JOINT, SEQUENTIAL PRODUCTION WITH PARTICIPATION AND EFFORT CHOICE*

David Godes†
University of Maryland

Niladri Syam
University of Missouri

June 9, 2023

Abstract

Nearly all organizations of any type, size or scale produce their final output via a process characterized in some form by (i) joint contributions from multiple individual agents and (ii) a sequential work flow in which "upstream" agents take actions that may impact the choices made by those "downstream," at later stages. This presents unique contracting challenges given the interactions among the agents. Consistent with the existing contracting literature, we demonstrate in a linear production process that the power of the optimal linear scheme is increasing along the process. Our main contribution is to show that this is driven by a unique mechanism based on the principal's efforts to induce each agent to internalize their "disutility externality" on downstream agents. While robust to a number of extensions, we show that the increasing power result is not generally robust to different technologies. Via a general model, we are able to demonstrate that the power of schemes may vary from stage to stage arbitrarily as a function of the super- or sub-modularity of the technology with respect to equilibrium effort. In this sense, we are able to explain a broader range of pay patterns than existing theories. Finally, we argue that the schemes we study, even those structured with differential levels of power, promise all identical agents the same expected utility once we account for the differential levels of effort and should thus not be seen as "discriminatory."

JEL Classification Codes: D86, J33, L22, L23
Keywords: Contracts, Principal Agent, Compensation, Incentives, Internal Organization

---

*Declarations of interest: none

†Corresponding author: David Godes dgodes@umd.edu. The authors are grateful to Birger Wernerfelt, Bob Gibbons, Roi Orzach, Dina Mayzlin, Jiwoong Shin, and Eyal Winter, as well as seminar participants at the MIT Organizational Economics Workshop and the Smith School Marketing Quant Lunch for comments on earlier versions of the paper.

1 Introduction

Multi-agent joint, sequential production presents unique contracting challenges due to the dynamic and potentially complex dependencies among the agents. In particular, in such a setting, the expected impact of a given agent's action is a function of other agents' previous actions. In writing a contract with each agent, the principal's challenge is to endow them with the desired objective function, given their specific role in a process, the details of which may not be fully determined at the time of contracting. This has implications both for each agent's participation decision and for their action (e.g., effort) selection.

Our aim in this paper is to characterize the optimal linear contract in this setting, and to assess the relative power of the scheme across agents at different stages in the process. We make several points: (i) Even under risk-neutrality, making agents the residual claimants on the firm output in a joint, sequential setting does not yield first-best outcomes even though it does so in a setting with atomistic agents; (ii) With a linear production technology throughout the organization, the power of the optimal linear scheme increases through the production process from start to finish; (iii) While (ii) is consistent with the literature, we demonstrate that it is driven in our model by a novel and general "disutility externality;" (iv) Though the increasing-power result is robust to a number of extensions, it is not robust to different technologies. In fact, contrary to the literature, the relative power of the optimal linear scheme may vary arbitrarily from stage to stage, driven by the super- or sub-modularity of the technology across stages. Thus, our model can explain a broad range of intra-organizational pay patterns; (v) In spite of the fact that the optimal linear scheme reflects differential *incentives* across identical agents, it may be seen as non-discriminatory in that it results in identical *expected utility*. In particular, since the equilibrium effort varies across agents, once we account for the associated disutility, all agents are paid identically; and (vi) in a context with a probabilistic dichotomous outcome, each of these results holds and the contract is unique.

Our contributions to the existing literature derive from the fact that we consider both continuous effort choices and participation decisions. As described by Winter (2006), a joint sequential production model is appropriate to capture many real-world contracting settings such as modern division-based corporations, R&D and software development. Each of these settings may be characterized by upstream agents selecting from a rich choice set the action that maximizes their own expected utility, while these choices, in turn, also drive the relative attractiveness of the action choices available to downstream agents. Employees in a marketing department, for example, may invest in the creation of email marketing campaigns designed to create leads. Higher effort levels in marketing should result in better leads. That lead quality, in turn, will affect how productive or impactful various levels of sales effort might be on customer decisions. Based on this, salespeople decide how much effort to invest in meeting with, and selling to, prospects. This example highlights several distinct features of our setting and model. First, each agent makes choices, the relative appeal of which will vary from setting to setting. Different products, economic or seasonal cycles and other firm or competitor tactics each could impact the relative returns of different choices. However, contracts are written beforehand, so the firm may not be able to specify at the time of contracting what specific level of effort is desired.

Second, the marketing decisions made upstream impact how much effort the salespeople will contribute, again from a rich choice set. However, the disutility experienced by the salesperson is not directly relevant to the marketer, who balances their own expected compensation against their own disutility. The impact of he marketer on the disutility borne by the salesperson represents an intra-organizational externality that is of course relevant to the principal and which must be addressed by the compensation scheme in order to achieve first-best outcomes. This externality and its relationship to incentives has not been discussed in the literature.

Finally, equilibrium effort levels will differ across agents. Salespeople, for example, may exert more effort than marketing people in equilibrium, or vice versa. This complicates somewhat the analysis of the potentially discriminatory aspect of the compensation scheme.

However, with a binding participation constraint, identical agents working in different stages of the process will still all receive the same net expected utility.

We model an arbitrarily-large set of agents producing, through their sequential, interdependent effort, a single contractible outcome. We focus on a continuous outcome but demonstrate that our results apply equally well to the commonly-studied context with a dichotomous, probabilistic outcome. Agents earlier in the process select effort that drives the marginal impact of effort ("productivity") selected by agents later in the process. We first show, using a simple linear technology, that the optimal linear contract has power that is increasing as we move from earlier to later agents in the process. With a dichotomous outcome, this contract is unique. This finding echoes that of Winter (2006), however we demonstrate that it is driven here by a very different force not present in his model. One might expect that, given risk neutrality, a candidate for the optimal contract would make agents residual claimants on the firm output, as this might ensure that each agent would choose first-best effort. However, such a contract is ineffective in the joint, sequential setting because it does not account for the disutility externality. That is, under residual claimancy, the principal's and agents' objective functions are still misaligned in that the latter do not account for the impact of their effort choice on the disutility faced by those agents downstream from them. As a result, with a linear technology, the principal "meters," or attenuates, the incentive in each contract in proportion to this disutility. Only the final agent in the process is offered a maximum-powered contract because they don't generate any such externality. Note that these forces don't appear in Winter (2006) because (a) on-path effort costs are fixed in the binary action setup, and (b) "high effort" is exogenously given as optimal for all agents.

Our analysis also reveals that, since each agent is bound to their reservation wage, a context with identical agents would imply that each is paid the same expected wage, though via a different mix of salary and commission. Unlike the existing literature, the agents in this setting would not, for example, prefer to be in a different role in equilibrium. In this sense, we argue that the schemes may be seen as non-discriminatory.

We build on this result in a number of directions including allowing for heterogeneity across agents with respect to both productivity and efficiency; uncertainty around these factors; dichotomous outcomes; a setting in which total payout is constrained; and risk aversion. In each of these settings, we find consistent results. Finally, we investigate joint, sequential production with a general technology, while maintaining the focus on a linear contract. Here, we find that the increasing-power result is not general. In fact, the power of the scheme is higher in stage $n+1$ relative to stage n if and only if the technology in stage $n+1$ is *supermodular* in effort. That is, if and only if an increase in stage n effort leads to higher optimal effort in stage $n+1$. This generalizes the core result with respect to the disutility externality in the sense that when the technology is supermodular – which is the case for linear technology – the externality is a positive one, meaning that an increase in the agent's effort level will increase efforts downstream. Thus, if they don't account for the externality, this cost is ignored and effort choice will be supra-optimal, requiring metering. The converse of this argument is that under sub-modular technologies, the agent will choose effort that is too low if they ignore the externality. Thus, the principal will "boost" the power of the scheme to induce them to internalize it. In this way, we see our analysis as explaining a wide range of observed pay patterns throughout and across organizations as a function of the nature of the production technologies.

In the rich literature on contracting with multiple agents, most of the early focus had been on the principal's problem of designing schemes for "teams" comprised of agents who make simultaneous effort choices either in static (Holmstrom, 1982) or dynamic (Che and Yoo, 2001) settings. These studies have considered, for example, the effectiveness of tournaments and/or relative performance evaluation (RPE) when individual outcomes are at least partially observable by the principal. More broadly, these schemes have been studied in multi-agent settings in which there isn't necessarily joint production but correlated shocks. That is, there is in these models a consideration of incentive interdependence but not necessarily production interdependence (Lazear and Rosen, 1981; Nalebuff and Stiglitz, 1983; Green and Stokey, 1983). Closely related to the tournament and RPE literatures is the extensive literature on job design. These studies ask how the principal should optimally assign tasks to agents given considerations associated with preferences, technology and risk aversion. These models typically abstract away from any explicit technological interdependence (Corts, 2007) and/or focus on joint production in a simultaneous setting (Corts, 2007; Holmstrom and Milgrom, 1991; Itoh, 1992). On the contrary, here we take task assignment and job design as exogenous but focus on an organization with explicit and sequential interdependence among the tasks.

Segal (1999, 2003) investigates a broad range of multi-agent contexts with externalities in which agents make participation decisions. A general finding in these settings is that identical agents may earn different returns in equilibrium. Winter (2004) studies a simultaneous collaborative process with multiple agents making binary effort choices. He finds that the optimal scheme must necessarily be discriminatory in the sense that the principal will pay some agents more than others in order to signal that there will be sufficient effort provided to ensure a successful outcome. Halac et al. (2021) extend these ideas by considering a context in which contracts are not public and, thus, one is uncertain as to whether and which others are paid more. They demonstrate that, in this setting, again simultaneous, there is no discrimination and bonuses are identical. While this stream is characterized by models of binary participation decisions associated with a simultaneous production process, an exception is Edmans et al. (2013) who allow for a richer set of effort choices, again in a simultaneous-move setting. Our specification of a sequential process is critical to our findings because, by assumption, agents are able to observe the outcome of previous actions before making their own effort decisions. Returning to our example above, salespeople are able to observe the quality of the leads before deciding how much outreach they should invest in. Without this feature, the disutility externality at the center of our results would not be present.

Most-closely related to the context we study here are Strausz (1999) and Winter (2006), each of which departs from the previous literature by focusing on a sequential technology. The former models a partnership and seeks to find an efficient budget-balanced method for dividing the revenue generated by a sequential process while ensuring each agent puts forth first-best effort. The author presents a novel solution in which the deterministic nature of the technology allows for precise identification of any shirker. In contrast, our focus here is on identifying the principal's profit-maximizing contract in a context with stochastic output. Moreover, we demonstrate that our result is robust to arbitrary constraints on compensation pay out, including budget balance.

Winter (2006) models an exogenously-given joint, sequential production process yielding a dichotomous outcome (success or failure). Each of the agents in his model has a task to complete and makes an observable (but not contractible) binary effort choice. High effort leads to a successful task outcome with higher probability. The overall collaborative process is successful if and only if each task is successful. The cost of high effort is fixed for each agent and not a function of other agents' choices. The principal's objective is to attain high effort choices by all agents at minimum cost. He derives a result that is similar in spirit to our supermodular case, yet as noted above, and discussed in more detail below, is based on a different set of forces, owing to our modeling of both participation and continuous effort.

The rest of our paper is structured as follows. In Section 2, we focus on a linear model of joint, sequential production and demonstrate that the optimal linear scheme has a sharing rule that increases monotonically in power from the beginning to the end. We demonstrate how this result is distinct from the literature in that it is driven by the presence of a disutility externality. We extend the linear model in a number of directions including allowing for rich heterogeneity across the agents as well as uncertainty about the agents at the time of contracting (Section 2.2). We also demonstrate that the core result persists under an arbitrary payout constraint (Section 2.3) and with a dichotomous outcome (see 2.1). In Section 3, we provide a counter-example to the increasing-power result and in Section 4 we present a general model that incorporates and provides insight into both the linear model and the counter-example. In short, under joint, sequential production, the power of the scheme may change arbitrarily from stage to stage with the direction dictated by the supermodularity (increasing power) or sub-modularity (decreasing power) of the technology. We conclude in Section 5 with a summary and suggestions for future work.

2 A Linear Model of Joint Sequential Production

Consider an organization comprised of N risk-neutral agents producing a single continuous output in a sequential process. Each agent is indexed by n with agent 1 representing the beginning, and agent N the end, of the process. Each agent chooses effort level e_n which is associated with a private cost $G(e_n) = \frac{1}{2}e_n^2$. Let $X(\boldsymbol{e}) + \varepsilon$ represent the stochastic output of the joint sequential production technology where $\boldsymbol{e} \equiv (e_1, e_2, ..., e_N)$ and $E[\varepsilon|\boldsymbol{e}] = 0$. We assume that only the Nth agent's effort has a direct impact on the outcome, while the contribution made by each agent $n < N$ is via the creation of an interim output which enters the production function for agent $n + 1$. We refer to this as "productivity" $p_n(e_{n-1}, p_{n-1})$.

In this section, we incorporate a specific, linear specification of the production process:

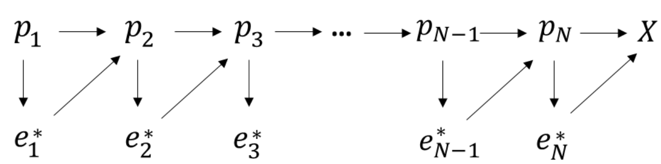
$$p_1 \textit{ given} \tag{1}$$

$$p_n = \sqrt{p_{n-1}} \cdot e_{n-1}, \quad 1 < n \leq N - 1 \tag{2}$$

$$X = \sqrt{p_N} \cdot e_N \tag{3}$$

We assume here that p_1 is exogenous, fixed and known to all parties. Higher values of p_1 imply a production process in which the principal wants the agents to exert higher levels of effort. We present below in Proposition 4 a version of the model in which this is instead a random variable at the contracting stage, but known before effort is chosen. Note that this specification is supermodular in effort in the sense that higher effort at stage n implies higher effort at stage $n+1$ *ceteris paribus*. We investigate different specifications in Sections 3 and 4. See Figure 1 for a graphical representation of the process.

Figure 1: Joint, Sequential Technology



We model a joint, sequential technology comprised of N agents jointly producing output X. Higher effort e_n put forth by each agent $n < N$ increases the productivity p_{n+1} of the next agent in the production process: $\frac{\partial p_{n+1}}{\partial e_n} > 0$.

We focus on linear contracts with salary S_n and commission $C_n \in [0, 1]$. Each agent's total payoff (and, thus, utility, given our assumption of risk neutrality) W_n is then:

$$W_n = S_n + C_n \cdot \left[X(\boldsymbol{e}) - \frac{1}{2}e_n^2 + \varepsilon \right] \quad (4)$$

which each agent maximizes sequentially with respect to e_n, conditional on observing p_n. Owing to the sequential nature of the process, defined in Equations (1) - (3), we can rewrite (4) to reflect the fact that each agent differs in terms of her impact on the expected outcome. Next, we define the stage-n output function, $X_n(\boldsymbol{e^n}(e_n)|p_n)$, as agent n's expectation of the output X, conditional on their own observed productivity and as a function of their effort choice. Here, $\boldsymbol{e^n} \equiv \{e_{n+1}, ..., e_N\}$, that is the vector of all effort levels downstream from agent n. Thus, (4) becomes, in expectation:

$$E[W_n] = S_n + C_n \cdot X_n(\boldsymbol{e^n}(e_n)|p_n) - \frac{1}{2}e_n^2 \quad (5)$$

which each agent maximizes at effort level e_n^{**}. We use e_n^{**} to capture agents' utility-maximizing effort choices, and e_n^* for first-best effort levels.

In this model setup, the principal implements the first-best outcome by selecting the compensation parameters that solve the following problem:

$$Max_{\{C_n,S_n\}} \;\; X(e_1, e_2, ..., e_N) - \frac{1}{2}\sum_{n=1}^{N} e_n^2 \tag{6}$$

$$s.t. \;\; E[W_n] \geq \underline{W}_n \tag{7}$$

$$e_n = e_n^{**} \tag{8}$$

The $\underline{W}_n$ terms capture each agent's reservation wage. Before analyzing the scheme, we solve for the the first-best effort levels by maximizing the following unconstrained problem:

$$Max_{e_n} \;\; X(e_1, e_2, ..., e_N) - \frac{1}{2}\sum_{n=1}^{N} e_n^2 \tag{9}$$

In the following lemma, we characterize the first-best effort e_n^* and associated productivity levels, as well as stage-n output function, $X_n(\boldsymbol{e^n}(e_n)|p_n)$. We will, where possible, suppress the arguments to X_n for parsimony.

Lemma 1. *The first-best productivity, effort levels, and stage-n output functions, associated with the problem in (9), are given by:*

1. $p_n^* \cdot \psi_{n-1} = p_{n-1}^*$
2. $e_n^* = \frac{\sqrt{p_n}}{\psi_n}$
3. $X_n = \dfrac{\sqrt{p_n}}{\prod_{m=n+1}^{N} \psi_m} e_n$

where $\psi_n = 2^{k(n)}, \; k(N) = 0, \; k(n-1) = 2 \cdot k(n) + 1$

The lemma makes clear that in this linear production technology, first-best productivities decline along the process, from beginning to end. We are also able to use Lemma 1 to derive implications for the dynamics of optimal effort:

$$\frac{e_n^*}{e_{n-1}^*} = \frac{\psi_{n-1}}{\psi_n} \cdot \sqrt{\frac{p_n}{p_{n-1}}} > 1$$

Though productivity declines, first-best effort levels *increase* monotonically along the sequence.

The stages of the game are as follows: (i) the principal makes a take-it-or-leave-it contract offer of $\{C_n, S_n\}$ to each agent n. (ii) Conditional on each agent accepting the offer, the production process unfolds as each agent first observes their own productivity and then selects their effort which drives the next agent's productivity. (iii) The final output is determined by agent N's effort choice, and (iv) agents are paid and firm profits are realized. We find the unique set of contract parameters that yield a subgame-perfect equilibrium.

The following lemma presents useful results on agents' optimal response to arbitrary linear contracts in our current specification. Our focus is on the slope parameters $\{C_n\}$ as the "salaries" will be chosen to bind the agents to their reservation wages.

Lemma 2. *Consider the problem defined in (1)-(3). For a given compensation scheme* $\big\{S_n, C_n\big\}$ *the unique subgame-perfect set of efforts levels, total sales and costs of effort are given by:*

1. $$e_n^{**}(p_n) = C_{n+1} C_n \sqrt{p_n} \prod_{j=2}^{N-n} C_{n+j}^{2^{j-1}} \tag{10}$$
2. $$X = \prod_{n=1}^{N} C_n^{2^{n-1}} \tag{11}$$
3. $$\frac{1}{2} e_n^{**2} = \frac{C_n}{2} \prod_{j=1}^{N} C_j^{2^{j-1}} \tag{12}$$

Equation (10) highlights the forces driving each agent's effort choices. In addition to their own incentive C_n, they also consider both their productivity p_n and the incentives of all agents downstream from them. This is a direct implication of the interdependence of the production process. Using the results from Lemma 2, we can rewrite Equation (6) as an unconstrained optimization problem in which the firm chooses $\big\{C_n\big\}$ to maximize the following objective:

$$E[\Pi] = X - \frac{1}{2}\sum_{n=1}^{N} e_n^2 - \sum_{n=1}^{N} \underline{W}_n = \prod_{n=1}^{N} C_n^{2^{n-1}} - \sum_{n=1}^{N} \frac{C_n}{2} \prod_{j=1}^{N} C_j^{2^{j-1}} - \sum_{n=1}^{N} \underline{W}_n \tag{13}$$

We are now able to solve in closed form for the unique linear scheme implementing the first-best outcome, which in our setting is surprisingly simple:

Proposition 1. *The subgame-perfect linear contract that maximizes (13) is* $C_n^* = 2^{n-N}\ \forall n$

In a four-stage sequential technology, for example, the commission rates would be given by (from upstream to downstream): $\frac{1}{8}$, $\frac{1}{4}$, $\frac{1}{2}$, 1. This provides a parsimonious explanation, for example, for the commonly-observed phenomenon that, in the same organization, employees in marketing are typically paid via lower-powered schemes than those in sales. Along the same lines, it is not uncommon for functional managers (upstream) to be paid according to lowered-powered schemes relative to their direct reports (downstream). The following is a direct corollary of Proposition 1:

Corollary 1.1. *There does not exist a linear compensation scheme satisfying* $\sum_n C_n \leq 1$ *that implements the first-best outcome in this model*

While this is an unattractive property, we show later that the qualitative features of the result in particular, the increasing power – persist even when we constrain the solution to $\sum_n C_n^* \leq 1$.

We highlight several aspects of Proposition 1. First, we note that this is inconsistent a heuristic that suggests that a principal should always make risk-neutral agents residual claimants on revenue – i.e., that the principal should "sell the firm" to risk-neutral agents. Typically, in a risk-neutral setting, the assignment of the firm's revenue to the agent has been shown to endow them with the firm's objective function. However, this is not the case with joint, sequential production. Specifically, even if we were to assign all revenues to any agent $n < N$, they would not choose e_n^*. The reason for this relates to the fact that an upstream agent's effort choice imposes two externalities on downstream agents, one related to revenue, and thus monetary compensation, and the other related to disutility. The former is aligned via residual claimancy, but the latter is not. To see this in a simple example, note that first-best effort at the final stage of the process is, from Lemma 1, $e_N^* = \sqrt{p_N}$. From Lemma 2, this agent's effort choice is $e_N^{**} = C_N\sqrt{p_N}$, so indeed the principal would optimally sell them the firm. However, consider now agent $N-1$. Conditional on the agent's productivity level p_{N-1}, the firm's problem is:

$$Max_{e_{N-1}} \ \sqrt{\sqrt{p_{N-1}}e_{N-1}} \cdot e_N(p_N) - \frac{1}{2}e_N^2(p_N) - \frac{1}{2}e_{N-1}^2 \tag{14}$$

Substituting in the equilibrium choice by agent N yields:

$$Max_{e_{N-1}} \ \frac{1}{2}\sqrt{p_{N-1}} \cdot e_{N-1} - \frac{1}{2}e_{N-1}^2 \tag{15}$$

which implies the first-best effort level of $e_{N-1}^* = \frac{1}{2}\sqrt{p_{N-1}}$. However, if the firm sets $C_{N-1} = 1$, the agent chooses, again from Lemma 2, the supra-optimal level of $e_{N-1}^{**} = \sqrt{p_{N-1}}$. The agent works "too hard" as do all subsequent agents as a result. To appreciate why this the case, note that in selecting their effort, agents trade off two factors: (i) the marginal impact of their effort on the expected revenue (and, ultimately, their expected payout), and (ii) the marginal cost of *their* effort. Since it does not affect their utility, agent $N-1$ in this example does not consider the impact their effort choice has on the effort costs experienced by the downstream agent N. Moreover, endowing them with a claim on the firm's revenue does not accomplish this either. Selecting $C_{N-1} = 1$ is not tantamount to endowing the agent with the firm's objective function because it doesn't ensure they internalize this "disutility externality."

To illustrate this property more generally, we compare here the principal's optimal effort choice with an arbitrary agent's choice. The principal's solution must satisfy a system of N

necessary first-order conditions of the form:

$$\left[\frac{dX}{dp_N}\frac{dp_N}{dp_{N-1}}\cdots\frac{dp_{n+2}}{dp_{n+1}}\frac{\partial p_{n+1}}{\partial e_n^*}\right]-\left[e_n^*+\sum_{m>n}e_m\frac{\partial e_m}{\partial p_m}\frac{dp_m}{dp_{m-1}}\cdots\frac{dp_{n+2}}{dp_{n+1}}\frac{\partial p_{n+1}}{\partial e_n^*}\right]=0 \qquad (16)$$

The first bracketed term represents the marginal impact of agent n's effort choice on the outcome X, propagated down through the entire organization via the effort levels of all downstream agents. Note that we indicate the full derivatives of both X and the productivities p_n, reflecting the fact that changes in productivity at stage n have both direct and indirect effects on productivity at stage $n+1$. The second bracketed term reflects the marginal impact of agent n's decision on their cost of effort as well as that of those agents later in the process. Our choice of functional form for the technology makes this straightforward to solve in closed form. We return below to a characterization of a more-general solution. Agent n's utility-maximizing effort choice, on the other hand, satisfies the following necessary first-order condition.

$$C_n\cdot\left[\frac{dX}{dp_N}\frac{dp_N}{dp_{N-1}}\cdots\frac{dp_{n+2}}{dp_{n+1}}\frac{\partial p_{n+1}}{\partial e_n^{**}}\right]-e_n^{**}=0 \qquad (17)$$

It is clear in comparing (16) and (17) at $C_n=1$, that they're aligned on revenues, which is what one would expect from the literature on atomistic, or separable, production. However, there is clear misalignment on costs due to the disutility externality. Moreover, as n decreases – i.e., as one goes farther back in the process – this misalignment becomes more and more pronounced because this implies that there are more agents on whom one is imposing the disutility externality.

Note that Winter (2006) presents a set of results that appear quite similar. However, there are a number of important distinctions. Most critically, our results are driven by very different forces. His result is driven by a declining intrinsic cost to shirking. Specifically, shirking by agents causes, in equilibrium, all later agents to shirk. This of course is less harmful to the probability of a successful outcome (and, thus, more tempting) the later an agent sits in the process. To offset this, the principal increases the reward for a successful outcome paid to agents closer to the end, in order to ensure they don't shirk. From (16) and (17), it is clear that, at maximum power, this revenue- (or outcome-) based misalignment is not present here.[1]

Another point of distinction, to which we return in detail in Section 4, our setup accommodates a range of technology specifications that can give rise to arbitrary patterns of contracts, including those that may decrease in power from one agent to the next. The core driving factor in all will be the nature of the disutility externality. In this linear case, the externality leads to supra-optimal effort causing the principal to meter, via dampened incentives, the effort levels. However, we will both give an example of, and more-fully characterize, a context in which the externality may lead to sub-optimal effort, necessitating boosted incentives earlier in the process. Again, owing to the different drivers, such a solution is not shown by Winter (2006).

Finally, it is worth comparing the optimal linear contract here with that in Winter (2004) and Halac et al. (2021), who investigate the optimality of "discrminatory" contracts, defined as those that pay *ex ante* identical agents different levels of compensation. On one hand, this is indeed the case here as well, with agents earlier in the process being paid via lower-powered schemes. However, given our consideration of both effort and participation, by construction, each agent in our model is bound to their reservation wage. Put differently, while they receive different compensation structures, they experience exactly the same expected utility. In this sense, then, we would argue that the contracts are in fact not discriminatory. Put simply, if given the option, no agent would prefer to move from one position to another.

2.1 Unique Contract with Dichotomous Outcome

Our core model specifies a continuous outcome as a function of the continuous, sequential effort of N agents. An alternative approach, common in the literature (see, for example,

[1]Moreover, making any agent the residual claimant in Winter (2006) would ensure that they choose the principal's desired effort level. As outlined above, this is not the case here due to the disutility externality.

Baker et al. (1994)), would be to conceptualize the process as producing with probability $\sqrt{p_N}e_N$ a single observable output of value normalized to 1. An appeal of this extension, besides a broader set of applications, is that there is, by construction, a unique optimal sharing scheme.

To see how naturally such a specification fits here, we can maintain (1) and (2) exactly. In place of (3), we specify a dichotomous outcome (e.g., success or failure):

$$X \in \{0, 1\} \tag{18}$$

$$Pr(X = 1) = \sqrt{p_N} \cdot e_N \tag{19}$$

Then, all that remains is to constrain $\sqrt{p_N} \cdot e_N \in [0, 1]$. Per the discussion following Lemma 1, a sufficient condition for this would be $p_1 \leq 1$, though this could be relaxed significantly. We thus state the following:[2]

Proposition 2. *Consider a joint, sequential process with* N *agents producing a single output of value* 1 *with probability* $\sqrt{p_N} \cdot e_N$*, with agents* $1, ..., N-1$ *acting according to (1) and (2). Then, with* $p_1 \leq 1$*, the unique subgame perfect compensation scheme is characterized by* $C_n^* = 2^{n-N}$*.*

Note that the driver of this result is, again, the principal's effort to encourage the agent to internalize the disutility externality.

2.2 Heterogeneity in the Linear Model

Our core model is parsimonious in that it lacks heterogeneity across either agents or tasks. In this section, we consider a richer a version of (2):

$$p_{n+1} = b_n \cdot \sqrt{p_n} \cdot e_n \tag{20}$$

[2]The authors are sincerely grateful to Bob Gibbons and Roi Orzach for suggesting this approach.

Here, the b_n parameters reflect the fact that some functions in an organization have more impact on the production process than others. Alternatively, some agents are inherently more productive than others. Moreover, rather than a homogeneous disutility of effort $G(e_n) = \frac{1}{2}e_n^2$, we allow heterogeneity here as well:

$$G_n(e_n) = \frac{\gamma_n}{2}e_n^2 \tag{21}$$

For example, some work is inherently more difficult or perhaps generates more disutility than other work. Indeed, many management decisions focus on the magnitude, and the allocation, of investments associated with the productivity of agents – as captured in the b_n parameters – as well as their efficiency, as captured in the γ_n's.

As the following proposition demonstrates, this richer specification nonetheless yields an optimal linear contract that is identical to that in the simpler set up.

Proposition 3. *Incorporating into (9) the agent-level heterogeneity as described in (20) and (21) yields the same optimal linear compensation plan as presented in Proposition 1.*

As shown in the proof of Proposition 3, the fully-parameterized version of the profit function in (13) is given by:

$$E[\Pi] = \Big(\prod_{j=1}^{N} \frac{b_j^{2^j}}{\gamma_j^{2^{j-1}}}\Big)\Big(\prod_{n=1}^{N} C_n^{2^{n-1}} - \sum_{n=1}^{N} \frac{C_n}{2}\prod_{j=1}^{N} C_j^{2^{j-1}}\Big) - \sum_{n=1}^{N} \underline{W}_n \tag{22}$$

It is immediately apparent that these parameters serve only to shift up or down the firm's revenue, as well as the equilibrium effort levels.[3] To develop some intuition around this result, consider a simple ($N = 2$) joint, sequential process. Let $b_1 = \gamma_1 = 1$, and allow b_2 and γ_2 to be free. It is easily verified that $e_2^* = \frac{b_2}{\gamma_2} \cdot \sqrt{p_2}$ and $C_2^* = 1$. Given this, the principal's problem is:

$$Max_{C_1} \frac{b_2^2}{\gamma_2}p_2(e_1) - \frac{1}{2}\frac{b_2^2}{\gamma_2}p_2(e_1) - \frac{1}{2}e_1^2 \tag{23}$$

[3] Note that this is also true in standard models of non-joint, sequential production under risk neutrality.

while the agent's problem is:

$$Max_{e_1}\ C_1 \frac{b_2^2}{\gamma_2} b_2 p_2(e_1) - \frac{1}{2}e_1^2 \tag{24}$$

So, while the parameters b_2 and γ_2 are incorporated into agent 1's problem, they only account for "half" of the effect. In particular, they only account for the revenue effect, but not the disutility. Thus, the pay scheme C_1 must compensate for this, which it does by metering the incentive in the agent's utility function by a factor of one half. Moreover, this is true regardless of the values of b_2 and γ_2. As we add more and more stages, the metering must increase, reflecting more misalignment between the agent and the principal. However, in every case, this misalignment – which the compensation scheme is selected to rectify – is a fixed proportion, independent of the heterogeneity parameters.

Given the invariance of the optimal scheme to different values of the $\{b_n, \gamma_n\}$ parameters, the next proposition is a natural extension in which we allow them to be random variables at the contracting stage:

Proposition 4. *Assume the $\{b_n, \gamma_n\}$ parameters are distributed according to a $2N$-dimensional joint distribution with support $\mathbb{R}^{2N+}$. If an expected value exists for*

$$\prod_{j=1}^{N} \frac{\tilde{b}_j^{2^j}}{\tilde{\gamma}_j^{2^{j-1}}} \tag{25}$$

then, a model in which the values of $\{b_n, \gamma_n\}$ are realized after contracting, but before effort selection, yields the same optimal linear scheme as in Proposition 1.

So, our result is robust not only to the presence of heterogeneity in the desired effort levels, but also to *ex ante* uncertainty.

Given that neither the equilibrium compensation scheme nor reservation wages (the latter by assumption) are impacted by these parameters, our result in Proposition 3 makes straightforward the investigation of a range of organizational design decisions. As an example, we look here at the allocation of productivity (b_n) and efficiency (γ_n) improvements across an organization.

It is obvious from (22) that the firm's profit improves as the productivity and efficiency of its agents improve (i.e., as b_n increases and γ_n decreases, for any n). However, it is less obvious how an organization should distribute its investments in productivity or efficiency improvements. Here, we abstract away from the decision as to *how much* to invest and focus instead on *where* to invest.[4]

Consider the impact on the principal's return of increasing the productivity of a single agent n by a small amount:

$$\frac{\partial E[\Pi]}{\partial b_n} = 2^n \cdot b_n^{-1} \cdot K \equiv d_n \tag{26}$$

where $K \equiv E[\Pi] + \sum_n \underline{W}_n > 0$. Now, for any two agents m and n, we have:

$$d_m - d_n = K \cdot \left[\frac{2^m}{b_m} - \frac{2^n}{b_n} \right] \tag{27}$$

An analogous analysis can be performed for efficiency γ_n. Several interesting insights can be gleaned from this. First, in a hypothetical context with *ex ante* homogeneous agents such that $b_n = b \ \ \forall n$, productivity-increasing investments should always be made downstream (in the productivity of agents $m > n$) before they're made upstream. At first, this may seem surprising given the multiplicative effect of effort provision through the process. However, it relates to a point made above regarding the monotonic increase in first-best effort levels across the process, driven again by the firm's desire to attenuate upstream effort relative to downstream. As a result, a given investment in productivity improvement yields lower gains upstream, relative to downstream. Of course, in a standard atomistic, additive, separable model with homogeneous agents, there would be no difference in the impact of productivity investment across different stages.

More generally, allowing for heterogeneity across stages, the allocation decision involves

[4] Moreover, let h_t represent funds budgeted for productivity improvements in period t. For the sake of parsimony, we assume here that the marginal response of productivity $\frac{\partial b_n}{\partial h_t}$ is constant across n.

a trade off between this preference for downstream investments, on one hand, and current productivity levels, on the other. To the latter point, it is clear from (26) that such investments yield concave gains, limiting the optimal disparity in downstream vs. upstream investments. Moreover, given the joint nature of production in our setting, an improvement in an agent's productivity will have a bigger impact on profits the relatively more-productive the *other* agents are. To complete this analysis, we can state formally the following result on the long-run allocation of productivity in an organization. Note that for this result, we assume the random arrival of funds for productivity improvement. Since we don't model the activities throughout the entire organization, this is meant to capture the idea that productivity improvements (e.g., training) may in a given period be expected to yield higher or lower marginal returns relative to other investment opportunities.

Proposition 5. *In the unique stable long-run allocation of productivity (efficiency) investments across the organization, for any pair of agents m and n: $b_m = 2^{m-n} \cdot b_n \ \ \forall m, n$ ($\gamma_m = 2^{m-n} \cdot \gamma_n \ \ \forall m, n$).*

While we're aware of no rigorous empirical analysis of relative intra-organizational investment into productivity or efficiency, this result is consistent with casual observation. For example, the average spending on training for marketing employees (upstream) is generally significantly below that spent on (downstream) sales training.[5]

2.3 Payout Constraint

As noted in Corollary 1.1, a clear implication of Proposition 1 is that the first-best plan implies that the principal pays out more in commission than the value of their output. More precisely, it pays out a proportion of revenue equal to $2 \cdot (1 - 2^{-N}) \in (1, 2)$ for $N > 1$. We consider here a version of our model in which we impose an additional constraint such that

[5] See, for example, cxl.com/blog/state-of-marketing-training-2019/ which reports annual marketing training to be around $1,000 per employee, and www.ama.org/listings/2020/04/26/sales-training-2/ which reports sales training averages over $2,200 per employee.

the principal maximizes expected return as specified in Equation (13) subject to:

$$\sum_{n=1}^{N} \overline{C}_n \leq \overline{C}, \quad \overline{C} \in (0,1] \tag{28}$$

where $\overline{C}_n$ represents the constrained-optimal agent-level commission rates and $\overline{C}$ is the overall payout cap. As the following result demonstrates, while of course the overall level of payout is lower, the constrained model yields precisely the same relationship among adjacent sharing rules as we obtain above in the unconstrained model.

Proposition 6. *In the constrained model, in which we impose (28) on the core model, we find the following unique optimal linear sharing rule:*

(i) $\overline{C}^*_{n+1} = 2\overline{C}^*_n, \ \forall\, n$

(ii) $\overline{C}^*_n = \dfrac{2^{n-1}}{2^N - 1}\overline{C}, \ \forall\, n$

Part (i) of the result shows that again we see a doubling of the power of the scheme from n to $n+1$. It is straightforward to verify that, as one would expect:

$$\overline{C}^*_n = C^*_n \cdot \left[\frac{\overline{C}}{2 \cdot (1 - 2^{-N})}\right]$$

In other words, the constrained-optimal sharing rates are equal to the unconstrained values, each scaled downward by a fixed proportion. Moreover, the scaling factor is intuitive in that it reflects the relative magnitudes of the constrained and the unconstrained payouts. By referring to part (i) of Lemma 2, it can be verified that the effects on *effort levels* of imposing a payout constraint are more severe with respect to upstream agents than downstream agents. As a result, such a constraint leads to more variation in effort levels across stages. This is a direct result of the multiplicative effects of incentives. As a final note on this constrained model, it is straightforward to demonstrate that the core results in Proposition 5 carry over exactly to this constrained setting.

2.4 Risk Aversion in a Linear Joint Sequential Process

To this point, for the sake of parsimony, we have maintained an assumption of risk neutrality, which has been the focus of much of the contracting literature. In order to capture this in our modeling framework, we adopt the CARA framework, including an exponential utility function such that the utility for monetary payoff W is given by:

$$U(W) = -e^{-RW} \tag{29}$$

where R captures the agent's constant absolute risk aversion. As is well known (Pratt, 1964; Holmstrom and Milgrom, 1991), assuming that the error term in the final stage-N output is normally distributed with zero mean and variance σ^2 gives rise to the tractable transformation of expected utility into a certainty equivalent CE_n:

$$CE_n = S_n + C_n \cdot X - \frac{1}{2}e_n^2 - \frac{1}{2}R_n C_n^2 \sigma^2 \tag{30}$$

In turn, incorporating the "risk premium" into the firm's objective function yields:

$$E[\Pi] = X - \frac{1}{2}\sum_{n=1}^{N} e_n^2 - \frac{1}{2}\sigma^2 \sum_{n=1}^{N} R_n C_n^2 - \sum_{n=1}^{N} \underline{W}_n \tag{31}$$

Since, conditional on the compensation parameters, agent n's maximization of (30) yields identical effort to her solution to (4), the results in Lemma 2 define optimal effort choices e_n^{**} in the risk-averse context as well. What remains then is to characterize the optimal sharing scheme, which will *not* be identical to the risk-neutral setting, since the firm will no longer implement first-best outcomes. Substituting the results from Equations (1)-(3), we can rewrite the firm's objective as a function of the compensation parameters:

$$E[\Pi] = X - \frac{X}{2}\sum_{n=1}^{N} C_n - \frac{\sigma^2}{2} \sum_{n=1}^{N} R_n C_n^2 - \sum_{n=1}^{N} \underline{W}_n \tag{32}$$

The firm's first-order conditions for (32) with respect to N commission rates $\{C_n\}$ are:

$$\frac{\partial X}{\partial C_n^*}\Big(1-\sum_{n=1}^{N}\frac{C_n^*}{2}\Big)-\frac{X}{2}-R_n\sigma^2 C_n^*=0 \tag{33}$$

From (11), one can substitute in for:

$$\frac{\partial X}{\partial C_n}=\frac{2^{n-1}}{C_n}\cdot X \tag{34}$$

and thus rewrite the necessary first-order conditions for an optimal set of commission rates:

$$2^{n-1}X\Big(1-\sum_{n=1}^{N}\frac{C_n^*}{2}\Big)=\frac{XC_n^*}{2}+R_n\sigma^2 C_n^{*2} \tag{35}$$

Note that $X\Big(1-\sum_{n=1}^{N}\frac{C_n^*}{2}\Big)$ is constant across n.

Starting with homogeneous risk aversion in which $R_n \equiv R\ \forall n$, one can see that, as in the risk-neutral case, the power of the compensation scheme will increase monotonically across the risk-averse production process, from beginning to end. Having said this, the introduction of risk aversion mitigates the slope of incentives along the process. That is, the increase from C_n^* to C_{n+1}^* is attenuated, relative to the risk-neutral case, due to risk aversion. This can be seen by re-arranging (35) to form the ratio $\frac{C_n^*}{C_{n+1}^*}$:

$$\frac{C_n^*}{C_{n+1}^*}=\frac{\frac{X}{2}+R\sigma^2 C_{n+1}^*}{X+2R\sigma^2 C_n^*} \tag{36}$$

Then, assuming that the claim is false, that is

$$\frac{\frac{X}{2}+R\sigma^2 C_{n+1}^*}{X+2R\sigma^2 C_n^*}<\frac{1}{2}$$

yields an immediate contradiction. Intuitively, since optimal schemes under risk aversion are lower powered relative to risk neutrality, the extent to which agents will overproduce effort if they ignore the externality is already dampened. Thus, there is less of a need for the firm to further attenuate the incentive.

These ideas are readily extended to cases of heterogeneous risk aversion. On one hand, it can be verified from Equation (35) that, if risk aversion is declining across the sequential process (for example, if salespeople are less risk averse than marketing, who are in turn less risk averse than human resources), then again it will always be the case that the power of incentives is monotonically increasing. Intuitively, the heterogeneity in risk-aversion reinforces in such a setting the decreasing need to attenuate effort levels. However, this is not necessarily the case if risk aversion is increasing, or perhaps non-monotonic. We summarize this discussion with the following Proposition:

Proposition 7. *Assuming agents of level n are endowed with risk aversion R_n, then the following claims hold:*

1. *Under homogeneous risk aversion, $C_n^* < C_{n+1}^*$ $\forall n$*
2. *Under homogeneous risk aversion, it is always the case that $\frac{C_{n+1}^*}{C_n^*}$ is declining in R*
3. *If risk aversion is declining in n, then $C_n^* < C_{n+1}^*$ $\forall n$ though this is not necessarily true for increasing risk aversion.*

Similar to the discussion in Section 2.2, we can also apply our results here to the question of organizational design. Specifically, we can ask where in the organization should one optimally "locate" risk aversion? For example, in an organization with N agents differing only on the dimension of risk aversion, how should they be assigned across the production process? It is immediately clear by application of the envelope theorem that an increase in risk aversion has a smaller effect upstream than it does downstream. From Equation (32):

$$\frac{d\Pi(C_n^*)}{dR_n} = \frac{\partial\Pi(C_n^*)}{\partial R_n} = -\frac{1}{2}\sigma^2 C_n^{*2}$$

Thus, by Proposition 7, in contexts with either homogeneous or declining risk aversion, it is optimal to assign more risk-averse agents upstream in the process, rather than downstream.

We note that this provides a novel explanation for the lay belief that salespeople are generally more risk tolerant than are those in other, upstream roles like marketing or HR. Typically, this is explained by taking as given the compensation scheme and arguing that those that are less risk averse are more likely to be drawn to jobs characterized by higher-powered incentive schemes like commissioned sales. Our result here doesn't account for any choices made by agents as to where to work (indeed, all agents earn in expectation their reservation wage regardless of pay plan structure). Instead, it is driven by the fact that risk premia are proportional to the power of the pay plan. Thus, the organizational designer places the risk averse agent upstream where the pay plan is, in equilibrium, lower powered due to the sequential nature of production. Put differently, this is a supply-side explanation for this phenomenon, as compared with the typical demand-side explanation.

3 A Counter-Example

So far, we have demonstrated that in a context with a linear, joint sequential production process each agent is paid via a higher-powered scheme than the previous agent. Further, we have explained this result as a being driven by a disutility externality which the principal induces agents to internalize via the optimal linear scheme. If ignored, the externality leads to supra-optimal effort choices and sub-optimal profit as agents demand higher salaries to compensate for the disutility. In order to induce the upstream agent to account for their externality, the principal meters, or attenuates, the schemes. This metering is more pronounced the earlier in the process the agent appears.

While we have shown the result to be robust to heterogeneity, homogeneous or declining risk aversion and an alternative discrete probabilistic outcome, we demonstrate by example that it is not generally robust to the nature of the production technology. Specifically, we provide an example of an organization with a stage characterized by a submodular technology, defined as one in which higher effort in stage n – which leads to higher productivity in stage

$n+1$ – leads the stage $n+1$ agents to exert *less* effort. While this may appear somewhat counterintuitive on the surface, it isn't very difficult to imagine contexts in which more-productive agents work less, rather than more. Imagine, for example, a calculus instructor (the upstream agent) who trains the math student (the downstream agent) for a math olympiad. It seems reasonable to imagine that the more effort the former exerts in the training process, the less difficult the task faced by the latter in solving the problems.

To appreciate the contracting problem in such an environment, consider a three-stage set-up as follows:

$$\begin{aligned} X &= \sqrt{p_3}e_3 \\ p_3 &= 1 - \frac{1}{e_2 p_2} \\ p_2 &= \sqrt{p_1}e_1 \\ p_1 &= 1 \\ C(e_n) &= \frac{1}{2}e^2 \ \ \forall n \end{aligned}$$

The only difference between this process and that in Section 2 is that, here, the agent in stage 2 drives the stage 3 agent's productivity via a submodular technology. We know that $C_3^* = 1$, $e_3^* = \sqrt{p_3}$ and thus $X = p_3$. First-best second-stage effort level maximizes the following:

$$Max_{e_2} \ \ X(e_2) - \frac{1}{2}e_2^2 - \frac{1}{2}e_3(e_2)^2$$

Substituting in the known quantities allows us to transform the problem:

$$Max_{e_2} \ \ \frac{1}{2}p_3(e_2) - \frac{1}{2}e_2^2$$

Agent 2 solves the following problem:

$$Max_{e_2} \ \ C_2 p_3(e_2) - \frac{1}{2}e_2^2$$

One can readily ascertain that $C_2^* = \frac{1}{2}$, $e_2^* = (2p_2)^{-\frac{1}{3}}$ and, thus, $p_3 = 1 - \sqrt[3]{2}p_2^{-\frac{2}{3}}$. Note that effort here declines in productivity. Finally, moving to stage 1, where $p_1 = 1$ and, thus, $p_2 = e_1$, first best effort levels are found by solving:

$$Max_{e_1} \;\; p_3 - \frac{1}{2}e_1^2 - \frac{1}{2}e_2(e_1)^2 - \frac{1}{2}e_3(e_1)^2$$

By substituting quantities already calculated we can transform this into:

$$Max_{e_1} \;\; \frac{1}{2}\Big(1 - \sqrt[3]{2}p_2^{-\frac{2}{3}} - 2^{-\frac{2}{3}}p_2^{-\frac{2}{3}}\Big) - \frac{1}{2}e_1^2$$

The necessary first-order condition for the first-best stage-1 effort choice is given by:

$$\frac{1}{2}\Big(\frac{2}{3}\sqrt[3]{2}e_1^{*-\frac{5}{3}} + \frac{2}{3}2^{-\frac{2}{3}}e_1^{*-\frac{5}{3}}\Big) = e_1^* \tag{37}$$

The agent chooses her effort level that maximizes:

$$Max_{e_1} \;\; C_1\Big(1 - \sqrt[3]{2}e_1^{-\frac{2}{3}}\Big) - \frac{1}{2}e_1^2$$

The associated necessary first-order condition[6] for this problem is:

$$C_1\Big(\frac{2}{3}\sqrt[3]{2}e_1^{**-\frac{5}{3}}\Big) = e_1^{**} \tag{38}$$

It can be readily determined that the value of C_1^* that ensures that the agent's effort choice e_1^{**} is that which the firm prefers (e_1^*) is: $C_1^* = \frac{3}{4}$. Thus, owing to the submodularity of the stage-2 technology, we find that $C_1^* > C_2^* < C_3^*$.

This simple illustration makes two points. First, we reinforce the idea that the nature of the production technology is a critical determinant of the relative power of schemes in an interdependent organization. We emphasize that this is not the case in traditional models of atomistic agents producing joint outputs via additive technologies. Given the first point, this analysis also implies that simple linear (and thus supermodular) models of joint

[6] It is straightforward to demonstrate that both the firm's and the agent's problems yield unique solutions.

production, while tractable, are *not* without loss of generality, at least with respect to optimal compensation schemes. Specifically, in the presence of submodular technologies, we would expect that the power of the scheme may instead *decrease* along part of the process. Again, this is driven by the principal's desire to induce the agent to internalize their disutility externality. The only difference here is that the externality leads to sub-optimal effort, requiring a "boosting" of the power of the scheme, rather than metering. In the final section of the paper, we solve a version of the model with a more-general technology, allowing us to more-fully characterize the optimal linear scheme across both super- and sub-modular settings.

4 A General Model of Joint Sequential Production

To specify a general model of joint sequential production which can accommodate either sub- or super-modular technologies, we again start with an organization with a production process comprised of N risk-neutral agents. Each agent chooses effort level e_n which imposes on her a private, non-monetary, monotonically-increasing, twice-differentiable and convex cost $G_n(e_n)$. Let $X(\boldsymbol{e})+\varepsilon$ represent the output of the joint sequential production technology where $\boldsymbol{e} \equiv (e_1, e_2, ..., e_N)$ and $E[\varepsilon|\boldsymbol{e}] = 0$. The main distinction between this set-up and that in previous sections is that we leave the relationship among effort $e_n(p_n)$, productivity $p_n(e_{n-1}, p_{n-1})$ and outcome $X(e_N, p_N)$ unspecified. However, it remains true that each agent makes their effort choice in sequence and conditional on their productivity, which is the output of the previous agent's effort choice. Thus, the depiction in Figure 1 captures this general process as well.

We assume $p_n(e_{n-1}, p_{n-1})$ is twice differentiable, and (weakly) increasing and concave in each argument. Thus, higher effort by the agent in stage n implies that the agent in stage $n+1$ will be more productive. Similarly, for a given level of e_n, higher productivity in stage n implies that the agent in stage $n+1$ will be more productive. Since outcome $X(\cdot)$ is the analog of the productivity functions, we assume again that it is twice differentiable, and (weakly) increasing and concave in each argument. Moreover, we continue to assume that only X is contractible.

Finally, we assume that the total impact of productivity shifts in stage n on productivity changes – direct and indirect – in stage $n+1$, for $n < N$ is positive. That is:

$$\frac{dp_{n+1}(e_n^*(p_n), p_n)}{dp_n} = \frac{\partial p_{n+1}}{\partial p_n} + \frac{\partial p_{n+1}}{\partial e_n}\frac{\partial e_n^*}{\partial p_n} > 0 \tag{39}$$

The analogous assumption holds for the total derivative of the final output $X(e_N(p_N), p_N)$. While we assume that, in (39), both the direct effect of productivity $\frac{\partial p_{n+1}}{\partial p_n}$ and the effect of effort $\frac{\partial p_{n+1}}{\partial e_n}$ are positive., we'll investigate the impact of different assumptions on the relationship between stage n effort and productivity. Specifically, We'll formally define the context in which $\frac{\partial e_n^*}{\partial p_n} > (<) \ 0$ as "super- (sub-) modularity." To appreciate why this is appropriate here, note that e_n impacts e_{n+1} only through its effect on productivity p_{n+1}. Thus $\frac{de_{n+1}}{de_n} = \frac{\partial e_{n+1}}{\partial p_{n+1}}\frac{\partial p_{n+1}}{\partial e_n}$.

The first-best effort choice by agent n, conditional on her observed productivity p_n is given by the effort level e_n^* that satisfies the following, which is a general version of (16):

$$\left[\frac{dX}{dp_N}\frac{dp_N}{dp_{N-1}}\cdots\frac{dp_{n+2}}{dp_{n+1}}\frac{\partial p_{n+1}}{\partial e_n}\right] - \left[\frac{dG_n}{\partial e_n} + \sum_{m>n}\frac{dG_m}{de_m}\frac{\partial e_m}{\partial p_m}\frac{dp_m}{dp_{m-1}}\cdots\frac{dp_{n+2}}{dp_{n+1}}\frac{\partial p_{n+1}}{\partial e_n}\right] = 0 \tag{40}$$

We again focus here on linear contracts in order to facilitate a straightforward comparison of the power of schemes across agents. Thus, given risk neutrality, and conditional on observing their productivity p_n and contract parameters C_n and S_n, each agent chooses their utility-maximizing effort level e^{**} by solving the following:

$$\text{Max}_{\ e_n} \ \ C_n \cdot X_n(\boldsymbol{e^n}(e_n)|p_n) - G_n(e_n) \tag{41}$$

where $X_n(\cdot)$ is the stage-n output function defined above. Agent n's utility-maximizing effort choice e_n^{**} satisfies the following necessary first-order condition.

$$C_n \cdot \left[\frac{dX}{dp_N} \frac{dp_N}{dp_{N-1}} \cdots \frac{dp_{n+2}}{dp_{n+1}} \frac{\partial p_{n+1}}{\partial e_n} \right] - \left[\frac{dG_n}{de_n} \right] = 0 \tag{42}$$

Recall that $X(\cdot)$ is a composition of concave functions – $X \equiv X \circ p_{N-1} \circ \cdots \circ p_{n+1}$ – and is thus concave, and that $G_n(\cdot)$ is convex. Thus, the agent's effort choice is unique. A comparison of (40) and (42) reinforce our main insight: even under risk neutrality, the firm and the agent solve problems that differ substantively in terms of their consideration of the impact of the agent's decision on the costs of effort of downstream agents. That is, the disutility externality extends to this general setting, and can be seen in the summation on the left-hand side of (40).

Reflecting the principal's objective that $e_n^{**} = e_n^*$, we set (40) equal to (42) and re-arrange, which yields, for each n:[7]

$$\left[1 - C_n^* \right] \left[\frac{dX}{dp_{N-1}} \frac{dp_{N-1}}{dp_{N-2}} \cdots \frac{dp_{n+2}}{dp_{n+1}} \frac{\partial p_{n+1}}{\partial e_n} \right] = \sum_{m>n} \frac{\partial G_m}{\partial e_m} \frac{\partial e_m}{\partial p_m} \frac{dp_m}{dp_{m-1}} \cdots \frac{dp_{n+2}}{dp_{n+1}} \frac{\partial p_{n+1}}{\partial e_n} \tag{43}$$

Immediately, we see that the only case in which the agent is made the residual claimant – where $C_n^* = 1$ – is for $n = N$. Of course, this is the only agent that doesn't create a disutility externality. The following claim speaks to the power of the sharing schemes for agents along the general production process. It demonstrates that the critical factor in predicting their relative magnitudes – do they increase or decrease? – is the extent to which effort levels in adjacent stages are super- or sub-modular.

Proposition 8. *In the general, joint sequential model with a linear sharing scheme, there is a unique subgame perfect equilibrium in which, for all n, $C_n^* < (>)\, C_{n+1}^*$ if and only if the technology in stage $n+1$ is super- (sub-) modular.*

When increased stage $n+1$ productivity leads to an increase in equilibrium effort in stage $n+1$ (as it is in our core linear model in Section 2) we expect to observe lower-powered

[7] Of course, in addition to the incentive compatibility constraint in (43), there also exists a participation constraint for each n. Since, in this risk neutral context with a linear scheme, this implies a fixed transfer, we focus on the power, or "commission" as we did in Section 2.

schemes for the stage n agent than for the those in stage $n+1$. On the other hand, when it leads to a decrease in stage $n+1$ effort (as it does in the example in Section 3), then the power should decrease. In other words, the driving factor in determining super- or sub-modularity is the sign of $\frac{\partial e_n^*}{\partial p_n}$. It is straightforward to demonstrate that a general analog of Proposition 2 can be derived under a set of assumptions that bound the value of the final agent's output to lie within $[0, 1]$.

Empirically, the determination of sub- vs. super-modular technologies is likely to vary significantly across industries and perhaps firms and work processes within industry. Moreover, it is likely to vary within organization along the production process. Thus, we argue that Proposition 8 represents an empirically-testable hypothesis regarding the relative power of compensation schemes *ceteris paribus*.

5 Final Remarks

We investigate a context in which a principal designs a compensation plan for a production process that reflects the sequential and joint efforts of an arbitrary number of agents. We derive a number of key insights. First, the relative power of the compensation scheme is contingent on the nature of the underlying technology. When agent $n+1$'s production process is super-modular – as in, for example, a simple, linear process – agent n is paid via a lower-powered scheme than is agent $n+1$. The opposite is true in a submodular context. This is independent of the number of agents involved in the process as well as their relative productivity or efficiency. The driving force behind this result is the existence in joint processes of a "disutility externality" through which upstream agents impact the disutility experienced by those downstream. If ignored, as is the case under residual claimancy, this leads to sub-optimal returns. In order to induce the agent to internalize the externality, the principal either meters or boosts the power of the scheme, depending on the nature of the technology.

Second, our set-up allows us to assess the extent to which the compensation policy is discriminatory from the perspective of total expected utility, rather than just pay. We find that this is not necessarily the case. Because agents in equilibrium exert different levels of effort, they will experience different disutility, for which they must be compensated. However, by binding them to their participation constraint, the principal ensures that no agent would experience an increase in expected utility by moving from one stage of the process to another, and thus the scheme should not be seen as discriminatory.

As a final comment, we believe the simplicity of the linear set-up we propose may be applied to many other interesting contexts. In particular, we expect that the disutility externality would be present in the joint, simultaneous case as well. Since this captures the reality in many contexts, this would be worth studying. Similarly, the assumption of a simple sequential line of production might be extended to a variety of settings. As just a few examples, multiple outputs, turn-taking, and the participation of multiple agents at a given stage would each capture realistic and potentially-insightful nuances that may yield additional and valuable insights.

References

Baker, G., R. Gibbons, and K. J. Murphy (1994). Subjective performance measures in optimal incentive contracts. *The Quarterly Journal of Economics 109*(4), 1125–1156.

Che, Y.-K. and S.-W. Yoo (2001). Optimal incentives for teams. *American Economic Review 91*(3), 525–541.

Corts, K. (2007). Teams versus individual accountability: solving multitask problems through job design. *RAND Journal of Economics 38*(2), 467–479.

Edmans, A., I. Goldstein, and J. Zhu (2013). Contracting with synergies. CEPR Discussion Papers 9559, C.E.P.R. Discussion Papers.

Green, J. and N. Stokey (1983). A comparison of tournaments and contracts. *Journal of Political Economy 91*(3), 349–64.

Halac, M., E. Lipnowski, and D. Rappoport (2021). Rank uncertainty in organizations. *American Economic Review 111*(3), 757–86.

Holmstrom, B. (1982, Autumn). Moral Hazard in Teams. *Bell Journal of Economics 13*(2), 324–340.

Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization 7*(0), 24–52.

Itoh, H. (1992). Cooperation in hierarchical organizations: An incentive perspective. *Journal of Law, Economics, and Organization 8*(2), 321–45.

Lazear, E. and S. Rosen (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy 89*(5), 841–64.

Nalebuff, B. and J. Stiglitz (1983). Prices and incentives: Towards a general theory of compensation and competition. *Bell Journal of Economics 14*(1), 21–43.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica 32*(1/2), 122–136.

Segal, I. (1999). Contracting with externalities. *The Quarterly Journal of Economics 114*(2), 337–388.

Segal, I. (2003). Coordination and discrimination in contracting with externalities: divide and conquer? *Journal of Economic Theory 113*(2), 147–181.

Strausz, R. (1999). Efficiency in sequential partnerships. *Journal of Economic Theory 85*(1), 140–156.

Winter, E. (2004). Incentives and discrimination. *American Economic Review 94*(3), 764–773.

Winter, E. (2006). Optimal incentives for sequential production processes. *RAND Journal of Economics 37*(2), 376–390.

A Appendix

Proof of Lemma 1. We'll demonstrate these claims via induction. Starting at N, the first-best outcome, conditional on p_N solves:

$$Max_{e_N} \ \sqrt{p_N}e_N - \frac{1}{2}e_N^2$$

which, of course, yields $e_N^* = \sqrt{p_N}$. Moving back to $n = N-1$, this implies that:

$$\begin{aligned}
\Pi_{N-1}(e_{N-1}; p_{N-1}) &= \sqrt{p_N}e_N - \frac{1}{2}e_N^2 - \frac{1}{2}e_{N-1}^2 \\
&= \frac{1}{2}p_N - \frac{1}{2}e_{N-1}^2 \\
&= \frac{1}{2}\sqrt{p_{N-1}}e_{N-1} - \frac{1}{2}e_{N-1}^2
\end{aligned}$$

First-best effort for agent $N-1$, in turn, is thus $e_{N-1}^* = \frac{1}{2}\sqrt{p_{N-1}}$. This establishes the initial condition for effort (ii), since $\psi_N = 1$ and $\psi_{N-1} = 2$. Moreover, since $p_N = \sqrt{p_{N-1}}e_{N-1} = \frac{1}{2}p_{N-1}$, this also establishes the initial condition for productivity (i). Finally, note that $E[X|p_N]$ is simply, by definition, $\sqrt{p_N}e_N$. Given that $\psi_N = 1$, this completes the initial condition for (iii).

Before proceeding, it's useful to first prove the following:

Claim: For all n, $\Pi_n = \frac{1}{\psi_n}\sqrt{p_n}e_n - \frac{1}{2}e_n^2$

Proof: This holds trivially for $n = N$. Now, assume it holds for general n. This assumption implies that $e_n^* = \frac{1}{\psi_n}\sqrt{p_n}$. Thus:

$$\begin{aligned}
\Pi_{n-1}(e_{n-1}; p_{n-1}) &= \frac{1}{\psi_n}\sqrt{p_n}e_n - \frac{1}{2}e_n^2 - \frac{1}{2}e_{n-1}^2 \\
&= \frac{1}{\psi_n^2}p_n - \frac{1}{2\psi_n^2}p_n - \frac{1}{2}e_{n-1}^2 \\
&= \frac{1}{\psi_{n-1}}\sqrt{p_{n-1}}e_{n-1} - \frac{1}{2}e_{n-1}^2
\end{aligned}$$

This completes the proof of the claim.

Now, returning to the proof of the lemma, we assume that claims (i) - (iii) hold for a given n. Given the above claim, we know that agent $n-1$'s first-best effort maximizes:

$$\Pi_{n-1}(e_{n-1}; p_{n-1}) = \frac{1}{\psi_{n-1}}\sqrt{p_{n-1}}e_{n-1} - \frac{1}{2}e_{n-1}^2$$

This implies directly that $e_{n-1}^* = \frac{1}{\psi_{n-1}}\sqrt{p_{n-1}}$, which proves (ii). By definition,

$$\begin{aligned} p_n &= \sqrt{p_{n-1}}e_{n-1} \\ &= \frac{1}{\psi_{n-1}}p_{n-1} \end{aligned}$$

which proves (i). Finally, if (iii) holds for a given n, then from (ii) we know that:

$$\begin{aligned} X_{n-1}(\boldsymbol{e^{n-1}}(e_{n-1})|p_{n-1}) =& \frac{p_n}{\prod_{m=n}^{N}\psi_m} \\ =& \frac{\sqrt{p_{n-1}}}{\prod_{m=n}^{N}\psi_m}e_{n-1} \end{aligned}$$

which completes the proof. □

Proof of Lemma 2. We first prove the following claim:

Claim: $E[X+\varepsilon|p_n, e_n] = \sqrt{p_n}e_n\prod_{j=1}^{N-n}C_{n+j}^{2^{j-1}}$ Proof: We'll show this via induction in n, given an N. It holds trivially for n=N so we move to n=N-1 (i.e., one step up from the sales equation). We know that sales is given by:

$$\sqrt{p_N}e_N + \varepsilon_N$$

so the level-N agent chooses effort by maximizing:

$$
\begin{aligned}
W_N &= S_N + C_N\sqrt{p_N}e_N - \frac{1}{2}e_N^2 \\
\Rightarrow e_N^* &= C_N\sqrt{p_N} \\
\Rightarrow X(p_N) &= p_N C_N
\end{aligned}
$$

Moving up one step, we can specify output as:

$$X(p_{N-1}, e_{N-1}) = \sqrt{p_{N-1}}e_{N-1}C_N$$

which shows the claim for $n = N - 1$. Now, we assume it holds for arbitrary n and test $n - 1$. Since we know that the agent in level n will maximize the following:

$$W_n = S_n + C_n\sqrt{p_n}e_n \prod_{j=1}^{N-n} C_{n+j}^{2^{j-1}} - \frac{1}{2}e_n^2$$

then we know that her optimal effort level will be given by

$$e_n^{**} = C_n\sqrt{p_n} \prod_{j=1}^{N-n} C_{n+j}^{2^{j-1}} \tag{A.1}$$

and therefore

$$
\begin{aligned}
X(p_n) =& \sqrt{p_n}\Big[C_n\sqrt{p_n} \prod_{j=1}^{N-n} C_{n+j}^{2^{j-1}}\Big] \prod_{j=1}^{N-n} C_{n+j}^{2^{j-1}} \\
=& p_n C_n \prod_{j=1}^{N-n} C_{n+j}^{2^j}
\end{aligned}
$$

All that is left at this point is to substitute the expression for $p_n = \sqrt{p_{n-1}}e_{n-1}$:

$$X(p_n) = \sqrt{p_{n-1}}e_{n-1}C_n \prod_{j=1}^{N-n} C_{n+j}^{2^j}$$

and to re-organize and re-index:

$$\begin{aligned} X(p_n) =& \sqrt{p_{n-1}} e_{n-1} \prod_{j=0}^{N-n} C_{n+j}^{2^j} \\ =& \sqrt{p_{n-1}} e_{n-1} \prod_{j=1}^{N-(n-1)} C_{(n-1)+j}^{2^{j-1}} \end{aligned}$$

which completes the proof of the Claim. Part (i) is shown in Equation(A.1). Part (ii) follows directly from the conditional sales equation in the claim above with n=1:

$$X(p_1, e_1) = \sqrt{p_1} e_1 \prod_{j=1}^{N-1} C_{j+1}^{2^{j=1}}$$

Substituting the expression for optimal effort from (i):

$$e_1^{**} = C_1 C_2 \sqrt{p_1} \prod_{j=2}^{N-1} C_{j+1}^{2^{j-1}}$$

yields:

$$\begin{aligned} X &= \sqrt{p_1} \Big[C_1 C_2 \sqrt{p_1} \prod_{j=2}^{N-1} C_{j+1}^{2^{j-1}} \Big] \prod_{j=1}^{N-1} C_{j+1}^{2^{j-1}} \\ &= \prod_{j=1}^{N} C_j^{2^{j-1}} \end{aligned}$$

which completes the proof of part (ii), recalling that $p_1 \equiv 1$ by assumption. Finally, to show part (iii), we again prove a subsidiary claim first, this time with respect to the unconditional productivity parameters.

Claim: $p_{n+1} = \left(\prod_{k=1}^{n} C_k^{2^{k-1}} \right) C_{n+1}^{2^n - 1} \left(\prod_{j=2}^{N-n} C_{n+j}^{(2^n-1)2^{j-1}} \right) \quad \forall n$

Proof: We again proceed via induction in n. First, we check that it holds for $n = 1$:

$$p_2 = \sqrt{p_1} e_1 = e_1$$

where we can now substitute in the optimal effort from part (i) of the lemma:

$$p_2 = C_1 C_2 \prod_{j=2}^{N-1} C_{j+1}^{2^{j-1}}$$

which, it can easily be determined, satisfies the claim. Now, we assume it holds for given n and solve for p_{n+1}:

$$p_{n+1} = \sqrt{p_n} e_n = p_n C_n C_{n+1} \prod_{j=2}^{N-n} C_{n+j}^{2^{j-1}}$$

We now substitute in the expression for p_n, given by the premise:

$$\begin{aligned} p_{n+1} &= \Big(\prod_{k=1}^{n-1} C_k^{2^{k-1}}\Big) C_n^{2^{n-1}-1} \Big(\prod_{j=2}^{N-n+1} C_{n-1+j}^{(2^{n-1}-1)2^{j-1}}\Big) C_n C_{n+1} \prod_{j=2}^{N-n} C_{n+j}^{2^{j-1}} \\ &= \Big(\prod_{k=1}^{n} C_k^{2^{k-1}}\Big) C_{n+1} \Big(\prod_{j=2}^{N-n+1} C_{n-1+j}^{(2^{n-1}-1)2^{j-1}}\Big) \prod_{j=2}^{N-n} C_{n+j}^{2^{j-1}} \end{aligned}$$

We now multiply the two rightmost terms, being careful to keep track of indexes. We pull out the C_{n+1} term and then re-index:

$$\begin{aligned} &= \Big(\prod_{k=1}^{n} C_k^{2^{k-1}}\Big) C_{n+1}^{2^n-1} \Big(\prod_{j=3}^{N-n+1} C_{n-1+j}^{(2^{n-1}-1)2^{j-1}}\Big) \prod_{j=2}^{N-n} C_{n+j}^{2^{j-1}} \\ &= \Big(\prod_{k=1}^{n} C_k^{2^{k-1}}\Big) C_{n+1}^{2^n-1} \Big(\prod_{j=2}^{N-n} C_{n+k}^{(2^{n-1}-1)2^{j}}\Big) \prod_{j=2}^{N-n} C_{n+j}^{2^{j-1}} \\ &= \Big(\prod_{k=1}^{n} C_k^{2^{k-1}}\Big) C_{n+1}^{2^n-1} \Big(\prod_{j=2}^{N-n} C_{n+j}^{(2^{n-1}-1)2^{j-1}}\Big) \end{aligned}$$

which completes the proof. To make the final step, note that:

$$\frac{2^{n-1}-1}{2} 2^{j+1} = (2^{n-1}-1)2^j + \frac{1}{2}2^j = 2^j(2^{n-1}-1+\frac{1}{2}) = 2^j(2^{n-1}-\frac{1}{2}) = 2^j \frac{2^n-1}{2}$$

Now, to complete the lemma, we solve directly for the costs using the claim on productivity:

$$\frac{1}{2}e_n^{**2} = \frac{1}{2}C_{n+1}^2 C_n^2 p_n \prod_{j=2}^{N-n} C_{n+j}^{2^j}$$

$$\frac{1}{2}e_n^{**2} = \frac{1}{2}C_{n+1}^2 C_n^2 \Big[\Big(\prod_{k=1}^{n-1} C_k^{2^{k-1}-1}\Big) C_n^{2^{n-1}-1}\Big(\prod_{j=2}^{N-n+1} C_{n-1+j}^{(2^{n-1}-1)2^{j-1}}\Big)\Big] \prod_{j=2}^{N-n} C_{n+j}^{2^j}$$

All that remains is to simplify, which, to aid the reader, we do sequentially in the index from 1 to N:

$$\begin{aligned}
\frac{1}{2}e_n^{**2} &= \frac{C_n}{2}\Big(\prod_{k=1}^{n} C_k^{2^{k-1}}\Big) C_{n+1}^2 \Big(\prod_{j=2}^{N-n+1} C_{n-1+j}^{(2^{n-1}-1)2^{j-1}}\Big) \prod_{j=2}^{N-n} C_{n+j}^{2^j} \\
&= \frac{C_n}{2}\Big(\prod_{k=1}^{n} C_k^{2^{k-1}}\Big) C_{n+1}^{2^n} \Big(\prod_{j=3}^{N-n+1} C_{n-1+j}^{(2^{n-1}-1)2^{j-1}}\Big) \prod_{j=2}^{N-n} C_{n+j}^{2^j} \\
&= \frac{C_n}{2}\Big(\prod_{k=1}^{n+1} C_k^{2^{k-1}}\Big) \Big(\prod_{j=3}^{N-n+1} C_{n-1+j}^{(2^{n-1}-1)2^{j-1}}\Big) \prod_{j=2}^{N-n} C_{n+j}^{2^j}
\end{aligned}$$

We re-write the middle product to facilitate the multiplication:

$$\Big(\prod_{j=3}^{N-n+1} C_{n-1+j}^{(2^{n-1}-1)2^{j-1}}\Big) = \Big(\prod_{j=2}^{N-n} C_{n+j}^{(2^{n-1}-1)2^j}\Big)$$

and substitute this in:

$$\begin{aligned}
\frac{1}{2}e_n^{**2} &= \frac{C_n}{2}\Big(\prod_{k=1}^{n+1} C_k^{2^{k-1}}\Big)\Big(\prod_{j=2}^{N-n} C_{n+j}^{(2^{n-1}-1)2^j}\Big) \prod_{j=2}^{N-n} C_{n+j}^{2^j} \\
&= \frac{C_n}{2}\Big(\prod_{k=1}^{n+1} C_k^{2^{k-1}}\Big) \prod_{j=2}^{N-n} C_{n+j}^{2^{n+j-1}} \\
&= \frac{C_n}{2}\Big(\prod_{k=1}^{n+1} C_k^{2^{k-1}}\Big) \prod_{k=n+2}^{N} C_k^{2^{k-1}} \\
&= \frac{C_n}{2}\Big(\prod_{k=1}^{N} C_k^{2^{k-1}}\Big)
\end{aligned}$$

where the final step follows from a change of variable $k = n + j$. □

Proof of Proposition 1. Based on the results in Lemma 2, the firm solves the following problem:

$$Max_{\{C_n\}_{n=1}^{N}} \prod_{n=1}^{N} C_n^{2^{n-1}} - \Big[\sum_{n=1}^{N} \frac{C_n}{2}\Big]\Big[\prod_{j=1}^{N} C_j^{2^{j-1}}\Big] \tag{A.2}$$

The N first-order conditions are given by:

$$2^{n-1} C_n^{*2^{n-1}-1} \prod_{m \neq n}^{N} C_m^{2^{m-1}} - \frac{2^{n-1}+1}{2} C_n^{*2^{n-1}} \prod_{m \neq n}^{N} C_m^{2^{m-1}} - 2^{n-1} C_n^{*2^{n-1}-1} \sum_{m \neq n}^{N} \frac{C_m}{2} \prod_{w \neq n}^{N} C_w^{2^{w-1}} = 0 \tag{A.3}$$

In order to simplify the analysis, we divide both sides by the following quantity:

$$2^{n-2} C_n^{2^{n-1}-1} \prod_{m \neq n} C_m^{2^{m-1}}$$

yielding a system of N equations:

$$2 - \frac{2^{n-1}+1}{2} C_n^* - \sum_{m \neq n} C_m = 0 \quad n = 1, ..., N \tag{A.4}$$

We first show that the result must be unique and then derive the optimal sharing rule in closed form. To do so, we take the system of equations and express them in matrix-vector form:

$$F\xi = 2 \tag{A.5}$$

where ξ is a column vector of commissions and F is an N-by-N matrix. Each row of F corresponds to the first-order condition for the sharing rule for agent n. Using Equation (A.4), we define F as follows:

$$F_{ij} = \begin{cases} \frac{2^{i-1}+1}{2} & j = i \\ 1 & j \neq i \end{cases}$$

Since F is of full rank, the solution vector ξ is necessarily unique. Consider now the following solution:

$$\xi_i = 2^{i-N} \tag{A.6}$$

To see that this solves the system in (A.5), we plug it into (A.4) for a given n and assume that it is not a solution:

$$\sum_{i \neq n} 2^{i-N} + \frac{2^{n-1}+1}{2^{n-1}} 2^{n-N} \neq 2 \tag{A.7}$$

Noting that:

$$\sum_{i \neq n} 2^{i-N} = 2 - \left(\frac{1}{2}\right)^{N-1} - \left(\frac{1}{2}\right)^{N-n}$$

We can substitute into (A.7) to yield the contradiction. This completes the proof. □

Proof of Proposition 3. In order to demonstrate this result, we will show that the model specification incorporating a rich set of parameters as captured in (20) and (21) collapses to the set-up of the core model, absent a constant shifting term. First, we'll establish the impact on revenue, relative to the result in Lemma 2.

Claim: For any n, $X(p_n, e_n) = \sqrt{p_n} \cdot e_n \cdot \prod_{j=0}^{N-n} b_{n+j}^{2^j} \frac{C_{n+j}^{\upsilon(j)}}{\gamma_{n+j}^{\upsilon(j)}}$, where $\upsilon(0) = 0$ and $\upsilon(j > 0) = 2^{j-1}$.

We show this via backward induction in n. To see that the claim holds for n=N, note that $X(p_N, e_N)$ is given by $\sqrt{p_N} e_N b_N$. Assuming now that the claim holds for $X(p_n, e_n)$, we check whether this implies it holds for $X(p_{n-1}, e_{n-1})$. Specifically, we assume that

$$X(p_n, e_n) = \sqrt{p_n} e_n \prod_{j=0}^{N-n} b_{n+j}^{2^j} \frac{C_{n+j}^{\upsilon(j)}}{\gamma_{n+j}^{\upsilon(j)}} \tag{A.8}$$

Agent n chooses e_n to maximize the following:

$$E[U_n | p_n, e_n] = S_n + C_n \sqrt{p_n} e_n \left(\prod_{j=0}^{N-n} b_{n+j}^{2^j} \frac{C_{n+j}^{\upsilon(j)}}{\gamma_{n+j}^{\upsilon(j)}} \right) - \frac{\gamma_n}{2} e_n^2$$

which yields

$$e_n^* = \frac{C_n}{\gamma_n}\sqrt{p_n}\Big(\prod_{j=0}^{N-n} b_{n+j}^{2^j}\frac{C_{n+j}^{\upsilon(j)}}{\gamma_{n+j}^{\upsilon(j)}}\Big) \tag{A.9}$$

Substituting (A.9) into (A.8) yields:

$$X(p_n) = p_n\frac{C_n}{\gamma_n}\Big(\prod_{j=0}^{N-n} b_{n+j}^{2^{j+1}}\frac{C_{n+j}^{2\upsilon(j)}}{\gamma_{n+j}^{2\upsilon(j)}}\Big)$$

All that is left is to substitute for $p_n = \sqrt{p_{n-1}}b_{n-1}e_{n-1}$:

$$X(p_{n-1}, e_{n-1}) = \sqrt{p_{n-1}}b_{n-1}e_{n-1}\frac{C_n}{\gamma_n}\Big(\prod_{j=0}^{N-n} b_{n+j}^{2^{j+1}}\frac{C_{n+j}^{2\upsilon(j)}}{\gamma_{n+j}^{2\upsilon(j)}}\Big)$$

and to re-index:

$$X(p_{n-1}, e_{n-1}) = \sqrt{p_{n-1}}e_{n-1}\prod_{j=0}^{N-(n-1)} b_{n-1+j}^{2^j}\frac{C_{n-1+j}^{\upsilon(j)}}{\gamma_{n-1+j}^{\upsilon(j)}}\Big)$$

which completes the proof of the claim.

We can now solve for revenue in equilibrium, which simply follows from the conditional sales equation with n=1:

$$X(p_1 = 1, e_1) = e_1\cdot\prod_{j=1}^{N} b_j^{2^{(j-1)}}\frac{C_j^{\upsilon(j-1)}}{\gamma_j^{\upsilon(j-1)}}$$

Substituting the optimal effort level from (A.9) yields:

$$X = \prod_{j=1}^{N} b_j^{2^j}\frac{C_j^{\upsilon(j)}}{\gamma_j^{\upsilon(j)}} \tag{A.10}$$

Next, we show that the cost of the sales effort is shifted by the same constant factor, relative to the Equation (A.2) above. We do so, again, via induction in n. Starting at $N = 1$, using

Equation (A.9) and shifting the index for clarity:

$$e_1^* = \frac{C_1}{\gamma_1} \prod_{j=1}^{N} b_j^{2^{j-1}} \frac{C_{j+1}^{v(j-1)}}{\gamma_{j+1}^{v(j-1)}}$$

and thus:

$$\begin{aligned}\frac{\gamma_1}{2} e_1^{*2} &= \frac{\gamma_1}{2} \frac{C_1^2}{\gamma_1^2} \prod_{j=1}^{N} b_j^{2^j} \frac{C_{j+1}^{2v(j-1)}}{\gamma_{j+1}^{2v(j-1)}} \\ &= \frac{C_1}{2} \Big(\prod_{j=1}^{N} \frac{b_j^{2^j}}{\gamma_j^{2^{j-1}}} \Big) \Big(\prod_{j=1}^{N} C_j^{2^{j-1}} \Big)\end{aligned}$$

This proves the claim for $N = 1$ (to verify, compare this expression with point 3 in Lemma 2). In order to complete the induction argument, we use Equation (A.9) to form the ratio of successive effort costs:

$$\frac{\frac{\gamma_{n+1}}{2} e_{n+1}^{*2}}{\frac{\gamma_n}{2} e_n^{*2}} = \frac{\frac{C_{n+1}^2}{2\gamma_{n+1}} \Big(\prod_{j=0}^{N-(n+1)} b_{n+j+1}^{2^{j+1}} \frac{C_{n+j+1}^{2(j)}}{\gamma_{n+j+1}^{2(j)}} \Big)}{\frac{C_n^2}{2\gamma_n} \Big(\prod_{j=0}^{N-n} b_{n+j}^{2^{j+1}} \frac{C_{n+j}^{2(j)}}{\gamma_{n+j}^{2(j)}} \Big)} \cdot \frac{p_{n+1}}{p_n} \tag{A.11}$$

We note that:

$$\begin{aligned} & p_{n+1} = \sqrt{p_n} b_n e_n \\ \Leftrightarrow \quad & \frac{p_{n+1}}{p_n} = \frac{b_n C_n}{\gamma_n} \prod_{j=0}^{N-n} b_{n+j}^{2^j} \frac{C_{n+j}^{v(j)}}{\gamma_{n+j}^{v(j)}} \end{aligned}$$

One can verify that (A.11) reduces to:

$$\frac{\frac{\gamma_{n+1}}{2} e_{n+1}^{*2}}{\frac{\gamma_n}{2} e_n^{*2}} = \frac{C_{n+1}}{C_n}$$

To complete the induction proof, note that the premise that:

$$\frac{\gamma_n}{2} e_n^{*2} = \frac{C_n}{2} \prod_{j=1}^{N} \frac{b_j^{2^j}}{\gamma_j^{2^{j-1}}} C_j^{2^{j-1}}$$

implies, by the ratio analysis above, that:

$$\begin{aligned}\frac{\gamma_{n+1}}{2} e_{n+1}^{*2} &= \frac{C_n}{2} \prod_{j=1}^{N} \frac{b_j^{2^j}}{\gamma_j^{2^{j-1}}} C_j^{2^{j-1}} \cdot \frac{C_{n+1}}{C_n} \\ &= \frac{C_{n+1}}{2} \prod_{j=1}^{N} \frac{b_j^{2^j}}{\gamma_j^{2^{j-1}}} C_j^{2^{j-1}}\end{aligned}$$

Combined, these analyses prove that, in this richer specification, the firm's compensation scheme is chosen to maximize:

$$X = \left(\prod_{j=1}^{N} \frac{b_j^{2^j}}{\gamma_j^{2^{j-1}}}\right)\left[\left(\prod_{j=1}^{N} C_j^{2^{j-1}}\right) - \sum_{n=1}^{N} \frac{C_n}{2} \prod_{j=1}^{N} C_j^{2^{j-1}}\right]$$

which differs from our core model in Proposition 1 only by a constant shifting term which is a function of the $n-$level parameters. □

Proof of Proposition 4. Following the proof of Proposition 3, all that remains is to show that the participation constraint is satisfied. Proposition 3 shows that the expected incentive compensation will shift linearly with the quantity in (25). In the induction argument in the proof of Proposition 3, we show that the same is true of disutility. Thus, the agent's payout W_n shifts linearly with (25) and the expected return can be found trivially as long as (25) has an expected value. □

Proof of Proposition 5. We consider the random arrival in time t of funds available for productivity improvement. The claim is that in the limit as enough funds arrive for such investment, the distribution of productivity approaches that characterized in the Proposition. In order to prioritize such investments, we begin by assigning agent m the rank $r(m)$ in order of the expected marginal return on improved productivity. This assignment is accomplished in such a way that the set of agents with the highest values of d_m are assigned rank $r(m) = 1$ and so forth, without gaps. Formally, define the correspondence $y(k) \equiv r^{-1}(k)$ as the set of agents with marginal returns on productivity investments that rank kth among agents. Define $y(1) \equiv \{m | d_m \in max_{n \in N} \; d_n\}$. Generally, define $y(k) \equiv \{m | d_m \in max_{n \in N \setminus \bigcup_{h=1}^{k-1} y(h)} \; d_n\}$. Now, consider the following optimal investment policy: *Step (i):* Divide funds equally across $y(1)$ agents until funds run out or $d_{y(1)} = d_{y(2)}$. *Step (ii):* If funds run out, stop and wait for the arrival of more investment funds. *Step (iii):* If funds remain, re-calculate all $r(i)$ and return to Step (i) with remaining funds. Clearly, this policy gets (weakly) closer to the distribution of productivity described in the proposition each time new funds arrive. As the available funds gets large enough, we approach this distribution. □

Proof of Proposition 6. Formally, the firm solves the following problem:

$$Max_{C_n} \prod_{n=1}^{N} C_n^{2^{n-1}} - \sum_{n=1}^{N} \frac{C_n}{2} \prod_{j=1}^{N} C_j^{2^{j-1}}$$
$$s.t. \sum_{n}^{N} C_n \leq C, \;\; C \in (0,1]$$

Note that this specification already reflects the imposition of both the participation and individual rationality constraints for the N agents. We now form the Lagrangian to incorporate the constraint into the objective:

$$L = \prod_{n=1}^{N} C_n^{2^{n-1}} - \sum_{n=1}^{N} \frac{C_n}{2} \prod_{j=1}^{N} C_j^{2^{j-1}} + \lambda(C - \sum_{n}^{N} C_n)$$

The necessary conditions for a constrained maximum are:

$$\frac{\partial L}{\partial C_n} \leq 0 \ \forall n \tag{A.12}$$

$$C_n \in (0,1] \ \forall n \tag{A.13}$$

$$\lambda \geq 0 \ \forall n \tag{A.14}$$

$$\lambda \cdot (C - \sum_{n}^{N} C_n) = 0 \tag{A.15}$$

$$C_n \cdot \frac{\partial L}{\partial C_n} = 0 \ \forall n \tag{A.16}$$

Since we know that the unconstrained maximum (i.e., $\lambda = 0$) violates the sharing constraint for $C \in (0,1]$, we know that the constraint will bind. This implies that $(C - \sum_{n}^{N} C_n) = 0$. Noting that for this to be a meaningful solution, we further impose $C_n > 0 \ \forall n$ so that:

$$\frac{\partial L}{\partial C_n} = 0 \ \forall n$$

Or:

$$\Big[2^{n-1} - \frac{2^{n-1}+1}{2} C_n^*\Big] C_n^{*(2^{n-1}-1)} \prod_{m \in N \setminus n} C_m^{2^{m-1}} - 2^{n-1} C_n^{*(2^{n-1}-1)} \sum_{m \in N \setminus n} \frac{C_m}{2} \prod_{j \in N \setminus n} C_j^{2^{j-1}} = \lambda \ \forall n$$

We divide both sides of the nth equation by:

$$2^{n-1} C_n^{*(2^{n-1}-1)} \prod_{m \in N \setminus n} C_m^{2^{m-1}}$$

which allows us to derive the following from the nth and $(n+1)$th equations:

$$2 - \frac{2^{n-1}+1}{2^{n-1}} C_n^* - \sum_{m \in N \setminus n} C_m = \frac{C_{n+1}^*}{2C_n^*} \Big[2 - \frac{2^n+1}{2^n} C_{n+1}^* - \sum_{m \in N \setminus (n+1)} C_m\Big] \tag{A.17}$$

Assume that $C^*_{n+1} \neq 2C^*_n$, which implies that:

$$\frac{2^{n-1}+1}{2^{n-1}}C^*_n + \sum_{m<(n-1)} C_m + C^*_{n+1} + \sum_{m>n+1} C_m \neq \frac{2^n+1}{2^n}C^*_{n+1} + \sum_{m<(n-1)} C_m + C^*_n + \sum_{m>n+1} C_m$$
$$\Leftrightarrow \frac{2^{n-1}+1}{2^{n-1}}C^*_n + C^*_{n+1} \neq \frac{2^n+1}{2^n}C^*_{n+1} + C^*_n$$
$$\Leftrightarrow \frac{2^{n-1}+1}{2^{n-1}}C^*_n + C^*_n \neq \frac{2^n+1}{2^n} \cdot 2C^*_n$$

which can be readily verified to present a contradiction. All that remains is to show that $\zeta \equiv \frac{C^*_{n+1}}{C^*_n} = 2\ \forall$n is the (unique) solution to the problem. We start by re-writing Equation (A.17):

$$2 - \frac{2^{n-1}+1}{2^{n-1}}C^*_n - \sum_{m\in N\setminus n} C_m = \frac{\zeta}{2}\Big[2 - \frac{2^n+1}{2^n}\zeta C^*_n - \sum_{m\in N\setminus(n+1)} C_m\Big]$$
$$\Leftrightarrow 2 - \frac{2^{n-1}+1}{2^{n-1}}C^*_n - \sum_{m<n} C_m - \zeta C^*_n - \sum_{m>n+1} C_m = \frac{\zeta}{2}\Big[2 - \frac{2^n+1}{2^n}\zeta C^*_n - \sum_{m<n} C_m - C^*_n - \sum_{m>n+1} C_m\Big]$$
$$\Leftrightarrow K - \frac{2^{n-1}+1}{2^{n-1}}C^*_n - \zeta C^*_n = \frac{\zeta}{2}\Big[K - \frac{2^n+1}{2^n}\zeta C^*_n - C^*_n\Big]$$

where $K \equiv 2 - \sum_{m<n} C_m - \sum_{m>n+1} C_m$. This yields the quadratic equation:

$$\zeta^2 \frac{2^{n-1}+1}{2^{n+1}}C^*_n - \zeta\Big[\frac{C^*_n}{2} + \frac{K}{2}\Big] + K = 0 \qquad \text{(A.18)}$$

Given that $\frac{2^{n-1}+1}{2^{n+1}}C^*_n > 0$ and that $-\Big[\frac{C^*_n}{2} + \frac{K}{2}\Big] < 0$, this can yield at most one positive solution.

In order to complete the proof, we need to solve the finite geometric sum, noting again that the constraint will bind in equilibrium:

$$
\begin{aligned}
&C_1^* + C_2^* + C_3^* + ... + C_N^* = C \\
\Leftrightarrow\ &C_1^* + 2 \cdot C_1^* + 2^2 \cdot C_1^* + ... + 2^{N-1} \cdot C_1^* = C \\
\Leftrightarrow\ &C_1^* \sum_{n=1}^{N} 2^{n-1} = C \\
\Leftrightarrow &C_1^* = \frac{C}{2^N - 1}
\end{aligned}
$$

Finally, since $C_n^* = 2^{n-1} C_1^*$, it follows that $C_n^* = \frac{2^{n-1}}{2^N-1} C$ □

Proof of Proposition 7. Part 1 of the proof is shown in the main body of the paper. To see part 2, we can complete the analysis of (2.4) and note that the premise implies that

$$X + 2R\sigma^2 C_{n+1}^* < X + 2R\sigma^2 C_n^*$$

which, given part 1, is clearly contradicted. Finally, to see part 3, note that, in equilibrium:

$$\frac{C_n^*}{C_{n+1}^*} = \frac{\frac{X}{2} + R_{n+1}\sigma^2 C_{n+1}^*}{X + 2R_n\sigma^2 C_n^*}$$

and, note that:

$$\frac{\frac{X}{2} + R_{n+1}\sigma^2 C_{n+1}^*}{X + 2R_n\sigma^2 C_n^*} < \frac{\frac{X}{2} + R_n\sigma^2 C_{n+1}^*}{X + 2R_n\sigma^2 C_n^*} < 1$$

Where the first inequality is implied by declining risk aversion and the second is proven in part 1 of the proposition. □

Proof of Proposition 8. We begin by writing an expression analogous to (43) for the agent in stage $n + 1$:

$$\left[1 - C_{n+1}^*\right] \left[\frac{dX}{dp_N} \frac{dp_N}{dp_{N-1}} \cdots \frac{dp_{n+3}}{dp_{n+2}} \frac{\partial p_{n+2}}{\partial e_{n+1}^*}\right] = \sum_{m>n+1} \frac{\partial G_m}{\partial e_m} \frac{\partial e_m}{\partial p_m} \frac{dp_m}{dp_{m-1}} \cdots \frac{dp_{n+3}}{dp_{n+2}} \frac{\partial p_{n+2}}{\partial e_{n+1}^*} \quad \text{(A.19)}$$

Next, we multiply both sides of (A.19) by:

$$\frac{\frac{dp_{n+2}}{dp_{n+1}} \cdot \frac{\partial p_{n+1}}{\partial e_n^*}}{\frac{\partial p_{n+2}}{\partial e_{n+1}^*}}$$

which yields:

$$\left[1 - C_{n+1}^*\right]\left[\frac{dX}{dp_N}\frac{dp_N}{dp_{N-1}}\cdots\frac{dp_{n+2}}{dp_{n+1}}\frac{\partial p_{n+1}}{\partial e_n^*}\right] = \sum_{m>n+1}\frac{\partial G_m}{\partial e_m}\frac{\partial e_m}{\partial p_m}\frac{dp_m}{dp_{m-1}}\cdots\frac{dp_{n+2}}{dp_{n+1}}\frac{\partial p_{n+1}}{\partial e_n^*} \quad \text{(A.20)}$$

Subtracting (A.20) from (43) yields the following:

$$C_{n+1}^* - C_n^* = \frac{\frac{dG_{n+1}}{de_{n+1}}\frac{\partial e_{n+1}^*}{\partial p_{n+1}}\frac{\partial p_{n+1}}{\partial e_n^*}}{\frac{dX}{dp_N}\frac{dp_N}{dp_{N-1}}\cdots\frac{dp_{n+2}}{dp_{n+1}}\frac{\partial p_{n+1}}{\partial e_n^*}} \quad \text{(A.21)}$$

Our assumptions ensure that the denominator is positive, and G_m is strictly increasing, which completes the proof. □