

Repeated Principal-Agent Games with Discounting

Author(s): Roy Radner

Source: *Econometrica*, Vol. 53, No. 5 (Sep., 1985), pp. 1173-1198

Published by: [The Econometric Society](#)

Stable URL: <http://www.jstor.org/stable/1911017>

Accessed: 07/12/2010 13:16

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=econosoc>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



The Econometric Society is collaborating with JSTOR to digitize, preserve and extend access to *Econometrica*.

REPEATED PRINCIPAL-AGENT GAMES WITH DISCOUNTING

BY ROY RADNER¹

In a repeated principal-agent game (supergame) in which each player's criterion is his long-run average expected utility, efficient behavior can be sustained by a Nash equilibrium if it is Pareto-superior to a one-period Nash equilibrium. Furthermore, if the players discount future expected utilities, then for every positive epsilon, and every pair of discount factors sufficiently close to unity (given epsilon), there exists a supergame equilibrium that is within epsilon (in normalized discounted expected utility) of the target efficient behavior. These supergame equilibria are explicitly constructed with simple "review strategies."

1. INTRODUCTION

1.1. *Some Background*

IN A PRINCIPAL-AGENT SITUATION, the agent chooses an action "on behalf of" the principal. The resulting consequence depends on a random state of the environment as well as on the agent's action. After observing the consequence, the principal makes a payment to the agent according to a pre-announced reward function, which depends directly only on the observed consequence. This last restriction expresses the fact that the principal cannot directly observe the agent's action, nor can the principal observe the information on which the agent bases his action. This situation is one of the simplest examples of decentralized decision-making in which the interests of the decision-makers do not coincide.²

If this action-reward situation occurs only once, I shall call it a *short-run principal-agent relationship*. The situation can be naturally modeled as a two-move game, in which the principal first announces a reward function to the agent, and then the agent chooses an action (or decision function if he has prior information about the environment).

The Nash (or perfect Nash) equilibria of such a game are typically inefficient (unless the agent is neutral towards risk), in the sense that there will typically be another (but nonequilibrium) reward-decision pair that yields higher expected utilities to both players.

In order to increase the efficiency of short-run equilibria, the principal could monitor (at least ex post) the information and decision of the agent. However such monitoring would typically be costly, so that *net* efficiency need not be increased by monitoring.

Another approach to increasing efficiency is suggested by the theory of repeated games. If a game with two or more players is repeated, the resulting situation can be modeled naturally as a game ("supergame") in which the players' actions in any one repetition are allowed to depend on the history of the previous repetitions. In the principal-agent situation, the repetition of the game would

¹ I am grateful to R. A. Aumann, R. W. Rosenthal, and A. Rubinstein for helpful discussions of the topic of this paper, and to A. Rubinstein, J. Mirrlees, and the referees for comments on a previous draft. The views expressed here are those of the author, and do not necessarily reflect the views of AT&T Bell Laboratories.

² The references relevant to the Introduction are gathered in the Bibliographic Notes in Section 9.

give the principal an opportunity to observe the results of the agent's actions over a number of periods, and use some statistical test to infer whether or not the agent was choosing the appropriate actions. The repetition of the game would also provide the principal with opportunities to "punish" the agent for apparent departures from the appropriate actions. Thus, roughly speaking, the principal could employ the analogue of a "statistical quality control chart" to deter "cheating" by the agent. However, since the accumulation of reliable statistical evidence takes time, the threat of future punishment would be a less effective deterrent the more the agent discounts future utility.

A formal analysis suggested by the preceding intuitive ideas is carried out in the present paper, for the case in which the game is repeated infinitely often. I shall show that the less the players discount future utility the closer they can approach efficiency with equilibria of the supergame. These equilibria can all be achieved by a family of relatively simple strategy-pairs that I shall call *review strategies*. Roughly speaking, in a review strategy the principal periodically evaluates the cumulative performance of the agent since the last review. If a review results in a satisfactory evaluation, a new review phase is begun; if not, the players enter a penalty phase, after which a new review phase is begun. During each entire review phase the principal pays the agent according to the target efficient reward function. During the penalty phases the players revert to the short-run equilibrium. A particular review strategy is characterized by the lengths of the review and penalty phases, and by the criterion for satisfactory performance at the times of review.

I should emphasize that, by definition, the equilibrium strategy pairs are self-enforcing, and thus do not rely on any binding contracts or other pre-commitments. In particular, the agent induces the principal to follow an equilibrium review strategy by threatening to initiate a phase of myopic optimization (short-run equilibrium) following any departure by the principal from the target reward function during a review phase. Without such a threat, the principal might be tempted to shorten a particular review phase if, for example, the agent had already attained a very high cumulative performance and hence could safely "coast" to the end of the current review phase with minimal effort.

1.2. *Summary of the Main Results*

Let u^* and v^* be the one-period expected utilities of the principal and agent, respectively, corresponding to an inefficient short-term equilibrium, and let \hat{u} and \hat{v} be respective one-period expected utilities corresponding to an efficient reward-decision pair that is more efficient than the short-run equilibrium (i.e. $\hat{u} > u^*$ and $\hat{v} > v^*$); such an improvement will always exist in the present model. Let γ and δ be the players' respective discount factors; then for every $\varepsilon > 0$ there will exist γ_ε and δ_ε less than 1 such that for each (γ, δ) with $\gamma \geq \gamma_\varepsilon$ and $\delta \geq \delta_\varepsilon$ there exists an equilibrium of the corresponding supergame that yields the players' (normalized) discounted expected utilities at least $\hat{u} - \varepsilon$ and $\hat{v} - \varepsilon$, respectively. In particular, there will be critical discount factors above which there exist supergame equilibria that are strictly more efficient than (u^*, v^*) .

In the remainder of this introductory section I shall discuss some aspects of these results more fully, before proceeding to a more formal presentation and analysis of the model. For this discussion, however, some minimum of formal notation will be helpful. To simplify the exposition, suppose for the time being that both players have the same discount factor, say δ ($0 \leq \delta < 1$). If, for a given pair of supergame strategies, the principal's expected utility in period t is u_t , then his normalized discounted expected utility for the supergame is defined to be

$$u = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} u_t;$$

a corresponding formula defines the agent's supergame payoff, say v . For each δ , let $W(\delta)$ denote the set of pairs (u, v) of normalized discounted expected utilities of the players corresponding to equilibria of the supergame.

Let \hat{W} denote the set of efficient one-period expected utility pairs. The first main result of the paper can be paraphrased as follows: for every pair (\hat{u}, \hat{v}) in \hat{W} that is superior to (u^*, v^*) , i.e., $\hat{u} > u^*$ and $\hat{v} > v^*$, one can get arbitrarily close to (\hat{u}, \hat{v}) with points in $W(\delta)$, by taking δ sufficiently close to 1. In addition, an explicit construction of such "approximately efficient" supergame equilibrium strategies is given, namely the "review strategies" described above.

1.3. Multiplicity of Equilibria of the Supergame

For every pair of the players' discount factors, the corresponding supergame equilibrium is not unique, provided the discount factors are not too small. For example, the pair of supergame strategies in which each player stubbornly sticks to his short-term-equilibrium strategy is a supergame equilibrium, whatever the players' discount factors. Thus, for each δ , (u^*, v^*) is in $W(\delta)$. In addition, there will be many equilibria in review strategies, as described above. Indeed, it can be shown that for sufficiently large discount factors the set of equilibria has the cardinality of the continuum. Following standard terminology, I shall call the mapping from δ to $W(\delta)$ the *equilibrium utility correspondence*.

For a fixed discount factor δ , a utility-pair in $W(\delta)$ is called *second-best-efficient* if there is no other utility-pair in $W(\delta)$ that is at least as large in each coordinate and strictly larger in one. It would be of interest to characterize such second-best-efficient utility-pairs and the associated supergame equilibrium strategies, but I have not attempted to do so in this paper. It does not appear that the review strategies alluded to above are second-best-efficient, even though they are approximately efficient (and therefore approximately second-best-efficient) for discount factors close to unity. This point is discussed more fully in Section 8.

1.4. The Case of No Discounting

It is natural to try to define a "limit supergame" as the discount factor δ approaches 1. One way to do this is suggested by Abel's Theorem, which states that

$$\lim_{\delta \rightarrow 1} (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} u_t = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T u_t,$$

provided the right-hand limit exists. Thus, for the case in which $\delta = 1$, define the principal's supgame payoff to be

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T u_t,$$

which is well-defined and finite for every bounded sequence (u_t) , and define the agent's supgame payoff analogously. It can be shown (see Section 9) that, for the case $\delta = 1$, the corresponding set $W(1)$ of supgame equilibria contains every efficient pair in \hat{W} that is superior to a short-run equilibrium pair (u^*, v^*) . In other words, if the players do not discount the future at all, then they can attain *exact* efficiency with supgame equilibria. On the other hand, there will also be inefficient supgame equilibrium payoff pairs, e.g., (u^*, v^*) .

1.5. A Continuity Property

Taking the two preceding results together, we see that the equilibrium payoff correspondence $W(\cdot)$ has a continuity-like property, with respect to efficient payoff pairs, at $\delta = 1$. This property of the repeated principal-agent game is apparently not shared by more general repeated games under uncertainty (see Bibliographic Notes, Section 9). Because of this ability to exploit long-term relationships to increase efficiency, the principal-agent mechanism may be particularly important in decentralized organizations. The implications of this for organization theory will be explored elsewhere.³

1.6. Outline of the Paper

Section 2 defines the one-period principal-agent game and reviews those of its properties that are relevant to the present paper. Section 3 defines the repeated game, and Section 4 describes a family of review strategies for the repeated game. In Section 5 the discounted expected utilities of such strategies are derived. In Section 6, as a preparation for the first main result on equilibria, I derive information about the agent's optimal response to a review strategy of the principal; in particular I derive lower bounds on the players' expected utilities corresponding to such an optimal response by the agent. Section 7 gives the main result on equilibria for the discounting case, and in particular shows that the principal's optimal response to a review strategy is itself a review strategy.

To make the exposition less abstract and thus appeal to the reader's imagination, the argument in Sections 2-7 is developed entirely in the context of a simple example in which the agent's action is a level of "effort," the stochastic consequence of the agent's action is either success or failure, and the principal is neutral towards risk. The extension of the argument to a more general model is sketched in Section 8 and in the Appendix.

³ For a partial treatment see [15], where the reader will also find other references to the literature on theories of economic decentralization and incentives.

References to the related literature, as well as other bibliographic notes, are gathered in Section 9.

2. THE ONE-PERIOD GAME

I shall start with a description of the one-period game. First, the principal announces a reward function, ω , which is a pair of numbers (w_0, w_1) . Second, the agent chooses an action, a , which is a real number. Third, there is a consequence, which is a random variable taking on the values 1 (“success”) or 0 (“failure”); the probability of success is an increasing function of the agent’s action. Finally, the agent receives a monetary reward, w_1 or w_0 , according as the consequence of his action is success ($C = 1$) or failure ($C = 0$); the principal receives the remainder, $C - w_C$. Without loss of generality, one can take the agent’s action to be the probability of success, i.e.

$$\text{Prob}(C = 1) = a,$$

$$\text{Prob}(C = 0) = 1 - a.$$

Assume that the resulting utility to the principal is

$$U = C - w_C,$$

and the utility to the agent is

$$V = P(w_C) - Q(a),$$

where P and Q are differentiable and strictly increasing, P is strictly concave, and Q is strictly convex. Make the convention that

$$P(0) = Q(0) = 0.$$

It is typically realistic to impose two constraints on the rewards. The first constraint is that the principal may not impose arbitrarily large penalties on the agents; in other words, the rewards are constrained so that the agent’s disutility is bounded from below. The second constraint expresses the condition that the agent is free to refuse to enter into the relationship (i.e., to play the game). For this, the rewards w_0 and w_1 must be such as to enable the agent to achieve some minimum *expected* utility. For the purposes of this paper, it is sufficient to impose a constraint of the first type; the addition of the second constraint would slightly complicate the exposition, but would not change the results in any essential way. To express the first constraint, one can assume that the rewards are bounded below (and that the function P is finite everywhere); without loss of generality I assume that they are nonnegative:

$$\omega \equiv (w_0, w_1) \geq 0.$$

Note that it has been assumed that the principal is neutral towards risk, whereas the agent is averse to risk.

In this game, the principal's pure strategy is the reward function, ω , and the agent's pure strategy is a mapping, α , from reward functions to actions:

$$a = \alpha(\omega).$$

An equilibrium of the game is a pair of strategies, (ω^*, α^*) , such that (i) ω^* maximizes the principal's expected utility, $u = EU$, given that the agent uses α^* , and (ii) $\alpha^*(\omega^*)$ maximizes the agent's expected utility, $v = EV$, given ω^* . In this paper I shall consider only *perfect* equilibria, in which, for *every* ω (not just ω^*), $\alpha^*(\omega)$ is an optimal action for the agent given ω . Thus "equilibrium" is henceforth to be understood here as "perfect equilibrium." (See Section 8.3.)

It is not necessary for the purpose of this paper to give a complete analysis of the one-period game. However, an understanding of a few aspects of the game will clarify the main issues that are addressed in subsequent sections.⁴ Given the reward function ω , if the agent chooses the action a his expected utility will be

$$(2.1) \quad v = aP(w_1) + (1-a)P(w_0) - Q(a),$$

and the principal's expected utility will be

$$u = a(1 - w_1) - (1 - a)w_0.$$

The principal can achieve at least $u = 0$ by taking $w_1 = w_0 = 0$. One can easily verify from (2.1) that if $w_0 = w_1$ the agent will have no incentive to work, i.e., $\alpha^*(w, w) = 0$. In addition, one sees from (2.1) that if

$$Q'(0) \leq P(1),$$

then $\alpha^*(w_0, w_1) = 0$ for all w_0 and w_1 between 0 and 1; in this case the only equilibrium has $\omega^* = (0, 0)$ and $a^* = 0$. On the other hand, if

$$Q'(0) > P(1),$$

then the equilibrium is characterized by

$$(2.2) \quad 0 = w_0^* < w_1^* < 1, \\ a^* > 0;$$

also, $\alpha^*(0, w_1)$ is strictly increasing in w_1 whenever $\alpha^*(0, w_1)$ is strictly between 0 and 1. This is the case I shall discuss from now on.

A pair $(\hat{\omega}, \hat{a})$ is *efficient* (Pareto optimal) if no other pair (ω, a) yields each player as much expected utility and at least one player strictly more. It is easy to see that, for the same level of effort, a , the agent prefers the compensation function (\bar{w}, \bar{w}) to the compensation function (w_0, w_1) , where

$$\bar{w} = aw_1 + (1-a)w_0,$$

⁴ Since the main focus of the present paper is the repeated game, it is neither necessary nor appropriate here to give a thorough and rigorous treatment of the one-period game. Most of the facts about the latter that are alluded to here are discussed more systematically in [2] and in the references cited there.

whereas the principal is indifferent between the two (recall that the agent is risk-averse and the principal is risk-neutral). Hence, if $[(\hat{\omega}_0, \hat{\omega}_1), \hat{a}]$ is efficient, then $\hat{\omega}_0 = \hat{\omega}_1$. Together with (2.2) this shows that an equilibrium is not efficient.

There are, of course, many efficient pairs $[(\hat{\omega}, \hat{\omega}), \hat{a}]$; one can verify that for $0 < \hat{a} < 1$ they are characterized by the condition

$$P'(\hat{\omega}) = Q'(\hat{a}).$$

In summary, we shall be concerned with an equilibrium (ω^*, α^*) of the one-period game that is inefficient, and for which there is no Pareto-superior equilibrium. Since (ω^*, α^*) is not efficient, it follows from the structure of the one-period game that there is a pair $(\hat{\omega}, \hat{a})$ that is efficient and is strictly better than (ω^*, α^*) for both players. Thus let \hat{u} and \hat{v} be the respective one-period expected utilities yielded by $(\hat{\omega}, \hat{a})$; then $\hat{u} > u^*$ and $\hat{v} > v^*$.

3. THE REPEATED GAME

I shall now describe the infinitely-repeated game, or *supergame*. Roughly speaking, during each period the principal and agent play a one-period game, with a new random environment each time. Each period each player's action can depend on what he has observed up to that point in time, his *information history*. For the principal, this is the history of his own previous actions (i.e., announced reward-pairs), and the history of previous successes and failures. For the agent this is the history of his own and the principal's previous actions, the history of previous successes and failures, and the reward-pair that the principal has just announced. Neither player ever observes the random environments, which are assumed to be independent and identically distributed. At the end of each period, after having observed the current success or failure, the principal compensates the agent according to the reward-pair that he announced at the beginning of the period. A supergame strategy for a player is a sequence of decision-rules that determine his action at each period as a function of his information history at that point of time. The supergame payoff for a player is the normalized sum of his discounted expected one-period utilities.

Here is a more formal definition of the supergame. For $t = 1, 2, \dots$, ad inf., let ω_t be the reward function announced by the principal at the beginning of period t , let A_t be the action chosen by the agent in period t , let C_t be the corresponding realized consequence, and let $W_t = \omega_t(C_t)$ be the agent's reward. Define, for $t \geq 1$,

$$H_t^C = (C_1, \dots, C_t),$$

$$H_t^A = (A_1, \dots, A_t),$$

$$H_t^\omega = (\omega_1, \dots, \omega_t);$$

it is convenient to define H_0^C , A_0^A , and H_0^ω to be arbitrary one-element sets. The information available to the principal when he makes his decision in period t is

$$(3.1a) \quad I_t^C = (H_{t-1}^C, H_{t-1}^\omega),$$

and the information available to the agent when he makes his decision in period t is

$$(3.1b) \quad I_t'' = (H_{t-1}^C H_t^\omega, H_{t-1}^A).$$

The difference between I_t' and I_t'' expresses the assumption that the principal cannot learn the past actions of the agent.

A *supergame strategy* for the principal is a sequence $\sigma = (\sigma_t)$ in which σ_t maps the information I_t' into a reward function ω_t . Similarly, a supergame strategy for the agent is a sequence $\tau = (\tau_t)$ in which τ_t maps the information I_t'' into an action A_t . Assume that $C_t = 1$ or 0 , and that

$$\text{Prob}(C_t = 1 | I_t'', A_t) = A_t.$$

The realized utilities of the principal and agent, respectively, in period t are

$$U_t = C_t - W_t,$$

$$V_t = P(W_t) - Q(A_t).$$

The corresponding discounted expected utilities are

$$u(\gamma) = (1 - \gamma) \sum_{t=1}^{\infty} \gamma^{t-1} E U_t, \quad 0 \leq \gamma < 1,$$

$$v(\delta) = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} E V_t, \quad 0 \leq \delta < 1.$$

Since the realized utilities in each period are uniformly bounded, the discounted expected utilities are well-defined and finite for all the permissible discount factors γ and δ .

One can now define (Nash) equilibrium for the supergame in the usual way, namely, a *supergame equilibrium* is a pair of supergame strategies (one for each player) such that no player can increase his own supergame payoff by unilaterally changing his strategy.

One equilibrium of the supergame is the pair (σ^*, τ^*) in which, for all t and all information histories,

$$\sigma_t(I_t') = \omega^*,$$

$$\tau_t(I_t'') = \alpha^*(\omega_t),$$

where (ω^*, α^*) is the (inefficient) equilibrium of the one-period game that was described in Section 2. The discounted expected utilities yielded by this equilibrium are u^* and v^* . I shall say that the agent *optimizes myopically* during any period in which he uses α^* .

4. REVIEW STRATEGIES

As in Section 2, let (ω^*, α^*) be associated with an equilibrium of the one-period game, yielding one-period expected utilities u^* and v^* to the principal and agent

respectively, and let $(\hat{\omega}, \hat{a})$ be an efficient pair in the one-period game, with $\hat{\omega} = (\hat{w}, \hat{v})$, yielding respective one-period expected utilities \hat{u} and \hat{v} . Furthermore, suppose that $(\hat{\omega}, \hat{a})$ is more efficient than (ω^*, a^*) , so that $\hat{u} > u^*$ and $\hat{v} > v^*$. I shall now describe a class of sequential strategy pairs, (σ, τ) , from which equilibrium strategy-pairs in the supergame will be constructed; these strategy-pairs will be called *review strategies*. (I emphasize that the review-strategies that I construct will be equilibria in the space of *all* strategy-pairs, not just in the space of review strategies.)

Recall that $H_t^C = (C_1, \dots, C_t)$, and define $S_t = C_1 + \dots + C_t$. Recall that

$$\text{Prob}(C_t = 1 | H_{t-1}^C, A_t) = A_t,$$

$$\text{Prob}(C_t = 0 | H_{t-1}^C, A_t) = 1 - A_t.$$

Roughly speaking, a review strategy for the principal pays the agent the reward \hat{w} during periods 1 through R , whatever the agent's performance during this time, and then reviews the agent's cumulated performance, S_R . If S_R is large enough (the agent "passes the review"), then the process is repeated. If S_R is too small (the agent "fails the review"), then the principal uses the one-period equilibrium reward function ω^* for M periods, where M is a number to be determined; after period $(R + M)$ the process is repeated. Passing the review is defined by the condition

$$S_R \geq R\hat{a} - B,$$

where \hat{a} is the agent's efficient action, and B is also a (positive) parameter yet to be determined. The quantity $q = R\hat{a} - B$ may be interpreted as the agent's quota of successes during the review phase. Notice that if the agent were to use the action \hat{a} in each period, then the expected number of successes in R periods would be $R\hat{a}$, so that B can be interpreted as the "margin of error" in cumulated performance that is allowed by the principal at the time of review. Thus the principal's review strategy is determined by the parameters R (the length of a review phase), B (the allowable margin of error at review), and M (the length of the "penalty phase"). Notice that the "penalty" consists in reverting to short-run noncooperative behavior for some length of time, so a better term might be "noncooperative" phase.

If $(\tilde{\sigma}, \tilde{\tau})$ is to be an equilibrium pair of strategies in the supergame, with $\tilde{\sigma}$ a review strategy, then (by definition of equilibrium) $\tilde{\tau}$ must be optimal for the agent given $\tilde{\sigma}$. In particular, it is clear that it will be optimal for the agent to optimize myopically (see Section 3) against the reward function during each period of a penalty phase, since the agent cannot influence the principal's one-period strategies during the remainder of the penalty phase. Section 6 provides further information about the agent's optimal response to $\tilde{\sigma}$.

In addition, to further specify the agent's sequential strategy, one must describe what the agent would do if the principal did not pay the agent the (constant) reward \hat{w} at the end of some period during a review phase. In this case, the agent optimizes myopically during the remainder of the review phase and for M'

additional periods. (This will also be called a penalty phase.) The number M' is another parameter of the pair of review strategies.

Finally, for completeness one should specify what each player would do if he did not himself follow his own strategy at some previous period. (This specification, which may appear arcane to many readers, will be used to show that the equilibrium satisfies a criterion like perfectness; this point will be discussed in Section 8.4.) If the principal does not pay the constant reward $\hat{\omega}$ during some period in a review phase, then he will use the reward function ω^* during the remaining periods of what would have been the review phase plus additional M' periods. Otherwise, if either player has not followed his strategy at some period, as described above, then he will nevertheless continue to follow it in subsequent periods.

In order to give a precise description of the players' strategies, it is desirable to have some notation for the successive review and penalty phases, although this notation is somewhat complicated. (The reader who is satisfied with the preceding heuristic description of review strategies can proceed at this point to the next section without loss of continuity.) Define:

$$\begin{aligned}
 (4.1) \quad q &= \hat{a}R - B, \\
 D(1) &= \min \{t: \omega_t \neq \hat{\omega}\}, \\
 H^0(1) &= \text{the event } \{S_R < q\}, \\
 H^1(1) &= \text{the event } \{D(1) \leq R\}, \\
 N'(1) &= \min (D(1), R), \\
 N(1) &= \begin{cases} R + M', & \text{if } H^1(1), \\ R + M, & \text{if } H^0(1) \text{ but not } H^1(1), \\ R, & \text{otherwise.} \end{cases}
 \end{aligned}$$

Similarly, recursively define for $n \geq 1$:

$$\begin{aligned}
 (4.2) \quad D(n+1) &= \min \{t: t \geq N(n) \text{ and } \omega_t \neq \hat{\omega}\}, \\
 H^0(n+1) &= \text{the event } \{C_{N(n)+1} + \dots + C_{N(n)+R} < q\}, \\
 H^1(n+1) &= \text{the event } \{D(n+1) \leq N(n) + R\}, \\
 N'(n+1) &= \min \{D(n+1), N(n) + R\}, \\
 N(n+1) &= \begin{cases} N(n) + M', & \text{if } H^1(n), \\ N(n) + M, & \text{if } H^0(n) \text{ but not } H^1(n), \\ N(n) + R, & \text{otherwise.} \end{cases}
 \end{aligned}$$

The n th review phase consists of the periods $n(n-1)+1, \dots, N'(n)$, and the n th penalty phase, which may be empty, consists of the periods (if any) $N'(n)+1, \dots, N(n)$. By convention, $N(0) = 0$.

The principal's review strategy $\sigma(R, B, M, M')$ is defined by

$$(4.3) \quad \omega_t = \begin{cases} \hat{\omega} & \text{during review phases,} \\ \omega^* & \text{during penalty phases.} \end{cases}$$

The agent's review strategy $\tau(R, B, M, M')$ is defined to be a sequential strategy that is optimal against $\sigma(R, b, M, M')$ and, in particular, satisfies

$$(4.4) \quad A_t = \{\text{one-period-optimal action against } \omega_t \text{ during penalty phases}\} \\ = \alpha^*(\omega_t).$$

(Recall that condition (4.4) defines *myopic optimization*.) In addition, the agent uses the same strategy in each review period.

5. DISCOUNTED EXPECTED UTILITIES FOR REVIEW STRATEGIES

Let the *n*th epoch be the periods included in the *n*th review phase, together with the following penalty phase if there is one. (In the notation of the end of the previous section these are the periods from $N(n-1)+1$ through $N(n)$.) The beginnings of successive epochs are points of "renewal" or "regeneration" if review strategies are used, in the sense that events within different epochs are statistically independent. This fact facilitates the calculation of the players' discounted expected utilities.

As in Section 3, for any pair (σ, τ) of review strategies let (U_t) and (V_t) denote the corresponding sequences of realized utilities for the principal and the agent, respectively and let the respective (normalized) *discounted expected utilities* be denoted by

$$u(\gamma) = (1 - \gamma) \sum_{t=1}^{\infty} \gamma^{(-)} EU_t, \\ v(\delta) = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} EV_t,$$

where γ and δ are the discount factors for the principal and the agent, respectively. The normalization factors, $(1 - \gamma)$ and $(1 - \delta)$, are used to keep the discounted expected utilities bounded uniformly in γ and δ , where $0 \leq \gamma < 1$ and $0 \leq \delta < 1$. These are the natural normalization factors, since by Abel's Theorem (see [5, Theorem 55, p. 108]),

$$\lim_{\substack{\gamma \rightarrow 1 \\ \gamma < 1}} u(\gamma) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T EU_t,$$

provided the limit on the right-hand side exists.

During the first review phase, the principal's realized utility in period t is $C_t - \hat{w}$; during each period of the penalty phase (if any), the principal's expected utility is u^* . Let ϕ denote the probability of $H^0(1)$, i.e. the probability that the agent fails the first review. By the strong Markov property, the conditional discounted expected utility of the principal from period $N(1)+1$ on, given the history of the process during the first epoch, is $u(\gamma)$. Hence $u(\gamma)$ satisfies the

equation

$$(5.1) \quad u(\gamma) = (1 - \gamma) \sum_{t=1}^R \gamma^{t-1} (EC_t - \hat{w}) + \phi [\gamma^R (1 - \gamma^M) u^* + \gamma^{R+M} u(\gamma)] \\ + (1 - \phi) \gamma^R u(\gamma).$$

Hence, solving (5.1) for $u(\gamma)$, one gets

$$(5.2) \quad u(\gamma) = \frac{(1 - \gamma) \sum_{t=1}^R \gamma^{t-1} (EC_t - \hat{w}) + \phi \gamma^R (1 - \gamma^M) u^*}{1 - \phi \gamma^{R+M} - (1 - \phi) \gamma^R}.$$

Similarly, the agent's discounted expected utility, $v(\delta)$, satisfies the equation

$$v(\delta) = (1 - \delta) \sum_{t=1}^R \delta^{t-1} EV_t + \phi [\delta^R (1 - \delta^M) v^* + \delta^{R+M} v(\delta)] \\ + (1 - \phi) \delta^R v(\delta),$$

so that

$$(5.3) \quad v(\delta) = \frac{(1 - \delta) \sum_{t=1}^R \delta^{t-1} EV_t + \phi \delta^R (1 - \delta^M) v^*}{1 - \phi \delta^{R+M} - (1 - \phi) \delta^R}.$$

6. OPTIMAL REVIEW STRATEGIES FOR THE AGENT

In this section I derive some useful information about the agent's optimal response to a review strategy of the principal. In particular, I derive lower bounds on both players' expected utilities corresponding to such an optimal response by the agent. These preliminary results have an independent interest in the case in which the principal can make a binding commitment to follow an announced review strategy, although this case is not explicitly treated in the present paper (however see Section 9).

For any specification of R , B , and M , the agent will choose the strategy τ to maximize (5.3); call the resulting maximum $\tilde{v}(\delta)$. This optimization problem can be formulated as a standard dynamic program with a finite set of states,⁵ and the existence of a stationary optimal strategy can be demonstrated by standard techniques. However, we shall need here only a few properties of the optimal strategy.

Notice that the agent has only to decide what to do during the review phase. One strategy that is available to the agent is to use the "efficient" action \hat{a} throughout the review phase; call this the "good faith" strategy and denote it by

⁵ At date t , take the state of the system to be the history of the process during the current epoch. A stationary strategy would then use the same strategy within each epoch.

$\hat{\tau}$. From (5.3) one obtains the following formula for the discounted expected utility to the agent under $\hat{\tau}$, where $\hat{\phi}$ denotes the value of ϕ under $\hat{\tau}$:

$$(6.1) \quad \hat{v}(\delta) = \frac{(1 - \delta^R)\hat{v} + \hat{\phi}\delta^R(1 - \delta^M)v^*}{1 - \hat{\phi}\delta^{R+M} - (1 - \hat{\phi})\delta^R}.$$

This formula provides, therefore, a lower bound for the optimal value, $\tilde{v}(\delta)$. It will be useful to have in mind the limit of $\hat{v}(\delta)$ as δ approaches unity:

$$(6.2) \quad \lim_{\substack{\delta \rightarrow 1 \\ \delta < 1}} \hat{v}(\delta) = \frac{R\hat{v} + \hat{\phi}Mv^*}{R + \hat{\phi}M}.$$

Equation (6.2) has an easy interpretation. The expected length of the penalty phase is $\hat{\phi}M$ (if we make the convention that the penalty phase has length zero when the agent passes review), the expected length of an epoch is $(R + \hat{\phi}M)$, and (6.2) is a weighted average of \hat{v} and v^* , with the weights proportional to the expected lengths of the corresponding phases. If the expected length of the penalty phase can be made small, then $\hat{v}(\delta)$ will be close to \hat{v} when δ is close to unity. I shall show that this can be done by an appropriate choice of R , B , and M . This will imply that $\tilde{v}(\delta)$ can be made at least close to \hat{v} when δ is close to unity.

A precise statement of this result is embodied in Lemma 6.1 below. Lemma 6.2 gives a corresponding result for the principal's expected utility. I shall lead up to these lemmas with some preliminary work.

The next step in the study of review strategies is to derive an upper bound on the agent's discounted expected utility. From the convexity and monotonicity of the function Q it follows that there is a positive number K such that, for any strategy of the agent, any review period t , and any history H_{t-1}^C ,

$$E(V_t | H_{t-1}^C) - \hat{v} + K(A_t - \hat{a}) \leq 0,$$

or

$$(6.3) \quad E(V_t | H_{t-1}^C) \leq \hat{v} + K(\hat{a} - A_t).$$

It follows from this and (5.3) that

$$(6.4) \quad v(\delta) \leq \frac{(1 - \delta^R)\hat{v} + (1 - \delta)K \sum_{t=1}^R \delta^{t-1}(\hat{a} - EC_t) + \phi\delta^R(1 - \delta^M)v^*}{1 - \phi\delta^{R+M} - (1 - \phi)\delta^R}.$$

Define

$$(6.5) \quad f(x, y) \equiv \sum_{t=1}^y (1 - x^{t-1}) = y - \frac{1 - x^y}{1 - x}, \quad 0 \leq x < 1.$$

Since $|\hat{a} - EC_t| \leq 1$, one can verify that

$$(6.6) \quad \left| \sum_{t=1}^R \delta^{t-1}(\hat{a} - EC_t) - \sum_{t=1}^R (\hat{a} - EC_t) \right| \leq f(\delta, R),$$

so that

$$(6.7) \quad \sum_{t=1}^R \delta^{t-1}(\hat{a} - EC_t) \leq R\hat{a} - ES_R + f(\delta, R).$$

Since

$$\begin{aligned} \phi &= \text{Prob}(S_R < R\hat{a} - B), \\ S_R &\geq 0, \end{aligned}$$

it follows that

$$(6.8) \quad R\hat{a} - ES_R \leq \phi R\hat{a} + (1 - \phi)B.$$

Thus (6.4)-(6.8) imply that

$$(6.9a) \quad v(\delta) \leq v_0(\delta),$$

where

$$(6.9b) \quad v_0(\delta) \equiv \frac{(1 - \delta^R)\hat{v} + \phi\delta^R(1 - \delta^M)v^*}{(1 - \delta^R) + \phi\delta^R(1 - \delta^M)} + \frac{(1 - \delta)K[\phi R\hat{a} + (1 - \phi)B + f(\delta, R)]}{(1 - \delta^R) + \phi\delta^R(1 - \delta^M)}.$$

Compare this with (6.1), which can be rewritten as

$$(6.10) \quad \hat{v}(\delta) = \frac{(1 - \delta^R)\hat{v} + \hat{\phi}\delta^R(1 - \delta^M)v^*}{(1 - \delta^R) + \phi\delta^R(1 - \delta^M)}.$$

I shall now specify the following relations among R , B , and M :

$$(6.11) \quad B = \beta R^\rho, \quad \beta > 0, \quad \frac{1}{2} > \rho < 1.$$

$$(6.12) \quad M = \mu R, \quad \mu > 0.$$

Define

$$(6.13) \quad \hat{k} \equiv \text{Var}(C_t | A_t = \hat{a}) = \hat{a}(1 - \hat{a}).$$

By Chebychev's Inequality,

$$(6.14) \quad \hat{\phi} \leq \frac{R\hat{k}}{B^2} = \frac{\hat{k}}{\beta^2 R^{2\rho-1}},$$

so that $\hat{\phi}$ approaches zero as R increases without limit. Hence, from (6.10)-(6.14) one sees that $\hat{v}(\delta)$ approaches $\hat{v}(1)$ uniformly in R , as δ approaches 1 (for fixed μ , β , and ρ), and

$$(6.15) \quad \hat{v}(1) = \frac{\hat{v} + \hat{\phi}\mu v^*}{1 + \hat{\phi}\mu}.$$

This is equivalent to (6.2). In particular, for every $\varepsilon < 0$ there exist R_ε and δ_ε such that

$$(6.16) \quad \hat{v}(\delta) > \hat{v} - \varepsilon, \quad \text{for } R \geq R_\varepsilon \text{ and } \delta \leq \delta_\varepsilon.$$

If in (6.9b) one lets δ approach unity, for fixed R , one obtains

$$(6.17) \quad v_0(1) = \frac{\hat{v} + \phi\mu v^* + K[\phi\hat{a} + (1 - \phi)\beta R^{\rho-1}]}{1 + \phi\mu}.$$

Let η be a number, fixed throughout the remainder of the discussion, such that

$$(6.18a) \quad 0 < \eta < \hat{v} - v^*,$$

and fix μ so that

$$(6.18b) \quad \mu > \frac{K\hat{a}}{\hat{v} - v^* - \eta}.$$

With an elementary calculation one can verify that for any positive $\varepsilon' < \eta$,

$$(6.19) \quad \phi \geq \frac{K\beta R^{\rho-1} + \varepsilon'}{\mu(\hat{v} - v^* - \varepsilon') - K\hat{a} + K\beta R^{\rho-1}}$$

implies that

$$(6.20) \quad v_0(1) \leq \hat{v} - \varepsilon'.$$

With a slight change in notation, let $\tilde{v}(\delta, R)$ denote the maximum discounted expected utility of the agent, and let $\tilde{\phi}(\delta, R)$ be the corresponding probability of failing the review.

LEMMA 6.1: *For every positive $\varepsilon < 2\eta$, there exist R_ε and $\delta_\varepsilon < 1$ such that $\delta \geq \delta_\varepsilon$ implies that*

$$(6.21) \quad \tilde{v}(\delta, R_\varepsilon) > \hat{v} - \frac{\varepsilon}{4},$$

$$(6.22) \quad \hat{\phi}(\delta, R_\varepsilon) < \varepsilon.$$

PROOF. Let R_ε be a value of R such that $R_\varepsilon \geq R'_{\varepsilon/4}$ in (6.16) and the right-hand side of (6.19) does not exceed ε , with $\varepsilon' = \varepsilon/2$. Let δ_ε be a value of δ such that $\delta_\varepsilon \geq \delta'_{\varepsilon/4}$ in (6.16) and such that, for $R = R_\varepsilon$ and $\delta \geq \delta_\varepsilon$,

$$(6.23) \quad |v_0(\delta) - v_0(1)| \leq \frac{\varepsilon}{4}.$$

Inequality (6.21) in the conclusion of the lemma follows immediately from (6.16). On the other hand, if $\delta > \delta_\varepsilon$ and $\tilde{\phi}(\delta, R_\varepsilon) \geq \varepsilon$, then (6.19) and (6.20) would imply

$$v_0(1) \leq \hat{v} - \frac{\varepsilon}{2},$$

which with (6.23) would imply

$$(6.24) \quad v_0(\delta) \leq \hat{v} - \frac{\varepsilon}{4}.$$

But $\tilde{v}(\delta, R_\varepsilon) \leq v_0(\delta)$, so that (6.24) contradicts (6.21), which proves (6.22), and completes the proof of the lemma.

If the probability that the agent fails the review is small, and B is small relative to $R\hat{a}$, then the principal's utility will be close to the efficient level most of the time. This is formalized in the proof of the next lemma, which shows that review strategies yield the principal discounted expected utility close to \hat{u} if the review phase is long enough and if the discount factors of the principal and agent are high enough.

LEMMA 6.2 *Let $\tilde{u}(\gamma, \delta, R)$ denote the principal's discounted expected utility yielded by review strategies given R , if the discount factors of the principal and agent are γ and δ , respectively; then for every $\varepsilon > 0$ there is a $\gamma'_\varepsilon < 1$ such that if $R = R_\varepsilon$ and $\delta \geq \delta_\varepsilon$ (as in Lemma 6.1), and $\gamma \geq \gamma'_\varepsilon$, then*

$$\tilde{u}(\gamma, \delta, R_\varepsilon) \geq \frac{\hat{u} - \varepsilon\mu u^* - \varepsilon(\hat{a} + \beta R_\varepsilon^{\rho-1}) - \beta R_\varepsilon^{\rho-1}}{1 + \varepsilon\mu} - \varepsilon.$$

PROOF. First note that

$$\begin{aligned} EC_t - \hat{w} &= (\hat{a} - \tilde{w}) + (EC_t - \hat{a}) \\ &= \hat{u} + (EC_t - \hat{a}). \end{aligned}$$

Hence, from (5.2), and writing $u(\gamma)$ for $u(\gamma, \delta, R)$ and ϕ for $\tilde{\phi}(\delta, R)$,

$$u(\gamma) = \frac{(1 - \gamma^R)\hat{u} + \phi\gamma^R(1 - \gamma^M)u^* + (1 - \gamma) \sum_{t=1}^R \gamma^{t-1}(EC_t - \hat{a})}{(1 - \gamma^R) + \phi\gamma^R(1 - \gamma^M)}.$$

By an argument that parallels the one leading from (6.4) to (6.9b) one can show that

$$(6.25) \quad u(\gamma) \geq u_0(\gamma),$$

where

$$(6.25b) \quad u_0(\gamma) \equiv \frac{(1 - \gamma^R)\hat{u} + \phi\gamma^R(1 - \gamma^M)u^*}{(1 - \gamma^R) + \phi\gamma^R(1 - \gamma^M)} - \frac{(1 - \gamma)[\phi R\hat{a} + (1 - \phi)B + f(\gamma, R)]}{(1 - \gamma^R) + \phi\gamma^R(1 - \gamma^M)},$$

and that for fixed R and ϕ ,

$$(6.26) \quad \lim_{\gamma \rightarrow 1} u_0(\gamma) = u_0(1) \\ = \frac{\hat{u} + \phi\mu u^* - \phi\hat{a} - (1 - \phi)\beta R^{\rho-1}}{1 + \phi\mu}.$$

The conclusion of the lemma now follows from (6.26) and Lemma 6.1.

7. EQUILIBRIUM REVIEW STRATEGIES

In this section I give the main results on equilibria for the case of discounting, namely, that efficiency can be approached with equilibria of the supergame as the players' discount factors approach unity, and such equilibria can be attained with review strategies. These results are stated formally in Theorems 7.1 and 7.2. An immediate corollary is that for all discount factors above some critical values there are equilibria in review strategies that yield the principal and agent discounted expected utilities strictly greater than u^* and v^* , respectively.

Recall that a pair of review strategies, (σ, τ) , is characterized by the parameters R, B, M , and M' (see Section 3). The relationships among R, B , and M were specified by (6.11) and (6.12), which I repeat here:

$$B = \beta R^\rho,$$

$$M = \mu R,$$

where β, ρ , and μ are fixed parameters satisfying

$$\beta > 0, \quad \frac{1}{2} < \rho < 1,$$

and, from (6.18ab),

$$0 < \eta < \hat{v} - v^*,$$

$$\mu > \frac{K\hat{a}}{\hat{v} - v^* - \eta}.$$

Finally, the agent's review strategy is specified to be optimal against the principal's review strategy, and satisfy (4.4). Thus we can adequately denote the principal's review strategy by $\sigma(R, M')$ and the agent's review strategy by $\tau(\delta, R, M')$, which represents a natural change of notation from that of Section 3. As in Section 4, let $\tilde{u}(\gamma, \delta, R)$ and $\tilde{v}(\delta, R)$ denote the corresponding discounted expected utilities of the principal and agent, respectively.

Since it is part of the definition of $\tau(\delta, R, M')$ that it be optimal against $\sigma(R, M')$, in order to show that a particular pair is an equilibrium it is sufficient to show that $\sigma(R, M')$ is optimal against $\tau(\delta, R, M')$. Sufficient conditions for this will be given in this section.

By (4.1) and (4.2), if the principal departs from the constant reward $\hat{\omega} = (\hat{w}, \hat{w})$ during a review phase, then this action—which I shall call *stopping*—initiates a

penalty phase. By (4.4), the agent optimizes myopically against ω_t during a penalty phase, so that it is optimal for the principal to set $\omega_t = \omega^*$ during a penalty phase; hence the second line of (4.3). Hence to show that $\sigma(R, M')$ is optimal against $\tau(\delta, R, M')$ it is sufficient to show that the principal should not stop during a review phase.

For the time being, let $u(\gamma)$ denote the principal's maximum discounted expected utility against $\tau(\delta, R, M')$, and $H_t = H_t^C$. If the principal stops just after period $t < R$ during the first review phase, and then follows an optimal policy thereafter, his discounted conditional expected utility after period t , given H_t , will be

$$(7.1) \quad (1 - \gamma^{R-t+M'})u^* + \gamma^{R-t+M'}u(\gamma).$$

If the principal continues without stopping from t to the end of the first review phase and follows an optimal policy thereafter, his discounted conditional expected utility after period t , given H_t , will be

$$(7.2) \quad E \left\{ (1 - \gamma) \left(\sum_{n=1}^{R-t} \gamma^{n-1} U_{t+n} \right) + \gamma^{R-t} (1 - \gamma^{M(H_R)})u^* + \gamma^{R-t+M(H_R)}u(\gamma) \mid H_t \right\},$$

where

$$M(H_R) = \begin{cases} 0, & \text{if } S_R \geq R\hat{a} - B, \\ M, & \text{if } S_R < R\hat{a} - B. \end{cases}$$

By the optimality principle of dynamic programming, $u(\gamma)$ is at least as large as the maximum of (7.1) and (7.2), so a sufficient condition for $\sigma(R, M')$ to be optimal is that (7.1) be less than (7.2) for each $t = 0, \dots, R - 1$.

A feasible strategy for the principal is to use $\omega_t = \omega^*$ for all t ; hence

$$u(\gamma) \geq u^*.$$

Also, in (7.2),

$$(7.3) \quad U_{t+n} \geq -\hat{w} \quad (n = 1, \dots, R - t),$$

so that replacing $M(H_R)$ by M cannot increase (7.2). Hence (7.2) is at least as large as

$$(7.4) \quad -(1 - \gamma^{R-1})\hat{w} + \gamma^{R-t}(1 - \gamma^M)u^* + \gamma^{R-t+M}u(\gamma).$$

Hence a sufficient condition for $\sigma(R, M')$ to be an equilibrium is that (7.1) be strictly less than (7.4), or equivalently,

$$\gamma^{R-t}(\gamma^M - \gamma^{M'})(u(\gamma) - u^*) > (1 - \gamma^{R-t})(\hat{w} + u^*), \text{ for } t = 0, \dots, R - 1.$$

Since $u(\gamma) \geq u^*$, this is equivalent to

$$(7.5a) \quad \gamma^R(\gamma^M - \gamma^{M'})(u(\gamma) - u^*) > (1 - \gamma^R)(\hat{w} + u^*),$$

$$(7.5b) \quad u(\gamma) > u^*.$$

Let us first investigate the second line of (7.5). For this it is sufficient that $u_0(\gamma) > u^*$, where $u_0(\gamma)$ is defined in (6.25b); see the proof of Lemma 6.2. By (6.26), for given δ and R , this last inequality is satisfied for γ sufficiently large, provided

$$(7.6) \quad \phi(\delta, R)\hat{a} + [1 - \phi(\delta, R)]\beta R^{\rho-1} < \hat{u} - u^*.$$

This last inequality is satisfied for R sufficiently large, and δ sufficiently close to 1 given R , as was shown in Lemma 6.1.

Let us now examine the first line of (7.5), which (if the second line holds) is implied by

$$(7.7) \quad \gamma^{M'} < \gamma^{\mu R} - \frac{(1 - \gamma^R)Z(\gamma)}{\gamma^R},$$

where

$$(7.8) \quad Z(\gamma) \equiv \frac{\hat{w} + u^*}{u_0(\gamma) - u^*}.$$

(Keep in mind that $u_0(\gamma)$ and $Z(\gamma)$ also depend on δ and R .) Inequality (7.7) is satisfied for sufficiently large M' provided that the right-hand side of (7.7) is positive, or that

$$(7.9) \quad \frac{\gamma^{(1+\mu)R}}{1 - \gamma^R} > Z(\gamma).$$

For every R and δ satisfying (7.6), this last inequality is in turn satisfied for γ sufficiently close to 1, since, by (6.26), the limit of $Z(\gamma)$ exists and is positive as γ tends to 1.

To summarize the analysis of (7.5), for each δ and R let $\Gamma(\delta, R)$ be the infimum of all $\gamma' < 1$ such that for all $\gamma \geq \gamma'$, both $u_0(\gamma) > u^*$ (see (6.25)) and inequality (7.9) is satisfied; if no such γ' exists assign the value infinity to $\Gamma(\delta, R)$. In particular, $\Gamma(\delta, R)$ is finite (and less than 1) for R sufficiently large, and δ sufficiently close to 1 given R . In this case, if $\gamma > \Gamma(\delta, R)$ there is a number M' satisfying (7.7); call the infimum of such numbers $MM'(\delta, \gamma, R)$. Thus we have demonstrated the first main result.

THEOREM 7.1: *If δ and R satisfy (7.6), $\gamma > \Gamma(\delta, R)$, and $M' > MM'(\delta, \gamma, R)$, then the review strategies $\sigma(R, M')$ and $\tau(\delta, R, M')$ form an equilibrium of the supgame.*

The second main result concerns the existence of equilibria in review strategies that are arbitrarily close to efficient for discount factors sufficiently close to unity. Note that (7.7) is only needed to assure that the principal's review strategy is optimal against the agent's review strategy, but the discounted expected utilities of the players do not depend on M' .

THEOREM 7.2: *For every positive $\varepsilon < 2\eta$ there exist R_ε , δ_ε , and γ_ε such that, for every $\delta > \delta_\varepsilon$, $\gamma > \gamma_\varepsilon$, and $M' > MM'(\delta, \gamma, R_\varepsilon)$, the review strategies $\sigma(R_\varepsilon, M')$ and $\tau(\delta, R_\varepsilon, M')$ form an equilibrium of the supergame and yield discounted expected utilities*

$$\tilde{u}(\gamma, \delta, R_\varepsilon) \geq \hat{u} - \varepsilon,$$

$$\tilde{v}(\delta, R_\varepsilon) \geq \hat{v} - \varepsilon;$$

furthermore, $\tilde{\phi}(\delta, R_\varepsilon)$, the probability that the agent fails a review, does not exceed ε .

PROOF: The Theorem follows immediately from Lemmas 6.1 and 6.2, and Theorem 7.1.

COROLLARY: *For all discount factors above some critical values there are equilibria in review strategies that yield the principal and agent discounted expected utilities strictly greater than u^* and v^* , respectively.*

It is of some interest to note that, for equilibria that are close to efficient with discount factors close to 1, it is sufficient to take M' proportional to R . To see this, write the first line of (7.5) as

$$(7.10) \quad \gamma^{\mu R} - \gamma^{M'} > \left(\frac{1 - \gamma^R}{\gamma^R}\right) \left(\frac{\hat{w} + u^*}{u(\gamma) - m^*}\right).$$

Under the hypothesis of Theorem 7.2, with $\varepsilon < (\hat{u} - u^*)/2$,

$$u(\gamma) - u^* = \tilde{u}(\delta, \gamma, R_\varepsilon) - u^* > \frac{\hat{u} - u^*}{2}.$$

Hence (7.10) will be satisfied if

$$(7.11) \quad \gamma^{\mu R_\varepsilon} - \gamma^{M'} > 2 \left(\frac{1 - \gamma^{R_\varepsilon}}{\gamma^{R_\varepsilon}}\right) \left(\frac{\hat{w} + u^*}{\hat{u} - u^*}\right).$$

Letting γ tend to 1 in (7.11), one sees that (7.11) is satisfied for γ sufficiently close to 1 if

$$(7.12) \quad M' - \mu R_\varepsilon = \kappa R_\varepsilon$$

where κ is a number such that

$$\kappa > 2 \left(\frac{\hat{w} + u^*}{\hat{u} - u^*}\right).$$

One can rewrite (7.12) as

$$M' = (\mu + \kappa) R_\varepsilon.$$

8. GENERALIZATIONS AND FURTHER REMARKS

8.1. *A More General Model*

Theorems 7.1, 7.2, and 8.1 can be generalized, with relatively minor modifications of the proofs, to cases in which (i) the consequences C_i are bounded random variables, rather than variables taking on only the values 0 and 1, (ii) the agent bases his choice of action in each period on some (possibly noisy) information about the environment, and (iii) the principal is averse to risk. The second extension is, of course, important if one wants to apply the results to situations in which the principal employs the agent precisely because the agent has “superior” information in each particular period. The players’ utility functions need not be additively separable.

This more general model contains as special cases all of the one-period principal-agent models in the literature of which I am aware, except those in which the principal can base the agent’s reward on partial information about the agent’s action or the random environment. I believe that the analysis could be extended to cover such cases as well, but I have not yet attempted to do this.

On the other hand, the assumption that the successive random environments are independent and identically distributed plays an important role in the present analysis of the supergame, which could not be carried through in general if the agent observed a random event at the beginning of the supergame, and this random event remained payoff-relevant throughout the supergame.

A precise description of the more general model is given in the Appendix.

8.2. *Review-Strategy Equilibria Are Not Second-Best*

I have not attempted here to characterize those supergame equilibria that are most efficient *within* the set of equilibria that are attainable with fixed discount factors. It does not appear that the equilibria described in Section 7 are in fact efficient in this second-best sense. For these equilibria, the agent has a quota of successes to fulfill in each review phase, namely $q = R\hat{a} - B$. Once the agent has fulfilled his quota, he has no incentive to put in any effort during the remainder of the review phase. On the other hand, if the agent is too far from the quota at some point in the review phase (i.e., $S_t + R - t < q$), then he also has no incentive to put in any effort during the remainder of the review phase.

To induce the agent to over-fulfill his quota, the principal could share with him the proceeds from successes over and above the quota. Thus, if $S_t = q$, then the agent might receive $W_n = \hat{w} + fC_n$, for $n = t + 1, \dots, R$, where f is a number between 0 and 1. To induce the agent not to “give up” after unfavorable experience, the length of the penalty phase could be made an increasing function of the amount by which the agent falls short of the quota, e.g., $M = m_0 + m_1(q - S_R)$, where m_0 and m_1 are positive parameters. These modifications would appear to increase efficiency, but it is not known to me whether they are sufficient to actually attain second-best efficiency.

8.3. *Precommitment by the Principal*

In the one-period game studied here, the principal moves first by announcing a reward function, which he is committed to use at the end of the period. In the corresponding supergame, these commitments last one period at a time, and the principal cannot bind himself in advance to a sequence of two or more reward functions. A variation of this game would have the principal move second, and announce (and pay) the reward after observing the consequence of the agent's action. In the one-period game it would, of course, be optimal for the principal to pay the agent nothing, but in the supergame the situation would be analogous to the one studied here. The important aspect of the problem studied here, which is common to the two variations, is that neither player can make binding commitments to follow any particular sequence of one-period strategies.

One could also consider a variation of the supergame in which the principal can enter into a binding commitment to follow his pre-announced supergame strategy. In this case the principal would choose a supergame strategy that is best for him, given that the agent would optimize against it. Since the review-strategy equilibria studied here are not second-best efficient, the optimal supergame strategy for the principal would not be a review strategy. On the other hand, Lemmas 6.1 and 6.2 would still be valid for the agent's optimal responses to review strategies of the principal, and so optimal equilibria of the supergame with precommitment by the principal would also be approximately first-best efficient for discount factors close to unity.

8.4. *Credible Equilibria*

It is known that one must typically impose some further restriction on Nash equilibria of sequential games in order to assure that they will be "credible," i.e., that any threats that are implicit in the players' strategies are credible. Space limitations allow me only a few remarks here for the already initiated reader. For recent discussions, see [8, 9].

One such restriction is sub-game perfectness. The criterion of subgame-perfectness in the supergame has not been formally invoked here because, strictly speaking, there are typically no proper subgames after the principal's first move. This is implied by the fact that the principal can never observe the agent's actions directly, and if A_i is never 0 nor 1 then all finite sequences of consequences have positive probability. Nevertheless, both the agent and the principal can immediately detect any departure from a pre-announced strategy of the principal. An alternative concept of "credible equilibrium" that seems useful in this situation, called *sequential equilibrium*, has been proposed by Kreps and Wilson [9]. In fact, one can show that the review-strategy equilibria constructed in Sections 4-7 satisfy their definition.

9. BIBLIOGRAPHIC NOTES

In the case of a one-period game, Groves [3, 4] studied a more general decentralized decision-making situation with several risk-neutral agents and an organizer

(principal). Using the assumption of risk-neutrality of the agents, he constructed reward functions that induce efficient equilibria. Early analyses of short-run principal-agent relationships (with one agent) were provided by Spence and Zeckhauser [22], Ross [18], and Mirrlees [12]. For later work see Hurwicz and Shapiro [7], Holmström [6], Shavell [23], Grossman and Hart [2], and the many references cited in those papers. An early forerunner of this literature was Simon's 1953 paper on the employment relation [24].

The properties of the set of supergame equilibrium payoff vectors are well-understood for the case of repeated games under certainty in which (i) the players can monitor the actions of the other players after each one-period game ("perfect monitoring") and (ii) the players do not discount their future utilities ("no discounting"). In this case, the set of supergame equilibrium payoff vectors is the same as the set of feasible, individually rational payoff vectors in the one-period game. (This is the so-called Folk Theorem.) The same conclusion can be derived for perfect equilibria of the supergame; this deeper result is due to Aumann and Shapley (unpublished) and to A. Rubinstein; see [20] for references and a related result. The case of perfect monitoring with discounting has not been so well explored; see Kurz [10] and Lockwood [11].

Unfortunately, the condition of perfect monitoring is ruled out by the informational structure of the principal-agent situation (without additional cost, as noted above). I am not aware of any previous treatment of the infinite principal-agent supergame with discounting. For the case of no discounting, Radner [13] has shown that for sufficiently long but *finite* principal-agent supergames one can sustain approximate efficiency by means of *approximate* equilibria. Particular infinite principal-agent supergames have been analyzed in a similar spirit, again for the no-discounting case, by Rubinstein [19] and by Rubinstein and Yaari [21]. However, the strategies used for the no-discounting case in the above-cited literature do not appear to be applicable to the case of discounting.

The use of review strategies permits a more-or-less unified treatment of the two cases; for the no-discounting case one can construct equilibria with review strategies in which the review periods are progressively longer. Furthermore, the use of review strategies permits a more elementary mathematical analysis. For a full treatment of the no-discounting case, see Radner [16].

As mentioned in Section 1.5, the continuity-like property of the equilibrium payoff correspondence at efficient payoff-pairs, as the players' discount factors approach unity, is apparently not a general property of repeated games with imperfect monitoring. An example is provided by repeated partnership games (Radner [14]), for which there are fully efficient supergame equilibria without discounting, whereas in the case of discounting the set of supergame equilibrium payoff vectors may be bounded away from efficiency *uniformly in the players' discount factors*, provided these are strictly less than unity (see Radner, Myerson, and Maskin [17]).

AT & T Bell Laboratories, Murray Hill, New Jersey.

APPENDIX

In this Appendix I sketch the general model described informally in Section 8.1, and indicate the corresponding changes in the proofs of the main results of the paper.

THE ONE-PERIOD GAME

The principal moves first, announcing a nonnegative reward function, ω . The environment is a random variable, X . The agent observes a random variable, Y , and then chooses in action, A (the agent's move). The consequence, also a random variable, is

$$(A.1) \quad C = C(A, X).$$

The principal pays the agent

$$(A.2) \quad W = \omega(C).$$

The resulting (stochastic) utilities to the principal and agent, respectively, are

$$(A.3) \quad U = U(C, W),$$

$$(A.4) \quad V = V(A, W).$$

The principal's strategy is the same as his move, ω . The agent's strategy is a mapping, α , that determines his action as a function of the announced reward function and his observation:

$$(A.5) \quad A = \alpha(\omega, Y).$$

For given strategies ω and α , equations (A.1)-(A.5) determine the respected expected utilities for the principal and agent:

$$(A.6) \quad u = EU, \quad v = EV.$$

All of the above data, including the joint probability distribution of X and Y , are common knowledge. Assume that the utility functions, U and V , are bounded. To avoid technical issues of measure theory, it is convenient to assume that X , Y , and C have only finitely many values. (This assumption could be replaced by various regularity conditions; I omit the details.)

I assume further that the following situation obtains. There is a (perfect) Nash equilibrium, (ω^*, α^*) , of the one-period game, yielding an expected-utility pair (u^*, v^*) ; in particular, for each ω and Y the action $\alpha^*(\omega, Y)$ maximizes the agent's conditional expected utility. Since the principal's utility is bounded, one can without loss of generality make the convention that $U_i \geq 0$. Assume that $u^* > 0$. Let $(\hat{\omega}, \hat{\alpha})$ be efficient (Pareto-optimal), with corresponding expected-utility pair (\hat{u}, \hat{v}) , such that $\hat{u} > u^*$ and $\hat{v} > v^*$. Finally, assume that there is a positive number K such that, for any strategy α of the agent, if the principal uses $\hat{\omega}$ then the corresponding expected utility of the agent and expected consequence satisfy

$$(A.7) \quad (v - \hat{v}) + K(u - \hat{u}) \leq 0.$$

This assumption states that the set of feasible expected-utility pairs (u, v) is supported at (\hat{u}, \hat{v}) by a line of finite and strictly negative slope, given that the principal uses $\hat{\omega}$.

THE REPEATED GAME

The repeated game is constructed from the one-period game just as in Section 3, except that to the agent's available information at date t is added the history of observations, $H_t^Y = (Y_1, \dots, Y_t)$, through period t ; cf. (3.1b). The successive pairs (X_t, Y_t) are independent and identically distributed. Define

$$G_t \equiv U \left[C_t, \hat{\omega}(C_t) \right],$$

and note that, for every history H_{t-1}^C ,

$$E(G_t | H_{t-1}^C, \alpha_t = \hat{\alpha}) = \hat{u}.$$

Also, note that G_t is observable by both players. Review strategies can now be defined as in Section 4, except that $S_R = G_1 + \dots + G_R$, and the criterion for passing review is $S_R \geq R\hat{u} - B$. Thus, during the review phase, the principal gauges the agent's performance directly by his own utility, G_t , rather than by the consequence, C_t .

From (A.7) it follows that for any strategy of the agent and any period t during a review phase,

$$(A.8) \quad E(V_t | H_{t-1}^C) \leq \hat{v} + K[\hat{u} - E(G_t | H_{t-1}^C)],$$

which is the analogue of (6.3). Also, note that in (5.2) and elsewhere the principal's expected utility during a review period becomes EG_t instead of $EC_t - \hat{w}$. One can replace the inequality preceding (6.6) by $|\hat{u} - EG_t| \leq L$. Similarly, (7.3) can be replaced by $U_{t+n} \geq 0$, and correspondingly in (7.5b), (7.8), (7.10), and elsewhere, $\hat{w} + u^*$ can be replaced by u^* , which is positive.

With these changes and corresponding modifications elsewhere, the arguments of Sections 5-7 can be carried out in a straightforward manner.

REFERENCES

- [1] CHUNG, K. L.: *A Course in Probability Theory*, Second Edition. New York: Academic Press, 1974.
- [2] GROSSMAN, S. J., AND O. D. HART: "An Analysis of the Principal-Agent Problem," *Econometrica*, 51(1983), 7-45.
- [3] GROVES, T.: "The Allocation of Resources under Uncertainty: the Informational and Incentive Roles of Prices and Demands in a Team," unpublished Ph.D. dissertation, Department of Economics, University of California, Berkeley, 1969.
- [4] ——— "Incentives in Teams," *Econometrica*, 41(1973), 617-631.
- [5] HARDY, G. H.: *Divergent Series*. Oxford: Clarendon Press, 1949.
- [6] HOLMSTRÖM, B: "Moral Hazard and Observability," *Bell Journal of Economics* 10(1979), 74-91.
- [7] HURWICZ, L., AND L. SHAPIRO: "Incentive Structures Maximizing Residual Gain under Incomplete Information," *Bell Journal of Economics*, 9(1978), 180-191.
- [8] KOHLBERG, E., AND J. F. MERTENS: "On the Strategic Stability of Equilibria," CORE Discussion Paper 8248, Université Catholique de Louvain, Louvain-la-Neuve, November, 1982 (unpublished).
- [9] KREPS, D. M., AND R. B. WILSON: "Sequential Equilibria," *Econometrica*, 50(1982), 863-894.
- [10] KURZ, M.: "Altruism as an Outcome of Social Interaction," *American Economic Review*, Papers and Proceedings, 68(1978), 216-222.
- [11] LOCKWOOD, B.: "Perfect Equilibria in Repeated Games with Discounting," Department of Applied Economics, University of Cambridge, 1983 (unpublished).
- [12] MIRLEES, J.: "Notes on Welfare Economics, Information, and Uncertainty," *Essays on Economic Behavior under Uncertainty*, ed. by M. D. Balch, D. McFadden, and Wu. Amsterdam: North-Holland, 1974.
- [13] RADNER, R.: "Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship," *Econometrica*, 49(1981), 1127-1148.
- [14] ——— "Optimal Equilibria in a Class of Repeated Games with Imperfect Monitoring," Bell Laboratories, Murray Hill, N. J., 1981, to appear in the *Review of Economic Studies*.
- [15] ——— "Decentralization and Incentives," to appear in *Information, Incentives, and Economic Mechanisms*, ed. by T. Groves, R. Radner, and S. Reiter. Minneapolis: University of Minnesota Press, 1984.
- [16] ——— "Repeated Principal-Agent Games with Discounting," Bell Laboratories Discussion Paper, Murray Hill, N.J., May, 1981.
- [17] RADNER, R., R. MEYERSON, AND E. MASKIN: "An Example of a Repeated Partnership Game with Discounting and with Uniformly Inefficient Equilibria," Bell Laboratories, Murray Hill, N.J., 1983, to appear in the *Review of Economic Studies*.
- [18] ROSS, S.: "The Economic Theory of Agency: the Principal's Problem," *American Economic Review*, 63(1973), 134-139.
- [19] RUBINSTEIN, A.: "Offenses that May Have Been Committed by Accident—an Optimal Policy of Retribution," in *Applied Game Theory*, ed. by S. J. Brams, A. Schotter, and G. Schrödäuer. Würzburg: Physica-Verlag, 1979.
- [20] ——— "Equilibrium in Supergames with the Overtaking Criterion," *Journal of Economic Theory*, 21(1979), 1-9.

- [21] RUBINSTEIN, A., AND M. E. YAARI: "Repeated Insurance Contracts and Moral Hazard," *Journal of Economic Theory*, 30(1983), 74-97.
- [22] SPENCE, M., AND R. ZECKHAUSER: "Insurance, Information, and Individual Action," *American Economic Review*, 61(1971), 380-387.
- [23] SHAVELL, S.: "Risk-sharing and Incentives in the Principal and Agent Relationship," *Bell Journal of Economics*, 10(1979), 55-73.
- [24] SIMON, H.: "A Formal Theory of the Employment Relation," *Econometrica*, 19(1953), 293-305.