}}(z)$ then $\overline{\hat{F}}(z)$ fails to be lower-hemi continuous. The result then follows from Proposition 5.4. That proposition asserts that at every point of the complement of a (relatively) closed, lower-dimensional, semi-algebraic subset of Z the correspondence $\hat{F}$ is continuous, hence so is the projection $\overline{\hat{F}}$.

To show that when there are multiple solutions in $\overline{\hat{F}}(z)$ then $\overline{\hat{F}}(z)$ fails to be lower-hemi continuous, it suffices to show that a solution can be perturbed away. That is, it suffices to exhibit a sequence $z^m \to z$, a $\overline{\sigma} \in \overline{\hat{F}}(z)$, and a closed (and therefore compact) set H not containing $\overline{\sigma}$ such that $\overline{\sigma}^m \in \overline{\hat{F}}(z^m)$ implies $\overline{\sigma}^m \in H$.

For an aggregate mixed strategy $\overline{\sigma}$ it will be useful to let $\pi(\overline{\sigma})$ denote the probability distribution over outcomes induced by $\overline{\sigma}$. To carry out the construction, the perturbations z^m will perturb the payoffs of player i by $u^i(s) + \lambda^m v^i(s^{-i})$, where $\lambda^m > 0$ and $\lambda^m \to 0$. Hence, by Proposition 5.7 the feasible sets $F(z^m) = F(z)$ are constant. Let $x(s) = \sum_i v^i(s^{-i})/n$. The perturbed objective function is then $w(\overline{\sigma}, z^m) = w(\overline{\sigma}, z) + \lambda^m x \cdot \pi(\overline{\sigma})$.

Suppose that $\overline{\hat{F}}(z)$ has at least two points, $\overline{\hat{\sigma}} \neq \overline{\tilde{\sigma}}$. Then for some player i and $\hat{s}^i \in S^i$ it must be that $\overline{\hat{\sigma}}^i(\hat{s}^i) > \overline{\tilde{\sigma}}^i(\hat{s}^i)$. For $j \neq i$ take

$$v^j(s^{-j}) = \begin{cases} 1 & s^i = \hat{s}^i \\ 0 & s^i \neq \hat{s}^i \end{cases}$$

and take $v^i(s^{-i}) = -(n-1)\overline{\hat{\sigma}}^i(\hat{s}^i)$. Then $x \cdot \pi(\overline{\sigma}) = ((n-1)/n)\left(\overline{\sigma}^i(\hat{s}^i) - \overline{\hat{\sigma}}(\hat{s}^i)\right)$. In particular, $x \cdot \pi(\hat{\sigma}) = 0$ and $x \cdot \pi(\tilde{\sigma}) < 0$. The idea is to reward other players when player i plays $\hat{s}^i$ so as not to change incentive constraints, while making $\hat{s}$ more welfare attractive than $\tilde{s}$.

For the space H take the $\overline{\sigma}$ such that $x \cdot \pi(\overline{\sigma}) \geq 0$. Since π is continuous, this space is closed. Suppose that $\overline{\sigma}^m \in \overline{\hat{F}}(z^m)$. It must be that it arises from $\sigma^m \in F(z^m) = F(z)$, and since σ^m maximizes $w(\overline{\sigma}, z^m)$ in $F(z)$ and $\hat{\sigma} \in F(z)$

$$w(\overline{\sigma}^m, z) + \lambda^m x \cdot \pi(\overline{\sigma}^m) \geq w(\hat{\sigma}, z) + \lambda^m x \cdot \pi(\hat{\sigma})$$

$$= w(\hat{\sigma}, z) \geq w(\overline{\sigma}^m, z)$$

where the final inequality follows from $\sigma^m \in F(z)$ and $\hat{\sigma} \in \hat{F}(z)$. Hence $\lambda^m x \cdot \pi(\overline{\sigma}^m) \geq 0$ and $\lambda^m > 0$ implies $\overline{\sigma}^m \in H$. □

Note that this result holds if payoffs and largesse are symmetric and only symmetric equilibria are allowed. Here the perturbation must be chosen to preserve symmetry. Rather than picking a player and a pure strategy that has a higher probability under one solution than the other, pick a pure strategy $\hat{s}^i$ that has a higher probability for all players under one common mixed strategy than the other. The perturbation is then that for each player that plays $\hat{s}^i$ all other players are rewarded with a bonus of 1, with the corresponding normalized expected value being subtracted from all player payoffs. The rest of the proof is unchanged.

5.7. Behavior

At bifurcation points behavior changes abruptly. From Proposition 5.4 this happens only on a lower dimensional subset of the parameter space. A model is, however, at best a good idealization. There is no sense in thinking that in reality parameters fall exactly into a lower dimensional set. Near a bifurcation, at best, behavior can be described by a probability distribution over some nearby set of parameters. This means that to a good approximation behavior is described by a point in the convex hull of the limit points of nearby parameters. I have accordingly drawn the "theoretical predictions," in Figures 2.1 and 4.1, with vertical segments at the bifurcation points.

There is a second point that is useful in stating (and proving) theorems about solutions sets. Call a subset $\tilde{Z} \subseteq Z$ *comprehensive* if it is the complement of a (relatively) closed, lower-dimensional, semi-algebraic subset of Z and $\hat{F}(z)$ is a single-valued continuous function on $\tilde{Z}$. From Proposition 5.11 such sets exist if we allow utility perturbations, and they exist in many examples even without utility perturbations. When there is a comprehensive $\tilde{Z}$ it is adequate from the point of view of behavior to describe $\hat{F}(z)$ on $\tilde{Z}$ where it is a continuous function. From a behavioral point of view, the bifurcation set should be filled with the convex hull of limit points of $\hat{F}(z)$

6. Reward Games

I continue to consider unitary largesse, that is, I will assume that each player has only one type. I first consider a class of simplified versions of the trust and gift exchange games, that I shall refer to as reward games. These are 2 × 2 games with the payoff matrix given below in Table 6.1.

	S	R
K	$E, 0$	$E, 0$
I	$0, B$	$r, B-c$

Table 6.1: The Reward Game

The payoff parameters in the parameter space Z are restricted to $B > c > 0, r > E > 0, r \geq c$. The interpretation is this. The first player has an endowment of E. They may either K(eep) their endowment or I(nvest) it with player 2. If they keep the endowment they consume it and player 2 gets nothing. Player 2 may either elect to R(eciprocate) an investment by returning r at a cost of c or S(elfishly) keep the entire investment which has a value to them of B. An example of the reward game is a simplified version of the trust game of Berg, Dickhaut and McCabe (1995). Here the first player has an endowment of $E = 10$. In this simplified version, they can either invest all or none. If the investment is made, the value to player 2 is tripled so that $B = 30$. If they reciprocate, in this simplified version, they can return only $r = 15$ which costs $c = 15$.

6.1. The Reward Game Paradox

Analyzing Nash equilibrium, the unique best response by 2 to a positive probability on I is S and unique best response to S is K, hence the unique Nash equilibrium is KS with welfare E. This is strictly less than the maximum achievable welfare which is at IR and is $r - c + B \geq B$. The dilemma is similar to the Prisoner's dilemma in that there is a unique Nash equilibrium that is welfare sub-optimal, and indeed IR Pareto dominates IR.

In the laboratory, what is seen is that in fact player 1's sometimes play I and player 2's sometimes return R. The usual interpretation is that this is some sort of fairness or reciprocity. Fairness says that since IS is unfair to player 1 player 2 will return something, and this gives an incentive for player 1 to play I. Reciprocity says that player 1 playing I is a kind act and should be rewarded by returning something.

By contrast largesse design says that player 2 plays R some of the time both because it is raises welfare, and because it provides incentives to player 1 to play I further raising welfare.

6.2. The Solution to the Largesse Design Problem

Denote by σ^1_I, σ^2_R the probability that 1 plays I and 2 plays R respectively.

Proposition 6.1. *The set $r > c$ and $\gamma^1\gamma^2 > 0$ is comprehensive and has three regions.*

i. $\gamma^2 < c$ and $\gamma^1 \leq E - \gamma^2 r/c$ then the unique solution to the largesse design problem is

$$\sigma_I^1 = \frac{\gamma^1 c + \gamma^2 r}{cE},$$

$$\sigma_R^2 = \frac{\gamma^2 E}{\gamma^1 c + \gamma^2 r},$$

and

$$2W = E + (\gamma^1 c + \gamma^2 r)\frac{B-E}{cE} + (\gamma^2/c)(r-c).$$

ii. $\gamma^2 < c$ and $\gamma^1 > E - \gamma^2 r/c$, then the unique solution to the largesse design problem is $\sigma_I^1 = 1$, $\sigma_R^2 = \gamma^2/c$, and $2W = B + (\gamma^2/c)(r-c)$.

iii. $\gamma^2 \geq c$, then the unique solution to the largesse design problem is $\sigma_I^1 = \sigma_R^2 = 1$, and $2W = B + r - c$, so the first best is obtained.

Just to emphasize: $\hat{F}(z)$ is single-valued and continuous on the entire set $B > c > 0, r > E > 0, r > c$ and $\gamma^1\gamma^2 > 0$ and there are no bifurcations on this set.

Some insight into how largesse design works can be had by considering the case $\gamma^1 = 0$. In this case, the first player will choose I with positive probability only if $\sigma_R^2 \geq E/r$. Take the case where $\sigma_R^2 = E/r$ and player 1 is indifferent. How can player 2 be induced to play R this frequently? Playing R means a loss of c. However, player 2 chooses to be vulnerable to that loss only $\sigma_R^2 = E/r$ of the time, and player 1 only asks them to consider taking the loss only σ_I^1 of the time. That is, the expected loss to player 2 is $\sigma_I^1(E/r)c$ and the incentive constraint says only that this expected loss must be no greater than γ^2. If player 1 randomizes onto I sufficiently infrequently $\sigma_I^1(E/r)c \leq \gamma^2$ will be satisfied, and the solution is to take $\sigma_I^1(E/r)c = \gamma^2$ (provided this is no greater than 1). This idea, that reducing the frequency with which a player is asked to sacrifice, makes feasible sacrifices by that player, is fundamental to largesse design.

A related fact is that the probability of the first best outcome IR in cases (i) and (ii) is $\sigma_I^1\sigma_R^2 = \gamma^2/c$, that is, it does not depend on γ^1. As γ^1 increases in case (i) σ_I^1 increases, but σ_R^2 must decline to maintain the incentive constraint for player two. Hence, the probability of IS increases, while the probability of IR remains the same.

A final and also related observation is that, in the symmetric largesse case in which $\gamma^1 = \gamma^2$, the probability of player 2 playing R is $\sigma_R^2 = E/(c+r)$, independent of largesse. Here again, the strategy of player 1 must adjust so that the largesse constraint for player 2 is satisfied.

Proof. The incentive constraint for player 1 can be written as the expected loss from I being no greater than γ^1

$$E\sigma_I^1 - \sigma_I^1\sigma_R^2 r \le \gamma^1,$$

and that for player 2 as the expected loss from R being no greater than γ^2

$$\sigma_I^1\sigma_R^2 c \le \gamma^2.$$

Consider the second player constraint $\sigma_I^1\sigma_R^2 c \le \gamma^2$. If this is strictly satisfied then σ_R^2 can be increased resulting in a weak welfare improvement, while additional slack is weakly added to the first constraint. That is, more R makes I more attractive to player 1. Hence, there is a solution to the largesse design problem in which either $\sigma_I^1\sigma_R^2 c = \gamma^2$ or $\sigma_R^2 = 1$.

Take first the "easy case," $\sigma_R^2 = 1$. Then the first constraint is $E\sigma_I^1 - \sigma_I^1 r \le \gamma^1$. As $E - r < 0$ this is always satisfied, and as σ_I^1 strictly increases welfare with $\sigma_R^2 = 1$ it must be that $\sigma_I^1 = 1$. However, the second constraint is satisfied only when $c \le \gamma^2$. Moreover, if this is the case then indeed IR is a first-best equilibrium. This is case (iii).

Suppose, then, that the second constraint does bind so that $\sigma_I^1\sigma_R^2 c = \gamma^2$. Then the first constraint can be written as

$$E\sigma_I^1 - (\gamma^2/c)r \le \gamma^1.$$

Welfare is given by

$$2W = E + \sigma_I^1(B - E) + (\gamma^2/c)(r - c).$$

This is increasing in σ_I^1, so either $E\sigma_I^1 - (\gamma^2/c)r = \gamma^1$, or $\sigma_I^1 = 1$. The former case is $\gamma^1 \le E - \gamma^2 r/c$, that is, case (i), and the latter case (ii). □

7. Public Goods

I continue to examine unitary largesse. In 2 × 2 games there is no issue about where to spend largesse, but only how much can be achieved by spending it. I now turn to a 3 × 3 class of games in which a decision has to be made on which of the two sub-optimal actions the largesse should be used for. Specifically, I consider a symmetric two player public goods game in which players can choose to free ride, F, contribute to a public good, C, or both contribute to the public good and punish free riders, P. I continue to examine the unitary case, and now also make the obvious symmetry assumption that $\gamma^1 = \gamma^2 = \overline{\gamma}$.

Contributing to the public good has a cost of $v + c$ where $v > c > 0$, while contribute to the public good and punishing free riders has a higher $v + c + d$ where $d > 0$. You can think of d as a simplified model of costs that arise from trembling.

The public good, if produced, it is worth v to each player. This is additive, so if both produce the public good, each gets $2v$ from public good output. If a free rider is punished, they suffer a cost of $p > c$, so that it is better to contribute than to free ride if the opponent is punishing. To avoid an uninteresting case, I assume also that $p > d$.

7.1. Dominance Solvability

Suppose that $\gamma = 0$. Then C strictly dominates P. Once P is eliminated F strictly dominates C. Hence FF is the unique solution of iterated strict dominance.

In the laboratory, what is seen is that in fact free riders are sometimes punished. Once again, the usual interpretation is that this is some sort of fairness or reciprocity. Fairness free riding is unfair, and fairness can be restored through punishment. Reciprocity says that free rising is an unkind act and should be reciprocated by punishment.

By contrast largesse design says that punishment is an effective way to provide incentives for socially desirable contributions.

7.2. Overview of the Solution

The big picture is that at $\overline{\gamma} = 0$ there are only free-riders. As $\overline{\gamma}$ is increased from zero, initially some free riders use their largesse to produce, and welfare increases linearly. If the cost of punishment is low in the sense that $d/(p - d) < (v - c)/c$ then there is a bifurcation*,* a value of $\overline{\gamma} = (d/p)c$, where a new type of equilibrium is possible. In this equilibrium there are no free-riders, but punishers provide the incentives for production, and largesse is used to compensate punishers for the cost of punishment. Welfare jumps up discontinuously.

Bifurcation, the emergence of a new type of equilibrium, is a key feature of models like this. There are two ways in which largesse can be "spent." It can be used directly to compensate for the cost of production or it can be used indirectly to compensate for the providing incentives for others. Once enough largesse is available, if the cost of punishment is not too high, it is better to compensate for providing incentives.

As largesse is increased further, past the bifurcation point, less incentive is needed for producers, and the number of punishers falls, the number of producers increases, and welfare increases, until, when there is enough largesse, $\overline{\gamma} = c$, the first best of $v - c$ is attained, and production $\sigma^1_C = 1$, takes place entirely out of kindness.

7.3. Solution of the Largesse Design Problem

Proposition 7.1. *There is a comprehensive set consisting of three different regions with qualitatively different solutions:*

i. if $\overline{\gamma} < c$ and either $\overline{\gamma} < dc/p$ or $d/(p-d) \geq (v-c)/c$ then $\sigma^1_P = 0$, $\sigma^2_C = \overline{\gamma}/c$, and welfare

$$W = (\overline{\gamma}/c)(v-c).$$

ii. if $dc/p \leq \overline{\gamma} < c$ and $d/(p-d) \leq (v-c)/c$ then $\sigma^1_C + \sigma^1_P = 1$ (no free riders),

$$\sigma^1_P = \frac{c - \overline{\gamma}}{p - d},$$

and welfare

$$W = (v-c) - \frac{c-\overline{\gamma}}{p-d} d.$$

iii. if $\overline{\gamma} \geq c$ then $\sigma^1_C = 1$ and the solution is first best with welfare $W = v - c$.

In particular, if $d/(p-d) < (v-c)/c$ welfare jumps up at $\overline{\gamma} = dc/p$, that is, there is a welfare bifurcation. At this point it becomes possible to entice all the free-riders to contribute and with a low cost of punishment, $d/(p-d) < (v-c)/c$, it is desirable to do so.

Proof. As usual it is useful to dispose of the large largesse, first best, case (iii) first.

If $\overline{\gamma} < c$ then in the solution the constraint must bind and $\sigma^1_C < 1$, while if $\overline{\gamma} \geq c$ then the solution of the largesse design problem is $\sigma^1_C = 1$.

Suppose that the constraint does not bind. In this case shifting a small amount of weight to C increases welfare without violating the constraint. If $\overline{\gamma} < c$ and the constraint does not bind then $\sigma^1_C = 1$. This implies that the best response is F, with a gain of c over C. Hence, if $\overline{\gamma} < c$, the incentive constraint is violated. On the other hand, if $\overline{\gamma} \geq c$ then $\sigma^1_C = 1$ satisfies the incentive constraint, and, as this is the unique first best, it must be the solution.

We may now assume that $\overline{\gamma} < c$ and that the constraint binds with $\sigma^1_C < 1$. Since P is never a best response, this implies that the best response is F. Observe that c and $c + d$ is what is lost by C and P respectively compared to F if there is no punishment, but that both avoid the punishment cost of $\sigma^1_P p$. Hence the incentive constraint is

$$(\sigma^1_C + \sigma^1_P)c + \sigma^1_P d - (\sigma^1_C + \sigma^1_P)\sigma^1_P p = \overline{\gamma}.$$

The expected social surplus from output of the public good is $\sigma^1_C(v-c)+\sigma^1_P(v-c-d)$, while the expected cost of punishment is $\sigma^1(F)\sigma^1_P p$, so welfare is given by

$$W = (\sigma^1_C + \sigma^1_P)(v - c) - \sigma^1_P d - (1 - \sigma^1_C - \sigma^1_P)\sigma^1_P p.$$

The incentive constraint can be rewritten in the convenient form

$$(\sigma^1_C + \sigma^1_P - d/p)(c - \sigma^1_P p) = \overline{\gamma} - dc/p.$$

This highlights the semi-algebraic nature of the constraint discussed in Section 5.1. In particular there are two distinct possibilities. If $c - \sigma^1_P p = 0$ then it must be $\overline{\gamma} = dc/p$, so this is a candidate for a bifurcation point.

Suppose, indeed, that $\sigma^1_P = c/p$ and $\overline{\gamma} = dc/p$. In this case any $\sigma^1_P \leq \sigma^1_C + \sigma^1_P \leq 1$ is feasible. Increasing $\sigma^1_C + \sigma^1_P$ holding fixed σ^1_P preserves the incentive constraint while reducing the number of free-riders: this unambiguously increases welfare, so it must be in this case that $\sigma^1_C + \sigma^1_P = 1$. Welfare is $v - c - dc/p$.

If $\sigma^1_P > c/p$ then C is a best response, which is already ruled out, so the other case is $\sigma^1_P < c/p$, in which case we can solve

$$\sigma^1_C + \sigma^1_P - d/p = \frac{\overline{\gamma} - dc/p}{c - \sigma^1_P p}.$$

If $\overline{\gamma} < dc/p$ then $\sigma^1_C+\sigma^1_P$ is a decreasing function of σ^1_P meaning that as shift players from C to P the number of free-riders actually increases, so this is unambiguously bad for welfare. Hence $\sigma^1_P = 0$ and we can solve the incentive constraint to find $\sigma^1_C = \overline{\gamma}/c$.

If $\overline{\gamma} \geq dc/p$ then $\sigma^1_C+\sigma^1_P$ is an non-decreasing function of σ^1_P. We plug the solution of the incentive constraint into the objective function to find

$$W = \frac{\overline{\gamma} - \sigma_P^1 d}{c - \sigma_P^1 p} v - \gamma - \sigma_P^1 p.$$

The first derivative is

$$\begin{aligned} \frac{dW}{d\sigma_P^1} &= \left(\frac{-d\,(c - \sigma_P^1 p)}{(c - \sigma_P^1 p)^2} + \frac{(\overline{\gamma} - \sigma_P^1 d)\, p}{(c - \sigma_P^1 p)^2} \right) v - p \\ &= \left(\frac{\overline{\gamma} p - dc}{(c - \sigma_P^1 p)^2} \right) v - p \end{aligned}$$

The second derivative is

$$\frac{d^2 W}{d(\sigma_P^1)^2} = 2 \left(\frac{\overline{\gamma} p - dc}{(c - \sigma_P^1 p)^3} \right) pv.$$

Since $\overline{\gamma} \geq dc/p$ and $c - \sigma_P^1 p < 0$ the function is weakly convex. Hence we must check the endpoints to see where the welfare maximum is.

Since $\sigma_C^1 + \sigma_P^1$ is an non-decreasing function of σ_P^1 the endpoints are where $\sigma_C^1 + \sigma_P^1 = 1$ and where $\sigma_P^1 = 0$. In the former case

$$\sigma_P^1 = \frac{c - \overline{\gamma}}{p - d},$$

and in the latter $\sigma_C^1 + \sigma_P^1 = \overline{\gamma}/c$. Plugging into welfare gives

$$W_P = (v - c) - \frac{c - \overline{\gamma}}{p - d} d$$

in the former case and

$$W_C = (\overline{\gamma}/c)(v - c).$$

in the latter. The condition $W_P > W_C$ can be written as

$$\frac{d}{p - d} < \frac{v - c}{c},$$

meaning the cost of punishment relative to the size of punishment is not too great. □

8. Leveraging Reputation

Commitment plays a crucial role in reputational models. That is, while, in a sense, the goal of these models is to show how reputation can substitute for commitment, underlying it is a small probability that some players are actually committed. This low probability of commitment is then bootstrapped to show that non-committed players never-the-less act committed to keep a good reputation. In the context of largesse design, players can commit, but they can incur only limited losses from doing so. My goal here is to explain how a few committed players with limited and possibly very small largesse can act like "behavioral types" in reputational models to provide good solutions to the largesse design problem.

A simple setting for studying reputation are the reward games of Section 6, with payoffs shown again below in Table 8.1, where recall that $B > c > 0, r > E > 0$ and now $r > c$.

	S	R
K	$E, 0$	$E, 0$
I	$0, B$	$r, B - c$

Table 8.1: The Reward Game

Notice that player 2 would like to commit to R to force the first best outcome IR. However, while limited commitment is available due to largesse, as discussed in Section 6, I want to examine what happens when the game is played T times between a patient player 2 who receives the time average payoff and a sequence of short-run player 1's. I assume that every player 1 has the same distribution of largesse. The main result is then this:

Proposition 8.1. *Fix any largesse distribution such that for some τ both $\phi^2_\tau > 0$ and $\gamma^2_\tau > 0$. Let W_T denote the solution to the largesse design problem. Then $\lim_{T\to\infty} W_T \to r + B - c$, the first best.*

Proof. Note that certainly $W_T \leq r + B - c$.

By Proposition 5.8 it suffices to prove this in the special case where player 1 is selfish and player 2 is either type τ or selfish. That is, there is a single type of player 1 with $\gamma^1 = 0$ and two types of player 2, one with $\phi^2_\tau, \gamma^2_\tau$ and the other with $\phi^2_\sigma = (1 - \phi^2_\tau), \gamma^2_\sigma = 0$.

Rather than trying to compute the solution to the largesse design problem, I use the technique developed Fudenberg and Levine (1989) for analyzing reputational models to get a lower bound on welfare. This is done by showing that type τ can find a largesse feasible strategy that gives nearly first best welfare, when the selfish players all best respond. The idea is similar to that used in the reputational literature, that is, type τ (with high probability) forever plays the "Stackelberg" action of R. From Fudenberg and Levine (1989) this implies $\lim_{T\to\infty} W_T \to r + B - c$. However, it is necessary to show that this strategy also satisfies the largesse constraint. This requires an upper as well as the usual lower bound on the payoffs of the "rational type," here the selfish type of player 2.

As indicated, type τ with high probability plays R always. In addition, however, with positive probability type τ also plays each strategy of the form, play R until t then play S. The latter part of type τ's strategy makes it easy to get an upper bound on the payoff of the selfish type of player 2. Specifically, it forces the short-run players, after observing S,to conclude either that they face a selfish type or a type committed to subsequently playing S. Hence, it is a Nash equilibrium for the selfish player always to play S and the short-run player always to play K after observing S. With this short-run player strategy, the selfish type of player 2 must play R, except, possibly, for some fixed number of periods near the end of the game, where they switch to S.

Since payoff is time average, the loss to type τ playing R during the final periods goes to zero as $T \to \infty$, so the largesse constraint will be satisfied. □

I want to emphasize that the *ex ante* nature of the largesse constraint is here again crucial. That is, type τ of player 2 has a limited amount of largesse, but they can spend it in whatever periods they choose. Hence, they are free to lose in a few periods, provided they do not lose in too many. This is closely connected to way in which Radner (1980) exploits ϵ-equilibrium to generate cooperation in the finitely repeated prisoner's dilemma game.

9. Conclusion

The largesse design problem is economically relevant, but involves maximizing a function that may not be concave over a set that need not be convex, and which need not be lower-hemi continuous in parameters. Despite this, solutions exist and are well behaved, generically being single valued and continuous. In simple examples, it is not difficult to characterize solution sets analytically.

From an economic perspective, what these examples show, is that largesse design is about being opportunistic. It is about using low cost punishments, rewards, and reputation to encourage other players to pro-social behavior, rather than blindly acting in a pro-social way. The theory of largesse design provides a sensible and tractable alternative to psychological theories of preferences, for explaining non-selfish behavior inside and outside the laboratory.

References

Andrei, J. (1990): “Impure altruism and donations to public goods: A theory of warm-glow giving,” *Economic Journal* 100: 464-477.

Battalio, R., L. Samuelson and J. Van Huyck (2001): “Optimization Incentives and Coordination Failure in Laboratory Stag Hunt Games,” *Econometrica* 69: 749-764.

Berg, Joyce, John Dickhaut and Kevin McCabe (1995): “Trust, Reciprocity, and Social History,” *Games and Economic Behavior* 10: 123, 122-142.

Plume, L. E. and W. R. Zane (1994): ”The algebraic geometry of perfect and sequential equilibrium,” *Econometrica* 783-794.

Bolton, G. E., and A. Cockerels (2000): “REC: A theory of equity, reciprocity, and competition,” *American Economic Review* 91: 166-193.

Carlson, H. and E. van Dame (1993): “Global Games and Equilibrium Selection,” *Econometrica* 61: 989-1018.

Charness, G. (1998): “Attribution and Reciprocity in a Simulated Labor Market,” Unpublished manuscript, Pompey Sabra.

Case, R. H. (1960): “The Problem of Social Cost,” *Journal of Law and Economics* 3: 1-44.

Cote, Michel (2002): *An Introduction to Semi algebraic Geometry*, mimeo, Institute DE Recherche Mathematics DE Rennet

Cote, S. and M. Colin (2004): “A Group Rule-Utilitarian Approach to Voter Turnout: Theory and Evidence,” *American Economic Review* 94: 1476-1504.

Cox, J. C. and D. James (2015): “On replication and perturbation of the McKelvey and Palfrey Centipede game experiment,” In *Replication in Experimental Economics* (pp. 53-94). Emerald Group Publishing Limited.

Dal B, P., G. R. Brochette and J. Kim (2021): “The determinants of efficient behavior in coordination games,” *Games and Economic Behavior* 130: 352-368.

Dufwenberg, M., and G. Kirchsteiger (2004): “A theory of Sequential Reciprocity,” *Games and Economic Behavior* 47: 268-298.

Dutta, R., D. K. Levine and S. Modica (2021): “The Whip and the Bible: Punishment Versus Internalization,” *Journal of Public Economic Theory* 23: 858-894

Talk, A., and U. Fischer (2006): “A Theory of Reciprocity,” *Games and Economic Behavior* 54: 293-315.

Defense, T., A. Sandra (2006): “A Theory of Participation in Elections,” *American Economic Review* 96: 1271–1282.

Fehr, E. and G. Charness (2024): “Social Preferences: Fundamental Characteristics and Economic Consequences,” *Journal of Economic Literature*, forthcoming.

Fehr, Ernst and Klaus M. Schmidt (1999): “A Theory of Fairness, Competition and Cooperation”, *Quarterly Journal of Economics* 114: 817-868..

Fudenberg, D. and D. K. Levine (1997): “Measuring Players’ Losses in Experimental Games,” *Quarterly Journal of Economics* 112: 507-536.

Fudenberg, D. and D. K. Levine (1993): “Self-Confirming Equilibrium,” *Econometrica* 61: 523-546

Fudenberg, D. and D. K. Levine (2011): “Risk, Delay, and Convex Self-Control Costs,” *AEJ Micro* 3: 34–68.

Fudenberg, D. and D. K. Levine (2012): “Fairness and Independence: An Impossibility Theorem,” *Journal of Economic Behavior and Organization*, 81: 606-12

Fudenberg, D. and D. K. Levine (1989): “Reputation and Equilibrium Selection in Games with a Patient Player,” *Econometrica* 57: 759-778.

Fudenberg, D., D. Levine and E. Masking (1994): “The Folk Theorem With Imperfect Public Information,” *Econometrica* 62: 997-1039.

Garner, Richard T. and Bernard Rose (1967): *Moral Philosophy: A Systematic Introduction to Normative Ethics and Meta-ethics* New York: Macmillan.

Mindanao, S. and R. Wilson (2009): “On forward induction,” *Econometrica* 77: 1-28.

Harsanyi, J. C. (1982): *Rule utilitarianism, rights, obligations and the theory of rational behavior*, Springer.

Harsanyi, John C. and Reinhardt Tense (1988): *A General Theory of Equilibrium Selection in Games*, MIT Press.

Herrera, H., M. Morelli, M. and S. Nunnari (2016): “Turnout Across Democracies,” *American Journal of Political Science* 60: 607-624.

Kandori, M., F. Mailath and R. Rob (1993): “Learning, Mutation, and Long Run Equilibria in Games,” *Econometrica:* 29-56.

Kreps, D. and R. Wilson (1982): “Reputation and Imperfect Information,” *Journal of Economic Theory* 50: 253-79.

Cockcrow, E. M., A. M. Coleman, and B. D. Wilford (2016): “Cooperation in repeated interactions: A systematic review of Centipede game experiments, 1992–2016,” *European Review of Social Psychology* 27: 231-282.

Levine, D. K. (1986): “Modeling altruism and spitefulness in experiments,” *Review of Economic Dynamics* 1: 593-622.

Levine, D. K. (2024): “Behavioral Mechanism Design in the Repeated Prisoner’s Dilemma,” Levine’s Working Paper Archive

Levine, D. K. (2025): “Behavioral Mechanism Design as a Benchmark for Experimental Studies,” mimeo RHUL.

Levine, D. K. [2024]: "Method of Moments and Maximum Likelihood in the Laboratory," RHUL

Levine, D. K. and A. Mattozzi (2020): "Voter Turnout with Peer Punishment," forthcoming, *American Economic Review.*

Levine, D. K., A. Mattozzi and S. Modica (2022): *Social Mechanisms and Political Economy: When Lobbyists Succeed, Pollsters Fail and Populists Win*, mimeo RHUL.

Maniadis, Zacharias (2011): “Aggregate Information and the Centipede Game: a Theoretical and Experimental Study,” University of Southampton

McKelvey, R. D. and T. R. Palfrey (1992): “An Experimental Study of the Centipede Game,” *Econometrica*: 803-836.

McKelvey, R. D. and T. R. Palfrey(1995): “Quantal response equilibria for normal form games,” *Games and Economic Behavior* 10: 6-38.

McKelvey, R. D. and T. R. Palfrey (1998): “Quantal response equilibria for extensive form games,” *Experimental Economics* 1: 9-41.

Milgrom, P. and J. Roberts (1982): “Predation, reputation, and entry deterrence,” *Journal of Economic Theory* 27: 280-312

Mill, J. S. (1861): *Utilitarianism.*

Ostrom, Elinor (1990): *Governing the commons: The evolution of institutions for collective action*, Cambridge university press.

Ostrom, E., J. Walker and R. Gardner (1992): “Covenants with and without a sword: Self-governance is possible,” *American Political Science Review* 86: 404-417.

Palfrey, T. R., and J. E. Prisbrey (1996): “Altruism, Reputation and Noise in Linear Public Goods Experiments,” *Journal of Public Economics* 61: 409-427.

Palfrey, T. R. and J.E. Prisbrey (1997): “Anomalous behavior in public goods experiments: How much and why?” *American Economic Review*, 829-846.

Darner, R. (1980): “Collusive behavior in non cooperative epsilon-equilibrium of oligopolies with long but finite lives,” *Journal of Economic Theory* 22: 136-154.

Roemer, J. E. (2010): “Kantian equilibrium,” *Scandinavian Journal of Economics* 112: 1-24.

Rousseau, Jean-Jacques(1754): *Discourse on Inequality.*

Strauss, P. G. (1995): “Risk dominance and coordination failures in static games,” *Quarterly Review of Economics and Finance* 35: 339–363.

Townsend, R. M. (1994): “Risk and insurance in village India,” *Econometrica*, 539-591.

Van Huyck, J. B., R. C. Battalio and R.O. Beil (1990): “Tacit coordination games, strategic uncertainty, and coordination failure,” *American Economic Review* 80: 234-248.

Young, H. P. (1993): “The Evolution of Conventions,” *Econometrica*: 57-84.