Contracts and Principal-Agent Problems

Game Theory · Module 10 — Strategic Systems Integration

Node 10.3

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Strategic Systems Integration **Node:** 10.3

A contract is a **private institution** (10.1) that binds two parties to a set of conditional obligations, transforming their payoffs and constraining their strategies. The principal-agent problem is the canonical strategic situation in which contracts arise: one party (the principal) wants another party (the agent) to take an action, but the agent's action is imperfectly observable and the agent's interests may diverge from the principal's. This creates two distinct information problems — **moral hazard** (hidden action: the principal cannot observe the agent's effort) and **adverse selection** (hidden type: the principal cannot observe the agent's quality). The theory of optimal contracts, developed by Holmstrom (1979), Grossman and Hart (1983), and the mechanism-design tradition of Module 6, asks: what contract maximizes the principal's expected payoff subject to the agent's incentive-compatibility and participation constraints? 10.3 develops this theory formally, derives the key results (Holmstrom's informativeness principle, optimal linear contracts, the tradeoff between risk-sharing and incentive provision), and connects them to the layered-mechanism framework of 10.2 — a contract is the simplest layered mechanism, with the legal system as the outer layer and the bilateral agreement as the inner layer.

Formal Definition

A **principal-agent problem** is defined by:

- **Principal** P and **agent** A, with outside options $\bar{u}_P$ and $\bar{u}_A$ (reservation utilities).
- **Agent's action** $e \in E \subseteq \mathbb{R}_+$ (effort), chosen by the agent after the contract is signed.
- **Output** $x \in X$, a random variable with distribution $F(x|e)$ depending on effort. Output is observable by both parties; effort is **not** observable by the principal (moral hazard).
- **Contract** $w : X \to \mathbb{R}$, a compensation schedule specifying the agent's pay as a function of output.
- **Payoffs:** Principal receives $x - w(x)$; agent receives $w(x) - c(e)$, where $c(e)$ is the cost of effort with $c'(e) > 0, c''(e) > 0$.
- **Agent's preferences:** $U_A = \mathbb{E}[v(w(x))|e] - c(e)$ where v is a von Neumann-Morgenstern utility (possibly concave: risk-averse agent).
- **Principal's preferences:** $U_P = \mathbb{E}[x - w(x)|e]$ (risk-neutral principal, WLOG by standard arguments).

The **moral hazard problem** is:

$$\max_{w(\cdot),e} \ \mathbb{E}[x - w(x)|e]$$

subject to: - **Incentive compatibility (IC):** $e \in \arg\max_{e'} \mathbb{E}[v(w(x))|e'] - c(e')$. - **Participation (IR):** $\mathbb{E}[v(w(x))|e] - c(e) \geq \bar{u}_A$.

The **adverse selection** variant replaces hidden action with hidden type: the agent has a privately known type $\theta \in \Theta$ that determines the cost function $c(e, \theta)$, and the principal designs a menu of contracts $\{(w_\theta(\cdot), e_\theta)\}_{\theta \in \Theta}$

such that each type self-selects the intended contract (IC across types) and participates (IR for each type). This is a screening problem in the sense of Module 6.

Intuition

Every employment relationship, every staking deal, every consulting contract, every delegation of authority is a principal-agent problem. The principal wants something done. The agent does it. The problem is that the principal cannot perfectly monitor the agent's effort, and the agent — rationally — will shirk unless the contract provides sufficient incentive not to.

The central tension of contract design is **risk versus incentives**. A risk-neutral agent could simply be "sold the firm" (pay a fixed fee to the principal, keep all output): this provides perfect incentives (the agent bears all the consequences of their effort) but also transfers all risk to the agent. If the agent is risk-averse, this is costly — the agent demands a risk premium to accept the contract, and the principal ends up paying for both effort and risk-bearing. The optimal contract trades off these two forces: some risk is shifted to the agent to provide incentives, but not all risk, because the agent's risk premium would be too expensive.

Holmstrom's informativeness principle (1979) says: a performance measure should be included in the contract if and only if it is **informative** about the agent's effort — that is, if observing the measure changes the posterior distribution over effort levels. Including uninformative signals adds noise (increasing the risk premium) without providing incentive value (no information about effort). Including informative signals tightens the inference about effort, allowing the principal to provide incentives with less risk.

Adverse selection adds a second layer: even before the effort choice, the principal does not know the agent's type (quality, ability, cost structure). The principal must design a menu of contracts that induces truthful revelation of type, which typically requires leaving **information rents** to high-ability types (they could mimic low-ability types and do less work for the same pay). The tradeoff here is between extraction (getting value from each type) and incentives (getting each type to reveal themselves truthfully).

Structural Role

10.3 sits at the intersection of mechanism design (Module 6) and institutional embedding (10.1). A contract is:

- A **mechanism** in the sense of 6.1: it specifies a message space (reports, output) and an outcome function (payment).
- A **private institution** in the sense of 10.1: it transforms the raw bilateral interaction into a constrained game with legally enforced payoff transfers.
- A **layered mechanism** in the sense of 10.2: the outer layer is the legal system that makes contracts enforceable, and the inner layer is the specific contract terms.

The principal-agent framework also underlies: - **10.4 Commitment and credibility:** Whether a contract is enforceable — whether the parties will actually follow through — is a commitment problem. - **10.5 Political economy:** Many political relationships (voters and politicians, citizens and bureaucrats) are principal-agent problems where the "contract" is the electoral mechanism. - **10.6 Strategic ecosystems:** Staking relationships in poker are principal-agent problems embedded in the broader poker ecosystem.

Mathematical Development

First-best (observable effort)

If the principal can observe and contract on effort directly, the problem reduces to choosing (e^*, w^*) to maximize:

$$\mathbb{E}[x|e] - w^* \quad \text{subject to} \quad v(w^*) - c(e^*) \geq \bar{u}_A$$

The IR constraint binds: $w^* = v^{-1}(\bar{u}_A + c(e^*))$. The optimal effort satisfies $\mathbb{E}[x'|e^*] = c'(e^*)/v'(w^*)$ — marginal product of effort equals marginal cost, adjusted for risk. With a risk-neutral agent ($v(w) = w$): $\mathbb{E}[x'|e^*] = c'(e^*)$, and the agent is paid a fixed wage equal to their reservation utility plus effort cost. No performance pay is needed because effort is directly observable.

Second-best (moral hazard)

With hidden effort, the principal cannot contract on e and must use the output x as a noisy signal. The Lagrangian of the principal's problem yields the first-order condition for the optimal contract:

$$\frac{1}{v'(w(x))} = \lambda + \mu \frac{f_e(x|e)}{f(x|e)}$$

where λ is the multiplier on IR, μ is the multiplier on IC, and f_e/f is the **likelihood ratio** — the sensitivity of the output density to effort. This is the **Holmstrom condition**: optimal pay at output x is increasing in the likelihood ratio $f_e(x|e)/f(x|e)$. Outputs that are more "indicative" of high effort receive higher pay.

Theorem (Holmstrom Informativeness Principle, 1979). Let x and y be two observable signals. If y is informative about e conditional on x — that is, $f(y|x, e)$ depends on e — then the optimal contract uses both x and y. If y is uninformative ($f(y|x, e) = f(y|x)$ for all e), then the optimal contract does not depend on y.

Corollary. Relative performance evaluation (comparing an agent's output to peers') is optimal when peer output is informative about common shocks but not about the agent's effort. This filters out common noise, tightening the signal-to-noise ratio on effort.

Linear contracts

In many applications (and for tractability), attention is restricted to **linear contracts**:

$$w(x) = \alpha + \beta x$$

where α is a fixed salary and $\beta \in [0, 1]$ is the **incentive intensity** (piece rate, bonus coefficient, equity share).

Assume $x = e + \varepsilon$ with $\varepsilon \sim N(0, \sigma^2)$, $c(e) = e^2/2$, and agent utility $U_A = \mathbb{E}[w] - (r/2)\text{Var}(w) - c(e)$ (mean-variance with risk aversion r). Then:

- Agent's optimal effort: $e^*(\beta) = \beta$ (from FOC of agent's problem).
- Agent's certainty equivalent: $\alpha + \beta^2 - \beta^2/2 - (r/2)\beta^2\sigma^2 = \alpha + \beta^2/2 - (r/2)\beta^2\sigma^2$.
- Principal's expected payoff: $e^*(\beta) - \alpha - \beta e^*(\beta) = \beta - \alpha - \beta^2$.

Substituting the binding IR constraint $\alpha = \bar{u}_A - \beta^2/2 + (r/2)\beta^2\sigma^2$:

$$U_P = \beta - \beta^2 - \bar{u}_A + \beta^2/2 - (r/2)\beta^2\sigma^2 = \beta - \beta^2/2 - (r/2)\beta^2\sigma^2 - \bar{u}_A$$

Maximizing over β:

$$\beta^* = \frac{1}{1 + r\sigma^2}$$

This is the **canonical incentive-intensity formula**. Incentive intensity is: - **Decreasing in risk aversion** r: more risk-averse agents get lower-powered incentives. - **Decreasing in noise** σ^2: noisier environments get lower-powered incentives. - Equal to 1 (sell the firm to the agent) only when $r = 0$ or $\sigma^2 = 0$.

Adverse selection: the screening problem

When the agent's type θ is private (e.g., θ determines the cost function $c(e, \theta)$ with $c_\theta < 0$ — higher types have lower cost), the principal designs a menu $\{(w(\theta), e(\theta))\}_\theta$ subject to:

$$\text{IC:} \quad v(w(\theta)) - c(e(\theta), \theta) \geq v(w(\theta')) - c(e(\theta'), \theta) \quad \forall \theta, \theta'$$
$$\text{IR:} \quad v(w(\theta)) - c(e(\theta), \theta) \geq \bar{u}_A \quad \forall \theta$$

Standard results (Mirrlees, Mussa-Rosen): IC implies monotonicity ($e(\theta)$ is nondecreasing in θ), the IR constraint binds only for the lowest type, and all higher types earn **information rents** $\int_{\underline{\theta}}^{\theta} (-c_\theta(e(\tilde{\theta}), \tilde{\theta}))d\tilde{\theta}$. The principal distorts effort downward for all types below the top (to reduce information rents) and assigns efficient effort only to the highest type. This "no distortion at the top" result is the screening analogue of Myerson's optimal auction (6.3).

Worked Example

Poker staking as a principal-agent problem

Setup. A backer (principal) stakes a player (agent) for a 100-session online poker campaign. The backer puts up the buy-ins; the player keeps a share of the profits.

Moral hazard. The backer cannot observe the player's effort directly — they cannot see whether the player is studying, table-selecting carefully, quitting when tilted, or playing their A-game. They observe only the output: session results $x_1, x_2, \ldots, x_{100}$.

Contract structure. A standard staking contract is linear: the player receives fraction β of net profits above a "makeup" threshold (cumulative losses that must be repaid before profit-sharing begins). The backer receives $1 - \beta$. Typical splits are $\beta \in [0.4, 0.6]$.

Applying the linear-contract formula. Let r be the player's risk aversion, σ^2 be the variance of session results. The optimal split is:

$$\beta^* = \frac{1}{1 + r\sigma^2}$$

For a high-variance game (PLO, high stakes) with large σ^2, the optimal β is lower — the player bears less of the variance and the backer bears more, but this weakens the player's incentive to exert effort. For a low-variance game (tight NLH, low stakes) with small σ^2, the optimal β is higher — stronger incentives with less risk.

Informativeness principle applied. Should the backer include non-financial performance measures in the contract? For instance: hours played, hands per hour, number of tables, study-session logs. If these are informative about effort conditional on results (a player who plays 40 hours and runs bad looks different from a player who plays 5 hours and runs bad), then yes — the optimal contract should depend on both results and hours. If hours are uninformative (all players play the same hours regardless of effort, and effort manifests only in decision quality), then no — adding hours to the contract adds noise without information.

Adverse selection. Before signing the contract, the backer does not know the player's true skill level θ. A high-skill player (θ high) produces higher expected output for the same effort. The backer must screen: offer a menu of contracts — e.g., "50/50 split with no makeup" for players who claim high skill (willing to accept because they expect high profits) and "70/30 split with makeup" for players who claim low skill (need protection). If the menu is designed correctly, each type self-selects the intended contract. The high-skill player earns an information rent (they could have taken the low-skill contract and gotten an easy deal). The backer accepts this rent as the cost of learning the player's type.

Moral hazard in session selection. The most insidious moral hazard in staking is not lazy play but **adverse session selection**: the player, staked for cash games, might choose softer games where they have a smaller edge but lower variance (comfortable but low-EV for the backer) rather than tougher games where their edge is larger but variance is higher. The backer's contract should, by the informativeness principle, include measures that are informative about game selection: average stakes played, player-pool quality metrics, table-selection data. Without these, the player's optimal response to a profit-sharing contract is to minimize their own risk (play soft games) rather than maximize expected profit (play the highest-EV games).

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Principal	Backer, staking stable owner, coaching investor
Agent	Staked player, horse
Action e	Effort: study hours, A-game discipline, table selection, tilt control
Output x	Session results (profit/loss in BB or dollars)
Contract $w(x)$	Staking deal: split, makeup, bonuses, volume requirements
Moral hazard	Backer cannot observe effort quality — only results
Adverse selection	Backer cannot observe player's true skill level before signing
Incentive intensity β	Player's profit share (typically 40–60%)
Informativeness principle	Include non-result metrics (hours, game selection) only if informative about effort

Concrete situation: the makeup trap

A player is down $20,000 in makeup (cumulative losses the backer has absorbed). Under a standard makeup contract, the player must earn back $20,000 before any profit-sharing kicks in. The player's marginal incentive for the next session is nearly zero: even if they win $500, it goes entirely to reducing makeup, not to their pocket. The rational (if cynical) response: reduce effort, play carelessly, or quit the staking deal and find a new backer who will offer a clean slate.

This is a textbook moral hazard problem created by the contract structure. The makeup provision is designed to protect the backer (ensuring the player repays losses), but it destroys incentives when losses are large. The contract-design fix: either (a) cap makeup at a level where the player's incentive is still meaningful, (b) offer periodic "makeup resets" conditional on demonstrated effort (a stick-and-carrot structure from 5.3), or (c) switch to a "no-makeup" contract with a lower split, accepting more backer risk in exchange for preserving player incentives.

What this understanding buys you

Every staking deal is a contract-design problem. Negotiating a staking deal without understanding principal-agent theory is like playing poker without understanding pot odds — you might get it right by intuition, but you cannot diagnose mistakes. The informativeness principle tells you which metrics to include. The risk-incentive tradeoff tells you how to set the split. The adverse-selection framework tells you how to screen players before staking them.

Makeup destroys incentives at scale. The deeper the makeup hole, the weaker the player's incentive. This is a direct consequence of the moral-hazard model: when the agent's marginal return to effort goes to zero (because all incremental gains go to repaying debt), effort goes to zero. Backers who do not understand

this will attribute the player's deteriorating performance to "tilt" or "bad attitude" when it is actually a rational response to a broken contract.

Relative performance evaluation applies to staking stables. A backer with multiple horses can apply the informativeness principle by comparing results across players in similar games. If all players in the stable run bad simultaneously, it is a common shock (bad run), not individual shirking. If one player runs bad while others run well, it is more likely individual — the backer should investigate.

Cross-Links

- **6.1–6.3 Mechanism design** — contracts as mechanisms; adverse selection as screening.
- **4.1–4.3 Bayesian games** — the agent's type is private information; BNE is the solution concept.
- **10.1 Games embedded in institutions** — the legal system as the outer institution that makes contracts enforceable.
- **10.2 Layered mechanisms** — contracts as the simplest layered mechanism (legal system + bilateral terms).
- **10.4 Commitment and credibility** — contract enforcement requires credible commitment by both parties.
- **5.3 Trigger strategies** — informal "contracts" sustained by repeated-game punishment rather than legal enforcement.
- **Economics / contract theory** — Holmstrom, Hart, Tirole: the full contract-theory canon.

Structural Compression

Invariant: A principal-agent problem arises whenever one party delegates an action to another under asymmetric information; the optimal contract balances incentive provision (motivating effort) against risk-sharing (insuring the agent), with the informativeness principle determining which signals belong in the contract, and adverse selection requiring menus that trade information rents for truthful type revelation.

Mechanism: The principal designs a contract $w(x)$ mapping observable output to compensation; the agent chooses effort to maximize their expected utility net of effort cost; the optimal contract satisfies the Holmstrom condition (pay proportional to likelihood ratio) under moral hazard and the monotonicity/no-distortion-at-the-top conditions under adverse selection; the incentive intensity of linear contracts is $\beta^* = 1/(1 + r\sigma^2)$, trading off risk aversion and noise against incentive strength.

Cross-System Echo: The principal-agent problem is structurally identical to the **signal-extraction problem** in information theory (the "agent's effort" is the signal, the "output" is the noisy observation, and the "contract" is the decoder that maps observations to actions). Shannon's separation theorem (source coding and channel coding can be optimized independently) fails here: the "encoder" (agent) is strategic and chooses the signal (effort) to maximize their own utility, not the principal's. This strategic encoding is what makes contract theory harder than information theory and what connects it to the mechanism-design tradition.

Exercises

1. Derive the first-best effort level when both principal and agent are risk-neutral and effort is observable. Show that the first-best contract is a "franchise fee" — the agent pays a fixed fee to the principal and keeps all output.

2. For the linear contract model with $x = e + \varepsilon$, $\varepsilon \sim N(0, \sigma^2)$, $c(e) = e^2/2$, and CARA utility with risk aversion r, derive the optimal incentive intensity β^* and the principal's expected payoff. Show that the principal's payoff is decreasing in both r and σ^2.

3. A backer stakes two players with identical observable characteristics. Player 1 generates results $x_1 = e_1 + \varepsilon_1 + \eta$ and Player 2 generates $x_2 = e_2 + \varepsilon_2 + \eta$, where η is a common shock. Show that the optimal contract for Player 1 depends on x_2 (relative performance evaluation) and derive the optimal linear contract $w_1 = \alpha + \beta_1 x_1 - \gamma x_2$.

4. In the adverse-selection staking problem, the backer faces two types: high-skill (θ_H, win rate 5 BB/100) and low-skill (θ_L, win rate 1 BB/100). Design a separating menu of contracts $\{(\beta_H, \alpha_H), (\beta_L, \alpha_L)\}$ such that each type self-selects. Compute the information rent earned by the high-skill type.

5. Show that in the moral-hazard model, a contract that pays the agent a fixed wage regardless of output (pure insurance) leads to zero effort. Then show that a contract that gives the agent 100% of output (pure incentives) provides first-best effort but requires the agent to bear all risk. Conclude that the second-best contract is strictly interior: $0 < \beta^* < 1$ whenever $r > 0$ and $\sigma^2 > 0$.

6. A staking contract has a makeup provision: the player receives $\beta \cdot \max(X - M, 0)$ where X is cumulative profit and M is cumulative losses. Show that as M grows large, the player's marginal incentive to exert effort in the next session approaches zero. Propose a contract modification that preserves incentives.

7. **Argue, structurally,** why the informativeness principle implies that a staking contract should include hours-played data if and only if hours are informative about effort conditional on results. Give a concrete scenario where hours are informative and one where they are not, and explain what the optimal contract looks like in each case.

Game Theory · Module 10 — Strategic Systems Integration · Node 10.3
Concepts: principal-agent, incentive-compatibility, incomplete-information, mechanism-design
Prerequisites: 4.1, 4.3, 6.1, 10.1, 10.2
Enables: 10.4, 10.6

hbar.university — own.your.cognition