INCENTIVE DESIGN WITH SPILLOVERS

KRISHNA DASARATHA, BENJAMIN GOLUB, AND ANANT SHAH

Abstract. A principal uses bonuses conditioned on stochastic outcomes of a team project to elicit costly private efforts from the team members. We characterize the optimal allocation of incentive pay across agents and outcomes under arbitrary smooth team production functions. It is optimal to make the strength of an agent's incentives proportional both to marginal productivity and to a measure of organizational centrality that reflects the strength of complementarities with productive colleagues. Insights from the theory of network games play a crucial role in analyzing how incentives given to one agent spill over to others and shape the optimal contract. The results generalize Holmstrom's characterization of optimal single-agent contracts under uncertainty to the multi-agent case.

Current version of paper available at bengolub.net/incentivedesign

Dasaratha: krishnadasaratha@gmail.com, Boston University. Golub: ben.golub@gmail.com, Northwestern University. Shah: anantshah2026@u.northwestern.edu, Northwestern University. We thank Yann Calvó Lopez, Yu-Chi Hsieh and Vitalii Tubdenov for excellent research assistance. We are grateful (in random order) to Ilya Segal, George Mailath, Aravindan Vijayaraghavan, Stephen Morris, Omer Tamuz, Marina Halac, Thomas Steinke, Marzena Rostek, Jacopo Perego, Juan Ortner, Alex Wolitzky, Jason Hartline, Evan Sadler, Adam Szeidl, Michael Ostrovsky, Elliot Lipnowski, Melika Liporace, Piero Gottardi, and Michael Powell, as well as many seminar and conference participants, for valuable conversations and comments.

Extended abstract

A popular method of motivating the members of a team to work toward a common goal is giving them performance incentives that depend on jointly achieved outcomes. Compensation instruments of this form commonly used in practice include options on the firm's stock, bonuses for achieving sales targets, and profit-sharing. How should such incentive schemes be designed, and how should they take into account the structure of production on the team?

Model. We examine these questions in a model of a team working on a joint project. Each worker chooses a level of effort at an increasing marginal cost. These efforts contribute to a real-valued *team output*, such as the quality of a product. That output, in turn, determines the distribution of an *outcome* that can take a finite number of possible values. The principal can design a contract contingent on realized outcomes to maximize profit.

We now formalize this setup. There are n agents, $N = \{1, 2, \ldots, n\}$, and one principal. The agents take real-valued actions $a_i \geq 0$, which can be interpreted as effort levels. These jointly determine a team *output*, given by a function $Y : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$ which we assume is twice differentiable and strictly increasing in each of its arguments. The team output determines the project *outcome*, an element of the finite set $\mathcal{S}$. The probability of outcome s is $P_s(Y)$, where for any $s \in \mathcal{S}$, the function $P_s(\cdot)$ is strictly positive and twice differentiable. The principal receives *revenue* v_s from the outcome s.[1]

The principal observes the project outcome but does not observe agents' actions or the output Y. (When we use pronouns, we use "she" for the principal and "he" for an agent.) To maximize revenue by incentivizing agents' actions, the principal makes a non-negative payment contingent on the outcome. Upon realization of outcome s, agent i receives payment $\tau_i(s)$. The payments are denoted by $\boldsymbol{\tau} : \mathcal{S} \to \mathbb{R}^n_{\geq 0}$; such a function is called a *contract*.

[1]This should be interpreted as the principal's valuation of that state realizing, gross of any payments she will make to the agents.

We consider risk-averse agents and a risk-neutral principal.[2] The utility to agent i from a monetary transfer is given by the function $U_i : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, which is strictly increasing, concave and differentiable. Each agent also has a private cost function $C_i : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, which is strictly increasing, strictly convex and twice differentiable in that agent's action. The marginal cost at zero action is zero, that is, $C_i'(0) = 0$. Agent i maximizes the expected payoff from payments minus the private cost of the action a_i,

$$\mathcal{U}_i = \sum_{s \in \mathcal{S}} P_s\left(Y\right) U_i\left(\tau_i(s)\right) - C_i(a_i).$$

The payoff for the principal given a contract $\boldsymbol{\tau}$ and team output Y is the expected payoff of the outcome minus transfers to agents:

$$\sum_{s \in \mathcal{S}} \left(v_s - \sum_i \tau_i(s) \right) P_s(Y).$$

The timing is as follows: The principal commits to a contract $\boldsymbol{\tau}$, following which agents' simultaneously choose actions. Our solution concept for the game among the agents is pure strategy Nash equilibrium, which we refer to as the *equilibrium* for the remainder of the paper.

There may be multiple equilibria under some contracts. Given a contract $\boldsymbol{\tau}$, we assume that agents play an equilibrium $\mathbf{a}^*(\boldsymbol{\tau})$ maximizing the principal's expected payoff. Under this selection, a principal's payoff under a contract is well-defined if at least one equilibrium exists. Among such contracts, a contract $\boldsymbol{\tau}$ is *optimal* if no other contract $\widetilde{\boldsymbol{\tau}}$ gives the principal a higher payoff. Implicit in this definition is the assumption that contracts without equilibria can never be optimal.

Example. An example inspired by the literature on network games will be useful to keep in mind.

There is a symmetric matrix $\mathbf{G}$, representing an undirected network; so $G_{ij} \geq 0$ is the weight of the link from agent i to j, and $G_{ii} = 0$ for each i. The output is the sum of a term that is linear in actions—corresponding to agents' standalone

[2]The modelling assumption that a principal is risk-neutral is not crucial to the results. The characterization of an optimal contract and its consequences holds for a risk-averse principal as well.

contributions—and a quadratic complementarity term:

$$Y(\mathbf{a}) = \sum_{i \in N} k_i a_i + \frac{\beta}{2} \sum_{i,j \in N} G_{ij} a_i a_j.$$

Here $k_i > 0$ are arbitrary constants reflecting agents' standalone productivities.

There are two possible outcomes $s \in \{0, 1\}$. The revenues from these outcomes are normalized so that $v_1 = 1$ and $v_0 = 0$. These can be interpreted as success or failure of the project. The probability of success is $P(Y)$, where the function $P(\cdot)$ is strictly increasing, concave, and twice differentiable.

Agents have arbitrary, strictly concave, and identical utility functions for money.

Main results. A principal designing a contract in this environment should take into account the team's production function. To illustrate why, imagine that the principal changes the payments promised to a particular agent in a way that motivates this agent to put in more effort. In team production, changing one team member's effort level typically affects some other players' productivities—i.e., the marginal effect of their effort on team output. That, in turn, changes the incentives that these other agents face, even though the payments promised to them in various outcomes did not change. Thus, payments to one agent can motivate or deter effort by others, and this makes the design of all the different agents' contracts interconnected.

We do not have a good understanding of how the principal should account for the structure the team's production function in this design problem. Several recent works have studied special cases of this problem—e.g. Shi (2022), Dasaratha et al. (2023), and Mayol (2023). These studies have been conducted assuming specific, parametric production functions featuring complementarities, often inspired by the literature on network games. At a conceptual level, these contributions have shown that contract design in this setting is related to optimally controlling (certain aspects of) spillovers in a network game. But little is known about the general incentive design problem.

Our contribution in this paper is a characterization, without parametric assumptions, of optimal incentives in the environment we have described. In brief, the results state that optimal contracts must allocate steeper incentives to agents who have higher productivity and those who are organizationally central, in the sense that they have high direct and indirect complementarities with productive agents. We now formalize this statement.

In order to do this, we define two key concepts that play a central role in our main result: an agent's *productivity* and *centrality*. An agent i's productivity measures how much an incremental increase in i's effort would increase the team's output, holding all other agents' actions fixed. Formally, fixing an action profile $\mathbf{a}$, agent i's productivity is:

$$\kappa_i(\mathbf{a}) = \frac{\partial Y}{\partial a_i}(\mathbf{a}).$$

Note that an agent's marginal productivity can depend on others' effort levels through the team production function $Y(\cdot)$.

To understand an agent's centrality, imagine starting at an equilibrium action profile $\mathbf{a}$ and then slightly perturbing i's incentives, as follows:

$$\mathcal{U}_i(a_i; \tau_i) = \sum_{s \in \mathcal{S}} P_s(Y) U_i(\tau_i(s)) - C_i(a_i) + \delta_i a_i.$$

Here $\delta_i > 0$ represents a small increase in agent i's marginal returns to effort. In the perturbed game, agents play a nearby equilibrium $\widetilde{\mathbf{a}}(\delta_i)$. We define agent i's *centrality* as the rate at which the team's output increases due to this perturbation:

$$c_i(\mathbf{a}) = \frac{d}{d\delta_i} Y(\widetilde{\mathbf{a}}(\delta_i)),$$

where the derivative is evaluated at 0. Though the connection to a network notion is not obvious at this stage, an agent's centrality captures not only his direct effect on output by increasing his own effort, but also his indirect effects by motivating changes in others' effort levels—ripple effects through a network of strategic interactions.

To state the main result, we need one more notion—a measure of an agent's responsiveness to monetary compensation. Define an agent's *compensation index* as

$$\beta_i(\tau_i) = 1/U_i'(\tau_i).$$

Notice that if the agent is paid τ_i, then in order to increase U_i by ϵ, the agent must be given a transfer of $\epsilon \beta_i(\tau_i)$ (for small enough ϵ). We term this quantity the compensation index, since it measures how much extra money the agent must be give to compensate him for a one-unit increase in the cost of effort $C(a_i)$. Note that if U_i is strictly concave, then the compensation index in a given state s is increasing in $\tau_i(s)$: i.e., better-compensated agents have a higher compensation index.

Our main result states that the following holds as long as the equilibrium at an optimal contract is differentiable in the incentive payments.[3]

at any optimal contract $\boldsymbol{\tau}$ with equilibrium action profile $\mathbf{a}^*$, at every outcome where any agent i receives a positive payment $\tau_i(s) > 0$, it holds that

$$\beta_i(\tau_i(s)) = \lambda_s \kappa_i(\mathbf{a}^*) c_i(\mathbf{a}^*), \tag{1}$$

where λ_s is an outcome-specific constant.

Thus, the first-order conditions require paying agents by setting their compensation indices to be equal to "productivity times centrality." In particular, if all agents have the same utility function for money, agents with a higher value of the product $\kappa_i(\mathbf{a}^*)c_i(\mathbf{a}^*)$ (which does not depend on the outcome) should be paid more in at all outcomes where any agent receives incentive pay. Our result also implies simple "hill-climbing" algorithms for improving suboptimal incentive pay by reallocating incentives across individuals.

Each of the factors on the right-hand side of (1) has an interesting reason behind it. Recall the example with a quadratic Y,

$$Y(\mathbf{a}) = \sum_{i \in N} k_i a_i + \frac{\beta}{2} \sum_{i,j \in N} G_{ij} a_i a_j. \tag{2}$$

First, take the case where $\mathbf{G}$ is the zero matrix. In that case, (1) says that agents' compensation indices should be proportional to their individual productivities k_i. More productive agents should be paid more because the outcome is highly responsive to their effort, so incentive pay that is based on the outcome is highly motivating for them, making them responsive recipients of incentive pay. This force is interesting in

[3]A technical contribution of our work is formulating conditions to ensure that this differentiability condition holds generically, in a suitable sense, over primitives of our problem, so that no assumptions on endogenous objects need to be made. That is, we give conditions ensuring that the "first-order approach" may be used to study the response of equilibrium play to the contract. This involves some subtleties that do not arise in the single-agent version.

itself. The centrality effect—the c_i factor in the formula—then accounts for the fact that motivating central agents more results in a higher team output due to strategic ripple effects emanating from that agent. The centrality statistics we use are closely related to those arising in the theory of network games—in particular, Bonacich centrality and a "Leontief"-type matrix accounting for indirect propagation plays an important role in our techniques.

Related to this point, at a technical level, we analyze the optimality of contracts by locally approximating a general smooth Y with a quadratic polynomial. This turns out to give an approximation that is correct for computing derivatives of the principal's payoff in perturbations of a given contract, around an arbitrary equilibrium—even accounting for all strategic spillovers of agents' efforts on others' incentives and actions. That, in turn, allows us to leverage insights from the network games literature for analyzing these spillovers. In effect, we show that we can analyze spillover effects in a general environment using formulas based on a game arising from the very simple production function (2), which affords a great deal of tractability.

Applications and implications. Building on our main result, we explore several implications and extensions.

First, we apply the main characterization (1) in some specific environments to obtain quite explicit descriptions of optimal compensation. For example, when all standalone productivities are equal in the simple parametric setting of (2), our analysis prescribes that equity should be allocated so that each agent who puts forth positive effort has an equal value of $\sum_j G_{ij}\tau_j(1)$. That is, bonuses allocated to the collaborators of working agents are equalized in this weighted sense, implying a remarkably "balanced" distribution of incentive pay throughout the organization.

We also study the model with agents who are nearly risk-neutral,[4] and show that in this case, at an optimal contract, more productive agents must necessarily be *less* organizationally central. In effect, in the risk-neutral case, agents *either* have high marginal products of their own effort or induce large ripple effects when they contribute, but not both. These consequences show that our main result has strong

[4]Note that under that assumption, compensation indices cannot vary much by definition.

substantive implications once the model is applied to settings with some additional structure.

We also characterize how bonuses should be allocated across different outcomes, showing that under concave utility, agents should be paid more in less likely and more sensitive to changes in output. This shows that our characterization generalizes the famous Holmström (1979) result on optimally targeting performance pay across outcomes and collapses to it in the case of one agent. In the case of many agents, Holmstrom's result characterizes how to allocate pay across outcomes, and ours says how to allocate it across agents.

We also study settings where contracts are constrained to take specific forms, such as equity pay—a contract linear in the principal's revenue. Even in this more restrictive contracting environment, our results imply that a compensation index computed at an agent's optimal equity share is proportional to the product of his productivity and centrality. This finding shows our insights are applicable in realistic settings where contracts cannot be perfectly tailored to states (for example because states are difficult to individuate and condition on in an enforceable way). Despite this, firms should allocate equity not just based on an agent's direct contribution to firm value, but also based on their role in shaping the productivity of their colleagues, and our formula characterizes how this should work.

Finally, we investigate how changes in the production technology affect the principal and the agents under optimal contracts. Interestingly, we find that increasing the complementarity between two agents' efforts can sometimes make one of them worse off by reducing their bonus and expected utility. This somewhat counterintuitive result arises because the strengthening of a complementarity can lead to a redistribution of incentives towards the more central agent involved, leaving the other agent with lower-powered incentives. This finding highlights a potential misalignment of interests between the principal and the agents when the production technology changes, and suggests that agents may sometimes resist seemingly beneficial changes to the team's structure.

Related literature. Broadly, we contribute to the literature on incentive design when production features spillovers across agents. This is related to the literature on moral hazard in teams, going back to the classic work of Holmström (1982). We adapt that model in a way that accommodates a flexible form of uncertainty in what the principal observes about the team's output, generalizing the "imperfect observability" modeling of Holmström (1979) to the multi-agent case. This modeling approach allows us to obtain much richer and more specific predictions about how a principal allocates pay among members of a team.[5]

The general topic of optimally setting incentives in the presence of spillovers has recently attracted interest in the literature on networks. This includes, in addition to the work cited above, papers such as Bloch (2016); Galeotti et al. (2020) and Belhaj and Deroïan (2018). Our main contribution to that literature is a study of a natural and non-parametric formulation, both in terms of the production function and the form of incentives. We show that network game techniques permit some general characterizations of optimal outcomes without the strong parametric assumptions common in the network games literature.

The problem of designing multi-agent contracts has also recently attracted attention in the algorithmic game theory community. Dütting et al. (2023) consider the problem of efficiently computing an optimal *linear* contract in the multi-agent setting for a specific class of output functions. Importantly, the class of output functions they consider restricts the structure of complementarities; in contrast, allowing for essentially arbitrary complement and substitute relationships among agents is a key feature of our results. However, in other respects the focus of this literature overlaps with ours; in particular, we also study the class of equity contracts (restricting pay to be a constant share of principal revenue). The other main contrast between our work and this literature is that the computational literature has focused more on the *extensive margin* question of which agents should be included in a team—i.e., given any incentive to work (Ezra et al., 2023, 2024). This literature shows this problem

[5]We are also related to papers in this literature on partnerships, such as Legros and Matsushima (1991) and Levin and Tadelis (2005) which analyze optimal sharing of project returns to provide incentives, but which ask questions different from ours.

is computationally hard in general but admits acceptable approximations in some cases. We address the important complementary question of optimizing on the intensive margin of exactly how much incentive pay to give agents, with or without linearity restrictions on the contract. Our intensive-margin optimization can be carried out adaptively via "hill climbing," without full information on the production function. Thus, we see our approach as offering a new set of questions and techniques to this emerging literature.

Finally, there is a considerable amount of recent pure and applied theoretical work in economics under the general umbrella of contract design for teams. We give just a few examples: Rayo (2007) consider a relational contract setting where soft information about agents' effort levels is observable and used in relational enforcement. Dai and Toikka (2022) study robust multi-agent contracts and give foundations for a principal's use of linear contracts such as equity. Starmans (2022) is motivated by questions related to ours, examining how moral hazard affects the type of team a principal prefers; the modeling approach there is different, with particular additive specifications of effort, in contrast to the flexible technologies we study. Sugaya and Wolitzky (2023) focus on issues of dynamic enforcement in team projects. Our main contribution is a simple static model of optimal allocation of incentives across agents, with obvious potential to interact with the many questions—especially dynamic ones—that are of interest in this literature.

References

Belhaj, M. and F. Deroïan (2018): "Targeting the key player: An incentive-based approach," *Journal of Mathematical Economics*, 79, 57–64.

Bloch, F. (2016): "Targeting and pricing in social networks," Oxford University Press, 504–542.

Dai, T. and J. Toikka (2022): "Robust incentives for teams," *Econometrica*, 90, 1583–1613.

Dasaratha, K., B. Golub, and A. Shah (2023): "Equity Pay in Networked Teams," New York, NY, USA: Association for Computing Machinery, EC '23, 512.

Dütting, P., T. Ezra, M. Feldman, and T. Kesselheim (2023): "Multi-agent Contracts," in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, New York, NY, USA: Association for Computing Machinery, STOC 2023, 1311–1324.

Ezra, T., M. Feldman, and M. Schlesinger (2023): "On the (In)approximability of Combinatorial Contracts," .

——— (2024): "Sequential Contracts," .

Galeotti, A., B. Golub, and S. Goyal (2020): "Targeting interventions in networks," *Econometrica*, 88, 2445–2471.

Holmström, B. (1979): "Moral Hazard and Observability," *The Bell Journal of Economics*, 10, 74–91.

——— (1982): "Moral hazard in teams," *The Bell Journal of Economics*, 324–340.

Legros, P. and H. Matsushima (1991): "Efficiency in partnerships," *Journal of Economic Theory*, 55, 296–322.

Levin, J. and S. Tadelis (2005): "Profit sharing and the role of professional partnerships," *The Quarterly Journal of Economics*, 120, 131–171.

Mayol, M. C. (2023): "Networks and Moral Hazard," .

Rayo, L. (2007): "Relational Incentives and Moral Hazard in Teams," *The Review of Economic Studies*, 74, 937–963.

Shi, X. (2022): "Optimal compensation scheme in networked organizations," Available at SSRN: https://ssrn.com/abstract=4023019 or http://dx.doi.org/10.2139/ssrn.4023019.

Starmans, J. (2022): "Team Design under Moral Hazard," Available at SSRN https://ssrn.com/abstract=3614519.

Sugaya, T. and A. Wolitzky (2023): "Moral Hazard and Repeated Games with Many Agents," Mimeo., Stanford.