Journal of Accounting Research
Vol. 27 No. 1 Spring 1989
Printed in U.S.A.

Escalation Errors and the Sunk Cost Effect: An Explanation Based on Reputation and Information Asymmetries

CHANDRA KANODIA,* ROBERT BUSHMAN,† AND JOHN DICKHAUT††

1. Introduction

This paper is concerned with explaining a paradox of human behavior. Consider the following example.

A department manager makes a large investment in new production equipment and, soon after, learns of different equipment that could perform the same operations at lower cost. Incremental analysis favors switching, but the manager refuses saying he does not want to waste the investment already made.

Such real-world examples have been extensively documented and studied by social scientists (see Thaler [1980] and Staw [1981]). In these examples, the decision maker, having commited to a course of action, subsequently discovers new information that indicates that continuing the earlier commitment would likely result in worse consequences than switching. In spite of this he clings to and even escalates his earlier

* University of Minnesota; †University of Chicago; ††University of Minnesota. This paper is an extension of our earlier working paper, "Sunk Costs: Is There a Fallacy?" (January 1984). We would like to thank Amin Amershi, Stan Baiman, Joyce Berg, Nick Dopuch, Mark Penno, and Mark Wolfson for insightful suggestions on earlier drafts of this paper. Helpful comments were also provided by participants in workshops at Stanford University, Washington University at St. Louis, the Wharton School, New York University, and the University of Pittsburgh. All errors and omissions are solely the responsibility of the authors. [Accepted for publication May 1988.]

commitment, a commitment often involving large expenditures of resources. In this case, escalation has been interpreted as evidence that decision makers do not ignore sunk costs, and is part of a general phenomenon carrying various titles such as "the sunk cost effect," "escalation behavior," and "escalation error."

Why does this seemingly irrational behavior occur? Existing explanations rely exclusively on psychological factors such as psychic accounting (Thaler [1980]), framing (Laughhunn and Payne [forthcoming]), and need for internal justification (Staw [1976]). Staw and Ross [1986] provide a comprehensive review of this psychological literature. The purpose of this paper is to present an alternative explanation that relies solely on economic rationality.

We demonstrate that escalation behavior can be explained as part of a larger phenomenon of hiding private information on human capital. When information on his human capital is private to a manager, and can only be inferred by others from observation of the former's actions and their consequences, these actions acquire a reputation value. In general, reputation effects distort preferences over actions. Kreps and Wilson [1982*a*] showed how reputation could be used to explain the chain store paradox, and Milgrom and Roberts [1982] explain predatory pricing in this manner. In our model, reputation is used in the following way. The manager chooses between escalation and switching in the light of information he *privately* observes. Human capital is introduced in such a manner that this private information is also informative with respect to the manager's talent level. In this situation, when the manager switches, he reveals information which damages his reputation for talent and thereby hurts his opportunities in a labor market. This precipitates escalation behavior which is puzzling if reputation effects are ignored.

Section 2 contains a model of managerial labor markets which shows how a desire for reputation building could arise endogenously. The key result from this section is that, in equilibrium, the expected wage of a manager is strictly increasing in his reputation for talent. Readers who are willing to assume this result and are primarily interested in our explanation of escalation behavior may skip this section without loss of continuity.

Section 3 specifies the economy in which escalation vs. switching behavior is studied, and section 4 defines and motivates the notion of equilibrium that is used. Section 5 contains the main results of the paper. We demonstrate there that escalation errors do not occur in a public-information world, i.e., the manager always switches when subsequent information contradicts his earlier commitment. Next, we show how in a private-information world the desire to build reputation distorts the manager's incentives. Precise conditions are derived under which the manager's equilibrium decision strategy will, and will not, manifest escalation errors. Section 6 concludes the paper.

2. Reputation in Managerial Labor Markets

A key ingredient of our explanation of escalation behavior is that the equilibrium in the labor market is such that expected wages of managers are strictly increasing in their reputation for talent. For this property to hold it must be the case that there do not exist contracts, contingent on future performance, that separate manager types through self-selection. (See Rothschild and Stiglitz [1976].) The purpose of this section is to show that two simple assumptions can be introduced in the Rothschild and Stiglitz analysis to yield a pooling equilibrium. These assumptions are limited liability and imperfect observability of the manager's stochastic marginal product. Given these assumptions, we derive a pooling equilibrium in which reputation has economic value.

Consider a market in which two or more identical firms bid in a Bertrand fashion to hire a single manager who could be one of two types. The manager could either be talented, type z_T, or untalented, type z_N. In general, let $z \in Z$ denote manager type. The manager knows his type, but bidding firms do not. Let ρ, $0 < \rho < 1$, be the latter's assessed probability that the manager is of type z_T. The marginal product of the manager in a firm that employs him is an unobservable random variable, but its distribution conditional on talent is known. These distributions satisfy first-order stochastic dominance, in the sense that higher levels of talent shift the distribution to the right. Let $G(z)$ be the expected marginal product of manager type $z \in Z$ where:

$$G(z_T) > G(z_N) > 0. \tag{1}$$

There is an observable random variable (or performance index), $v \in V$, that is correlated with the manager's marginal product. For simplicity, we assume that V is a two element set (v_H, v_L). The marginal probabilities of v conditional on z satisfy:

$$1 > P(v_H \mid z_T) > P(v_H \mid z_N) > 0. \tag{2}$$

The relationship between $G(z)$ and $P(v \mid z)$ is assumed to satisfy:

$$\frac{G(z_T)}{P(v_H \mid z_T)} > \frac{G(z_N)}{P(v_H \mid z_N)}. \tag{3}$$

This last assumption is crucial to our results regarding equilibrium wage contracts. For assumptions (2) and (3) to be satisfied simultaneously, it is necessary that v is positively correlated with but is an imprecise measure of the manager's marginal product.

Contracts offered by firms are allowed to be contingent on observable performance subsequent to employment, i.e., contingent on v. We assume that managers have limited liability; in particular, negative wages are infeasible. Thus, the set of feasible wage contracts is the set of functions,

$w : V \rightarrow R$ such that:

$$w(v) \geq 0, \qquad \forall\, v \in V. \tag{4}$$

Finally, we assume that the manager and firm are risk neutral. This is not a key assumption. We make it to simplify the analysis and abstract from risk-sharing considerations.

2.1 NONEXISTENCE OF SEPARATING CONTRACTS

Suppose that firms offered a pair of contracts in the hope that a z_N manager would choose one contract and a z_T manager would choose the other. Competition among firms will ensure that each such contract pays the manager his expected marginal product conditional on type. Thus, in equilibrium, these contracts must satisfy:

$$\sum_{v \in v} P(v \mid z) w(v) = G(z), \qquad z = z_N, z_T. \tag{5}$$

In figure 1, $L^N M^N$ and $L^T M^T$ represent the sets of contracts that satisfy (4) and (5) conditional on $z = z_N$ and $z = z_T$, respectively. Assumption (3) ensures that L^T lies above L^N. Further, $L^T M^T$ must be less steep than $L^N M^N$, since the slopes of these lines are $-P(v_L \mid z)/P(v_H \mid z)$. Therefore, $L^T M^T$ must lie above $L^N M^N$ throughout the nonnegative orthant as depicted in figure 1. But this means that both manager types would strictly prefer every contract on $L^T M^T$ to any contract on $L^N M^N$. Therefore, in equilibrium, there can be no separating contracts.

Now consider pooling contracts. Recall that the manager has prior probability ρ of being talented. Therefore, pooling contracts must satisfy:

$$\sum_{v \in V} [\rho P(v \mid z_T) + (1 - \rho) P(v \mid z_N)] w(v) = \rho G(z_T) + (1 - \rho) G(z_N). \tag{6}$$

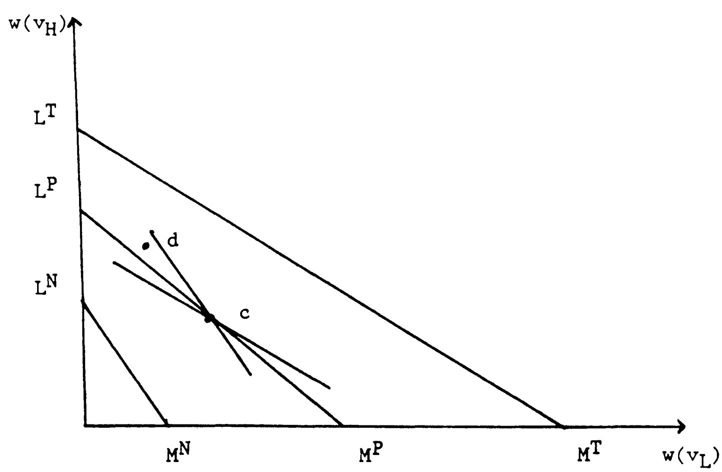

FIG. 1

The set of contracts satisfying (6) and (4) is described by the line $L^P M^P$, in figure 1. We show that competition among firms results in the manager obtaining the pooling contract L^P, in equilibrium. Consider any other contract on $L^P M^P$, such as contract c, in figure 1. Through c, we have drawn indifference curves for the two manager types. The z_N manager has the steeper indifference curve. Suppose one firm offers c and the other responds with contract d. The z_T manager would choose d, while the z_N manager would prefer c. Clearly d is a profitable defection since it lies below $L^T M^T$, while c, if accepted, would result in losses to the firm offering it. This type of argument, first developed by Rothschild and Stiglitz (1976), can be repeated for every pooling contract other than the contract L^P.

Note that L^P maximizes the expected wage of the z_T manager and minimizes the expected wage of the z_N manager, subject to (6) and (4). But, since L^P lies strictly above $L^N M^N$ and strictly below $L^T M^T$, the z_N manager is overpaid and the z_T manager is underpaid relative to his expected marginal product. Thus, in this equilibrium, there are incentives for firms to take into account any other observables that may contain information on manager type.

We now claim that, given a pooling equilibrium, the expected wage of the manager is strictly increasing in the assessed probability that he is talented. The pooling contract L^P is characterized by:

$$\begin{aligned} w(v_L) &= 0, \\ w(v_H) &= \frac{\rho G(z_T) + (1-\rho)G(z_N)}{\rho P(v_H \mid z_T) + (1-\rho)P(v_H \mid z_N)}. \end{aligned} \tag{7}$$

The derivative of $w(v_H)$ with respect to ρ is positive if:

$$\begin{aligned} &[G(z_T) - G(z_N)][\rho P(v_H \mid z_T) + (1-\rho)P(v_H \mid z_N)] \\ &\qquad > [\rho G(z_T) + (1-\rho)G(z_N)][P(v_H \mid z_T) - P(v_H \mid z_N)]. \end{aligned}$$

Canceling common terms and rearranging yields:

$$G(z_T)P(v_H \mid z_N) > G(z_N)P(v_H \mid z_T),$$

which is true from assumption (3).

The above analysis establishes the equilibrium relationship between reputation and expected wages in the labor market we have described. Clearly, if the manager could take actions, prior to contracting, that impact the assessed probability he is talented (i.e., his reputation), he would prefer the action that maximizes his reputation, other things being the same. The reader can verify that all of the above results continue to hold if the manager does not know for sure whether he is talented or untalented but has private information on the *probability* that he is talented.

We now turn to the main object of analysis in this paper and show how the reputation value of actions could lead to the occurrence of escalation behavior. We begin by describing the economy in which escalation vs. switching decisions are studied.

3. *Structure of the Economy*

The economy consists of two periods. In the first period the manager is self-employed and manages his own small private firm. In the second period the manager shuts down his private firm and seeks employment. Two or more large firms bid to hire him. As before, the manager could be one of two types, talented or untalented, and all parties are risk neutral.

In his self-employed capacity, the manager is engaged in choosing among investment projects and implementing them to produce inflows. The distribution of inflows depends on whether there is a match between the project implemented and some underlying state of nature. Assume there are two possible states of nature, θ_A and θ_B, and the manager chooses between two projects A and B. Project A is more desirable if the underlying state is θ_A, while B is favored if the state is θ_B. Initially, the manager observes some information signal on the state of nature, picks an investment project, and proceeds to implement it. Some time during implementation the state of nature is revealed. In the light of this new information the manager decides whether to continue (escalate) the project earlier chosen or switch projects at some cost. Period one ends with realization of project inflows. In the next period the manager seeks employment.

The above scenario is modeled in the following way (refer to the time line in figure 2). We think of period one as consisting of three stages. At the beginning of the first stage the manager either sees the state of nature or fails to see it. If he sees state θ_A, (θ_B), we say he has received the signal y_A, (y_B). If he fails to see the state of nature, we say he has received the signal y_0, in which case the posterior probabilities of θ_A and θ_B are equal to their priors. In general, let y represent the first-stage signal. Thus $y \in \{y_A, y_B, y_0\} \equiv Y$. Let $P_1(y|\theta,z)$ be the conditional probabilities of first-stage signals, conditioned by state and managerial talent.

It is convenient to assume that everything about A and B is symmetric. Thus, we assume that prior probabilities on states are:

$$P(\theta_A) = P(\theta_B) = 0.5. \tag{8}$$

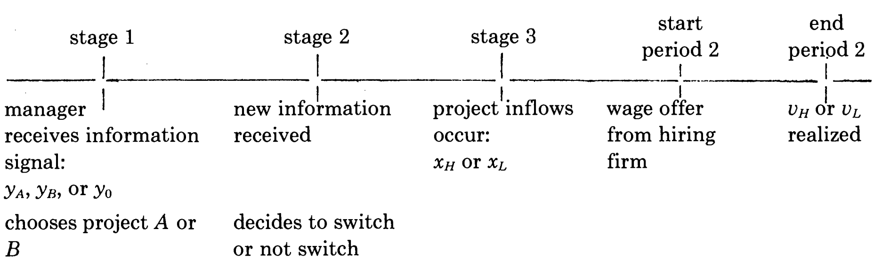

FIG. 2

Since y_0 is uninformative, the conditional probabilities of y_0 satisfy:

$$P_1(y_0 \mid \theta_A, z) = P_1(y_0 \mid \theta_B, z), \qquad \forall\, z \in Z. \tag{9}$$

This in turn implies that:

$$P_1(y_A \mid \theta_A, z) = P_1(y_B \mid \theta_B, z), \qquad \forall\, z \in Z, \tag{10}$$

and by earlier assumption:

$$P_1(y_B \mid \theta_A, z) = P_1(y_A \mid \theta_B, z) = 0, \qquad \forall\, z \in Z. \tag{11}$$

Having observed $y \in Y$, the manager chooses either project A or B and proceeds to implement it. At the end of stage one the chosen project is only partially implemented. If the manager failed to see the state of nature in stage one, he discovers the true state at the beginning of stage two. In the light of this new information, the manager decides whether to switch projects or continue with the previously chosen project. Switching projects entails switching costs, defined by:

$$C(AB) = C(BA) > 0, \text{ and } C(AA) = C(BB) = 0, \tag{12}$$

where $C(AB)$ is the cost of switching from A to B, and so on.

Implementation of projects is completed in stage three, and project inflows, x, are realized. There are only two possible inflows: high inflows, denoted x_H; or low inflows, denoted x_L. The probability of high inflows depends only on whether there is a match between the project finally implemented and the state of nature. Let:

$P_m = P(x_H \mid \theta_A, A) = P(x_H \mid \theta_B, B) =$ probability of high inflows given a match between the state of nature and the project finally implemented.

$P_u = P(x_H \mid \theta_A, B) = P(x_H \mid \theta_B, A) =$ probability of high inflows if project and state are unmatched.

We assume that:

$$0 < P_u < P_m < 1. \tag{13}$$

Finally, to allow for the possibility of escalation error, we assume that, in the presence of disconfirming information, the incremental expected project inflows from switching exceed the cost of switching, i.e.:

$$\sum_x xP(x \mid \theta_B, B) - \sum_x xP(x \mid \theta_B, A) > C(AB); \tag{14}$$

or equivalently:

$$(x_H - x_L)(P_m - P_u) - C(AB) > 0. \tag{15}$$

3.1 MANAGERIAL FORESIGHT

The role of managerial talent is central to our analysis of escalation behavior. Typically, managerial talent is thought of as the ability to organize, supervise, and in general "manage" the various inputs needed for investment and production. Here, we suppress this production aspect of talent since it is not a key factor in explaining escalation behavior. In our model, managerial talent represents "foresight," in the following sense. A farsighted manager is quicker to anticipate future developments than a manager without foresight, who requires more observations to arrive at the same conclusions. For example, a chess master, observing a configuration of pieces on the chessboard, foresees the same opponent's strategy that the novice is defeated by ten moves later. An engineer supervising the construction of a dam would typically uncover unanticipated technological complications as the construction proceeds. The engineer with foresight would likely discover these complications earlier than the engineer with less foresight.

In our model, foresight ability is captured by the assumption that, on average, the talented manager would discover the state earlier than the untalented manager, i.e.:

$$0 < P(y_0 \mid z_T) < P(y_0 \mid z_N) < 1, \tag{16}$$

where $P(y_0 \mid z) = P(\theta_A)P(y_0 \mid \theta_A, z) + P(\theta_B)P(y_0 \mid \theta_B, z)$.

The information structure in the economy is as follows. The talent of the manager is unknown to everybody in the economy, including the manager himself. Talent levels are learned over time by Bayesian updating conditioned on observable variables. The information signal in stage one and the state revealed in stage two are *privately* observed by the manager. Only the sequence of projects chosen by the manager and the inflows realized in stage three are publicly observed.

4. *The Equilibrium Concept*

The equilibrium concept used is Kreps and Wilson's [1982*b*] sequential equilibrium. This equilibrium consists of several components: decision strategies for the manager in stages one and two, consistent beliefs for firms contingent on each feasible decision and outcome, and wage strategies for firms that compete to hire the manager.

4.1 DECISION STRATEGIES OF THE MANAGER

In order to formalize the manager's decision strategy, let $d_1 \in \{A, B\}$, $d_2 \in \{A, B\}$ be the manager's decisions in stages one and two respectively. The manager's decision strategies are described by a pair of mappings:

$$\mu_1\colon \{A, B\} \times Y \to [0, 1], \qquad \mu_2\colon \quad \{A, B\} \times Y \times \{A, B\} \times \Theta \to [0, 1],$$

where:

$\mu_1(d_1 \mid y) =$ probability of choosing d_1 in stage one, given that signal y was observed, and

$\mu_2(d_2 \mid y, d_1, \theta) =$ probability of choosing d_2 in stage two, given that signal y was observed and d_1 was chosen in stage one, and θ was observed in stage two.

Let d be the vector of decisions, (d_1, d_2), and define:

$$\mu(d \mid y,\theta) \equiv \mu_1(d_1 \mid y)\mu_2(d_2 \mid y\ d_1,\theta).$$

Let $D \equiv \{AA,AB,BB,BA\}$ be the set of feasible decision vectors.

4.2 CONSISTENT BELIEFS

As noted earlier, firms observe the pair (d, x) before offering contracts contingent on v (future observable performance). Since the choice of d is contingent on y, and y contains information on talent, (d,x) is potentially informative with respect to talent. In a sequential equilibrium, wage contracts offered by firms in response to each observed (d, x) must be sequentially rational given beliefs conditional on (d, x). These beliefs must satisfy Bayes' Rule at each information set that lies on the equilibrium path. Let:

$\pi(d,x\mu)$ = the posterior probability that the manager is of type z_T, given the manager's decision strategy μ, and given that (d, x) is observed.

From Bayes' Rule:

$$\pi(d,\ x;\ \mu) = \frac{P(d,\ x \mid z_T)\rho}{P(d,\ x \mid z_T)\rho + P(d,\ x \mid z_N)(1-\rho)}, \tag{17}$$

where:

$$P(d,\ x \mid z) = \sum_{\theta\in\Theta} P(\theta) \sum_{y\in Y} P_1(y \mid \theta,\ z)\mu(d \mid y,\ \theta)P(x \mid \theta,\ d_2). \tag{18}$$

Expanding (18) and using (8) yields:

$$\begin{aligned} P(d, x \mid z) = 0.5[1 - P(y_0 \mid z)][&\mu(d \mid y_A, \theta_A)P(x \mid \theta_A, d_2) \\ &+ \mu(d \mid y_B, \theta_B)P(x \mid \theta_B, d_2)] \\ &+ 0.5P(y_0 \mid z)[\mu(d \mid y_0, \theta_A)P(x \mid \theta_A, d_2) \\ &+ \mu(d \mid y_0, \theta_B)P(x \mid \theta_B, d_2)], \end{aligned} \tag{19}$$

where $P_1(y_A \mid \theta_A,z) = P_1(y_B \mid \theta_B,z) = 1 - P(y_0 \mid z)$.

Posterior beliefs can be calculated in this manner only when $\mu(d \mid y,\theta) > 0$ for some (y,θ). When d is off the equilibrium path, beliefs will be defined as the limit of a sequence of beliefs derived from a sequence of completely mixed strategies, as required in sequential equilibria.

4.3 WAGE STRATEGIES OF FIRMS

In general, wage contracts offered by firms are described by real-valued functions, $w(v,\ d,\ x)$. Since in a sequential equilibrium, firms' responses to any observed $(d,\ x)$ must be sequentially rational, wage offers are contingent on (d, x) pairs only through their impact on the assessed distribution on manager types.

In our earlier discussion of the managerial labor market, we derived equilibrium wage contracts for any arbitrary set of beliefs. In that discussion the assessed probability of type z_T was ρ. Replacing ρ by the posterior assessment $\pi(d, x)$, we have established that equilibrium wage contracts must satisfy:

$$w(v_L, d, x) = 0, \forall (d, x); \tag{20}$$
$$w(v_H, d, x) = \frac{\pi(d,x)G(z_T) + \{1-\pi(d,x)\}G(z_N)}{\pi(d,x)P(v_H \mid z_T) + \{1-\pi(d,x)\}P(v_H \mid z_N)}, \forall (d,x).$$

It was also established that $w(v_H, d, x)$ is strictly increasing in $\pi(d, x)$, so that the greater is the assessed probability that the manager is talented the greater is his expected wage.

4.4 THE MANAGER'S OPTIMIZATION PROBLEM

The decision problem faced by the manager can now be specified. Given the hiring strategies of firms, described by the schedule $w(v,d,x)$, the manager chooses projects in stages one and two to maximize his expected income over two periods. Clearly, the manager's optimal strategy in stage one depends on his strategy for stage two. Therefore, his choice problem is described recursively, as follows.

Consider, first, the manager's decision problem in stage two. The manager has received a signal $y \in Y$ in stage one and has privately calculated a posterior distribution on his types. Let $\psi(y)$ be the privately calculated posterior probability of type z_T, where:

$$\psi(y) = \frac{P(y \mid z_T)\rho}{P(y \mid z_T)\rho + P(y \mid z_N)(1-\rho)}.$$

It is easy to verify that $\psi(y_0) < \psi(y_A) = \psi(y_B)$, so that after receipt of y there are two manager types. The manager then assesses a conditional distribution on his future performance. This is described by:

$$P(v \mid y) = \psi(y)P(v \mid z_t) + \{1 - \psi(y)\}P(v \mid z_N).$$

Now suppose that having received $y \in Y$ the manager chose $d_1 \in \{A,B\}$ in stage one and has further discovered state θ in stage two. The manager's optimal stage-two decision is described by:

$$\Omega_2(y, d_1, \theta) = \underset{d_2 \in \{A,B\}}{\text{Max}} \sum_x xP(x \mid \theta, d_2) - C(d_1, d_2) + \sum_x \sum_v P(x \mid \theta, d_2)P(v \mid y)w(v, d, x). \tag{21}$$

His optimal stage-one decision, after observing $y \in Y$, is described by:

$$\Omega_1(y) = \underset{d_1 \in \{A,B\}}{\text{Max}} \sum_\theta P(\theta \mid y)\Omega_2(y, d_1, \theta). \tag{22}$$

In the above expressions, Ω_1 and Ω_2 represent the manager's expected utility from behaving optimally, given his state variables.

The above discussion motivates the following definition of equilibrium, for the economy under study.

4.5 DEFINITION OF EQUILIBRIUM

An equilibrium is a pair of value functions, $\Omega_2(y, d_1, \theta)$ and $\Omega_1(y)$, a pair of decision strategies, $\mu_2(d_2 \mid y, d_1, \theta)$ and $\mu_1(d_1 \mid y)$, beliefs, $\pi(d, x)$, and a wage function, $w(v, d, x)$, such that:

(*i*) Given $w(v, d, x)$, $\Omega_2(y, d_1, \theta)$ satisfies (21) and $\Omega_1(y)$ satisfies (22).

(*ii*) $\mu_2(d_2 \mid y, d_1, \theta) \geq 0$, $\sum_{d_2 \in \{A,B\}} \mu_2(d_2 \mid y, d_1, \theta) = 1$.
$\mu_2(d_2 \mid y, d_1, \theta) > 0$ if and only if d_2 attains Ω_2.

(*iii*) $\mu_1(d_1 \mid y) \geq 0$, $\sum_{d_1 \in \{A,B\}} \mu_1(d_1 \mid y) = 1$.
$\mu_1(d_1 \mid y) > 0$ if and only if d_1 attains Ω_1.

(*iv*) $\mu(d \mid y, \theta) \equiv \mu_1(d_1 \mid y)\mu_2(d_2 \mid y, d_1, \theta)$.

(*v*) $\pi(d, x)$ is consistent with $\mu(d \mid y, \theta)$ in the sense of Kreps and Wilson [1982*b*].

(*vi*) $w(v, d, x)$ satisfies (20), $\forall (v, d, x)$.

The above definition of equilibrium incorporates the requirements that the decision strategy of the manager is optimal given the wage responses of employing firms, wage offers are sequentially rational and competitive, and beliefs are consistent. Condition (*v*) requires that for all decisions that lie on the equilibrium path, $\pi(d, x)$ is obtained from $\mu(d \mid y, \theta)$ via Bayes' Rule, as described in (17) and (18). For those decisions that lie off the equilibrium path, $\pi(d, x)$ is required to be the limit of a sequence of beliefs derived from a sequence of completely mixed strategies that converge to the equilibrium strategy, $\mu(d \mid y, \theta)$. In (*vi*) we have made use of our earlier results from the managerial labor market.

To capture precisely what it means for escalation errors to occur in our economy, we use the following definitions.

(*i*) Information is *disconfirming* when the manager selects project A in stage one and subsequently discovers that the state of nature is θ_B, or selects B in stage one and subsequently discovers θ_A.

(*ii*) Conversely, information is *confirming* when the discovered state is consistent with the project chosen in stage one.

(*iii*) The manager *escalates* when, at stage two, he continues with the project chosen in stage one. The manager *switches* when the project chosen at stage two is different from the project chosen at stage one.

(*iv*) Escalation errors do not occur when the manager's equilibrium decision strategy assigns probability 1 to escalation in the presence of confirming information and probability 0 to escalation in the presence of disconfirming information.

(*v*) Escalation errors occur when the manager's equilibrium decision strategy assigns probability 1 to escalation when information is confirming and strictly positive probability to escalation when information is disconfirming.

5. *Equilibrium Decision Rules*

First we characterize equilibrium decision rules in a world in which the information signals received by the manager are also publicly observed. In such a world (d, x) would contain no information on talent, since all of this information has already been inferred from y. The manager's decisions in the current period have no impact on his reputation and therefore on future expected wages from employment. Hence, his optimal decision strategy maximizes current-period expected income.

It is easy to verify that current-period expected income is maximized by the decision vector AA in response to signal y_A, and BB in response to y_B. When y_0 is received the manager would be indifferent between A and B, and would choose between them randomly. If subsequent information were confirming, the manager would escalate. On the other hand, if subsequent information were disconfirming, the manager would switch. The latter claim follows from assumption (14).

The above decision rules are referred to as "first-best." Let δ be the probability of choosing project A in stage one contingent on signal y_0. A family of first-best decision rules is generated by varying δ over the closed interval [0,1].

We turn now to the private-information world, in which the signals observed by the manager in stages one and two are not publicly observed. In such a world (d, x) is potentially informative with respect to managerial talent. Therefore, the manager will be concerned with the reputation value of his decisions. We show that privacy of information and the resultant concern for reputation are the key factors that result in escalation error.

Consider the family of first-best decision rules index by $\delta \in [0,1]$. Any particular member of this family will be denoted $\mu_*(\delta)$. First, we show that from this family of first-best decision rules the only candidate for equilibrium is $\mu_*(0.5)$. All other first-best decision rules can be ruled out from further consideration.

THEOREM 1. No first best decision rule, $\mu_*(\delta)$, other than $\mu_*(0.5)$, can be an equilibrium strategy in the private-information economy.

A formal proof is available from the authors on request. The intuition underlying the above result is as follows. If, contingent on y_0, the manager assigns greater probability to A than to B, then the decision vector AA makes it more likely that the manager received y_0 than does the decision vector BB. Since y_0 is bad news, this makes the reputation value of AA smaller than the reputation value of BB. This, in turn, makes the manager prefer B to A, contingent on y_0, which is a deviation from the candidate decision rule. For a first-best decision rule to be an equilibrium strategy, it is necessary (but not sufficient) that the reputation values of

AA and BB are equal; this requires that projects A and B be treated in a completely symmetrical fashion, which occurs only when $\delta = 0.5$.

Our search for equilibrium decision rules, in this private-information economy, is restricted to the following two questions:

(*i*) What are necessary and sufficient conditions under which $\mu_*(0.5)$ is an equilibrium decision rule in the private-information economy? Since this is the only equilibrium from the public-information economy that could possibly survive in the private-information economy, these conditions would determine when escalation errors do not occur.

(*ii*) When $\mu_*(0.5)$ is not an equilibrium decision rule in the private-information economy, are there equilibria that exhibit escalation errors? (Note that we are not seeking to characterize all possible equilibria, but only those that exhibit escalation errors.)

To answer the above questions, we analyze a family of decision rules that includes $\mu_*(0.5)$ as a limiting case. Each member of this family differs from other members only in the degree to which escalation errors are manifest. Hereafter the notation $\mu_*(0.5)$ will be abbreviated to μ_*. The family of decision rules to be analyzed is described as follows. When y_A is observed, the manager chooses project A in stage one and escalates in stage two. When y_B is observed, the manager chooses B and escalation. When y_0 is observed, the manager chooses between A and B with equal probability. If subsequent information is confirming, the manager escalates. If subsequent information is disconfirming, the manager switches projects with probability ϵ and escalates with probability $1-\epsilon$.

A family of decision rules is generated by varying ϵ over the closed interval, $[0,1]$. Note that $\epsilon = 1$ defines the first-best decision rule earlier identified as μ_*. The smaller the value of ϵ, the greater the severity with which escalation errors are manifest. When $\epsilon = 0$, the phenomenon is manifest in its strongest form since project switches never occur in the presence of disconfirming information. We will establish that there always exists an equilibrium that lies in this family of decision rules, and within this family equilibrium is unique. A member of this family will be identified by the notation μ_ϵ. Figure 3 summarizes this decision rule.

5.1 CONSISTENT BELIEFS FOR THE ABOVE FAMILY OF DECISION RULES

To analyze this family of decision rules, first calculate beliefs given μ_ϵ. For $\epsilon > 0$, each decision vector $d \in D$ has positive probability, so beliefs can be calculated from Bayes' Rule. Calculating posterior probabilities from (17) and (19), for $\epsilon > 0$, yields:

$$\pi(AA,\, x_H;\, \mu_\epsilon) = \pi(BB,\, x_H;\, \mu_\epsilon) = \left[1 + q_H(\epsilon)\,\frac{(1-\rho)}{\rho}\right]^{-1}, \qquad (23)$$

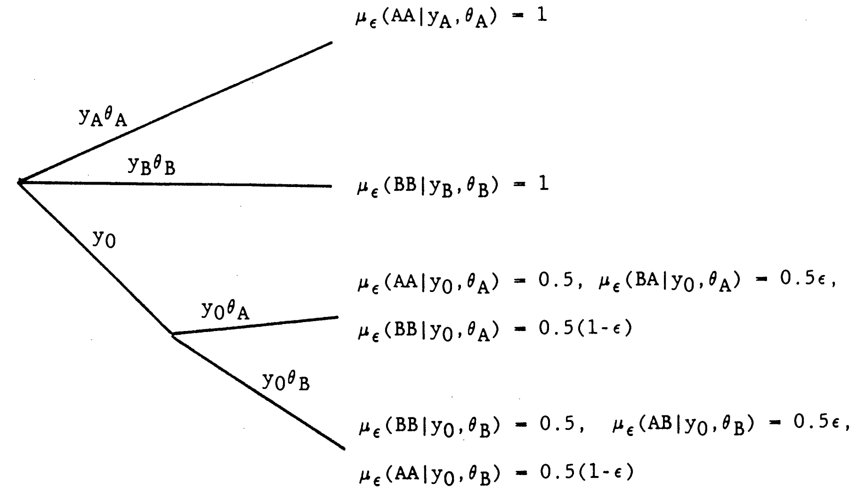

FIG. 3

where:

$$q_H(\epsilon) = \frac{P_m - 0.5P(y_0\,|\,z_N)[P_m - P_u(1-\epsilon)]}{P_M - 0.5P(y_0\,|\,z_T)[P_m - P_u(1-\epsilon)]}, \tag{24}$$

$$\pi(AA,\, x_L;\, \mu_\epsilon) = \pi(BB,\, x_L;\, \mu_\epsilon) = \left[1 + q_L(\epsilon)\,\frac{(1-\rho)}{\rho}\right]^{-1}, \tag{25}$$

where:

$$q_L(\epsilon) = \frac{(1-P_m) - 0.5P(y_0\,|\,z_N)[(1-P_m) - (1-P_u)(1-\epsilon)]}{(1-P_m) - 0.5P(y_0\,|\,z_T)[(1-P_m) - (1-P_u)(1-\epsilon)]}, \tag{26}$$

and:

$$\pi(AB,\, x;\, \mu_\epsilon) = \pi(BA,\, x;\, \mu_\epsilon) = \left[1 + \frac{P(y_0\,|\,z_N)}{P(y_0\,|\,z_T}\,\frac{(1-\rho)}{\rho}\right]^{-1} = \psi(y_0), \tag{27}$$

for each $x \in \{x_H,\, x_L\}$.

Note from (27) that the manager's reputation conditional on switching is independent of ϵ and x. The intuition underlying this is that any observed switch in projects reveals unambiguously that the state of nature was discovered late. Hence, the probability of switching and the realized value of x are irrelevant.

Now consider beliefs given the manager's decision rule when $\epsilon = 0$. Call this decision rule μ_0. Given μ_0, beliefs conditional on observing a switch cannot be calculated from Bayes' Rule, since switches have zero probability. To get around this problem, beliefs given μ_0 are determined as the limit of a sequence of beliefs derived from a sequence of completely mixed strategies that converges to μ_0, as required in Kreps and Wilson

[1982*b*]. It can be shown that the following beliefs satisfy the definition of consistency:

$$\pi(AB, x; \mu_0) = \pi(BA, x; \mu_0) = \psi(y_o), \qquad \text{for each } x \in \{x_H, x_L\}. \tag{28}$$

$$\pi(AA, x_H\ \mu_0) = \pi(BB, x_H; \mu_0) = \left[1 + q_H(0)\,\frac{(1-\rho)}{\rho}\right]^{-1} \tag{29}$$

where q_H (0) is calculated from (24) when $\epsilon = 0$.

$$\pi(AA, x_L; \mu_0) = \pi(BB, x_L; \mu_0) = \left[1 + q_L(0)\,\frac{(1-\rho)}{\rho}\right]^{-1}, \tag{30}$$

where $q_L(0)$ is calculated from (26) when $\epsilon = 0$.

The above calculations show that for any given decision rule, μ_ϵ, $\epsilon \geq 0$, posterior beliefs depend primarily on whether switching or escalation was observed. To streamline notation, let S represent the event "switch," and E the event "escalation." Define:

$$\pi(S) \equiv \pi(AB, x; \mu_\epsilon) = \pi(BA, x; \mu_\epsilon)\ \forall \epsilon \geq 0.$$

As shown above, $\pi(S)$ is a constant independent of x and ϵ. Also define:

$$\pi_\epsilon(E, x) \equiv \pi(AA, x; \mu_\epsilon) = \pi(BB, x; \mu_\epsilon),\ \forall \epsilon \geqslant 0.$$

Theorem 2, below, characterizes those properties of beliefs that are important for derivation of equilibrium decision rules.

THEOREM 2.

(*i*) Given any μ_ϵ, $0 \leq \epsilon \leq 1$, $\pi(S) < \pi_\epsilon(E, x_L) \leq \pi_\epsilon(E, x_H)$. The last inequality is strict for every $\epsilon < 1$.

(*ii*) $\pi_\epsilon(E, x_L)$ and $\pi_\epsilon(E, x_H)$ are strictly increasing and continuous in ϵ over the closed interval, [0, 1].

Proof (sketch). The first part of the theorem is proved by showing that $q_L(\epsilon) > q_H(\epsilon)$ for each $\epsilon < 1$. To do this employ the definitions of $q_L(.)$ and $q_H(.)$ and use the facts that $P_m > P_u$ and $P(y_0 \mid z_N) > P(y_0 \mid z_T)$. Next, $\pi(S) < \pi_\epsilon(E, x_L)$ follows from $P(y_0 \mid z_N) > P(y_0 \mid z_T)$. To establish that $\pi_\epsilon(E, x)$ is strictly increasing in ϵ, differentiate $q_H(\epsilon)$ and $q_L(\epsilon)$ with respect to ϵ and use $P(y_0 \mid z_N) > P(y_0 \mid z_T)$ to show that the derivatives are negative. A formal proof is available from the authors, on request.

Since the equilibrium wage function $w(v_H, d, x)$ depends on (d, x) only through beliefs conditional on (d, x), we shall henceforth use the notation $w(v_H\ \pi(S))$ and $w(v_H, \pi_\epsilon(E, x))$ to denote equilibrium wage contracts. The strict monotonicity of w in π, together with Theorem 2(*i*) implies that given any decision rule μ_ϵ, $\epsilon \geq 0$:

$$w(v_H, \pi(S)) < w(v_H, \pi_\epsilon(E, x_L)) \leq w(v_H, \pi_\epsilon(E, x_H)). \tag{31}$$

A decision rule μ_ϵ is an equilibrium decision rule if μ_ϵ is the manager's best response to the wage contracts implied by beliefs $\pi(S)$ and $\pi_\epsilon(E, x)$. To determine whether μ_ϵ is an equilibrium decision rule, consider the manager's decision problem in the presence of confirming and disconfirming information. Given confirming information, his decision problem has the form:

$$\Omega_2(y_0, A, \theta_A) = \text{Max}\left\{\sum_x xP(x \mid \theta_A, A) + P(v_H \mid y_0) \sum_x P(x \mid \theta_A, A) w(v_H, \pi_\epsilon(E, x)),\right.$$

$$\left.\sum_x xP(x \mid \theta_A, B) - C(AB) + P(v_H \mid y_0) w(v_H, \pi(S))\right\}.$$

Clearly, current-period expected income is bigger from escalation than from switching. For each $\epsilon \geq 0$, (31) implies that future expected wages from employment are also bigger if the manager escalates than if he switches. Therefore, for each $\epsilon \geq 0$, escalation is unambiguously preferred. Similar results hold when y_A or y_B is observed in stage one.

Thus, the sole factor that determines whether μ_ϵ is an equilibrium is the manager's behavior in the presence of disconfirming information. Suppose the manager observed y_0 in stage one, chose project A, and subsequently discovered that the state was θ_B. The manager's decision problem is described by:

$$\Omega_2(y_0, A, \theta_B) = \text{Max}\left\{\sum_x xP(x \mid \theta_B, B) - C(AB) + P(v_H \mid y_0) w(v_H, \pi(S)),\right.$$

$$\left.\sum_x xP(x \mid \theta_B, A) + P(v_H \mid y_0) \sum_x P(x \mid \theta_B, A) w(v_H, \pi_\epsilon(E, x))\right\}.$$

The above maximization is equivalent to:

$$\begin{aligned} \text{Max } \{&(1 - P_m)x_L + P_m x_H - C(AB) + P(v_H \mid y_0) w(v_H, \pi(S)), \\ &(1 - P_u)x_L + P_u x_H + P(v_H \mid y_0)[(1 - P_u) w(v_H, \pi_\epsilon(E, x_L)) \\ &+ P_u w(v_H, \pi_\epsilon(E, x_H))]\}. \end{aligned}$$

Rearranging and canceling common terms yields:

$$\text{Max}\{J, F(\epsilon)\}, \text{ where:}$$

$$J \equiv (x_H - x_L)(P_m - P_u) - C(AB), \text{ and}$$

$$\begin{aligned} F(\epsilon) \equiv P(v_H \mid y_0)[(1 - P_u) w(v_H, \pi_\epsilon(E, x_L)) \\ + P_u w(v_H, \pi_\epsilon(E, x_H)) - w(v_H, \pi(S))]. \end{aligned}$$

When $J > F(\epsilon)$, the manager prefers switching to escalation and when $J < F(\epsilon)$, the manager would escalate rather than switch. When $J = F(\epsilon)$, the manager is indifferent between switching and escalation and would randomly choose one or the other.

Theorem 3, below, establishes existence and uniqueness of equilibrium in the family of decision rules under consideration. The theorem also provides necessary and sufficient conditions under which escalation errors will occur, and the severity with which they occur.

THEOREM 3. There is a unique equilibrium decision rule in the family μ_ϵ, $0 \leq \epsilon \leq 1$.

(*i*) $J \geq F(1) => \mu_*$ is the unique equilibrium decision rule; escalation errors do not occur.

(*ii*) $J \geq F(0) => \mu_0$ is the unique equilibrium decision rule; escalation errors occur in their severest form.

(*iii*) $F(0) < J < F(1) =>$ there exists some unique ϵ, $0 < \epsilon < 1$, such that μ_ϵ is an equilibrium decision rule; escalation errors occur with positive probability.

Proof. From Theorem 2, $\pi_\epsilon(E, x_L)$ and $\pi_\epsilon(E, x_H)$ are continuous and strictly increasing in ϵ, and $\pi(S)$ is a constant independent of ϵ. Also, $w(v_H, \pi)$ is continuous and strictly increasing in π. Therefore $F(\epsilon)$ is continuous and strictly increasing in ϵ. Since $F(\epsilon)$ is strictly increasing:

$$J \geq F(1) => J > F(\epsilon) \text{ ; } \epsilon \text{ satisfying } 0 \leq \epsilon < 1.$$

In this case μ^* is the unique equilibrium decision rule:

$$J \leq F(0) => J < F(\epsilon) \text{ ; } \epsilon \text{ satisfying } 0 < \epsilon \leq 1.$$

In this case μ_0 is the unique equilibrium. Now consider the remaining case where $F(0) < J < F(1)$. Since $F(.)$ is continuous over (0, 1), it follows from the Intermediate Value Theorem that there exists an ϵ, $0 < \epsilon < 1$, such that $J = F(\epsilon)$. Further, since $F(.)$ is strictly increasing, for each $J \in (F(0), F(1))$, there is a unique ϵ satisfying $J = F(\epsilon)$. In this case the manager is indifferent between switching and escalation in the face of disconfirming information. Hence, switching with probability ϵ is an optimal response, which makes μ_ϵ the unique equilibrium decision rule. This completes the proof of the theorem.

The necessary and sufficient conditions described in Theorem 3 have a simple and intuitive interpretation. When the manager switches projects in the presence of disconfirming information, he maximizes his expected current-period income. J measures the increase in current-period income from switching as opposed to escalation. However, a switch reveals for sure that he discovered the state of nature late. Since this is more likely to happen for the untalented than the talented manager, such revelation damages the manager's reputation. $F(\epsilon)$ measures the loss in future expected wages from such damage to reputation. When this loss is sufficiently large, first-best decision rules fail, and escalation errors occur with positive probability.

If $G(z_T)$ is large relative to $G(z_N)$, $F(1)$ will be large and (*i*) of Theorem 3 would likely be violated. Thus, escalation errors occur when managerial talent makes a big difference to firms that seek to recruit managers. This results in considerably higher wages for managers who are evaluated more highly in the managerial labor market. In turn, this big differential in offered wages makes managers very sensitive to "looking good or bad."

Switching projects makes them "look bad" since it communicates lack of foresight, and therefore managers refrain from switching even at the cost of sacrificing current-period income.

Theorem 3 also suggests that escalation errors will not occur when the prior probability that the manager is talented, ρ, is very close to 0 or 1. In these cases, the difference between the posterior probabilities, $\pi(E, x)$ and $\pi(S)$ will be so small that $F(1)$ will be small even if $G(z_T)$ is much larger than $G(z_N)$. Again, this result is quite intuitive and consistent with casual observation. When it is almost certain that the manager is talented (or untalented) his reputation is no longer sensitive to new information, and therefore he is not so concerned about "looking good or bad." The gains from manipulating reputation are small and more than offset by the loss in current income.

6. *Conclusion*

In this paper we have used an aspect of human capital (foresight) to show how escalation errors could occur in a world of rational decision makers. While we do not deny the importance of psychological factors, we have deliberately avoided using them in order to show that alternative explanations of escalation errors can be formulated that rely solely on economic rationality. The main ingredients of our explanation are (*i*) information on the desirability of switching is private to the decision maker; (*ii*) this information is also related to the unobservable talents of the decision maker; (*iii*) these talents are inferred by others in society from observation of the decision maker's actions; and (*iv*) these inferences impact the future opportunities of the decision maker.

The last ingredient above is crucial to our analysis and merits further comment. Its existence requires either an inability or a lack of incentives for firms to precommit to ignore information revealed by managers' switching decisions. If such commitments were possible, escalation errors could be avoided. In our model, the incentive to make such commitments is lacking since the costs of not switching are borne entirely by the manager. However, even if such incentives were present at some point in time, it is unrealistic to assume that firms would continue to ignore information on their managers' talents for all future time. We do not know of a single firm with binding contracts that specify probabilities of promotion to various hierarchical positions contingent on long histories of actions and outcomes. For reasons not fully understood, contracts in real organizations are typically incomplete. (See Kreps [1984] for an exploratory discussion of incomplete contracts and their consequences.) Although our analysis formally applies to managers who are currently not under contract or who aspire to move from one firm to another, it is clear that if firms can only partially commit, our analysis would apply to intrafirm settings as well.

What implications does our model have for data collection? Our model suggests that documented cases of escalation errors should consist only of cases where the decision maker has been shielded from reputation effects. This could happen when the researcher has guaranteed anonymity or the data are collected long after their reputational impact has deteriorated. An alternative way to obtain data on the phenomenon would be to conduct a controlled experiment in which the frequency of switching can be predicted by the researcher.

We conclude by indicating an extension of our model that would capture the idea that incentives for switching or escalation would vary with the magnitude of sunk costs. Suppose that implementation of projects requires more than two stages. The state of nature could be discovered at any stage, but the later it is discovered the more likely it is that the manager is untalented. In this case, a switch at a later date would damage the manager's reputation more than a switch at an early date. At the same time, sunk costs would be bigger at the later date since the project would be more complete. Thus, if disconfirming information were received in the early stages of implementation, when sunk costs are small, the manager would switch. Alternatively, if information were received in later stages, when sunk costs are large, the manager would not switch.

REFERENCES

KANODIA, C., R. BUSHMAN, AND J. DICKHAUT. "Sunk Costs: Is There a Fallacy?" Working paper, University of Minnesota, January 1984.

KREPS, D. "Corporate Culture and Economic Theory. " Working paper, Stanford University, May 1984.

———, AND R. WILSON. "Reputation and Imperfect Information." *Journal of Economic Theory* 27 (1982*a*): 253–79.

———. "Sequential Equilibria." Econometrica (July 1982*b*): 863–94.

LAUGHHUNN, D., AND J. W. PAYNE. "The Impact of Sunk Outcomes on Risky Choice Behavior." *Canadian Journal of Operational Research and Information Processing* (forthcoming).

MILGROM, P., AND J. ROBERTS. "Predation, Reputation and Entry Deterrence." *Journal of Economic Theory* 27 (1982): 280–312.

ROTHSCHILD, M., AND J. E. STIGLITZ. "Equilibrium in Competitive Insurance Markets: An Essay in Economics of Imperfect Information." *Quarterly Journal of Economics* (November 1976): 629–49.

STAW, B. M. "Knee Deep in the Big Muddy: A Study of Escalating Commitment to a Chosen Course of Action." *Organizational Behavior and Human Performance* 16 (1976): 26–44.

———. "The Escalation of Commitment to a Course of Action." *Academy of Management Review* 6, no. 4 (1981): 577–87.

———, AND J. ROSS. "Behavior in Escalation Situations: Antecedents, Prototypes, and Solutions." In *Research in Organizational Behavior*, edited by S. Cummings and B. Staw. Greenwich, Conn.: JAI Press, 1986.

THALER, R. "Towards a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization* (March 1980): 39–60.