Behavioral Mechanism Design in the Repeated Prisoner's Dilemma

David K. Levine[1]

Abstract

The infinitely repeated prisoner's dilemma game provides a paradigm for long-term cooperation in the face of short-term incentives for free riding. However, the extent to which players cooperate in long-term play in repeated indefinite horizon prisoner's dilemma games in the laboratory depends on the specific payoff parameters of the game. Attempts to explain this have employed a variety of *ad hoc* models of individual learning. Here I take the simpler more direct approach of behavioral mechanism design. I hypothesize that there ethical players who actively seek to maximize social welfare and that players tremble. I use a simple benchmark calibration that uses no data from strategic experiments and which is consistent with data from a variety of one-shot games. I show that this simple calibrated model makes sharp predictions and does a good job both qualitatively and quantitatively in explaining experimental data in repeated indefinite horizon prisoner's dilemma games.

[1]Department of Economics, RHUL

Acknowledgements: First version: March 25, 2024. I would like to thank Alessandra Casella, Gary Charness, Michalis Drouvelis, Rohan Dutta, Martin Dufwenberg, Drew Fudenberg, Ernst Fehr, John Mair, Andrea Mattozzi, Agnieszka Mensfelt, Salvatore Modica, Tom Palfrey, Klaus Ritzberger, Larry Samuelson, Arthur Schram, Francesco Squintani, Catharina Tilmans, and Myrna Wooders. The Leverhulme Trust provided financial support for which I am grateful. A particular debt is owed to Pedro Dal Bo and Guillaume Frechette not only for their intellectual contributions to this work, but for providing me with their original experimental data.

1. Introduction

The infinitely repeated prisoner's dilemma game provides a paradigm for long-term cooperation in the face of short-term incentives for free riding. Such an environment is induced in the laboratory by ending the game with a role of the dice. In a repeated indefinite horizon prisoner's dilemma game, players are then rematched with new opponents to play another indefinite horizon prisoner's dilemma game. This enables us to see in the laboratory the long-term play of experienced players - the type of play that is most relevant in most applications outside the laboratory.

Theoretically, when the discount factor is sufficiently high, on account of the folk theorem of Fudenberg and Maskin (1986), any degree of cooperation is possible in equilibrium. However, game theorists have long believed, or at least hoped, that players will find their way to cooperative equilibria. This hope has proven false in experimental studies such as Dal Bo (2005) and Dal Bo and Frechette (2011).

In order to explain how the degree of long-term cooperation depends on the specific payoff parameters of the game researchers, such as Dal Bo and Frechette (2011), Blonski, Ockenfels and Spagnolo (2011), Blonski and Spagnolo (2015) and Fudenberg and Rehbinder (2024), have employed a variety of *ad hoc* models of individual learning. Here I take the simpler more direct approach of behavioral mechanism design. I hypothesize that there ethical players who actively seek to maximize social welfare and that players tremble. I show that this theory does a good job of explaining both qualitatively and quantitatively long-term play in repeated indefinite horizon prisoner's dilemma games.

Behavioral mechanism design draws on two facts about behavior in the experimental laboratory. First, many participants in laboratory experiments strive to achieve a social objective: this is the basis of much work on psychological theories beginning with Levine (1986) and Fehr and Schmidt (1999). Second, participants are sometimes inattentive or follow other objectives than utility maximization: the success of the quantal response model of McKelvey and Palfrey (1995) in explaining experimental data is due to the prevalence of such trembling behavior. The goal of this paper is to show that a systematic combination of social goals with trembling can explain levels of cooperation in repeated prisoner's dilemma experiments.

This paper builds on earlier work in Levine (2024). That paper developed a simple benchmark model of behavioral mechanism design. It showed that, when this model is calibrated to data on play in non-strategic settings, the model successfully predicts out-of-sample welfare for a variety of one-shot games. These games include stag hunt, ultimatum bargaining, and public goods games with punishment. Here I use this same calibration to predict welfare and cooperation in the repeated indefinite horizon prisoner's dilemma experiments of Dal Bo and Frechette (2011).

As indicated, in behavioral mechanism design, players are assumed to tremble. As in Levine (2024) there is a 1/3rd chance that the a player trembles by playing uniformly with equal chance of cooperating or defecting. An indefinite horizon prisoner's dilemma game is then one of imperfect public signals and I study the strongly symmetric equilibria of Abreu, Pearce and Stacchetti (1986). Following Levine (2024) I look for the welfare maximum among these equilibria.

I apply this benchmark calibration to study long-run behavior in the six repeated prisoner's dilemma treatments of Dal Bo and Frechette (2011). In five of the six treatments the benchmark calibration predicts welfare to within a single penny. In the sixth the calibration fails.

I then study the anomalous sixth case. While other theories do not provide a correct qualitative prediction of welfare either, one learning based theory does provide a correct qualitative prediction about cooperation rates. With this in mind I propose that in the anomalous sixth treatment it is difficult for selfish players to learn the mechanism the ethical players are trying to teach them. Instead, the ethical players provide a second best mechanism in which they provide incentives only for each other. This predicts welfare in the anomalous sixth case to within a penny. I confirm this theory with an out of sample study of a public goods contribution game with punishment.

2. The Experimental Game

The experimental game and data are from Dal Bo and Frechette (2011). A population of players is randomly matched to play a repeated prisoner's dilemma game with probability $\delta \in \{1/2, 3/4\}$ of continuing each period. Payoffs in pennies are given by $U \in \{32, 40, 48\}$ and the payoff matrix in Table 2.1 below.

	C	D
C	U, U	$12, 50$
D	$50, 12$	$25, 25$

Table 2.1: Prisoner's Dilemma Payoffs

3. The Benchmark Model

Denote the pure strategies in the repeated game for player $i \in \{1,2\}$ by $s^i \in S$ and mixed strategies by σ^i. Players are risk neutral. In each period player i has a 2/3rd chance of playing according to their intended strategy σ^i and a 1/3rd chance of trembling according to an iid uniform random variable. The solution concept is the strongly symmetric equilibrium that maximizes *ex ante* social welfare. This requires that players always intend to take the same action as each other and that the signals DC and CD are punished in the same way.

Related Literature

The analysis here is closely related to that in Levine (2024). There are four differences.

First, players here "tremble" rather than draw "noise types." As shown in the earlier paper this does not matter.

Second, in Levine (2024), there were ethical players who were willing to sacrifice an amount γ (their *largesse*) for the common good. As the dollar equivalent of γ is $\$1.00/T$ where $T \approx 160$ is the number of times a player gets to play over the course of the experiment it is reasonable to ignore this in studying the Dal Bo and Frechette (2011) data. Moreover, when cooperation is supported as a strongly symmetric equilibrium largesse can play no role: when both players are cooperating an ethical player can do no more, and the punishment equilibrium is calibrated to be as generous as possible consistent with incentive compatibility for the selfish agents. Any effort by ethical players to improve on the punishment equilibrium will undermine the entire equilibrium. However, under certain circumstances, the largesse of the ethical players in the form of γ can play a role in promoting cooperation: how and why this is examined at the end of the paper.

Third, risk aversion is ignored, as the stakes are, as shown in the earlier paper, too small for it to play a significant role.

Finally, attention is restricted to strongly symmetric equilibria an assumption that has no analog in the earlier calibrations. It is chosen because it is the only tractible case, and it is unlikely that even in 40 to 80 attempts players in the laboratory could realistically hope to discover, learn, and teach each other a mechanism that relies on asymmetric punishments based on scoring schemes. Moreover, the advantage of such schemes with these relatively low discount factors is not great. Readers interested in learning about other types of equilibria are referred to Fudenberg, Levine and Maskin (1994), Mailath and Samuelson (2006) and Sannikov (2007).

Payoffs with Trembles

The general form of the payoff matrix with trembles is below in Table 3.1 and the specific matrices corresponding to the different games are Table 3.2

	C	D
C	u, u	$n-\ell, u+g$
D	$u+g$,$n-\ell$	n, n

Table 3.1: Prisoner's Dilemma Perturbed Payoffs General From

	C	D		C	D		C	D
C	$32, 32$	$18, 43$	C	$37, 37$	$19, 44$	C	$43, 43$	$20, 45$
D	43,18	$27, 27$	D	44,19	$27, 27$	D	45,20	$27, 27$

Table 3.2: Prisoner's Dilemma Perturbed Payoffs: $U = 32, 40, 48$

Strongly Symmetric Equilibrium

I now solve to find the best strongly symmetric equilibrium. It is convenient to define the probability of a deviation due to trembling as $\pi = 1/6$. Strongly symmetric equilibrium makes use of three signals: CC, DD and DC/CD. If the actual play is CC the probability of the DD signal is $p_{DD} = \pi^2$ and the probability of the DC/CD signal is $p_{DC} = 2\pi(1-\pi)$. Similarly if actual play is DC the probability of the DD signal is $q_{DD} = (1-\pi)\pi$ and the DC/CD signal is $q_{DC} = (1-\pi)^2 + \pi^2$.

The equilibrium is defined by the punishment issued for the "bad" signals DD and DC/CD and the equilibrium average present value. Denote by P_{DD} the average present value punishment for DD and P_{DC} that for DC/CD signal. Equilibrium average present value is denoted by v.

The average present value v in the best strongly symmetric that supports cooperation is characterized by solving the LP problem of choosing v, P_{DD}, P_{DC} to maximize v subject to the two incentive constraints

$$v = (1-\delta)u + \delta((1 - p_{DD} - p_{DC})v + p_{DD}(v - P_{DD}) + p_{DC}(v - P_{DC}))$$

$$v = (1-\delta)(u+g) + \delta((1 - q_{DD} - q_{DC})v + q_{DD}(v - P_{DD}) + q_{DC}(v - P_{DC}))$$

and the feasibility constraints $v - P_{DD}, v - P_{DC} \geq n$ and $P_{DD}, P_{DC} \geq 0$. Note that because $\delta \geq 1/2$ and there are two states there is no need for a public randomizing device: any feasible punishment can be attained by a deterministic (intended) sequence of play by the two players.

Define the relative gain to cooperation

$$R \equiv \frac{u-n}{g}$$

the interest rate

$$\rho \equiv \frac{1-\delta}{\delta}$$

and the noisy interest rates

$$\eta_{DD} \equiv \frac{\rho + p_{DD}}{q_{DD} - p_{DD}}, \eta_A \equiv \frac{\rho + p_{DD} + p_{DC}}{(q_{DD} + q_{DC}) - (p_{DD} + p_{DC})}.$$

.

The LP problem is solved in Appendix I. The solution is given by

Proposition 3.1. *If $\eta_A \leq R < \eta_{DD}$ and $q_{DC} - ((1+\eta_{DD})/\eta_{DD})p_{DC} \leq 0$ the solution is the the static Nash equilibrium $P_{DC} = P_{DD} = 0$ and $v = n$. Otherwise it is given below*

	P_{DC}	v	P_{DD}
$R < \eta_A$	0	n	0
$\eta_A \leq R < \eta_{DD}$	$\rho\left(\frac{1-\eta_{DD}R}{q_{DC}-p_{DC}-\eta_{DD}p_{DD}}\right)g$	$u - \left(\frac{p_{DD}(u-n)+p_{DC}P_{DC}}{\rho+p_{DD}}\right)$	$v - n$
$R \geq \eta_{DD}$	0	$u - \left(\frac{p_{DD}}{q_{DD}-p_{DD}}\right)g$	$\left(\frac{\rho}{q_{DD}-p_{DD}}\right)g$

Table 3.3: Best Strongly Symmetric Equilibrium

Note that the columns are ordered so that the solutions in each column depend only on the results from the previous columns.

To gain a little intuition, observe that the efficiency of using a signal $K \in \{DD, DC\}$ for punishment is measured by $(q_K - p_K)/p_K$. This is greatest for DD so this signal should be used until the constraint binds. In the intermediate case

$\eta_A \leq R < \eta_{DD}$ this constraint does bind so it is necessary also to punish on the less efficient signal DC.

4. Theory vs. Data

From the theory the best equilibrium average present value expected utility (welfare) can be computed. Below in Table 4.1 this is reported in cents along with the actual average present value welfare for all matches starting with the tenth. Note that starting with the tenth match is to assure that the players are experienced and is consistent with past practice including Levine (2024). The RD column is discussed subsequently.

δ	U								
	32			40			48		
	theory	data	RD	theory	data	RD	theory	data	RD
1/2	26.9	25.9	25	27.1	27.1	25	**41.6**	**30.9**	48
3/4	26.9	27.2	25	33.5	34.2	40	42.0	43.0	48

Table 4.1: Welfare

Note that when $\delta = 3/4, U = 40$ the CC strategy yields $u = 37.1$ so that the loss from punishment in that case is non-trivial. By contrast when $\delta = 3/4, U = 48$ the CC strategy yields $u = 42.6$ so that the loss from punishment is small.

The bottom line here is that in five out of six cases the theory does extremely well predicting welfare (within one cent). Moreover it gets a key and non-trivial comparative static prediction right: it predicts that when $\delta = 3/4$ increasing U from 40 to 48 should increase welfare by 8.5 while the actual increase is 8.8.

There is one glaring anomaly: $\delta = 1/2, U = 48$ where according to the benchmark players should have been able to find an equilibrium earning more than ten cents greater than they actually got. Notice that they played around 160 times so over the entire session this is a substantial amount of money. Notice, moreover, that players have not simply reverted to static Nash equilibrium with noise: that payoff is 27.3, which, although a more accurate prediction of welfare than the benchmark, is not particularly close.

Risk Dominance

The best subgame perfect equilibrium (in the usual sense) in all the games except for $\delta = 1/2, U = 32$ has all players cooperating on the equilibrium path. This theory does not do well and as an alternative risk dominance was introduced by Blonski and Spagnolo (2015) and was also used by Dal Bo and Frechette (2011). It is assumed that players consider only two strategies: always defect or grim-trigger. Grim trigger cooperates in the first period and cooperates if both players cooperated last period, otherwise defects. The best response against a population playing these two strategies with equal probability is called risk dominant. The RD column in Table 4.1 shows the conclusions of that theory: it does better than the best subgame perfect equilibrium but does not do particularly well.

The concept of risk dominance has been weakened to propose measures of risk dominance and hypothesize that higher values of this measure result in a better outcomes. Two such measures have been proposed: the basin β and a discount factor based cutoff Δ^{RD}. To define these denote the static Nash utility as $N = 25$, the gain to defecting against cooperation as $G = 50 - U$ and the loss to cooperating against defect as $L = 13$.

The basin is defined by

$$\beta = \frac{(1-\delta)G - \delta(U-N)}{(1-\delta)(G-L) - \delta(U-N)}.$$

It measures the share of always defect players in the population that makes a player indifferent between always defect and grim trigger. Hence the criterion for risk dominance is $\beta > 1/2$. This measure was introduced by Dal Bo and Frechette (2011) and was also used in Fudenberg and Rehbinder (2024).

The discount factor cutoff is defined by

$$\Delta^{RD} = \delta - \frac{G+L}{U-N+G+L}.$$

It measures the difference between the discount factor and the discount factor that makes a player indifferent between always defect and grim trigger when the facing a population with equal probability of playing the two strategies. Hence the criterion for risk dominance is $\Delta^{RD} > 0$. This measure was introduced by Blonski, Ockenfels and Spagnolo (2011) who suggest it is a empirically better than the basin. It was also used used by Fudenberg and Rehbinder (2024) to explain learning about first period play. They indicate it does indeed work better than the basin of attraction β.

Relative Welfare

To measure the quality of an outcome observe that welfare levels differ between games because of changes in behavior but also because of changes in payoffs: cooperation is much less valuable when $U = 32$ than when $U = 48$, for example. Hence to examine qualitative theories it makes sense to introduce a measure of *relative welfare*: the increase in welfare achieved divided by the greatest possible such increase. If w denotes welfare, then relative welfare is $\alpha \equiv (w - N)/(U - N)$. The empirical value of this quantity for all matches starting with the tenth is denoted by $\hat{\alpha}$.

Table 4.2 below summarizes the situation where the α column is the computed for the benchmark behavior mechanism design analysis. Negative values of Δ^{RD} or $1/2 - \beta$ imply always defect is risk dominant and conversely.

δ	$U = 32$			
	Δ^{RD}	$1/2 - \beta$	α theory	$\hat{\alpha}$ data
1/2	−0.32	−0.50	0.27	0.13
3/4	**−0.07**	**−0.31**	**0.27**	***0.31***

δ	$U = 40$			
	Δ^{RD}	$1/2 - \beta$	α theory	$\hat{\alpha}$ data
1/2	−0.11	−0.22	0.14	0.14
3/4	0.14	0.23	0.57	0.61

δ	$U = 48$			
	Δ^{RD}	$1/2 - \beta$	α theory	$\hat{\alpha}$ data
1/2	**0.11**	**0.12**	**0.72**	***0.26***
3/4	0.36	0.34	0.74	0.79

Table 4.2: Relative Welfare

From the perspective of relative welfare the benchmark theory does slightly worse than it does for absolute welfare: for the $\delta = 1/2, U = 32$ game it gets a value of $\alpha = 0.27$ that is much larger than the empirical relative welfare of $\alpha\hat{}$=0.13. There is not much reason for concern over this: players only lose a penny compared to the best mechanism and few mechanism designers would be concerned with improving a mechanism to squeeze out an extra penny.

The basin theory, as indicated in Fudenberg and Rehbinder (2024), does slightly worse than Δ^{RD} : it suggests that $\hat{\alpha}$ should be smaller at $\delta = 3/4, U = 32$ than at $\delta = 1/2, U = 40$ and it is not. All three theories also have an anomaly at $\delta = 1/2, U = 48$. The bold highlighting shows why: both qualitative theories and mechanism design agree that α should be higher in the anomalous case when $\delta = 1/2, U = 48$ than when $\delta = 3/4, U = 32$. It is not.

Cooperation Rate

While relative welfare α is a measure of "the quality of the outcome" it is not the only one: a more traditional measure is the cooperation rate φ. This differs from relative welfare in the way that off diagonal outcomes DC/CD are scored. Both measures agree that DD counts 0 and CC counts 1, but the cooperation rate scores DC/CD as 0.5 while relative welfare scores it as $0.86, 0.33$ or 0.26 respectively for the three payoff matrices. The cooperation rate is not clearly a superior measure: it scores fifty-fifty between CC and DD the same as fifty-fifty between DC and CD although most would agree that these are rather different outcomes.

Cooperation rates are reported below in Table 4.3. The value φ is computed from the benchmark behavioral mechanism design model while $\hat{\varphi}$ is the data from the tenth and subsequent rounds. All six treatments are reported along with the corresponding Δ^{RD}.

With respect to the cooperation rate the basin and Δ^{RD} do not suffer a qualitative anomaly between $\delta = 3/4, U = 32$ ($\Delta^{RD} = -0.07$) and $\delta = 1/2, U = 48$ ($\Delta^{RD} = 0.11$). The reason for this is that while the former has a higher relative welfare of 0.31 than the latter's of 0.26 it has a lower cooperation rate of 0.21 than the latters 0.34. The reason for the discrepency is that from a welfare point of view DC/CD are almost as good as CC when $U = 32$, while when $U = 48$ the outcomes DC/CD are not all that much better than DD.

The second anomaly involving the basin remains: while the cooperation rate should be smaller at $\delta = 3/4, U = 32$ than at $\delta = 1/2, U = 40$ in fact it is 0.21 at the former but only 0.17 at the latter.

With respect to cooperation rates Δ^{RD} is qualitatively correct in all cases as shown below in Table 4.3.

Δ^{RD}	-0.32	-0.11	-0.07	0.11	0.14	0.36
$\hat{\varphi}$	0.08	0.17	0.21	0.35	0.65	0.83
φ	0.17	0.17	0.17	0.79	0.60	0.81

Table 4.3: Cooperation Rates

The bottom line is that there is a theory, Δ^{RD}, and there is a measure of behavior, $\hat{\varphi}$, which for which the qualitative predictions of the theory about behavior are correct while the benchmark theory fails qualitatively (by any measure) to predict what happens at $\delta = 1/2, U = 48$.

5. *Ad Hoc* Learning

Risk dominance is derived from evolutionary game theory, such as Kandori, Mailath and Rob (1993) and Young (1993), which suggests that in 2×2 games risk dominant outcomes should be observed in the long-run. Fudenberg and Rehbinder (2024)'s simulation study of a simple learning procedures shows that risk dominance measures such as Δ^{RD} are useful in predicting initial round cooperation in the long-run. However, the restriction to two strategies, always defect and strong-grim-trigger, is not innocuous.

To see the issue, consider the anti-grim-trigger strategy of defecting in the first period and defecting forever if any player cooperated in the first period, otherwise cooperating forever. Here is the point: in a population with equal fractions of all three strategies regardless of the discount factor always defect is the strict best response. The idea is that always defect does better against anti-grim-trigger than grim-trigger does against itself. Indeed, in any population where there are at least as many anti-grim-trigger as grim trigger always defect is the strict best response. More broadly, it is well known in the literature that evolution and learning theory is fraught when a broad range of strategies including strategies like anti-grim-trigger are allowed. One example is Johnson, Levine and Pesendorfer (2001) who study a simpler setting of a gift giving game.

One argument that can be used against anti-grim-trigger is that it is weakly dominated by always defect. Empirically, however, in the first match of a one-shot repeated prisoner's dilemma Dal Bo (2005) observes more than 25% of the population are playing the strictly dominated strategy of cooperating. This makes it hard to argue that players will not use weakly dominated strategies.

Consider, by contrast, the behavioral mechanism design approach. Here it is supposed that ethical players understand that they must create incentives for others to get a good outcome. Naturally they focus on strategies such as grim-trigger that provide incentives. Unless they are awfully clever, they cannot be sure what the right incentives are, or even if adequate incentives can be provided: this they must learn. In this sense behavioral mechanism design and learning theory are complements rather than substitutes.

It is doubtful that many real players in real laboratories consider or use strategies such as anti-grim-trigger. From the perspective of mechanism design it makes little sense as it is not part of a welfare maximizing mechanism. While ethical players might be unsure how much to punish and whether it will work, it is unlikely they spend a lot of time experimenting with strategies that fail obvious tests of contributing to the goal of achieving high welfare. In other words, the behavioral mechanism design approach provides a rationale for limiting strategic options. On the other hand, learning theory may explain why best mechanisms may not be observed: it may be to hard for the ethical players to teach the selfish players.

Learning

Consider first what is known about learning, Figure 5.1 is reproduced from Fudenberg and Rehbinder (2024). As indicated, they use a simple learning model to predict first period cooperation over repeated indefinite horizon prisoner's dilemma games for a large data set that includes Dal Bo and Frechette (2011). What is striking is that for $0.00 < \Delta^{RD} < 0.15$, despite the fact that good mechanisms with high cooperation are available, players seem to have no success in finding them. Specifically, in the range $0.00 < \Delta^{RD} \leq 0.15$ initial cooperation in the first supergame is a bit less than 0.6 and and it remains a bit less than 0.6 even in the 20th match.

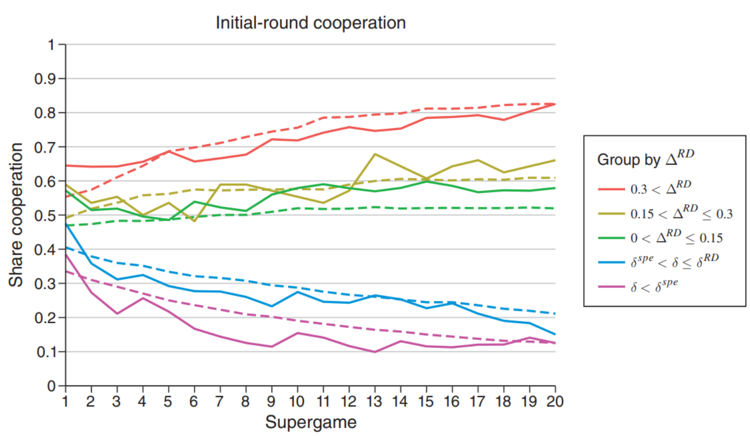

Figure 5.1: Learning: solid line is data

The next question is: what sort of mechanism would be consistent with the type of behavior seen when $\delta = 1/2, U = 48$ where some incentives are provided, but not "enough?" The Δ^{RD} cutoff presumes that some players are always defecting while others provide incentives, and this appears to be what happens in the intermediate range. From Table 4.3 there is a sharp discontinuity between $\Delta^{RD} = 0.11$ and $\Delta^{RD} = 0.14$. From this a possible learning hypothesis is to choose a cutoff $\overline{\Delta}^{RD} \in (0.11, 0.14)$ and assume that for $\Delta^{RD} \leq \overline{\Delta}^{RD}$, in the long run players can at best attain only a second best mechanism in which a substantial fraction always defect. Alternatively, this can be framed in terms of the basin with a cutoff $\overline{\beta} \in (0.62, 0.73)$ with the criterion $\beta \leq \overline{\beta}$.

The Second Best

In the calibration of Levine (2024) there are hypothesized to be equal numbers of ethical players who actively strive to find a good mechanism and of selfish players who simply are out for what they can get. Suppose that in the lower range $\Delta^{RD} \leq \overline{\Delta}^{RD}$ the selfish players are unable to learn about the incentives provided by the ethical players so they intend to always defect. In other words, suppose that in an *ad hoc* way the theory is modified to say that for $\Delta^{RD} \leq \overline{\Delta}^{RD}$ what is observed is the best mechanism subject to the constraint that half the players intend to always defect. Note that when optimal incentives are provided players are in fact indifferent between always defect and following the equilibrium strategy so there is no loss to selfish players in playing always defect.

To minimize the departure from the base model I assume that although a player knows their own type before deciding how to play, types are independently redrawn each period. This makes it possible to continue to use strongly symmetric equilibrium. An alternative model would assume that types are persistent. In this case, in repeated play, ethical types will update their beliefs about facing a selfish type and there will be a semi-separating equilibrium. In Appendix II I show this does not make much difference.

Below in Table 5.1 theoretical welfare is recomputed using the ad hoc model.

δ	U					
	32		40		48	
	theory	data	theory	data	theory	data
1/2	26.9	25.9	27.1	27.1	**31.7**	**30.9**
3/4	26.9	27.2	33.5	34.2	42.0	43.0

Table 5.1: *Ad Hoc* Welfare

The only game in which $\Delta^{RD} \leq \overline{\Delta}^{RD}$ and a non-trivial mechanism is feasible is the anomolous $\delta = 1/2, U = 48$ game highlighted in bold: the remaining numbers do not change. The results are fairly striking: the *ad hoc* $50-50$ hypothesis for $\Delta^{RD} \leq \overline{\Delta}^{RD}$ predicts welfare to within a cent in every case.

Largesse

As indicated above, in Levine (2024) the population of ethical players has largesse equal to about \$1.00/160 per round. This is too small to make much difference to the static Nash equilibrium (with trembles) and cannot make a difference when the ethical players are providing incentives to the selfish players. For this reason I have so far ignored it. However, it could make difference if the ethical players provide incentives only to each other: the punishment need not be so great as ethical players do not feel the need to deviate if the gain is less than their largesse.

Specifically, with largess $\gamma = 1.00$ and $T = 160$ rounds, the second incentive constraint becomes

$$v + \gamma/T = (1-\delta)(u+g) + \delta((1 - q_{DD} - q_{DC})v + q_{DD}(v - P_{DD}) + q_{DC}(v - P_{DC}))$$

which is the same as reducing g by $\gamma/(T(1-\delta))$,

To see the effect of largesse I recomputed the *ad hoc* welfare in Table 5.2 below. Note that welfare accounts for the fact that the selfish players do a bit better (γ/T) than the ethical players.

For the $\delta = 1/2, U = 48$ game largesse makes little difference: welfare increases only slightly from 31.7 to 32.3. The interesting thing that emerges is that while the "full monte" solution is infeasible when $\delta = 1/2, U = 40$ the "half monte" solution of letting the selfish players defect is. (Without largesse the "half monte" solution is not feasible.) While this may seem a bit puzzling, bear in mind that with $U = 40$ the gain to deviating is reduced by having many defections. This makes it a bit easier to sustain cooperation among the ethical players.

δ	U					
	32		40		48	
	theory	data	theory	data	theory	data
1/2	26.9	25.9	**28.3**	**27.1**	**32.3**	**30.9**
3/4	26.9	27.2	33.5	34.2	42.0	43.0

Table 5.2: *Ad Hoc* Welfare with Largesse

In all cases the theory remains within a cent and a half of the data.

6. Retrospective Study of a Public Goods Game With Punishment

The theory that for $\Delta^{RD} \leq \overline{\Delta}^{RD}$ or $\beta \leq \overline{\beta}$ the ethical players cannot teach incentives to the selfish players and so can only provide incentives to each other is *ad hoc* in the sense that it is invented after discovering an anomaly in the data. To validate the theory it would have to make predictions for other games on which the theory was not based. In fourteen treatments involving experienced players Levine (2024) found one important anomaly. Can the teaching/learning theory explain that?

The Game

The anomaly occurs in a game studied by Nikiforakis and Normann (2008). This is a public good contribution game in which players have the opportunity to punish one another and in which the cost of punishment is varied. Specifically the game has two stages. In the first stage money payoffs are given by

$$m^i(1) = 1.50 - q^i + 0.4 \sum_{j=1}^{n} q^j$$

where $q^i \in \{0, 0.075, 0.150, \ldots, 1.50\}$ and $n = 4$. There is a punishment factor $\lambda \in \{0, 1, \ldots, 4\}$ and if $\lambda > 0$ there is a second stage

$$m^i(2) = m^i(1) - \sum_{j \neq i} p^{ij} - \lambda \sum_{j \neq i} p^{ji}$$

where $p^{ij} \in \{0, 0.075, 0.150, \ldots\}$ is a punishment assigned by player i to player j. There is also a constraint on individual punishment $\sum_{j \neq i} p^{ij} \leq m^i(1)$.

The Anomaly

Table 6.1 below summarizes the anomaly. The theory column is welfare predicted by the benchmark theory, the second best is predicted by the "second best" learning theory described below, the "se" is the standard error due to sampling, and the "jungle" column is welfare when no incentives are provided.

λ	welfare				
	theory	second best	data	se	jungle
2	1.88	1.88	1.78	0.10	1.67
1	**1.80**	1.67	**1.64**	0.07	1.67

Table 6.1: Public Goods Contribution with Punishment

As highlighted in bold, for punishment factor $\lambda = 1$ the theory predicts that players should get substantially more than they actually get. By contrast actual welfare is much higher when the punishment factor $\lambda = 2$ and the difference with the theoretical prediction may well be due to sampling error.

Risk Dominance

Note that Δ^{RE} is specific to a repeated game with discounting and has no analog here, so I focus on $\beta \leq \overline{\beta}$ where recall that $\overline{\beta} \in (0.62, 0.73)$. What is the analog of β for $\lambda \in \{1, 2\}$ and is it smaller than $\overline{\beta}$ for $\lambda = 1$ and bigger for $\lambda = 2$? If so what is predicted for $\lambda = 1$?

First let me indicate that in public goods with punishment games the "half monte" solution is the same as the "law of the jungle" where no incentives are provided: incentives that only get 1/3 the population to increase their contributions are too welfare costly. Hence, as indicated in Table 6.1, the *ad hoc* teaching/learning theory would predict that when $\lambda = 1$ welfare should be 1.67, quite close to actual welfare of 1.64 and this would be scored as a success for the *ad hoc* theory.

To figure risk dominance, I proceed by analog to the repeated prisoner's dilemma game. Players contributing nothing to the public good can be regarded as "all defect." Contributing the maximum of \$1.50 with maximum punishment for doing less can be regarded as "grim trigger." The former is subgame perfect. The latter is not, but it is Nash. The idea then is to compute for each λ compute the basin β which is fraction of "all defect" which leads to indifference between "all defect" and contributing \$1.50. Note that the players who are being "persuaded" to contribute by the ethical players are selfish players who will not punish, that is, they find it a best response not to play "grim trigger" but to play "always cooperate."

Finding the Basin

For each β a player faces three opponents who are either playing "all defect" or "grim trigger." The probability of each combination is shown below in Table

6.2. Denote by Π the maximum punishment that can be sent by all "grim trigger" opponents divided by the number of "all defect" plus one: this is also shown below in Table 6.2.

"all defect"	"grim trigger"	probability	Π
3	0	β^3	0
2	1	$3\beta^2(1-\beta)$	0.2
1	2	$2\beta(1-\beta)^2$	0.6
0	3	$(1-\beta)^3$	5.4

Table 6.2: Player Matching

Construct $P(\beta)$ the probability weighted average of these punishments and let $Q(\beta)$ be the expected contribution of the opponents. A player who plays "all defect" against such a population gets utility $1.50 - \lambda P(\beta) + 0.4Q(\beta)$, while a player who contributes 1.50 and does not punish gets $0.4(Q(\beta) + 1.50)$. Hence the basin is given by solving $\lambda P(\beta) = 0.90$. The solution for $\lambda = 1$ is $\beta = 0.57$ and for $\lambda = 2$ it is $\beta = 0.71$. This gives the range that separates $\lambda = 1$ from $\lambda = 2$ as $\overline{\beta} \in (0.57, 0.71)$. The corresponding range for the repeated PD was $\overline{\beta} \in (0.62, 0.73)$. Hence the range $\overline{\beta} \in (0.62, 0.71)$ is thus consistent with both public goods with punishment and the repeated prisoner's dilemma.

7. Conclusion

There are very few benchmark models that yield predictions about experimental outcomes based only on the experimental instructions. Subgame perfect Nash equilibrium with risk-neutral selfish players is one such, but is known to do a poor job in many circumstances. It is rather a low bar these days to conduct an experiment and find that it is not explained well by Nash equilibrium. In Levine (2024) I proposed an alternative benchmark theory that is behavioral having both ethical and noisy players. This yielded few anomalies when applied to a number of treatments of stag-hunt, ultimatum bargaining, and public goods contribution games with and without punishment.

If the benchmark is to be truly useful it must apply broadly. One particularly demanding study is that of repeated games. In a sense this is a natural setting to think of players acting as mechanism designers since the data suggests that players are struggling to find good solutions, but sometimes fail to do so. Can the benchmark theory explain their successes and failures? In five out of six treatments it does exactly that. In the sixth treatment it does not. Further examination of this anomaly suggests that this may be due to difficult in learning and a simple, but *ad hoc*, adjustment of the theory accounts well for the anomalous case. A retrospective study of an anomaly in a public goods game with punishment provides an out-of-sample confirmation of the *ad hoc* theory.

The learning idea is an important one. In the broad sample in Figure 5.1 it can be seen that repeated prisoner's dilemma games with experienced players tend to fall into three categories: high cooperation, low cooperation, and intermediate cooperation. The "half monte" mechanism in which ethical players blow off the selfish players and provide incentives only for themselves captures this intermediate level of cooperation (and in the anomalous case does so with high quantitative accuracy). As the analysis of largesse shows, in some cases the "half monte" mechanism may be available even if the "full monte" is not. In the anomalous case the "full monte" is available, but learning models suggest that it is too difficult to teach it to the selfish players. In this sense the behavioral mechanism design approach and the learning approach may be seen to be complementary. It also points the way towards hybrid theories, for example, learning models in which players strategic options are limited to always defect, the half monte mechanism and the full monte mechanism or in which they adjust the punishment parameters of the mechanism.

Appendix I: Solving the LP Problem

Observe that the LP problem makes players indifferent between cooperating and defecting. Hence they could mix: but doing so changes the incentive constraints by worsening the signal, reducing the payoff to cooperating and for $U \in \{40, 48\}$ increasing the gain to defecting so can never be part of the best equilibrium. For $U = 32$ to make sure that mixing did not make a cooperative equilibrium feasible the incentive constraints were checked where the gain to defecting was computed under the assumption that the opponent intended to defect, that is, is as small as possible. Hence, the computations below are for pure strategy equilibria.

Recall that

$$v = (1-\delta)u + \delta((1 - p_{DD} - p_{DC})v + p_{DD}(v - P_{DD}) + p_{DC}(v - P_{DC}))$$

and rewrite it as

$$\rho(u - v) = p_{DD}P_{DD} + p_{DC}P_{DC}.$$

Similarly the second incentive constraint can be rewritten as

$$\rho(u + g - v) = q_{DD}P_{DD} + q_{DC}P_{DC}.$$

The two can be combined to give a second punishment constraint entirely in terms of the punishments

$$\rho g = (q_{DD} - p_{DD})P_{DD} + (q_{DC} - p_{DC})P_{DC}.$$

Maximizing v is the same as minimizing the aggregate on-path punishment cost $p_{DD}P_{DD} + p_{DC}P_{DC}$ subject to the punishment constraint. Equivalently Q_{DD}, Q_{DC} can be chosen to minize

$$\frac{p_{DD}}{q_{DD} - p_{DD}}Q_{DD} + \frac{p_{DC}}{q_{DC} - p_{DC}}Q_{DC}$$

subject to $\rho g = Q_{DD} + Q_{DC}$. Hence the the maximum Q (and P) should be allocated to the smallest $p_K/(q_K - p_K)$.

Computing the ratios for the two signals gives

$$\frac{p_{DD}}{q_{DD}-p_{DD}} = \frac{1}{1-2\pi}\pi$$

$$\frac{p_{DC}}{q_{DC}-p_{DC}} = \frac{1}{1-2\pi}\pi\frac{2-2\pi}{1-2\pi}$$

so that DD is more efficient.

Only DD Signal Used

If only DD is punished and the $v - P_{DD} \geq n$ constraint does not bind, then the two equality constraints can be solved to find

$$P_{DD} = \rho\frac{1}{q_{DD}-p_{DD}}g$$

$$v = u - \frac{p_{DD}}{q_{DD}-p_{DD}}g.$$

Hence

$$v - P_{DD} = u - \frac{p_{DD}}{q_{DD}-p_{DD}}g - P$$

$$= u - \frac{\rho + p_{DD}}{q_{DD}-p_{DD}}g \geq n$$

or

$$\frac{\rho + p_{DD}}{q_{DD}-p_{DD}} \leq R. \tag{7.1}$$

Both Signals Used

If inequality 7.1 fails then both signals must be used and $P_{DD} = v - n$. The two equalities are now

$$\rho(u-v) = p_{DD}(v-n) + p_{DC}P_{DC}$$

$$\rho g = (q_{DD}-p_{DD})(v-n) + (q_{DC}-p_{DC})P_{DC}.$$

Notice that increasing P_{DC} has two effects: it increases the punishment received when DC/CD occur, but it also reduces $v-n$ which is the punishment received when DD

occurs. If $q_{DC} - p_{DC} - \frac{q_{DD}-p_{DD}}{\rho+p_{DD}} p_{DC} \leq 0$ the latter effect weakly dominates the first so the only solution is that static Nash. Otherwise

$$P_{DC} = \rho g \frac{1 - \frac{q_{DD}-p_{DD}}{\rho+p_{DD}} R}{q_{DC} - p_{DC} - \frac{q_{DD}-p_{DD}}{\rho+p_{DD}} p_{DC}}$$

$$v = \frac{\rho u + p_{DD} n}{\rho + p_{DD}} - \frac{p_{DC} P_{DC}}{\rho + p_{DD}}.$$

It remains to check that $v - P_{DC} \geq n$. This is

$$\frac{1}{\rho + p_{DD}} \left(\rho(u-n) + p_{DC} P_{DC}\right) \geq P_{DC}$$

which can be written as

$$R \geq \frac{\rho + p_{DD} + p_{DC}}{(q_{DD} + q_{DC}) - (p_{DD} + p_{DC})}.$$

If this fails then the only strong symmetric equilibrium is the static Nash.

Appendix II: Persistent Types

In the $\delta = 1/2, U = 48$ game consider the "half monte" mechanism with persistent types.

In the first round (Table 3.2) an ethical player gets an equally weighted average of 43 and 20.

Ethical types cooperate in the first period and the selfish types defect and both have the same probability of trembling. Hence if the opponent cooperates in the first period the posterior probability of the opponent is ethical is 5/6, while if the opponent defects it is 1/6.

To get an idea of what happens from the second round on take the extreme assumption that the type is perfectly revealed after the first round. There is a 1/2 chance that the opponent is ethical, resulting in an average present value payoff of 41.6 (Table 4.1) and a 1/2 chance the opponent is selfish resulting in an average expected present value payoff of 27 (Table 3.2). Taking expectations and present value this yields 32.9.

Selfish players, not knowing they can do better, get less. With probability 1/2 in the first period selfish players get 45, but in all other cases they get 27 giving 31.5

Averaging over the two types gives welfare as 32.2. With impersistent types this number (Table 5.1) is 31.7, so the increase due to type persistence is not large.

References

Abreu, D., D. Pearce, and E. Stacchetti (1986): “Optimal Cartel Equilibria with Imperfect Monitoring," *Journal of Economic Theory* 39: 251-269.

Blonski, M., P. Ockenfels and G. Spagnolo (2011): “Equilibrium Selection in the Repeated Prisoner’s Dilemma: Axiomatic Approach and Experimental Evidence,” *American Economic Journal: Microeconomics* 3: 164–92.

Blonski, M. and G. Spagnolo (2015): "”Prisoners’ other dilemma,”" *International Journal of Game Theory* 44: 61-81.

Bó, P. Dal (2005): “Cooperation under the shadow of the future: experimental evidence from infinitely repeated games,” *American Economic Review* 95: 1591-1604.

Dal Bó, P. and G. R. Fréchette, G. R. (2011): “The evolution of cooperation in infinitely repeated games: Experimental evidence,” *American Economic Review* 101: 411-429.

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

Fudenberg, D., and G. Karreskog Rehbinder (2024): “Predicting Cooperation with Learning Models,” *American Economic Journal: Microeconomics* 16: 1-32.

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

Fudenberg, D., and E. Maskin (1986), “The Folk Theorem in Repeated Games with Discounting or with Incomplete Information,” *Econom*etrica 54: 533.

Johnson, P., D. K. Levine and W. Pesendorfer (2001): “Evolution and Information in a Gift Giving Game,” *Journal of Economic Theory* 100: 1-22.

Kandori, Michihiro, George Mailath, and Rafael Rob (1993): “Learning, Mutation, and Long Run Equilibria in Games,” *Econometrica* 61: 29-56.

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

Levine, D. K. (2024): “Behavioral Mechanism Design as a Benchmark for Experimental Studies,” mimeo RHUL.

Mailath, G. J., and L. Samuelson, L. (2006), *Repeated Games and Reputations: Long-run Relationships*, Oxford university press.

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

Nikiforakis, N. and H. T. Normann (2008): “A Comparative Statics Analysis of Punishment in Public-good Experiments,” *Experimental Economics* 11: 358-369.

Sannikov, Y. (2007): “Games with Imperfectly Observable Actions in Continuous Time,” *Econometrica* 75: 1285-1329.

Young, P. (1993): “The Evolution of Conventions,” *Econometrica* 61: 57-83