Trigger Strategies and Punishment

Game Theory · Module 5 — Repeated Games

Node 5.3

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Repeated Games **Node:** 5.3

A trigger strategy is the canonical device by which long-run interaction disciplines short-run behavior. The name comes from the structural form: a player "triggers" a punishment phase as soon as a designated deviation is observed, and some variant of that punishment is carried out before returning (or not) to the cooperative phase. The three classical variants — grim trigger, tit-for-tat, limited punishment — all share the same architecture: a state-machine with a cooperative state and one or more punishment states, transitions driven by observed play, and payoffs that reward the cooperative state and punish deviation. What makes the architecture interesting is not its simplicity but its power: for sufficiently patient players, a small set of trigger strategies can sustain essentially any individually rational cooperative outcome as an SPE. This node develops the three main trigger strategies, derives their incentive-compatibility conditions, discusses their credibility (are the punishments themselves SPE continuations?), and introduces the **Abreu stick-and-carrot** construction of optimal penal codes. Trigger strategies are the bridge from 5.2's abstract machinery to 5.4's folk theorem.

Formal Definition

Fix an infinitely repeated game G^∞ with discount factor $\delta \in (0,1)$. A **trigger strategy** is a strategy specified by a finite state machine:

- A finite set of **internal states** Q, with designated **cooperative** and **punishment** states.
- An **initial state** $q^1 \in Q$.
- An **action mapping** $\alpha : Q \to S_i$ specifying what action to play in each state.
- A **transition function** $\tau : Q \times S \to Q$ updating the state based on the realized stage profile.

The strategy proceeds by: play $\alpha(q^t)$ in period t, observe the realized profile s^t, update $q^{t+1} = \tau(q^t, s^t)$.

Three canonical trigger strategies.

Grim trigger. States $Q = \{C, D\}$ (cooperate, punish). Initial $q^1 = C$. Action: $\alpha(C)$ = cooperative action, $\alpha(D)$ = stage NE action. Transition: $\tau(C, s) = C$ if s is the target cooperative profile, $\tau(C, s) = D$ if anyone deviated, $\tau(D, \cdot) = D$. The punishment is forever.

Tit-for-tat. States $Q = \{C, D\}$. Action: $\alpha(C) = C$, $\alpha(D) = D$. Transition: $\tau(q, s)$ depends on what the *opponent* played in the previous period. In the two-player Prisoner's Dilemma: $\tau(\cdot, (C, C)) = C$, $\tau(\cdot, (C, D)) = D$ (if I am player 1 and opponent played D), $\tau(\cdot, (D, C)) = C$, $\tau(\cdot, (D, D)) = D$. Copy the opponent's last move.

Limited punishment (k-period reversion). States $Q = \{C, P_1, P_2, \ldots, P_k\}$. Actions: $\alpha(C)$ = cooperate, $\alpha(P_j)$ = punish. Transition: $\tau(C, s) = C$ if cooperation, $\tau(C, s) = P_1$ if deviation; $\tau(P_j, \cdot) = P_{j+1}$ for $j < k$; $\tau(P_k, \cdot) = C$. Punish for exactly k periods then return.

Credibility condition (SPE). A trigger strategy profile is SPE iff at every history — including histories reached by deviations — the prescribed action is a best response given the strategy's future play. This requires that the *punishment* phase itself be SPE: no player can profitably deviate from the punishment prescription.

Credibility fails for certain naive trigger strategies. For example, "player 2 punishes player 1 by taking an action that gives player 2 payoff less than what they would get by not punishing" — this is only SPE if the punishment phase's future itself rewards player 2 for punishing. Constructing credible punishment phases is the subject of **Abreu's optimal penal codes**.

Intuition

Trigger strategies formalize the intuition: "if you deviate, I will punish you." For the threat to be effective, two things must be true. First, the punishment must be severe enough (relative to the one-shot gain from deviation) to deter the deviation. This is the **incentive-compatibility condition**. Second, the punishment must be credible — I must actually be willing to carry it out when the time comes, which means the punishment phase itself must be an SPE continuation. This is the **credibility condition**.

Grim trigger is the simplest trigger strategy: threaten the most severe punishment (reversion to stage NE forever) and it is always credible (the stage NE forever is itself an SPE continuation, because the stage NE is a Nash of the one-shot game at every period). The cost of grim trigger is that once the punishment is triggered, cooperation is lost forever — there is no way back. This is unrealistic for many applications and motivates more forgiving strategies.

Tit-for-tat is the opposite extreme: punish for exactly one period, and be willing to return to cooperation if the opponent returns. Tit-for-tat is famous from Axelrod's tournaments: in the repeated Prisoner's Dilemma, tit-for-tat won Axelrod's classic tournament against dozens of alternative strategies. Its appeal is that it is "nice" (never defects first), "retaliatory" (punishes defection), "forgiving" (stops punishing when the opponent returns), and "clear" (easy to recognize and respond to). However, tit-for-tat is **not** always SPE — it often fails the credibility condition when both players are using it, because the one-period punishment is not always severe enough to deter the one-period deviation gain. Axelrod's tournament used a different criterion (average payoff against a population of opponents), not SPE.

Limited-punishment strategies interpolate: punish for k periods then return. By choosing k appropriately, you can tune the severity of punishment to just deter deviation while preserving the possibility of future cooperation. These strategies tend to be SPE in richer parameter regimes than tit-for-tat but are more complex.

The deepest insight of 5.3 is **Abreu's optimal penal code construction** (1986–1988): to support a given cooperative profile as SPE with the weakest possible requirements on δ, you do not want a punishment phase that reverts to stage NE; you want a punishment phase that minimizes the deviator's continuation payoff while remaining SPE itself. This is typically accomplished by a "stick-and-carrot" scheme: after deviation, first punish the deviator harshly for a few periods (the "stick"), then return to cooperation (the "carrot"). The stick gives the punishment bite; the carrot makes cooperation attractive enough to restore. The optimal penal code is the one that squeezes the deviator's continuation payoff down to the minmax value — which is the tight bound from 5.2 on SPE continuation payoffs.

Structural Role

5.3 is the constructive heart of Module 5. It provides the concrete strategies by which 5.4's folk theorem is proved, 5.5's reputation arguments are operationalized, and 5.6's cooperation in the Prisoner's Dilemma is sustained. Without trigger strategies, the infinite-horizon framework of 5.2 is a setup without a payoff; with them, we get a toolkit for constructing any SPE we want (subject to the folk theorem constraints).

The three canonical strategies form a hierarchy of structural complexity:

- **Grim trigger:** simplest, always credible (reverts to stage NE), but unforgiving.
- **Tit-for-tat:** forgiving but not always credible; more subtle.
- **Limited punishment / optimal penal codes:** most flexible, can push to the folk-theorem bound.

They are not just alternative implementations of the same idea. The choice of trigger strategy matters for the tightness of the incentive-compatibility bound, the robustness to errors, the computational complexity of reasoning about the strategy, and the empirical likelihood of human players using it.

Mathematical Development

Incentive compatibility of grim trigger

Fix a target cooperative profile $\bar{s}$ with payoff $u_i(\bar{s}) = v_i^c$ for each player. Let $v_i^d = \max_{s_i} u_i(s_i, \bar{s}_{-i})$ be player i's best one-period deviation payoff, and let $v_i^n = u_i(s^*)$ be the stage Nash payoff (the punishment payoff under grim trigger).

Incentive compatibility. For grim trigger to be SPE, at every on-path history:

$$v_i^c \;\geq\; (1-\delta)\, v_i^d + \delta\, v_i^n$$

which rearranges to

$$\delta \;\geq\; \frac{v_i^d - v_i^c}{v_i^d - v_i^n}.$$

The right-hand side is the critical discount factor δ^*_{grim}. For $\delta \geq \delta^*_{\text{grim}}$, grim trigger sustains $\bar{s}$ as SPE.

Credibility. The punishment phase is "play s^* forever." Is this SPE? At any history in the punishment phase, the continuation is constant s^* forever, with payoff v_i^n. A one-period deviation gives at most $\max_{s_i} u_i(s_i, s^*_{-i}) = v_i^n$ (by definition of Nash equilibrium). So no profitable deviation. The punishment phase is SPE. Grim trigger is credible.

Incentive compatibility of tit-for-tat in the Prisoner's Dilemma

Stage payoffs as before: $(C, C) = (3, 3), (C, D) = (0, 5), (D, C) = (5, 0), (D, D) = (1, 1)$.

Tit-for-tat: player 1 plays C if player 2 played C last period, else D; symmetrically for player 2. Initial state is cooperate.

On-path check (both cooperating). Is it best to continue cooperating? If I deviate to D: payoff this period is 5, opponent (tit-for-tat) plays D next period, I respond to opponent's D by playing D (my tit-for-tat response). So we are now in the (D, D) state. But then my strategy says D is the response to opponent's previous D, and opponent plays D because I played D last period, etc. So we are stuck at (D, D) indefinitely, same as grim trigger.

Wait — actually, in the next period after my defection, opponent plays D and I play D (copying opponent's earlier C from two periods ago — no, tit-for-tat copies the opponent's *most recent* action, which is now D). So both play D. In the period after that, opponent plays D (copying my D) and I play D (copying opponent's D). Still (D, D). We have hit an absorbing state.

Under this particular implementation of tit-for-tat (copy opponent's most recent action), the "punishment" is effectively forever — so the analysis is identical to grim trigger. Critical discount factor: $\delta \geq 1/2$.

But wait, that is not the usual intuition of tit-for-tat. The intuition is "punish for one period then return to cooperation." That requires a different automaton: punish once, then regardless of what happens next, return to cooperating. This "one-period limited punishment" is what is usually meant by tit-for-tat-style forgiveness.

Let me redo with limited punishment: $Q = \{C, P\}$. $\alpha(C) = C, \alpha(P) = D$. Transition: from C, if opponent cooperated, stay in C; else go to P. From P, go to C (regardless).

On-path check (state C, cooperation). If player 1 deviates, the state transitions to P for player 2 (who will punish), but player 1's state also transitions based on opponent's action (which was C, so player 1 stays in C). Next period: player 1 plays C (their state is still C), player 2 plays D (their state is P because player 1 defected). Payoffs this period: player 1 gets 0, player 2 gets 5. Period after: player 2 returns to C (one-period punishment done), player 1 was in state C and opponent played D, so player 1 transitions to P (punishment). So the punishment roles swap. This is not symmetric and does not generally converge back to cooperation without extra structure.

The moral: tit-for-tat in its various incarnations is surprisingly subtle to analyze as SPE. Axelrod's tournament used a different criterion (total payoff over a fixed horizon against a fixed opponent population), and the intuitive appeal of tit-for-tat does not always translate to SPE in the game-theoretic sense. For a clean SPE analysis of Prisoner's Dilemma, grim trigger and symmetric limited-punishment strategies are the standard tools.

Limited-punishment SPE

Consider a symmetric strategy: both players cooperate in state C; upon any deviation, both players play D for exactly k periods, then both return to C. Transitions: from state C, go to P_1 if anyone deviates; from P_j ($j < k$) go to P_{j+1}; from P_k go to C. Both players' states are synchronized.

Incentive compatibility (on-path cooperation).

$$v_i^c \;\geq\; (1-\delta)\, v_i^d + (1-\delta^k)\, v_i^n \cdot \text{normalized term} + \delta^k v_i^c,$$

where the middle term is the punishment phase's normalized contribution. Equivalently:

$$v_i^c - v_i^n \;\geq\; \frac{(1-\delta)(v_i^d - v_i^c)}{\delta(1-\delta^k)/(1-\delta)} = \cdots\,,$$

and the tidiest form is to compare the present value of cooperation v_i^c with the present value of "deviate now, punished for k periods, return to cooperation":

$$v_i^c \;\geq\; (1-\delta)v_i^d + \delta(1-\delta^k)v_i^n \cdot \frac{1-\delta}{1-\delta} + \delta^{k+1} v_i^c.$$

Simpler: the one-period gain from deviation is $v_i^d - v_i^c$, the k-period cost of being punished is $(v_i^c - v_i^n)\sum_{j=1}^{k} \delta^j = (v_i^c - v_i^n) \cdot \delta \cdot (1-\delta^k)/(1-\delta)$. Incentive compatibility:

$$(v_i^d - v_i^c) \;\leq\; (v_i^c - v_i^n) \cdot \frac{\delta(1-\delta^k)}{1-\delta}.$$

Critical discount factor as a function of k. As $k \to \infty$, the right-hand side approaches $(v_i^c - v_i^n) \cdot \delta/(1-\delta)$, which is the grim-trigger bound. For finite k, the bound is looser (higher δ required), but finite k has the advantage that play eventually returns to cooperation — a more realistic model for many applications.

Abreu's optimal penal codes

Abreu (1986, 1988) showed that for any target SPE payoff vector v feasible and individually rational, the tightest SPE construction uses a "stick-and-carrot" two-phase punishment:

- **Stick phase:** play a profile that gives the deviator the lowest possible one-period payoff (typically very negative) for a number of periods.
- **Carrot phase:** return to a cooperative phase that rewards the players (including the deviator) with the target payoff v.

The stick is the mechanism by which the deviator is punished; the carrot is what makes the stick credible (the punishers are willing to carry out the stick because they expect the return to cooperation afterward). The length of the stick phase and the identity of the punishment profile are optimized to minimize the deviator's continuation payoff subject to the SPE constraint.

Theorem (Abreu). For δ sufficiently close to 1, any individually rational feasible payoff v can be sustained as SPE by a suitable optimal penal code. The minimum required δ is characterized by the structure of the stage game.

The Abreu construction is the technical engine behind the folk theorem of 5.4. Its structural importance is that it shows there is no loss of generality in using simple trigger-based strategies: anything achievable by arbitrary strategies is achievable by a stick-and-carrot construction.

Worked Example

Critical discount factor for various trigger strategies in the Prisoner's Dilemma

Stage payoffs as before: $(C,C)=(3,3),(C,D)=(0,5),(D,C)=(5,0),(D,D)=(1,1)$. We have $v^c=3, v^d=5, v^n=1$.

Grim trigger. $\delta \geq (5-3)/(5-1)=1/2$. Critical $\delta^*_{\text{grim}}=0.5$.

Limited punishment, $k=1$ period. Incentive-compatibility:

$$(5-3) \;\leq\; (3-1)\cdot\frac{\delta(1-\delta)}{1-\delta}=2\delta.$$

This gives $\delta \geq 1$, which is not satisfied for any $\delta<1$. So one-period punishment does not sustain cooperation in this Prisoner's Dilemma. The punishment is too mild.

Limited punishment, $k=2$ periods.

$$(5-3) \;\leq\; (3-1)\cdot\delta(1+\delta)=2\delta+2\delta^2.$$

Solve $2\delta^2+2\delta-2\geq 0$, i.e. $\delta^2+\delta-1\geq 0$. $\delta \geq (\sqrt{5}-1)/2 \approx 0.618$.

Limited punishment, $k=3$ periods.

$$2 \;\leq\; 2\delta(1+\delta+\delta^2)=2\delta+2\delta^2+2\delta^3.$$

$\delta^3+\delta^2+\delta-1\geq 0$, $\delta\approx 0.544$.

As $k\to\infty$. Approaches grim trigger's bound $\delta=0.5$.

Summary. Longer punishments sustain cooperation at lower discount factors, with the limit being grim trigger at $\delta=0.5$. Shorter punishments require more patience.

Abreu's stick-and-carrot in a coordination game

Stage game with multiple equilibria, so there is room for harsher punishment than the stage NE. Say:

	L	M	R
L	$(6,6)$	$(-5,7)$	$(-5,-5)$
M	$(7,-5)$	$(4,4)$	$(-5,-5)$
R	$(-5,-5)$	$(-5,-5)$	$(1,1)$

Stage NE: $(M,M)=(4,4)$ and $(R,R)=(1,1)$. Target: sustain $(L,L)=(6,6)$.

Grim trigger to (M,M). Critical $\delta=(7-6)/(7-4)=1/3$. **Grim trigger to (R,R) (harsher punishment).** Critical $\delta=(7-6)/(7-1)=1/6$.

Stick and carrot: punish with (R, R) for $k = 2$ periods, then return to (L, L).

Punishment is credible because after the punishment, players return to (L, L) which gives them payoff 6 — more than the (R, R) stage NE. They are willing to play the punishment because it is followed by a carrot. Incentive:

$$(7 - 6) \;\leq\; (6 - 1) \cdot \delta(1 + \delta) = 5\delta + 5\delta^2.$$

$5\delta^2 + 5\delta - 1 \geq 0$, $\delta \geq 0.175$. Much tighter bound than grim trigger.

Credibility of the punishment. In the stick phase, player 1 is playing R and getting payoff -5 (if the other is also playing R). Why would player 1 play this? Because the continuation is: more punishment, then return to (L, L) with payoff 6. Is this preferred to a one-period deviation? Deviation from R gives payoff -5 anyway (any alternative against R is dominated), so no profitable deviation even at the stage level. The punishment phase is trivially credible because (R, R) is a stage NE.

The stick-and-carrot power here is that (R, R) is a worse stage NE than (M, M), so it bites harder. Because it is still a Nash, it is credible; because it is below the "natural" grim-trigger punishment, it sustains cooperation at a lower δ. This is the essence of Abreu's insight.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Cooperative state	Both players playing balanced, exploitative adjustments off
Punishment state	Player punishes opponent for deviating from norm — aggressive counter-exploitation
Grim trigger	"Once you angle-shoot me, I will 4-bet bluff your range forever"
Limited punishment	"Punish for 100 hands then go back to normal"
Stick and carrot	Harsh short-term exploit to punish, then return to cooperative balance
Credibility	Am I actually willing to execute my threat when the time comes?

Concrete situation: coaching setting with a student

You are coaching a student through long-term poker development. The student occasionally makes unstudied moves (the "deviation"). Your strategy as coach is some form of trigger strategy: in the cooperative phase, you give full attention and deep feedback; in the punishment phase (after a careless mistake), you give reduced feedback, fewer examples, less encouragement; after a few sessions of this the punishment ends and you return to cooperation. The student is deterred from careless moves by the prospect of reduced attention.

This is not metaphor. It is the literal structure of a trigger strategy: finite state machine, deviations transition to punishment, punishment phase of bounded length, return to cooperative state. The parameters of the student's discount factor (how much they care about future coaching quality), the size of the punishment (how much worse the punished phase is), and the one-shot gain from deviating (what they "save" by not doing their homework) determine whether the trigger strategy sustains cooperation.

Concrete situation: tournament cooperation at the bubble

At the tournament bubble, multiple players have an incentive to "cooperate" by not knocking each other out. Big stacks can in principle pressure short stacks but typically do not, because a bubble is a finite repeated interaction with reputational stakes. A big stack that starts maximally pressing the bubble can be "punished" by other bigs targeting them when they become vulnerable. This is a trigger-strategy norm: cooperate on the bubble, and if you defect, you will be punished later.

The credibility question is important: if the punishment phase is "we will target you later," that is only credible if the punishers expect to actually be in position to punish (same tournament later, or next tournament). The discount factor here is the probability that the same configuration of players will recur, which is low in a single tournament but higher on a tour with recurring regulars. At the high-stakes live circuit, these norms are well-developed; at random online MTTs, they barely exist.

What this understanding buys you

Every tacit poker norm is a trigger strategy. When you observe that "regulars don't 3-bet-light each other at the bubble," or "no one uses the cooler line against a friend," or "we don't discuss solver output during live sessions" — these are all trigger-strategy equilibria. They are sustained by the credible threat of future retaliation or exclusion if violated. Recognizing them as such lets you predict when they will and will not hold: when the discount factor (probability of future interaction with the same players) drops, the norms break down.

Choosing your punishment is a design problem. If you want to design an informal agreement with an opponent (or within a group of regulars), you need a punishment that is both severe enough to deter deviation and credible enough to be executed. Grim-trigger-like punishments ("never cooperate again") are easy to make credible but may be overkill. Short punishments may be too mild. Stick-and-carrot structures — harsh but brief, then return to cooperation — are often the best compromise.

Credibility is where most informal agreements fail. The "I will punish you later" threat is only useful if you will actually execute it. Most failures of informal cooperation are failures of credibility, not of severity: the punishment is threatened but not carried out when the time comes, because the punisher would rather move on than incur the cost of executing the punishment. Abreu's stick-and-carrot insight — punishment must be followed by a carrot to make it credible — is useful here: the punisher is willing to pay the cost of the stick because they expect the carrot of restored cooperation afterward.

Cross-Links

- **5.2 Infinitely repeated games and discounting** — the framework in which trigger strategies live.
- **5.4 Folk theorem** — the characterization of all SPE payoffs, proved via Abreu-style stick-and-carrot constructions.
- **5.5 Reputation and chain-store paradox** — an alternative route to cooperation, without explicit trigger strategies.
- **5.6 Cooperation in repeated Prisoner's Dilemma** — the canonical case study for trigger strategies.
- **10.4 Commitment and credibility** — the institutional structures that give trigger-strategy punishments bite.
- **9.5 Convergence to equilibrium** — learning dynamics that can implement trigger-like behavior without explicit planning.

Structural Compression

Invariant: A trigger strategy is a finite-state automaton alternating between cooperative and punishment phases, with incentive compatibility requiring that the one-shot deviation gain be dominated by the discounted cost of the punishment, and credibility requiring that the punishment phase itself be a subgame-perfect continuation; the optimal trigger strategies (Abreu's stick-and-carrot) minimize the discount factor threshold subject to these constraints.

Mechanism: State machine with cooperative and punishment states; transitions driven by observed deviations; incentive compatibility enforced by comparing one-shot deviation gain to discounted punishment cost; credibility enforced by embedding the punishment phase in an SPE continuation (typically via a return-to-cooperation carrot that makes the stick credible).

Cross-System Echo: Trigger strategies are the game-theoretic analogue of **feedback controllers with discrete operating modes**. A feedback controller with a nominal operating mode and a "fault recovery" mode that activates upon observing a deviation is a trigger strategy in everything but name. The literature on hybrid dynamical systems — switched controllers, mode-predictive control — has essentially the same structural problem: design a switching rule such that the closed-loop is stable in the nominal mode, the switching is triggered by observable deviations, and the deviation mode eventually returns to the nominal mode. The analogy is deep and will be referenced in Module 10.

Exercises

1. Derive the critical discount factor for grim trigger in a general two-player game with target cooperative payoff v^c, one-shot deviation payoff v^d, and stage NE payoff v^n.
2. For the Prisoner's Dilemma with payoffs as in the worked example, derive the critical discount factor for k-period limited punishment as a function of k. Confirm that the sequence converges to grim trigger's $\delta = 0.5$ as $k \to \infty$.
3. Show that tit-for-tat (copy opponent's most recent action) can fail to be SPE in the Prisoner's Dilemma even for δ above 0.5. (Hint: check credibility of the punishment phase when both players are using tit-for-tat.)
4. Construct a two-player stage game with at least two stage Nash equilibria (differing in payoff) and compute the Abreu stick-and-carrot critical discount factor for sustaining a non-equilibrium cooperative profile. Compare to grim trigger using the stage NE as punishment.
5. A three-player repeated game has target profile $\bar{s}$ with payoff 4 each and one-shot deviation gain 2 for any player. Two stage NE exist with payoffs $(2, 2, 2)$ and $(1, 1, 1)$. Compare critical discount factors for grim trigger using each punishment.
6. Find a repeated game where the one-period limited punishment strategy is an SPE but grim trigger is not because the grim trigger "punishment forever" is so severe that players prefer to deviate and then accept the punishment over continuing cooperation. Interpret the structural reason.
7. **Argue, structurally,** why credibility, not severity, is the binding constraint in most real-world trigger-strategy equilibria. Give an example from outside game theory (economics, politics, biology) where an informal agreement broke down because the punishment threat was not credible.

Game Theory · Module 5 — Repeated Games · Node 5.3
Concepts: nash-equilibrium, repeated-interaction, feedback
Prerequisites: 5.1, 5.2
Enables: 5.4, 5.5, 5.6

hbar.university — own.your.cognition