 DOI: 10.1111/j.1539-6924.2011.01736.x

Shortcuts in Complex Engineering Systems: A Principal-Agent Approach to Risk Management

Russ Garber* **and Elisabeth Paté-Cornell**

In this article, we examine the effects of shortcuts in the development of engineered systems through a principal-agent model. We find that occurrences of illicit shortcuts are closely related to the incentive structure and to the level of effort that the agent is willing to expend from the beginning of the project to remain on schedule. Using a probabilistic risk analysis to determine the risks of system failure from these shortcuts, we show how a principal can choose optimal settings (payments, penalties, and inspections) that can deter an agent from cutting corners and maximize the principal's value through increased agent effort. We analyze the problem for an agent with limited liability. We consider first the case where he is risk neutral; we then include the case where he is risk averse.

KEY WORDS: Corner cutting/shortcuts; incentive structure; principal-agent formulation; probalistic risk analysis; project management

1. INTRODUCTION AND BACKGROUND

Shortcuts in the development and operations of complex engineered systems can present serious problems and cause costly system failures. Examples include the Westray coal mine explosion in Nova Scotia(1) and a shutdown of the San Francisco Bay Bridge in 2005 because of the use of inferior, recycled concrete. Real problems arise when an agent causes undue risks of system failure by taking shortcuts that higher managers are unaware of and would not accept.

To understand the problem and prevent an agent from cutting corners, managers have to realize how their policies affect their employees' actions. An example of this is emphasized in the following excerpt from the congressional hearings following the failure of the shuttle Columbia:

> NASA Administrator Sean O'Keefe explained how a culture of denial forms from pressure to show progress within approved budgets that prove unrealistic using approved tools and methods people like. People gradually begin ignoring requirements as mere guidelines, because managers are driven to cut corners to show the illusion of progress in order to maintain program funding.(2)

This article analyzes the problem of corner cutting from a manager's point of view[1] through a combination of game theory(3–5) and systems analysis. The principal-agent theory with hidden actions(6,7) provides a framework for this analysis. Hidden actions (here, to the principal) include the level of effort that the agent is willing to expend to complete the project on schedule, and his decision to take shortcuts in the development phase to remain on schedule if he falls behind. In this model, the principal reaps the benefits of the system's operation and the agent is responsible for developing the system or part of it. Some risks of system failure depend on the

Management Science and Engineering, Stanford University, CA, USA.

*Address correspondence to Russ Garber, tel: 650-387-1590; rgarber@alumni.stanford.edu.

[1] The model can also be extended, with appropriate modifications, to the decisions of government regulators.

 0272-4332/11/0100-0001$22.00/1

agent's decision to follow strictly the procedures, or to cut corners if he is late. Probabilistic risk analysis (PRA)[8–10] is used in this article to determine the probability of system failure, with or without shortcuts.

The problems that can occur in project management have been studied extensively in their generality,[11,12] and more specifically, the dynamics of systems development and the challenges that can occur in that phase.[13,14] We employ a more static analysis and do not use feedback loops to update the decisions of both the principal and the agent. For example, we assume that the agent chooses his effort level at the start of the project and does not change it. There may be, however, situations where the agent might alter his effort level depending on the project's progress. In the same way, the principal may decide to inspect the agent's progress at multiple points and reevaluate her parameter settings as the project progresses. These dynamic aspects can be introduced in the model presented here. Our objective here is to present the basis for a more complex analysis if needed.

The time frame of the project is an important factor. Some failures in operation that can be traced back to employees' actions with the potential for severe punishments (including jail sentences) may occur far into the future. Therefore, these penalties are discounted by the agent and may not constitute an effective deterrent to shortcuts. Furthermore, punishing the agent at that time may not compensate for the principal's losses in the failure of a valuable engineering project because the agent generally cannot pay for the damage, either because he does not have the funds or because of legal limits. Therefore, the principal may want to set the agent's incentives (rewards and punishments) so that he does not take shortcuts if he[2] is behind schedule.[3]

In this analysis, the principal has to make five decisions that can influence the agent's actions. These are: the payment to the agent for delivering the system on time, the amount to be spent on inspection of the agent's work, and three penalty amounts to the agent (for being delayed, for being caught cutting corners, or for causing a system failure due to shortcuts). When considering the problem from the agent's perspective, we focus on two decisions: how much effort to invest in the project from the beginning so as to remain on schedule, and whether or not to take shortcuts to catch up if he falls behind.

The problem addressed here is to determine how the principal's decisions affect the agent's actions. From the manager's perspective, the purpose of the penalty for falling behind schedule is to create an incentive for the agent to exert effort from the onset of the project in an attempt to complete it on time. The level of agent effort, in turn, affects significantly the principal's expected benefits from the project because if the system is on schedule, the agent does not have to choose between shortcuts and a delay.[4]

To permit mathematical modeling of this complex problem and derivation of closed-form solutions, a number of (reasonable) assumptions regarding rationality, knowledge, and probabilistic relationships were made. The main ones are:

(1) Both the principal and the agent act to maximize their expected returns (i.e., they are risk indifferent). The model structure requires that the agent's utility function be wealth independent (here: risk neutral) to allow for separation of the agent's initial determination of the efforts that he will deploy and his decision to take shortcuts or not if his project is late.
(2) The agent is aware of the principal's decisions regarding penalties and monitoring, which determines his effort level, if he takes shortcuts, and the probability that he is caught cutting corners.
(3) The probabilities of system failure with and without shortcuts can be determined by a PRA, and are known to both the principal and the agent.
(4) Because of the high value of the project, the principal sets the penalties to the agent for being caught taking shortcuts so that the rational employee chooses not to do it.[5]
(5) The agent has limited liability, i.e., there is a limit to all penalties that he can incur.

[2] The agent is referred to by the masculine pronoun and the principal by the feminine.

[3] In the model presented here, it is assumed for simplicity that failure, if it occurs, happens at the time of the start of system operations. The model can be extended to consider failures at any given operation time and include the principal's and the agent's discount factors.

[4] We assume that the principal prefers delaying the project to launching with shortcuts.

[5] Even though the incentives are set so that the rational, risk-indifferent agent will not cut corners in the optimal solution, we still need to calculate the probability of system failure with shortcuts in order to determine the principal's optimal settings.

(6) If the agent cuts corners, he does it so that he regains the time needed to get back on schedule, and he chooses the shortcuts that minimize the resulting increase in the failure probability. This implies that he is experienced and capable of anticipating the potential effects of his actions on the system's performance.
(7) The agent knows how each effort level determines the probability that he later falls behind schedule, and the principal knows how the amount she spends on inspection affects the agent's behavior and the probability that the agent is caught if he takes shortcuts.
(8) We do not consider here the possibility of type II errors (false positive) associated either with monitoring or with attribution of system failure responsibility. That is, the agent can never be accused of corner cutting when he has not, and of having caused system failure if that is not the case.[6]

These assumptions allow using standard optimization techniques to obtain closed-form solutions for the decisions of both the agent and the principal.

One goal of this article is to create a model supporting both decisions of the agent and the principal. More importantly perhaps, it allows the principal (and to a lesser extent, the agent) to understand the relationships between the incentive system that she sets and its link to the system's performance. It also provides insights into the agent's decisions regarding the efforts that he deploys, the shortcuts that may be taken if he is late, and how these decisions affect the principal's expected value of the project.

The model also emphasizes the importance of PRA in determining both the optimal decisions and the value of the project as a whole. Among other insights, our analysis shows that the shortcut issue cannot be separated from how principal-set parameters affect the agent's effort and how such effort affects the need, or lack thereof, for corner cutting or costly system delay. Another interesting finding, when we relaxed the assumption of risk neutrality, was that, from the principal's point of view, there exists an optimal level of agent risk aversion that maximizes her expected value for the project.

The framework presented here can be applied to many real-world problems, but the model may have to be adjusted if the assumptions made here do not hold for a particular problem. In many such cases, the closed-form solutions found here will not apply and the model will have to involve simulation or other numerical techniques to identify optimum solutions.

This article is structured as follows. Section 2 describes the structure of the overall model along with a timeline and the principal's decision diagram. Section 3 presents a PRA of the system to determine the probability of system failure associated with shortcuts or not. Section 4 shows an analysis of the agent's decision problem regarding his level of effort and whether or not to cut corners if he falls behind schedule. Based on the results from Sections 3 and 4, Section 5 presents an optimization model for the principal's decisions. Section 6 expands the model to the case of a risk-averse agent with constant absolute risk aversion. Section 7 discusses the model's limitations due to the strict assumptions that were made to allow for a closed-form solution and Section 8 presents a summary of our findings.

2. MODEL STRUCTURE

The model is structured as follows. The principal sets the payments and penalties for a rational agent in charge of the system's development, and her investment in inspection of the agent's work at a given time. The agent is made aware of these decisions. If the agent decides to accept the project, he commits to a costly effort level that determines the likelihood that he completes his task within the accepted time constraints.[7] At some critical time, the agent observes whether or not the project is late. At that point, he has the choice to admit the delay and pay a penalty. Alternatively, he can take shortcuts that will bring him back on schedule but may be exposed through inspection. In addition, the shortcuts will increase the probability of system failure later in operation.[15–17] If the system fails due to his actions, the agent will be held accountable at that time.

The goal of the model is to determine the principal's optimal settings for the payments and penalties to the agent, and for the optimal inspection amount that maximizes the principal's expected value. That value is what she receives from the system's operation minus the amount she spends on inspection costs and payments to the agent.

[6] One can certainly think of cases where inspection can be imperfect or unwarranted blame assessed in the event of a failure but we do not consider these examples in this article.

[7] The time constraints themselves are assumed to have been set externally, sometimes based on uncontrollable factors such as the position of the planets or the seasonal temperatures.

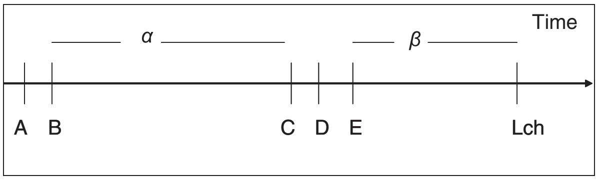


Fig. 1. Timeline of events in principal-agent-system interaction.

The principal's dilemma is that the agent's actions affect both the probability of system failure and whether or not system operation is delayed, both of which influence her expected value of the system.[8] Therefore, in order to determine that expected value for different options and solve for the optimal payment and inspection levels, she needs to consider the agent's responses to these settings in making her decisions.

The supporting analysis involves the following steps:

(1) Determine the probability of system failure with and without shortcuts using a PRA model. Note again that if the agent has any incentive to cut corners, we assume that he will choose the subset of shortcuts that allows him to catch up with the schedule with the lowest increase in system failure probability.[16]
(2) Analyze the agent's optimal decision to cut corners (and which ones) or not if he finds out at a specified time that he is behind schedule given the different potential payments, penalties, and inspection levels set by the principal.
(3) Solve for the agent's optimal effort commitment decision, given his optimal shortcut decision and the principal's setting (incentives and monitoring). This resource commitment determines the probability that the system remains on time.
(4) Solve the principal's optimization model to determine the optimal settings given the agent's preferences (steps 2 and 3 above) and the probability of system failure (step 1 above).

Timeline and decision tree:
The order of events is shown below by the timeline of Fig. 1:

Time 0:

(A) The principal sets payments, penalties, and inspection level.
(B) The agent accepts the project and commits his effort resources.
(Assume that events A and B take place at approximately the same time.)

After α time units:

(C) At time α (e.g., a critical step in the system's development) and after decisions A and B, the agent discovers whether the project is delayed, and by how much.
(D) If delayed, the agent decides whether or not to take shortcuts and which ones, depending on how many time units he must regain and minimizing the increase in system failure probability.
(E) The principal inspects the agent's project. If shortcuts are detected, the agent is penalized and the system launch is delayed.[9]
(Assume that events C, D, and E take place at approximately the same time.)

After $\alpha + \beta$ time units of total development time, the system is "launched" (start of operation).

Lch: After β time units following C, D, and E, the system starts operation ("launch"). We assume that if the system is going to fail, it will happen at that time.

The decision tree for the principal is presented in Fig. 2, with the following notation:

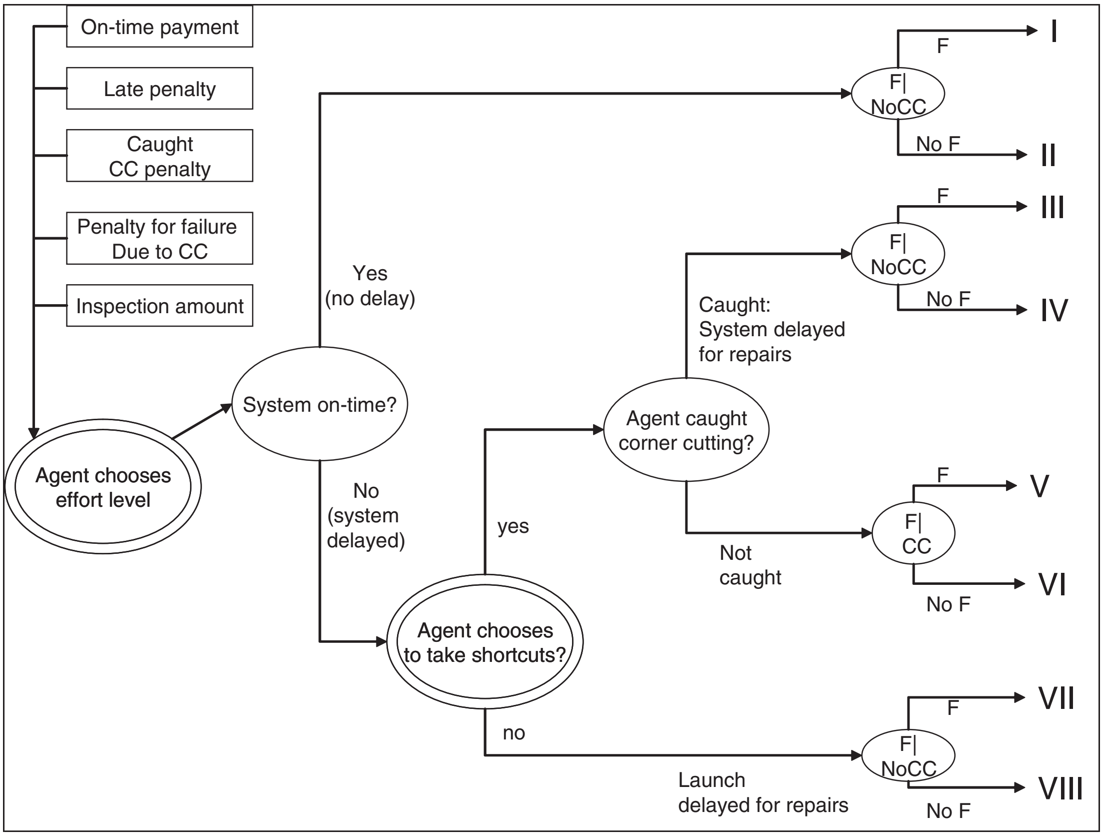
CC: corner cutting;
F: system failure.

The roman numerals on the right side are the consequences to the principal of the potential outcome scenarios and are defined as follows:

(I) On-time system; failure.
(II) On-time system; no failure.
(III) Agent caught corner cutting; system delayed and no failure.
(IV) Agent caught corner cutting; system delayed and system failure.
(V) Agent not caught corner cutting; system failure (either due to corner cutting or regular failure).
(VI) Agent not caught corner cutting; no failure.

[8] The principal's expected value is also affected by the loss associated with a system failure, but this value is not affected by the agent's actions.

[9] We assume that the agent does not incur further costs in the event of a delay. An example of this could be if his initial "effort" cost was comprised of materials, such as the number of trucks purchased, or workers who were paid a one-time fee until project completion.



Legend: The rectangle represents the principal's decisions. The double ovals represent the agent's decisions that are deterministic for the principal based on her decisions. The ovals represent uncertainties.

Fig. 2. Principal's decision tree at time 0.

(VII) Agent admits delay; system delayed and system failure.
(VIII) Agent admits delay; system delayed and no failure.

(The agent finds out about a delay at time α in scenarios III–VIII; the system is originally on time in scenarios I and II.)

Notations:
Principal's decision variables:

x_0 : reward to the agent for delivering an on-time system,
x_L : penalty to the agent for delivering a late system,
x_c : penalty to the agent for being caught cutting corners,
x_F : penalty to the agent for being responsible for a system failure, and
s : amount spent on inspection by the principal; this amount determines the probability that the agent is caught cutting corners.

Agent's decision variables:

r : agent's effort expense; this variable is translated into dollars and determines the probability that the system is on time without shortcuts, and
cc : corner cutting; if the agent is behind schedule, he decides whether to cut corners and which ones based on how many time units he has to regain, the corresponding increase in the probability of system failure, and the principal's settings of incentives and monitoring.

System variables (data):

k : index of subsystems in series,

i : index of components in parallel in each subsystem,
$p_{k,i}$: probability of failure of component k,i *without* shortcut,
$p'_{k,i}$: probability of failure of component k,i *given* shortcut,
p_F : probability of system failure with no shortcuts,
p'_F : probability of system failure for the optimal set of shortcuts that return the system to the original schedule,
p''_F : increase in the probability of system failure caused by agent shortcuts ($p''_F = p'_F - p_F$),
V_{OT} : value to the principal of an on-time, functional system,
V_L : value to the principal of a delayed but functional system,
V'_{OT} : expected value to the principal of an on-time system with no shortcuts, accounting for the possibility of system failure,
V'_L : expected value to the principal of a delayed system with no shortcuts, taking into account the possibility of system failure,
p_{OT} : probability that the system is on time with no shortcuts; this variable depends upon the agent's effort level r and is known to both the principal and the agent, and
p_C : probability that the agent is caught cutting corners; this variable depends upon the amount s that the principal spends on inspection and is known to both the principal and the agent.

Other:

$E_A(\cdot)$ agent's expectation of a random variable,
$E_P(\cdot)$ principal's expectation of a random variable,
$\{(k,i)\}^*_{SC}$: agent's optimal set of shortcuts if he decides to cut some corners,
G : number of time units to be regained at time α (through shortcuts) in order to be back on time,
$g(k,i)$: time gained (number of time units) by cutting corners in component k,i (one option per component),
κ : agent's reservation value, i.e., the threshold of the agent's expected value for the whole project below which he will not accept the job (opportunity cost of his involvement/effort in the project),
a : constant relating the agent's effort to the probability that the system is on time, $p_{OT} \equiv 1 - e^{-r/a}$ with p_{OT}: probability that at time α the project is on schedule and r: amount of effort on the part of the agent,
b : constant relating the principal's amount spent on inspection, s, to the probability that the agent is caught if he cuts corners: $p_C = \frac{s}{b}$,
X_c : maximum penalty to the agent for being caught corner cutting, and
X_F : maximum penalty to the agent for causing a system failure due to shortcuts.

3. PROBABILITIES OF SYSTEM FAILURE WITH AND WITHOUT SHORTCUTS

The probability of system failure with or without shortcuts is determined by a PRA model. Assume that the system contains K subsystems in series (indexed in k) and that each subsystem k is formed of I_k components in parallel (indexed in i). The probability of system failure is given by the sum of the probabilities of failure of each subsystem, minus the probabilities of two subsystems failing (doubles) plus the probability of three subsystems failing (triples), etc. It is given by Equation (1):

$$p_F = \sum_{k=1}^{K}\left(\prod_{i=1}^{I_k}(1 - p_{k,i})\right) - \text{doubles} + \text{triples} \pm \text{etc.}, \tag{1}$$

in which "doubles" and "triples" represent the conjunctions of two and three failure modes, respectively. For any component where the agent has taken shortcuts, $p_{k,i}$ is replaced by $p'_{k,i}$ in the computation of the overall system failure probability.

As mentioned above, it is assumed that the agent, if he cuts corners, does it so as to minimize the probability of system failure given the number of time units that he has to regain. The agent's problem is thus:

Find the set of shortcuts $\{(k,i)\}$ associated with probabilities of failure $\{p'_{ki}(g(k,i))\}$ that yields the minimum p'_F subject to $\Sigma_{k,i}\, g(k,i) \geq G$ in which p'_F is computed by the function of Equation (1) replacing p_{ki} by p'_{ki}.

Numerical illustration 1: Computation of the probability of system failure with and without

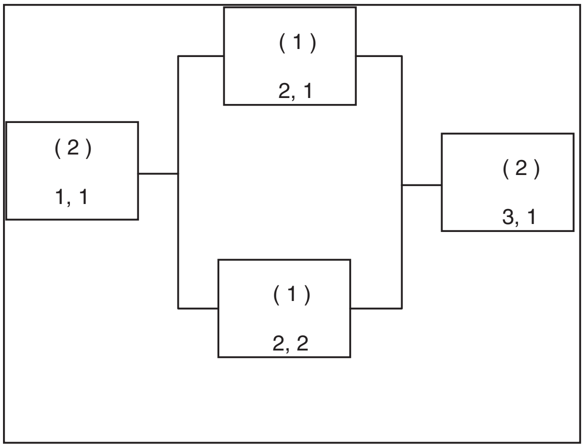


Fig. 3. A simple system with subsystems in series and parallel. The numbers in parentheses represent the amount of time units saved through corner cutting in each component.

shortcuts. An illustration is provided by the simple system shown in Fig. 3: three systems in series, the second composed of two redundant elements in parallel.

Assume the following data:

Probabilities of failure without and with shortcuts (marginal and conditional)

$$\begin{aligned} \text{Without shortcuts}: \quad & \mathrm{p}_{11} = 0.006 \quad \mathrm{p}_{21} = 0.05 \\ & \mathrm{p}_{22|21} = 0.3 \quad \mathrm{p}_{31} = 0.004. \\ \text{With shortcuts}: \quad & \mathrm{p}'_{11} = 0.15 \quad \mathrm{p}'_{21} = 0.4 \\ & \mathrm{p}'_{22|21} = 0.9 \quad \mathrm{p}'_{31} = 0.2. \end{aligned}$$

Based on these data and using Equation (1), the probability of failure given no corner cutting is 0.0248. Further assume that if the system is behind schedule, that it is delayed by two time units ($G = 2$) and that the times gained by shortcuts in each component are:

$$g_{11} = 2, \qquad g_{21} = 1, \qquad g_{22} = 1, \qquad g_{31} = 2.$$

Therefore, to regain two time units, the agent must cut corners on either components (1,1), (3,1), or on both components (2,1) and (2,2). The probability of system failure associated with each shortcut alternative is given in Table I.

Table I shows that the optimal shortcut set $\{(k, i)\}^*_{SC}$ is in component $(1, 1)$ alone. Note that without the PRA model, the agent may choose to cut corners on both components $(2, 1)$ and $(2, 2)$. This would be suboptimal because the failures of these redundant elements are dependent, which would increase the probability of system failure to 0.367 compared to 0.166 with the optimal choice. For this example, the results associated with the agent's optimal choice are thus:

$$\mathrm{p_F} = 0.0248, \quad \mathrm{p'_F} = 0.1661, \quad \mathrm{p''_F} = 0.1413.$$

4. THE AGENT'S DECISION PROBLEM

This section analyzes the agent's two decisions: how much effort to exert from time 0, and which shortcuts, if any, to take if the system is behind schedule at time α. Because the shortcut decision depends upon the agent's effort, we examine the shortcut decision first.

4.1. The Agent's Decision to Cut Corners

This decision occurs at time α if the agent realizes that his project is behind schedule. At this point, he must decide whether or not to admit delay and accept the penalty for it, x_L (payment to the agent $= x_0 - x_L$), or take shortcuts in the hope of not getting caught and not causing a system failure. Again, the assumption here is that if the agent cuts corners, he takes the optimal set of shortcuts (see Section 2). Given that the system is delayed and that the agent decides to cut corners, the expected value of the project to the risk-indifferent agent is, at that time, the weighted average of the three potential outcomes. These are: (1) he is not caught and there is no system failure attributed to his shortcuts; (2) he is caught cutting corners; and (3) he is not caught but the system fails at the beginning of operation because of the shortcuts he took, and he is identified as the culprit.

Given that he takes the optimal set of shortcuts that allow him to get back on time, his expected value at time α is:

$$\begin{aligned} EV_A(\{(k, i)\}^*_{SC}) = {} & (1 - \mathrm{p}_C)(1 - \mathrm{p}''_F)(x_0) \\ & + (\mathrm{p}_C)(x_0 - x_c) + (1 - \mathrm{p}_C)(\mathrm{p}''_F) \\ & \times (x_0 - x_F). \end{aligned} \tag{2}$$

Table I. Probability of Failure and Time Units Regained for Each Corner-Cutting Alternative

Corners Cut	Probability of Failure	Time Units Regained
None	0.0248	N/A
(1, 1)	**0.1661**	2
(3, 1)	0.2167	2
(2, 1) and (2, 2)	0.3667	2

Note: Bold type indicates optimal corner-cutting selection.

Comparing this quantity to expected benefits to the agent of the option to accept the late penalty, yields the following decisions:

$$x_L \leq p_C x_c + (1 - p_C)p''_F x_F$$
$$\Rightarrow \text{the agent takes no shortcuts.} \tag{3}$$

$$x_L > p_C x_c + (1 - p_C)p''_F x_F$$
$$\Rightarrow \text{the agent takes shortcuts the optimal set of shortcuts.} \tag{4}$$

4.2. Agent's Optimal Effort Level

The previous subsection shows whether or not the rational, behind-schedule agent wants to take shortcuts (given the principal's settings and the system's failure probabilities). One can now determine through the maximization of the agent's expected value, the agent's optimal effort expense, r^* (which he decides at the onset of the project), and the subsequent probability that the system is on time, p_{OT}. The agent's effort can be described as a function of many factors: how many people he hires for the tasks, how much of his time he invests in the project, how much equipment he buys or rents, etc. The agent's effort is thus described here as the sum of the monetary value of each of these factors.

To describe p_{OT} as a function of r, we have chosen the negative (concave) exponential function shown in Fig. 4:

$$p_{OT} \equiv 1 - e^{-r/a}, \tag{5}$$

where a is a conversion coefficient that describes the effectiveness of additional agent efforts in keeping the project on time.[10] In the Fig. 4 illustration, we assumed that a is equal to \$25,000. The form of the function itself is reasonable for illustrative purposes because (1) the probability of being on time increases with agent effort; (2) it shows diminishing marginal returns with respect to the agent's effort level; and (3) it makes it impossible for the agent to be absolutely sure that the system will be on time, which is reasonable given that some development problems may be out of his control.

[10] The value of a is related to the efficiency of agent effort. A low value of a implies that a relatively low r can result in a high value for p_{OT} while a high value of a means that more effort is needed to increase p_{OT}.

4.2.1. *Optimal Effort Expense*

Case 1: No shortcuts.

If the agent does not intend to take shortcuts in case of delay at time α, the expected benefits to him associated with the entire project at time 0 are his expected payment minus the cost of effort (scenarios I, II, VII, and VIII in Fig. 2):

$$\begin{aligned} E_A(\text{No CC}) &= p_{OT}x_0 + (1 - p_{OT})(x_0 - x_L) - r \\ &= x_0 - p_{OT}x_L - r. \end{aligned} \tag{6}$$

Using the probability of an on-time outcome shown in Equation (5) ($p_{OT} = 1 - e^{-r/a}$) yields the objective function, of which the agent's effort is the decision variable:

$$\max_r E_A(\text{No CC}) = x_0 - x_L e^{-r/a} - r, \tag{7}$$

$$\frac{dE_A(\text{No CC})}{dr} = \frac{x_L(e^{-r/a})}{a} - 1. \tag{8}$$

Setting to 0 the second term of Equation (8), one finds the value of r that yields the maximum of E_A[11] ($r^*_{\text{No CC}}$ represents the optimal resource allocation given that the agent does not cut corners):

$$r^*_{\text{No CC}} = -a \ln\left(\frac{a}{x_L}\right) \tag{9}$$

(note that if $a > x_L$ then $r^*_{\text{No CC}} = 0$).

Introducing the agent's optimal effort level from Equation (9) into the probability that the system is on time (Equation (5)) allows computation of the probability that the system is on schedule at time α as a function of the penalty for being late:

$$p_{OT} = 1 - \frac{a}{x_L}. \tag{10}$$

As expected, Equations (9) and (10) show that both the agent's effort and the probability that the system is on time are increasing functions of the late penalty x_L.

Case 2: The agent takes shortcuts.

If the incentives are such that the agent is going to cut corners, his expectation for the project is again the expected payment that he will normally receive minus the cost of his effort. The expected value of the benefits to the agent is based on the assumption that he receives the on-time payment but is penalized if he is caught taking shortcuts or if he causes a system

[11] Equation (9) is a global maximum because the function is concave, i.e., the second derivative test is always negative: $\frac{d^2E_A(\text{No CC})}{dr^2} = -\frac{(x_L)(e^{-r/a})}{a^2} < 0$ for all r.

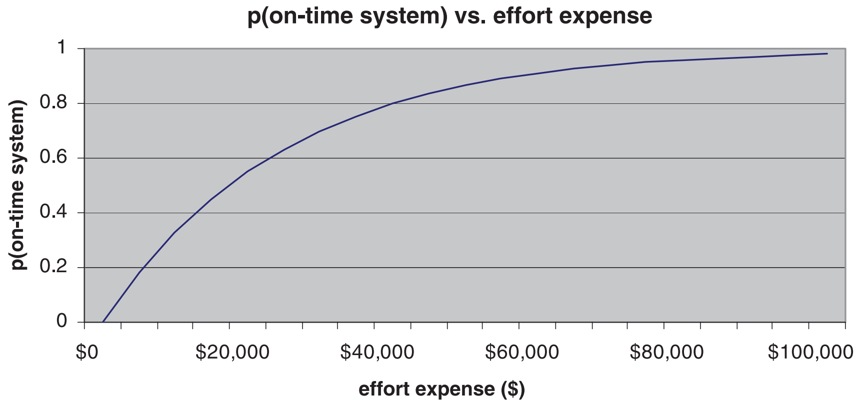


Fig. 4. Illustration of the probability that the agent's system is on time as a function of his effort expense.

failure. The optimum level of effort for this case is given by the equation:

$$\operatorname*{Max}_{r}\ \mathrm{E}_A(\mathrm{CC}) = x_0 - (1-\mathrm{p}_{OT})[\mathrm{p}_C(x_c) + (1-\mathrm{p}_C) \times \mathrm{p}''_F(x_F)] - r. \quad (11)$$

Introducing $\mathrm{p}_{OT} = 1 - \mathrm{e}^{-r/a}$ in this equation yields:

$$\operatorname*{Max}_{r}\ \mathrm{E}_A(\mathrm{CC}) = x_0 - (\mathrm{e}^{-r/a})[\mathrm{p}_C x_c + (1-\mathrm{p}_C) \times \mathrm{p}''_F x_F] - r, \quad (12)$$

$$\frac{\mathrm{d}E_A(CC)}{\mathrm{d}r} = (\mathrm{e}^{-r/a})\frac{[\mathrm{p}_C x_c + (1-\mathrm{p}_C)\mathrm{p}''_F x_F]}{a} - 1. \quad (13)$$

Setting to 0 the terms of Equation (13) and solving for r yields[12] (r^*_{CC} represents the optimal resource allocation given that the agent does not cut corners):

$$r^*_{\mathrm{CC}} = -a\ln\left(\frac{a}{\mathrm{p}_C x_c + (1-\mathrm{p}_C)\mathrm{p}''_F x_F}\right). \quad (14)$$

The resulting probability of an on-time system is:

$$\mathrm{p}_{OT} = 1 - \frac{a}{\mathrm{p}_C x_c + (1-\mathrm{p}_C)\mathrm{p}''_F x_F}. \quad (15)$$

Equations (14) and (15) show, as expected, that the agent's effort and the subsequent probability of an on-time system are increasing functions of x_c and x_F. This means that as the penalties for being caught cutting corners and for causing a failure increase, the agent has more incentive to exert greater effort in an attempt to complete the project on schedule.

[12] Equation (14) yields a global maximum because the second derivative is always negative: $\frac{\mathrm{d}^2\mathrm{E}_A(\mathrm{CC})}{\mathrm{d}r^2} = -\frac{(\mathrm{e}^{-r/a})(\mathrm{p}_C x_c + (1-\mathrm{p}_C)\mathrm{p}''_F x_F)}{a^2} < 0$ for all r.

4.2.2. The Agent's Individual Rationality Constraint (Reservation Threshold)

The previous two subsections assumed that the agent has accepted the contract between him and the manager to proceed to the development of the subsystem under the conditions set by the principal. Yet, the agent will only accept the contract if his overall expected value is greater than his reservation price noted κ. This opportunity cost is a characteristic of the agent's preferences, i.e., how much he values his alternative use of time and effort. Therefore, for the agent to accept the project, the following must be true:

$$\mathrm{E}_A(r^*, \text{optimal corner cutting decision}) \geq \kappa. \quad (16)$$

If the agent's optimal decision is not to take any shortcuts, Equation (16) becomes:

$$\mathrm{p}_{OT}x_0 + (1-\mathrm{p}_{OT})(x_0 - x_L) - r^* \geq \kappa, \quad (17)$$

where r^* here is $r^*_{\mathrm{No}\,CC}$.

5. THE PRINCIPAL'S OPTIMIZATION PROBLEM

The principal's optimization problem is to maximize her expected value through the choices of incentives and monitoring, i.e., x_0, x_c, x_L, x_F, and s. To solve this problem, we make two key assumptions:

(1) *The principal sets incentives so that the rational, risk-indifferent agent will not take shortcuts* (as mentioned earlier). This assumption is reasonable in the case of a system with great value, where the principal under no circumstances wants to risk the increased failure rate associated with any shortcut. This assumption implies that the agent's incentives are such

that Equation (3) is always true, i.e., that $x_L \leq \mathrm{p}_C x_c + (1 - \mathrm{p}_C)\mathrm{p}''_F x_F$ (the penalty for admitting lateness is less than or equal to the expected costs to the agent of cutting corners).

(2) *The agent's liability is limited.* This means that there is an upper limit to the agent's penalties set by the principal (x_c, x_L, and x_F). We assume that these upper limits are externally fixed and based upon a combination of the agent's ability to pay, legal issues, and corporate policies, etc. We also assume that there are different maximum penalties for being caught taking shortcuts and for causing a system failure, the latter generally being more severe than the former, and that the late penalty is limited by Equation (3). Notations X_c and X_F represent the maximum penalties to the agent for being caught cutting corners and for causing a system failure, respectively.

Furthermore, we assume that the probability of the agent being caught cutting corners (given that he did it) is directly proportional to the amount spent on inspection by the principal:

$$\mathrm{p}_C = \frac{s}{b}, \tag{18}$$

where b is a constant characteristic of the effectiveness of monitoring.

5.1. Mathematical Formulation of the Principal's Decision

With the above assumptions, we can now solve for the principal's five decision variables (x_0, x_c, x_L, x_F, and s) by examining her objective function and the problem's constraints.

The principal's objective function is based on her expectation regarding the agent's behavior as characterized in Section 4. It is equal to the principal's expected benefit from the system's operation minus the expected payments to the agent and the money spent on inspections:

$$\underset{\{x_0, x_L, x_c, x_F, s\}}{\mathrm{Max}\ \mathrm{E}_P} = \mathrm{p}_{OT}(V'_{OT} - x_0) + (1 - \mathrm{p}_{OT}) \times (V'_L - (x_0 - x_L)) - s, \tag{19}$$

where V'_{OT} and V'_L are the expected values to the principal of the system launched on time and launched late, respectively, including the probability of failure given that the agent has taken no shortcuts:

$$V'_{OT} = (1 - \mathrm{p}_F)(V_{OT}), \tag{20}$$

$$V'_L = (1 - \mathrm{p}_F)(V_L). \tag{21}$$

Given the relationship between agent effort and the probability that the system is on time ($\mathrm{p}_{OT} = 1 - \mathrm{e}^{-r/a}$), p_{OT} is a function of the late penalty (Equation (12)): $\mathrm{p}_{OT} = 1 - \frac{a}{x_L}$.

Introducing this value of p_{OT} in the principal's objective function yields:

$$\underset{\{x_0, x_L, x_c, x_F, s\}}{\mathrm{Max}\ \mathrm{E}_P} = \left(1 - \frac{a}{x_L}\right)(V'_{OT} - x_0) + \left(\frac{a}{x_L}\right) \times (V'_L - (x_0 - x_L)) - s. \tag{22}$$

Maximizing the principal's objective function is subject to the following constraints (based on previous assumptions):

(1) The agent does not cut corners (Equation (3)):

$$x_L \leq \mathrm{p}_c x_c + \mathrm{p}''_F x_F. \qquad \text{Constraint 1}$$

(2) The agent's threshold constraint for accepting the contract (Equation (17)):

$$\mathrm{p}_{OT} x_0 + (1 - \mathrm{p}_{OT})(x_0 - x_L) - r^*_{\text{No CC}} \geq \kappa. \qquad \text{Constraint 2}$$

(3) The upper limits on penalties to the agent:

$$x_c \leq X_c$$
$$x_F \leq X_F. \qquad \text{Constraint 3}$$

5.2. Solution for the Principal's Decisions

We now show how using the principal's objective function, the expression of the constraints, and the rational agent's actions, the principal's decision problem can be reduced from five decision variables to one variable, leading to a closed-form solution for her optimal decision.

The principal's objective function (Equation (22)) is increasing in x_L (the penalty to the agent for being late), which means the principal prefers the greatest possible value of x_L[13] and that Constraint 1 becomes an equality:

$$x^*_L = \mathrm{p}_C x_c + (1 - \mathrm{p}_C)\mathrm{p}''_F x_F. \tag{23}$$

It follows from Equation (23) that the maximum values of x_c and x_F should be chosen to maximize the

[13] We do not consider the degenerate case where the principal prefers a delayed system because the amount she can penalize the agent is greater than her loss due to a delayed system.

value of x_L. As such, x_c^* and x_F^* are set at their maximum value and can be eliminated as free variables:

$$x_c^* = X_c \text{ and } x_F^* = X_F.$$

Because the probability that the agent is caught taking shortcuts is linked to the cost of monitoring by the relation $p_C = \frac{s}{b}$, the optimal value to the principal for the penalty to the agent for being late, x_L^*, can be written as:

$$x_L^* = \frac{s X_c}{b} + \left(1 - \frac{s}{b}\right) p_F'' X_F. \tag{24}$$

The principal's objective can then be written as a function of two free variables: the on-time payment, x_0, and the amount spent on monitoring, s:

$$\begin{aligned}\max_{\{x_0,s\}} \mathrm{E}_P = {} & \left(1 - \frac{ab}{s X_c + b\mathrm{p}_F'' X_F - s\mathrm{p}_F'' X_F}\right) \\ & \times (V_{OT}' - x_0) \\ & + \frac{ab}{s X_c + b\mathrm{p}_F'' X_F - s\mathrm{p}_F'' X_F}\Bigg(V_L' \\ & - \left(x_0 - \frac{s X_c + b\mathrm{p}_F'' X_F - s\mathrm{p}_F'' X_F}{b}\right)\Bigg) - s.\end{aligned} \tag{25}$$

Equation (25) shows that the principal's expected value is decreasing with respect to the payment x_0 that she makes to the agent for a successful project. Therefore, the principal wants to set x_0 at the lowest possible value that satisfies the agent's rational threshold for accepting the contract. Rewriting Equation (17) as an equality and introducing the values of $r^*_{\text{No CC}}$, p_{OT} and x_L yields:

$$x_0^* = \kappa + a - a \ln \frac{ab}{s X_c + b\mathrm{p}_F'' X_F - s\mathrm{p}_F'' X_F}. \tag{26}$$

Introducing this value $x_0{}^*$ in the principal's objective function (Equation (25)) leaves only one decision variable, s, which is the amount that the principal dedicates to monitoring the progress of the agent's work. For mathematical convenience, define a subset w of the principal's objective function as:

$$w = \frac{ab}{s X_c + b\mathrm{p}_F'' X_F - s\mathrm{p}_F'' X_F}. \tag{27}$$

The principal's objective can now be written as:

$$\begin{aligned}\max_s \mathrm{E}_P &= (1-w)(V_{OT}' - \kappa - a + a\ln w) \\ &\quad + w\left(\left(V_L' - \kappa - a + a \ln w + \frac{a}{w}\right)\right) - s. \\ &= V_{OT}' - \kappa + a \ln w - w(V_{OT}' - V_L') - s\end{aligned} \tag{28}$$

To solve this problem simply requires setting to 0 the first derivative of E_P with respect to s:

$$\frac{\mathrm{dE}_P}{\mathrm{d}s} = \frac{a}{w}\frac{\mathrm{d}w}{\mathrm{d}s} - \frac{\mathrm{d}w}{\mathrm{d}s}(V_{OT}' - V_L') - 1 = 0, \tag{29}$$

$$\frac{\mathrm{d}w}{\mathrm{d}s} = -\frac{ab(X_c - \mathrm{p}_F'' X_F)}{(s X_c + b\mathrm{p}_F'' X_F - s\mathrm{p}_F'' X_F)^2}. \tag{30}$$

Combining these two equations yields the optimal level of monitoring (see Appendix A for the derivation and Appendix B for conditions of optimality):

$$s^* = -\frac{a}{2} \pm \frac{\sqrt{a^2(X_c - \mathrm{p}_F'' X_F)^2 + 4(V_0' - V_L')ab(X_c - \mathrm{p}_F'' X_F)} - 2b\mathrm{p}_F'' X_F}{2(X_c - \mathrm{p}_F'' X_F)}. \tag{31}$$

Equation (31) yields two possible solutions. Because s cannot be less than 0 or more than b (at $s = 0$, $\mathrm{p}_c = 0.0$ and at $s = b$, $\mathrm{p}_c = 1.0$), the principal chooses the value of s^* that satisfies these conditions. One can then determine the principal's remaining decision variables, $x_L{}^*$ and $x_0{}^*$, using Equations (24) and (26), respectively.

All the principal's optimal decision variables have now been identified. With these values, one can identify the optimal level of the agent's effort, the resulting probability that the system is on schedule at time α, and subsequently the principal's expected value for the project.

5.3. Numerical Illustration

Solving for the principal's decision variables:

Data:

$$\begin{array}{llllll} V_{OT} = & \$100\text{M} & V_L = & \$97\text{M} & \kappa = & \$100\text{k} \\ X_c = & \$100\text{k} & X_F = & \$200\text{k} & a = & \$25\text{k} \quad b = \$600\text{k} \\ \mathrm{p}_F = & 0.0248 & \mathrm{p}_F' = & 0.1661 & \mathrm{p}_F'' = & 0.1413 \end{array}$$

$$V_{OT}' = (1 - \mathrm{p}_F)(V_{OT}) = \$97{,}520{,}000$$

$$V_L' = (1 - \mathrm{p}_F(V_L) = \$95{,}594{,}400.$$

Solution:

First, from the data, knowing that $x_c^* = X_c$ and $x_F^* = X_F$, yields:

$$x_c^* = \$100\text{k and } x_F^* = \$200\text{k}.$$

With these values and the given illustrative data, we use Equation (31) to solve for s^*:[14]

$$s^* = \$\ 533{,}365.$$

We can now solve for the principal's other decision variables (optimal penalty for delay and basic payment for successful performance):

$$x_L^* = \frac{s x_c^*}{b} + \left(1 - \frac{s}{b}\right) p''_F x_F^* = \$\ 92{,}033,$$

$$x_0^* = \kappa + A - A \ln \frac{ab}{s x_c^* + b p''_F x_F^* - s p''_F x_F^*}$$
$$= \$157{,}582.$$

These values result in agent effort costs of $r^* = -a \ln(\frac{a}{x_L}) = \$32{,}582$ and a resulting probability that the system is on time = 0.728. Note that if the maximum penalty to the agent for being caught corner cutting, X_c, is set at \$100,000, the maximum effort on the part of the agent that the principal can achieve through incentives is \$34,657, resulting in a probability 0.75 that the system is on time. This occurs if the principal sets her monitoring budget at $s^* = \$600{,}000$, so that the probability that the agent is caught cutting corners if he did is 1.0. The penalty for being late can then be set as equal to the penalty for the agent being caught cutting corners. Table II shows the effect of the monitoring level s^* on the decision variables, as well as the optimal level of the agent's effort, the probability that the system is on time, and the principal's expected benefits.

Table II shows that as the amount spent on inspection increases, the agent expends more effort, resulting in a greater probability that the system is on time. This is true because with a higher probability of catching the agent if he cuts corners, at the optimum, the principal sets a higher penalty for a project delay at time α while still maintaining an incentive for the agent not to take shortcuts if he finds out that the project is late. At some point, however, there are diminishing returns for inspection because beyond the optimum, the principal's expected benefits decrease even though there is a greater probability that the system is on time. Note also that as the inspection amount increases, the principal needs to raise the on-time payment to meet the agent's reservation threshold.

6. EXTENSIONS

6.1. Discounting of the Failure Penalty by the Agent

In the previous example, it was assumed that the time between an agent's acceptance of the project and any potential failure did not affect the effectiveness of the penalty for causing a failure due to corner cutting. Yet, the time β elapsed between the agent's delivery of his project and the start of system operation (and potential system failure) could be substantial, and the agent could discount heavily the failure penalty. The maximum penalty for a failure, X_F, needs to be discounted at the agent's discount rate d, yielding the operational value X'_F:

$$X'_F = \frac{X_F}{(1+d)^\beta}. \qquad (32)$$

We can now substitute X'_F for X_F in Equations (23) to (31) and solve for the principal's optimal settings. Table IV shows the principal's optimal settings as a function of β.

Table IV shows that the principal's expected benefits decrease as β increases because the principal loses some of the deterrence value of the penalty to the agent associated with causing a system failure by taking shortcuts (i.e., the right side of Equation (23) is decreased due to the discounting of the failure penalty). In our example, as β increases, the principal raises both the inspection level and the penalty for project delay. The result is a greater probability that the system is on time, but at a greater price to the principal.

In the case of a long time to operation β and/or a high discount rate d, X'_F approaches \$0 and the optimal penalty for being late (Equation (23)) becomes:

$$x_L^* = p_C x_c. \qquad (33)$$

6.2. The Case of a Risk-Averse Agent

So far in this article, we considered an agent who was risk indifferent. In this section, we consider a risk-averse agent and a principal who is only responsible for setting the agent's penalty for being late; the inspection level and the other incentives (payments and penalties) are assumed to be externally set. Furthermore, the agent is assumed to have

[14] The other value of s was negative, so it was rejected. We also tested the boundary conditions of $s = \$0$ and $s = \$600{,}000$ and found that the current answer was optimal.

Table III. Effect of the Amount that the Principal Spends on Monitoring on the Probability that the Agent is Caught Corner Cutting, the Late Penalty, the On-Time Payment, the Agent's Effort Cost, the Probability the System is On Time, and the Principal's Expected Value

s	p_C	x_0	x_L	R	p_{OT}	E_P
\$0	0.00	\$128,064	\$28,260	\$3,064	0.115	\$94.828M
\$300,000	0.61	\$148,551	\$64,130	\$23,551	0.610	\$95.956M
\$533,365	0.89	\$157,582	\$92,032	\$32,581	0.728	\$96.059M
\$600,000	1.00	\$159,657	\$100,000	\$34,657	0.750	\$96.054M

constant absolute risk aversion (i.e., an exponential disutility function), which, as in the risk-neutral case, is wealth independent. The principal's objective, within the externally set parameters, is to set the penalty for a late project so as to maximize the probability that the system is on time, while discouraging shortcuts. We show here how to calculate the optimal penalty for being late as a function of the agent's risk tolerance.[15] We then use this late penalty to establish the agent's effort as a function of this risk tolerance.

For constant absolute risk aversion, the agent's utility function for any amount X can be written:

$$U_A(X) = 1 - e^{-X/\rho}. \tag{34}$$

X in the model presented earlier represents the present value of his payment less the amount spent on efforts (the former is denoted by x_g):

$$X = x_g - r. \tag{35}$$

Because the agent is assumed to have a constant absolute risk aversion, his decisions are assumed to be independent of his current wealth. As a result, as in the analysis of the risk-neutral agent, his decision to cut corners can be considered separately from his optimal effort calculation. It follows that for the risk-averse agent, the late-payment constraint becomes a comparison between the utility of accepting the penalty for a delay and the expected utility of cutting corners:

$$\begin{aligned} U_A(x_0 - x_L) \le\ & (1 - p_C)(1 - p''_F)U_A(x_0) \\ & + p_C \cdot U_A(x_0 - x_C) + (1 - p_C p''_F \\ & \times U_A(x_0 - x_F) \end{aligned} \tag{36}$$

Eliminating x_0 from both sides of the equation (because the utility is independent of the reference point) and substituting the exponential utility function (Equation (34)) into Equation (36) results in:

$$1 - e^{x_L/\rho} \le p_C(1 - e^{x_C/\rho}) + p''_F(1 - p_C)(1 - e^{x_F/\rho}). \tag{37}$$

The agent's optimal resource commitment is derived from his expected utility for the project, given that he admits delay if the project falls behind schedule:

$$\begin{aligned} EU_A = & \ p_{OT} \cdot U_A(x_0 - r) \\ & + (1 - p_{OT}) \cdot U_A(x_0 - x_L - r). \end{aligned} \tag{38}$$

[15] Risk tolerance, ρ, is defined as the first derivative of one's utility function divided by the second derivative of that utility function ($\rho = -\frac{u'}{u''}$). It is the inverse of the Arrow-Pratt coefficient of risk aversion. An estimate of risk tolerance is the amount that one would be willing to accept a wager of winning ρ and losing $\rho/2$, each with $P = 0.50$.

Table IV. Effect of the Time Elapsed (β) Between Project Inspection (α) and the Launch Time and of the Agent's Discount Rate on the Principal's Optimal Decisions and on the Probability that the Agent is Caught Corner Cutting, the Probability that the System is On Time, and the Principal's Expected Benefits

β	X'_F	p_C	S	x_0	x_L	R	p_{OT}	E_P
0	\$200,000	0.889	\$533,365	0.889	\$157,582	\$92,033	0.728	\$96,059,336
1	\$181,818	0.914	\$548,643	0.914	\$158,014	\$93,639	0.733	\$96,057,262
2	\$165,289	0.936	\$561,449	0.936	\$158,395	\$95,076	0.737	\$96,055,872
3	\$150,263	0.954	\$572,285	0.954	\$158,731	\$96,362	0.741	\$96,054,968
4	\$136,603	0.969	\$581,525	0.969	\$159,028	\$97,515	0.744	\$96,054,410
5	\$124,184	0.982	\$589,460	0.982	\$159,293	\$98,552	0.746	\$96,054,098
6	\$112,895	0.994	\$596,316	0.994	\$159,528	\$99,484	0.749	\$96,053,962
7	\$102,632	1.000	\$600,000	1.000	\$159,657	\$100,000	0.750	\$96,053,943
8	\$93,301	1.000	\$600,000	1.000	\$159,657	\$100,000	0.750	\$96,053,943

Substituting $(1 - e^{-r/a})$ for p_{OT}, and introducing the negative exponential utility argument results in:

$$\mathrm{EU}_A = (1 - e^{-r/a}) \cdot (1 - e^{-(x_0 - r)/\rho}) + e^{-s/a} \cdot (1 - e^{-(x_0 - x_L - r)/\rho}). \qquad (39)$$

The agent's optimal effort level can then be determined by setting the derivative of EU_A to 0 (see Appendix C for the derivation and proof of concavity):

$$r^* = a \ln\left(1 - \frac{\rho}{a} - e^{x_L/\rho} + \frac{\rho}{a} e^{x_L/\rho}\right). \qquad (40)$$

As in the risk-indifferent case, Equation (40) shows that the agent's effort increases as the penalty for being late increases, so x_L is set to its maximum allowable value, which occurs when Equation (37) is an equality. From this result, we can now compute the principal's optimal penalty for being late as a function of the agent's risk tolerance. Fig. 5 displays this optimal penalty as a function of agent risk tolerance.

Fig. 5 shows that for low levels of risk tolerance, the optimal penalty for being late is relatively high. As risk tolerance increases, this optimal penalty decreases, approaching the risk-indifferent optimal late penalty for high values of ρ (i.e., when the agent approaches risk indifference). This is true because as the agent becomes more risk averse, he is more willing to accept a certain payment for being late rather than expend resources in efforts. Therefore, he needs a greater incentive for being on time in order to elicit his maximum resource commitment. Increasing the penalty for being late for less risk-averse agents to a similar amount is not considered feasible as soon as it causes the agent to cut corners.

Knowing the late payment as a function of agent risk tolerance, one can now compute through Equation (44) the agent's optimal effort as a function of his risk tolerance. This relationship is shown in Fig. 6.

Fig. 6 shows that as risk tolerance increases, the agent's effort decreases, until, for large values of risk tolerance, it approaches the effort expended by the risk-neutral agent. Because the principal prefers that the agent expend more effort, she is best off with a relatively risk-averse agent because he will expend more effort resources and the system is more likely to be on time.

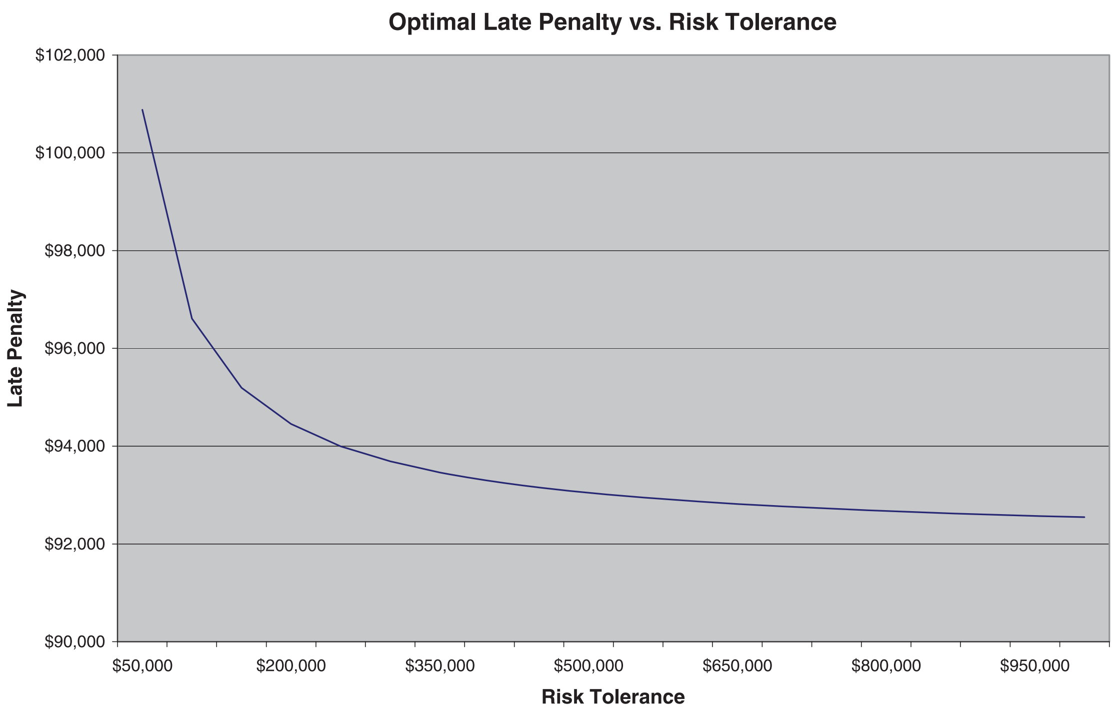


Fig. 5. Optimal penalty to agent as set by the principal for being late as a function of the agent's risk tolerance. This figure uses the same illustrative data as in the numerical example of Section 5.

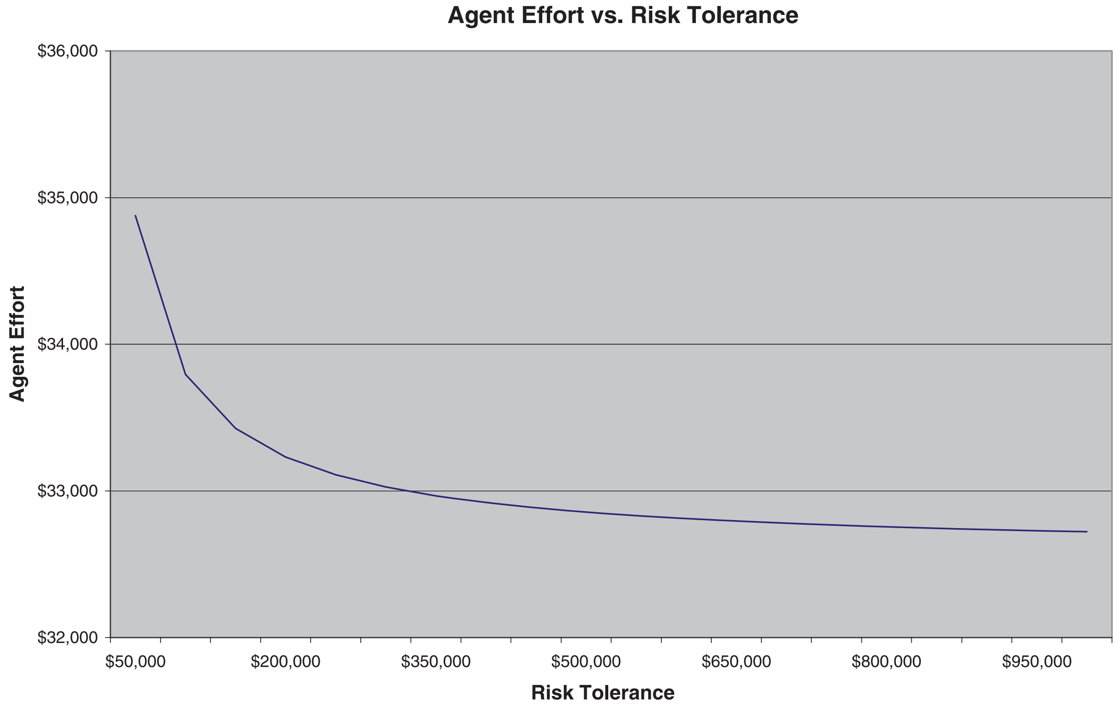


Fig. 6. Agent's optimal effort level as a function of his risk tolerance.

Due to the scale used, Fig. 6 does not show precisely the effects of very low levels of risk tolerance on the agent's effort. At such levels, the agent's effort decreases because his primary concern is spending too much on effort and still being late. Therefore, his effort approaches $0. Fig. 7 displays this behavior as it shows the agent's effort versus risk tolerance for ρ between $25,000 and $50,000 (note that the agent's effort is undefined if $\rho \leq a$ ($a = \$25{,}000$)).

The reason for the drop in resource commitment for efforts is that the maximum possible loss weighs heavily for the highly risk-averse agent. The more resources the agent commits, the greater his maximum loss. Therefore, these results indicate that to the principal, there is an optimal level of agent risk aversion. Knowing this optimum is important to the principal because she can tailor her hiring practices to recruit a type of agent with the desired risk attitude.

7. LIMITATIONS

The solutions presented in this model depend on our ability, first, to calculate the probability of system failure with and without corner cutting; second, to analyze the agent's corner cutting decision; third, to determine the agent's effort decision given that he will not cut corners; and fourth, to assess the relationships between the agent's effort and the chances that he will be late, and between inspection and discovering shortcuts. These assumptions were made because they yielded closed-form solutions. In real applications, the model may have to be adapted and include numerical methods or simulation. Assume now that some of the assumptions that we made do not hold. What changes in the model would be required?

The principal knows the agent's preferences and probability assessments: If this were not the case, the principal would have to put a probability distribution on the agent's preferences and probability assessments and calculate the optimal incentive structure on that basis. This would call for either an exhaustive analysis of possibilities or simulation.

The agent knows the probabilities of system failure with and without shortcuts: In the same way, if the agent does not optimize his choice of shortcuts, the principal needs to put a probability distribution on the agent's possible actions, and use the same kinds of computational method as above.

The agent is rational: Our model assumed a rational agent in the classical economic sense. The agent,

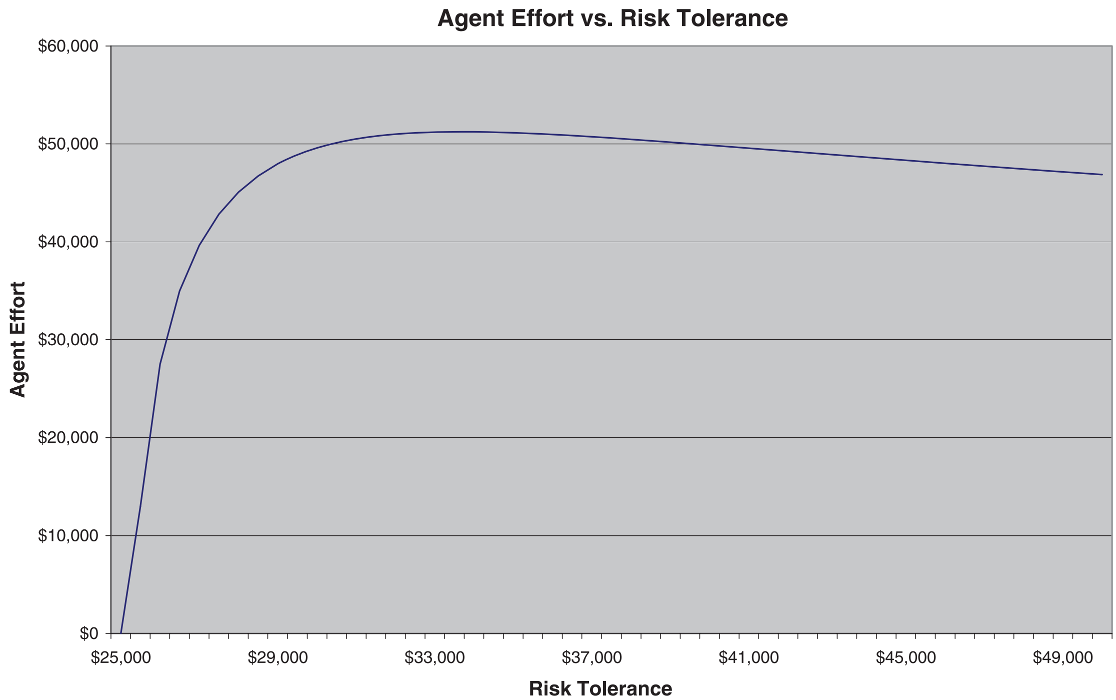


Fig. 7. Agent's effort versus risk tolerance for low levels of risk tolerance.

however, may not be strictly rational and his behavior may be explained by an alternative descriptive model (e.g., a change of preferences over time, or the kind of preferences described by prospect theory).[(18)] In these cases, the principal needs to anticipate the agent's behavior given what she knows about him.

The agent's risk attitude is wealth independent: In our model, we assumed that the agent had constant absolute risk aversion (risk neutrality is a particular case of this profile). If this is not the case,[16] we would not be able to separate the agent's corner cutting decision from his effort decision because the amount of effort expense would influence the agent's choice if the system fell behind schedule. To determine the optimal payments, penalties, and inspection level, the principal would then have to group the agent's two decisions and use numerical methods or simulation to determine the optimal incentives.

There is only one agent: We modeled the engineering system as involving only one agent. In a large system, there may be multiple agents whose actions affect its performance in operation. The principal would then need to take into account the dependencies among the effects of agents' decisions and potential corner cutting on the probability of delay and system failure. Another issue may be whether one set of incentives must (or should) be used for all agents. In that case, again, numerical methods or simulation may be required to determine the optimal incentive structure.

8. SUMMARY

This article has addressed the possibility that an agent cut corners in the development of engineered systems to meet constraints (here assumed to be a deadline), and the risk management problems posed by these shortcuts. As shown by many accidents rooted in human errors, shortcuts can increase considerably the probabilities and the costs of system failures. Often, the employees whose job is to build the system do not and cannot bear the full weight of the failures that they may cause, but they respond to the incentive structure set by their managers. Therefore, we used a principal-agent model to describe this interaction between the manager and the employee in order to understand how the principal can prevent corner cutting.

This article has examined the link between PRA, agent effort, incentives, constraints, and the expected value of the benefits of a project to the principal. We found that by using the principal-agent format, we could determine how the principal's settings for payments, penalties, and inspections influence the

[16] An example of such a risk preference would be a logarithmic utility function, displaying constant *relative* risk aversion.

agent's behavior and determine how these values affect the agent's decisions and actions.

This study thus illustrates the critical balance between the principal's decisions and the optimal behavior of the agent in the case of an engineered system for which we can compute the failure risk with and without shortcuts. It also assesses the principal's dilemma and the losses to her if she makes suboptimal decisions: if the penalty for being late is too high, the rational agent will take shortcuts, but if that penalty is too low, the agent will reduce his effort level and is more likely to fall behind.

We made a number of critical assumptions in terms of knowledge of the principal and the agent. The principal was assumed to have full knowledge of the agent's preferences. The agent was assumed to know to what extent his efforts affected the probabilities that the system be on time, that he be caught cutting corners, and that shortcuts cause a system failure.

An important property of this model is that it can also be used to compute the shadow price of the constraints that the agent faces: What difference will it make in the probability of system failure, thus in the value of the system to the manager, if the schedule is extended by one time unit? This model can thus be a valuable risk management tool to assess the risk-reduction benefits that might be derived from relaxing the constraints—schedule or budget—by one or more units. This is an important feature to the extent that there is generally a link between the schedule constraint and the cost of the project. Schedule constraints can thus be set at a very tight level to increase project revenues. The cost of doing so, however, may be an unacceptable rise in the probability of system failure if the agent(s) tries to satisfy the deadline by taking shortcuts, for example, eliminating some tests, ignoring their results, or using inappropriate material because it is readily available. This can happen in construction, in the oil industry, and in many other fields, including medicine, with catastrophic consequences. Yet these problems can be avoided by anticipating the effects of these constraints and of the incentive system on agent behaviors, and later on system safety. The model that was presented here can thus be an important decision support in proactive risk management for critical systems.

ACKNOWLEDGMENTS

This work was funded in part by a McMurtry student fellowship in the Department of Management Science and Engineering, for which we are grateful. We also thank the anonymous referees of this article for their insightful suggestions, as well as Dr. John Hatfield, from the Stanford Graduate School of Business, for numerous helpful discussions regarding this study.

APPENDIX A: DERIVATION OF THE OPTIMAL MONITORING EXPENSE *S*

From Equation (28), the principal's expected value from the entire project is:

$$\operatorname*{Max}_{s}\ \mathrm{E}_P = V'_{OT} - \kappa + a\ln w - w(V'_{OT} - V'_L) - s,$$

$$\text{where } w \text{ is defined as} : w = \frac{ab}{sX_c + b\mathrm{p}''_F X_F - s\mathrm{p}''_F X_F}.$$

It follows that

$$\frac{\mathrm{dE}_P}{\mathrm{d}s} = \frac{a}{w}\frac{\mathrm{d}w}{\mathrm{d}s} - \frac{\mathrm{d}w}{\mathrm{d}s}(V'_{OT} - V'_L) - 1 = 0,$$

and

$$\frac{\mathrm{d}w}{\mathrm{d}s} = -\frac{ab\,(X_c - \mathrm{p}''_F X_F)}{(sX_c + b\mathrm{p}''_F X_F - s\mathrm{p}''_F X_F)^2}.$$

Let $z = (sX_c + b\mathrm{p}''_F X_F - s\mathrm{p}''_F X_F)\frac{\mathrm{d}w}{\mathrm{d}s} =$ and $-\frac{ab(X_c - \mathrm{p}''_F X_F)}{z^2}$.
Therefore,

$$\frac{\mathrm{dE}_P}{\mathrm{d}s} = -\frac{a\,(X_c - \mathrm{p}''_F X_F)}{z} + \frac{ab\,(X_c - \mathrm{p}''_F X_F)\,(V'_{OT} - V'_L)}{z^2} - 1 = 0.$$

Multiplying both sides of the above equation by z^2 to eliminate the denominator yields:

$$\frac{\mathrm{dE}_P}{\mathrm{d}s} = -z^2 - a(X_c - \mathrm{p}''_F X_F)z + ab\,(X_c - \mathrm{p}''_F X_F) \times (V'_{OT} - V'_L) = 0.$$

The above is solved through a quadratic equation with $a = -1$, $b = -a(X_c - \mathrm{p}''_F X_F)$, and $\mathrm{c} = ab(X_c - \mathrm{p}''_F X_F)(V'_{OT} - V'_L)$.

The solution of this quadratic equation is:

$$z = \frac{-a\,(X_c - \mathrm{p}''_F X_F) \pm \sqrt{a^2\,(X_c - \mathrm{p}''_F X_F)^2 + 4(V'_0 - V'_L)ab\,(X_c - \mathrm{p}''_F X_F)}}{2}.$$

Returning to the original definition of $z = (sX_c + b\mathrm{p}''_F X_F - s\mathrm{p}''_F X_F)$ allows us to solve for s.

$$z = (sX_c + b\mathrm{p}''_F X_F - s\mathrm{p}''_F X_F)\,;\ \text{solving for s} : s = \frac{z - b\mathrm{p}''_F X_F}{X_c - \mathrm{p}''_F X_F},$$

$$s = -\frac{a}{2} \pm \frac{\sqrt{a^2\,(X_c - \mathrm{p}''_F X_F)^2 + 4(V'_0 - V'_L)ab\,(X_c - \mathrm{p}''_F X_F)} - 2b\mathrm{p}''_F X_F}{2\,(X_c - \mathrm{p}''_F X_F)}.$$

APPENDIX B: DERIVATION OF THE SECOND-ORDER CONDITIONS FOR *S*

From Equation (29):

$$\frac{\mathrm{dE}_P}{\mathrm{d}s} = \frac{a}{w}\frac{\mathrm{d}w}{\mathrm{d}s} - \frac{\mathrm{d}w}{\mathrm{d}s}(V'_{OT} - V'_L) - 1,$$

and

$$w = \frac{ab}{sX_c + b\mathrm{p}''_F X_F - s\mathrm{p}''_F X_F}.$$

Therefore:

$$\frac{\mathrm{d}^2\mathrm{E}_P}{\mathrm{d}s^2} = \frac{a}{w}\frac{\mathrm{d}^2 w}{\mathrm{d}s^2} - \frac{a}{w^2}\frac{\mathrm{d}w}{\mathrm{d}s} - \frac{\mathrm{d}^2 w}{\mathrm{d}s^2}(V'_{OT} - V'_L)$$

$$\frac{\mathrm{d}w}{\mathrm{d}s} = -\frac{ab\,(X_c - \mathrm{p}''_F X_F)}{(sX_c + b\mathrm{p}''_F X_F - s\mathrm{p}''_F X_F)^2},$$

$$\frac{\mathrm{d}^2 w}{\mathrm{d}s^2} = \frac{2ab\,(X_c - \mathrm{p}''_F X_F)^2}{(sX_c + b\mathrm{p}''_F X_F - s\mathrm{p}''_F X_F)^3}.$$

Therefore, the following must be negative to ensure that the value of s is a maximum:

$$\frac{\mathrm{d}^2\mathrm{E}_P}{\mathrm{d}s^2} = \frac{2a\,(X_c - \mathrm{p}''_F X_F)^2}{(sX_c + b\mathrm{p}''_F X_F - s\mathrm{p}''_F X_F)^2} - \frac{X_c - \mathrm{p}''_F X_F}{B} - \frac{2ab\,(X_c - \mathrm{p}''_F X_F)^2\,(V'_{OT} - V'_L)}{(sX_c + b\mathrm{p}''_F X_F - s\mathrm{p}''_F X_F)^3}.$$

APPENDIX C: DERIVATION OF *R* FOR A RISK-AVERSE AGENT WITH A CONSTANT RISK AVERSION (THUS AN EXPONENTIAL UTILITY FUNCTION)

From Equation (39), we know that the agent's expected utility is equal to:

$$\mathrm{EU}_A = (1 - \mathrm{e}^{-r/a}) \cdot (1 - \mathrm{e}^{-(x_0 - r)/\rho}) + \mathrm{e}^{-r/a} \cdot (1 - \mathrm{e}^{-(x_0 - x_L - r)/\rho}).$$

For simplicity, we can recalibrate the exponential utility function and remove the constant x_0 from the equation. Additionally, let:

$$\mu = 1/a \text{ and } \omega = 1/\rho.$$

We can now rewrite the agent's expected utility as:

$$\begin{aligned}\mathrm{EU}_A &= (1 - \mathrm{e}^{-\mu r}) \cdot (1 - \mathrm{e}^{\omega r}) + \mathrm{e}^{-\mu r} \cdot (1 - \mathrm{e}^{\omega(x_L + r)}) \\ &= 1 - \mathrm{e}^{-\mu r} + \mathrm{e}^{\omega r - \mu r} - \mathrm{e}^{\omega x_L + \omega r - \mu r}.\end{aligned}$$

Setting to zero the first derivative (first-order conditions):

$$\frac{\mathrm{dEU}_A}{\mathrm{d}r} = 0 = -\omega \mathrm{e}^{\omega r} + (\omega - \mu)(\mathrm{e}^{\omega r - \mu r}) - (\omega - \mu)(\mathrm{e}^{\omega x_L + \omega r - \mu r}).$$

Dividing both sides by $\mathrm{e}^{\omega r - \mu r}$ and reducing the equation yields:

$$\omega \mathrm{e}^{\mu r} = (\omega - \mu) - (\omega - \mu)(\mathrm{e}^{\omega x_L}).$$

The result is thus the following:

$$r^* = \frac{1}{\mu}\ln\left(1 - \frac{\mu}{\omega} - \mathrm{e}^{\omega x_L} + \frac{\mu}{\omega}\mathrm{e}^{\omega x_L}\right),$$

or (returning to A and ρ):

$$r^* = a\ln\left(1 - \frac{\rho}{a} - \mathrm{e}^{x_L/\rho} + \frac{\rho}{a}\mathrm{e}^{x_L/\rho}\right).$$

This is the unique optimal solution (provided that $\rho > a$, otherwise $s^* = \$0$) because the second-order conditions hold for all cases:

$$\frac{\mathrm{d}^2\mathrm{EU}_A}{\mathrm{d}r^2} = -\omega^2\mathrm{e}^{\omega r} + (\omega - \mu)^2(\mathrm{e}^{\omega r - \mu r} - \mathrm{e}^{\omega x_L + \omega r - \mu r}),$$

and this second derivative is negative for all positive values of ω and μ.

REFERENCES

1. Richard Justice PK. The Westray Story: A Predictable Path to Disaster, Report of the Westray Mine Public Inquiry. Halifax, Nova Scotia: Government Printer, 1997.
2. Welch R. NASA Culture of Denial Contributed to Loss of Columbia Space Shuttle, 2003. Available at: www.welchco.com/sd/08/00101/02/03/08/26/162321.HTM#8K4G, Accessed on June 16, 2011.
3. Nash J. Non-cooperative games. Annals of Mathematics, 1951; 54(2):286–295.
4. Harasanyi JC. A bargaining model for the cooperative n-person game. Pp. 325–355 in Tucker AW, Luce RD (eds). Contributions to the Theory of Games IV, Annals of Mathematical Studies, Vol. 40, Princeton, NJ: Princeton University Press, 1959.
5. Gibbons R. Game Theory for Applied Economics. Princeton, NJ: Princeton University Press, 1992.
6. Milgrom PR, Roberts J. Economics, Organization and Management. Upper Saddle River, NJ: Prentice-Hall, 1992.
7. Laffont J, Martimort D. The Theory of Incentives: The Principal-Agent Model. Princeton, NJ: Princeton University Press, 2001.
8. Kumamoto H, Henley EJ. Probabilistic Assessment and Management for Engineers and Scientists. New York: IEEE Press, 1996.
9. Bier VM, Nagaraj A, Abhichandani V. Protection of simple series and parallel systems with components of different values. Reliability Engineering and System Safety, 2005; 87:315–323.
10. Paté-Cornell ME. Rationality and imagination: The engineering risk analysis method and some applications. In Edwards W, Miles R, von Winterfeldt D (eds). Advances in Decision Analysis. Cambridge, MA: Cambridge University Press, 2007.
11. Flyvbjerg B, Bruzelius N, Rothengatter W. Megaprojects and Risk. Cambridge, UK: Cambridge University Press, 2003.
12. Meredith JR, Mantel, SJ. Project Management: A Managerial Approach. New York: Wiley, 2009.
13. Dillon RL, Paté-Cornell ME, Guikema SD. Programmatic risk analysis for critical engineering systems under tight resource constraints. Operations Research, 2003; 51(3):354–370.
14. Lyneis JM, Ford DN. System dynamics applied to project management: A survey, assessment, and directions for future research. System Dynamics Review, 2007; 23(2/3) (Summer/Fall):157–189.
15. Garber RG, Paté-Cornell ME. Managing shortcuts in engineering systems and their safety effects: A management science perspective. Proceedings of PSAM8, Paper # 0036, 2006.
16. Garber RG. Corner Cutting in Complex Engineering Systems: A Game Theoretic and Probabilistic Modeling Approach. Doctoral Thesis, Department of Management Science and Engineering, Stanford University, 2007.
17. Paté-Cornell ME (with the collaboration of Garber RG, Guikema SD, Kucik P). Games and risk analysis: Three examples of single and alternate moves. In Bier V, Azaiez MN (eds). Risk Analysis and Game Theory. New York: Springer, 2008.
18. Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. Econometrica, 1979; 47:263–291.