Optimization Incentives, Learning, and Strategy Selection in Indefinitely Repeated Prisoner's Dilemmas —Pre-Analysis Plan—

Maximilian Andres Yongping Bao Yves Breitmoser
Sebastian Fehrler Juri Nithammer

June 10, 2026

1 Conceptual Framework

This section defines the repeated-game environment, the optimization incentive, the strategy representations, and the competing behavioral interpretations.

1.1 Stage games and standard thresholds

Subjects play indefinitely repeated Prisoner's Dilemmas. The row player's stage-game payoff matrix is

$$\begin{array}{c|cc} & C & D \\ \hline C & R & S \\ D & T & P \end{array},$$

where $T > R > P > S$ and $2R > T + S$. The continuation probability is fixed at

$$\delta = \frac{2}{3}.$$

Following Stahl (1991) and the subsequent repeated-games literature, the normalized gain from unilateral defection and the normalized loss from unilateral cooperation are

$$g = \frac{T-P}{R-P} - 1, \qquad \ell = -\frac{S-P}{R-P}.$$

The standard subgame-perfection and risk-dominance thresholds for Grim are

$$\delta^{SPE} = \frac{T-R}{T-P}, \qquad \delta^{RD} = \frac{T-R+P-S}{T-S}.$$

The corresponding basin cutoff for Always Defect against Grim is denoted by

$$\omega^* = \frac{(1-\delta)\cdot(P-S)}{(R-P)-(1-\delta)\cdot(T+S-2\cdot P)}.$$

Within each strategic core of the experiment, g, ℓ, δ^{SPE}, δ^{RD}, and ω^* are held fixed while the cardinal optimization incentive varies. The payoff kernel is also held fixed across all treatments.

1.2 Optimization incentive

For the benchmark comparison between Grim and Always Defect, the optimization premium is

$$r(\omega) = V(GT,\omega) - V(AD,\omega) = \Delta(\omega - \omega^*),$$

where ω is the perceived probability that the opponent uses the cooperative strategy and

$$\Delta(T,R,P,S,\delta) = \frac{R-P}{1-\delta} - (T+S-2P)$$

is the optimization incentive. The optimization incentive captures the cardinal payoff distance between the better and worse strategic response, conditional on the same best-response geometry.

2 Experimental Design

This section presents the experimental design to study the effects of changes in the optimization incentives on round-1 cooperation, continuation play after cc, cd, dc, dd and the stratey-type shares.

2.1 Treatment structure and payoff construction

The treatment structure varies optimization incentives within fixed normalized strategic cores. The experiment uses a 2×2 between-subject design. The first dimension varies the optimization-incentive level,

$$\Delta_{Low}, \quad \Delta_{High},$$

and the second dimension varies the strategic core,

$$notRD, \quad RD.$$

Within each strategic core, the two optimization-incentive levels are constructed by multiplying the same payoff-difference vector by

$$3, \quad 12.$$

For each strategic core, write

$$b = R - P, \qquad x = T - P, \qquad y = P - S.$$

The payoffs are generated by

$$R = P + b, \qquad T = P + x, \qquad S = P - y,$$

where

$$(b, x, y) = s(b_0, x_0, y_0)$$

and

$$s \in \{3, 12\}.$$

The normalized parameters are

$$g = \frac{x}{b} - 1, \qquad \ell = \frac{y}{b}.$$

Within each core, scaling (b, x, y) preserves g, ℓ, δ^{SPE}, δ^{RD}, and ω^*, while scaling the optimization incentive by the same factor.

The payoff kernel is fixed at

$$\bar{u} = \frac{R + S + T + P}{4} = 50$$

in every treatment cell. For a given core k and scale s, this pins down the punishment payoff as

$$P_{k,s} = 50 - \frac{s(b_{0,k} + x_{0,k} - y_{0,k})}{4}.$$

Tables 1 and 2 report the full treatment design.

Table 1: Strategic cores and theoretical indicators

Core	b_0	x_0	y_0	g	ℓ	δ^{SPE}	δ^{RD}	ω^*
notRD	2	5	3	1.500	1.500	0.600	0.750	0.750
RD	2	3	1	0.500	0.500	0.333	0.500	0.250

Core	Δ_{Low}	Δ_{High}
notRD	12	48
RD	12	48

Table 2: Row player's payoff matrices by treatment

	not RD			RD		
		C	D		C	D
Δ_{Low}	C	53	38	C	53	44
	D	62	47	D	56	47
		C	D		C	D
Δ_{High}	C	62	2	C	62	26
	D	98	38	D	74	38

2.2 Supergames and matching

Subjects play many indefinitely repeated supergames. Because of time restrictions, they will play as many as fit given the time restrictions of experiments. In each round they choose between two neutrally labeled actions. After each round, the supergame continues with probability $\delta = 2/3$. Subjects are rematched across supergames using a stranger-matching protocol within matching groups.

3 Hypotheses

This section states the pre-registered hypotheses and separates the central repeated-game hypothesis from auxiliary benchmark predictions.

Hypothesis 1 (Effect of the optimization incentives)**.** *Optimization incentives affect convergence, but not continuation play or late within-core behavior.*

(a) For round-1 cooperation, higher optimization incentives move behavior more quickly toward the best-response prediction. The directional prediction is

$$C^1_{\Delta_{Low},k} \geq C^1_{\Delta_{High},k}$$

for $k \in \{NotRD\}$*, and*

$$C^1_{\Delta_{Low},k} \leq C^1_{\Delta_{High},k}$$

for $k \in \{RD\}$*.*

(b) For continuation play, cooperation after cc, cd, dc, and dd is invariant across optimization-incentive levels throughout the experiment.

Hypothesis 2 (Strategy representation)**.** *The behavior-strategy model* $AD/CSG/SSG$ *fits hot-play behavior better than the pure-strategy model* $AD/GT/TFT$*.*

The experiment compares two main strategy representations. The pure-strategy benchmark uses

$$\mathcal{S}^P = \{AD, GT, TFT\}.$$

This menu corresponds to the strategy-estimation tradition emphasizing Always Defect, Grim, and Tit-for-Tat (Dal Bó and Fréchette, 2011, 2019).

The behavior-strategy benchmark uses

$$\mathcal{S}^B = \{AD, CSG, SSG\}.$$

The cooperative strategies are memory-1 behavior strategies,

$$CSG = (0.33, 0.95, 0.33, 0.33, 0.05),$$

$$SSG = (0.95, 0.95, 0.33, 0.33, 0.05).$$

The entries are cooperation probabilities in states

$$(\emptyset, cc, cd, dc, dd),$$

where $\emptyset$ denotes the first round of a supergame and xy denotes that the player chose x and the co-player chose y in the previous round. The continuation parameters are chosen to be close to the recent estimates in Fudenberg and Karreskog Rehbinder (2024), who find high cooperation after cc, low cooperation after dd, and intermediate cooperation after mixed histories. The distinction between cautious and strong semi-grim follows Backhaus and Breitmoser (2026): the two cooperative types share continuation play and differ mainly in first-round cooperation.

We also estimate a flexible benchmark model,

$$AD + 2P5,$$

where the two cooperative types are unrestricted memory-1 behavior strategies. This model is included to test whether restriction to the strategy families $\mathcal{S}^P$ and $\mathcal{S}^B$ is too restrictive.

4 Power Analysis

We conducted a power analysis for Hypothesis 1 (a) based on the results of the meta-study presented in Andres and Nithammer (2026). With data from existing laboratory experiments (see Dal Bó and Fréchette (2018) and Fudenberg and Karreskog Rehbinder (2024)), we estimated the marginal effects of the optimization incentive on cooperation; see model (4) in Table 2 of Andres and Nithammer (2026). Using the estimated logistic function as a model for the underlying data-generating process, we ran a Monte Carlo simulation to estimate power. This thus takes as benchmark a (calibrated) logistic response model.

The Monte Carlo simulation takes the coefficients and individual-level clustered variance-covariance matrix from model (4) in Table 2 of Andres and Nithammer (2026) as input. We draw 200 plausible coefficient vectors from this distribution and, for each draw, simulate 500 hypothetical experiments—yielding 100,000 simulated experiments in total. In each simulated experiment, we compute the predicted cooperation probability in each treatment by plugging the respective optimization incentive (Δ_{Low} vs. Δ_{High}) and ω^* values into the logistic function, generate experimental data with additional within-matching-group dependence of 0.1, and estimate a logistic regression on first-round behavior with standard errors clustered at the matching group level, one model for NotRD and one for RD:

$$C_{i,s}^{1} = \beta_0 + \beta_1 \cdot \Delta_{High} + \epsilon_{i,s}$$

Power is the rejection rate of a one-sided 5% test across all 100,000 simulations. In the analysis of the experimental data, we will differentiate between early and late supergames, estimating the model both separately and pooled.

With $N = 8$ matching groups per each of the four treatments, at least eight participants per matching group and 40 supergames, we estimate a power of 82% in NotRD and 80% in RD. This gives us sufficient power to detect persistent effects on first round cooperation rates, if there are any.

References

Andres, M. and Nithammer, J. (2026). Optimization incentives and cooperation in infinitely repeated games. Manuscript.

Backhaus, T. and Breitmoser, Y. (2026). Individual strategy choice in the repeated prisoner's dilemma. Manuscript.

Dal Bó, P. and Fréchette, G. R. (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, 101(1):411–429.

Dal Bó, P. and 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. and Fréchette, G. R. (2019). Strategy choice in the infinitely repeated prisoner's dilemma. *American Economic Review*, 109(11):3929–3952.

Fudenberg, D. and Karreskog Rehbinder, G. (2024). Predicting cooperation with learning models. *American Economic Journal: Microeconomics*, 16(1):1–32.

Stahl, D. O. (1991). The graph of prisoners' dilemma supergame payoffs as a function of the discount factor. *Games and Economic Behavior*, 3(3):368–384.