DOI: 10.1111/j.1475-679X.2007.00253.x
Journal of Accounting Research
Vol. 45 No. 4 September 2007
Printed in U.S.A.

Performance Evaluations and Efficient Sorting

KOROK RAY*

Received 27 July 2005; accepted 6 February 2007

ABSTRACT

Much of the production in firms takes place over time. This paper seeks to understand the value of interim performance information on long projects. In particular, the model explores the sorting effects of performance evaluations. Conducting an interim performance evaluation increases efficiency by providing the option to end projects with low early returns. The main result: It is efficient to allocate more resources towards the end of a project. This result holds under a variety of scenarios: when the worker has unknown ability, when the outside options vary with output, and even under an agency framework with a risk-averse agent.

1. *Introduction*

The production of goods and services takes time. It takes time to build cars, write software, develop drugs, formulate strategy, market products, audit clients, and value companies. In fact, many firms today organize their production in the form of projects that run for weeks, months, or even years at a time before the firm can sell the finished product in a market.

*University of Chicago. This paper is based on my dissertation submitted to Stanford Graduate School of Business. I owe the greatest debt to my thesis advisor, Edward P. Lazear, for years of feedback, guidance, and encouragement. He is a true mentor, and I am honored to be his student. John Roberts, Andy Skrzypacz, and an anonymous referee vastly improved the paper. Thanks also to Stan Baiman, Navin Kartik, Rick Lambert, Paul Oyer, Haresh Sapra, Kathryn Shaw, and seminar participants at Aarhus, Chicago, INSEAD, IZA, Kellogg, Stanford, Stockholm, and Wharton for helpful comments. Maris Goldmanis and Sergey Mityakov provided excellent research assistance. Chicago, Graduate School of Business, Stanford Graduate School of Business, and the Olin Foundation provided generous financial support.

For example, software development runs through design, implementation, and testing stages before final release. This paper examines the problem of performance evaluation before a project is finished.

The main idea is that performance evaluations generate efficient sorting. They provide the firm with the option of ending projects with low early returns. A performance evaluation conducted at the interim stage sorts employees into two groups: stay or quit. It is efficient to quit if the early returns are weak, and to stay otherwise. If the agent's early output is low, it is in the interest of both the firm *and* the agent to discontinue work and collect their respective outside options. In this sense, the performance evaluation assigns an agent to his most efficient use. The role of the evaluation, therefore, is to perform this sorting.

Interim performance evaluations change the efficient allocation of resources over time in surprising ways. The main result shows that it is efficient to assign more resources to the later stages of the project. The evaluation creates the possibility of termination after the early stage, which reduces the marginal return from first-stage effort, and so the agent shades his effort downward in the early stage. Once the agent advances, the possibility of termination vanishes and the marginal return to effort rises, so he works harder. In sum, it is efficient to stay if the early returns are high, and for those that stay, it is efficient to work harder.

The usual analysis of performance evaluations takes place in agency models that span an enormous literature in accounting, economics, and finance (see Baiman [1982, 1990], Indjejikian [1999], Lambert [2001], Prendergast [1999] for reviews). Agency theory has enjoyed wide theoretical attention over the last 20 years, and has dramatically advanced our understanding of performance measurement. Unfortunately, its empirical support remains mixed, as Prendergast [2002] and Lazear [2003] document. The results of most agency models are highly sensitive to their details—the information structure, the timing of the game, the restrictions on the contracts, etc. So when the details of the game change, even slightly, so do the predictions of the model.

This paper takes a different approach. The focus here is not on incentives, but on sorting. I propose a simple and robust efficiency model that does not rely on a complex contracting game within the firm. The aim is to understand the value of interim performance information to all parties, not the opportunistic use of information by one side.[1] The analysis here shifts attention away from *distortions* from first-best to *improvements* in first-best. Performance evaluations don't just slice up a fixed pie, but make the pie itself bigger.

Exploring the sorting effects of interim performance evaluations in a first-best world abstracts from conflicts of interests between parties and makes more apparent the economic forces driving the effort allocation decision.

[1] See Lizzeri, Meyer, and Persico [2002] or Ray [2006] for agency models of interim performance evaluations.

Nonetheless, I show that the main result of the model is still robust within an agency context. In particular, when a risk-neutral principal cannot observe effort and contracts with a risk-averse agent, the agent still shifts his effort into the later stages. The difference in risk preferences between the two parties causes the agent to distort his effort choice within each period, but does not qualitatively change his effort decision between periods. The sorting effect of the interim evaluation only impacts the agent's dynamic decision problem, and does not interact with the difference in preferences between the principal and agent. Therefore, the result still holds within an agency context, even though it is not necessary to operate in a second-best world to generate the sorting effect.

The model consists of an agent working for a firm on a project over two stages. His effort levels across stages are perfect substitutes, so the production technology by itself does not skew his effort allocation. Production is *indivisible*: The firm cannot capture the full value of output until the end of the project. This is the sense in which production takes time. In particular, incomplete projects fetch a low market price, normalized to zero. The firm has invested so much firm-specific capital and labor into the project that it becomes too costly for other firms to understand and finish it. Finally, firms and workers have outside options, which include future profits from new projects for the firm and future wages from new jobs for the employee.

The existing body of theoretical literature on intertemporal allocation of effort in multistage projects falls into three categories: first-best analyses, models based on general agency considerations under discounting and/or risk aversion, and career-concerns models. The first strand of this literature, dating back to the early 1970s, is similar to my paper in that it examines *efficient* intertemporal effort allocation in research and development (R&D) projects, in the absence of any incentive problems. Examples include Kamien and Schwartz [1971], Luss [1975], and Grossman and Shapiro [1986]. Kamien and Schwartz [1971] study the time pattern of optimal R&D expenditures when the total cumulative effort E required to complete the project is not known. They find that per-period investment (effort) is increasing in time if the hazard rate associated with the cumulative distributive function of E is nondecreasing.

Luss [1975], on the other hand, examines a setup in which a project is completed in known finite time, but its returns depend on effort exerted each period, and the effectiveness of effort decreases over time. He finds that it is optimal to exert the highest effort in the first period, medium-level constant effort in the following intermediate periods, and the lowest effort in the last period. Grossman and Shapiro [1986] argue that when the total amount of effort required to complete a project is known, optimal effort increases over time. This result occurs due to discounting: The marginal product of effort falls at the discount rate. However, when time to completion is not known, effort is higher when the expected value of the project is higher, which does not necessarily imply a time-monotonic path of effort. This result is shown to be in line with Kamien and Schwartz's [1971] result: If the hazard rate is increasing over time, so is the expected project value.

Finally, it is interesting to note that in the discrete two-period case, their model predicts that second stage effort will be *lower* than first stage effort, provided that the second stage occurs. This is due to a disappointment effect: The occurrence of the second stage is unambiguously bad news; it shows that the project will take two periods rather than one.

The second strand of literature addresses optimal effort choice in an agency setting. A seminal paper in this line of work is Lambert [1984], who shows that when the project has time separable, mutually independent returns each subperiod, a risk-averse manager will choose effort to "smooth" income to keep it close to its ex ante expectation. In particular, second-stage effort is decreasing in observed first-stage output (but it can be higher or lower than first-stage effort). The author also shows that this result is an artifact of the incentive problem: Under the first-best, each period's effort is independent of the previous periods' output.

Lambert's [1984] main result seems to be at odds with mine, since in my model a satisfactorily high first-period output is followed by a high second-stage effort, while a low first-period effort results in termination and no further effort. This difference is a consequence of the independence of the two-period outcomes in Lambert [1984]. Since the expected gains from both periods' efforts in his model are the same and can be collected separately, the only consideration is intertemporal smoothing. A risk-averse agent smooths his effort choice to reduce variance in his consumption. In my model, both periods' efforts contribute only to the single final outcome and first period effort is useless if no second period occurs. Thus, unlike Lambert's [1984]model, the stages here are inherently linked. This plus the risk of termination causes my agent to work harder in later stages. This is true even if the agent is risk averse.

Other papers in the dynamic incentives literature do not deliver clear-cut predictions on intertemporal effort choice (such as Christensen, Feltham, and Şabac [2003a,b], Christensen et al. [2003], Feltham, Indjejikian, and Nanda [2003], Indjejikian and Nanda [1999]). For example, Feltham, Indjejikian, and Nanda [2003] use a dynamic agency model to compare two different performance measurement systems.[2] While their model is cast in a multiperiod LEN (Linear contracts, Exponential utility, Normal errors) framework, their focus is not on effort allocation but instead on which system generates more profit for the firm. As such, a comparison of effort levels over time is ambiguous and depends on a joint condition of several parameters. The authors do not solve for whether and when $e_1 < e_2$—nor should they, as that is not the aim of their paper. This speaks to a broader point: The dynamic incentives literature explores issues of commitment, renegotiation, short-term versus long-term contracts, measurement systems, etc., but not issues of dynamic effort allocation.

[2] They find that a single, dual purpose measurement system is simultaneously informative of past performance and future productivity, and hence dominates a differentiated measurement system.

The third strand of the relevant literature is the career-concerns literature, which is exemplified by the seminal work of Hölmstrom [1999]. He claims that agents (inefficiently) choose higher effort in earlier stages of their careers in order to send favorable signals to the market regarding their abilities. In a similar vein, Lewis [1986] shows that firms' reputation concerns lead to an effect similar to managers' career concerns. Gibbons and Murphy [1992] show that the optimal contract in the presence of career concerns is one that offers less steep compensation in earlier stages of the agent's career, when career concerns are the most severe. It is tempting to conclude that the dynamic effort allocation predicted by Hölmstrom [1999] and the other career-concerns papers rests on Bayesian updating of the agent's underlying ability. In fact, even if this updating takes place, sorting still guarantees that effort levels rise over time, as I show in section 6 of my paper.

The primary difference between all of the foregoing models and my model is that these papers do not explicitly account for the type of projects discussed here: namely, projects whose benefits can be reaped only after the completion of the *entire* project, which is not guaranteed (i.e., it is possible that the project will be terminated early). The combination of indivisible production and outside options in my model makes it efficient to abandon such projects if early returns are low, since the early returns are a signal of final returns. These assumptions are natural in the context of R&D-like projects and are used in models of research and development projects for at least a quarter of a century (see Roberts and Weitzman [1981], where such projects are termed SDPs or sequential development projects). However, this all-or-nothing nature of many research and development endeavors has not been captured by any of the existing intertemporal effort allocation papers.[3]

Overall, the disagreements found in the literature are not as fundamental as they seem. All papers agree that effort is higher when its return is higher. The only point of contention is the source of period-to-period variations in the return to effort: most frequently, such differences arise from discounting or preferences (Grossman and Shapiro [1986], Lambert [1984], Toxvaerd [2006]), or from career or reputation concerns (Gibbons and Murphy [1992], Hölmstrom [1999], Lewis [1986]). My paper shows that in the case of an SDP (Roberts and Weitzman [1981]) early effort is less valuable than late effort, because effort at each stage contributes equally to the final outcome if the final stage is reached, but the benefits from early effort are curtailed by the risk of early termination of the project.

The empirical literature on dynamic effort allocation shows that effort rises over time, consistent with the main prediction of this paper. The main literature comes from empirical operations management, which

[3] A notable exception is the paper by Harbaugh and Klumpp [2005], which models a two-round tournament where only first-stage winners advance to the second stage. They show that underdogs (i.e., players with a low probability of advancing to the second stage) exert higher effort in the first stage while favorites (players with a high probability of advancing) save their effort for the second stage.

documents that projects within firms often run late, over budget, and dedicate most resources towards the final stages (see Marshall and Meckling [1962] and Mansfield, Schnee, and Wagner [1995]). In particular, Barry, Mukhopadhyay, and Slaughter [2002] and Genuchten [1991] find that software projects are often subject to both delays and backloaded resource allocation. A second literature comes from psychology, which documents the tendency of managers to exert low effort well before the project deadline but high effort right before the deadline (see O'Donoghue and Rabin [1999] and citations within). These are precisely the production environments that fit my model: long-term projects combined with intermediate milestones that offer the option to quit before completion. Thus, at a first pass, the predictions of my model are consistent with existing evidence. Prior work either does not generate a precise prediction or generates one inconsistent with evidence.[4]

The paper is organized as follows. Section 2 outlines the basic model, and shows that gathering intermediate performance information raises total surplus. Section 3 establishes the main result: Collecting this intermediate performance information skews the efficient allocation of effort towards the end. Section 4 shows that the model is robust under three settings: the T-stage extension of the two-stage model that incorporates a finite number of evaluations, detailed evaluations when the agent conditions his effort on first-stage output, and outside options that vary with first-stage output. Section 5 lays out an agency model between a risk-neutral firm and risk-averse agent, and shows that the main result still holds in this environment. Section 6 includes an ability parameter for the agent that is unknown to either party and that persists across both stages. Under the monotone likelihood ratio property, the agent still shifts his effort into the later stages. Finally, I discuss an application to up-or-out schemes in accounting firms, and section 7 concludes.

2. *The Model*

Consider an agent working on a project for a firm. Production takes place across two stages, there is no discounting, and both parties are risk neutral. The agent exerts effort e_t in stage $t = 1, 2$ at cost $C(e_t)$, where C', C'' > 0. Although the model considers e_t as effort, it is possible to interpret it more generally as any costly resource (like capital). The agent's output from the project is $q_t = e_t + \varepsilon_t$. The noise terms ε_t are i.i.d., and distributed symmetrically around a mean of zero, with cumulative distributive function $G(\cdot)$ and density function $g(\cdot)$, which is continuously differentiable and has

[4] For example, the career-concerns models enjoy mixed empirical support. Hölmstrom's [1999] manager decreases effort over time. However, measures of labor supply over the life cycle rise over time up until the very last few years before retirement. See MaCurdy [1981], Heckman and MaCurdy [1980], and Heckman [1974] for life-cycle models; Imai and Keane [2004] for recent evidence; and Blundell and MaCurdy [1999] and Heckman [1993] for surveys.

support $(-\infty, \infty)$. Interpret ε_t as a stage-specific technology shock unknown to anyone. For now, assume effort e_t is observable to the agent *and the firm*, so the focus is on efficient resource allocation. Section 5 relaxes this assumption and considers moral hazard.

Let $V(q)$ be the value of the project after stage two, where $q = q_1 + q_2$. Thus, q_t is the project's *internal* output within the firm (prototypes, beta versions, etc.), while $V(q)$ measures the project's *external* value based on market prices. Production is *indivisible* in that q_1 fetches a market price of zero. Unfinished projects have no external value; the project cannot be sliced into pieces and sold separately. Work after the first stage alone has value only inside the firm, insofar as it becomes part of a finished product that the firm eventually sells in a market. Much of the investments into a project over time are firm specific, such as the organization of capital and labor surrounding the project. The market price of zero reflects the high costs for another firm to comprehend, evaluate, and ultimately complete another firm's unfinished project.

Assuming that production is a function of $q = q_1 + q_2$ is equivalent to assuming that effort levels across stages are perfect substitutes. This isolates the effects of performance evaluations on effort from the effects of technology on effort. So any discrepancy between e_1 and e_2 arises from information and not from technology. Other assumptions on production are standard. Let V be strictly increasing, weakly concave, and have a continuous first derivative everywhere.[5]

The purpose of an evaluation is to make knowledge of q_1 available after the first stage. While $V(q)$ is observable to all parties, q_1 is observable only if the firm conducts an evaluation. This describes any sort of production where output is not immediately observable, but requires an evaluation to make it observable. The model assumes the evaluation is costless, and focuses on the benefits of conducting a performance evaluation. Finally, once the firm conducts the evaluation, q_1 is observable to both parties. Since the agent is the sole worker on the project, an evaluation of the agent is synonymous with an evaluation of the project.

An evaluation after the first stage gives the firm and the agent the option to terminate the project and therefore end the agent's employment on the project. Let $\bar{u}_t^p$ and $\bar{u}_t^a$ be the outside options of the firm and agent in stage t, respectively. These outside options capture the value of the outside opportunities of both parties and are evaluated at the start of each stage. The agent can switch projects or jobs within the firm, or switch jobs or careers to other firms; the firm can reassign the agent to another project and deploy its resources elsewhere. Let $\bar{u}_t = \bar{u}_t^a + \bar{u}_t^p$, so $\bar{u}_t$ measures the opportunity cost of the project (time, labor, capital) to both parties; assume that $\bar{u}_t > 0$. Note

[5] A large class of production functions satisfy these assumptions. Consider, for instance, the affine and exponential examples $V(q) = \alpha + \beta q$ and $V(q) = \alpha - \beta \exp(-\gamma q)$, for any $\alpha \in \mathbb{R}$ and $\beta, \gamma > 0$. Note that since V is strictly increasing and weakly concave, $\lim_{q \to -\infty} V(q) = -\infty$.

that the outside options are independent of early-stage output. Section 4.3 relaxes this assumption.

An example of a production process that fits this model is software development. Suppose that stage 1 is the design phase, and stage 2 is the actual writing of software code. The firm can sell the complete software package to clients, but not the software design by itself. Moreover, the firm has invested specific capital into the design, in the form of the assignment of managers to teams, lines of communication between groups, a project timetable, etc. If the design fails, then so does the final software package. Evaluating the design/prototype is an extensive procedure involving many parties: discussions with clients, compatibility with other products and divisions, detailed cost accounting, compliance with standards, etc. Finally, both the firm and the programmer have outside options: The programmer can work on other software projects within the firm, and the firm can assign its capital and labor elsewhere.

Note that there are no explicit contracts in the benchmark model. Given that effort is observable, the firm can contract on effort to achieve the first-best, and therefore, it is easy to construct an efficient contract that implements first-best. For clarity, I suppress these efficient contracts from the analysis and focus on the induced effort choices. Throughout the paper, I assume all parties can fully commit to these efficient contracts.[6]

2.1 NO EVALUATION

As a benchmark, suppose that the firm does not conduct an evaluation after the first stage. Neither party is any better informed about output after the first stage than before. Importantly, there are no grounds for terminating the project after the first stage, since performance information is not gathered. The focus of this paper is on efficient resource allocation and thus the analysis centers on the social planner's problem. This is formally equivalent to the analysis of a self-employed agent who must decide when to stop working and how hard to work across stages. Since effort is observable, the firm can implement the first-best solution, and so the agent chooses the effort levels that solve the planner's problem. Let $\mathbb{E}$ denote the expectation taken over $\varepsilon_1 + \varepsilon_2$ in the following definition:

DEFINITION 1. *The efficient effort level under no evaluation is*

$$\hat{e}_t = \arg\max_{e_t} \mathbb{E}V(q_1 + q_2) - C(e_1) - C(e_2).$$

In words, the efficient effort level under no evaluation maximizes total surplus, which is the expected return less the cost of effort in each stage. The first-order conditions yield

$$\mathbb{E}V'(q_1 + q_2) = C'(\hat{e}_t).$$

[6] The issue of renegotiation and lack of commitment in contracts has been extensively explored elsewhere, as in Fudenberg, Holmstrom, and Milgrom [1990].

Notice that the left-hand side of the equality is independent of t, so the right-hand side must be as well. Hence $\hat{e}_1 = \hat{e}_2 \equiv \hat{e}$. It is efficient to split effort evenly across periods. Effort is the same because (1) convexity of the cost function guarantees a unique solution, (2) the marginal return to effort in each period is the same, and (3) the cost function is separable and identical across stages.

The only constraint is a bound on the reservation utilities. If the outside options are too high, then it is inefficient for the firm to employ the agent on the project at all. In particular, it must be that

$$\mathbb{E}V(\hat{e}_1 + \varepsilon_1 + \hat{e}_2 + \varepsilon_2) - C(\hat{e}_1) - C(\hat{e}_2) \geq \bar{u}_1 + \bar{u}_2,$$

where, as before, $\hat{e}_t$ is the efficient effort level under no evaluation. Call this the project feasibility constraint.

3. *Evaluation after the First Stage*

Now suppose that the firm conducts a performance evaluation after the first stage. The main reason for such an evaluation is that it provides the option to abandon the project if the early returns are low. This is efficient because of (1) indivisible production and (2) outside options. Precisely because production is indivisible, the firm and the agent cannot capture the market price for the unfinished project. Instead, they use first-stage output to compute the expected project value after the second stage. Because of the outside options, it is efficient to continue only if this value exceeds these outside options.

Consider two types of evaluations: coarse and detailed. Call an evaluation *detailed* if it returns the exact first-stage output level q_1, and call an evaluation *coarse* if it only returns whether q_1 exceeds a specified threshold, but does not specify the exact magnitude of q_1. The second-stage effort choice depends on whether the evaluation is coarse or detailed. Specifically, if the evaluation is coarse, then the agent chooses an effort *level* e_2, but if the evaluation is detailed, he chooses an effort *function* $e_2(q_1)$.

In this section, and for most of the rest of the paper, I assume coarse evaluations instead of detailed evaluations. First, it may be difficult to assess quality perfectly before the project is finished, especially since production is indivisible and the external market prices projects at zero. Second, coarse evaluation allows for sensible comparisons of effort across stages, since e_1 and e_2 are both numbers, whereas under detailed evaluations, e_1 is a number and $e_2(\cdot)$ is a function. Section 4.2 solves for detailed evaluations.

3.1 QUIT WHILE YOU'RE BEHIND

The linchpin of the analysis here and throughout the paper is the continuation surplus function

$$S(q_1, e_2) = \mathbb{E}_2 V(q_1 + e_2 + \varepsilon_2) - C(e_2),$$

where $\mathbb{E}_t$ denotes the expectation taken over ε_t. This function gives the expected value of continuing the project after the first stage, given that first-stage output is q_1 and planned second-stage effort is e_2. Let e_t^* denote the efficient effort levels when a coarse intermediate performance evaluation *is* conducted; In general this differs from $\hat{e}_t$, the efficient effort *without* an evaluation. For clarity, let $S^*(q_1) \equiv S(q_1, e_2^*)$ be the continuation surplus evaluated at the efficient effort level. The continuation decision rests entirely on this function. In particular, it is clearly efficient to continue if and only if $S^*(q_1) \geq \bar{u}_2$. The first result shows that the continuation surplus function is strictly increasing.

This means there exists a unique cutoff output level q^* such that $S^*(q^*) \geq \bar{u}_2$ if and only if $q_1 \geq q^*$. In words, it is efficient to continue the project if and only if first-stage output exceeds the target q^*, at which point the social planner is indifferent between advancing and retaining the agent at the target. All proofs are in the appendix.

PROPOSITION 1. *Under coarse evaluations, there exists a target q^* such that it is efficient only for agents with $q_1 \geq q^*$ to advance to the second stage.*

The key assumption in this result is that the stages are connected, so output in the early stage signals final project value. Indivisible production guarantees this. If production were divisible, so that the firm could collect value after stage one, it would *always* be efficient to employ the agent in the second stage. For *any* level of q_1, the firm would sell q_1 on the market, pay the agent, and start from scratch in the second stage, since the shocks, ε_t, are independent. The ability to quit halfway is also key, as it is a necessary condition for the existence of a cutoff, q^*.

Proposition 1 shows that the role of an interim performance evaluation is to sort projects and agents into two groups: stay or quit. The target, q^*, conducts the sorting, in that it allows only agents with high output to proceed. Imagine that after stage one, an agent with output q_1 must be assigned to one of two jobs. The first job pays him $S^*(q_1)$ in stage two, and the second job pays $\bar{u}_2$. The interim performance evaluation does this assignment, and does it efficiently.

COROLLARY 1. *Interim performance evaluations strictly increase total surplus.*

The intuition for the corollary is immediate. Whenever the first stage output is so low that the expected contribution to the total surplus is less than the outside options, evaluation allows termination of the project and collection of the outside options instead (which is impossible with no evaluations). When the first-stage output is at least as high as the outside options, the evaluation has no effect. Therefore, as long as low first-stage output occurs with positive probability, conducting an evaluation increases total surplus.

3.2 SETTING THE TARGET

The previous section shows that the continuation decision takes the form of a cutoff rule. Precisely, the planner sets some target (or milestone) q

after the first stage, and advances the agent only if $q_1 > q$. The probability of clearing the target is

$$P_1 = \Pr(q_1 > q) = 1 - G(q - e_1).$$

As expected, this probability increases in first-stage effort for each given target q, since $\partial P_1/\partial e_1 = g(q - e_1) > 0$. If the agent fails to clear the target, both parties capture their outside options, so expected surplus is $(1 - P_1)\bar{u}_2$. If he clears the target, he produces the surplus $S(q_1, e_2)$ for each output level $q_1 > q$. So the planner solves

$$\max_{e_t, q} \int_q^\infty S(q_1, e_2) g(q_1 - e_1)\, dq_1 + (1 - P_1)\bar{u}_2 - C(e_1), \tag{1}$$

where effort levels, e_t, and the target, q, are the planner's choice variables. Let (e_t^*, q^*) denote the efficient choices, that is, the variables that solve the planner's problem (under coarse evaluation). The first term is the expected value of continuing: the continuation surplus function integrated over all realizations of $q_1 > q$. The middle term $(1 - P_1)\bar{u}_2$ is the expected value of abandoning the project. Note that $C(e_2)$ does not appear in the objective function explicitly because it is embedded in $S(q_1, e_2)$. The planner bears the cost of $C(e_2)$ only in the event that the agent advances.

Note that in this program the efficient target, q^*, is set jointly with the efficient effort levels, e_t^*. To get traction on this, consider the first-order condition with respect to e_1 in the planner's problem. After integrating by parts, this condition is

$$C'(e_1^*) = \int_{q^*}^\infty S^{*\prime}(q_1) g(q_1 - e_1^*)\, dq_1 + g(q^* - e_1^*)\big[S^*(q^*) - \bar{u}_2\big],$$

where $S^*(q_1) \equiv S(q_1, e_2^*)$. This equation gives the efficient first-stage effort as a function of the efficient target, q^*, and the efficient second-stage effort, e_2^*. Increasing first-stage effort has two effects: It increases project value, $V(q_1 + q_2)$, and it increases the chances of clearing the target given by $P_1 = \Pr(q_1 > q^*)$. The integral above is the marginal increase in expected value of the project, $V(q_1 + q_2)$, arising from increased e_1. The term in brackets is the expected return from clearing the target. This is the change in the probability of clearing the target $g(q^* - e_1^*) = \frac{\partial P_1}{\partial e_1}$ times the return $S^*(q^*) - \bar{u}_2$, which is the benefit of continuing, $S^*(q)$, less the opportunity cost from quitting.

The presence of the effect represented by the second term might suggest that it is optimal to work harder in the first stage than in the second: If it is known that poor performance in the first stage leads to termination of the project and that working hard reduces the probability of poor first-stage performance, then it seems perfectly reasonable to work hard initially to avoid termination. While this logic is seductive, it is misleading. At the *efficient* target, the planner is indifferent between continuing and terminating the project ($S^*(q^*) = \bar{u}_2$ by the first order condition with respect to q), so that the second term in the expression for $C'(e_1^*)$ vanishes. At the same time, a performance evaluation creates a possibility of early termination, which lowers the first term and thus reduces the marginal return to effort in stage 1 (relative to stage 2). Once the project advances to the next stage, the possibility of termination vanishes, and it becomes efficient to work harder. Performance evaluations offer the firm an option to quit, which generates uncertainty in stage 1. This uncertainty lowers the marginal return to effort early on, and thus skews the allocation of effort towards the end of the project.

3.3 MAIN RESULT: EFFICIENT DYNAMIC EFFORT ALLOCATION

The following proposition solves for the efficient allocation of effort across stages, and is the main result.

THEOREM 1. *It is efficient to work harder in the second stage* ($e_1^* < e_2^*$).

Since the planner sets the target optimally, the marginal return to effort after clearing the target exceeds the marginal return to effort in the first stage. Formally,

$$C'(e_1^*) = \mathbb{E}\big[S^{*\prime}(q_1)\big] < \mathbb{E}\big[S^{*\prime}(q_1) \mid q_1 \geq q^*\big] = C'(e_2^*).$$

The mean marginal return conditional on $q_1 \geq q^*$ exceeds the unconditional mean. Since marginal costs are increasing, this implies that $e_1^* < e_2^*$.

The marginal return to effort is lower in stage 1 precisely because the agent may not advance to the second stage. In this case, the planner bears the cost of effort, $C(e_1)$, but acquires the benefit, $V(q_1 + q_2)$, not with certainty, but with probability less than one. This lowers the marginal return to effort in stage 1 relative to stage 2. At the optimum, the planner sets the marginal costs equal to the marginal returns, consequently shading effort downward in the early stages. More effort is allocated in the later stages of the project, when its marginal return is higher. This logic bears some similarity to the tournament model of Lazear and Rosen [1981]. Recall that in a tournament, an increase in noise chokes off effort for each agent. Extreme noise essentially determines the outcome of the tournament, and hence each agent has little incentive to work hard. Here, the possibility of termination essentially brings additional noise into the first stage, and this noise causes the social planner to shade effort downwards.

4. *Robustness*

4.1 FINITE NUMBER OF EVALUATIONS

A natural question is whether theorem 1 is an artifact of the two-stage setting. For example, if production takes place over three stages, is it still efficient to evaluate performance after the first and/or second stage? And if so, do the efficient effort levels monotonically increase in t? It turns out that Theorem 1 generalizes in the finite stage game. In particular, if a project with indivisible output extends over an arbitrary finite number of stages (instead of just two, as in the benchmark model), it is efficient to conduct performance evaluations after each stage and to hold back effort in the early stages, continually increasing it as time goes on. The result is driven by the same basic forces as in the benchmark model: Because of the indivisibility of the project's outcome, it is efficient to terminate the project early if its expected output is lower than the outside options. Also, it is efficient to shade early effort downward because of the possibility of early termination, which decreases the expected return to effort.

More precisely, consider a model with T interim evaluations. There are, therefore, $T + 1$ time periods. In time period t the agent exerts effort e_t at cost $C(e_t)$ with C', $C'' > 0$. As before, there is a random shock each period, ε_t. The shocks are distributed independently and identically over time with density function $g(\cdot)$. The functions g and C are the same each time period. The intermediate output at period t is $q_t = e_t + \varepsilon_t$. However, just as in the benchmark model, the intermediate outputs have no external value: If terminated before the end of period $T + 1$, the project is worth nothing. If brought to completion, the project is worth $V(Q)$, where $Q = q_1 + \cdots + q_{T+1}$. As before, assume that V is a strictly increasing and weakly concave C^1 function. Evaluation of the project occurs after each of the first T periods (at the time of the t^{th} evaluation, only the first t shocks have been realized). Evaluations are coarse in the sense that they only show whether the cumulative intermediate output to date, $\sum_{s=1}^{t} q_t$, exceeds a specified threshold.[7] Because evaluations are coarse, effort levels cannot be conditioned on the observed intermediate outputs.

The rest of the assumptions are a straightforward extension of the two-stage model. After each evaluation, the social planner decides whether or not to continue the project. If the project is terminated, it cannot be restarted later. The total outside option in period t is $\bar{u}_t > 0$. Outside options are expressed in per-period terms and are therefore additive: If the project is terminated after period t, the total outside options collected are $\sum_{s=t+1}^{T+1} \bar{u}_s$. These outside options are evaluated at the start of each stage. And just as before, effort is contractible, so the planner can implement the first-best with an efficient contract under full commitment. Finally, the project is assumed to be feasible, that is, $\mathbb{E}_0 V(q_1 + \ldots q_T) > \sum_{s=1}^{T+1} \bar{u}_s$, where $\mathbb{E}_0$ denotes the time-zero expectation, given efficient effort levels.

As a concrete example, consider the three-stage game, depicted in figure 1. There are two evaluation points, after the first and second stages. The project is worth $V(q_1 + q_2 + q_3)$ if completed. The payoffs in the figure reflect the joint payoffs of all parties. In stage 1, the agent selects effort e_1, and nature delivers ε_1, realizing q_1. If there is an interim evaluation after the first stage, it takes the form of a target, Q_1^*, such that the agent advances if $q_1 \geq Q_1^*$ (this is proven in the upcoming proposition). If the agent does

[7] Alternatively, we can consider evaluations that also determine whether a particular intermediate output, q_t, exceeds a given threshold. However, as Proposition 2 shows, the intermediate continuation surpluses depend solely on the cumulative outputs, so that obtaining additional information about individual thresholds is redundant.

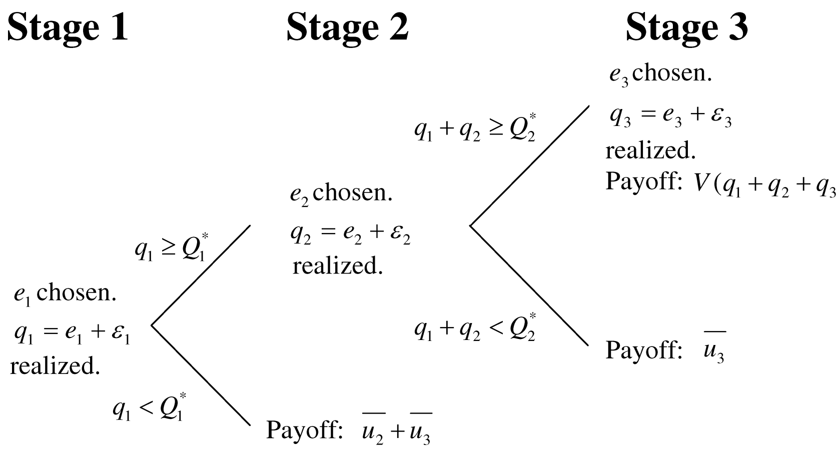

FIG. 1.—The model with three stages and two evaluations.

not clear the first target, he leaves the project and both parties collect their outside options for the remaining stages, $\bar{u}_2 + \bar{u}_3$. If the agent *does* clear the first target, he advances to the second stage, where again he exerts effort and q_2 is realized. Now if his cumulative output to date, $q_1 + q_2$, clears a second target, Q_2^*, he advances to stage 3. If not, he leaves the project and both parties collect their remaining outside option, $\bar{u}_3$. In stage 3, he exerts effort e_3, q_3 is realized, and both parties share the project value, $V(q_1 + q_2 + q_3)$.

Returning to the T-stage model, let $S_t(q_1, \ldots, q_t; e_{t+1}, \ldots, e_{T+1})$ be the continuation surplus at period t: The expected value of continuing the project after the t^{th} evaluation, given the first t intermediate outputs and the planned effort levels in each period.[8] Clearly, the planner chooses to continue after the t^{th} evaluation if and only if the continuation surplus is greater than or equal to the outside options that are obtained if the project is terminated that period, namely, if and only if $S_t(q_1, \ldots, q_t; e_{t+1}, \ldots, e_{T+1}) \geq \sum_{s=t+1}^{T+1} \bar{u}_s$. Thus, the value of having reached the end of the t^{th} period with outputs $q_1, \ldots, q_t$ and a future effort plan $(e_{t+1}, \ldots, e_{T+1})$ is $\max\{S_t(q_1, \ldots, q_t; e_{t+1}, \ldots, e_{T+1}), \sum_{s=t+1}^{T+1} \bar{u}_s\}$.

Consequently, for all $t \in \{1 \ldots T\}$, define recursively

$$S_t(q_1, \ldots, q_t; e_{t+1}, \ldots, e_{T+1}) = \int_{-\infty}^{\infty} \max\left\{ S_{t+1}(q_1, \ldots, q_{t+1}, ; e_{t+2}, \ldots, e_{T+1}), \sum_{s=t+2}^{T+1} \bar{u}_s \right\} \times g(q_{t+1} - e_{t+1})\, dq_{t+1} - C(e_{t+1}),$$

[8] Note that planned effort levels for stages one through t are irrelevant for the continuation surplus after q_t has been realized, since they have no effect on future outputs.

where $S_{T+1}(q_1, \ldots, q_{T+1}) \equiv V(q_1 + \cdots + q_{T+1})$.

While the continuation surplus function has more terms and is more complex than in the single-evaluation case, the same forces identified earlier apply here. In particular, the main result holds.

PROPOSITION 2. *For any finite number of evaluations, T, efficiency requires that:*

1. *There exist cutoff points $\{Q_t^*\}_{t=1}^T$ such that the project is terminated after evaluation t if and only if $\sum_{s=1}^t q_s < Q_t^*$;*
2. *The efficient effort levels are monotonically increasing over time ($e_t^* < e_{t+1}^*$ for all $1 \leq t \leq T$).*

This proposition shows that the result is robust to finitely many stages, and is not simply an artifact of a two-stage model. As before, the same intuition applies. In the early stages, there is a low probability of finishing the game, and this reduces the marginal return to effort, causing the planner to shade effort downward. As the project advances through the stages, it clears successive hurdles, and therefore the probability of collecting the final output, $V(q_1, \ldots, q_T)$, increases. This raises the marginal return to effort and causes more effort to be invested in later stages. However, note that the sequence of hurdles, $\{Q_t^*\}_{t=1}^T$, need not be monotonic in T: Their relative magnitudes depend on the nature of the sequence of outside options, $\{\bar{u}_t\}_{t=1}^T$.

4.2 DETAILED EVALUATION

Return now to the two-stage model, here and for the rest of the paper. As mentioned earlier, the assumption of coarse evaluation eases analysis considerably. Under detailed evaluation, Theorem 1 still holds, but an additional technical assumption is needed and the proof is slightly more complex. Instead of an effort level, e_2, the agent now selects an effort function, $e_2(q_1)$.

First note that the surplus function is still strictly increasing without any additional assumptions. Proposition 1 implicitly assumes a coarse evaluation, but the assumption is without loss of generality as the following corollary shows.

COROLLARY 2. *Under detailed evaluation, there exists a cutoff point, q^*, such that it is efficient only for agents with $q_1 \geq q^*$ to advance.*

An additional condition is necessary for Theorem 1 to remain valid under detailed evaluations. The *marginal* cost function cannot be "too convex." More precisely, the condition is that $C''' \leq 0$. Note in particular that this holds with quadratic costs (since $C''' = 0$ if $C(e) = \frac{c}{2}e^2$), the most common cost of effort function used in applied models.

PROPOSITION 3. *Under detailed evaluation, it is efficient to work harder on average in the second stage ($e_1^* < \mathbb{E}[e_2^*(q_1) \mid q_1 \geq q^*]$), if $C'''(e) \leq 0$ for all e.*

The reason for the additional condition is as follows. Effort is increasing in the main model because the marginal cost of effort is higher in the second period due to higher marginal return to effort (no failure possibility).

The same basic driving force continues to work under detailed evaluations. However, because the second-period effort is now a function of first-period output, it is a random variable from a time-zero perspective. Therefore, the comparison is now between first-stage effort and second-stage *expected* effort. If $C''' \leq 0$, then

$$C'\big(\mathbb{E}\big[e_2^*(q_1) \,\big|\, q_1 \geq q^*\big]\big) \geq \mathbb{E}\big[C'\big(e_2^*(q_1)\big) \,\big|\, q_1 \geq q^*\big] > C'\big(e_1^*\big),$$

where the first inequality is Jensen's inequality. Increasing costs then imply $e_1^* < \mathbb{E}[e_2^*(q_1) \mid q_1 \geq q^*]$. But if $C''' > 0$, Jensen's inequality operates in reverse, and the first inequality flips. In words, the sufficient condition is that the marginal cost of expected effort in stage 2 (the leftmost term above) exceeds the marginal cost of effort in stage 1 (the rightmost term above). If marginal costs are convex, then the former term sinks relative to the latter. If they are sufficiently convex, it is possible that $e_1^* > \mathbb{E}[e_2^*(q_1) \mid q_1 \geq q^*]$. But if the marginal cost function is only slightly convex, the main result still holds.[9]

4.3 OUTSIDE OPTIONS AND FIRST-STAGE OUTPUT

Previously, I assumed that the outside options, $\bar{u}_2$, were independent of the first-stage output, q_1. Now I explore what happens when the outside options vary with q_1. The outside options are now a function of q_1, so let $\bar{u}_2(q_1)$ denote that function. The most natural assumption is that this function is increasing; for higher levels of first-stage output, the firm and/or agent earn better opportunities outside. This occurs if information from the early stage transfers to an external market, as when the market can observe the quality of an intermediate good. To keep the model consistent, here I assume that if the market can condition its outside options on q_1, the agent can also. Therefore, I explore detailed evaluation when the second-stage effort is a function, $e_2(q_1)$.

In general, arbitrary nonlinear functions, $\bar{u}_2(q_1)$, may cross $S^*(q_1)$ many times, implying that the continuation set $X = \{q_1 \mid S^*(q_1) \geq \bar{u}_2(q_1)\}$ may be a union of disjoint intervals. To make the analysis tractable, assume that the continuation surplus function, $S^*(q_1)$, and the outside option function, $\bar{u}_2(q_1)$, satisfy the single crossing property (the functions cross once). In this case, let q^* satisfy $S^*(q^*) = \bar{u}_2(q^*)$, so the two functions cross at q^*. Because of single crossing, the continuation set is simply $X = [q^*, \infty)$.

Even under the single crossing property, arbitrary nonlinear outside option functions can cause second-stage effort to be higher than first-stage effort at the optimum. Intuitively, if $\bar{u}_2(q_1)$ increases rapidly with q_1, small increases in early-stage effort can lead to large increases in second-stage outside opportunities. Therefore, it is efficient for the agent to work hard early on in order to generate these outside opportunities. If, on the other

[9] Unfortunately, it is not possible to generate an exogenous condition on the primitives of the model that details how convex marginal costs can be before the main result reverses.

hand, these outside opportunities increase slowly, it is still efficient to work harder in the second stage. The next proposition gives conditions for the main result to hold under nonconstant outside option functions.

PROPOSITION 4. *Let $S^*(\cdot)$ and $\bar{u}_2(\cdot)$ satisfy single crossing and let $C'''(e) \leq 0$ for all e. If $\bar{u}_2{}'(q_1) < C'(e_1^*)$ for all $q_1 < q^*$, it is efficient to work harder in stage 2, so $e_1^* < \mathbb{E}[e_2^*(q_1) \mid q_1 \geq q^*]$.*

As long as the marginal outside opportunities do not exceed the marginal costs of continuing the project, it is still efficient to work harder in the later stages. For example, if costs are quadratic $(C(e) = e^2)$ and outside options are linear and increase slowly $(\bar{u}_2(q_1) = \gamma + \eta q_1$ for $\gamma > 0, \eta \approx 0)$, then the proposition holds. Of course, it is possible to construct an outside option function that increases so rapidly that it is always efficient to work harder in the early stages. But aside from these manufactured examples, the main result is still robust.

5. *An Agency Model*

This section explores an agency model where the firm can no longer observe the agent's effort. Instead the firm elicits effort through explicit contracts on output. Following the standard assumption in agency models, suppose the agent is now risk averse, while the firm is (still) risk neutral. It is tempting to believe that the difference in risk preferences and the new incentive problem causes the agent to change his dynamic effort allocation, possibly reversing the main result. While his effort choices change from first-best, his dynamic effort allocation is subject to the same forces discussed earlier. In fact, the main result of the paper is robust within an agency context: The agent still works harder in the second stage.

Because the principal can no longer observe effort, it is necessary to formulate a contract on observable output, q_t. Since the principal's final profit is a function of $V(q_1 + q_2)$ and not a function of q_t individually, the optimal contract is also a function of $q_1 + q_2$. Moreover, any general contract, $t(q_1 + q_2)$, can be represented as a flat salary, α_t, in stage t and a nonlinear sharing rule, $\beta(q_1 + q_2)$, on production.[10] If the agent advances to the second stage, his total compensation is $\alpha_1 + \alpha_2 + \beta(q_1 + q_2)V(q_1 + q_2)$, and otherwise he gets only α_1. As before, all parties can fully commit to the contract. Both parties have outside options in each stage, so let $\bar{u}_t^a > 0$ and $\bar{u}_t^p > 0$ denote the outside options in stage t of the agent and principal, respectively.

The principal and agent now have different preferences and different outside options, so they, in general, disagree on whether to continue the project for any given first-stage output level, q_1. Let the firing rule, q_p, denote

[10] To see this, suppose $t(q)$ is a general contract where $q = q_1 + q_2$. Write this as $t(q) = F + T(q)$, decomposed into its fixed and variable components. Let $\alpha_1 + \alpha_2 = F$, and $\beta(q) = T(q)/V(q)$.

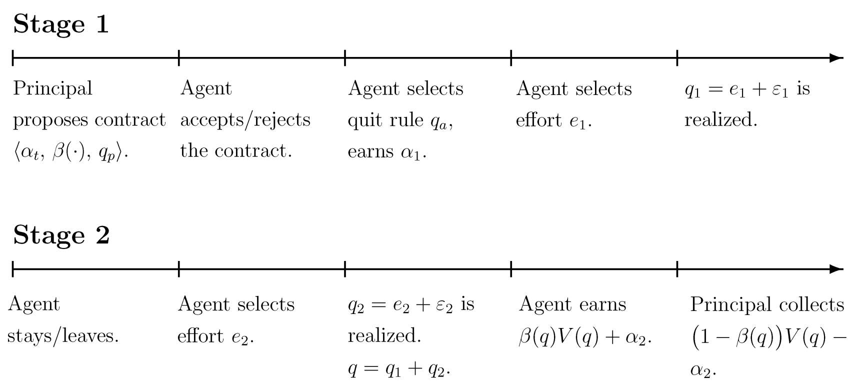

FIG. 2.—Timeline.

the principal's hurdle for continuation: The principal fires the agent if $q_1 < q_p$. Similarly, let the quit rule, q_a, denote the agent's hurdle for continuation: The agent quits if $q_1 < q_a$. I assume that q_a and q_p are set independently and that either party can discontinue the project at will, a feature of at-will employment.[11] Both parties must agree to continue for the project to move forward, so early-stage output, q_1, must clear the hurdle $\tau = \max\{q_a, q_p\}$ to justify continuation. This is simply a consequence of the ability of either party to walk away if they so choose, and doing so dissolves the project. The probability of continuation is now $P_1 = \Pr(q_1 > \tau)$.

Figure 2 denotes the timeline of the game. In stage 1, the principal proposes the contract $(\alpha_t, \beta(\cdot), q_p)$. The agent earns α_1 and selects an effort level, e_1, and a quit rule, q_a. Then $q_1 = e_1 + \varepsilon_1$ is realized. In stage 2, the agent stays if $q_1 \geq \max\{q_a, q_p\}$ and leaves otherwise. If he stays, the agent exerts effort e_2, output $q_2 = e_2 + \varepsilon_2$ is realized, and the project value $V(q_1 + q_2)$ is determined. The principal collects his share of the output, $(1 - \beta(q))V(q) - \alpha_2$, and pays the agent $\beta(q)V(q) + \alpha_2$, where $q = q_1 + q_2$. I adopt the standard assumptions that the agent's utility function is time-separable and separable in wages, w, and effort, e, and that there is no discounting. Thus his final utility is $[u(\alpha_1) - C(e_1)] + \bar{u}_2^a$ if he quits at the end of stage 1 and $[u(\alpha_1) - C(e_1)] + [u(\beta(q)V(q) + \alpha_2) - C(e_2)]$ if he completes the project. Suppose u is an increasing and concave utility function. The agent's continuation utility for each $q_1 > \tau$ is

$$U(q_1, e_2) = \mathbb{E}_2[u(\beta(q_1 + e_2 + \varepsilon_2)V(q_1 + e_2 + \varepsilon_2) + \alpha_2)] - C(e_2).$$

The agent receives this continuation utility for each $q_1 > \tau$, and receives his

[11] For example, in the United States at-will employment is the dominant form of employment in the private sector. Workers can quit firms, and firms can fire workers, subject to two weeks notice.

outside option $\bar{u}_2^a$ otherwise. Therefore, the agent's problem is

$$\max_{e_t, q_a} \int_{\tau}^{\infty} U(q_1, e_2) g(q_1 - e_1)\, dq_1 + (1 - P_1)\bar{u}_2^a - C(e_1) + u(\alpha_1).$$

Similarly, the principal's continuation profit for each $q_1 > \tau$ is

$$\pi(q_1) = \mathbb{E}_2[(1 - \beta(q_1 + e_2 + \varepsilon_2))V(q_1 + e_2 + \varepsilon_2)] - \alpha_2.$$

The principal earns this continuation profit for every $q_1 > \tau$ and captures his outside option, $\bar{u}_t^p$, otherwise, so the principal solves

$$\max_{\alpha_t, \beta, q_p} \int_{\tau}^{\infty} \pi(q_1) g(q_1 - e_1)\, dq_1 + (1 - P_1)\bar{u}_2^p - \alpha_1. \tag{2}$$

The principal and the agent both have individual rationality constraints, so their equilibrium utility and profit levels must exceed the value of their outside options: $(\bar{u}_1^a + \bar{u}_2^a)$ for the agent and $(\bar{u}_1^p + \bar{u}_2^p)$ for the principal.

Before stating the main result for this setting, one additional assumption is required. Recall that the players set their continuation thresholds, q_a and q_p, independently, and the relevant threshold for continuing the project is $\tau = \max\{q_a, q_p\}$. Intuitively, the relative levels of these thresholds move in the same direction as the relative magnitudes of the two players' outside options. To guarantee that the agent still works harder in the second stage than in the first, it must be that the *agent*'s threshold matters (which means that the agent has high outside options relative to the principal). When the agent's threshold is the relevant one, the margin of continuing versus quitting does not contribute to the effort decision, since the agent is indifferent between staying and leaving at the threshold. However, when the principal's target is the relevant one, the agent strictly prefers staying to leaving at the threshold. Thus, he has an additional incentive to work hard in the first period to be able to clear the threshold. This additional effect might cause the agent's first period effort to be higher than that of the second period. Assuming $\tau = q_a$ prevents this from happening.

PROPOSITION 5. *If $q_a \geq q_p$, then a risk-averse agent works harder in the second stage.*

The same forces as in the benchmark case apply here: The possibility of not advancing to the second stage reduces the marginal return to effort in the first stage, causing the agent to shade his effort downward. The difference in risk preferences between the two parties does not change the agent's dynamic effort allocation decision. Risk aversion causes the principal to withhold some incentives from the agent ($\beta < 1$), but it does not change the agent's effort level between periods. Ultimately, this shows that the agent's effort allocation decision depends on the possibility of discontinuing after the first stage and collecting an outside option. It does not depend on differential preferences between the two parties.

The agency model outlined above shows that while incentive problems may change the effort choice *within* each period, they do not qualitatively change the effort choice *between* periods. To be precise, let e_t^* be the first-best effort level selected by the planner in the absence of agency problems. The main theorem earlier in the paper shows that $e_1^* < e_2^*$. Now let $\tilde{e}_t$ be the equilibrium effort level the agent selects in the agency model of this section. This is an action chosen by the agent in a noncooperative game with the principal, where the principal selects a contract and the agent simultaneously selects effort. As in all agency models, risk aversion leads to a distortion in effort levels, so $e_t^* \neq \tilde{e}_t$ for each t. In this sense, incentive problems change the effort choices *within* each stage. However, both $e_1^* < e_2^*$ and $\tilde{e}_1 < \tilde{e}_2$, so agency problems do not qualitatively change the allocation of effort *between* stages. Again, this happens because the sorting effect still operates even when the agent is risk averse and the principal cannot observe his effort and so must induce it through an output contingent contract.

Finally, the main result also holds in an agency model with renegotiation, as long as a renegotiation-proof contract exists. Observe that the agency-model result (Proposition 5) is obtained with no reference to the principal's problem. As long as a time-invariant optimal contract exists, the original proof of Proposition 5 shows that the agent chooses $\tilde{e}_1 < \tilde{e}_2$. Now, in a model with renegotiation, this optimal contract is different from the model with full commitment, and the actual levels of $\tilde{e}_1$ and $\tilde{e}_2$ are different as well. But that is of no consequence here: What matters is that $\tilde{e}_1$ is still lower than $\tilde{e}_2$. The sorting effect still causes the agent to allocate more effort into later stages, for the same reasons outlined earlier: The possibility of termination reduces the agent's marginal return to effort in stage 1, so he works less. Risk aversion and renegotiation do not alter this force.

6. *Ability*

Return now to the original setting without the agency problem, that is, assume that the firm can observe the agent's effort. When projects fail, firms have difficulty separating technology failure from employee failure. Did the software fail because the programmer was incompetent, or because of a negative technology shock? This happens because output measures cannot disentangle the agent's ability from technological uncertainty when assessing the performance of a project. This section models dynamic decision making when ability and uncertainty are impossible to observe separately. I use ability as shorthand for project-specific ability, measuring the quality of the match between the agent and the project.

I extend the benchmark model to include an underlying ability parameter that persists through both stages. Ability, $\tilde{a}$, is distributed according to a prior distribution, $f(\cdot)$, with support A.[12] Since neither the firm nor the agent knows the agent's ability, the uncertainty is symmetric. Suppose that

$$\tilde{q}_t = \tilde{a} + e_t + \tilde{\varepsilon}_t.$$

Clearly, higher ability increases output. More precisely, ability and effort are substitutes, so more able agents can work less to generate the same output as less able agents. Including a persistent ability term induces correlation in output across stages. High ability, which generates high output today, also generates high output tomorrow. This is the main intuition that drives the results of this section.

The posterior mean, $\mathbb{E}[\tilde{a} \mid \tilde{q}_1 = q_1]$, measures all parties' revised estimate of the agent's ability after the first stage. If the posterior mean increases with q_1, then high early output signals high ability. Let $\bar{f}(\cdot \mid \cdot)$ be the posterior density of $\tilde{a}$, given the value of $\tilde{q}_1$. A fairly weak condition that guarantees that $\mathbb{E}[\tilde{a} \mid \tilde{q}_1 = q_1]$ increases with q_1 is the monotone likelihood ratio property of this posterior density.

DEFINITION 2. *Let $\bar{\varphi}(\cdot \mid \cdot)$ be the conditional probability density function of a random variable $\tilde{X}$ given the realization of another random variable $\tilde{Y}$. $\bar{\varphi}$ satisfies the strict monotone likelihood ratio property (MLRP) if and only if the likelihood ratio function $LR_{(x_1,x_0)}(y) \equiv \frac{\bar{\varphi}(x_1 \mid y)}{\bar{\varphi}(x_0 \mid y)}$ is strictly increasing in y for each pair of values $x_1 > x_0$ in the support of $\tilde{X}$.*

It is quite easy to see that MLRP does indeed guarantee that the posterior mean $\mathbb{E}[\tilde{X} \mid \tilde{Y} = y]$ is increasing in y.[13] In fact, the following stronger result holds:

LEMMA 1. *Let $\bar{\varphi}$ satisfy the strict MLRP. If $z(x)$ is a strictly increasing function, then*

$$\int_{\mathrm{supp}(\tilde{X})} z(x)\bar{\varphi}_y(x \mid y)\, dx > 0.$$

Since first-stage output, $\tilde{q}_1$, is the sum of a scalar (e_1), ability $(\tilde{a})$, and an error term $(\tilde{\varepsilon}_1)$, the posterior of $\tilde{a}$ given $\tilde{q}_1 = q_1$ takes a particularly simple form. In this special case, the MLRP reduces to conditions on just the distribution of the noise term:

LEMMA 2. *$\bar{f}(\cdot \mid \cdot)$ satisfies the strict MLRP if and only if the ratio $\frac{g(\varepsilon - \delta)}{g(\varepsilon)}$ is strictly increasing in ε for all positive δ.*[14]

[12] To avoid confusion, in this section I denote random variables by letters with tildes $(\tilde{x})$ and their values by the same letters without tildes (x). Also, conditional distribution functions are denoted by an over bar $(\bar{f}(\cdot \mid \cdot))$, while unconditional distributions have no bar $(f(\cdot))$.

[13] This readily follows from Lemma 1; just choose $z(x) \equiv x$.

[14] It has been noted elsewhere (Mattsson, Voorneveld, and Weibull [2004]) that this property holds if the error distribution is normal or Gumbel, but not if it is multimodal or fat-tailed (such as the t-distribution).

Now, let $S(q_1, e_1, e_2)$ be the continuation surplus function, that is, the expected utility of continuing given that first-stage output is q_1 and the effort plan is e_1 and e_2 (note that e_2 is a *function* in the case of detailed evaluations). This is the same function as before, except now it is necessary to take expectations over $\tilde{a}$. In the case of coarse evaluations, the surplus is

$$S(q_1, e_1, e_2) = \int_A \int_{\mathbb{R}} V(q_1 + a + e_2 + \varepsilon_2) g(\varepsilon_2) \bar{f}(a \mid q_1)\, d\varepsilon_2\, da - C(e_2);$$

in the case of detailed evaluations, it is

$$S(q_1, e_1, e_2(\cdot)) = \int_A \int_{\mathbb{R}} V(q_1 + a + e_2(q_1) + \varepsilon_2) g(\varepsilon_2) \bar{f}(a \mid q_1)\, d\varepsilon_2\, da - C(e_2(q_1)).$$

It is efficient to continue the project if the continuation surplus exceeds the outside option, or if $S^*(q_1) \equiv S(q_1, e_1^*, e_2^*) \geq \bar{u}_2$, where e_1^* and e_2^* are the efficient effort levels. As before, a cutoff strategy is a target q^* that the planner sets, such that the agent continues to work if $q_1 \geq q^*$, but not otherwise. If $S^*(q_1)$ is strictly increasing, then the planner uses a cutoff strategy. It is easy to see that MLRP guarantees this: as Lemma 1 shows, MLRP implies that a higher first-stage output signals higher ability and thus higher expected second-stage output. Since a higher first-stage output also gives a higher direct contribution to the final output, it is clear that high first-stage output is unambiguously good news. Formally, this gives us the following result:

PROPOSITION 6. *Let g satisfy the condition in Lemma 2. Then, it is efficient to use a cutoff strategy under both coarse and detailed evaluations.*

To illustrate this proposition, consider the simple example of a linear production function, $V(q) = Vq$, and suppose evaluations are coarse.

The continuation surplus simplifies to

$$S^*(q_1) = V\big(q_1 + e_2^* + \mathbb{E}[\tilde{a} \mid \tilde{q}_1 = q_1]\big) - C\big(e_2^*\big).$$

Now MLRP guarantees that the posterior mean increases in q_1, or

$$\frac{\partial}{\partial q_1}\mathbb{E}[\tilde{a} \mid \tilde{q}_1 = q_1] = \int_A a \bar{f}_{q_1}(a \mid q_1)\, da > 0.$$

In words, high realizations of output signal that the underlying ability of the agent is high. Therefore

$$S^{*\prime}(q_1) = V\left[1 + \int_A a \bar{f}_{q_1}(a \mid q_1)\, da\right] > 0.$$

That is, the continuation surplus increases in first-stage output. In addition, since $\lim_{q_1 \to -\infty} S^*(q_1) = -\infty$, $\lim_{q_1 \to \infty} S^*(q_1) = \infty$, and S^* is continuous, the Intermediate Value Theorem states that there exists a q^* such that $S^*(q^*) = \bar{u}_2$. Hence the planner uses a cutoff strategy: The project is continued if $q_1 \geq q^*$, or $S(q_1) \geq \bar{u}_2$. The proof of Proposition 6 is a more general version of this example. The entire decision on whether to terminate low output rests on the continuation surplus function, $S^*(q_1)$. As long as this function varies in q_1, it is efficient to terminate some agents. In particular, if the continuation surplus is monotonic, then the decision takes the form of a cutoff.

Introducing persistent ability induces correlation across stages. This correlation makes it worthwhile to cut projects with low output *even if* production is divisible. The posterior mean $\mathbb{E}[\tilde{a} \mid \tilde{q}_1 = q_1]$ increases in q_1 under MLRP. If a high q_1 is observed, all parties learn (in a precise Bayesian sense) that the agent's ability is likely to be high. Because ability persists into the second stage, this most likely results in a high q_2. Therefore, it makes sense to continue the project after a high q_1. It is efficient to retain the agent not only because output is high today, but because it is expected that his ability is high, which is likely to lead to high output tomorrow. This ability-induced correlation across stages is sufficient to guarantee the use of a cutoff strategy.

6.1 EFFICIENT DYNAMIC EFFORT SUPPLY

This section analyzes the efficient supply of effort across stages in the persistent-ability setting. The main result of the paper is robust even after including ability in the model. Specifically, the sorting effect of the efficient target, q^*, biases effort upwards in the second stage.

PROPOSITION 7. *Given persistent ability, the following results hold:*

1. *Under coarse evaluation, it is efficient to work harder in the second stage* ($e_1^* < e_2^*$).
2. *Under detailed evaluation, it is efficient to work harder on average in the second stage* ($e_1^* < \mathbb{E}[e_2^*(q_1) \mid q_1 \geq q^*]$), *provided that* $C'''(e) \leq 0$ *for all* e.

To gain intuition on the problem, suppose that errors are distributed $\tilde{\varepsilon}_t \sim N(0, s^2)$, and the ability parameter is distributed $\tilde{a} \sim N(a_0, t^2)$. Call s^2 the error variance and t^2 the prior variance, that is, the variance on the prior distribution of ability. Calculation shows that the posterior density is normal with moments

$$\mathbb{E}[\tilde{a} \mid \tilde{q}_1 = q_1] = \frac{t^2}{t^2 + s^2}(q_1 - e_1) + \frac{s^2}{t^2 + s^2} a_0;$$

$$\mathrm{Var}(\tilde{a} \mid \tilde{q}_1 = q_1) = \frac{s^2 t^2}{t^2 + s^2}. \tag{3}$$

The posterior mean is a linear function of the output realization, $q_1 - e_1$, and the prior mean, a_0, placing more weight on the term with smaller variance. For example, as the prior variance decreases to zero, the posterior mean converges to the prior mean. So as the available prior information on the agent's ability improves, the posterior places little weight on the observed output realization and more weight on the prior mean. Similarly, as the error variance decreases to zero, the posterior mean converges to the realization $q_1 - e_1$. With a low error variance, the output realization gives an accurate signal of ability, and the posterior mean reflects this. Solving the social planner's problem shows the paper's main result is robust. The same intuition from the main result earlier in the paper holds here: The possibility of halfway termination lowers the marginal return to effort in stage 1, and so it is efficient to shade effort downward in the first stage.

This result gives a view of dynamic labor supply different from the large career concerns literature that began with Hölmstrom [1999], who argues that managers work hard early in their careers to influence perceptions of their ability. These perceptions of high ability translate into higher future wages through a competitive labor market. The key assumption there is that the market cannot observe the manager's effort, and so the market "confuses" effort with ability. More precisely, if the market observes high output, there is positive probability that the high output comes from high ability. So the market updates its posterior distribution on ability in a Bayesian manner. Consequently, managers work hard *knowing* that this updating takes place. Thus the manager supplies high effort early on because the market uses output to make inferences on ability.

In contrast, effort is observable in my model, and this eliminates the career concerns effect, leaving only the sorting effect. Because effort is observable, the firm/market no longer interprets high output for high effort. It knows effort precisely, and so can see through any attempt by the agent to work harder to increase output and thus "fool the market" into believing he has high ability. To see this formally, observe that from equation (3) the posterior mean is a weighted average of the prior mean, a_0, and the output realization, $q_1 - e_1 = a + \varepsilon_1$. Notice that the output realization is not just q_1 but in fact is output less effort. Essentially the planner removes the effect of effort on output and considers only the random component, $a + \varepsilon_1$. Thus the agent has no reason to work hard simply to influence perceptions of his ability, as such perceptions take his effort directly into account.[15] Instead, only the sorting effect remains. Thus, it is efficient to work harder once an agent learns of his higher output and hence higher ability. This arises from sorting, not career concerns.

Outside of Hölmstrom [1999], other career-concerns papers give mixed predictions on dynamic effort allocation. For example, Dewatripont, Jewitt, and Tirole [1999] extend Hölmstrom [1999] to a multitasking framework and indeed find that effort falls over time. Similarly, Ghatak, Morelli, and Sjöström [2001] show that agents work hard early in their careers because of borrowing constraints: They exert high effort to produce high output and accumulate savings that allow them to become entrepreneurs later in their careers and subsequently acquire rent. But some papers find the opposite result, that effort increases over time. Meyer and Vickers [1997] and Gibbons

[15] To see this more intuitively, recall that because effort is observable, the planner's problem is equivalent to a self-employed agent working over time. Clearly such an agent never works harder just to fool himself into thinking he is high ability, as Holmstrom's [1999] manager does with the market. So career concerns are not relevant to a self-employed agent, nor are they relevant for a social planner.

[1987] show that the ratchet effect can cause the principal to increase incentives in later stages if the principal uses early output to infer productivity. Knowing that the principal does this, the agent withholds effort in the early stage to prevent the "ratcheting up" of incentives.

Gibbons and Murphy [1992] build on Hölmstrom [1999] to include both explicit and implicit incentives in a model of career concerns. In this case, agents work hard both because they are paid to do so (explicit incentives) and because the firm updates its prior on their ability, which leads to higher future wages (implicit incentives). The implicit incentives operate exactly as in Hölmstrom's [1999] model: The market observes output but not effort or ability, and uses output to update its perceptions on ability. Gibbons and Murphy [1992] find that optimal explicit incentives rise over time, serving as a substitute instrument for implicit incentives. Yet the total effect on effort allocation is ambiguous, and depends on the parameters of the ability distribution.[16] My model differs from Gibbons and Murphy [1992] for the same reasons it differs from Hölmstrom [1999]: Observability of effort eliminates the signal-jamming career-concerns effect, while outside options and the ability to quit halfway create a sorting effect.[17]

6.2 UP-OR-OUT SCHEMES AS INTERIM PERFORMANCE EVALUATIONS

The model can also explain up-or-out promotion policies. These policies are common in many professional services (accounting, law, consulting, investment banking), as well as in the military and academia. Up-or-out is a particular form of promotion in which employees are fired if they fail a single, major performance evaluation (the decision to grant partner or permanent tenure). They are fired even if their performance is just slightly below the firm's standard. This is puzzling because the firm could always retain the agent at lower wages that reflect his lower productivity.

Previous explanations of up-or-out contracts rely on asymmetric information. Kahn and Huberman [1988] argue that a firm chooses an up-or-out policy to force itself to behave honestly towards the employee, thus maintaining his incentive to work hard and acquire specific human capital. It is critical that the firm observes the agent's productivity but the agent himself does not. Waldman [1990] claims that firms use up-or-out to signal the agent's productivity between firms and thus induce human capital acquisition. The employee's general human capital is private information to his

---

[16] In particular, the effort allocation decision depends on the variance of the ability distribution. If ability is known precisely (low variance), implicit incentives are strong and the agent works hard early in his career in response to these incentives. On the other hand, if ability is not known precisely (high variance), implicit incentives are weak and the agent primarily responds to explicit incentives. Since these explicit incentives rise over time, so does his effort allocation.

[17] Career-concerns models include agency problems as well as outside options that vary with output. The ability model of this section is robust to these extensions, that is, the sorting effect ($e_1 < e_2$) still holds under some additional technical assumptions on the functions involved, which are similar to those in Proposition 4. Details are available from the author upon request.

firm, but the promotion decision of the firm is public information available to all firms. Levin and Tadelis [2002] claim that partnerships use up-or-out to maintain a commitment to high product and employee quality, where the firm observes output (product quality) and employee ability, but the market (including clients) does not.

This paper avoids a complicated asymmetric information argument, and instead uses symmetric information to cast up-or-out schemes as examples of efficient sorting. This stands in stark contrast to previous work, in particular to Kahn and Huberman [1988], who view up-or-out contracts as involuntary layoffs that occur because of information imperfections. Layoffs here are the outcome of an efficient termination decision: Performance evaluations efficiently sort agents into their most productive use, which may be to collect their outside option. It is not necessary to resort to a special and stylized information structure to explain a common form of performance evaluation, the up-or-out scheme.

Why doesn't every firm use up-or-out rules? Firms use them if (1) ability is a key factor of production and (2) both parties have outside options. These conditions are consistent with production in professional services (like accounting firms). These firms recruit highly educated workers because ability is a key input in production. Workers in these occupations enjoy high market wages, so their outside options are high. Firms in which these assumptions do not hold (such as manufacturing) do not employ up-or-out schemes.

7. *Conclusion*

Given that so much of today's production takes place in long-term projects, it is natural to explore the role and consequences of performance evaluations on projects before they finish. By their very definition, projects take time, and firms cannot collect full value from them until they are completed. So interim performance evaluations give firms the option to end projects (or fire employees on existing projects) if the early returns are low. Performance evaluations efficiently sort projects and employees into those that quit and those that stay. A surprising consequence of this sorting is its effect on dynamic resource allocation: It is efficient to allocate more effort towards the later stages.

The logic of the main result runs as follows. Because production is indivisible, firms and agents cannot sell the project after the first stage, and instead use early-stage output to calculate future project value and ultimately the value of continuing. If the continuation value is sufficiently low, both parties prefer to capture their outside options. This is the role of interim performance evaluations. The possibility of termination after the first stage reduces the expected project return in stage 1, and hence lowers the marginal return to effort. Thus the agent invests more of his resources (effort) later in the project. This is the main consequence of interim performance evaluations.

The paper shows that the main result is robust under a variety of permutations. First, the result generalizes beyond the simple two-stage model to the finite stage game, proving that restriction to two stages is without loss of generality. Second, even if the evaluation returns detailed performance information, so that the agent can condition his second-stage effort on first-stage output, the agent efficiently works harder on average in stage 2 than in stage 1. Third, if the outside options vary with output such that the marginal increase in outside options does not exceed the marginal cost of effort, the agent exerts more effort on average in stage 2 than in stage 1. And most importantly, this result holds within an agency framework. With a risk-neutral principal and a risk-averse agent, the agent still shades his effort downward in the early stage. The differential risk preferences do not change the allocation of effort between stages even though they do change the allocation of effort within each stage.

The purpose here is to show that production that takes place over time gives firms the option of ending projects before they finish. Exercising this option and collecting intermediate performance information skews the efficient allocation of resources in surprising ways. The prior agency literature on career concerns gives a mixed verdict on optimal effort allocation; some papers predict increasing effort over time while others predict decreasing effort over time. The aim of my analysis is to argue that sorting delivers a strong prediction that resource allocation increases over time. I present the majority of the analysis in a first-best world to show the main forces at work and highlight the key economic intuition.

APPENDIX

Proof of Proposition 1. Let e_1^* and e_2^* be the efficient effort levels. Let $S^*(q_1) \equiv S(q_1, e_2^*)$, and $q_1^* = e_1^* + \varepsilon_1$.

Since V is strictly increasing, $V' > 0$, and therefore,

$$S^{*\prime}(q_1) = \frac{\partial S\big(q_1, e_2^*\big)}{\partial q_1} = \int_{-\infty}^{\infty} V'\big(q_1 + e_2^* + \varepsilon_2\big) g(\varepsilon_2)\, d\varepsilon_2 > 0.$$

So continuation surplus is strictly increasing and continuous in q_1. Since V is strictly increasing and weakly concave, $\lim_{q\to-\infty} V(q) = -\infty$. Since $\bar{u}_2 > 0$, there exists an x low enough such that $0 < S^*(x) < \bar{u}_2$. Now

$$\begin{aligned}
\mathbb{E}_1 S^*\big(q_1^*\big) - C\big(e_1^*\big) &= \mathbb{E}_1\big[\mathbb{E}_2 V\big(q_1^* + e_2^* + \varepsilon_2\big) - C\big(e_2^*\big)\big] - C\big(e_1^*\big) \\
&= \mathbb{E}\, V\big(q_1^* + q_2^*\big) - C\big(e_2^*\big) - C\big(e_1^*\big) \\
&\geq \bar{u}_1 + \bar{u}_2,
\end{aligned}$$

where the inequality follows from project feasibility. Therefore $\mathbb{E}_1 S^*(q_1^*) > \bar{u}_2$. Consequently, there exists a $y \in \mathbb{R}$ such that $S^*(y) > \bar{u}_2 > S^*(x)$. By the Intermediate Value Theorem, there exists a $q^* \in (x, y)$ such that $S^*(q^*) = \bar{u}_2$.

Now note that $S^*(q_1)$ is precisely the value of continuing the project, given that the planner chooses effort levels optimally. While e_2^* is not necessarily optimal *given* q_1, we must recall that q_1 is unknown under coarse evaluations, so it is impossible to condition effort on it. Consequently, the same effort level, e_2^*, is exerted regardless of q_1; it is impossible to do better than to obtain $S^*(q_1)$ for each realization of q_1. It follows that it is efficient to continue the project if and only if $S^*(q_1) \geq \bar{u}_2$. Because $S^*(q_1)$ is monotonically increasing in q_1 and $S^*(q^*) = \bar{u}_2$, we know that this holds if and only if $q_1 \geq q^*$. ■

Proof of Corollary 1. Define

$$\mathcal{S}_n(e_1, e_2) = \int_{-\infty}^{\infty} S(q_1, e_2)g(q_1 - e_1)\, dq_1 - C(e_1);$$

$$\mathcal{S}_e(e_1, e_2, q) = G(q - e_1)\bar{u}_2 + \int_q^{\infty} S(q_1, e_2)g(q_1 - e_1)\, dq_1 - C(e_1).$$

Under no evaluation, the expected surplus is $\hat{\mathcal{S}}_n = \mathcal{S}_n(\hat{e}_1, \hat{e}_2)$, where $(\hat{e}_1, \hat{e}_2) = \arg\max \mathcal{S}_n(e_1, e_2)$ are effort levels that are efficient under no evaluation.

By Proposition 1, evaluation results in termination of projects with first-stage output below some threshold q. The surplus under evaluation is therefore $\mathcal{S}_e^* = \mathcal{S}_e(e_1^*, e_2^*, q^*)$, where $(e_1^*, e_2^*, q^*) = \arg\max \mathcal{S}_e(e_1, e_2, q)$.

Let $\hat{q}$ be defined by $S(\hat{q}, \hat{e}_2) = \bar{u}_2$. By arguments identical to those in Proposition 1, this is well defined, and $S(q_1, \hat{e}_2) \geq \bar{u}_2$ if and only if $q_1 \geq \hat{q}$.

By definition, $\mathcal{S}_e^* \geq \mathcal{S}_e(\hat{e}_1, \hat{e}_2, \hat{q})$, so that

$$\begin{aligned}\mathcal{S}_e^* - \hat{\mathcal{S}}_n &\geq \left[G(\hat{q} - \hat{e}_1)\bar{u}_2 + \int_{\hat{q}}^{\infty} S(q_1, \hat{e}_2)g(q_1 - \hat{e}_1)\, dq_1 - C(\hat{e}_1)\right] \\ &\quad - \left[\int_{-\infty}^{\infty} S(q_1, \hat{e}_2)g(q_1 - \hat{e}_1)\, dq_1 - C(\hat{e}_1)\right] \\ &= \int_{-\infty}^{\hat{q}} (\bar{u}_2 - S(q_1, \hat{e}_2))g(q_1 - \hat{e}_1)\, dq_1 > 0.\end{aligned}$$

The last inequality follows because $\bar{u}_2 > S(q_1, \hat{e}_2)$ for all $q_1 < \hat{q}$ and $g(\varepsilon) > 0$ for all $\varepsilon \in \mathbb{R}$. ■

Proof of Theorem 1. The planner solves

$$\max_{e_i, q} \int_q^{\infty} S(q_1, e_2)g(q_1 - e_1)\, dq_1 + (1 - P_1)\bar{u}_2 - C(e_1).$$

The first order conditions with respect to q, e_2, e_1 are

$$S^*(q^*) = \bar{u}_2;$$

$$\int_{q^*}^{\infty} \frac{\partial S(q_1, e_2^*)}{\partial e_2} g(q_1 - e_1^*)\, dq_1 = 0;$$

$$C'(e_1^*) = -\int_{q^*}^{\infty} S^*(q_1)g'(q_1 - e_1^*)\, dq_1 - g(q^* - e_1^*)\bar{u}_2,$$

where $S^*(q_1) = S(q_1, e_2^*)$, and $S^{*\prime}(q_1) = \frac{\partial S(q_1, e_2^*)}{\partial q_1}$. Substituting $S^*(q^*) = \bar{u}_2$ into the last equation and integrating by parts gives

$$C'(e_1^*) = \int_{q^*}^{\infty} S^{*\prime}(q_1) g(q_1 - e_1^*)\, dq_1.$$

From the continuation surplus function $S(q_1, e_2) = \mathbb{E}_2 V(q_1 + e_2 + \varepsilon_2) - C(e_2)$,

$$S^{*\prime}(q_1) = \mathbb{E}_2 V'(q_1 + e_2^* + \varepsilon_2);$$

$$\frac{\partial S(q_1, e_2^*)}{\partial e_2} = \mathbb{E}_2 V'(q_1 + e_2^* + \varepsilon_2) - C'(e_2^*).$$

Combining these gives

$$\frac{\partial S(q_1, e_2^*)}{\partial e_2} = S^{*\prime}(q_1) - C'(e_2^*).$$

Integrating both sides and combining with the FOC with respect to e_2 yields

$$0 = \int_{q^*}^{\infty} \frac{\partial S(q_1, e_2^*)}{\partial e_2} g(q_1 - e_1^*)\, dq_1 = \int_{q^*}^{\infty} S^{*\prime}(q_1) g(q_1 - e_1^*)\, dq_1 - P_1 C'(e_2^*),$$

where $P_1 = \Pr(q_1 \geq q^*)$. Now combining with FOC with respect to e_1 gives

$$C'(e_1^*) = \int_{q^*}^{\infty} S^{*\prime}(q_1) g(q_1 - e_1^*)\, dq_1 = P_1 C'(e_2^*) < C'(e_2^*).$$

Since marginal costs are increasing, this means $e_1^* < e_2^*$. ■

Proof of Proposition 2. Let us define, for all $t \in \{1, 2, \ldots, T+1\}$, $Q_t \equiv \sum_{s=1}^{t} q_s$. Observe that the continuation surplus at time t depends only on Q_t and the planned effort levels from period $t+1$ onwards. Therefore, we can write the continuation surplus function in terms of Q_t. Call this new function $\tilde{S}_t$, where

$$\tilde{S}_t(Q_t; e_{t+1}, \ldots, e_{T+1}) = S_t(q_1, q_2, \ldots q_t; e_{t+1}, \ldots e_{T+1}).$$

The proof of this observation is simple backwards induction on t. Clearly, the statement holds for $t = T+1$, where $\tilde{S}_{T+1}(Q_{T+1}) = V(Q_{T+1})$. Furthermore, if

$$S_t(q_1, q_2, \ldots q_t; e_{t+1}, \ldots e_{T+1}) = \tilde{S}_t(Q_t; e_{t+1}, \ldots e_{T+1}),$$

then

$$S_{t-1}(q_1, \ldots, q_{t-1}; e_t, \ldots, e_{T+1})$$
$$= \int_{-\infty}^{\infty} \max\left\{\tilde{S}_t(Q_t; e_{t+1}, \ldots, e_{T+1}), \sum_{s=t+1}^{T+1} \bar{u}_s\right\} g(q_t - e_t)\, dq_t - C(e_t)$$
$$= \int_{-\infty}^{\infty} \max\left\{\tilde{S}_t(Q_{t-1} + e_t + \varepsilon_t; e_{t+1}, \ldots, e_{T+1}), \sum_{s=t+1}^{T+1} \bar{u}_s\right\} g(\varepsilon_t)\, d\varepsilon_t - C(e_t)$$
$$\equiv \tilde{S}_{t-1}(Q_{t-1}; e_t, e_{t+1}, \ldots, e_{T+1}),$$

which completes the inductive step.

Let the efficient effort levels be $e_1^*, e_2^*, \ldots, e_{T+1}^*$. By the statement above, the continuation surplus at the efficient effort levels can be expressed as $S_t^*(Q_t) \equiv \tilde{S}_t(Q_t; e_{t+1}^*, \ldots, e_{T+1}^*)$.

Part 1: Cutoff Rule. Note that the statement to be proved is equivalent to the following: "There exist cutoff points $\{Q_t^*\}_{t=1}^T$ such that, for all t, $S_t^*(Q_t) \geq \sum_{s=t+1}^{T+1} \bar{u}_s$ if and only if $Q_t \geq Q_t^*$." The proof is again by backwards induction on t. I actually prove the following more detailed statement: "For all $t \leq T$, $S^{*\prime}(Q_t) > 0$ for all Q_t. Furthermore, there exist cutoff points $\{Q_t^*\}_{t=1}^T$ such that, for all $t \leq T$, $S_t^*(Q_t^*) = \sum_{s=t+1}^{T+1} \bar{u}_s$ and $S_t^*(Q_t) \geq \sum_{s=t+1}^{T+1} \bar{u}_s$ if and only if $Q_t \geq Q_t^*$."

For the base case, consider $t = T$. Note that

$$S_T^*(Q_T) = \int_{-\infty}^{\infty} V\big(Q_T + e_{T+1}^* + \varepsilon_{T+1}\big) g(\varepsilon_{T+1})\, d\varepsilon_{T+1} - C\big(e_{T+1}^*\big),$$

so that

$$S_T^{*\prime}(Q_T) = \int_{-\infty}^{\infty} V'\big(Q_T + e_{T+1}^* + \varepsilon_{T+1}\big) g(\varepsilon_{T+1})\, d\varepsilon_{T+1} > 0.$$

Thus $S_T^*(Q_T)$ is strictly increasing in Q_T. Note that it is also continuous. Since $\lim_{Q\to-\infty} V(Q) = -\infty$ and $\lim_{Q\to\infty} V(Q) > \bar{u}_{T+1}$ (by feasibility), the Intermediate Value Theorem implies that there exists a unique Q_T^* such that $S_T^*(Q_T^*) = \bar{u}_{T+1}$. Since S_T^* is increasing, this implies that $S_T^*(Q_T) \geq \bar{u}_{T+1}$ if and only if $Q_T \geq Q_T^*$.

Now, on to the inductive step. We know that $S_t^{*\prime}(Q_t) > 0$ for all Q_t, that $S_t^*(Q_t^*) = \sum_{s=t+1}^{T+1} \bar{u}_s$, and that $S_t^*(Q_t) \geq \sum_{s=t+1}^{T+1} \bar{u}_s$ if and only if $Q_t \geq Q_t^*$. We want to prove the equivalent statements for $t-1$. By the induction assumption and the definition of continuation surplus,

$$S_{t-1}^*(Q_{t-1}) = \int_{Q_t^*}^{\infty} S_t^*(Q_t) g\big(Q_t - Q_{t-1} - e_t^*\big)\, dQ_t + \left[G\big(Q_t^* - Q_{t-1} - e_t^*\big) \sum_{s=t+1}^{T+1} \bar{u}_s\right] - C\big(e_t^*\big).$$

Taking the derivative with respect to Q_{t-1}, integrating by parts, and noting that $S_t^*(Q_t^*) = \sum_{s=t+1}^{T+1} \bar{u}_s$ gives

$$S_{t-1}^{*\prime}(Q_{t-1}) = \int_{Q_t^*}^{\infty} S_t^{*\prime}(Q_t) g\big(Q_t - Q_{t-1} - e_t^*\big)\, dQ_t > 0,$$

since $S_t^{*\prime}(Q_t) > 0$ for all Q_t by the induction assumption. Since S_{t-1}^* is continuous, feasibility and the usual Intermediate Value Theorem argument guarantee the existence of Q_{t-1}^* such that $S_{t-1}^*(Q_{t-1}^*) = \sum_{s=t}^{T+1} \bar{u}_s$. Since S_{t-1}^* is increasing, $S_{t-1}^*(Q_{t-1}) \geq \sum_{s=t}^{T+1} \bar{u}_s$ if and only if $Q_{t-1} \geq Q_{t-1}^*$. This completes the inductive step and also the proof of part 1 of Proposition 2.

Part 2: Increasing Effort. Given the result from part 1, the planner's problem is

$$\max_{\{e_t\}_{t=1}^{T+1}} \tilde{S}_0(0; e_1, \ldots, e_{T+1}),$$

where, for all $t \in \{0, 1, \ldots, T\}$,

$$\begin{aligned} &\tilde{S}_t(Q_t; e_{t+1}, \ldots, e_{T+1}) \\ &\quad = \int_{Q_{t+1}^*}^{\infty} \tilde{S}_{t+1}(Q_{t+1}; e_{t+2}, \ldots, e_{T+1}) g(Q_{t+1} - Q_t - e_{t+1})\, dQ_{t+1} \\ &\qquad + \left[G\big(Q_{t+1}^* - Q_t - e_{t+1}\big) \sum_{s=t+2}^{T+1} \bar{u}_s \right] - C(e_{t+1}); \end{aligned} \tag{A1}$$

$$\tilde{S}_{T+1}(Q_{T+1}) = V(Q_{T+1}) \quad \text{and} \quad Q_{T+1}^* = -\infty; \tag{A2}$$

$$(\forall t \in \{1, 2, \ldots, T\}) \left(\tilde{S}_t\big(Q_t^*; e_{t+1}, \ldots, e_{T+1}\big) = \sum_{s=t+1}^{T+1} \bar{u}_s \right). \tag{A3}$$

The FOCs for the efficient effort levels $\{e_t^*\}_{t=1}^{T+1}$ in the planner's problem are

$$(\forall t \in \{1, 2, \ldots T+1\}) \left(\frac{\partial \tilde{S}_0}{\partial e_t}\big(0; e_1^*, \ldots, e_{T+1}^*\big) = 0 \right). \tag{A4}$$

In what follows, I omit the arguments of $\tilde{S}_t$ for better readability; thus, $\tilde{S}_t$ is used as shorthand for $\tilde{S}_t(Q_t; e_{t+1}, \ldots, e_{T+1})$.

Recall that the proof of Proposition 2 began by showing that $\tilde{S}_t$ depends only on Q_t and effort levels from period $t+1$ onwards. Using this observation and taking the derivative of equation (A1), we see that, for all $t \in \{0,$

$1, \ldots, T\}$,

$$\frac{\partial \tilde{S}_t}{\partial e_s} = \begin{cases} 0 & \text{if } s \le t \\ \displaystyle\int_{Q^*_{t+1}}^{\infty} \frac{\partial \tilde{S}_{t+1}}{\partial Q_{t+1}} g(Q_{t+1} - Q_t - e_{t+1})\, dQ_{t+1} & \\ \quad - C'(e_{t+1}) & \text{if } s = t+1 \\ \displaystyle\int_{Q^*_{t+1}}^{\infty} \frac{\partial \tilde{S}_{t+1}}{\partial e_s} g(Q_{t+1} - Q_t - e_{t+1})\, dQ_{t+1} & \text{if } s \ge t+2. \end{cases} \tag{A5}$$

Now, let S^*_t denote $\tilde{S}_t$ evaluated at efficient effort levels, that is, let $S^*_t \equiv \tilde{S}_t(Q_t; e^*_{t+1}, \ldots, e^*_{T+1})$. Applying equation (A5) repeatedly and evaluating at the efficient effort levels shows that

$$\frac{\partial S^*_0}{\partial e_t} = A_t - \alpha_t C'(e^*_t), \tag{A6}$$

where

$$A_t = \int_{Q^*_1}^{\infty} \cdots \int_{Q^*_t}^{\infty} \frac{\partial S^*_t}{\partial Q_t} \prod_{s=1}^{t} g(Q_s - Q_{s-1} - e^*_s)\, dQ_t \ldots dQ_1 \tag{A7}$$

and

$$\alpha_t = \int_{Q^*_1}^{\infty} \cdots \int_{Q^*_{t-1}}^{\infty} \prod_{s=1}^{t-1} g(Q_s - Q_{s-1} - e^*_s)\, dQ_{t-1} \ldots dQ_1. \tag{A8}$$

By differentiating equation (A1), integrating by parts, applying equation (A3) and evaluating at the efficient effort levels, we see that, for all $t \in \{0, 1, \ldots T\}$,

$$\frac{\partial S^*_t}{\partial Q_t} = \int_{Q^*_{t+1}}^{\infty} \frac{\partial S^*_{t+1}}{\partial Q_{t+1}} g(Q_{t+1} - Q_t - e^*_{t+1})\, dQ_{t+1}.$$

Plugging this into equation (A7) shows that $A_t = A_{t+1}$ for all $t \in \{0, 1, \ldots T\}$; that is, the sequence $\{A_t\}_{t=1}^{T}$ is in fact constant.

By equations (A6) and (A4), for all $t \in \{0, 1, \ldots T\}$,

$$A_t - \alpha_t C'(e^*_t) = 0 = A_{t+1} - \alpha_{t+1} C'(e^*_{t+1}).$$

Since $A_t = A_{t+1}$, this reduces to

$$\frac{C'(e^*_{t+1})}{C'(e^*_t)} = \frac{\alpha_t}{\alpha_{t+1}}. \tag{A9}$$

But, by equation (A8),

$$\begin{aligned} \alpha_{t+1} &= \int_{Q^*_1}^{\infty} \cdots \int_{Q^*_{t-1}}^{\infty} (1 - G(Q^*_t - Q_{t-1} - e^*_t)) \prod_{s=1}^{t-1} g(Q_s - Q_{s-1} - e^*_s)\, dQ_{t-1} \ldots dQ_1 \\ &< \int_{Q^*_1}^{\infty} \cdots \int_{Q^*_{t-1}}^{\infty} \prod_{s=1}^{t-1} g(Q_s - Q_{s-1} - e^*_s)\, dQ_{t-1} \ldots dQ_1 = \alpha_t, \end{aligned}$$

where the inequality holds because all terms are positive and $(1 - G(Q_t^* - Q_{t-1} - e_t^*)) < 1$.

Therefore, for all t, $\alpha_{t+1} < \alpha_t$, which (by equation (A9)) implies that $C'(e_{t+1}^*) > C'(e_t^*)$. Since costs are convex, this means that $e_{t+1}^* > e_t^*$, which completes the proof. ■

Proof of Corollary 2. Let $X \equiv \{q_1 : S^*(q_1) \geq \bar{u}_2\}$ be the social planner's continuation set: It is efficient to allow an agent with output q_1 to advance to the second stage if and only if $q_1 \in X$. Let $P_1 = \Pr(X)$. I show that X is in fact an interval.

Continuation surplus is

$$S^*(q_1) = \mathbb{E}_2 V(q_1 + e_2(q_1) + \varepsilon_2) - C(e_2(q_1)).$$

The social planner solves

$$\max_{e_1, e_2(\cdot)} \int_X S^*(q_1) g(q_1 - e_1)\, dq_1 + (1 - P_1)\bar{u}_2 - C(e_1).$$

Maximizing the integral pointwise yields

$$e_2(q_1) \in \arg\max S^*(q_1)$$

for almost every $q_1 \in X$. For each such q_1, the Envelope Theorem implies

$$S^{*\prime}(q_1) = \mathbb{E}_2 V'(q_1 + e_2(q_1) + \varepsilon_2).$$

Since $V' > 0$, $S^{*\prime}(q_1) > 0$. So, the ex post social value of an agent increases in his first-stage output. Since $S^*(\cdot)$ is continuous, there exists a q^* such that $S^*(q^*) = \bar{u}_2$ (by the Intermediate Value Theorem). Since $S^*(\cdot)$ is strictly increasing, this means that $X = \{q_1 : q_1 \geq q^*\}$. ■

Proof of Proposition 3. It was shown in Corollary 2 that the continuation surplus function $S^*(q_1) = S(q_1, e_2^*)$ and its derivative are

$$S^*(q_1) = S\big(q_1, e_2^*\big) = \mathbb{E}_2 V\big(q_1 + e_2^*(q_1) + \varepsilon_2\big) - C\big(e_2^*(q_1)\big); \qquad \text{(A10)}$$

$$S^{*\prime}(q_1) = \mathbb{E}_2 V'\big(q_1 + e_2^*(q_1) + \varepsilon_2\big) \quad \big(\text{for } q_1 \geq q^*\big). \qquad \text{(A11)}$$

Also by Corollary 2, the planner's problem is

$$\max_{e_1, e_2(\cdot), q} \int_q^\infty S(q_1, e_2(q_1)) g(q_1 - e_1)\, dq_1 + (1 - P_1)\bar{u}_2 - C(e_1),$$

where $P_1 = 1 - G(q - e_1)$.

The first-order conditions with respect to q, $e_2(\cdot)$, and e_1 are

$$S^*(q^*) = \bar{u}_2; \qquad \text{(A12)}$$

$$\frac{\partial S\big(q_1, e_2^*(q_1)\big)}{\partial e_2(q_1)} = 0 \quad \big(\text{for } q_1 \geq q^*\big); \qquad \text{(A13)}$$

$$C'\big(e_1^*\big) = -\int_{q^*}^\infty S^*(q_1) g'\big(q_1 - e_1^*\big)\, dq_1 - g\big(q^* - e_1^*\big)\bar{u}_2, \qquad \text{(A14)}$$

where $S^*(q_1) = S(q_1,\ e_2^*(q_1))$. Substituting $S^*(q^*) = \bar{u}_2$ from equation (A12) into equation (A14) and integrating by parts gives

$$C'(e_1) = \int_{q^*}^{\infty} S^{*\prime}(q_1) g(q_1 - e_1)\, dq_1. \tag{A15}$$

In what follows I write e_t for e_t^* for visual clarity. Taking the derivative of equation (A10) with respect to e_2 and combining with equation (A13) shows that (for $q_1 \geq q^*$)

$$\frac{\partial S(q_1,\ e_2(q_1))}{\partial e_2(q_1)} = \mathbb{E}_2 V'(q_1 + e_2(q_1) + \varepsilon_2) - C'(e_2(q_1)) = 0.$$

Combining this with equation (A11) yields (for $q_1 \geq q^*$)

$$S^{*\prime}(q_1) = C'(e_2(q_1)).$$

Multiply both sides by $g(q_1 - e_1)$ and integrate over $[q^*, \infty)$ to get

$$\int_{q^*}^{\infty} S^{*\prime}(q_1) g(q_1 - e_1)\, dq_1 = \int_{q^*}^{\infty} C'\,(e_2(q_1))\, g(q_1 - e_1)\, dq_1.$$

Combining with equation (A15) yields

$$C'(e_1) = \int_{q^*}^{\infty} C'\,(e_2(q_1))\, g(q_1 - e_1)\, dq_1.$$

Divide both sides by P_1 to get

$$\mathbb{E}\big[C'(e_2(q_1)) \,\big|\, q_1 \geq q^*\big] = \frac{C'(e_1)}{P_1} > C'(e_1).$$

When $C''' \leq 0$, Jensen's inequality gives

$$C'\big(\mathbb{E}\big[e_2(q_1) \,\big|\, q_1 \geq q^*\big]\big) \geq \mathbb{E}\big[C'(e_2(q_1)) \,\big|\, q_1 \geq q^*\big] > C'(e_1),$$

and increasing costs imply

$$e_1 < \mathbb{E}\big[e_2(q_1) \,\big|\, q_1 \geq q^*\big]. \qquad \blacksquare$$

Proof of Proposition 4. As before, the continuation surplus function, $S^*(q_1) = S(q_1,\ e_2^*)$, and its derivative are

$$S^*(q_1) = S\big(q_1, e_2^*\big) = \mathbb{E}_2 V\big(q_1 + e_2^*(q_1) + \varepsilon_2\big) - C\big(e_2^*(q_1)\big); \tag{A16}$$

$$S^{*\prime}(q_1) = \mathbb{E}_2 V'\big(q_1 + e_2^*(q_1) + \varepsilon_2\big) \quad \big(\text{for } q_1 \geq q^*\big). \tag{A17}$$

The planner solves

$$\max_{e_1, e_2(\cdot), q} \int_{q}^{\infty} S(q_1, e_2(q_1)) g(q_1 - e_1)\, dq_1 + \int_{-\infty}^{q} \bar{u}_2(q_1) g(q_1 - e_1)\, dq_1 - C(e_1).$$

The first-order conditions with respect to q, $e_2(\cdot)$, and e_1 are

$$S^*(q^*) = \bar{u}_2(q^*); \tag{A18}$$

$$\frac{\partial S(q_1, e_2^*(q_1))}{\partial e_2(q_1)} = 0 \quad (\text{for } q_1 \geq q^*); \tag{A19}$$

$$C'(e_1^*) = -\int_{q^*}^{\infty} S^*(q_1) g'(q_1 - e_1^*)\, dq_1 - \int_{-\infty}^{q^*} \bar{u}_2(q_1) g'(q_1 - e_1^*)\, dq_1, \tag{A20}$$

where $S^*(q_1) = S(q_1,\ e_2^*(q_1))$. Integrating by parts gives

$$\int_{-\infty}^{q^*} \bar{u}_2(q_1) g'(q_1 - e_1)\, dq_1 + \int_{-\infty}^{q^*} \bar{u}_2'(q_1) g(q_1 - e_1)\, dq_1$$
$$= \bar{u}_2(q^*) g(q^* - e_1);$$
$$\int_{q^*}^{\infty} S^*(q_1) g'(q_1 - e_1)\, dq_1 + \int_{q^*}^{\infty} S^{*\prime}(q_1) g(q_1 - e_1)\, dq_1$$
$$= -S^*(q^*) g(q^* - e_1) \overset{(A18)}{=} -\bar{u}_2(q^*) g(q^* - e_1).$$

Combine this with equation (A20) to get

$$C'(e_1^*) = \int_{q^*}^{\infty} S^{*\prime}(q_1) g(q_1 - e_1^*)\, dq_1 + \int_{-\infty}^{q^*} \bar{u}_2'(q_1) g(q_1 - e_1^*)\, dq_1. \tag{A21}$$

In what follows I write e_t for e_t^* for visual clarity. Taking the derivative of equation (A16) with respect to e_2 and combining with equation (A19) shows that (for $q_1 > q^*$)

$$\frac{\partial S(q_1, e_2(q_1))}{\partial e_2(q_1)} = \mathbb{E}_2 V'(q_1 + e_2(q_1) + \varepsilon_2) - C'(e_2(q_1)) = 0.$$

Combining this with equation (A17) yields (for $q_1 \geq q^*$)

$$S^{*\prime}(q_1) = C'(e_2(q_1)).$$

Multiply both sides by $g(q_1 - e_1)$ and integrate over $[q^*, \infty)$ to get

$$\int_{q^*}^{\infty} S^{*\prime}(q_1) g(q_1 - e_1)\, dq_1 = \int_{q^*}^{\infty} C'(e_2(q_1)) g(q_1 - e_1)\, dq_1.$$

Combining with equation (A21) yields

$$C'(e_1) = \int_{q^*}^{\infty} C'(e_2(q_1)) g(q_1 - e_1)\, dq_1 + \int_{-\infty}^{q^*} \bar{u}_2'(q_1) g(q_1 - e_1)\, dq_1. \tag{A22}$$

Now, $\bar{u}_2'(q_1) < C'(e_1^*) \forall q_1 < q^*$. Integrating over $q_1 < q^*$ gives

$$\int_{-\infty}^{q^*} \bar{u}_2'(q_1) g(q_1 - e_1)\, dq_1 < \int_{-\infty}^{q^*} C'(e_1) g(q_1 - e_1)\, dq_1 = (1 - P_1) C'(e_1),$$

where $P_1 = 1 - G(q^* - e_1)$ is the probability of clearing the target. Combining with equation (A22) gives

$$P_1 C'(e_1) < \int_{q^*}^{\infty} C'(e_2(q_1)) g(q_1 - e_1)\, dq_1.$$

Dividing both sides by P_1 gives

$$C'(e_1) < \mathbb{E}\big[C'(e_2(q_1)) \,\big|\, q_1 \geq q^*\big].$$

If $C''' \leq 0$, Jensen's inequality gives

$$C'(e_1) < \mathbb{E}\big[C'(e_2(q_1)) \,\big|\, q_1 \geq q^*\big] \leq C'\big(\mathbb{E}\big[e_2(q_1) \,\big|\, q_1 \geq q^*\big]\big),$$

and increasing costs imply

$$e_1 < \mathbb{E}\big[e_2(q_1) \,\big|\, q_1 \geq q^*\big]. \qquad \blacksquare$$

Proof of Proposition 5. Fix a contract $(\alpha_1, \alpha_2, \beta(\cdot), q_p)$. The agent's continuation utility is

$$U(q_1, e_2) = \mathbb{E}_2[u(\beta(q_1 + e_2 + \varepsilon_2)V(q_1 + e_2 + \varepsilon_2) + \alpha_2)] - C(e_2).$$

The agent solves

$$\max_{e_1, e_2, q_a} \int_{\tau}^{\infty} U(q_1, e_2) g(q_1 - e_1)\, dq_1 + G(\tau - e_1)\bar{u}_2^a - C(e_1) + u(\alpha_1).$$

Let $\tilde{e}_t$ denote the agent's optimal (equilibrium) effort choices. Given $\tau = q_a$, the first-order conditions are

$$U(q_a, \tilde{e}_2) = \bar{u}_2^a; \tag{A23}$$

$$\int_{q_a}^{\infty} \frac{\partial U}{\partial q_1}(q_1, \tilde{e}_2) g(q_1 - \tilde{e}_1)\, dq_1 = C'(\tilde{e}_1); \tag{A24}$$

$$\int_{q_a}^{\infty} \frac{\partial U}{\partial e_2}(q_1, \tilde{e}_2) g(q_1 - \tilde{e}_1)\, dq_1 = 0. \tag{A25}$$

Notice that

$$\frac{\partial U}{\partial q_1}(q_1, e_2) = \frac{\partial U}{\partial e_2}(q_1, e_2) + C'(e_2),$$

implying

$$\begin{aligned} C'(\tilde{e}_1) &= \int_{q_a}^{\infty} \frac{\partial U}{\partial q_1}(q_1, \tilde{e}_2) g(q_1 - \tilde{e}_1)\, dq_1 \\ &= \int_{q_a}^{\infty} \left[\frac{\partial U}{\partial e_2}(q_1, \tilde{e}_2) + C'(\tilde{e}_2)\right] g(q_1 - \tilde{e}_1)\, dq_1 \\ &\overset{(A25)}{=} \int_{q_a}^{\infty} C'(\tilde{e}_2) g(q_1 - \tilde{e}_1)\, dq_1 = [1 - G(q_a - \tilde{e}_1)]C'(\tilde{e}_2) < C'(\tilde{e}_2), \end{aligned}$$

so that if $C'' > 0$, then $\tilde{e}_1 < \tilde{e}_2$. $\blacksquare$

LEMMA 3. *If $\bar{\varphi}$ is everywhere differentiable and everywhere nonzero, then it satisfies the strict MLRP if and only if the function $\hat{L}(x) \equiv \frac{\bar{\varphi}_y(x \mid y)}{\bar{\varphi}(x \mid y)}$ is increasing in x for all y in the support of $\tilde{Y}$.*

Proof of Lemma 3. First note that in this case, $LR_{(x_1,x_0)}(y)$ is everywhere strictly increasing in y if and only if its first derivative is everywhere positive. Thus:

$$(MLRP) \Leftrightarrow (\forall x_1 > x_0; \forall y)\left(\frac{\partial}{\partial y}\left(\frac{\bar{\varphi}(x_1 \mid y)}{\bar{\varphi}(x_0 \mid y)}\right) > 0\right)$$

$$\Leftrightarrow (\forall x_1 > x_0; \forall y)\left(\frac{\bar{\varphi}_y(x_1 \mid y)\bar{\varphi}(x_0 \mid y) - \bar{\varphi}(x_1 \mid y)\bar{\varphi}_y(x_0 \mid y)}{\bar{\varphi}(x_0 \mid y)^2} > 0\right)$$

$$\Leftrightarrow (\forall y; \forall x_1 > x_0)\left(\frac{\bar{\varphi}_y(x_1 \mid y)}{\bar{\varphi}(x_1 \mid y)} > \frac{\bar{\varphi}_y(x_0 \mid y)}{\bar{\varphi}(x_0 \mid y)}\right). \qquad \blacksquare$$

Proof of Lemma 1. By Lemma 3 above, the function

$$\hat{L}(x) \equiv \frac{\bar{\varphi}_y(x \mid y)}{\bar{\varphi}(x \mid y)}$$

is strictly increasing for all y in the support of $\tilde{Y}$.

By Bayes's Rule,

$$\bar{\varphi}(x \mid y) = \frac{\bar{\xi}(y \mid x)\varphi(x)}{\displaystyle\int_S \bar{\xi}(y \mid x')\varphi(x')\,dx'},$$

where φ denotes the unconditional density function of $\tilde{X}$, $\bar{\xi}$ denotes the conditional density of $\tilde{Y}$ given the realization of $\tilde{X}$, and $S = \text{supp}(\tilde{X})$.

Now

$$\bar{\varphi}_y(x \mid y)$$

$$= \frac{\bar{\xi}'(y \mid x)\varphi(x)\displaystyle\int_S \bar{\xi}(y \mid x')\varphi(x')\,dx' - \bar{\xi}(y \mid x)\varphi(x)\int_S \bar{\xi}'(y \mid x')\varphi(x')\,dx'}{\left(\displaystyle\int_S \bar{\xi}(y \mid x')\varphi(x')\,dx'\right)^2}.$$

Integrating over S gives

$$\int_S \bar{\varphi}_y(x \mid y)\,dx = 0. \tag{A26}$$

So $\bar{\varphi}_y$ assumes both positive and negative values, and therefore so does $\hat{L}$. Let $\underline{x}$, $\bar{x}$ be the lower and upper limits of S, respectively (they may be $\pm\infty$). By (MLRP), $\hat{L}$ is strictly increasing, so there exists an x^* such that $\{x : \hat{L}(x) < 0\} = (\underline{x}, x^*)$ and $\{x : \hat{L}(x) > 0\} = (x^*, \bar{x})$. By definition of $\hat{L}(x)$, $\bar{\varphi}_y > 0$ if and only if $\hat{L}(x) > 0$. Hence $\{x : \bar{\varphi}_y(x \mid y) < 0\} = (\underline{x}, x^*)$ and $\{x : \bar{\varphi}_y(x \mid y) > 0\} = (x^*, \bar{x})$. Rewrite equation (A26) as

$$\int_{x^*}^{\bar{x}} \bar{\varphi}_y(x \mid y)\,dx = -\int_{\underline{x}}^{x^*} \bar{\varphi}_y(x \mid y)\,dx. \tag{A27}$$

Now, $\bar{\varphi}_y(x \mid y) < 0$ if $x < x^*$, so $|\bar{\varphi}_y(x \mid y)| = -\bar{\varphi}_y(x \mid y)$ for $x < x^*$. Integrate both sides over $(\underline{x}, x^*)$ and combine with equation (A27) to get

$$\int_{\underline{x}}^{x^*} |\bar{\varphi}_y(x \mid y)| dx = \int_{x^*}^{\bar{x}} \bar{\varphi}_y(x \mid y)\, dx. \tag{A28}$$

Now z is strictly increasing, so

$$\int_{\underline{x}}^{x^*} z(x)|\bar{\varphi}_y(x \mid y)| dx < \int_{\underline{x}}^{x^*} z(x^*)|\bar{\varphi}_y(x \mid y)| dx$$
$$= \int_{x^*}^{\bar{x}} z(x^*)\bar{\varphi}_y(x \mid y)\, dx < \int_{x^*}^{\bar{x}} z(x)\bar{\varphi}_y(x \mid y)\, dx.$$

Taking the left-hand side over to the right gives

$$0 < \int_{x^*}^{\bar{x}} z(x)\bar{\varphi}_y(x \mid y)\, dx - \int_{\underline{x}}^{x^*} z(x)|\bar{\varphi}_y(x \mid y)| dx$$
$$= \int_{x^*}^{\bar{x}} z(x)\bar{\varphi}_y(x \mid y)\, dx + \int_{\underline{x}}^{x^*} z(x)\bar{\varphi}_y(x \mid y)\, dx = \int_S z(x)\bar{\varphi}_y(x \mid y)\, dx. \quad \blacksquare$$

Proof of Lemma 2. Let Φ and ϕ be the unconditional cumulative and probability density function, respectively, of $\tilde{q}_1$. Let $\bar{\Phi}$ and $\bar{\phi}$ be the conditional cumulative and probability density function, respectively, of $\tilde{q}_1$, given $\tilde{a}$.

First, observe that

$$\phi(q_1) = \int_A \bar{\phi}(q_1 \mid a)\, f(a)\, da \quad \text{(by definition);} \tag{A29}$$

$$\bar{\Phi}(q_1 \mid a) = \Pr(\tilde{a} + e_1 + \tilde{\varepsilon}_1 < q_1 \mid \tilde{a} = a)$$
$$= G(q_1 - e_1 - a) \qquad \text{(by definition);} \tag{A30}$$

$$\bar{\phi}(q_1 \mid a) = g(q_1 - e_1 - a) \qquad \text{(by equation A30);} \tag{A31}$$

$$\bar{f}(a \mid q_1) = \frac{\bar{\phi}(q_1 \mid a)\, f(a)}{\phi(q_1)} \qquad \text{(by Bayes's Rule).} \tag{A32}$$

By equations (A29) through (A32),

$$\bar{f}(a \mid q_1) = \frac{g(q_1 - e_1 - a)\, f(a)}{\int_A g(q_1 - e_1 - a')\, f(a')\, da'}. \tag{A33}$$

Now,

$$(MLRP) \overset{\text{by def.}}{\Leftrightarrow}$$

$$(\forall a_1 > a_0; \forall q_1 \in \mathbb{R}; \forall \delta > 0) \left(\frac{\bar{f}(a_1 \mid q_1)}{\bar{f}(a_0 \mid q_1)} > \frac{\bar{f}(a_1 \mid q_1 - \delta)}{\bar{f}(a_0 \mid q_1 - \delta)} \right) \overset{\text{by (A37)}}{\Leftrightarrow}$$

$$(\forall a_1 > a_0; \forall q_1 \in \mathbb{R}; \forall \delta > 0)$$

$$\times \left(\frac{g(q_1 - e_1 - a_1) f(a_1)}{g(q_1 - e_1 - a_0) f(a_0)} > \frac{g(q_1 - \delta - e_1 - a_1) f(a_1)}{g(q_1 - \delta - e_1 - a_0) f(a_0)} \right) \overset{\varepsilon_i = q_1 - e_1 - a_i}{\Leftrightarrow}$$

$$(\forall \varepsilon_0 > \varepsilon_1; \forall \delta > 0) \left(\frac{g(\varepsilon_0 - \delta)}{g(\varepsilon_0)} > \frac{g(\varepsilon_1 - \delta)}{g(\varepsilon_1)} \right). \quad \blacksquare$$

Proof of Proposition 6. The continuation surplus is

$$S(q_1, e_1, e_2) = \int_A \int_{\mathbb{R}} V(q_1 + a + e_2 + \varepsilon_2) g(\varepsilon_2) \bar{f}(a \mid q_1)\, d\varepsilon_2\, da - C(e_2).$$

Note that while e_1 is a scalar under both coarse and detailed evaluations, e_2 is a scalar under coarse evaluations, but a nonconstant *function* of q_1 under detailed evaluations. Let the efficient effort schedule be (e_1^*, e_2^*). Let $S^*(q_1) \equiv S(q_1, e_1^*, e_2^*)$.

Let $X \equiv \{q_1 : S^*(q_1) \geq \bar{u}_2\}$ be the planner's continuation set. That is, the agent advances to the second stage if and only if $q_1 \in X$. Then the planner's problem is

$$\max_{e_1, e_2} \int_X S(q_1, e_1, e_2) \phi(q_1)\, dq_1 + (1 - \Pr(X)) \bar{u}_2 - C(e_1),$$

where ϕ is as defined in the proof of Lemma 2. Again, note that e_2 is a *function* in the case of detailed evaluations.

Next, note that

$$\frac{\partial}{\partial q_1} S(q_1, e_1, e_2) = W(q_1, e_1, e_2) + Y(q_1, e_1, e_2), \tag{A34}$$

where

$$W(q_1, e_1, e_2) = \int_A \int_{\mathbb{R}} V'(q_1 + a + e_2 + \varepsilon_2) g(\varepsilon_2) \bar{f}(a \mid q_1)\, d\varepsilon_2\, da; \tag{A35}$$

$$Y(q_1, e_1, e_2) = \int_A \int_{\mathbb{R}} V(q_1 + a + e_2 + \varepsilon_2) g(\varepsilon_2) \bar{f}_{q_1}(a \mid q_1)\, d\varepsilon_2\, da. \tag{A36}$$

Since $V' > 0$, $W(q_1, e_1, e_2)$ is everywhere positive. Furthermore, $V' > 0$ also implies that the function

$$z(a) \equiv \int_{\mathbb{R}} V(q_1 + a + e_2 + \varepsilon_2) g(\varepsilon_2)\, d\varepsilon_2$$

increases in a. By Lemma 1, this means

$$Y(q_1, e_1, e_2) = \int_A z(a)\, \bar{f}_{q_1}(a \mid q_1)\, da > 0.$$

Thus $\frac{\partial}{\partial q_1} S(q_1, e_1, e_2) = W(q_1, e_1, e_2) + Y(q_1, e_1, e_2) > 0$.

In the case of coarse evaluations, it immediately follows that $S^{*\prime}(q_1) = \frac{\partial}{\partial q_1} S(q_1, e_1^*, e_2^*) > 0$.

In the case of detailed evaluations, the conclusion is not immediate, because $S^*(q_1)$ also depends on $e_2^*(q_1)$, which in turn depends on q_1. However, the planner's problem shows that optimality requires that $e_2^*(q_1) = \arg\max_{e_2} S(q_1, e_1^*, e_2)$ for almost all q_1. Therefore, by the Envelope Theorem, even with detailed evaluations it is still true that $S^{*\prime}(q_1) = \frac{\partial}{\partial q_1} S(q_1, e_1^*, e_2^*) > 0$.

Now $S^{*\prime}(q_1) > 0$ with both coarse and detailed evaluations. Clearly S^* is continuous in both cases. By the Intermediate Value Theorem, there exists a q^* such that $S^*(q^*) = \bar{u}_2$. Hence $X = \{q_1 : q_1 \geq q^*\}$. ■

Proof of Proposition 7. By Proposition 6, the planner's problem is

$$\max_{e_1, e_2, q} \int_q^\infty S(q_1, e_1, e_2)\phi(q_1)\, dq_1 + \Phi(q)\bar{u}_2 - C(e_1),$$

where

$$S(q_1, e_1, e_2) = \int_A \int_{\mathbb{R}} V(q_1 + a + e_2 + \varepsilon_2) g(\varepsilon_2) \bar{f}(a \mid q_1)\, d\varepsilon_2\, da - C(e_2) \quad \text{(A37)}$$

and Φ and ϕ are as defined in the proof of Lemma 2. Note that e_2 is a *function* of q_1 in the case of detailed evaluations. Let the efficient argument values be e_1^*, e_2^*, and q^*. Denote $S^*(q_1) \equiv S(q_1, e_1^*, e_2^*)$.

First, it is immediate from Proposition 6 (and can also be seen by taking the first-order condition with respect to q in the planner's problem) that

$$S^*(q^*) = \bar{u}_2. \quad \text{(A38)}$$

The first-order condition with respect to e_1 (under both coarse and detailed evaluations) is

$$C'(e_1^*) = T_1 + T_2 + T_3,$$

where

$$T_1 = \int_{q^*}^\infty \frac{\partial}{\partial e_1} S^*(q_1)\phi(q_1)\, dq_1; \quad \text{(A39)}$$

$$T_2 = \int_{q^*}^\infty S^*(q_1) \frac{\partial}{\partial e_1}\phi(q_1)\, dq_1; \quad \text{(A40)}$$

$$T_3 = \frac{\partial}{\partial e_1}\Phi(q^*)\bar{u}_2. \quad \text{(A41)}$$

Now,

$$\begin{aligned}\frac{\partial}{\partial e_1} S^*(q_1) &\stackrel{(A37)}{=} \int_A \int_{\mathbb{R}} V\big(q_1 + a + e_2^* + \varepsilon_2\big) g(\varepsilon_2) \frac{\partial}{\partial e_1} \bar{f}(a \mid q_1)\, d\varepsilon_2\, da \\ &\stackrel{(A33)}{=} -\int_A \int_{\mathbb{R}} V\big(q_1 + a + e_2^* + \varepsilon_2\big) g(\varepsilon_2)\, \bar{f}_{q_1}(a \mid q_1)\, d\varepsilon_2\, da \\ &\stackrel{(A36)}{=} -Y\big(q_1, e_1^*, e_2^*\big),\end{aligned}$$

so that

$$T_1 = -\int_{q^*}^{\infty} Y\big(q_1, e_1^*, e_2^*\big)\phi(q_1)\, dq_1. \tag{A42}$$

Inserting equation (A31) into equation (A29) and taking derivatives, we see that

$$\frac{\partial}{\partial e_1}\Phi(q_1) = -\phi(q_1) \quad \text{and} \quad \frac{\partial}{\partial e_1}\phi(q_1) = -\phi'(q_1). \tag{A43}$$

Plugging equation (A43) into equations (A40) and (A41), integrating by parts and applying equation (A38) gives

$$T_2 + T_3 = \int_{q^*}^{\infty} S^{*\prime}(q_1)\phi(q_1)\, dq_1. \tag{A44}$$

In the proof of Proposition 6, under both coarse and detailed evaluations,

$$S^{*\prime}(q_1) = \frac{\partial}{\partial q_1} S\big(q_1, e_1^*, e_2^*\big) = W\big(q_1, e_1^*, e_2^*\big) + Y\big(q_1, e_1^*, e_2^*\big),$$

where the functions W and Y are as given in equations (A35) and (A36).

Plugging this into equation (A44) and adding to equation (A42), we obtain

$$C'\big(e_1^*\big) = \int_{q^*}^{\infty} W\big(q_1, e_1^*, e_2^*\big)\phi(q_1)\, dq_1. \tag{A45}$$

To complete the proof of the proposition, we need to consider the cases of coarse and detailed evaluations separately.

Part 1: Coarse Evaluations. Under coarse evaluations, the first-order condition with respect to e_2 is

$$\int_{q^*}^{\infty} \frac{\partial}{\partial e_2} S^*(q_2)\phi(q_1)\, dq_1 = 0.$$

Expanding this using equation (A37) and applying equation (A35) gives

$$\int_{q^*}^{\infty} \big[W\big(q_1, e_1^*, e_2^*\big) - C'\big(e_2^*\big)\big]\phi(q_1)\, dq_1 = 0.$$

Together with equation (A45), this becomes

$$C'\big(e_1^*\big) = (1 - \Phi(q^*))\, C'\big(e_2^*\big) < C'\big(e_2^*\big).$$

Increasing costs now imply $e_1 < e_2$. ■

Part 2: Detailed Evaluations. Under detailed evaluations, optimizing over $e_2(\cdot)$ yields, for almost all $q_1 \geq q^*$,

$$\frac{\partial}{\partial e_2} S\big(q_1, e_1^*, e_2^*\big) = 0.$$

Expanding this using equation (A37), the condition becomes (for almost all $q_1 \geq q^*$)

$$C'\big(e_2^*(q_1)\big)$$
$$= \int_A \int_{\mathbb{R}} V'(q_1 + a + e_2^* + \varepsilon_2)\, g(\varepsilon_2)\, \bar{f}(a \mid q_1)\, d\varepsilon_2\, da \overset{(A35)}{=} W\big(q_1, e_1^*, e_2^*\big).$$

Multiplying both sides by $\phi(q_1)$ and integrating over $[q^*, \infty)$ gives

$$\int_{q^*}^{\infty} C'\big(e_2^*(q_1)\big)\phi(q_1)\, dq_1 = \int_{q^*}^{\infty} W\big(q_1, e_1^*, e_2^*\big)\phi(q_1)\, dq_1 \overset{(A45)}{=} C'\big(e_1^*\big).$$

Dividing both sides by $1 - \Phi(q^*)$ gives

$$\mathbb{E}\big[C'\big(e_2^*(q_1)\big) \,\big|\, q_1 \geq q^*)\big] = \frac{C'(e_1)}{1 - \Phi(q^*)} > C'(e_1).$$

When $C''' \leq 0$, Jensen's inequality gives

$$C'\big(\mathbb{E}\big[e_2(q_1) \,\big|\, q_1 \geq q^*\big]\big) \geq \mathbb{E}\big[C'(e_2(q_1)) \,\big|\, q_1 \geq q^*\big] > C'(e_1),$$

and increasing costs imply

$$e_1 < \mathbb{E}\big[e_2(q_1) \,\big|\, q_1 \geq q^*\big].$$ ■

REFERENCES

BAIMAN, S. "Agency Research in Managerial Accounting: A Survey." *Journal of Accounting Literature* 1 (1982): 154–213.

BAIMAN, S. "Agency Research in Managerial Accounting: A Second Look." *Accounting, Organizations and Society* 15 (1990): 341–71.

BARRY, E. J.; T. MUKHOPADHYAY; AND S. A. SLAUGHTER. "Software Project Duration and Effort: An Empirical Study." *Information Technology and Management* 3 (2002): 113–36.

BLUNDELL, R., AND T. MACURDY. "Labor Supply: A Review of Alternative Approaches." *Handbook of Labor Economics* 3 (1999): 1559–89.

CHRISTENSEN, P. O.; G. A. FELTHAM; AND F. ŞABAC. "Correlated Performance Measures in Dynamic Incentives with Limited Commitment." Working paper, University of British Columbia, 2003a.

CHRISTENSEN, P. O.; G. A. FELTHAM; AND F. ŞABAC. "Dynamic Incentives and Responsibility Accounting: A Comment." *Journal of Accounting and Economics* 35 (2003b): 423–36.

CHRISTENSEN, P. O.; G. A. FELTHAM; C. HOFMANN; AND F. ŞABAC. "Timeliness, Accuracy, and Relevance in Dynamic Incentive Contracts." Working paper, University of British Columbia, 2003.

DEWATRIPONT, M.; I. JEWITT; AND J. TIROLE. "The Economics of Career Concerns, Part 1: Comparing Information Structures." *Review of Economic Studies* 66 (1999): 183–98.

FELTHAM, G.; R. INDJEJIKIAN; AND D. NANDA. "Dual Purpose Measures." Working paper, University of British Columbia, 2003.

FUDENBERG, D.; B. HOLMSTROM; AND P. MILGROM. "Short-Term Contracts and Long-Term Agency Relationships." *Journal of Economic Theory* 51 (1990): 1–31.

GENUCHTEN, M. V. "Why Is Software Late? An Empirical Study of Reasons for Delay in Software Development." *IEEE Transactions on Software Engineering* 17 (1991): 582–90.

GHATAK, M.; M. MORELLI; AND T. SJÖSTRÖM. "Occupational Choice and Dynamic Incentives." *Review of Economic Studies* 68 (2001): 781–810.

GIBBONS, R. "Piece-Rate Incentive Schemes." *Journal of Labor Economics* 5 (1987): 413–29.

GIBBONS, R., AND K. J. MURPHY. "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence." *Journal of Political Economy* 100 (1992): 468–505.

GROSSMAN, G. M., AND C. SHAPIRO. "Optimal Dynamic R&D Programs." *The RAND Journal of Economics* 17 (1986): 581–93.

HARBAUGH, R., AND T. KLUMPP. "Early Round Upsets and Championship Blowouts." *Economic Inquiry* 43 (2005): 316–29.

HECKMAN, J. J. "Life Cycle Consumption and Labor Supply: An Explanation of the Relationship Between Income and Consumption Over the Life Cycle." *The American Economic Review* 64 (1974): 188–94.

HECKMAN, J. J. "What Has Been Learned About Labor Supply in the Past Twenty Years?" *The American Economic Review* 83 (1993): 116–21.

HECKMAN, J. J., AND T. E. MACURDY. "A Life Cycle Model of Female Labour Supply." *Review of Economic Studies* 47 (1980): 47–74.

HÖLMSTROM, B. "Managerial Incentive Problems: A Dynamic Perspective." *Review of Economic Studies* 66 (1999): 169–82. Originally published in "Essays in Economics and Management in Honor of Lars Wahlbeck." Helsinki: Swedish School of Economics, 1982.

IMAI, S., AND M. P. KEANE. "Intertemporal Labor Supply and Human Capital Accumulation." *International Economic Review* 45 (2004): 601–41.

INDJEJIKIAN, R. J. "Performance Evaluation and Compensation Research: An Agency Perspective." *Accounting Horizons* 13 (1999): 147–57.

INDJEJIKIAN, R. J., AND D. NANDA. "Dynamic Incentives and Responsibility Accounting." *Journal of Accounting & Economics* 27 (1999): 177–201.

KAHN, C., AND G. HUBERMAN. "Two-Sided Uncertainty and "Up-or-Out" Contracts." *Journal of Labor Economics* 6 (1988): 423–43.

KAMIEN, M. I., AND N. L. SCHWARTZ. "Expenditure Patterns for Risky R and D Projects." *Journal of Applied Probability* 8 (1971): 60–73.

LAMBERT, R. A. "Income Smoothing as Rational Equilibrium Behavior." *The Accounting Review* 59 (1984): 604–18.

LAMBERT, R. A. "Contracting Theory and Accounting." *Journal of Accounting & Economics* 32 (2001): 3–87.

LAZEAR, E. P. "Output-Based Pay: Incentives, Retention or Sorting?" IZA Discussion Paper 761, Institute for the Study of Labor (IZA), Bonn, Germany, 2003.

LAZEAR, E. P., AND S. ROSEN. "Rank-Order Tournaments as Optimum Labor Contracts." *The Journal of Political Economy* 89 (1981): 841–64.

LEVIN, J. D., AND S. TADELIS. "A Theory of Partnerships." Stanford Law and Economics Olin Working Paper 244, Stanford University, Stanford, CA, 2002.

LEWIS, T. R. "Reputation and Contractual Performance in Long-Term Projects." *RAND Journal of Economics* 17 (1986): 141–57.

LIZZERI, A.; M. A. MEYER; AND N. PERSICO. "The Incentive Effects of Interim Performance Evaluations." Penn CARESS Working Paper 02-09, University of Pennsylvania, Philadelphia, PA, 2002.

LUSS, H. "Optimization of a Multiperiod Resource Allocation Model." *Operational Research Quarterly* 26 (1975): 83–90.

MACURDY, T. "An Empirical Model of Labor Supply in a Life-Cycle Setting." *Journal of Political Economy* 89 (1981): 1059–85.

MANSFIELD, E.; J. SCHNEE; AND S. WAGNER. "Overruns and Errors in Estimating Development Cost, Time, and Outcome," in *Innovation, Technology and the Economy: The Selected Essays of Edwin Mansfield*, Volume 1. Brookfield, UK: Edward Elgar Publishing Limited, 1995: 86–123.

MARSHALL, A. W., AND W. H. MECKLING. "Predictability of the Costs, Time, and Success of Development," in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, edited by R. Nelson. Princeton, NJ: Princeton University Press, 1962: 461–75.

MATTSSON, L. G.; M. VOORNEVELD; AND J. W. WEIBULL. "Information-Constrained Discrete Choice." Working Paper 558, Stockholm School of Economics/EFI, 2004.

MEYER, M. A., AND J. VICKERS. "Performance Comparisons and Dynamic Incentives." *Journal of Political Economy* 105 (1997): 547–81.

O'DONOGHUE, T., AND M. RABIN. "Doing It Now or Later." *The American Economic Review* 89 (1999): 103–24.

PRENDERGAST, C. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37 (1999): 7–63.

PRENDERGAST, C. "The Tenuous Trade-off between Risk and Incentives." *Journal of Political Economy* 110 (2002): 1071–102.

RAY, K. "The Retention Effect of Withholding Performance Information." Working paper, University of Chicago Graduate School of Business, 2006.

ROBERTS, K., AND M. L. WEITZMAN. "Funding Criteria for Research, Development, and Exploration Projects." *Econometrica* 49 (1981): 1261–88.

TOXVAERD, F. "A Theory of Optimal Deadlines." *Journal of Economic Dynamics & Control* 31 (2007): 493–513.

WALDMAN, M. "Up-or-Out Contracts: A Signaling Perspective." *Journal of Labor Economics* 8 (1990): 230–50.