I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15023

It's Payback Time: New Insights on Cooperation in the Repeated Prisoners' Dilemma

Maria Bigoni
Marco Casari
Andrea Salvanti
Andrzej Skrzypacz
Giancarlo Spagnolo

JANUARY 2022

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15023

It's Payback Time: New Insights on Cooperation in the Repeated Prisoners' Dilemma

Maria Bigoni
University of Bologna and IZA

Marco Casari
University of Bologna and IZA

Andrea Salvanti
Universitat Pompeu Fabra

Andrzej Skrzypacz
Stanford GSB

Giancarlo Spagnolo
Stockholm University, University of Rome Tor Vergata and EIEF

JANUARY 2022

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

It's Payback Time: New Insights on Cooperation in the Repeated Prisoners' Dilemma

In an experiment on the repeated prisoner's dilemma where intended actions are implemented with noise, Fudenberg et al. (2012) observe that non-equilibrium strategies of the "tit-for-tat" family are largely adopted. Furthermore, they do not find support for risk dominance of TFT as a determinant of cooperation. This comment introduces the "Payback" strategy, which is similar to TFT but is sustainable in equilibrium. Using the data from the original article, we show that Payback captures most of the empirical support previously attributed to TFT, and that the risk dominance criterion based on Payback can explain the observed cooperation patterns.

JEL Classification: C72, C73, C91, D82

Keywords: asymmetric strategies, imperfect monitoring, indefinitely repeated games, risk dominance, strategic risk

Corresponding author:
Maria Bigoni
Department of Economics
University of Bologna
Piazza Scaravilli 2
40126 Bologna
Italy
E-mail: maria.bigoni@unibo.it

In an important contribution for understanding cooperation in the repeated prisoner's dilemma, Fudenberg et al. (2012) [FRD, henceforth] study experimentally a setting where intended actions are implemented with noise. Their main goals are to understand whether and when cooperation emerges, and to identify which strategies participants use. FRD show that – in the treatments with cooperative equilibria – successful strategies were "lenient" in not retaliating for the first defection, and many were "forgiving" in trying to return to cooperation after inflicting a punishment.

This note focuses on two other findings of FRD that we consider somewhat puzzling. We address those puzzles and suggest a novel interpretation of the data that may have far-reaching implications well beyond their specific setting. The key to explaining these findings is the introduction in the theoretical and empirical analyses of an often-neglected strategy that we label "Payback" strategy. The Payback strategy triggers punishment after any observed unilateral deviation but reverts to mutual cooperation as soon as the deviator "pays the other player back". In particular, it prescribes players to cooperate in period 1; if a unilateral defection is observed, the player who suffered the deviation should defect, while the original deviator should cooperate to compensate the opponent for the original loss. Once this happens, both players return to mutual cooperation. In this note, we argue that in a noisy environment, the Payback strategy has theoretical advantages compared to other cooperative strategies, such as Grim Trigger and Tit-for-Tat, and that it finds strong empirical support in the data.

1 The FRD experiment

The FRD setup is an infinitely repeated prisoner's dilemma where intended actions are implemented with noise. The duration of a supergame is set by a random draw following each round. The continuation probability is $\delta = 0.875$. Each agent makes errors with an independent probability E. When an error occurs, the agent's intended move is changed to the opposite move. Agents know when their own move is changed by an error, but not when this happens to their opponent. In other words, agents observe their opponents' actual move but not their intended one. Under this form of imperfect observability, strategies can only be conditional on realized actions, which may differ from the intended ones. In their six experimental treatments, FRD vary the payoff specifications and the level of noise. They parametrize the game payoffs in terms of a "benefit/cost" (b/c) ratio. "Cooperation means paying a cost c to give a benefit b to the other player, while defection gives 0 to each party." (p.723). In their treatments, the implemented values of b/c are 1.5, 2, 2.5, and 4, while the error probability takes values: $E = 0, 1/16, 1/8$.

The original paper reports that subjects use more lenient and forgiving strategies with imperfect private monitoring than with perfect monitoring. Besides this main result, the authors highlight two other findings that are the focus of this note.

Finding A *Although not an equilibrium strategy, TFT is empirically widely used both with perfect and imperfect monitoring (p.734 and Table 3 in FRD).*

The relative success of TFT in categorizing behavior is puzzling in our view because the strategy is not subgame perfect Nash equilibrium in any treatment of FRD.[1] In the Appendix, FRD show that "TFT is never an equilibrium in the presence of noise" (p.48 Online Appendix). One can also show that TFT is not an equilibrium in the treatment with perfect monitoring.

[1] We say for short that a strategy is an equilibrium if both players using that strategy forms a subgame perfect Nash equilibrium. We say that a strategy is not an equilibrium otherwise.

We provide a possible explanation for this finding. We show that the Payback strategy mentioned earlier is an equilibrium in all FRD treatments, including the treatment with b/c = 1.5, where cooperation is most challenging. We then show that a re-estimation of strategy frequency that also includes Payback considerably reduces, and sometimes wipes out completely, the empirical relevance of TFT.

Finding B *"The data do not show the strong support for risk dominance of TFT as the key determinant of the level of cooperation in games with noise that was seen in studies of games without noise" (p.733 in FRD).*

FRD ask whether risk dominance of TFT over ALLD, in the sense of Blonski and Spagnolo (2015) and Blonski et al. (2011), is a good predictor of cooperation (Question 1, p. 726 in FRD). They observe that TFT is risk dominant only in treatments where b/c$\geq$ 2.5, but "the largest difference in cooperation occurs between b/c=1.5 and b/c=2, as opposed to between b/c=2 and b/c=2.5" (p. 733). Thus they conclude that, in contrast to what observed in games without noise, "the risk-dominance criterion has at best limited predictive power regarding cooperation in games with noise" (p.742).

We argue that considering Payback instead of TFT sheds new light on this finding too. As we show, when we look at Payback vs. ALLD, risk dominance measures are much more aligned with the pattern of cooperation. The same happens when we use the basin of attraction of ALLD (BAD). Therefore, the experimental results in FRD are consistent with the findings for games without noise indicating that measures of strategic risk are effective predictors of cooperation.

2 The Payback strategy

Figure 1 describes the Payback strategy using an automaton with three states. The agent starts cooperating (c) and remains in the "Cooperation" state after observing mutual cooperation (cc) or mutual defection (dd). If the implemented actions are cooperation by the agent and defection by

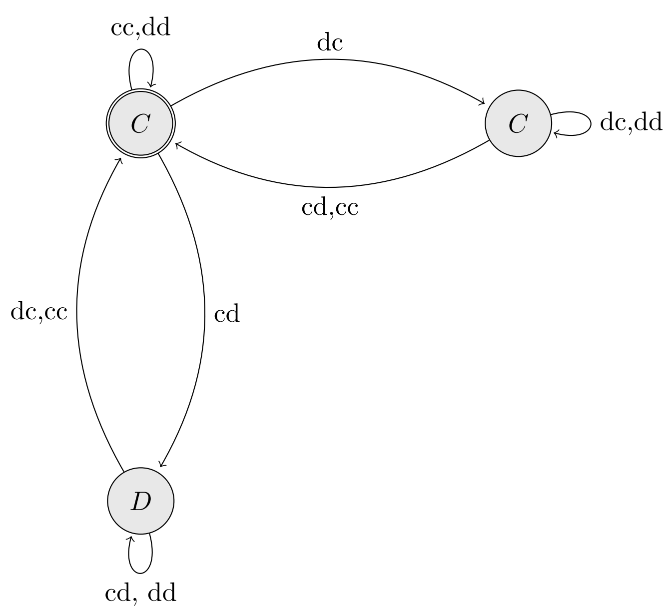

Figure 1. Automaton of "Payback strategy"

the opponent (cd), then the agent transitions to a Punishment state, where the prescribed action is defection; the agent continues defecting until observing a cooperative action by the opponent (dc or cc), then reverts to Cooperation.[2] If in the Cooperation state the implemented actions are defection by the agent and cooperation by the opponent (dc), then the agent transitions to the Payback state, where the prescribed action is cooperation; if in the Payback state the implemented actions are cc or cd, the agent transitions back to the Cooperation state, and continues cooperating.

A comparison with the TFT and Grim strategies can further clarify the Payback strategy. While Payback can be described by a three-state automaton, TFT and Grim have only two states. All strategies start with cooperation. While Payback allows for forgiveness, as TFT does, Grim is unforgiving. At first sight, Payback may look very similar to TFT. To highlight the specificities of Payback, we first mention two instances that set it apart from both TFT and Grim. First, if the observed outcome is cd, the prescription of Payback depends on the previous state: cooperation in the Payback state and defection in the other states. Second, if the agent observes dd, Payback prescribes defection only if the agent was in the Punishment state. Otherwise, it prescribes cooperation. Instead, in both instances, TFT and Grim always prescribe defection.

The Payback strategy differs notably from the Grim strategy because it punishes deviations asymmetrically: when player 1 defects (intentionally or not), his continuation payoff goes down. At the same time, the continuation payoff of player 2 goes up. In the continuation game, player 2 is paid back for the bad outcome this period. In contrast, Grim (and any strongly symmetric strategy, like the ones described in Green and Porter, 1984) punishes both players when one of them defects.

The Payback strategy is the simplest (represented by just a three-state automaton) example of a larger family of cooperative strategies that use such asymmetric punishments. Such strategies have been shown to have good theoretical properties and empirical support. For example, such strategies have been studied by Van Damme (1989) in the context of renegotiation-proofness in the repeated Prisoner's Dilemma, and in models of repeated games with monetary payments like relational contracts (see, for example, Levin, 2003). In the context of tacit collusion in repeated auctions, the Payback strategy is similar to the "chips mechanism" studied in Skrzypacz and Hopenhayn (2004). Interpreted through this lens, Payback strategy is also similar to the strategies used by agents in monetary exchange experiments, see for example Bigoni et al. (2019). In pricing games, asymmetric strategies similar to the Payback strategy have also been shown to perform better than strongly symmetric equilibria (see, for example, Athey and Bagwell, 2001, and Harrington and Skrzypacz, 2007). Empirically, there is a lot of evidence that real-life pricing cartels use asymmetric punishments – see, for example, Harrington (2006).

The differences between Payback, Grim, and TFT become particularly consequential in an environment with noisy implementation of actions, like the one studied by FRD. While in games without noise how off-path punishments are structured may not matter a lot (provided they are strong enough), in games with noise, unintended deviations happen on the equilibrium path, and symmetric punishments become very costly for the players. In contrast, asymmetric punishments in Payback can keep the total payoff high while still providing sufficient incentives for cooperation.

Consider two players who wish to coordinate on mutual cooperation, and both start playing c in period 1. With probability E, the action of either of them is changed to d. With a Grim Trigger strategy, they would immediately switch to mutual defection until the end of the supergame. With TFT, a sequence of asymmetric cd-dc outcomes would follow. The players would be able to revert to

[2] The outcome cc in the Punishment state happens only in case the agent's own action is implemented with noise.

mutual cooperation only when a second random error materializes and changes the intended action of one of them from d to c. With Payback, instead, they would be able to revert to mutual cooperation after a single period of asymmetric punishment unless a second random error takes place. This intuitive consideration clearly illustrates why – on the equilibrium path – the value of cooperation under Payback is much higher than under TFT and Grim, as soon as we allow for a positive probability of implementation errors. The first panel of Table 1 presents the value of cooperation for Payback, Grim Trigger and Tit-for-Tat, under the six treatments included in FRD's experiment: in all noisy treatments, the expected ex-ante payoff is higher if players were to use Payback than if they were to use Grim or TFT strategies.

The value of cooperation, however, is only one of the elements that contribute to the success of a cooperative strategy. Incentive compatibility constraints are also important. One of the reasons TFT and Payback have higher expected on-path payoffs than Grim is that they are more forgiving. In a game with noise, forgiveness is helpful to reduce the cost of false-positive punishments (i.e., punishments after unintended deviations). However, inappropriate forgiveness can undermine the incentives to follow the equilibrium strategies.

How do these strategies compare in terms of satisfying equilibrium (incentive compatibility) conditions? To answer that, we report in Panel 2 of Table 1 the critical discount factors (δ_{SPE}) for these strategies to form a subgame perfect Nash equilibrium in the different treatments. The discount factor (or continuation probability) must be above the critical one for all the IC constraints to be satisfied. For the continuation probability adopted in the FRD experiment (87.5%), we find that Payback is an

Table 1. Thresholds for cooperation in FRD treatments

Treatment:	b/c=1.5 E=1/8	b/c=2 E=1/8	b/c=2,5 E=1/8	b/c=4 E=1/8	b/c=4 E=1/16	b/c=4 E=0
Strategies						
Vcoop						
Payback	6.5	12.9	19.4	38.8	42.9	48
Grim	3.3	6.5	9.8	19.6	25.7	48
TFT	5.1	10.2	15.3	30.6	35.2	48
δ_{SPE}						
Payback	0.85	0.67	0.55	0.36	0.30	0.25
Grim	0.91	0.70	0.57	0.37	0.30	0.25
TFT	/	/	/	/	/	/
δ_{RD}						
Payback	0.96	0.83	0.73	0.54	0.46	0.4
Grim	>1	0.91	0.80	0.57	0.47	0.4
TFT	>1	0.89	0.76	0.53	0.46	0.4
Size-BAD						
Payback	0.81	0.40	0.27	0.13	0.08	0.05
Grim	>1	0.57	0.38	0.19	0.11	0.05
TFT	>1	0.52	0.35	0.17	0.10	0.05
Cooperation rates						
First rounds	54%	75%	79%	76%	87%	83%
All rounds	32%	49%	61%	59%	82%	78%

Note: Cooperation rates from Tables 4 and 6 in FRD. Critical thresholds from our calculations (Appendix B).

equilibrium in all treatments. By contrast, as already noticed by FRD, TFT is not an equilibrium in any of the treatments. It is not an equilibrium for the discount factor in the experiment, and it is not an equilibrium for any discount factor. In fact, it would not be an equilibrium for any continuation probability: it does not resolve the balance between punishments and forgiveness correctly in this game. Finally, Grim is an equilibrium in all but one treatment (b/c=1.5).

Comparing the critical discount factors for Grim and Payback we find a new regularity: they are the same in the treatment without noise and uniformly lower for Payback in all treatments with noise. In Appendix B we prove a stronger result: these numerical findings are not a coincidence: for all b/c and $E \in (0, 0.5)$ the critical discount factor for Payback is smaller than for Grim.[3]

In summary, these theoretical findings suggest that Payback may be a particularly attractive strategy for subjects to try – it is simple, if subjects manage to coordinate on it, it results in high payoffs, and the temptation to deviate from that strategy is weaker than for the other two strategies.

3 An alternative interpretation of the findings

FRD find that TFT is an empirically common strategy with perfect and imperfect monitoring even though it is never an equilibrium. The Maximum Likelihood Estimation of a set of eleven strategies shows that with perfect monitoring, 14% of subjects are classified as playing TFT, and with imperfect monitoring, the share is 4-19% across five treatments (Tables 3 and 5 in FRD, reported in Appendix A). We repeated the MLE analysis of strategy choices reported in FRD, including Payback in the strategy set.[4] With perfect monitoring, the MLE assigns a share of 17% to Payback, and with imperfect monitoring, the share is 7-15% across treatments. The introduction of Payback reduces the share assigned to TFT in all treatments: with perfect monitoring, the estimated share of TFT falls to zero, while with imperfect monitoring, it falls to 0-14%. The introduction of Payback makes TFT essentially disappear in four out of the six treatments.

Table 2. Maximum Likelihood Estimates on the full strategy set

Strategies	**bc=1.5** ε=1/8	**bc=2** ε=1/8	**bc=2.5** ε=1/8	**bc=4** ε=1/8	**bc=4** ε=1/16	**bc=4** ε=0
ALLC	0.00	0.03	0.00	0.06**	0.00	0.20
TFT	0.14***	0.00	0.04	0.00	0.00	0.00
TF2T	0.04	0.00	0.11	0.16**	0.21**	0.00
TF3T	0.01	0.03	0.03	0.09***	0.37*	0.00
2TFT	0.06*	0.07*	0.02	0.03	0.00	0.15**
2TF2T	0.01	0.10**	0.12**	0.09**	0.08	0.00
Grim	0.14***	0.07	0.11**	0.04	0.03	0.15**
Grim2	0.06*	0.20***	0.02	0.05***	0.09*	0.16
Grim3	0.05*	0.26***	0.23***	0.11	0.00	0.00
ALLD	0.29***	0.17***	0.14***	0.23***	0.05**	0.07**
D-TFT	0.13***	0.00	0.05	0.00	0.05**	0.09
Payback	0.08*	0.07	0.11***	0.15***	0.12**	0.17**
Gamma	0.46***	0.50***	0.49***	0.42***	0.44***	0.35***

Note: Estimates based only on the last four interactions in each session.

To further investigate the empirical role of Payback vs. TFT, we consider a new estimate based on

[3] There are versions of the PD game in which this ranking is reversed for small amount of noise, but for the subset of games studied in FRD the ranking always holds.

[4] We thank Fudenberg, Rand and Dreber for having shared with us the code for the MLE analysis of strategies.

a restricted set of strategies: ALLC, ALLD, GRIM, and TFT. Removing the TFT variants from the strategy set provides the best conditions for TFT to classify behavior and allows for a more balanced comparison with Payback. Table 3 reports these estimates articulated into three panels. Panel A estimates a subset of the strategies in FRD; Panel B adds Payback to the subset; and Panel C replaces TFT with Payback in the subset. The estimated share of TFT is 40% with perfect monitoring, and with imperfect monitoring, the share is 27-48% across treatments (Table 3, panel A). We then replicate

Table 3. Maximum Likelihood Estimates on the restricted strategy set

	Strategies	**bc=1,5** ε=1/8	**bc=2** ε=1/8	**bc=2,5** ε=1/8	**bc=4** ε=1/8	**bc=4** ε=1/16	**bc=4** ε=0
A	ALLC	0.00	0.12***	0.36***	0.26***	0.65***	0.26***
	TFT	0.45***	0.48***	0.37***	0.45***	0.27***	0.40***
	Grim	0.18***	0.23***	0.12***	0.06*	0.03	0.27***
	AllD	0.37***	0.16***	0.15***	0.23***	0.06	0.07*
	Gamma	0.53***	0.64***	0.60***	0.52***	0.50***	0.39***
B	ALLC	0.00	0.11**	0.27***	0.19***	0.54**	0.23***
	TFT	0.31***	0.24***	0.17***	0.21***	0.14***	0.19
	Grim	0.19***	0.25***	0.12***	0.06***	0.03	0.28***
	AllD	0.37***	0.16***	0.15***	0.23***	0.05*	0.07*
	Payback	0.13***	0.24***	0.29***	0.31***	0.23***	0.23**
	Gamma	0.51***	0.62***	0.58***	0.49***	0.49***	0.38***
C	ALLC	0.00	0.12***	0.28***	0.19***	0.61***	0.24***
	Grim	0.32***	0.34***	0.15***	0.15***	0.05	0.33***
	AllD	0.38***	0.16***	0.16***	0.23***	0.05*	0.07*
	Payback	0.30***	0.38***	0.41***	0.44***	0.29***	0.36***
	Gamma	0.56***	0.64***	0.59***	0.53***	0.51***	0.38***

Notes: Panel A estimates a subset of the strategies in FRD; Panel B adds Payback to the subset; Panel C replaces TFT with Payback in the subset. Estimates are based only on the last four interactions in each session.

these estimates adding Payback to the strategy set. The estimated share of Payback is 23% with perfect monitoring, and with imperfect monitoring, the share is 13-31% across treatments (Table 3, panel B). Again, the introduction of Payback significantly reduces the share of subjects classified as TFT in all treatments. In particular, with perfect monitoring, the estimated share of TFT halves and is no longer statistically significant, while with imperfect monitoring it falls to 14-31%.

Payback appears to effectively replace TFT in the Maximum Likelihood Estimates on the experimental data. Support for this statement comes from comparing Panel A with Panel C, where we substitute TFT with Payback. Consider the parameter Gamma, which shows the goodness of fit of subjects' classification (0=perfect classification). The reported values are similar in Panels A and C: absolute differences in Gamma range from -0.01 through +0.03. Estimates precision declines in three treatments and improves or stays the same in the other three. With perfect monitoring, the estimated share of Payback is 36% and with imperfect monitoring the share is 29-44% across treatments (Table 3, panel C).

4 Risk dominance and the Payback strategy

Let's now turn to Finding B, which concerns the lack of predicting power of risk dominance of TFT in games with noise. It contrasts with the evidence emerging from prisoners' dilemmas played with perfect monitoring, where risk dominance of Tit-for-Tat appears to be one of the strongest predictors of cooperation (Dal Bó and Fréchette, 2018). As illustrated by FRD, TFT becomes risk dominant only in the treatments with $b/c \geq 2.5$, which is misaligned with the largest increase in cooperation observed in the experiment, especially in terms of first-round cooperation, which is observed when b/c increases from 1.5 to 2 (Table 1, last panel).

If we consider Payback instead of TFT, we notice that the largest increase in cooperation occurs precisely when it becomes risk dominant. As illustrated in panel 3 of Table 1, the Payback strategy is not risk dominant for b/c = 1.5 ($\delta_{RD} = 0.95 \geq 0.875$) but becomes risk dominant for $b/c \geq 2$. This suggests that, contrary to Finding B, risk dominance retains predictive power even with imperfect monitoring.

Another way to capture the riskiness of a cooperative strategy is by using the basin of attraction of ALLD when playing against the strategy (BAD). In their meta-analyses, Dal Bó and Fréchette (2018) report that BAD is a reliable predictor of the aggregate cooperation level in repeated PD experiments with perfect monitoring and that it is highly correlated with the measure of risk dominance, δ_{RD}. The fourth panel of Table 1 reports the basin of attraction of ALLD against agents playing Grim, TFT, and Payback. Using this measure reinforces our conclusions above about Finding B. The measure for TFT is below 0.5 only in the treatments where $b/c > 2$ (the 0.5 threshold is conventionally adopted as a predictor of cooperation). This pattern would not explain why the largest increase in cooperation observed in the experiment takes place in the treatment b/c = 2. For Payback, instead, BAD is 0.8 for b/c = 1.5 and 0.4 for b/c = 2, marking a theoretical discontinuity, which predicts a strong increase in cooperation as we observe in the experiment (Table 1, last panel).[5]

5 Conclusions

In this note, we documented the relevant role that the Payback strategy can have in interpreting experimental results in the repeated prisoner's dilemma, in particular with imperfect monitoring. A re-analysis of the FRD data suggests that introducing the Payback strategy sheds light on two puzzles that emerged in their original work. First, Payback is a subgame-perfect equilibrium strategy in the FRD setup, and it constitutes a good empirical alternative to TFT in classifying subjects' choices in the experiment, which explains the sustained rate of cooperation observed in treatments where TFT (and sometimes Grim) are not equilibrium strategies. Second, risk dominance is a relevant predictor of cooperation also under imperfect monitoring when we consider Payback instead of TFT. In other words, in the FRD setting, the introduction of Payback downsizes the role of TFT and restores the ability of equilibrium conditions and risk dominance to predict cooperation also with imperfect monitoring.

A number of other experiments have reported an important role for TFT in categorizing behavior (for example, Bigoni et al., 2013; Dal Bó and Fréchette, 2011, 2019; Dvorak and Fehrler, 2018; Embrey et al., 2016; Romero and Rosokha, 2019) and Payback can offer a new perspective on their results. In general, Payback has been overlooked in the empirical literature. Introducing Payback into the

[5] In the Appendix we show that for all b/c and E the BAD is smaller for Payback than for Grim, providing another theoretically attractive feature of the Payback strategy in these games.

analysis could shed new light on cooperative behavior in many different settings, but we leave this for future work.

Appendix

A Original tables from FRD

Table 3 from FRD. Maximum Likelihood Estimates
Using the Last 4 Interactions of Each Session

	b/c = 1.5	b/c = 2	b/c = 2.5	b/c = 4
ALLC	0.00	0.03	0.00	0.06*
	(0.00)	(0.03)	(0.02)	(0.03)
TFT	0.19***	0.06	0.09**	0.07**
	(0.05)	(0.04)	(0.04)	(0.03)
TF2T	0.05	0.00	0.17**	0.20***
	(0.03)	(0.00)	(0.06)	(0.07)
TF3T	0.01	0.03	0.05	0.09**
	(0.01)	(0.03)	(0.05)	(0.04)
2TFT	0.06	0.07*	0.02	0.03
	(0.04)	(0.04)	(0.02)	(0.02)
2TF2T	0.00	0.11**	0.11*	0.12**
	(0.02)	(0.05)	(0.06)	(0.05)
Grim	0.14***	0.07	0.11**	0.04*
	(0.04)	(0.05)	(0.04)	(0.02)
Grim2	0.06*	0.18***	0.02	0.05*
	(0.03)	(0.06)	(0.03)	(0.03)
Grim3	0.06	0.28***	0.24***	0.11***
	(0.03)	(0.08)	(0.07)	(0.04)
ALLD	0.29***	0.17***	0.14***	0.23***
	(0.06)	(0.06)	(0.04)	(0.04)
D-TFT	0.14***	0.00	0.05*	0.00
	(0.05)	(0.00)	(0.03)	(0.00)
Gamma	0.46***	0.5***	0.49***	0.43***
	(0.02)	(0.03)	(0.03)	(0.02)

Notes: All payoff specifications use error rate $E = 1/8$. Bootstrapped standard errors (shown in parentheses) used to calculate p-values.

Table 5 from FRD. Maximum Likelihood Estimates for our E = 0, E = 1/16, and E = 1/8 Conditions Using the Last Four Interactions of Each Session

	E = 0	E = 1/16	E = 1/8
ALLC	0.24**	0.00	0.06*
	(0.10)	(0.04)	(0.03)
TFT	0.14*	0.04	0.07**
	(0.08)	(0.04)	(0.03)
TF2T	0.00	0.24**	0.20***
	(0.02)	(0.10)	(0.07)
TF3T	0.00	0.42***	0.09**
	(0.04)	(0.09)	(0.04)
2TFT	0.15**	0.00	0.03
	(0.07)	(0.00)	(0.02)
2TF2T	0.00	0.08	0.12**
	(0.00)	(0.06)	(0.05)
Grim	0.15*	0.03	0.04*
	(0.08)	(0.02)	(0.02)
Grim2	0.16*	0.09	0.05*
	(0.09)	(0.05)	(0.03)
Grim3	0.00	0.00	0.11***
	(0.06)	(0.00)	(0.04)
ALLD	0.07*	0.05	0.23***
	(0.04)	(0.03)	(0.04)
D-TFT	0.09**	0.05	0.00
	(0.04)	(0.03)	(0.00)
Gamma	0.35***	0.44***	0.43***
	(0.03)	(0.03)	(0.02)

Notes: All specifications use $b/c = 4$. Bootstrapped standard errors (shown in parentheses) used to calculate p-values.

B Theoretical analysis of the Payback strategy

In this section, we provide a theoretical analysis of the Payback strategy (and prove results comparing it to the Grim strategy).

B.1 Repeated PD without Noise

We start with a repeated PD without noise. The family of the PD games studied in FRD have stage game payoffs:

	C	D
C	$b-c, b-c$	$-c, b$
D	$b, -c$	$0, 0$

and discount factor (or the probability of continuation) δ.
Payoffs in any symmetric PD can be normalized to be of the form:

	C	D
C	$1, 1$	$-l, 1+g$
D	$1+g, -l$	$0, 0$

(1)

for some $g, l > 0$. Normalizing the FRD payoffs that way we get:

$$g = l = \frac{c}{b-c}.$$

Grim Strategy. For the Grim strategy, the equilibrium path payoff is

$$V_C = \frac{1}{1-\delta}.$$

Grim is an equilibrium if and only if the deviation payoff in the cooperative state is lower than the on-path payoff:

$$1 + g \leq V_C \iff \delta \geq \frac{g}{1+g}.$$

Grim has the lowest possible critical discount among all possible strategies that sustain cooperation in every period (for the game without noise). The reason is that it uses the harshest possible punishment for deviations (in PD, the static Nash payoffs are the min-max payoffs). Hence, the critical discount factor for the Payback strategy has to be weakly higher than for Grim.

Payback Strategy. The Payback Strategy has three states: (C, P, R) (Cooperation, Punishment and Repayment). The on-path payoffs in these three states are:

$$\begin{aligned} V_C &= \frac{1}{1-\delta}, \\ V_P &= 1 + g + \delta V_C, \\ V_R &= -l + \delta V_C. \end{aligned}$$

The IC constraint in the P state is always satisfied because the player is recommended the dominant stage-game action, and the continuation payoff does not depend on his action. The IC constraint in the C state is

$$1 + g + \delta V_R \leq 1 + \delta V_C \Longleftrightarrow \delta \geq \frac{g}{1+l}.$$

The IC constraint in the R state is

$$0 + \delta V_R \leq -l + \delta V_C \Longleftrightarrow \delta \geq \frac{l}{1+l}.$$

Which of the two constraints is binding depends on whether g is larger than l. The critical discount factor is

$$\frac{\max\{l, g\}}{1+l}.$$

This leads to the following observation:

Proposition 1 *Consider a repeated PD without noise with normalized payoffs as in* (1). *If* $g = l$, *then the critical discount factors for the Grim and Payback strategies are the same. If* $g \neq l$, *Grim's critical discount factor is strictly lower.*

So, for the class of games in FRD, in the case of no noise, Grim and Payback are equilibria for the same set of parameters.

B.2 Repeated PD with Noise

In the rest of this Appendix we focus on the class of games analyzed in FRD. We use their normalization $c = 2$. We denote $G \equiv b/c$. Let $E < \frac{1}{2}$ be the probability that the realized action is changed from the player's intended action (independently across the two players). The expected stage game payoffs of player 1 as a function of intended actions are then:

	C	D
C	$2(G-1)(1-E)$	$2(E(1+G)-1)$
D	$2G - 2E(1+G)$	$2E(G-1)$

These formulas can be used to replicate the expected payoffs in Figure 1 of FRD.

Payback Strategy. The expected payoffs in the three states are respectively:

$$\begin{aligned}
V_C &= 2(G-1)(1-E) + \delta(V_C(1-E)^2 + (1-E)EV_P + (1-E)EV_R + E^2V_C), \\
V_P &= 2G - 2E(1+G) + \delta(V_C(1-E)^2 + (1-E)EV_C + (1-E)EV_P + E^2V_P), \\
V_R &= 2(E(1+G)-1) + \delta(V_C(1-E)^2 + (1-E)EV_C + (1-E)EV_R + E^2V_R).
\end{aligned}$$

This can be simplified to:

$$\begin{aligned} V_C &= 2\frac{(G-1)(1-E)}{(1-\delta)(1+E\delta(1-2E))} \\ V_P &= \frac{2G(1-E)-2E+\delta(1-E)V_C}{1-\delta E} \\ V_R &= \frac{2(E(1+G)-1)+\delta(1-E)V_C}{1-\delta E} \end{aligned}$$

These formulas can be used to compute the on-path payoffs for that strategy reported in (1).
To check whether Payback is an equilibrium, we need to check the IC constraints in the three states:

Step 1: IC constraints in State C.

A defection in State C gives payoff

$$V_{DC} = 2G - 2E(1+G) + \delta(V_R(1-E)^2 + (1-E)EV_C + (1-E)EV_C + E^2V_P)$$

For this not to be profitable, we must have

$$V_{DC} \leq V_C$$

The critical discount factor solves:

$$\begin{aligned} & 2G - 2E(1+G) + \delta(V_R(1-E)^2 + (1-E)EV_C + (1-E)EV_C + E^2V_P) \\ = \; & 2(G-1)(1-E) + \delta(V_C(1-E)^2 + (1-E)EV_P + (1-E)EV_R + E^2V_C). \end{aligned}$$

The solution is:

$$\delta_C = \frac{2}{V_C - V_R + E(V_P + V_R - 2V_C)} \tag{2}$$

Plugging in the formulas for the payoffs and solving for the critical δ we get that it is:

$$\delta_C = \frac{-G(1-2E)(1-E) - E + \sqrt{(G(1-2E)(1-E)+E)^2 + 4E(1-2E)(1-E)}}{2E(1-2E)(1-E)}$$

Step 2: IC constraints in State P.

The IC constraint is always satisfied in state P. Note that the transitions from that state are independent of player 1 actions. Hence the continuation payoff of that player does not depend on his action and hence the best response is just myopic best response which is the recommended action in that state.

Step 3: IC constraints in State R.

In that state the recommended action is C. A defection to D yields expected payoff:

$$V_{DR} = 2E(G-1) + \delta(V_R(1-E) + EV_C)$$

For this not to be profitable, we must have

$$V_{DR} \leq V_R = 2(E(1+G)-1) + \delta(V_C(1-E) + EV_R)$$

The critical discount factor in that state is:

$$\delta_R = \frac{2}{V_C - V_R}$$

Recall from (2) that the critical discount factor in state C was

$$\delta_C = \frac{2}{V_C - V_R + E(V_P + V_R - 2V_C)}$$

Since

$$V_P + V_R - 2V_C = -2\frac{(G-1)(1-2E)}{1+\delta E - 2E^2\delta} < 0,$$

we have that $\delta_C \geq \delta_R$ for all parameters. Hence:

Proposition 2 *Payback is an equilibrium if and only if*

$$\delta \geq \delta_{SPE,Payback} \equiv \frac{-G(1-2E)(1-E) - E + \sqrt{(G(1-2E)(1-E)+E)^2 + 4E(1-2E)(1-E)}}{2E(1-2E)(1-E)}.$$

This formula can be used to verify our calculations in Table 1 for the critical discount factor for the Payback strategy.

Grim Strategy. The Grim Trigger Strategy has two states: (C,P) (Cooperate, Punishment). The expected payoffs in the two states are respectively:

$$\begin{aligned} V_C &= 2(G-1)(1-E) + \delta(V_C(1-E)^2 + (1-(1-E)^2)V_P), \\ V_P &= \frac{2E(G-1)}{1-\delta}. \end{aligned}$$

Solving for V_C we get:

$$V_C = 2(G-1)\frac{(1-\delta)(1-E) + E^2\delta(2-E)}{(1-\delta)(1-\delta+E\delta(2-E))}.$$

When is Grim an equilibrium? The IC constraint in the Punishment state is always satisfied since the players play infinite repetition of the static Nash equilibrium. The IC constraint in the Cooperate state is

$$2G - 2E(1+G) + \delta((1-E)EV_C + (1-(1-E)E)V_P) \quad \leq V_C.$$

It holds when:

$$\delta \quad \leq \quad \delta_{SPE,Grim} \equiv \frac{1}{(1-E)(G(1-2E)+E)}.$$

B.3 Comparing the critical factor for the Grim and Payback strategies

We claim that:

Proposition 3 *For all $G > 1$, and $E \in (0, \frac{1}{2})$, the critical discount factor for the Payback strategy is lower than for Grim: $\delta_{SPE,Payback} < \delta_{SPE,Grim}$.*

Proof. We first define $X \equiv (G(1-2E)(1-E)+E)$ and $A \equiv E(1-2E)(1-E)$. In this notation we can write:

$$\begin{aligned} \delta_{SPE,Payback} &\equiv \frac{-X+\sqrt{X^2+4A}}{2A}, \\ \delta_{SPE,Grim} &= \frac{1}{X-E^2}. \end{aligned}$$

Let Z be the ratio of these critical discount factors:

$$Z \equiv \frac{\delta_{SPA,PB}}{\delta_{SPA,Grim}} = \frac{(X-E^2)(-X+\sqrt{X^2+4A})}{2A}.$$

First, note that the limit of Z as $G \to \infty$ is 1

$$\lim_{X \to +\infty} \frac{(X-E^2)(-X+\sqrt{X^2+4A})}{2A} = 1.$$

We now show that $Z < 1$ for all parameters by showing that it is increasing in G (and using that as $G \to \infty$ the ratio converges to 1).

Second, we have that the derivative of the numerator of Z with respect to X is:

$$\begin{aligned} & \frac{\partial}{\partial X}((X-E^2)(-X+\sqrt{X^2+4A})) \\ = & -\frac{(2X-E^2)\sqrt{X^2+4A}-2X^2-4A+E^2X}{\sqrt{X^2+4A}} \end{aligned}$$

To show that this is positive, we note that $\frac{\partial Z}{\partial X}$ is zero if $A = 0$, and the numerator of $\frac{\partial Z}{\partial X}$ is increasing in A, so that the derivative is positive for all $A > 0$:

$$\begin{aligned} \frac{\partial}{\partial X}((X-E^2)(-X+\sqrt{X^2+4A})) \text{ at } A &= 0 \text{ is } 0 \\ \frac{\partial^2 Z}{\partial X \partial A} &= 2\frac{E^2-2X+2\sqrt{X^2+4A}}{\sqrt{X^2+4A}} > 0. \end{aligned}$$

So for all parameters $\frac{\partial Z}{\partial X} > 0$, which implies that indeed Z is strictly increasing in G and hence, for all parameters, $Z < 1$. ■

There are several reasons why Payback does better than Grim in the repeated game with noise. Most importantly, Payback uses asymmetric punishments while Grim uses symmetric punishments.

When the players move from (C,C) to (D,D) the sum of payoffs drops approximately from $2(b-c)$ to 0 (ignoring noise). In contrast, when they move to (D,C) or (C,D), the sum of payoffs drops approximately only to $(b-c)$, while providing even stronger punishment per period (the deviating player loses $1+l$ instead of 1). So the false-positive punishments caused by noise are just much more efficient (in terms of keeping a high sum of continuation payoffs while satisfying the IC constraints) in the Payback strategy than in Grim. That improves on-path expected payoffs in the cooperative state and relaxes the Payback strategy's IC constraints.

B.4 Basin of Attraction.

We now present the calculations for the size of the basin of attraction for the two strategies when compared to ALLD.

BAD for Payback. When I play Payback and meet another player that plays Payback, my expected payoff is

$$V_{Payback,Payback} = 2\frac{(G-1)(1-E)}{(1-\delta)(1+E\delta(1-2E))}.$$

When I meet a player that always plays D, my expected payoff is a solution to the system:

$$\begin{aligned} U_C &= 2(E(1+G)-1) + \delta((1-E)^2 U_P + 2E(1-E)U_C + E^2 U_R), \\ U_P &= 2E(G-1) + \delta((1-E)U_P + EU_C), \\ U_R &= 2(E(1+G)-1) + \delta((1-E)U_C + EU_R). \end{aligned}$$

The solution is:

$$V_{Payback,ALLD} = 2\frac{\delta(1-E)^2(2\delta E^2 - \delta E + 1) - 2\delta E^2 G(1-E) + E(1+G) - 1}{(1-\delta)(1-2\delta E(1-E))}.$$

If instead I play ALLD and I meet a player that plays ALLD, I get a payoff:

$$V_{ALLD,ALLD} = \frac{2E(G-1)}{1-\delta}.$$

Finally, if I play ALLD and I meet a player that plays Payback, I get a payoff that is a solution to the following system of equations (where states are now representing the other player).

$$\begin{aligned} U_C &= 2G - 2E(1+G) + \delta((1-E)^2 U_P + 2E(1-E)U_C + E^2 U_R), \\ U_P &= 2E(G-1) + \delta((1-E)U_P + EU_C), \\ U_R &= 2G - 2E(1+G) + \delta((1-E)U_C + EU_R). \end{aligned}$$

The solution is:

$$V_{ALLD,Payback} = 2\frac{G(1-\delta) - (1+G)E + EG(1-2E)(1-E)^2\delta^2 - E\delta(-2G + EG - 2E + 2E^2)}{(1-\delta)(1-2\delta E(1-E))}.$$

Let α be the fraction of other players using the Payback strategy. Player 1 is indifferent between ALLD and Payback if

$$\alpha V_{Payback,Payback} + (1-\alpha)V_{Payback,ALLD} = \alpha V_{ALLD,Payback} + (1-\alpha)V_{ALLD,ALLD}.$$

Solving for α we get:

$$\alpha^*_{Payback} = \frac{(1-\delta E(1-E))(1+\delta E(1-2E))(1-\delta(1-E))}{(1-2E)(1-E)(1-\delta^2 E^2(1-E))(G-1)\delta}.$$

This formula can be used to verify our calculations in Table 1.

BAD for Grim. When I play Grim and meet another player that plays Grim, my expected payoff is

$$V_{Grim,Grim} = 2(G-1)\frac{(1-\delta)(1-E)+E^2\delta(2-E)}{(1-\delta)(1-\delta+E\delta(2-E))}.$$

When I meet a player that always plays D, my payoff is a solution to the system:

$$\begin{aligned} U_C &= 2(E(1+G)-1)+\delta((1-E),EU_C+(1-(1-E)E)U_P) \\ U_P &= \frac{2E(G-1)}{1-\delta}, \end{aligned}$$

The solution is:

$$V_{Grim,ALLD} = 2\frac{\delta E^2(G-1)(-1+E)+(G+1-2\delta)E-1+\delta}{(1-\delta)(1-\delta E+\delta E^2).}$$

If instead I play ALLD and I meet a player that plays ALLD, I get a payoff:

$$V_{ALLD,ALLD} = \frac{2E(G-1)}{1-\delta}.$$

Finally, if I play ALLD and I meet a player that plays Grim, I get a payoff that is a solution to the following system of equations (where states are now representing the other player):

$$\begin{aligned} U_C &= 2G-2E(1+G)+\delta((1-E)EU_C+(1-(1-E)E)U_P), \\ U_P &= \frac{2E(G-1)}{1-\delta}. \end{aligned}$$

The solution is:

$$V_{ALLD,Grim} = 2\frac{G(1-\delta)-\delta E^2(G-1)(1-E)-(1+G(1-2\delta))E}{(1-\delta)(-\delta E+\delta E^2+1)}.$$

Let α be the fraction of other players using the Grim strategy. Player 1 is indifferent between ALLD and Grim if

$$\alpha V_{Grim,Grim} + (1-\alpha)V_{Grim,ALLD} = \alpha V_{ALLD,Grim} + (1-\alpha)V_{ALLD,ALLD}.$$

Solving for α we get:

$$\alpha^*_{Grim} = \frac{1-\delta+E\delta(2-E)}{\delta(1-2E)(1-E)(G-1)}.$$

Proposition 4 *The BAD for Grim is larger than for Payback.*

Proof. Using our formulas for α^*_{Grim} and $\alpha^*_{Payback}$ we get that the difference is:

$$\begin{aligned} & \alpha^*_{Grim} - \alpha^*_{Payback} \\ = \quad & E\frac{E^4\delta^2 + 2\delta(1-\delta)E^3 + 1 - \delta E(1 + E(1-\delta))}{(1-2E)(1-E)(1-\delta^2E^2 + \delta^2E^3)(G-1)} \end{aligned}$$

The denominator is positive, and the numerator is positive because

$$1 - \delta E(1 + E(1-\delta)) \geq 1 - E > 0.$$

■

References

Athey, S., & Bagwell, K. (2001). Optimal collusion with private information. *RAND Journal of Economics*, 428–465.

Bigoni, M., Camera, G., & Casari, M. (2013). Strategies of cooperation and punishment among students and clerical workers. *Journal of Economic Behavior & Organization*, *94*, 172–182.

Bigoni, M., Camera, G., & Casari, M. (2019). Partners or strangers? cooperation, monetary trade, and the choice of scale of interaction. *American Economic Journal: Microeconomics*, *11*(2), 195–227.

Blonski, M., Ockenfels, P., & Spagnolo, G. (2011). Equilibrium selection in the repeated prisoner's dilemma: Axiomatic approach and experimental evidence. *American Economic Journal: Microeconomics*, *3*(3), 164–92.

Blonski, M., & Spagnolo, G. (2015). Prisoners' other dilemma. *International Journal of Game Theory*, *44*(1), 61–81.

Dal Bó, P., & Fréchette, G. R. (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, *101*(1), 411–29.

Dal Bó, P., & Fréchette, G. R. (2018). On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature*, *56*(1), 60–114.

Dal Bó, P., & Fréchette, G. R. (2019). Strategy choice in the infinitely repeated prisoner's dilemma. *American Economic Review*, *109*(11), 3929–52.

Dvorak, F., & Fehrler, S. (2018). *Negotiating cooperation under uncertainty: Communication in noisy, indefinitely repeated interactions* (IZA Discussion Papers No. 11897).

Embrey, M., Mengel, F., & Peeters, R. (2016). *Eliciting Strategies in Indefinitely Repeated Games of Strategic Substitutes and Complements* (SSRN Scholarly Paper No. ID 2896321).

Fudenberg, D., Rand, D. G., & Dreber, A. (2012). Slow to anger and fast to forgive: Cooperation in an uncertain world. *American Economic Review*, *102*(2), 720–49.

Green, E. J., & Porter, R. H. (1984). Noncooperative collusion under imperfect price information. *Econometrica*, 87–100.

Harrington, J. E. (2006). How do cartels operate? *Foundations and trends in microeconomics* (pp. 1–105). Now Publishers Inc.

Harrington, J. E., & Skrzypacz, A. (2007). Collusion under monitoring of sales. *The RAND Journal of Economics*, *38*(2), 314–331.

Levin, J. (2003). Relational incentive contracts. *American Economic Review*, *93*(3), 835–857.

Romero, J., & Rosokha, Y. (2019). The Evolution of Cooperation: The Role of Costly Strategy Adjustments. *American Economic Journal: Microeconomics*, *11*(1), 299–328.

Skrzypacz, A., & Hopenhayn, H. (2004). Tacit collusion in repeated auctions. *Journal of Economic Theory*, *114*(1), 153–169.

Van Damme, E. (1989). Renegotiation-proof equilibria in repeated prisoners' dilemma. *Journal of Economic theory*, *47*(1), 206–217.